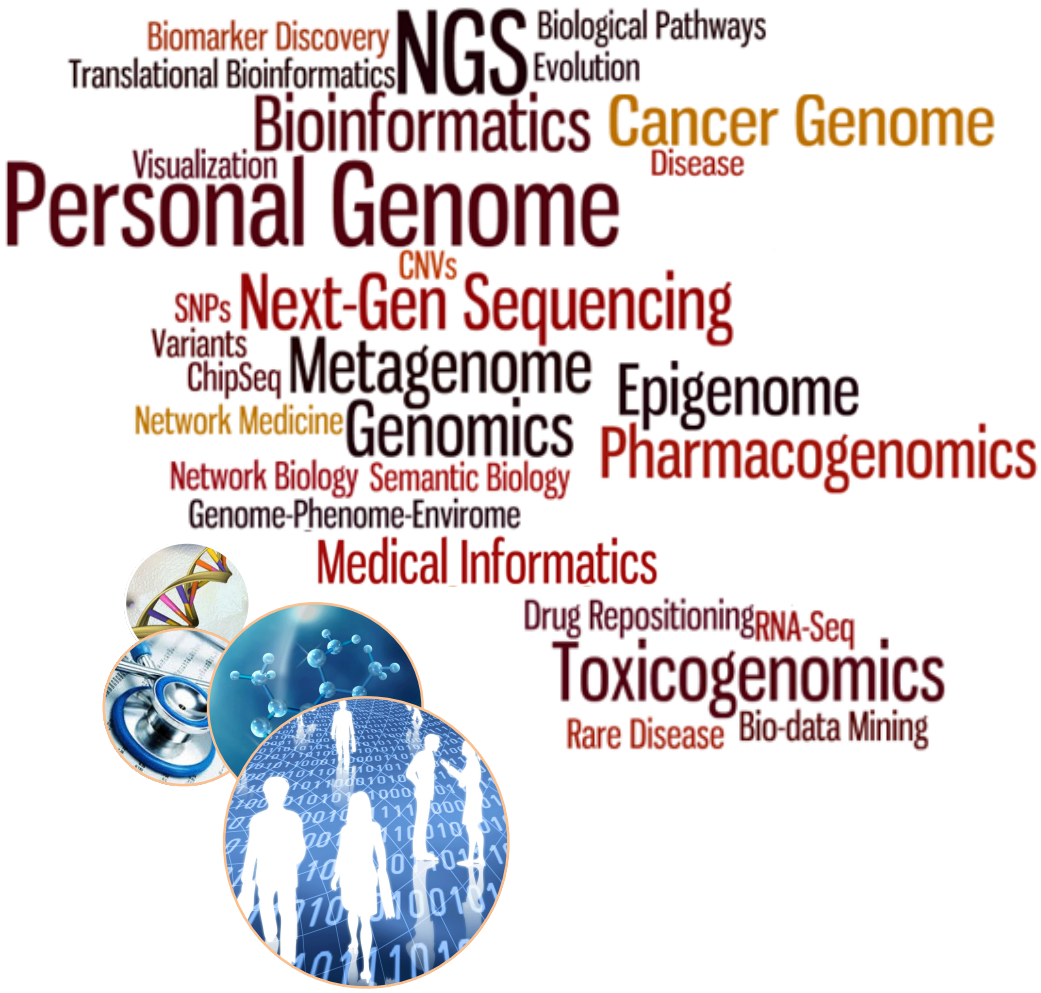
Translational Bioinformatics



TIME and | 15th – 17th , October, 2016
LOCATION | Hyatt Regency Jeju, Jeju Island, Korea
Conference | http://www.snubi.org/TBC2016/

Korean Society of Bioinformatics and Systems Biology
The Korean Society of Medical Informatics
Systems Biomedical Informatics Research Center
TBC 2016 Organizing Committee

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 and Friends,

Along with the Organizing Committee, I am delighted to welcome you to attend the 6th annual Translational Bioinformatics Conference (TBC 2016) back in Jeju Island in three years following the TBC 2013. TBC has provided for the last six consecutive years a general forum for disseminating the latest research in genomics, bioinformatics, translational research, and biomedical informatics.

As you may agree, 'Translational Bioinformatics' is even more appreciated than ever throughout most of the biomedical communities. We all have closely collaborated for promoting the field of translational bioinformatics. We have decided that the seventh annual TBC 2017 will be held out of the East Asian region either in Los Angeles, USA or in Goa, India. We welcome proposals for better TBCs for the following years.

Thanks to the invited speakers and presenters from all around the world to this conference, who are shaping the future of translational bioinformatics and genomics, I am sure that you will find an exciting atmosphere at TBC 2016. All participants are the ones who shape the future of translational bioinformatics. We are the one who makes the future of translational bioinformatics happen at TBC 2016.

I wish all participants of the conference have pleasant and memorable experience. Please enjoy TBC 2016 and the beautiful weather of Jeju Island again.

With my best regards,

Ju Han Kim, M.D., Ph.D., M.S.

Chair, TBC 2016 Organizing Committee

Organizing Committee Members

Ju Han Kim, M.D., Ph.D. (Korea)
Professor and Chair, Div. of Biomedical Informatics
Director, Systems Biomedical Informatics Research Center
Seoul National University College of Medicine

Atul Butte, M.D., Ph.D. (U.S.A.)
Director of the Institute of Computational Health Sciences (ICHS)
University of California, San Francisco

Luonan Chen, Ph.D. (China)
Key Laboratory of Systems Biology,
Shanghai Institute for Biological Sciences, China

Indira Ghosh, Ph.D. (India)
Dean and Professor, School of Informatics Technology,
Jawaharlal Nehru University, New Delhi, India

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

Yves A. Lussier, M.D. (U.S.A.)
Prof. of Medicine, Associate Vice President of Health Sciences
University of Arizona Attended Columbia University

Lucila Ohno-Machado, M.D., Ph.D. (U.S.A.)
Founding Chief, Division of Biomedical Informatics, UC San Diego
Director, Biomedical Research Informatics for Global Health Program

Marylyn DeRiggi Ritchie, Ph.D. (U.S.A.)
Director, Biomedical and Translational Informatics
Geisinger Health System

Tomohiro Sawa, M.D., Ph.D. (Japan)
Chief Information Officer, Headquarters, Teikyo University
Dept. of Anesthesiology, Teikyo University

Sangsoo Kim, Ph.D. (Korea)
Professor & Director, School of Systems Biomedical Sciences,
Soongsil University

Youngju Kim, Ph.D. (Korea)
Principal Researcher, Genome Resource Center,
Korea Research Institute of Bioscience & Biotechnology (KRIBB)

Kiejung Park, Ph.D. (Korea)
Director, Div. of Bio-Medical Informatics
National Institute of Health, Korea

Hyunjung Shin, Ph.D. (Korea)
Professor, Ajou University Datamining Lab,
Dept. of Industrial and Information Systems Engineering

Sanghyuk Lee, Ph.D. (Korea)
Director, Korean Bioinformation Center
Professor, Dept. of Life Sciences, Ewha Womans University
Director, Ewha Research Center for Systems Biology

Hojin Choi, Ph.D. (Korea)
Professor, Dept. of Computer Science
Korea Advanced Institute of Science and Technology (KAIST)

Pre-Congress Tutorials

Saturday, Oct. 15, 2016

Tutorial I. Translational Bioinformatics

- **14:00-16:00: A Protein-Domain Approach for Analysis of Disease Mutations.**

  Maricel G. Kann (U. of Marylnad Baltomore County)

- **16:00-18:00: Computational methods for precision medicine and single subject studies with genomes and transcritomes.**

  Yves A. Lussier & Colleen Kenost (U. of Arizona)

Tutorial II. Machine Learning Methods for Translational Bioinformatics

- **14:00-18:00: Machine Learning in Bioinformatics and Translational Disease Network Analysis**

  Helen Hyunjung Shin (Ajou University)

Program at a Glance

Day 1, Sunday, Oct. 16, 2016

	Regency Ballroom	Terrace Ballroom
08:00-09:30	Registration	
09:30-10:00	**Opening TBC 2016**	
10:00-10:50	**Keynote I:** Data storage and sharing in SCRUM-Japan: a nation-wide cancer genome screening project for drug development **Katsuya Tsuchihara**, National Cancer Center Japan	
Session	**S1. Multi-Omic Application** Chair: Kyung-Ah Sohn(Ajou Univ.)	**S2. Disease Genomics** Chair: Jung Kyoon Choi (KAIST)
11:00-11:25	**S1-1** N-of-1-pathways MixEnrich: advancing precision medicine via single-subject analysis in discovering dynamic changes of transcriptomes ***Qike Li***	**S2-1** Knowledge-driven binning approach for rare variant association analysis: Application to neuroimaging biomarkers in Alzheimer's disease ***Dokyoon Kim***
11:25-11:50	**S1-2** An Inference Method from Multi-Layered Structure of Omics ***Myungjun Kim***	**S2-2** Genotype Based Disease Similarity Matrix from Uniqueness of Shared Genes ***Hui Lu***
11:50-12:15	**S1-3** Identification of interactions between miRNA and DNA methylation associated with gene expression as potential prognostic markers in bladder cancer ***Manu Shivakumar***	**S2-3** Association analysis of rare variants near the APOE region with CSF and neuroimaging biomarkers of Alzheimer's disease ***Kwangsik Nho***
12:15-13:10	Lunch	
13:10-14:00	**Keynote II:** Significant Pattern Mining for Biomedical Applications **Koji Tsuda**, U. of Tokyo	
Session	**S3. Cancer Bioinformatics** Chair: Dongsup Kim (KAIST)	**S4. Bio/Medical Data Mining** Chair: Nigam Shah (Stanford Univ.)
14:00-14:25	**S3-1** Prediction of Recurrent Regulatory Mutations in Noncoding Cancer Genomes ***Jung Kyoon Choi***	**S4-1** Disease Causality Extraction based on Lexical Semantics and Clause Frequency from Biomedical Literature ***Dong-Gi Lee***
14:25-14:50	**S3-2** Identifying subtype-specific gene expressions explained by DNA methylation patterns in breast cancer ***Garam Lee***	**S4-2** ICU Event Prediction by integrating Sequential Patterns as Classification Features ***Shameek Ghosh***
14:50-15:15	**S3-3** Racial differences of intron retention and DNA methylation in breast cancer subtypes ***Younghee Lee***	**S4-3** Quad-phased Data Mining Modeling for Dementia Diagnosis ***Sunjoo Bang***
15:15-15:40	**S3-4** Identification of clinically relevant genes from mRNA and splicing changes of skin cutaneous melanoma ***Ji Yeon Park***	**S4-4** Medical Concepts Embedding ***Ting Chen***
15:40-16:10	Coffee Break	
16:10-17:00	**Keynote III:** Single cell genome analysis for precision cancer medicine **Woong Yang Park**, Samsung Medical Center	
17:00-18:00	**Keynote IV:** Evolution vs Disease: From Big Data and Text Mining to Personalized Genomics **Olivier Lichtarge**, Baylor College of Medicine	
18:00-	Dinner Beach Party	

Program at a Glance

Day 2, Monday, Oct. 17, 2016

	Regency Ballroom	Terrace Ballroom
09:00-10:00	**Keynote V:** Deploying Genomics and Immunology for Risk Assessment and Prevention **Olufunmilayo I. Olopade**, U. of Chicago	
Session	**H1. Highlight Research Tracks** Chair: Hyun Goo Woo (Ajou Univ.)	**S5. Network Biology and Medicine** Chair: Younghee Lee (Univ. of Utah)
10:00-10:25	**H1-1** Reproducibility in large in vitro drug screening: where do we stand? ***Benjamin Haibe-Kains***	**S5-1** Integrative Information Theoretic Network Analysis for GWAS of Aspirin Exacerbated Respiratory Disease in Korean Population ***Kyung-Ah Sohn***
10:25-10:50	**H1-2** ISOexpresso: a web-based platform for isoform-level expression analysis in human cancer ***Sangwoo Kim***	**S5-2** Taking promoters out of enhancers in sequence based predictions of tissue-specific mammalian enhancers ***Bartek Wilczynski***
10:50-11:15	**H1-3** Uncovering synthetic lethal interactions for therapeutic targets and predictive markers in lung adenocarcinoma ***Grace S. Shieh***	**S5-3** Modeling Long-Term Human Activeness Using Recurrent Neural Networks for Biometric Data ***Ho-Jin Choi***
11:15-11:40	**H1-4** Public health monitoring of drug interactions, patient cohorts, and behavioral outcomes via network analysis of Instagram and Twitter user timelines ***Luis Rocha***	**S5-4** Cascade Recurrent Deep Networks for Audible Range Prediction ***Yonghyun Nam***
11:40-13:00	Lunch	
13:00-14:00	**Keynote VI:** Medical data and text mining: Linking diseases, drugs, and adverse reactions **Lars Juhl Jensen**, U. of Copenhagen	
Session	**S6. PharmacoGenomic Application** Chair: Keun Woo Lee (Gyeongsang Nat'l Univ.)	**S7. Linking Phenotypes** Chair: Sangwoo Kim (Yeonsei Univ.)
14:00-14:25	**S6-1** Tissue specificity of in vitro drug sensitivity ***Benjamin Haibe-Kains***	**S7-1** An integrative approach for analyzing host factors during tuberculosis infection ***Indira Ghosh***
14:25-14:50	**S6-2** Genome Sequence Variability Predicts Pharmaceutical Withdrawals/Precautions from the Market ***Kye Hwa Lee***	**S7-2** A meta-analysis of gene expression profiles to discover obesity signatures in peripheral blood mononuclear cells ***Gwan-Su Yi***
14:50-15:15	**S6-3** Network Mirroring for Drug Repositioning ***Sunghong Park***	**S7-3** SEXCMD : Development of Sex Determination Markers for next-generation sequencing data ***Seongmun Jeong***
15:15-15:45	Coffee Break	
15:45-16:45	**Keynote VII:** Integrative genomics analyses unveil downstream biological effectors of disease-specific polymorphisms buried in intergenic regions **Yves A. Lussier**, U. of Arizona	
16:45-17:45	**Keynote VIII:** Using Electronic Health Records for Translational Science and Better Patient Care **Nigam Shah**, Stanford University	
17:45-18:00	Closing Ceremony	

Keynote Speakers

17:00-18:00 (Sunday, Oct. 16)

Olivier Lichtarge

Baylor College of Medicine, USA

Evolution vs Disease: From Big Data and Text Mining to Personalized Genomics

Abstract

Computational integration is essential to translate the buildup of biological data and publications into meaningful knowledge. But the complexity, heterogeneity, and sheer mass of information are daunting. Here, we split this long-term goal into small, tractable steps. One step integrates gene interaction networks over hundreds of species to predict gene function, including a possible new target of a leading anti-malarial drug. Another step mines the literature into a network that it then reasons over, leading in a case study to the discovery of novel p53 kinases. These examples fuse structured and unstructured data into novel networks amenable to automated hypotheses generation. But, they still lack individual patient information. As a potential solution, we introduce an analytic model of evolution. This model describes the genotype-phenotype relationship in terms of perturbations in the fitness landscape. Mutational, clinical, and population genetic data show that this approach predicts the effect of point mutations in diverse proteins, in vivo and in vitro; that it correlates disease-causing gene mutations with morbidity and mortality; and that it determines human coding polymorphism frequencies, respectively. Altogether, these studies point to an integrative network formalism that may soon reflect structured and unstructured personalized to the relevant mutational variations of any individual. Diverse applications in biology and precision medicine should follow.

09:00-10:00 (Monday, Oct. 17)

Olufunmilayo I. Olopade

University of Chicago, USA

Deploying Genomics and Immunology for Risk Assessment and Prevention

Abstract

Breast cancer is no longer defined as a single disease but rather a heterogeneous disease comprised of distinct sub-types with varied molecular, clinical and prognostic characteristics. In the Era of Precision Medicine and Cancer Moonshot, women at risk for the most aggressive forms of breast cancer can derive more benefit from innovative interventions to personalize risk assessment for early detection, and optimal use of molecularly-targeted therapies to improve clinical outcomes. We are performing whole genome sequencing of breast cancer cases on the Illumina platform, with neoplastic and non-neoplastic tissues sequenced to average depths of 90x and 30x, respectively. To handle the computational burden inherent to large-scale sequencing analyses, we have developed SwiftSeq, a modular, highly-parallel workflow for fast, efficient, and robust processing of DNA sequencing data. Using Genome Analysis Toolkit's best practices, SwiftSeq is able to completely align, process, genotype, and annotate a 30x genome in ~36-40 hours. By scaling with compute resources, our framework can analyze hundreds of genomes in days, rather than weeks. Gathering data to inform policy interventions for diverse populations of women with breast cancer is daunting. With continued sequencing, analysis, and comparison of tumor-normal genomes from The Cancer Genome Atlas, we will elucidate the unique characteristics of young onset breast cancer genomes to determine which of these alterations may be amenable to novel approaches for therapy and primary prevention.

Keynote Speakers

15:45-16:45 (Monday, Oct. 17)

Yves A. Lussier

University of Arizona, USA

Integrative genomics analyses unveil downstream biological effectors of disease-specific polymorphisms buried in intergenic regions

Abstract

Functionally altered biological mechanisms arising from disease-associated polymorphisms, remain difficult to characterise when those variants are intergenic, or, fall between genes. We sought to identify shared downstream mechanisms by which inter- and intragenic single-nucleotide polymorphisms (SNPs) contribute to a specific physiopathology. Using computational modelling of 2 million pairs of disease-associated SNPs drawn from genome-wide association studies (GWAS), integrated with expression Quantitative Trait Loci (eQTL) and Gene Ontology functional annotations, we predicted 3,870 inter-intra and inter-intra SNP pairs with convergent biological mechanisms (FDR<0.05). These prioritised SNP pairs with overlapping messenger RNA targets or similar functional annotations were more likely to be associated with the same disease than unrelated pathologies (OR>12). We additionally confirmed synergistic and antagonistic genetic interactions for a subset of prioritised SNP pairs in independent studies of Alzheimer's disease (entropy P=0.046), bladder cancer (entropyP=0.039), and rheumatoid arthritis (PheWAS case-control P<10-4). Using ENCODE data sets, we further statistically validated that the biological mechanisms shared within prioritised SNP pairs are frequently governed by matching transcription factor binding sites and long-range chromatin interactions. These results provide a 'roadmap' of disease mechanisms emerging from GWAS and further identify candidate therapeutic targets among downstream effectors of intergenic SNPs.

16:45-17:45 (Monday, Oct. 17)

Nigam Shah

Stanford University, USA

Using Electronic Health Records for Translational Science and Better Patient Care

Abstract

In the era of Electronic Health Records, it is possible to examine the outcomes of decisions made by doctors during clinical practice to identify patterns of care-generating evidence from the collective experience of patients. We will discuss methods that transform unstructured EHR data into a de-identified, temporally ordered, patient-feature matrix. We will review use-cases, which use the resulting de-identified data, for pharmacovigilance, to reposition drugs, build predictive models, and drive comparative effectiveness studies in a learning health system.

Keynote Speakers

13:00-14:00 (Monday, Oct. 17)

Lars Juhl Jensen

University of Copenhagen, Denmark

Medical data and text mining: Linking diseases, drugs, and adverse reactions

Abstract

Clinical data describing the phenotypes and treatment of patients is an underused data source that has much greater research potential than is currently realized. Mining of electronic health records (EHRs) has the potential for revealing unknown disease correlations and for improving post-approval monitoring of drugs for adverse drug reactions. In my presentation I will introduce the centralized Danish health registries and show how we use them for identification of temporal disease correlations and discovery of common diagnosis trajectories of patients. I will also describe how we perform text mining of the clinical narrative from electronic health records and use this for identification of new adverse reactions of drugs.

10:00-10:50 (Sunday, Oct. 16)

Katsuya Tsuchihara

Division of Translational Genomics, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center Japan

Data storage and sharing in SCRUM-Japan; a nation-wide cancer genome screening project for drug development

Abstract

SCRUM-Japan is a nation-wide cancer genome screening program including a lung cancer screening network, "LC-SCRUM" and a gastrointestinal cancer screening network, "GI-SCREEN". 4500 patients in total are planned to be collected from participating institutions extending from Hokkaido to the Kyushu regions from February 2015 to March 2017. Tumor samples are applied for the Oncomine Cancer Research Panel (Thermo Fischer Scientific) at CLIA-certified laboratories. Clinical information and annotated genome data are centralized to the SCRUM-Japan data center. The patients and physicians obtain individual profiles of actionable mutations and corresponding therapeutic arms. As well, the accumulated data are open for collaborating researchers in academia and industries to enhance the development of cancer therapies. As of September, 2016, 3559 cases of non-small non-squamous lung cancer, squamous cell lung cancer, colorectal cancer, and non-colorectal cancer have been enrolled. Based on the screening system, 34 clinical trials are on-going. Among them, LURET-study, a phase II study of vandetanib in patients with advanced RET-rearranged non-small cell lung cancer was successfully conducted.

Keynote Speakers

13:10-14:00 (Sunday, Oct. 16)

Koji Tsuda

The University of Tokyo, Japan

Significant Pattern Mining for Biomedical Applications

Abstract

Pattern mining techniques such as itemset mining, sequence mining and graph mining have been applied to a wide range of datasets. To convince biomedical researchers, however, it is necessary to show statistical significance of obtained patterns to prove that the patterns are not likely to emerge from random data. The key concept of significance testing is family-wise error rate, i.e., the probability of at least one pattern is falsely discovered under null hypotheses. In the worst case, FWER grows linearly to the number of all possible patterns. We show that, in reality, FWER grows much slower than the worst case, and it is possible to find significant patterns in biomedical data. The following two properties are exploited to accurately bound FWER and compute small p-value correction factors. 1) Only closed patterns need to be counted. 2) Patterns of low support can be ignored, where the support threshold depends on the Tarone bound. We introduce efficient depth-first search algorithms for discovering all significant patterns and discuss about parallel implementations.

16:10-17:00 (Sunday, Oct. 16)

Woong Yang Park

Sungkyunkwan University School of Medicine, Korea

Single cell genome analysis for precision cancer medicine

Abstract

Tumor-infiltrating lymphocytes (TILs) and the immune gene signature correlate with clinical progression in breast c ancer. We isolated single cells from four breast cancer patients with different molecular subtypes to analyze the wh ole transcriptome. Based on copy number alterations (CNAs) in gene expression patterns, tumor cells could be sepa rated from microenvironmental non-tumor cells. Although the pure population of tumor cells from four different su btypes displayed the characteristics of each subtype, heterogeneity was observed in the gene expression of cancer-r elated pathways. Most non-tumor cells were infiltrated immune cells, which showed the immune-suppressive signat ure in the triple-negative breast cancer-type sample. Immune cells for the luminal type of breast cancer consisted o f activated lymphocytes. In this study, we uncovered molecular characteristics of TILs in breast cancers by single c ell transcriptome analysis, especially through CNA-based separation of tumor and non-tumor cells.

Scientific Paper Sessions

S1. Multi-Omic Application

Room: Regency Ballroom
Date: Sunday, Oct. 16, 11:00 - 12:15

S1-1: N-of-1-pathways MixEnrich: advancing precision medicine via single-subject analysis in discovering dynamic changes of transcriptomes

Qike Li[1-4,§], A. Grant Schissler[1-4,§], Vincent Gardeux[1-3], Ikbel Achour[1-3], Colleen Kenost[1-3], Joanne Berghout[1-3], Haiquan Li[1-3,*], Hao Helen Zhang[4,5,*], Yves A. Lussier[1-4, 6-7,*]

1 Center for Biomedical Informatics and Biostatistics, The University of Arizona, Tucson, AZ, 85721, USA
2 Bio5 Institute, The University of Arizona, Tucson, AZ, 85721, USA
3 Department of Medicine, The University of Arizona, Tucson, AZ, 85721, USA
4 Graduate Interdisciplinary Program in Statistics, The University of Arizona, Tucson, AZ, 85721, USA
5 Department of Mathematics, The University of Arizona, Tucson, AZ, 85721, USA
6 University of Arizona Cancer Center, The University of Arizona, Tucson, AZ, 85721, USA
7 Institute for Genomics and Systems Biology, The University of Chicago, IL 60637, USA
§ Equal contribution

Abstract

Background: Transcriptome analytic tools are commonly used across patient cohorts to develop drugs and predict clinical outcomes. However, as precision medicine pursues more accurate and individualized treatment decisions, these methods are not designed to address single-patient transcriptome analyses. We previously developed and validated the N-of-1-pathways framework using two methods, Wilcoxon and Mahalanobis Distance (MD), for personal transcriptome analysis derived from a pair of samples of a single patient. Although, both methods uncover concordantly dysregulated pathways, they are not designed to detect dysregulated pathways with up- and down- regulated genes (bidirectional dysregulation) that are ubiquitous in biological systems.

Results: We developed N-of-1-pathways MixEnrich, a mixture model followed by a gene set enrichment test, to uncover bidirectional and concordantly dysregulated pathways one patient at a time. We assess its accuracy in a comprehensive simulation study and in a RNA-Seq data analysis of head and neck squamous cell carcinomas (HNSCCs). In presence of bidirectionally dysregulated genes in the pathway or in presence of high background noise, MixEnrich substantially outperforms previous single-subject transcriptome analysis methods, both in the simulation study and the HNSCCs data analysis (ROC Curves; higher true positive rates; lower false positive rates). Bidirectional and concordant dysregulated pathways uncovered by MixEnrich in each patient largely overlapped with the quasi-gold standard compared to

other single-subject and cohort-based transcriptome analyses.
Conclusion: The greater performance of MixEnrich presents an advantage over previous methods to meet the promise of providing accurate personal transcriptome analysis to support precision medicine at point of care.

S1-2: An Inference Method from Multi-Layered Structure of Omics

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

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

Abstract

Biological system is a multi-layered structure of omics with genome, epigenome, transcriptome, metabolome, proteome, etc., and can be further stretched to clinical/medical layers such as diseasome, drugs, and symptoms. One advantage of omics is that we can figure out an unknown component or its trait by inferring from known omics components. The component can be inferred by the ones in the same level of omics or the ones in different levels. To implement the inference process, an algorithm that can be applied to the multi-layered complex system is required. In this study, we develop a semi-supervised learning algorithm that can be applied to the multi-layered complex system. In order to verify the validity of the inference, it was applied to the prediction problem of disease co-occurrence with a two-layered network composed of symptom-layer and disease-layer. The symptom-disease layered network obtained a fairly high value of AUC, 0.74, which is regarded as noticeable improvement when comparing 0.59 AUC of single-layered disease network. If further stretched to whole layered structure of omics, the proposed method is expected to produce more promising results.

S1-3: Identification of interactions between miRNA and DNA methylation associated with gene expression as potential prognostic markers in bladder cancer

Manu Shivakumar[1,§], Younghee Lee[2,§], Lisa Bang[1], Tullika Garg[3], Kyung-ah Sohn[4,*], Dokyoon Kim[1,5,*]

1 Department of Biomedical & Translational Informatics, Geisinger Health System, Danville, Pennsylvania, USA

2 Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, Utah, USA

3 Mowad Urology Department, Geisinger Health System, Danville, Pennsylvania, USA

4 Department of Software and Computer Engineering, Ajou University, Suwon, South Korea

5 The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA

§ Equal contribution

Abstract

One of the fundamental challenges in cancer is to detect the regulators of gene expression changes during cancer progression. Through transcriptional silencing of critical cancer-related genes, epigenetic change such as DNA methylation plays a crucial role in cancer. In addition, miRNA, another major component of epigenome, is also a regulator at the post-transcriptional levels that modulate transcriptome changes. However, a mechanistic role of synergistic interactions between DNA methylation and miRNA as epigenetic regulators on transcriptomic changes and its association with clinical outcomes such as survival have remained largely unexplored in cancer. In this study, we propose an integrative framework to identify epigenetic interactions between methylation and miRNA associated with transcriptomic changes. To test the utility of the proposed framework, the bladder cancer data set, including DNA methylation, miRNA expression, and gene expression data, from The Cancer Genome Atlas (TCGA) was analyzed for this study. First, we found 120 genes associated with interactions between the two epigenomic components. Then, 11 significant epigenetic interactions between miRNA and methylation, which target E2F3, CCND1, UTP6, CDADC1, SLC35E3, METRNL, TPCN2, NACC2, VGLL4, and PTEN, were found to be associated with survival. To this end, exploration of TCGA bladder cancer data identified epigenetic interactions that are associated with survival as potential prognostic markers in bladder cancer. Given the importance and prevalence of these interactions of epigenetic events in bladder cancer it is timely to understand further how different epigenetic components interact and influence each other.

S2. Disease Genomics

Room: Terrace Ballroom
Date: Sunday, Oct. 16, 11:00 - 12:15

S2-1: Knowledge-driven binning approach for rare variant association analysis: Application to neuroimaging biomarkers in Alzheimer's disease

Dokyoon Kim[1,2], Anna O. Basile[2], Lisa Bang[1], Emrin Horgusluoglu[4], Seunggeun Lee[3], Marylyn D. Ritchie[1,2], Andrew J. Saykin[4], Kwangsik Nho[4,*], for the Alzheimer's Disease Neuroimaging Initiative (ADNI)

1 Department of Biomedical & Translational Informatics, Geisinger Health System, Danville, Pennsylvania, USA
2 The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA

3 Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA
4 Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana, USA

Abstract

Background: Rapid advancement of next generation sequencing technologies such as whole genome sequencing (WGS) has facilitated the search for genetic factors that influence disease risk in the field of human genetics. To identify rare variants associated with human diseases or traits, an efficient genome-wide binning approach is needed. In this study we developed a novel biological knowledge-based binning approach for rare-variant association analysis and then applied the approach to structural neuroimaging endophenotypes related to late-onset Alzheimer's disease (LOAD).

Methods: For rare-variant analysis, we used the knowledge-driven binning approach implemented in Bin-KAT, an automated tool, that provides 1) binning/collapsing methods for multi-level variant aggregation with a flexible, biologically informed binning strategy and 2) an option of performing unified collapsing and statistical rare variant analyses in one tool. A total of 750 non-Hispanic Caucasian participants from the Alzheimer`s Disease Neuroimaging Initiative (ADNI) cohort who had both WGS data and magnetic resonance imaging (MRI) scans were used in this study. Mean bilateral cortical thickness of the entorhinal cortex extracted from MRI scans was used as an AD-related neuroimaging endophenotype. SKAT was used for a genome-wide gene- and region-based association analysis of rare variants (MAF (minor allele frequency) < 0.05) and potential confounding factors (age, gender, years of education, intracranial volume (ICV), and MRI field strength) for entorhinal cortex thickness were used as covariates. Significant associations were determined using FDR adjustment for multiple comparisons.

Results: Our knowledge-driven binning approach identified 16 functional exonic rare variants in FANCC significantly associated with entorhinal cortex thickness (FDR-corrected p-value < 0.05). In addition, the approach identified 7 evolutionary conserved regions, which were mapped to FAF1, RFX7, LYPLAL1, and GOLGA3, significantly associated with entorhinal cortex thickness (FDR-corrected p-value < 0.05). In further analysis, the functional exonic rare variants in FANCC were also significantly associated with hippocampal volume and cerebrospinal fluid (CSF) Aβ1-42 (p-value < 0.05).

Conclusions: Our novel binning approach identified rare variants in FANCC as well as 7 evolutionary conserved regions significantly associated with a LOAD-related neuroimaging endophenotype. FANCC (fanconi anemia complementation group C) has been shown to modulate TLR and p38 MAPK-dependent expression of IL-1β in macrophages. Our results warrant further investigation in a larger independent cohort and demonstrate that the biological knowledge-driven binning approach is a powerful strategy to identify rare variants associated with AD and other complex disease.

S2-2: Genotype Based Disease Similarity Matrix from Uniqueness of Shared Genes

Matthew Carson[1], Cong Liu[2], Yao Lu[3], Caiyan Jia[4], Hui Lu[2,3,5]

1 Northwestern University, USA
2 Department of Bioengineering, University of Illinois at Chicago, USA
3 Center for Biomedical Informatics, Shanghai Children's Hospital, China
4 Department of Computer Science, Beijing Jiaotong University, China
5 SJTU-Yale Joint Center for Biostatistics, Shanghai Jiaotong University, China

Abstract

Diseases could be related to each other based on shared cause or symptoms. It has been long exploited to treat similar diseases with the similar therapies and drugs. Researchers have been exploring the disease similarities based on their shared genotype or phenotype data. Here we attempt to improve the similarity search by incorporating the uniqueness of the genes shared different diseases by construct a disease similarity matrix based on shared genes and their uniqueness defined in OMIM and DORIF annotation. By further investigating the resulting clusters, we identified several interesting links such as cancer and malaria. Our similarity matrix can be used to identify potential disease relationships and to motivate further studies into the elucidation of causal mechanisms in diseases.

S2-3: Association analysis of rare variants near the APOE region with CSF and neuroimaging biomarkers of Alzheimer's disease

Kwangsik Nho[1,3,12,*], Sungeun Kim[1,3,12], Emrin Horgusluoglu[2], Shannon L. Risacher[1,12], Li Shen[1,3,12], Dokyoon Kim[13], Seunggeun Lee[14], Tatiana Foroud[1,2,3,12], Leslie M. Shaw[4], John Q. Trojanowski[4], Paul S. Aisen[5], Ronald C. Petersen[6], , Clifford R. Jack, Jr.[7], Michael W. Weiner[8,9], Robert C. Green[10], Arthur W. Toga[11], and Andrew J. Saykin[1,2,3,12,*], 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 Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
5 Department of Neuroscience, University of California-San Diego, San Diego, CA, USA
6 Department of Neurology, Mayo Clinic Minnesota, Rochester, MN, USA
7 Department of Radiology, Mayo Clinic Minnesota, Rochester, MN, USA
8 Departments of Radiology, Medicine, and Psychiatry, University of California-San Francisco, San Francisco, CA, USA
9 Department of Veterans Affairs Medical Center, San Francisco, CA, USA
10 Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
11 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

12 Indiana Alzheimer's Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA
13 Department of Biomedical and Translational Informatics, Geisinger Health System, Danville, PA, USA
14 Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA

Abstract

Background: The APOE Îµ4 allele is the most significant common genetic risk factor for late-onset Alzheimer`s disease (LOAD). The region surrounding APOE on chromosome 19 has also shown consistent association with LOAD. However, no common variants in the region remain significant after adjusting for APOE genotype. We report a rare variant association analysis of genes in the vicinity of APOE with cerebrospinal fluid (CSF) and neuroimaging biomarkers of LOAD.

Methods: Whole genome sequencing (WGS) was performed on 817 blood DNA samples from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Sequence data from 757 non-Hispanic Caucasian participants was used in the present analysis. We extracted all rare variants (MAF (minor allele frequency) < 0.05) within a 312 kb window in APOE's vicinity encompassing 12 genes. We assessed CSF and neuroimaging (MRI and PET) biomarkers as LOAD-related quantitative endophenotypes. Gene-based analyses of rare variants were performed using the optimal Sequence Kernel Association Test (SKAT-O).

Results: A total of 3,334 rare variants (MAF < 0.05) were found within the APOE region. Among them, 72 rare non-synonymous variants were observed. Eight genes spanning the APOE region were significantly associated with CSF AÎ²1-42 (p<1.0x10-3). After controlling for APOE genotype and adjusting for multiple comparisons, 4 genes (CBLC, BCAM, APOE, and RELB) remained significant. Whole-brain surface-based analysis identified highly significant clusters associated with rare variants of CBLC in the temporal lobe region including the entorhinal cortex, as well as frontal lobe regions. Whole-brain voxel-wise analysis of amyloid PET identified significant clusters in the bilateral frontal and parietal lobes showing associations of rare variants of RELB with cortical amyloid burden.

Conclusions: Rare variants within genes spanning the APOE region are significantly associated These findings warrant further investigation and illustrate the role of next generation sequencing and quantitative endophenotypes in assessing rare variants which may help explain missing heritability in AD and other complex diseases.

S3. Cancer Bioinformatics

Room: Regency Ballroom
Date: Sunday, Oct. 16, 14:00 - 16:40

Scientific Paper Sessions

S3-1: Prediction of Recurrent Regulatory Mutations in Noncoding Cancer Genomes

Woojin Yang[1], Hyoeun Bang[1], Kiwon Jang[1], Min Kyung Sung[1], Jung Kyoon Choi[1,*]

1 Department of Bio and Brain Engineering, KAIST, Daejeon, Republic of Korea

Abstract

One of the greatest challenges in cancer genomics is to distinguish driver mutations from passenger mutations. Whereas recurrence is a hallmark of driver mutations, it is difficult to observe recurring noncoding mutations owing to a limited amount of whole-genome sequenced samples. We therefore developed a machine learning method to predict potentially recurrent mutations. In this work, we develop a random forest classifier that aims to predict regulatory mutations that may recur by learning the features of the mutations repeatedly appearing in a given cohort. With breast cancer as a model, we profiled 35 quantitative features describing genetic and epigenetic signals at the mutation site, transcription factors effected by the mutation, and genes targeted by long-range chromatin interactions. A true set of mutations for machine learning was generated by interrogating pan-cancer genomes based on our statistical model. The performance of our random forest classifier was evaluated by cross validations and showed an area under the curve of ~0.78. The variable importance of each feature in the classification of mutations was investigated. Chromatin accessibility at the mutation sites, the distance from the mutations to known cancer risk loci, and the role of the target genes in the regulatory or interaction network were among the most important variables in the classification. In conclusion, our methods enable to characterize recurrent regulatory mutations using a limited number of whole-genome samples, and based on the characterization, to predict potential driver mutations whose recurrence is not found in the given samples but likely observed with additional samples.

S3-2: Identifying subtype-specific gene expressions explained by DNA methylation patterns in breast cancer

Garam Lee[1], Lisa Bang[2], So Yeon Kim[1], Dokyoon Kim[2,3,*], Kyung-Ah Sohn[1,*]

1 Department of Software and Computer Engineering, Ajou University, Suwon 16499, South Korea
2 Department of Biomedical & Translational Informatics, Geisinger Health System, Danville, PA, USA
3 The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA

Abstract

Breast cancer is a complex disease in which different genomic patterns exists depending on different subtypes. Recent researches present that multiple subtypes of breast cancer occur at different rates, and

plays a crucial role in planning treatment. For understanding genomic mechanisms underlying breast cancer subtypes, investigating the specific gene regulatory system via different subtypes is desirable. In this paper, gene expression, as an intermediate phenotype, is estimated based on methylation profiles to identify the impact of epigenome on transcriptome in breast cancer. We propose a kernel weighted l1-regularized regression for incorporating subtype information to reveal gene regulations affected by different breast cancer subtypes. Comparing with typical method, our result shows prediction improvement of gene expression level over subtypes. Also, we identified subtype-specific network structure by carrying out the association study between gene expression and DNA methylation.

S3-3: Racial differences of intron retention and DNA methylation in breast cancer subtypes

Dongwook Kim[1], Manu Shivakumar[2], Michael Sinclair[1], Youngji Lee[3], Dokyoon Kim[2,4,*], Younghee Lee[1,*]

1 Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT 84102, USA
2 Department of Biomedical & Translational Informatics, Geisinger Health System, Danville, PA, USA
3 Department of Health and Community Systems, University of Pittsburgh School of Nursing, Pittsburgh, PA 15261, USA
4 The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA

Abstract

Regulation of gene expression by DNA methylation in gene promoter regions is well-studied; however, the effects on gene expression of methylation in the gene body (i.e., exons and introns) is comparatively understudied. Recently, hyper-methylation has been implicated in the inclusion of alternatively spliced exons; moreover, exon recognition can be enhanced by recruiting the methyl-CpG-binding protein (MeCP2) to hyper-methylated sites. In this study, we examined whether or not the level of methylation of an intron is correlated with how frequently that intron is retained during splicing. We analyzed DNA methylation and RNA sequencing data from breast cancer tissue samples in The Cancer Genome Atlas (TCGA). With breast cancer, most novel cancer-specific mRNA isoforms are due to intron retention. We found that hypo-methylation of introns is correlated with higher levels of intron expression in mRNA. In other words, the methylation level of an intron is inversely correlated with its retention in mRNA transcripts from the gene in which it is located. Furthermore, we observed significant racial difference in the methylation level of retained introns: In samples from African-American donors, retained introns were not only less methylated compared to Caucasian donors, but also were more highly expressed. Our findings have translational implications for improving diagnosis, prognosis, and treatment for breast cancer. Understanding racial epigenetic differences and their correlation with breast cancer is an important step toward achieving personalized cancer care. Moreover, IR is not only limited to breast cancer; transcriptomes from many different types of cancer show higher incidence of IR compared to

healthy controls.

S3-4: Identification of survival-associated genes from mRNA and splicing changes of cutaneous melanoma

Ji Yeon Park[1], Brian Y Ryu[1], Chan Hee Park[1], Bin Tian[2] and Ju Han Kim[1,*]

1 Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul, Republic of Korea
2 Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers New Jersey Medical School, Newark, New Jersey, USA

Abstract

Skin cutaneous melanoma (SKCM) is a cancer of the highest mutational load, and the DNA-level aberrations have been clarified through comprehensive genome sequencing. However, the transcriptional and posttranscriptional states by numerous genetic alterations remain to be fully characterized. In this study, using genomic data provided by The Cancer Genome Atlas (TCGA), we defined RNA-level genetic alterations at both transcript and exon levels between primary and metastatic samples of SKCM. Many genes related to immune response and epidermis development were significantly regulated at transcription. On the contrary, exon-level splicing changes were shown marginal, at least in the number of affected genes, but their functional link was predicted in cancer cell signaling. The mRNA expression of Epithelial Splicing Regulatory Protein 2 (ESRP2) was shown useful as a predictive marker of epithelial phenotype. To evaluate the clinical value of RNA-based measurements, we also tested the influence of somatic mutations and the correlation with patient survival times according to mRNA abundance. Our RNA-level findings promote the functional interpretation of genetic variants, and our exon-level analysis provides a more complete view of altered transcriptomics in SKCM.

S4. Bio/Medical Data Mining

Room: Terrace Ballroom
Date: Sunday, Oct. 16, 14:00 - 15:40

S4-1: Disease Causality Extraction based on Lexical Semantics and Clause Frequency from Biomedical Literature

Dong-gi Lee[1] and Hyunjung Shin[1,*]

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

Abstract

Motivation: Recently, research on human disease network has succeeded and has become an aid in figuring out the relationship between various diseases. In most disease networks, however, the relationship between diseases has been simply represented as an association. This representation results in the difficulty of identifying prior diseases and their influence on posterior diseases. In this paper, we propose a causal disease network that implements disease causality through text mining on biomedical literature.

Methods: To identify the causality between diseases, the proposed method includes two schemes: the first is the lexicon-based causality term strength, which provides the causal strength on a variety of causality terms based on lexicon analysis. The second is the frequency-based causality strength, which determines the direction and strength of causality based on document and clause frequencies in the literature.

Results: We applied the proposed method to 6,617,833 PubMed literature, and chose 195 diseases to construct a causal disease network. From all possible pairs of disease nodes in the network, 1,011 causal pairs of 149 diseases were extracted. The resulting network was compared with that of a previous study. In terms of both coverage and quality, the proposed method showed outperforming results; it determined 2.7 times more causalities and showed higher correlation with associated diseases than the existing method.

S4-2: ICU Event Prediction by integrating Sequential Patterns as Classification Features

Shameek Ghosh[1,*], Jinyan Li[1], Hung Nguyen[2], Kotagiri Ramamohanarao[3]

1 Advanced Analytics Institute, Faculty of Engineering and IT,

2 Centre for Health Technologies, Faculty of Engineering and IT, University of Technology Sydney, NSW 2007, Australia

3 Department of Computing and Information Systems, The University of Melbourne, Parkville, VIC, Australia, 3010, Australia

Abstract

Pattern mining algorithms have been previously utilized to extract informative rules in various clinical contexts. However, the number of generated patterns is numerous. In most cases, the extracted rules are directly investigated by clinicians for understanding disease diagnoses. As the elicitation of important patterns for clinical investigation places a significant demand for precision and interpretability, it is essential to obtain a set of interpretable patterns for building advanced learning models about a patient's physiological condition, especially in critical care units. In this study, a two stage sequential contrast patterns based classification framework is presented, which is used to detect critical patient events like hypotension and patient mortality. In the first stage, we obtain a set of sequential patterns by using a contrast mining algorithm. These sequential patterns undergo post-processing, for conversion to binary valued or frequency based features for developing a classification model in the second stage. Our results

on six critical care hypotension datasets and one large scale mortality prediction dataset demonstrate better predictive capabilities, when sequential patterns are used as features.

S4-3: Quad-phased Data Mining Modeling for Dementia Diagnosis

Sunjoo Bang[1], Hyunwoong Noh[2], Jihye Lee[3], Sungyun Bae[3], Kyungwon Lee[3], Changhyung Hong[2], Sangjoon Son[2,*], Hyunjung Shin[1,*]

1 Department of Industrial Engineering, Ajou University,
2 Department of Psychiatry, Ajou University School of Medicine,
3 Department of Digital Media, Ajou University, Wonchun-dong, Yeongtong-gu, Suwon 443-749, South Korea

Abstract

The number of people with dementia is increasing along with peopleâ€™s ageing trend worldwide. Therefore, there are various researches to improve a dementia diagnosis process in the field of computer-aided diagnosis (CAD) technology. The most significant issue is that the evaluation processes by physician which is based on medical information for patients and questionnaire from their guardians are time consuming, subjective and prone to error. This problem can be solved by an overall data mining modeling, which subsidizes an intuitive decision of clinicians. Therefore, in this paper we propose a quad-phased data mining modeling consisting of 4 modules. In Proposer Module, significant diagnostic criteria are selected that are effective for diagnostics. Then in Predictor Module, a model is constructed to predict and diagnose dementia based on a machine learning algorism. To help clinical physicians understand results of the predictive model better, in Descriptor Module, we interpret causes of diagnostics by profiling patient groups. Lastly, in Visualization Module, we provide visualization to effectively explore characteristics of patient groups. The proposed model is applied for CREDOS study which contains clinical data collected from 37 university-affiliated hospitals in republic of Korea from year 2005 to 2013.

S4-4: Medical Concepts Embedding

Xu Min[1], Xiaolei Xie[2], Haibo Wang[3,4], Ning Chen[1], Ting Chen[1,5]

1 MOE Key Lab of Bioinformatics; Bioinformatics Division and Center for Synthetic & Systems Biology, TNLIST; Department of Computer Science and Technology; State Key Lab of Intelligent Technology and Systems, Tsinghua University, Beijing 100084 China
2 Department of Industrial Engineering, Tsinghua University, Beijing 10084 China
3 Clinical Trial Unit, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510080 China
4 China Standard Medical Information Research Center, Shenzhen, Guangdong 518054 China

5 Program in Computational Biology and Bioinformatics, University of Southern California, Los Angels, CA 90089 USA

Abstract

One challenge in healthcare analytics is that there are a large number of medical concepts, such as clinical diagnoses and surgical operations. Proper low-dimensional representation of these medical concepts are necessary for subsequent tasks. In this paper, we propose a fast and efficient model that embeds these medical concepts into a low-dimensional Euclidean space using the Skip-gram algorithm based on the co-occurrence information to conserve the relatedness of these concepts. To prove the effectiveness of the learned embedded representation, we apply our embedding method into the patient expense prediction problem using the HQMS (Hospital Quality Monitoring System) data. In experiments, we compare our model with the one-hot vector representation method according to the prediction accuracy, showing a much improved R2 value. The embedded vectors are further visualized by the t-SNE technique to demonstrate the effectiveness of grouping related medical concepts. We also analyze the model sensitivity, and show that our model is not sensitive to the window size. Finally, we show that the embedding quality is positively correlated to the embedding dimension.

S5. Network Biology and Medicine

Room: Terrace Ballroom
Date: Monday, Oct. 17, 10:00 - 11:40

S5-1: Integrative Information Theoretic Network Analysis for GWAS of Aspirin Exacerbated Respiratory Disease in Korean Population

Sehee Wang[1], Hyun-hwan Jeong[2,3], Dokyoon Kim[4,5], Kyubum Wee[1], Hae-Sim Park[6], Seung-Hyun Kim[6,7,*], Kyung-Ah Sohn[1,*]

1 Department of Software and Computer Engineering, Ajou University, Suwon 16499, South Korea
2 Jan and Dan Duncan Neurological Research Institute at Texas Childrenâ€™s Hospital, Houston, Texas 77030, USA
3 Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, Texas 77030, USA
4 Department of Biomedical & Translational Informatics, Geisinger Health System, Danville, PA 17822, USA
5 The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA
6 Department of Allergy and Clinical Immunology, Ajou University School of Medicine, Suwon, Korea
7 Translational Research Laboratory for Inflammatory Disease, Clinical Trial Center, Ajou University Medical Center, Suwon, South Korea

Abstract

Aspirin Exacerbated Respiratory Disease (AERD) is a chronic medical condition that encompasses asthma, nasal polyposis, and hypersensitivity to aspirin and other non-steroidal anti-inflammatory drugs. Several previous studies have shown that part of the genetic effects of the disease may be induced by the interaction of multiple genetic variants. However, heavy computational cost as well as the complexity of the underlying biological mechanism has prevented a thorough investigation of epistatic interactions and thus most previous studies have typically considered only a small number of genetic variants at a time. In this study, we propose a gene network based analysis framework to identify genetic risk factors from a genome-wide association study dataset. We first derive multiple single nucleotide polymorphisms (SNP)-based epistasis networks that consider marginal and epistatic effects by using different information theoretic measures. Each SNP epistasis network is converted into a gene-gene interaction network, and the resulting gene networks are combined as one for downstream analysis. The integrated network is validated on existing knowledgebase of DisGeNET for known gene-disease associations and GeneMANIA for biological function prediction. We demonstrated our proposed method on a Korean GWAS dataset, which has genotype information of 440,094 SNPs for 188 cases and 247 controls. The topological properties of the generated networks are examined for scale-freeness, and we further performed various statistical analyses in the Allege and Asthma Portal (AAP) using the selected genes from our integrated network. Our result reveals that there are several gene modules in the network that are of biological significance and have evidence for controlling susceptibility and being related to the treatment of AERD.

S5-2: Taking promoters out of enhancers in sequence based predictions of tissue-specific mammalian enhancers

Julia Herman-Izycka[1], Michal Wlasnowolski[1], Bartek Wilczynski[1,*]

1 University of Warsaw, Krakowskie PrzedmieÅ›cie 26/28, 00-927 Warszawa, Poland

Abstract

Motivation: Many genetic diseases are caused by mutations in non-coding regions of the genome. These mutations are frequently found in enhancer sequences, causing disruption to the regulatory programme of the cell. Enhancers are short regulatory sequences in the non-coding part of the genome that are essential for the proper regulation of transcription. While the experimental methods for identification of such sequences are improving every year, our understanding of the rules behind the enhancer function has not progressed much in the last decade. This is especially true in case of tissue-specific enhancers, where there are clear problems in predicting specificity of enhancer function.

Results: We show a random-forest based machine learning approach capable of matching the performance of the current state-of-the-art methods for enhancer prediction. Then we show that it is,

similarly to other published methods, frequently cross-predicting enhancers as active in different tissues, making it less useful for predicting tissue specific activity. Then we proceed to show that the problem is related to the fact that the enhancer predicting models exhibit a bias towards predicting gene promoters as active enhancers. Then we show that using a two-step classifier can lead to lower cross-prediction between tissues.

Availability: The software needed to train the models is available at http://github.com/regulomics/enhancer prediction and the predictions themselves are available at http://regulomics.mimuw.edu.pl:8888

S5-3: Modeling Long-Term Human Activeness Using Recurrent Neural Networks for Biometric Data

Zae Myung Kim[1], Chae-Gyun Lim[1], Hyungrai Oh[2], Kyo-Joong Oh[1], Ho-Jin Choi[1,*]

1 School of Computing, KAIST, Daejeon 34141, South Korea
2 Samsung Seoul R&D Campus, Samsung Electronics, Seoul 06765, South Korea

Abstract

This paper explores the feasibility of modeling a person's "activeness" using biometric data retrieved from a fitness tracker. Currently, the notion of activeness of a user at a given period time is defined to be a tuple of three types of biometric data: heart rate, consumed calories, and the number of steps taken. Four recurrent neural network architectures are proposed to investigate the performance on predicting the activeness of the user under various length-related hyper-parameter settings. In addition, the learned model is tested to predict the time period when the user's activeness falls below a certain threshold. The dataset used in this study consists of several months of biometric time series data gathered by seven users independently. The experimental results show that forecasting the users' activeness is indeed feasible under suitable lengths of input and output sequences.

S5-4: Cascade Recurrent Deep Networks for Audible Range Prediction

Yonghyun Nam[1], Oak-Sung Choo[2], Yu-Ri Lee[2], Yun-Hoon Choung[2,*], Hyunjung Shin[1,*]

1 Department of Industrial Engineering, Ajou University, Suwon, Korea
2 Department of Otolaryngology, Ajou University School of Medicine, Suwon, Korea

Abstract

Hearing Aids amplify sounds at certain frequencies to help patients, who have hearing loss, to improve the quality of life. Variables affecting hearing improvement include the characteristics of the patients'

hearing loss, the characteristics of the hearing aids, and the characteristics of the frequencies. Although the two former characteristics have been studied, only few models reflect the characteristics of frequencies. Therefore, we propose a new machine learning algorithm that can present the degree of hearing improvement expected from the wearing of hearing aids. The proposed algorithm consists of cascade structure, recurrent structure and deep network structure. For cascade structure, it reflects correlations between frequency bands. For recurrent structure, output variables in one particular network of frequency bands are reused as input variables for other networks. Furthermore it is of deep network structure with many hidden layers. We denote such networks as cascade recurrent deep network where training consists of two phases; cascade phase and tuning phase. When applied to medical records of 2,182 patients treated for hearing loss, the proposed algorithm reduced the error rate by 58% from the other neural networks. The proposed algorithm is a novel algorithm that can be utilized for signal or sequential data. Clinically, the proposed algorithm can serve as a medical assistance tool that fulfill the patients' satisfaction.

S6. PharmagoGenomic Application

Room: Regency Ballroom
Date: Monday, Oct. 17, 14:00 - 15:15

S6-1: Tissue specificity of in vitro drug sensitivity

Fupan Yao[1,2,§], Zhaleh Safikhani[1,2,§]Seyed Ali Madani Tonekaboni[1,2], Petr Smirnov[3,4], Nehme ElÂ-Hachem[1], Mark Freeman[1,2], Venkata Satya Kumar Manem[1,2], Benjamin HaibeÂKains[1,2,5,6,*]

1 Princess Margaret Cancer Centre, Toronto, Ontario M5G 1L7, Canada
2 Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada 3 Integrative systems biology, Institut de Recherches Cliniques de Montreìal, Montreal, Quebec, Canada
4 Department of Medicine, University of Montreal, Montreìal, Quebec, Canada
5 Department of Computer Science, University of Toronto, Toronto, Ontario M5T 3A1, Canada 6 Ontario Institute of Cancer Research, Toronto, Ontario M5G 1L7, Canada

Abstract

Research in oncology traditionally focuses on specific tissue type from which the cancer develops. However, advances in highÂthroughput molecular profiling technologies have enabled the comprehensive characterization of molecular aberrations in multiple cancer types. It was hoped that these largeÂscale pharmacogenomic data would provide the foundation for a paradigm shift in oncology which would see

tumors being classified by their molecular profiles rather than tissue types, but tumors with similar genomic aberrations may respond differently to targeted therapies depending on their tissue of origin. There is therefore a need to reassess the potential association between pharmacological response and tissue of origin for cytotoxic and targeted therapies, as well as how these associations translate from preclinical to clinical settings. In this paper, we investigate the tissue specificity of drug sensitivities in largeÂscale pharmacological studies and compare these associations to those found in clinical trial descriptions. Our metaÂanalysis of the four largest i n vitro drug screening datasets indicates that tissue of origin is strongly associated with drug response. We identify novel tissueÂdrug associations, which may present exciting new avenues for drug repurposing. One caveat is that the vast majority of the significant associations found in preclinical settings do not concur with clinical observations. Accordingly, our results call for more testing to find the root cause of the discrepancies between preclinical and clinical observations.

S6-2: Genome Sequence Variability Predicts Drug Precautions and Withdrawals from the Market.

Kye Hwa Lee[1], Su Youn Baik[1], Soo Youn Lee[1], Chan Hee Park[1], Paul J. Park[3], 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 Biomedical Informatics Training and Education Center (BITEC), Seoul National University Hospital, Seoul 110744, Korea

3 Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, NV, USA

Abstract

Despite substantial premarket efforts, a significant portion of approved drugs has been withdrawn from the market for safety reasons. The deleterious impact of nonsynonymous substitutions predicted by the SIFT algorithm on structure and function of drug-related proteins was evaluated for 2504 personal genomes. Both withdrawn (n=154) and precautionary (Beers criteria (n=90), and US FDA pharmacogenomic biomarkers (n=96)) drugs showed significantly lower genomic deleteriousness scores ($P < 0.001$) compared to others (n=752). Furthermore, the rates of drug withdrawals and precautions correlated significantly with the deleteriousness scores of the drugs ($P < 0.01$); this trend was confirmed for all drugs included in the withdrawal and precaution lists by the United Nations, European Medicines Agency, DrugBank, Beers criteria, and US FDA. Our findings suggest that the person-to-person genome sequence variability is a strong independent predictor of drug withdrawals and precautions. We propose novel measures of drug safety based on personal genome sequence analysis.

S6-3: Network Mirroring for Drug Repositioning

Sunghong Park[1], Dong-gi Lee[1], Hyunjung Shin[1,*]

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

Abstract

Although drug discoveries can provide meaningful insights and significant enhancements in pharmaceutical field, the longevity and cost that it takes can be extensive where the success rate is low. In order to circumvent the problem, there has been increased interest in 'Drug Repositioning' where one searches for already approved drugs that have high potential of efficacy when applied to other diseases. To increase the success rate for drug repositioning, one considers stepwise screening and experiments based on biological reactions. Given the amount of drugs and diseases, however, the one-by-one procedure may be time consuming and expensive. In this study, we propose a machine learning based approach to efficiently selecting candidate disease and drugs. We assume that if two diseases are similar, then a drug for one disease can be applicable to other disease. For the procedure, we first construct two disease networks; one with disease-protein association and the other with disease-drug information. If two networks are dissimilar, it remains room for being either candidate disease for a drug or candidate drugs for a disease. The Kullback-Leibler divergence is employed to measure difference of connections in two constructed disease networks. Lastly, we perform repositioning of drugs to the top 20% ranked diseases. The results showed that F-measure of the proposed method was 0.75, outperforming 0.5 of greedy searching for the entire diseases.

S7. Linking Phenotypes

Room: Terrace Ballroom
Date: Monday, Oct. 17, 14:00 - 15:15

7-1: An integrative approach for analyzing host factors during tuberculosis infection

Rama Kaalia[1], Indira Ghosh1[1,*]

1 School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067, India

Abstract

Tuberculosis (TB) is an infectious disease caused due to Mycobacterium tuberculosis (MTB). Though pathogenic virulence is important, host response to MTB is known to play an important role in the manifestation of clinical symptoms of the disease. Not everyone exposed to the bacterium get sick with this disease. Identifying target genes in MTB is important, but need for completely eliminating TB requires focus on host-pathogen interactions. The main objective of the present work is to use a context-based approach to integrate different levels of information available for the disease and to study the factors associated with host response in TB infection. We have developed a Disease Association Ontology for Tuberculosis (DAO-tb) that provides a standard ontology-driven platform for describing host genes/proteins, pathways involved in tuberculosis, role of host genes during infection and for integrating functional associations from various interaction levels (gene-disease, gene-pathway, gene-function, gene-cellular component and protein-protein interactions). DAO-tb consists of 79 classes including 7 super classes. Our ontology provides a semantic based framework for querying and analyzing the disease associated information in the form of RDF graphs. Link analysis algorithms (PageRank, HITS (Hyperlink Induced Topic Search) and HITS with semantic weights) are used to score the host gene nodes on the basis of their functional associations during infection. The above developed protocol is used to predict novel potential host based targets for TB from the long list of loose gene- disease associations.

S7-2: A meta-analysis of gene expression profiles to discover obesity signatures in peripheral blood mononuclear cells

Haangik Park[1], Yul Kim[1], and Gwan-Su Yi[1,*]

1 Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea

Abstract

Obesity is typically defined as a state that abnormal amount of body fat accumulation. Due to its association with various disease pathogenesis, revealing biological mechanisms and constructing a model of obesity becomes popular. Among the studies about the obesity, gene signatures within blood tissue were often proceeded because of its interrelation with fat tissues. However, previous studies between obese patients and controls had limitation about lack of sample number and high model-dependent variability. These problems made severe difficulty for the construction of general obesity model in blood. To overcome this drawback, we constructed meta-dataset by merging four blood transcriptome microarray datasets between obesity patients and control subjects. Next we introduced a statistical testing and several classification task based on a combination of random partitioning, t-test and SVM-RFE. We ensured a validity of our own selection method by cross-validation. As a consequence of our approach, 50 differential gene expression signatures appeared among 124 obesity patients has been obtained.

Furthermore we demonstrated our finding was associated with key obesity mechanisms and some diseases caused by obesity. In conclusion, we revealed obesity signatures in blood tissues which can be applied to an effect of an obesity on the entire body and obesity-related disease pathogenesis studies.

S7-3: SEXCMD : Development of Sex Determination Markers for next-generation sequencing data

Seongmun Jeong[1], Jiwoong Kim[2], Won Park[1], Namshin Kim[1,*]

1 Personalized Genomic Medicine Research Center, Division of Strategic Research Groups, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Korea

2 Quantitative Biomedical Research Center, Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

Abstract

Generally, array-based single nucleotide polymorphism (SNP) genotyping technology uses a few markers or B-allele frequency of sex chromosome for sex determination. However, it is not applicable for the latest next- generation sequencing (NGS)-based data types because those markers should be known a priori. Also, one should align all reads onto reference genome to get B-allele frequency information on sex chromosomes. We developed novel approach to extract sex marker sequences from sex chromosomes in vertebrate and mammalian genomes. By simply counting total number of reads mapped onto each sex marker sequences, we can easily identify sex information in a very short time without aligning all sequence reads. We successfully tested out bioinformatics pipeline and sex marker sequences on human. Usually, we can identify human sex information from exome-sequencing data within a few minutes and an hour or so for high-coverage whole genome sequencing data. Y chromosome in human has psuedoautosomal regions (PAR) which are exactly duplicates from X chromosome. If we include Y chromosome for female genotyping, those sequences reads can be moved to Y chromosome instead. Here we report an open-source and easy-to-use program "SEXCMD" that can identify sex using user-created sex marker. It aligns reads onto the created sex marker sequences extracted from homologous regions between sex chromosomes, and counts the numbers of mapped reads. SEXCMD gives putative sex information within about 10 minutes for human whole genome sequencing data.

Highlight Research Tracks

Highlight Research 1.

Room: Grand Ballroom C
Date: Thursday, Oct. 3, 13:00 - 14:20

H1-1: Comparison and validation of genomic predictors for anticancer drug sensitivity

Simon Papillon-Cavanagh[1], Nicolas De Jay[1], Nehme Hachem[1], Catharina Olsen[2], Gianluca Bontempi[2], Hugo J W L Aerts[3,4], John Quackenbush[4], and Benjamin Haibe-Kains[1,*]

1 Bioinformatics and Computational Genomics Laboratory, Institut de recherches cliniques de Montréal, University of Montreal, Montreal, Quebec, Canada
2 Machine Learning Group, Université Libre de Bruxelles, Bruxelles, Belgium
3 Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts, USA
4 Department or Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts, USA

Abstract

Background: An enduring challenge in personalized medicine lies in selecting the right drug for each individual patient. While testing of drugs on patients in large trials is the only way to assess their clinical efficacy and toxicity, we dramatically lack resources to test the hundreds of drugs currently under development. Therefore the use of preclinical model systems has been intensively investigated as this approach enables response to hundreds of drugs to be tested in multiple cell lines in parallel.
Methods: Two large-scale pharmacogenomic studies recently screened multiple anticancer drugs on over 1000 cell lines. We propose to combine these datasets to build and robustly validate genomic predictors of drug response. We compared five different approaches for building predictors of increasing complexity. We assessed their performance in cross-validation and in two large validation sets, one containing the same cell lines present in the training set and another dataset composed of cell lines that have never been used during the training phase.
Results: Sixteen drugs were found in common between the datasets. We were able to validate multivariate predictors for three out of the 16 tested drugs, namely irinotecan, PD-0325901, and PLX4720. Moreover, we observed that response to 17-AAG, an inhibitor of Hsp90, could be efficiently predicted by the expression level of a single gene, NQO1.
Conclusion: These results suggest that genomic predictors could be robustly validated for specific drugs. If successfully validated in patients' tumor cells, and subsequently in clinical trials, they could act as companion tests for the corresponding drugs and play an important role in personalized medicine.
Keywords

Highlight Research Tracks

H1-2: ISOexpresso: a web-based platform for isoform-level expression analysis in human cancer

In Seok Yang[1], Hyeonju Son[1,2], Sora Kim[1,2] and Sangwoo Kim[1,2,*]

1 Severance Biomedical Science Institute, Yonsei University College of Medicine, 50-1 Yonsei-ro, 03722 Seoul, Korea
2 Brain Korea 21 PLUS Project for Medical Sciences, Yonsei University College of Medicine, 50-1 Yonsei-ro, 03722 Seoul, Korea

Abstract

Background: Alternative splicing events that result in the production of multiple gene isoforms reveals important molecular mechanisms. Gene isoforms are often differentially expressed across organs and tissues, developmental stages, and disease conditions. Specifically, recent studies show that aberrant regulation of alternative splicing frequently occurs in cancer to affect tumor cell transformation and growth. While analysis of isoform expression is important for discovering tumor-specific isoform signatures and interpreting relevant genomic mutations, there is currently no web-based, easy-to-use, and publicly available platform for this purpose.

Description: We developed ISOexpresso to provide information regarding isoform existence and expression, which can be grouped by cancer vs. normal conditions, cancer types, and tissue types. ISOexpresso implements two main functions: First, the Isoform Expression View function creates visualizations for condition-specific RNA/isoform expression patterns upon query of a gene of interest. With this function, users can easily determine the major isoform (the most expressed isoform in a sample) of a gene with respect to the condition and check whether it matches the known canonical isoform. ISOexpresso outputs expression levels of all known transcripts to check alterations of expression landscape and to find potential tumor-specific isoforms. Second, the User Data Annotation function supports annotation of genomic variants to determine the most plausible consequence of a variation (e.g., an amino acid change) among many possible interpretations. As most coding sequence mutations are effective through the subsequent transcription and translation, ISOexpresso automatically prioritizes transcripts that act as backbones for mutation effect prediction by their relative expression. By employing ISOexpresso, we could investigate the consistency between the most expressed and known canonical/principal isoforms, as well as infer candidate tumor-specific isoforms based on their expression levels. In addition, we confirmed that ISOexpresso could easily reproduce previously known isoform expression patterns: recurrent observation of a major isoform across tissues, differential isoform expression patterns in a given tissue, and switching of major isoform during tumorigenesis.

Conclusions: ISOexpresso serves as a web-based, easy-to-use platform for isoform expression and alteration analysis based on large-scale cancer database. We anticipate that ISOexpresso will expedite formulation and confirmation of novel hypotheses by providing isoform-level perspectives on cancer research. The ISOexpresso database is available online at http://wiki.tgilab.org/ISOexpresso/.

Highlight Research Tracks

H1-3: ISOexpresso: a web-based platform for isoform-level expression analysis in human cancer

Jan-Gowth Chang[1,#], Chia-Cheng Chen[2,#], Yi-Ying Wu[3,#], Ting-Fang Che[4], Yi-Syuan Huang[2], Kun-Tu Yeh[5,6], Grace S. Shieh[2,7,8,*] and Pan-Chyr Yang[4,9,10]

1 Department of Laboratory Medicine and Epigenome Research Center, China Medical University Hospital, China Medical University, Taichung, Taiwan

2 Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

3 Graduate Institute of Clinical Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan

4 Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

5 Department of Pathology, Changhua Christian Hospital, Changhua, Taiwan

6 Department of Pathology, School of Medicine, Chung Shan Medical University, Taichung, Taiwan.

7 Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan

8 Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taipei, Taiwan

9 Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan

10 Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

 # These authors have contributed equally to this work

Abstract

Two genes are called synthetic lethal (SL) if their simultaneous mutation leads to cell death, but mutation of either individual does not. Targeting SL partners of mutated cancer genes can selectively kill cancer cells, but leave normal cells intact. We present an integrated approach to uncover SL gene pairs as novel therapeutic targets of lung adenocarcinoma (LADC). Of 24 predicted SL pairs, PARP1-TP53 was validated by RNAi knockdown to have synergistic toxicity in H1975 and invasive CL1-5 LADC cells; additionally FEN1-RAD54B, BRCA1-TP53, BRCA2-TP53 and RB1-TP53 were consistent with the literature. While metastasis remains a bottleneck in cancer treatment and inhibitors of PARP1 have been developed, this result may have therapeutic potential for LADC, in which TP53 is commonly mutated. We also demonstrated that silencing PARP1 enhanced the cell death induced by the platinum-based chemotherapy drug carboplatin in lung cancer cells (CL1-5 and H1975). IHC of RAD54B↑, BRCA1↓-RAD54B↑, FEN1(N)↑-RAD54B↑ and PARP1↑-RAD54B↑ were shown to be prognostic markers for 131 Asian LADC patients, and all markers except BRCA1↓-RAD54B↑ were further confirmed by three independent gene expression data sets (a total of 426 patients) including The Cancer Genome Atlas (TCGA) cohort of LADC. Importantly, we identified POLB-TP53 and POLB as predictive markers for the TCGA cohort (230 subjects), independent of age and stage. Thus, POLB and POLB-TP53 may be used to stratify future non-Asian LADC patients for therapeutic strategies.

Highlight Research Tracks

H1-4: Public health monitoring of drug interactions, patient cohorts, and behavioral outcomes via network analysis of Instagram and Twitter user timelines

Rion Brattig Correia[1,2], Lang Li[3], Luis M. Rocha[1,4,*]

1 School of Informatics & Computing, Indiana University, Bloomington, IN 47408, USA
2 CAPES Foundation, Ministry of Education of Brazil, Brasília, DF 70040-020, Brazil
3 Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202 USA
4 Instituto Gulbenkian de Ciência, Oeiras 2780-156, Portugal

Abstract

Much recent research aims to identify evidence for Drug-Drug Interactions (DDI) and Adverse Drug reactions (ADR) from the biomedical scientific literature. In addition to this "Bibliome", the universe of social media provides a very promising source of large-scale data that can help identify DDI and ADR in ways that have not been hitherto possible. Given the large number of users, analysis of social media data may be useful to identify under-reported, population-level pathology associated with DDI, thus further contributing to improvements in population health. Moreover, tapping into this data allows us to infer drug interactions with natural products—including cannabis—which constitute an array of DDI very poorly explored by biomedical research thus far.

Our goal is to determine the potential of Instagram for public health monitoring and surveillance for DDI, ADR, and behavioral pathology at large. Most social media analysis focuses on Twitter and Facebook, but Instagram is an increasingly important platform, especially among teens, with unrestricted access of public posts, high availability of posts with geolocation coordinates, and images to supplement textual analysis.

Using drug, symptom, and natural product dictionaries for identification of the various types of DDI and ADR evidence, we have collected close to 7000 user timelines spanning from October 2010 to June 2015.We report on 1) the development of a monitoring tool to easily observe user-level timelines associated with drug and symptom terms of interest, and 2) population-level behavior via the analysis of co-occurrence networks computed from user timelines at three different scales: monthly, weekly, and daily occurrences. Analysis of these networks further reveals 3) drug and symptom direct and indirect associations with greater support in user timelines, as well as 4) clusters of symptoms and drugs revealed by the collective behavior of the observed population.

This demonstrates that Instagram contains much drug- and pathology specific data for public health monitoring of DDI and ADR, and that complex network analysis provides an important toolbox to extract health-related associations and their support from large-scale social media data.

Posters Session

TBC-1: A computational modeling for short term response predictive of long term response in the detection of Alzheimer's disease severity

Hyunjo Kim[1,*]

1 Department of Life Science, University of Gachon, Seungnam, Kyeonggido, Korea

Abstract

We have developed a computer based prediction model that is used to determine the severity of Alzheimer's disease (AD). To identify severity AD, we have analyzed the human based on these MRI images and data we have designed an automated system for the determination of AD severity. The algorithms described in this study may be used in clinical practice to validate or invalidate the diagnoses. Algorithms or method developed here may also be used for pooling diagnostic knowledge for serving mankind. Here we have described a computational based low cost AD diagnostic approach which can aid psychiatrists to quickly diagnose the various stages of AD. This system can accept AD and can successfully detect any pathological condition associated with AD.

TBC-2: An Approach to Function Prediction of Metabolites by Clustering the 3D- Chemical Structural Similarity Based Network

Md. Altaf-Ul-Amin[1,*], Nobutaka Wakamatsu[1], Shigehiko Kanaya[1]

1 Nara Institute of Science and Technology, 8916-5, Thakayama, Ikoma Nara 630-0192, Japan

Abstract

Secondary metabolites are used by humans as flavors, fragrances, medicines, biomarkers, fertilizers and for other purposes. They play important roles in ecological relationships between species. The broad functional spectrum of secondary metabolites is still not fully understood. A number of studies have investigated the relations between structures and functions of metabolites. It has been revealed that structural similarity between metabolites implies high possibility of functional similarity between them. In light of this fact we propose a method for function prediction of secondary metabolites based on guilt by association philosophy. First we determine the structural similarity scores of all possible metabolite pairs using COMPLIG algorithm and then select the metabolite pairs for which the similarity score is more than or equal to the threshold value of 0.95. To increase the possibility of clusters rich with known metabolites we then again select structurally similar metabolite pairs for which functions of both metabolites or at least one metabolite is known. The network of such metabolite pairs is then clustered using the DPClusO algorithm. Statistically significant cluster-function pairs are then selected using the concept of hypergeometric p-value and False Discovery Rate (FDR). Functions are then predicted for function unknown metabolites based on statistically significant cluster-function pairs.

TBC-3: VICTOR: a pipeline for Variant Interpretation in Clinical Testing Or Research

Bing-Jian Feng[1,*], Kristina Callis Duffin, Gerald Krueger, Wendy Kohlmann, Joshua Schiffman, Marjanka Schmidt, Alfon Meindl, Ricardo Berruti, Rita Schmutzler, Eric Hahnen, Maxime Vallee, Arnaud Droit, Douglas Easton, Sean Tavtigian, Jacques Simard and David Goldgar

1 Dermatology, University of Utah, Salt Lake City, UT, USA

Abstract

We have developed a variant interpretation pipeline that starts from a raw genotype file in VCF format. It conducts genotype-, variant-, and sample-wise quality control of data. This pipeline implements a novel functional consequence annotation program that annotates against the predominant transcripts whenever such information is available, chooses the most biologically relevant 5' or 3' representation for short insertions or deletions (InDel), merges multi-nucleotide polymorphisms (MNP), labels loss-of-function (LoF) variants, supports non-coding regions, and is robust to reference sequence errors. For clinical testing, this pipeline quantitatively integrates multiple deleteriousness scores, allele frequencies in different populations, co-segregation within pedigrees, and association among case-control samples to calculate a posterior probability of pathogenicity for each variant. For gene discovery research, it performs gene prioritization by integrating a region-based linkage analysis, a novel rare-variant association analysis where variants are weighted by deleteriousness and call quality, the relatedness of each gene to known disease genes within a gene-gene association network, and the differential gene expression between lesional and non-lesional tissues. In both scenarios, all components are combined in a quantitative fashion. Being light-weighted and fast with low demands on memory, this pipeline is scalable to whole genome sequencing (WGS) of a large sample of individuals that is typical of a complex disease research. Components of this framework can be assembled in various ways to accommodate different study designs and analysis goals. Using this pipeline, we have re-classified a TP53 variant of unknown significance (VUS) for Li-Fraumeni Syndrome (LFS), analyzed the whole exome sequencing (WES) of 1368 breast cancer cases and 3725 healthy controls from the PERSPECTIVE project (PErsonalised Risk Stratification for Prevention and Early deteCTIon of breast cancer), and analyzed the WES of 42 cases from 16 high-risk psoriasis pedigrees in the Utah Psoriasis Initiative project (UPI). The results demonstrated the value of the VICTOR pipeline in variant classification and gene discovery applications.

TBC-4: Identification of a Genetic Locus for Thoracic-to-Hip Ratio in a Large Family: a Genome-Wide Linkage and Targeted Re-sequencing Analyses

Seongwon Cha[1,*], and Changsoo Kang[2]

1 Mibyeong Research Center, Korea Institute of Oriental Medicine, Daejeon 34054, Korea

2 Department of Biology and Research Institute of Basic Sciences, College of Natural Sciences, Sungshin Women's University, Seoul 01133, Korea

Abstract

Increasing prevalence of cardiometabolic risks containing metabolic syndrome traits have affected the increased morbidity and mortality. The heritability of cardiometabolic risk factors has been known to be 31 - 77% by twin studies. Recently, various genome-wide association studies have been performed to elucidate single nucleotide polymorphisms (SNPs) associated with cardiometabolic risks. However, the heritability estimated with the genome-wide variants has been less than that from twin studies, although the genome-wide complex trait analysis has compensated a part of the missing heritability (approximately 20% - 50%). The genetic analysis combining genome-wide linkage analysis (GWLS) and next generation sequencing can facilitate the identification novel genetic loci on cardiometabolic risks from large family study. Here, we tried to find quantitative trait loci for anthropometric indices

including body mass index, waist-to-hip ratio, and thoracic-to-hip ratio (THR), systolic and diastolic blood pressures, lipid traits, and fasting blood glucose in a large family over three generations consisting of 171 individuals, using the GWLS and the followed targeted re-sequencing. After selecting 9,472 evenly distributed SNPs out of 500K genome-wide DNA chip, the significant linkage of the multiple SNPs to the THR was detected in the region of chromosome 5q12.3-31 (peak LOD score = 5.1 in the 5'-UTR of EPB41L4A). To identify causative variant associated with THR, we performed targeted sequencing of the 4.34-Mb region containing the peak SNP at its center in 31 individuals of the same family. Bioinformatic and statistical analyses showed that the most significant SNP for THR was localized in the intron of KCNN2 (p = 0.0000678). The second significant signal was detected in the SNP of EPB41L4A intron (p = 0.000636). In conclusion, we suggest that the variants in chromosome 5q21.3-22.3 may harbor genetic factors affecting THR and, by extension, cardiometabolic risk in Koreans.

TBC-5: An approach for inferring dynamic pathway interaction using cancer datasets

Shinuk Kim[1,*]

1 Sangmyung University

Abstract

In this paper we introduce an approach for inferring dynamic pathway interactions by converting static datasets to dynamic datasets using patients' clinical information. One such approach is using grade-and-stage based dynamic datasets. We generated six dynamic levels based on grades and stages, and obtained two pairs of positively related pathways among 12 enrichment pathways. The common genes of one pair of pathways consisting of LEISHMANIA INFECTION (21 overlapping genes) and ALLOGRAT REJECTION (12 overlapping genes) are four including HLA-DMB, HLA-DOA, HLA-DOB and IFNG with correlation coefficient 0.89. The other pair of pathways consists of SPLICESOME (32 overlapping genes) and PRIMARY IMMUNODEFICIENCY (15 overlapping genes) with 0.94 coefficient and no common genes.

TBC-6: Serum MicroRNA Expression Profiling of Prolonged Fatigue: RNA Sequencing and Quantitative PCR

Taehyeung Kim[1], Seongwon Cha[1,*]

1 Mibyeong Research Center, Korea Institute of Oriental Medicine, Daejeon 34054, Korea

Abstract

Background: Prolonged fatigue is defined by persistent fatigue lasting at least one month but not exceeding more than 6 months without evident clinical causes, leading to the temporary inability of physical activity or optimal cognitive performance. However, diagnostic markers of prolonged fatigue have been still unknown. Therefore, we aimed to find serum microRNAs associated with prolonged fatigue in this study.

Method: We performed small RNA-sequencing by the Illumina Nextseq 500 with serum of 10 prolonged fatigue subjects and of 10 healthy controls, both matched for age, gender, and BMI. After alignment and pairwise differential expression analysis, we identified 12 microRNA candidates whose expression were significantly altered in fatigue

subjects as compared with controls (P < 0.05 by DESeq2). At present, we are performing real-time quantitative PCR (RT-qPCR) in additional 80 fatigue subjects and 80 controls to validate differential expression of 12 microRNAs.
Results: Of 3 microRNAs having read counts of 1000 or more, miR-122-5p was down-regulated (P = 0.0088) only in women, while let-7f-5p and let-7a-5p were up-regulated (P = 0.0010 and 0.040, respectively) only in men. Of 9 microRNAs having read counts under 1000, 4 up-regulated and 1 down-regulated microRNAs were significant (P = 0.0031 - 0.044) in women, whereas 1 up-regulated and 2 down-regulated microRNAs were significant (P = 0.000056 - 0.027) in men. Exceptionally, miR-3605-5p was down-regulated (P = 0.043) in men + women. By using microRNA-seq browser (MiRGator v3.0), we checked that 12 serum microRNAs were mainly expressed in liver tissue or immune-related cells including peripheral blood mononuclear cell.
Conclusion: These findings emphasized that expression of several circulating microRNAs in fatigue individuals are significantly changed in a gender-specific manner. Furthermore, possible origin of the serum microRNAs would imply that fatigue associated microRNAs are involved in liver metabolism or immune regulation.

TBC-7: Analysis of mutation, copy number variation and DNA methylation in early breast tumorigenesis

Jong-Lyul Park[1,2], Yong-Sun Lee[3] and Seong-Young Kim[1,2,*]

1 Personalized Genomic Medicine Research Center, KRIBB, Daejeon 305-806, Korea
2 Department of Functional Genomics, University of Science and Technology, Daejeon, 305-806, Korea
3 Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX77555-1072, USA

Abstract

The timing and progression of mutation, copy number variation (CNV), and DNA methylation changes during carcinogenesis and metastasis are not completely understood. To inspect a timeline of aberrant mutation, CNV, and DNA methylation events during the carcinogenesis and metastasis progression, we analyzed normal human mammary epithelial cells (HMEC) (184D), four independent Benzo[a]pyrene (BaP)-derived immortal HMEC strains (184A1, 184AA4, 184B5, and 184BE1) and four HMEC strains immortalized with anchorage-independent growth (AIG) (184AA2, 184AA3, 184B5ME and 184FMY2) by Illumina whole genome sequencing and Epic 850K BeadChip. In carcinogenesis step coincident with immortalization, several of driver genes such as PCSK5, NFATC4, TAF1, AHNAK, CDKN2A, ASCL3, EPHB, KALRN, MED12, MTOR, ESCC5 and ESCC2 et al., were mutated. However, immortal with AIG HMEC strains acquired only one novel driver mutation (ADAM10 in 184FMY2) compared to immortal HMEC strains. For CNV, 0.06% and 0.36% of genome was amplified and deleted in the immortal HMEC strains, but percentage of amplification and deletion was 8.32 % and 18.08% in the immortal AIG HMEC strains, respectively. In case of DNA methylation, 6.91% and 5.74% of CpG sites were hypomethylated and hypomethylated in the immortal step compared to the normal HMECs but relatively small proportion (0.79% and 0.21% for hypomethylation and hypermethylation) of CpG sites were altered in the immortal with AIG step compared to the immortal HMECs. In summary, mutation and DNA methylation changes were dramatic in the carcinogenesis step, while CNV changes were dramatic in the metastasis progression. These results indicate that changes in mutation and DNA methylation may be significant in the carcinogenesis progression, while CNV changes may be important in the metastasis progression

Posters Session

TBC-8: GENT2: a platform for exploring Gene Expression patterns across Normal and Tumor tissues - newly updated

Seung-Jin Park[1,2], Seon-Kyu Kim[1], and Seon-Young Kim[1,2,*]

1 Personalized Genomic Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea
2 Department of Functional Genomics, University of Science and Technology, Daejeon, Korea

Abstract

Although distinct expression changes of a gene and its diagnostic or prognostic values in a specific cancer were frequently reported, exploring expression alterations of a gene across various tissues is still profoundly important for identifying its heterogeneous molecular behavior and clinically applying it to other types of cancer. Here, we intensely updated a searching platform of gene expression across normal and tumor tissues, namely GENT2. The system has several advanced features. First, currently, we generated a database using gene expression data obtained from more than 60,000 cancer patients, a significant increase in the number of samples than the previous platform. All data were obtained from the NCBI GEO repository and generated based on the Affymetrix U133A or U133plus2 experimental platforms (Accession numbers: GPL97 and GPL570, respectively). Second, in spite of numerous samples in the database, GENT2 shows a fast search result and provides an intuitive visualization of cancer tissue-wide gene expression patterns for utilizing Google Web Toolkit (GWT). Lastly, GENT2 also illustrates a difference of gene expression among known molecular subtypes of cancer, which were previously reported. In conclusion, with these significant improvements and plentiful user access (at least 2,000 visitors a month), GENT2 represents a promising cancer research supporting tool to provide a simple but best valuable information about genes across whole cancers. GENT2 is freely available at http://mgrc.kribb.re.kr/GENT.

TBC-9: Characterization of aging-related genes through network biology

Sang-Hun Bae[1,3], Han Wool Kim[1], Seo Jeong Shin[1], Jae Hyun Park[1], Chul Woo Lim[1], Jisook Moon[1,2,*]

1 College of Life Science, Department of Applied Bioscience, CHA University, Seoul, Korea
2 College of Life Science, Department of Bioengineering, CHA University, Seoul, Korea
3 General Research Institute, CHA general Hospital, Seoul, Korea

Abstract

Aging is an inevitable progressive decline in physiological functions and thus serves as a driver for disease and death. Due to complexity of it, the aging process needs to be understood in a systemic manner. To identify the general features of aging in the context of all the molecular interactions, we separated genes in the interactome that are associated with age into functionally distinct and physically connected modules using current biological knowledge. The modules are involved in immune process, metabolic process, developmental process, cancer pathway which are likely to be biological functions perturbed in the process of aging with the immune related sub-network showing the highest interconnectivity. Concerning relationships between age and diseases, we measured the network-based separation between the aging related modules and disease modules in the interactome. Certain disease modules are likely to be in the neighbourhood of some of the age related modules. Specifically, the age related immune module is

associated with multiple sclerosis, autoimmune disease of nervous system, demyelinating and glucose metabolism disorders. In consistent with it, genes associated with immune response and myelination in the aging hippocampus show similar co-expression patterns and tend to be expressed at higher levels with advancing age in human transcriptome data, suggesting a significant role of immune process in aging. In the study, several ageing-related modules were created by integrating biological annotation and interactome to lead to the identification of the processes driving aging and aging-disease relationships.

TBC-10: Compliance of Korean Patients with Inflammatory Bowel Disease to Colonoscopy

Jay Choi[1,2], Seungbin Oh[1,2], Eugene Jeong[1,2], and Hyun Wook Han[1,2,3,*]

1 CHA University Biomedical Informatics (CHABI),
2 Basic Medical Research Center, CHA University Graduate School of Medicine, Gyeonggi-do, Korea
3 Department of Preventive Medicine, CHA Bundang Medical Center, CHA University, Gyeonggi-do, Korea

Abstract

Patients with Inflammatory Bowel Disease in the United States (US) are recommended using surveillance colonoscopy at 2â€"3 year intervals beginning 8 years after diagnosis of IBD. However, one prior study showed that the use of surveillance colonoscopy in US Medicare patients with IBD was low. Meanwhile, it has long been commonly believed that IBD patients in Korea would have a facilitating access to healthcare, due to a lower medical treatment fee in Korea than in the United States, which would allow them to use surveillance colonoscopy more often than the patients in US have. Our aim was to study, through this retrospective, observational big data research, overall characteristics of IBD patients in Korea to challenge whether Korean IBD patients actually have a higher compliance to the surveillance colonoscopy. And then, with statistical analysis, we identified factors that affected the use of colonoscopy including sex, age, socioeconomic status, subtypes of IBD such as Crohnâ€™s Disease and Ulcerative Colitis, and the presence of Colorectal cancer, which IBD patients are at high risk to have. In conclusion, our research offers gastroenterologists in Korea more accurate views on overall IBD patients and a new clinical approach to take care of them.

TBC-11: A causal modeling approach to human disease using Korean claims data

Eugene Jeong[1,2,#], Kyungmin Ko[1,2,3,#], Seungbin Oh[1,2], Sangmin Nam[5], and Hyun Wook Han[1,2,4,*]

1 CHA University Biomedical Informatics (CHABI),
2 Basic Medical Research Center, CHA University Graduate School of Medicine, Gyeonggi-do, Korea
3 Korea Veterans Health Service Medical Center, Seoul, Korea
4 Department of Preventive Medicine, CHA Bundang Medical Center, CHA University, Gyeonggi-do, Korea
5 Department of Ophthalmology, CHA Bundang Medical Center, CHA University, Gyeonggi-do, Korea
 # Equally contributed

Abstract

In recent years, multiple risk factors of diseases are newly defined and the evidences of the relationship between

diseases have been discovered. Many researchers in the field of biological network science have presented several kinds of disease networks using big data to solve the mystery of disease-disease associations. However, there are a number of limitations to fully understand associations between human disease and apply in practice: important risk factors contributing to many human disease, such as age, sex and causality, are not considered in most of the disease networks. To bridge the gap between research findings and clinical practice, we constructed the casual network of human disease using National Health Insurance Service (NHIS) sample cohort data of approximately 2% of total Korean population from 2002 to 2013 in which disease terms are encoded according to ICD-10. The Fisher exact test with the Bonferroni correction was used to reduce the risk of obtaining false-positive results. To measure the weights of the associations, a relative risk or risk ratio(RR) was calculated and we considered significant only those combinations for which p-value < 0.001 and RR >4. Our network is composed of 798 nodes (diseases) and 6,089 links (disease-disease associations). We find that our network is a scale-free network, which suggests that a few diseases (hubs) have a large number of links while the most diseases have small degrees. By applying the clustering detection algorithm to identify highly connected local sub-networks, we present that diseases are clustered not by the ICD-10 disease classes but by the mean age at incidence, which indicates that the casual network is differentiated from other networks based on biological data. Ultimately, our network not only gives a guideline to many researchers in many fields for future researches but also help to turn the possibility of precision medicine into an achievable target.

TBC-12: A Disease Network representing Combinatorial Risk Ratio Calculations in a Large Sample Cohort

Kyungmin Ko[1,2,3,#], Eugene Jeong[1,2,#], Seungbin Oh[1,2], and Hyun Wook Han[1,2,3,4,*]

1 CHA University Biomedical Informatics (CHABI),
2 Basic Medical Research Center, CHA University Graduate School of Medicine, Gyeonggi-do, Korea
3 Korea Veterans Health Service Medical Center, Seoul, Korea
4 Department of Preventive Medicine, CHA Bundang Medical Center, CHA University, Gyeonggi-do, Korea
 # Equally contributed

Abstract

The list of medical problems that a patient has had is a very important and useful piece of information that is taken into consideration when forming a hypothesis explaining the patientâ€™s current symptoms or when planning treatment. The basis for this inference is provided in many cases by cohort studies, which are often used to test the association between an exposure and an outcome. There have been several disease networks called "comorbidity networks" based on relative risk calculations in cross-sectional models. Unlike this definition, risk ratios (also called relative risk) in cohort studies imply a temporal relationship between an exposure and outcome, which is a necessary condition for causality. We present a disease network representing combinatorial, pairwise risk ratio calculations as defined in the context of a cohort study. We used a sample cohort data provided by the National Health Insurance Service of South Korea where diagnoses are represented as ICD10 codes. A cutoff value of RR > 4 and FDR-corrected p-value of < 0.001 formed a network of 293 diseases and 3134 risk ratio relationships. We present some of the interesting and potentially useful properties of this network. Specifically, the disease nodes cluster into 4 major demographically distinct communities. In addition, the strength of the nodes calculated in the complete network aligns the disease nodes with respect to the age distribution of its patients and clusters the disease nodes into the

aforementioned communities. The combinatorial calculation and its network representation are easily scalable to the level of a clinic or hospital.

TBC-13: Mining Potential Inhibitors for Bcr-AblT315I Mutation from Chinese Traditional Medicine

Yali Xiao[1], Xin-Yi Liang[1], Ping-Ru Lai[1], and Pei-Chun Chang[1,*]

1 Department of Bioinformatics and Medical Engineering, Asia University, Taiwan

Abstract

Cancer has been ranked as one of the fatal causes since 1982. Recently, anticancer drug screening from the compounds of Chinese Traditional Medicine (TCM) has become a tendency in drug discovery. We build a drug screening process that focuses on the compounds from the formula of Chinese medicine. In this study, we focused on chronic myelogenous leukemia (CML) to mine the anticancer drug from herbs of Chinese Traditional Medicine. The gene Bcr-Abl in CML lost its regulation function of the tyrosine kinase that causes cells grow up continuously and inhibiting cells be withered. Currently, CML is treated by inhibiting the activity of Bcr-Abl tyrosine kinase that inhibits cell proliferation and induces apoptosis. Unfortunately, due to the variation of this cancer gene, the drugs such as imatinib, dasatinib, nilotinib, and bosutinib all have resistance effects for Bcr-AblT315I mutation. In addition, ponatinib has a deadly side effect. To overcome these problems, we proposed a filtering process to discover the potential drug from TCM compounds. The results show that salvianolic acid C, baicalin, 1, 4-dicaffeoylquinic acid, and dihydroisotanshinone I may have the potential for CML treatment with reducing side effects.

TBC-14: Short isoform of DNAJB6 protects against 1-methyl-4-phenylpridinium ion-induced apoptosis in LN18 cells via inhibiting ROS formation and mitochondrial membrane potential loss

Yeon-Mi Hong[1], Yohan Hong[1], Yeong-Gon Choi[1], Sujung Yeo[1], Hyejin Jung[1], Suk-Hyun Lee[1], Sae-Won Lee[1], Soo Hee Jin[1], and Sabina Lim[1,*]

1 Kyung Hee University, Korea

Abstract

In a previous study, we found that the short isoform of DNAJB6 (DNAJB6(S)) had been decreased in the striatum of a mouse model of Parkinson's disease (PD) induced by 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). DNAJB6, heat shock protein (HSP), has been implicated in the pathogenesis of Parkinson's disease (PD). In this study, we explored the cytoprotective effect of DNAJB6(S) against MPP+-induced apoptosis and the underlying molecular mechanisms in cultured LN18 cells from astrocytic tumors. We observed that MPP+ significantly reduced the cell viability and induced apoptosis in LN18 glioblastoma cells. DNAJB6(S) protected LN18 cells against MPP+-induced apoptosis not only by suppressing Bax cleavage, but also by inhibiting a series of apoptotic events including loss of mitochondrial membrane potential, increase in intracellular reactive oxygen species, and activation of caspase-9. These observations suggest that the cytoprotective effects of DNAJB6(S) may be mediated, at least in part, by the mitochondrial pathway of apoptosis.

Posters Session

TBC-15: A study of association of genetic variants with imaging phenotypes in multiple sclerosis

Kicheol Kim[1], Takuya Matsushita[1], Lohith Madireddy[1], Till Sprenger[2], Pouya Khankhanian[1], Stefano Magon[2], Yvonne Naegelin[3], Bruce A. Cree[1], Eduardo Caverzasi[1], Raija L.P. Lindberg[2], Laura E. Jonkman[3], Lisanne Balk[3], Jeroen J.G. Geurts[3], Ludwig Kappos[2], Stephen L. Hauser[1], Jorge R. Oksenberg[1], Roland G. Henry[1], Daniel Pelletier[1], Ari J. Green[1], Sergio E. Baranzini[1,*]

1 Department of Neurology, University of California, San Francisco (UCSF), San Francisco, USA
2 Department of Neurology, University Hospital of Basel, Basel, Switzerland
3 Dept. of Anatomy & Neuroscience of the VU University Medical Center, Amsterdam, Netherlands

Abstract

Multiple sclerosis (MS) is an autoimmune disorder caused by inflammatory demyelination of the central nervous system (CNS). GWAS have identified more than 140 loci that confer susceptibility to MS. However, a significant proportion of the heritability of MS remains to be explained. Relevant endophenotypes greatly empower the genetic analysis of complex diseases. In MS, brain, spinal cord, and retinal imaging represent valuable quantitative assessments of CNS integrity and function to help describe disease processes. Here we build on our experience in genetic analysis and acquisition of clinical datasets to: a) develop a set of neuroimaging-derived longitudinal endophenotypes that capture clinical relevant milestones associated with disease progression; b) test whether specific genetic variants associate with these endophenotypes. We conducted association studies using a UCSF (n=553) and two additional cohorts as replication (Amsterdam (n=205) and Basel (n=232)). Different MRI and OCT metrics were available in each of the cohorts, thus analyses were conducted on the appropriate datasets, but in all cases, at least two datasets were used. A GWAS with cortical thickness of 34 cerebral regions in UCSF and Basel datasets did not identify any significant association. However, in 9/34 regions in which thickness was found to be significantly different between cases and controls, significant associations with genes in pathways involved in neuronal differentiation were identified. Next, a focused genetic association study was conducted using only the known significant SNPs in a GWAS using MRI and OCT metrics as outcomes. Some MS-associated variants were modestly significant when tested for association with OCT metrics. These results are intriguing and warrant further exploration. The statistical significance of these associations are modest, highlighting the need to acquire even larger datasets. Efforts into merging these results with those of other groups in order to increase statistical power are underway.

TBC-16: Bioinformatics-based Analysis of Sepsis and CMap public microarray data

Seoungbin oh[1,2], Jongman Yoo[2,*], Hyun Wook Han[1,2,3,*]

1 CHA University Biomedical Informatics (CHABI),
2 Basic Medical Research Center, CHA University Graduate School of Medicine, Gyeonggi-do, Korea
3 Department of Preventive Medicine, CHA Bundang Medical Center, CHA University, Gyeonggi-do, Korea

Abstract

Sepsis is fatal systemic immune response triggered by microbial infection. A large number of specific single-targeted agents have been evaluated, but no specific agents currently approved to regulate immune system and improve the

survival rate effectively. To overcome the difficulties of anti-septic pharmacologic study, we attempted systemic and network-level analysis of sepsis to search anti-septic agents to reverse the pathological changes in sepsis patients. As one of the bioinformatics-based methods, we hypothesized that if sepsis-induced gene expression is signified by a specific set of mRNA expression signitures and exposure to a drug cause the opposite effect on the cell lines, then drug might have a therapeutic effect antagonising the disease process. We chose drug X from Connectivity mapping algorithm and we selected 100 most significant DE genes using FDR less than 0.05 on GEO data as criteria and visualized their logFC values in heatmap on both microarray data to investigate the relation between sepsis and drug X-induced expression data. After the heatmap visualization, we selected and mapped 94 inversely correlated genes to STRING database to investigate the connectivity among genes in protein-protein interaction network. And we queried OGEE database and summerized the ratio of interactions and essential genes. This bioinformatics methods can be exploited to analyse big public data and find candidate drugs for repurposing and we are developing computational and network level methods for systemic searching upon this concept.

TBC-17: Distributed computing performance adaptation for human long read sequence SNV analysis

Chang-Wei Yeh[1,#], Chieh-Wei Huang[1,#], Chao-Chun Chuang[1], Chang-Huain Hsieh[1], Yu-Tai Wang[1,*], and Chih-Min Yao[1]

1 National Center for High-performance Computing (NCHC), National Applied Research Laboratories (NARLabs), Taiwan
 # Equal co-first authors

Abstract

Single molecule sequencing long reads data computing demanding is emerging. The single molecule sequencing technology can be used for human genome single nucleotide variants calling. Despite using this technology with high quality sequencing result is expensive, long sequence reads can reduce many known issues, for example, phasing and identifying long structure variants. Base on above valuable advantages, we expect the long read sequencing technology will be getting popular and cost down continually. After the technology are popular used, long reads data type will cause what kind of computing impacts? There is no one knows. In this research, we use currently distributed computing facilities to emulate long reads data significantly emerging when identifying single nucleotide variants. In this emulating stress test, we will record each time benchmark logs. We analysed memory usages, CPU times, network traffics and file system capacities. That concludes, the computing and file system are still demanding. The memory requirement is still intensive. Those logs and parameters can help for designing next generation facilities for translational medicine users.

TBC-18: Big data analysis for chemical and protein interaction statistics

Hsuan-Feng Tseng[1,#], Chieh-Wei Huang[1,#], Chang-Wei Yeh[1,#], Chao-Chun Chuang[1], Chang-Huain Hsieh[1], Yu-Tai Wang[1,*] and Chih-Min Yao[1]

1 National Center for High-performance Computing (NCHC), National Applied Research Laboratories (NARLabs), Taiwan
 # Equal co-first authors

Abstract

There are 2,660 genes and gene products without any identification of chemical interaction in public domain. We merged Chembl, PubChem, ChEBI, Drugbank and FDA public information to be a huge data warehouse. We examined the data, in total 21,867 genes, each gene have found 13.2 interacting chemicals in average. The most interacting chemical amount genes are CASP3. It have 1,053 chemicals. The top 10 genes are CASP3, TNF, CYP3A4, CXCL8, MAPK1, MAPK3, BCL2, TP53, CYP1A1 and BAX. In the list, 7 genes are related to cancer, cell program death and cell activations. 2 genes are related to chemical detoxification and metabolism. 1 gene is for immune response. However, there are 2,660 genes can not be found any interacting chemical in our data set. In the list, there are 135 olfactory receptors without any chemical interacting information. Other genes are GTP binding proteins, gene transcription regulation pathway and unknown function. In the result, our statistics shows that we find cancer is the target of currently mankind medical resource. We also find the preferences for life scientists. However, there are many interested and important biological question, such as olfactory mystery still need to be explored.

TBC-19: Exploring Deep Learning for Making Sense of Biotech Data

Mijung Kim[1,2,*], Jasper Zuallaert[1,2], and Wesley De Neve[1,2]

1 Data Science Lab, Ghent University - iMinds, Belgium
2 Center for Biotech Data Science, Ghent University Global Campus, Korea

Abstract

Deep neural networks have recently proven to outperform different machine learning techniques. The usage of these neural networks has gained even more attention after Google DeepMindâ€™s AlphaGo managed to beat Sedol Lee in a five-game Go match in March 2016. In our research, we are applying deep learning techniques to vast sets of noisy biotech data, targeting four different use cases: (1) splice site detection in genomic data; (2) computer-aided drug discovery (CADD); (3) breast cancer detection and localization; and (4) sleep apnea detection. Each of these use cases leverages different deep learning techniques, given the different nature of the datasets involved. For splice site detection, we apply a convolutional neural network (CNN) to raw DNA sequences, with the goal of classifying candidate splice sites as true or pseudo splice sites. We combine CNNs with techniques that have already been successfully applied in the area of natural language processing, including long-short term memory networks (LSTMs) and word embeddings. For CADD, we leverage the publicly available PubChem database of chemical molecules and their activities against biological assays. In particular, we apply the ligand-based virtual screening method to detect interactions between drugs and targets, using a multi-task deep neural network. For breast cancer detection and localization, we apply a CNN to images belonging to a mammography dataset, with the goal of finding lesions. If a lesion is present in a given image, we then classify this lesion as either benign or malignant. Our neural network subsequently localizes where the lesion resides. For sleep apnea detection, we use clinical polysomnography data, for instance consisting of electroencephalograms (EEG), electrooculograms (EOG), electromyograms (EMG), and electrocardiograms (ECG). After processing of the raw data, we make use of a deep neural network to identify patterns in the cleansed data, with the aim of detecting sleep apnea.

TBC-20: A Cloud-Based Pathology Images Collaborative Platform for Medical Annotation, Analysis and Education

Chang-Wei Yeh[1], Yu-Tai Wang[1], Chao-Chun Chuang[1,*]

1 National Center for High-Performance Computing, Hsinchu 30076, Taiwan.

Abstract

The Cancer Genome Atlas (TCGA) data provides high-resolution digital whole-slide images (WSIs) for pathologists to make diagnoses directly and presents great opportunities to perform the studies of tissue morphology and development. Since different file formats and large-scale data, the integration between TCGA data and these WSIs from different laboratory are common challenges in pathology informatics. Thus, a suitable visualization and analysis platform is needed to integrate these vast and disparate images from TCGA and different laboratory. Here we developed a large scale and high performance storage system and created a web-based virtual microscopy platform for integrating WSIs and allowing users to view, search, annotate, and quantify high-resolution histology slides via the internet in real-time. Specially, users can easily keep their annotation personally or share these with other researchers by our protection system. For WSIs from different laboratory, this platform supports major Histology Image formats, including Aperio, Hamamatsu, Leica, MIRAX, Philips, Sakura, Trestle, Ventana, and Generic tiled TIFF. This platform is compatibility for any operating systems (OSX, Windows, iOS, and Android). In conclusion, while the basic purpose of the website is to provide a resource for the use of students in studying and analyzing pathological slides, it is being made available to the general pathology community and to interested clinicians everywhere.

TBC-21: Comparison of germline variant calling softwares from targeted next-generation sequencing

Minjung Kim[1], Taeheon Lee[1], Chae Hyun Lim[1], Junnam Lee[1], Guhwan Kim[1], Young-Eum Kim[1], Ja-Hyun Jang[1], Han-Wook Yoo[1], and Eun-Hae Cho[1,*]

1 Green Cross Genome, Yong-in 16924, Korea

Abstract

Background: Chromosomal microarray has been used as a first-tier diagnostic tool for microdeletion/microduplication syndromes in individuals with developmental delays or congenital anomalies. However, next generation sequencing (NGS) technology with rapid dropping of whole genome sequencing (WGS) cost, provides the possibility of low coverage WGS as an alternative method of copy number variation (CNV) detection in clinical cytogenetics. In this study, we developed bioinformatic pipeline for accurate CNV detection and compared the results of low coverage WGS with chromosomal microarray.

Methods: We analysed clinical samples of 62 patients who had been previously tested with chromosomal microarray (Affymetrix cytoscan 750K) due to congenital anomalies or developmental delays. The libraries prepared from these samples were pooled and sequenced with Nextseq 500 (Illumina) 75bp length. The average 3.6 million reads per sample were produced. Reads were aligned and curated to eliminate GC bias, mappability and high-order artifact using principle component analysis (PCA). For sample quality check, we developed Q-score system which used LOESS smoothing algorithm for elimination of locally clustered noise. Samples with higher Q-scores had more false segmentations.

Results: Compared with chromosomal microarray, we showed the possibility of low coverage WGS for detection of

clinically relevant CNVs with lower cost and higher capacity. This study was supported by R&D program of MOTIE/KEIT (10053626), Republic of Korea.

TBC-22: Low coverage sequencing for comprehensive screening of chromosomal Copy Number Variation relative disease

Junnam Lee[1], Young Joo Jeon[1], Chae Hyun Lim[1], Taeheon Lee[1], Minjung Kim[1], Young-Eun Kim[1], Ja-Hyun Jang[1], and Eun-Hea Cho[1,*]

1 Green Cross Genome, Yong-in 16924, Korea

Abstract

Background: Next Generation Sequencing (NGS) technologies enable fast and economic genome sequencing for clinical research and diagnosis. Accordingly, many analysis applications have been developed. But there have yet been high false positive rates in Insertion/deletion (INDEL) calling. So the aim of this study is to develop variant calling methods for targeted NGS.

Methods: For development of analysis pipeline, we compared performances of all possible combination of SNP and INDEL callers separately using the NA12878 whole exome sequencing data sets and NIST Genome in a Bottle validation call-set. And we tested this pipeline using patient samples of targeted NGS panels.

Results: The combination of UnifiedGenotyper and Samtools showed the best performance in SNP calling and a combination of three callers (UnifiedGenotyper, Freebayes and Scalpel) exhibited higher precision rate and sensitivity in INDEL calling using data sets of NA12878 whole-exome sequencing. This combination also detected about 90bp long deletion. We demonstrated 100% concordance in detecting 297 pathogenic single nucleotide variants and 33 pathogenic insertion-deletion mutations in 340 patients that were previously confirmed by Sanger sequencing. Intra- and inter-run reproducibility tests showed 100% of efficiency (sensitivity as well as precision rate).

Conclusion: This study provides accurate and reproducible analysis pipeline by targeted sequencing. We demonstrate 100% concordance in mutations identified. Clinical trials using these targeted panels and analysis pipeline will be conducted for the KFDA IVD approval. This study was supported by R&D program of MOTIE/KEIT (10053626), Republic of Korea.

TBC-23: Medical Examination Data Prediction with Missing Information Using Long Short-Term Memory

Han-Gyu Kim[1], Gil-Jin Jang[2], Ho-Jin Choi[1,*], Minho Kim[3], Young-Won Kim[3], and Jae-Hun Choi[3]

1 School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea
2 School of Electronics Engineering, Kyungpook National University, Daegu 41566, South Korea
3 Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

Abstract

In this work, we use recurrent neural network (RNN) to predict the medical examination data with missing parts. There often exist missing parts in medical examination data due to various human factors, for instance, because human subjects occasionally miss their annual examinations. Such missing parts make it hard to predict the future

examination data by machines. Thus, imputation of the missing information is needed for accurate prediction of medical examination data. Among various types of RNNs, we choose long short-term memory (LSTM) to predict the missing information as well as the future medical examination data, as LSTM shows good performance in many relevant applications. In our proposed method, the temporal trajectories of the medical examination measurements are modelled by LSTM with the missed measurements compensated, which is then used to predict the future measurements to be used as diagnosing the diseases of the subjects in advance. We have carried out experiments using a medical examination database of Korean people for 12 consecutive years with 13 medical fields. In this database, 11500 people took the medical check-up every year, and 7400 people missed their examination occasionally. We use complete data to train LSTM, and the data with missing parts are used to evaluate the imputation and future measurement prediction performance. In terms of root mean squared error (RMSE) between the prediction and the actual measurements, the experimental results show that the proposed LSTM network predicts medical examination data much better than the conventional linear regression in most of the examination items.

TBC-24: ConVarCal Facilitates Robust Massive Parallel Sequencing Variant Calling

Yonglan Zheng1, Alex Rodriguez2, Segun C. Jung2, Toshio F. Yoshimatsu1, Ravi K. Madduri2, Utpal J. Dave2, Ian Foster2, Olufunmilayo I. Olopade1,*

1 Center for Clinical Cancer Genetics, Department of Medicine, The University of Chicago, USA
2 Computation Institute and Argonne National Laboratory, The University of Chicago, USA

Abstract

Background: The vastly increasing implementation of massive parallel sequencing (MPS) in academic and clinical settings demands for reliable and reproducible variant calling methods. Precise detection of single nucleotide variants (SNVs), insertions and deletions (Indels), and structural variants (SVs) are required at both individual and population levels.

Methods: Our MPS variant identification platform, ConVarCal (Confident Variant Calling), compiles multiple tools in a malleable manner using the elastic computing capability of Globus Genomics built upon Amazon Web Services. FASTQ files are submitted for BWA-MEM or Bowtie2 alignment; BAM files generated are subsequently processed by highly parallelized workflows: GATK HaplotypeCaller, Platypus, FreeBayes, SAMtools mpileup, and Atlas2. The output VCF files are normalized, and a set of highly confident variants are obtained through refinement by Consensus Genotyper for ANNOVAR or VEP annotation. SVs are precisely detected with MetaSV workflow that integrates Pindel, BreakDancer, BreakSeq, CNVkit, and Manta. DELLY, LUMPY and CONTRA are also available. For dynamic parallel computing, wrapper script using Swift language was implemented for some callers.

Results: We tested the performance of ConVarCal by analyzing germline targeted sequencing data (1.3Mbp, ave. 260x) of 200 Nigerian breast cancer patients. The entire analysis was completed in a week, but the total processing time varied depending on the configuration and availability of cloud computing resources. ConVarCal confidently identified 25 deleterious SNVs/Indels in 29 subjects, and all have been confirmed experimentally. In addition, users can share and trace the analytic steps; further optimize the operations through adjustment of parameters, combination of job ordering for better parallelization, or properly allocating computing resources; and analyze the performance through

visualization of resource-performance plots.
Conclusion: ConVarCal takes full advantage of Globus Genomics for MPS variant calling in a reliable and robust manner. It has great scalability and its modular design allows building additional tools to further enhance the platform.

TBC-25: Creation and Validation of Metadata Registry based Personal Health Record

Hye Hyeon Kim[1], Ju Han Kim[1,*]

1 Division of Biomedical Informatics, College of Medicine, Seoul National University, Seoul, South Korea

Abstract

Personal health record (PHR) is a collection of information about individual health. It includes patient data that helps each individual and their health care providers manage their health as containing allergies, medications, family history, and so on. However, as kind of snapshot data, PHR has limitation to cover detail of patient data and to represent precise and semantic representation in the limited PHR model. To address this problem, we adopted ISO/IEC 11179 metamodel based MDR in PHR. We first adopted CCD/CCR standard models for representing standard based PHR. And we developed the process of how MDR based PHR is created with five steps; 1) Extracting individual health data from EMR/HER, 2) Determine whether MDR based PHR is developed, 3) Retrieval CDEs for MDR based PHR, 4) Creation and registration of PHR related CDEs, 5) Completion of MDR based PHR. We also developed the process of how MDR based PHR is validated as using value domain information of data element such as data type, min/max value, and so on as including three steps: 1) CCD/CCR XML Schema based validation; 2) CCD/CCR+ XML Schema based validation; 3) MDR based semantic validation. As a result, we specified how MDR based PHR in CCD/CCR model is represented with sample patient data. A data element in the MDR based PHR can be a medium to bring data semantics from rich semantic contents of MDR. It also provides several benefits including rich semantic representation with clear definition, semantic validation for patient data, improving semantic interoperability.

TBC-26: EasyFormBuilder: Form building tool based on standardized metadata repository to facilitate semantic interoperability

Hyeong Joon Kim[1], Hye Hyeon Kim[1], Ju Han Kim[1,*]

1 Division of Biomedical Informatics, College of Medicine, Seoul National University, Seoul, South Korea

Abstract

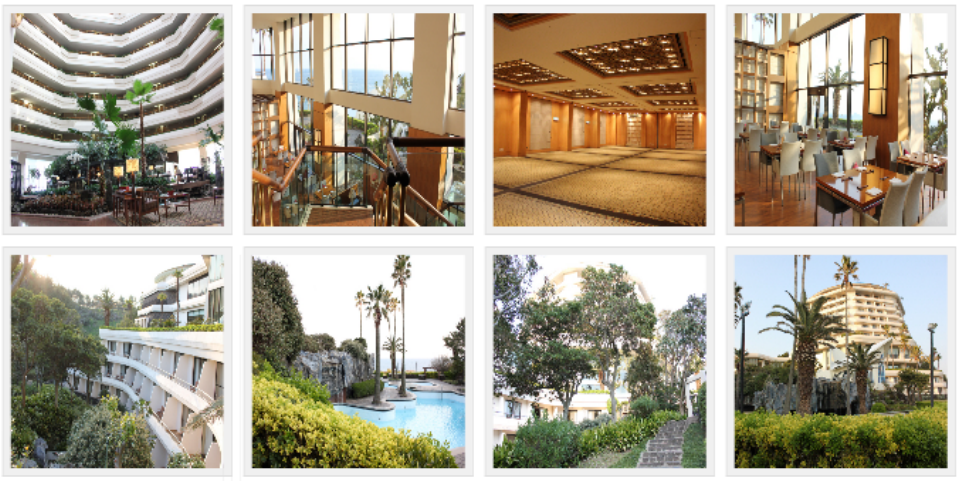
Clinical document is an effective tool for collecting patient data including demography, family history, and disease history. Though there is HL7 CDA standard to develop standard based clinical documents, clinical documents are developed differently and separately for each physician and for each hospital. So that, there are big variability among the same kind clinical documents such as the same admission notes among different hospitals. Paper or PDF based clinical documents are also problem to exchange and share clinical data. To address these problems, we developed semi-automatic and web-based application to build clinical forms, named EasyFormBuilder, composed by ISO/IEC 11179 based Common Data Elements (CDEs) for enhancing semantic interoperability. The process of generating forms

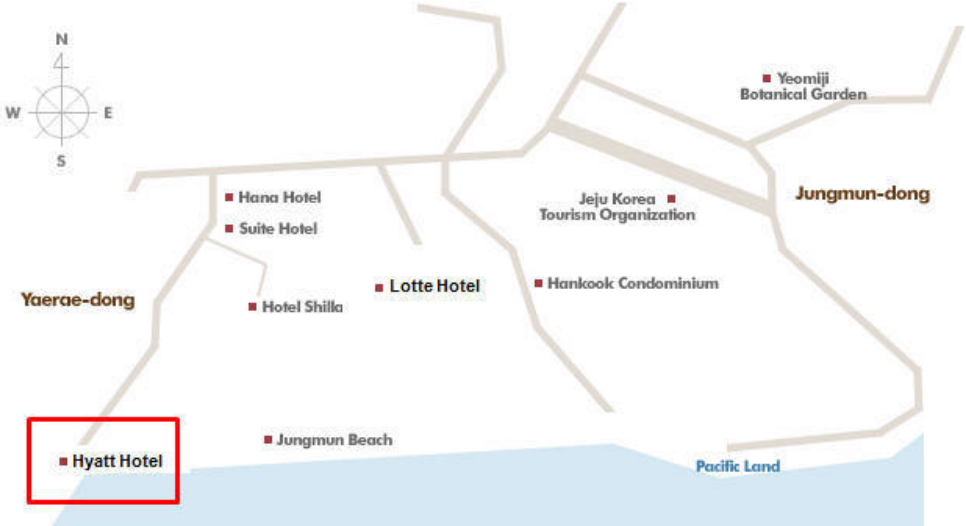
in EasyFormBuilder is summarized as follows; 1) User inserts the basic information about the form into the web page and upload CSV file template from EasyFormBuilder. The template is based items from ISO/IEC 11179 for storing information of CDE. 2) Our tool generate a form by using extracted information from CSV file. 3) The tool searches CDEs stored in metadata registry (MDR) to match questions and makes the 'List of recommended CDEs' for enhancing semantic interoperability. 4) For ensuring semantic representation, the user can choose most appropriate CDE in the list, and annotate the CDE to each user-defined question. 5) As last step, it is completed as building XML based form. As utilizing large scale of CDEs in MDR, we can build a forms easily. Through automatic form generation part, XML based form is created, and it gives machine and human readable documents, so it is useful to read, write, and reuse. Another benefit to use our fool is that it give rich semantic contents for each questions annotated CDEs as CDE has precise definition and information of concepts and representation.

Venue

Location & Map

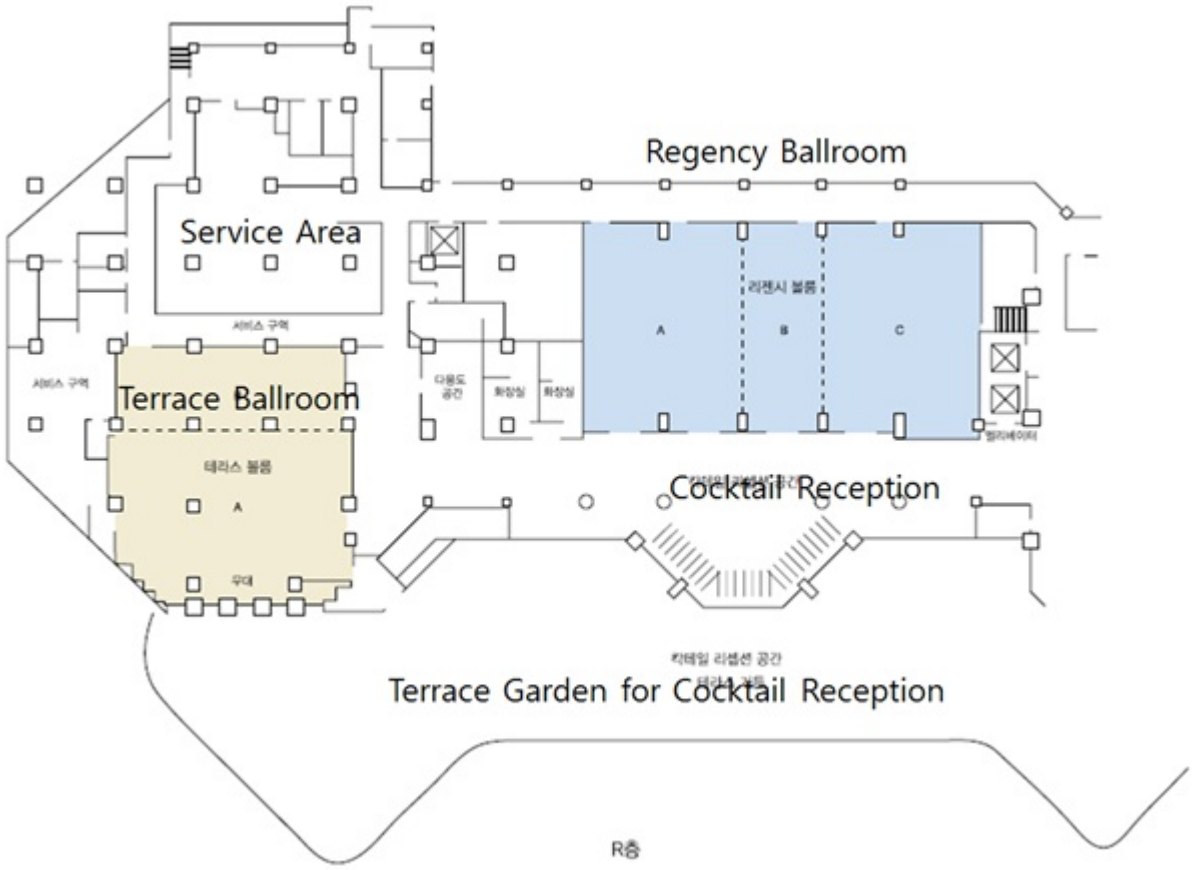
Regency & Terrace Ballroom, Hyatt Regency Jeju, Jeju Island, Korea





Venue

Floor Plan



Terrace & Regency Ballroom

Tours

Jeju Island - New 7 Wonders of Nature Jejudo is a volcanic island, 130 km from the southern coast of Korea. The largest island and smallest province in Korea, the island has a surface area of 1,846 sqkm. A central feature of Jeju is Hallasan, the tallest mountain in South Korea and a dormant volcano, which rises 1,950 m above sea level. 360 satellite volcanoes are around the main volcano.

Recommended Tour Courses

Seongsan Ilchulbong (Sunrise Peak)

99 rocky peaks surround the crater like a fortress and the gentle southern slope connected to water is a lush grassland.

On the grassland at the entrance of Sunrise Peak, you can enjoy horseback riding. Breathtaking scenic views while taking a rest in the middle of climbing up the peak such as Mount Halla, the deep blues of the ocean, the multi-colored coast line, and the picturesque neighboring villages will become unforgettable memories.

Jeju Olle Tour

"Olle" [Ole] is the Jeju word for a narrow pathway that is connected from the street to the front gate of a house. Hence, "Olle" is a path that comes out from a secret room to an open space and a gateway to the world. If the road is connected, it is linked to the whole island and the rest of the world as well. It has the same sound as "Would you come?" in Korean, so Jeju's "Olle" sounds the same as 'Would you come to Jeju?'.

[Route 8] Wolpyeong to Daepyeong(Port)

This route continues along the seashore through Jusangjeolli which is a formation of stone pillars piled up along the coast. The Jusangjeolli were created when the lava from Mt. Hallasan erupted into the sea of Jungmun. The sight of the abundant pampas grass makes your walk even more enjoyable. A pathway made of numerous rocks on the coastline was built by the marine corps for Jeju Olle, so it is called The Marine Corps Trail. The pathway used to be used only by local divers. The terminus of the route, Daepyeong Pogu (port), sits on the end of a valley and the open fields run towards the sea. Gun San, a mountain in Daepyeong, was created by the son of a sea god thank his master for his merciful attitude, according to local legend.

Halla Mountain

Mount Halla is the mountain of one of the three gods and is a notable mountain. It stands at the center of Jeju Island, spreading east and west. The east face is steep, the north side is gentle, and the east and west form a flat, wide highland. Mount Halla is a dormant volcano created by volcanic activities during the quaternary period of the Cenozoic era. It is primarily covered with basalt. On its top is a crater Baeknok Lake. This mountain is a home to alpine plants and houses as many as 1,800 species of flora. It also boasts luxuriant natural forests and vast grasslands.

Tours

In addition, precipitous cliffs and slopes, and unusual rock formations standing along valleys produce magnificent scenic views. The community of azaleas also adds to the beauty of Mount Halla. Mount Hallas autumnal tints and snow-covered scenes have been selected as the best of the best. It is possible to climb up to Wetse Oreum along Eorimok Trail and Youngsil Trail and to the top along Seongpanak Trail and Kwaneumsa Temple Trail.

Micheon Cave (Ilchul Land)

Filled with underground mystery, Micheon Cave has academic, tourism, and cultural value. Fresh air, crystal clear water, green fields, and a secondary volcanic cone (oreum) are nearby. This underground cave is nature awe-inspiring spot that provides an opportunity for contemplating human nature and the future.

You'll be fascinated with the nature beauty that simply cannot be felt in the city. Micheon Ilchulland Cave, where the new sun is rising, is here to make your tour more enjoyable.

Cheonjiyeon Waterfall

The waterfall falls from a precipice with thundering sounds, creating white water pillars. It has the name Cheonjiyon, meaning 'the heaven and the earth meet and create a pond'. At 22 m in height and 12 min width, the waterfall tumbles down to the pond to produce awe-inspiring scenery. The valley near the waterfall is home to Elaeocarpus sylvestrisvar, ellipticus, which is Natural Monument No. 163, Psilotum nudum, Castanopsis cuspidatavar.sieboldii, Xylosma congestum, Camellia and other subtropical trees. This place is also famous as home to the eel of Anguilla mauritiana, which is Natural Monument No. 27 and is active primarily at night. The Chilsipri Festival is held in every September at the falls.

This is a **Dol-Haru-bang**, and it means a grandfather made of stone. The people in Jeju Island believe that this Dol-Haru-bang is a guardian of the Jejudo to protect it. This souvenir is made a cute feature compare with the original Dor-haru-bang. Therefore, the real Dol-Haru-bang could be seen more fearful a bit rather than this one.




Conference App

Download now the TBC 2016 app for iOS and Android.

iOS

Android

Informatics Journals Supporting TBC

- **Journal of American Medical Informatics Association (JAMIA)**, IF: 3.974 (The first in the Medical Informatics Category)
- **Journal of Biomedical Informatics (JBI),** IF: 2.817
- **BMC Medical Genomics,** IF: 3.466
- **BMC Medical Informatics and Decision Making,** IF: 1.83
- **Healthcare Informatics Research**
- **Genomics & Informatics**

Sponsors

Co-sponsoring with AMIA Joint Summits

by AMIA

SEOUL NATIONAL UNIVERSITY

KAIST

ETRI
한국전자통신연구원

AJOU UNIVERSITY

대한의료정보학회
THE KOREAN SOCIETY OF MEDICAL INFORMATICS

한국생물정보시스템생물학회
Korean Society for Bioinformatics and Systems Biology

시스템 바이오 정보의학 국가핵심연구센터
Systems Biomedical Informatics National Core Research Center

KMA KOREAN MEDICAL ASSOCIATION

life technologies™