Oral Presentation 2

日時：2020 年 10 月 29 日 15:45-17:15
チャンネル：２

口頭発表２
『創薬応用（その他）』
Drug Discovery application

座長：　本間 光貴 （理化学研究所）
Teruki Honma (RIKEN)

O2-01
"Phenotypic Network Screening: A New High Throughput Screening Platform"
Hiroshi Kawai
(Socium Inc.)

O2-02
"Development of Monte Carlo Tree Search Based small organic compound generation system"
Daiki Erikawa
(Tokyo Institute of Technology, School of Computing)

O2-03
"Development of SSDB software for reproducing the ChEMBL database and its current status"
Akihiro Masuda
(Department of Biological Sciences, Graduate School of Science and Engineering, Chuo University)

O2-04
"Development of new AI-based cell visualization technology that can solve problems in drug discovery and regenerative medicine"
Tamio Mizukami
(Nagahama Institute of Bio-Science and Technology)

O2-01

Phenotypic Network Screening: A New High Throughput Screening Platform

Hiroshi Kawai[1]
hiroshi.kawai@socium.co.jp

Masashi Kitazawa[1]
masashi.kitazawa@socium.co.jp

Takayuki Kiboku[1]
takafumi.kiboku@socium.co.jp

Katsuhisa Horimoto[1,2,3]
katsuhisa.horimoto@socium.co.jp

[1] Socium Inc, AIST Waterfront build. 5F, 2-3-26, Aomi, Koto-ku, Tokyo,135-0064, Japan
[2] International Medical Center, Saitama Medical University, 1397-1, Yamane, Hidaka-City, Saitama-Pref, 350-1298, Japan
[3] Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, AIST Waterfront build. 10F, 2-3-26, Aomi, Koto-ku, Tokyo,135-0064, Japan

Keywords: high throughput screening, network analysis

The relation between gene expression and drug efficacy was uncovered by Broad Institute in the connectivity map [1] and the following LINCS [2]. It was embodied that the relation of gene expression between normal and disease was reverse to that before and after drug treatment even for a few kinds of commercial cell lines, by Gene Set Enrichment Analysis (GSEA) [3] that estimates the distribution bias of a set of genes against whole gene distribution. Due to GSEA methodology, the Broad system needs ca. 1000 genes to estimate its reverse relation. Therefore, it seems difficult to apply the relation to the high throughput screening (HTS) in terms of time and cost.

The basic approaches for recent HTS are divided into two trends: in silico screening and phenotype screening. The former can obtain computational solutions for optimized structure fitting between chemical compounds and a target protein, but should define the target protein before screening. The latter can obtain experimental results for drug efficacy, but demands frequently further different experiments to clarify the mechanism of drug efficacy.

Here, we have developed a new HTS platform, titled "*Phenotypic Network Screening*", to compensate for the pitfalls of the two screening approaches based on the reverse relation between disease and drug. The point is that we can uncover its relation by expression data of only 20 genes, instead of hundreds of genes: we estimate the drug candidates negatively correlated with disease by network analysis [4], followed by exploration of the marker of 20 genes from the gene signature between normal and patient with a target disease by our original method [5]. We have assessed the potential of our screening platform by means of diabetes, and the platform has successfully enabled us to select 4 of 5 drugs for diabetes among 3781 drugs using their expression data in commercial cell lines. This result indicates that our platform is feasible for HTS, concomitantly with the estimation of its mechanisms and of its stratification marker.

[1] Lamb, J., *et al.*, The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313:1929–1935, 2006.
[2] http://lincsproject.org/LINCS/
[3] Subramanian, A., *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, 102:15545-15550, 2005.
[4] Horimoto, K. Fukui, K. Device, method and program for discovering new drugs. Patent Number 6550571, Registered Jul. 12, 2019 in Japan.
[5] Horimoto, K. Fukui, K. Device, method and program for exploring biomarkers. Patent Number 6270221, Registered Jan. 30, 2018 in Japan and Aug. 9, 2018 in US

O2-02

Development of Monte Carlo Tree Search based small organic compound generation system

Daiki Erikawa[1] erikawa.d.aa@m.titech.ac.jp
Nobuaki Yasuo[2] yasuo.n.aa@m.titech.ac.jp
Masakazu Sekijima[1,2] sekijima@c.titech.ac.jp

[1] Tokyo Institute of Technology, School of Computing
[2] Tokyo Tech Academy for Convergence of Materials and Informatics

Keywords: Molecular Design, Machine Learning

Development of a new drug takes 13.5 years and costs an average of 2.6 billion dollars. Information technology has been used by many scientists to tackle the problem of the enormous amount of time and money involved in developing new drugs. In the process of new drug development, optimization tasks are important, starting with specific molecules and searching for molecules with desirable properties. In recent years, machine learning-based approaches have received much attention in the field of molecular design, and various methods such as VAE and GAN have been applied to molecular design. In this work, we have developed a method for generating derivatives of arbitrary molecules using MCTS (Monte Carlo tree search) and RNN (recurrent neural networks). A new molecule is generated by replacing a portion of the original SMILES with a partial SMILES generated by MCTS. In order to evaluate the performance of our method, optimizations of the octanol-water partition coefficient (LogP) and drug-likeness (QED) were carried out on molecules obtained from the ZINC database. As a result, well optimized molecules were generated for both QED and LogP as derivatives of specific molecules. Our approach can use any metric that can be calculated from SMILES and can be applied to a variety of fields, not just drug discovery.

O2-03

Development of SSDB software for reproducing the ChEMBL database and its current status

Akihiro Masuda[1]
a15.783r@g.chuo-u.ac.jp

Hideaki Umeyama[2]
umeyama@bio.chuo-u.ac.jp

Mitsuo Iwadate[1,2]
iwadate@bio.chuo-u.ac.jp

[1] Department of Biological Sciences, Graduate School of Science and Engineering, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo,112-8551, Japan
[2] Department of Biological Sciences, Faculty of Science and Engineering, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo,112-8551, Japan

Keywords: SSDB, Hydrophobic Interaction, Empirical Potential

The software for SSDB was developed considering both database search and physical interaction. The software was developed by combining Pyhon and RDkit library to easily add new features to my previous ChooseLD program. The function of hydrophobic interaction was added. This is a time-consuming process, but an attempt to save time is introduced. Many of the software's internal variables are in the process of being optimized in the ChEMBL database to make it suitable for the purpose of fast, comprehensive drug screening. A vast amount of data optimization is in progress, and we describe the current state of the art.

[1] Landrum G. RDKit http://www.rdkit.org
[2] Takaya, D., Takeda-Shitaka, M., Terashi, G., Kanou, K., Iwadate, M., and Umeyama, H., Bioinformatics based Ligand-Docking and in-silico screening, *Chem Pharm Bull(Tokyo)*, 56:742-744, 2008.
[3] Gaulton, A., Hersey, A., Nowotka, M., Bento, AP., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, LJ., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, MP., Overington, JP., Papadatos, G., Smit, I., and Leach, AR. The ChEMBL database in 2017. *Nucleic Acids Res.*, 45:D945-D954, 2017.

O2-04

Development of new AI-based cell visualization technology that can solve problems in drug discovery and regenerative medicine

Tamio Mizukami[1, 2]

mizukami@nagahama-i-bio.ac.jp

[1] Nagahama Institute of Bio-Science and Technology, 1266 Tamura, Nagahama, Shiga, 526-0829, Japan

[2] Frontier Pharma Inc., 1281-8 Tamura, Nagahama, Shiga, 526-0829, Japan

Keywords: AI drug discovery, Cell assay, Cell Imaging

Cell assays are widely used for drug discovery screening and activity evaluation, and are core and essential research tools for the drug discovery process. Recently, image analysis of cells using a high-content imaging system based on a fluorescence microscope has become widespread, and by using it, it is possible to observe cell characteristics such as life and death, undifferentiated and differentiated state with high precision and quantitatively. It was therefore, the cell assay using this system has been widely used.

However, in the cell observation method based on the fluorescence microscope, there is an unavoidable problem that the cells are invaded by the fluorescent labeling and the irradiation of the excitation light, and continuous observation of living cells is basically difficult. Further, a complicated experimental process of fluorescent labeling is required, which is a factor of lowering reliability and reproducibility.

Further, in regenerative medicine, at present, analysis such as cell counting is performed by sampling a part of cells in the cell manufacturing process, but sampling of cells at an intermediate stage is complicated and often difficult. Therefore, development of a technique that enables noninvasive and continuous analysis of living cells is expected.

We have developed a new cell visualization technology that can solve the above-mentioned problems in drug discovery and regenerative medicine.

This new technology has the following advantages.

・A high-precision fluorescent image showing the cell properties can be instantaneously generated non-invasively from an image of transmitted light without any experimental work such as fluorescent labeling.

・No cell fixing/staining work is required and therefore no damage to cells.

・Non-invasive, real-time, continuous observation of living cells becomes possible.

In this presentation, I will introduce the outline of this technology and an example of verification of its usefulness.