Oral Presentation 9

日時：2020 年 10 月 30 日 13:00-14:30
チャンネル：３

口頭発表９
『バイオインフォマティックス』
Bioinformatics

座長：　荻島 創一 （東北大学 東北メディカル・メガバンク機構）
Soichi Ogishima (Tohoku Medical Megabank Organization, Tohoku University)
茂櫛 薫 （富士通株式会社）
Kaoru Mogushi (Fujitsu Limited)

O3-08
"Systems biological approaches to understand the mechanisms of cancer drug resistance & cardiotoxicity"
Shigeyuki Magi
(Toho University)

O3-09
"Development of phenotyping algorithm of type 2 diabetes mellitus: A retrospective claims database analysis"
Satoshi Nagaie
(Department of Informatics for Genomic Medicine, Tohoku Medical Megabank Organization, Tohoku University)

O3-10
"Integrated Database "dbTMM" with family linkage for stratification of cohort participants toward drug development"
Naoko Kasahara
(Department of Informatics for Genomic Medicine, Tohoku Medical Megabank Organization, Tohoku University)

O3-11
"Estimation of relationships between chemical substructures and gene expression antibiotic-resistance of bacteria: Adapting canonical correlation analysis for small sample data by gathered features using consensus clustering"
Tsuyoshi Esaki
(Shiga University)

O3-08

Systems biological approaches to understand the mechanisms of cancer drug resistance & cardiotoxicity

Shigeyuki Magi[1] **Atsuhiko T Naito**[1]
Shigeyuki.magi@med.toho-u.ac.jp atsuhiko.naito@ med.toho-u.ac.jp

[1] Toho University, 5-21-16, Omori-Nishi, Ota-ku, Tokyo,143-8540, Japan

Keywords: drug resistance, cardiomyocyte dysfunction, mathematical modeling

Two major challenges in cancer drug therapy are the resistance of cancer cells and the toxicity on normal cells. Here, we introduce our approaches to understand the mechanisms of drug resistance against estrogen receptor antagonist and the cardiotoxicity against anti-HER2 monoclonal antibody.

The process of drug resistance in cancer cells can be understood as a series of changes in the molecular network in which cancer cells go through an unstable state where cells acquire heterogeneity due to the external stress of anticancer drugs and subsequent transition of surviving cells to another stable survival state. We tried to elucidate the molecular mechanisms contributing to the process of acquiring drug resistance by analyzing the changes in gene expression of breast cancer cells after the treatment with tamoxifen (TAM: estrogen receptor antagonist). We performed RNA-sequencing every week after the TAM treatment and identified that genes regulated by TGFβ, NOTCH, HIF-1 and cell adhesion molecules were increased after the growth rate of the cells was recovered, suggesting that these genes may characterize and contribute to the drug resistance against TAM. We also identified that genes related to FOXO signaling and autophagy were increased at the border period just before the growth rate of the cells was recovered, suggesting that these genes play role in “acquisition” of drug resistance and may become the biomarkers for pre-resistant state. To further analyze the transition trajectories to resistant cell types, we performed a single-cell RNA-sequencing after TAM treatment and revealed that resistant cells consisting two major subpopulations with different characteristics emerge from the pre-resistant state. We are currently constructing a mathematical model capable of reproducing the changes in subpopulation during the process of resistance acquisition.

The anti-HER2 antibody trastuzumab is known to have side effects that cause cardiac dysfunction and heart failure, but its effect on physiological HER2 signaling to human cardiomyocytes remains elusive. We report our preliminary results regarding the effects of trastuzumab on the HER2 signaling in human iPS cell-derived cardiomyocytes and human breast cancer cell lines and introduce our plan for the future research.

O3-09

Development of phenotyping algorithm of type 2 diabetes mellitus: A retrospective claims database analysis

Satoshi Nagaie[1]
nagaie@megabank.tohoku.ac.jp

Satoshi Mizuno[1]
samizuno@med.tohoku.ac.jp

Ryosuke Ishiwata[1]
rishiwata@megabank.tohoku.ac.jp

Naoko Kasahara[1]
nagaie@megabank.tohoku.ac.jp

Soichi Ogishima[1]
ogishima@megabank.tohoku.ac.jp

[1] Department of Informatics for Genomic Medicine,
Tohoku Medical Megabank Organization, Tohoku University,
Seiryo 2-1, Aoba-ku, Sendai Miyagi 980-8573 Japan

Keywords: Electronic Health Record, Phenotyping, Claims Database

As the research and development of genomic medicine based on individual genome information has progressed, highly accurate and deep genome information on genotype is being studied. On the other hand, for phenotypic information, more in-depth information on disease types as well as disease names is required. In such a situation, research is being conducted to obtain deeper information on disease types by secondary use of real-world data, such as medical information. The classification of disease types based on medical information is called “phenotyping”. In the U.S., the eMERGE project is constructing a database that aggregates 68 different pathological classification algorithms called PheKB, which is the result of extracting medical information from EHRs and classifying them into disease types.

In this study, we develop a phenotyping algorithm for type 2 diabetes mellitus (T2DM) using 3,014,656 subjects' receipt data and 3,156,110 subjects' specific health checkup data, especially medical practices (ICD-10 codes = E11)), laboratory test values (random glucose ≥ 200 mg/dL, HbA1c ≥ 6.5%) and medications (All drugs for the treatment of T2DM that have been reviewed for approval by the PMDA), registered from January 2005 to May 2019, among the data provided by JMDC, Inc. In order to evaluate the algorithm, we calculate sensitivity and specificity, which will need to be labeled with the correct answer. For this purpose, it is usually necessary for physicians to review medical records again, which is time-consuming and costly. In this study, we evaluate the performance of the algorithm using a prediction model of disease type classification as a probabilistic gold standard, without a physician's review of medical records.

O3-10

Integrated Database “dbTMM” with family linkage for stratification of cohort participants toward drug development

Naoko Kasahara[1]
n.kasahara@megabank.tohoku.ac.jp

Satoshi Mizuno[1]
samizuno@med.tohoku.ac.jp

Ryosuke Ishiwata[1]
rishiwata@megabank.tohoku.ac.jp

Satoshi Nagaie[1]
nagaie@megabank.tohoku.ac.jp

Yukie Konno[1]
y.konno@megabank.tohoku.ac.jp

Hiroshi Tanaka[2]
tanaka@megabank.tohoku.ac.jp

Soichi Ogishima[1]
ogishima@megabank.tohoku.ac.jp

[1] Department of Informatics for Genomic Medicine, Tohoku Medical Megabank Organization, Tohoku University, Seiryo 2-1, Aoba-ku, Sendai Miyagi 980-8573 Japan
[2] Department of Bioclinical Informatics, Tohoku Medical Megabank Organization, Tohoku University, Seiryo 2-1, Aoba-ku, Sendai Miyagi 980-8573 Japan

Keywords: Biobank, Cohort, Stratification, Drug Discovery

Tohoku Medical Megabank project was started to reconstruct from the Great East Japan Earthquake and Tsunami. We started cohort studies in 2013, and successfully recruited 84,000 residents for Community-Based cohort and 73,000 people for Birth and Three-Generation cohort, totally more than 150,000 participants for baseline assessment by 2017. Though these cohorts, biospecimens including serum, plasma, mononuclear cells, urine were collected more than 3.8 million sample storage (July 2020) and the answers to the questionnaires of their lifestyles including food, psychological condition, experiences of the disaster were also collected. A part of participants took more than 10 physiological examinations and also MRI examination in assessment centers and satellites.

We have developed the integrated database of Tohoku Medical Megabank project (dbTMM). In this July, we released collected data with family linkage of 68,000 subjects of Birth and Three-Generation cohort. Currently, the dbTMM stores large-scale data of 145,000 cohort participants including whole genome (n=3,500) and SNP genotyping data (n=80,000) (genetic factors), questionnaire data (environmental factors), family linkage data (n = 20,000), MRI data (n=4,300) and metabolome and proteome data (n=6,000) (molecular phenotypic factors), and clinical data (phenotypic factors). To utilize family linkage data, we developed a search function with family linkage to enable researchers to stratify cohort participants having a certain phenotype by genetic and environmental factors considering family linkage. As for clinical data, we collected medical charts and stored them into dbTMM. By using structured clinical data, we have conducted deep phenotyping of obstetric disease including hypertensive disorders of pregnancy. We believe our integrated database dbTMM will contribute to stratification of cohort participants toward drug development for common disease which is caused by complex interplay between genetic and environmental factors.

[1] Kuriyama, S. *et al.* The Tohoku Medical Megabank Project: design and mission. *J. Epidemiol.* 26, 493–511 (2016).

O3-11

Estimation of relationships between chemical substructures and gene expression antibiotic-resistance of bacteria: Adapting canonical correlation analysis for small sample data by gathered features using consensus clustering

Tsuyoshi Esaki[1]
tsuyoshi-esaki@biwako.shiga-u.ac.jp

Takaaki Horinouchi[2]
takaaki_horinouchi@rike.jp

Yayoi Natsume-Kitatani[3,4]
natsume@nibiohn.go.jp

Yosui Nojima[3]
yosui@nibiohn.go.jp

Iwao Sakane[5]
i-sakane@itoen.co.jp

Hidetoshi Matsui[1,6]
hmatsui@biwako.shiga-u.ac.jp

[1] The Center for Data Science Education and Research, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan
[2] Center for Biosystems Dynamics Research, RIKEN, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan
[3] Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-8-6 Saito Asagi, Ibaraki, Osaska 567-0085, Japan
[4] Center of Drug Design Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-8-6 Saito Asagi, Ibaraki, Osaka 567-0085, Japan
[5] Central Research Institute, ITO EN Ltd., 21 Megami, Makinohara, Shizuoka 421-0516, Japan
[6] Faculty of Data Science, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan

Keywords: canonical correlation analysis, clustering, antibiotic resistance

The emergence of antibiotic-resistant bacteria is a serious public health concern worldwide. To understand the drug resistance mechanisms employed by bacteria, quantitative phenotypic changes associated with drug resistance acquisition in *Escherichia coli* have been investigated. However, relationships between specific antibiotic structures and the acquisition of resistance by bacteria have not been clarified. Understanding such relationships will both provide new insights into the control of bacterial drug resistance and support the development of novel antibiotic drugs.

To investigate relationships between chemical structures and the acquisition of drug resistance, high dimensional datasets can be treated as two different datasets. To analyze such datasets, we applied canonical correlation analysis (CCA) and analyze the relationships between chemical substructures and phenotypic changes related to the acquisition of drug resistance, the mRNA transcript expression levels of 4444 genes from drug-resistant *E. coli* evolved *in vitro* for drug resistance to 10 antibiotic compounds. Structure data files (SDF) of 10 compounds were collected from ChEMBL and Morgan fingerprint with 2 radius was used to generate 2048 substructures from SDF in rdkit (ver. 2019.09.3) within Python.

To gather lots of features, consensus clustering was employed and proper cluster number were determined to be 5 for substructure and transcriptome. Average values in the same clusters were calculated to perform CCA.

From a two-dimensional scatter plot, we calculated two orthogonal canonical coefficient vectors under the correlation contributions of 0.901 and 0.822 for canonical coefficients 1 and 2, respectively. Overall, we observed clusters with large coefficient values in each axis. These findings suggest that these clusters represent different mechanisms of resistance for each class of antibiotics.

This work was supported in part by research grant for priority areas from Shiga University.