SORS/WomenInBSC: Multi-scale cancer signaling network modeling using natural language processing

Abstract

Patient heterogeneity makes cancer treatment and drug development difficult and costly. Therefore, the development of computational methods to find prognostic markers for individualised treatment and drug screening is urgently needed. To address this problem, we employ both network/pathway-based simulation (mechanistic modelling) and molecular-based simulation (molecular dynamics (MD) simulation).

Mechanistic modeling of gene networks using ordinary differential equations (ODEs) is regarded as a promising approach to uncover regulatory mechanisms and identify drug targets in human diseases. We developed Pasmopy (Patient-Specific Modelling in Python), a computational framework for patient stratification based on *in silico* signalling dynamics. With this framework, we constructed a comprehensive mechanistic model of the ErbB receptor - c-Myc signalling network, trained on phospho-proteomics data from breast cancer cell lines. We then performed a simulation of 377 breast cancer patients using transcriptome data from The Cancer Genome Atlas (TCGA) as the model's initial value. Through this approach, we successfully predicted the key maker genes associated with poor prognosis of triple-negative breast cancer (TNBC) based on the time-course patterns of *in silico* signalling dynamics, using deep learning method.

In this study, we also developed a new computational tool called Text2model, which transforms the descriptions of biochemical reactions into mechanistic models. Building on Text2Model, we are currently developing natural language processing (NLP) methods to automatically construct mathematical models from the literature and public databases, and to predict gene-drug interactions.

In mechanistic modeling, parameters such as association or dissociation constants are often altered by gene mutation. However, predicting these parameters within cells remains challenging. We are working to address this by combining molecular-based simulations with network/pathway-based simulations. Specifically, we are currently investigating how EGFR mutations alter the parameters of EGFR-adaptor proteins binding and influence downstream ERK activity. I will explain more about our ongoing cancer modeling study in this seminar.

References

1. Imoto H, Yamashiro S, Okada M. A text-based computational framework for patient -specific modeling for classification of cancers. ***iScience***, 25(3):103944, 2022.
2. Arakane K, Imoto H, Ormersbach F, Okada M. Extending BioMASS to construct mathematical models from external knowledge. ***Bioinformatics Adv.*** 4 (1), vbae042, 2024.

3. Tsutsui M, Okada M. DynProfiler: a Python package for comprehensive analysis and interpretation of signaling dynamics leveraged by deep learning techniques. ***Bioinformatics Adv.*** 4 (1), vbae145, 2024.
4. Yamagiwa H, Hashimoto R, Arakane K, Murakami M, Soeda S, Oyama M, Zhu Y, Okada M, Shimodaira H. Predicting Drug-Gene Relations via Analogy Tasks with Word Embeddings. ***arXiv***:2406.00984, 2024.

Short Bio

Mariko Okada is Professor at the Institute for Protein Research, Osaka University (Japan) (2016-Present) and also Provost's Visiting Professor of Systems Biology, Imperial College London (2024-2025). Previously, she was a Team Leader at the RIKEN Institute ; RIKEN Genomic Sciences Center (GSC), Research Center for Allergy and Immunology (RCAI) and Center for Integrative Medical Sciences (IMS), followed by the Visiting Scientist at the University California, Davis and Researcher position at Japanese R&D of Novo Nordisk. She is a member of Science Council of Japan.

Speakers

Speaker: Mariko Okada, Professor at the Institute for Protein Research, Osaka University (Japan)
Host: Josep Lluís Gelpí, INB Computational Node 2 group manager
Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---