Department of Systems Biology, School of Biomedical Sciences
Department of Bioinformatics, Medical Research Institute

1. Staffs and Students

<table>
<thead>
<tr>
<th>Role</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor</td>
<td>Hiroshi Tanaka</td>
</tr>
<tr>
<td>Associate Professor</td>
<td>Yoshihito Niimura</td>
</tr>
<tr>
<td>Assistant Professor</td>
<td>Soichi Ogishima (<del>April), Kaoru Mogushi (June</del>)</td>
</tr>
<tr>
<td>Project Associate Professor</td>
<td>Fengrong Ren, Takako Takai (~April), Jun Nakaya (~March)</td>
</tr>
<tr>
<td>Project Lecturer</td>
<td>Kanae Oda (~March)</td>
</tr>
<tr>
<td>Project Assistant Professor</td>
<td>Takeshi Hase, Naoki Hasegawa (~March), Jun Nakaya (~March)</td>
</tr>
<tr>
<td>Technical Staff</td>
<td>Ken Miyaguchi (~March)</td>
</tr>
<tr>
<td>Graduate Students</td>
<td>Hironobu Yamaguchi, Ryosuke Ishiwata, Satoru Suzuki, Tadashi Urashima, Masataka Kikuchi, Hajime Sawai, Taro Kishimoto, Syed Ali Zaidi, Hiroaki Hasegawa, Noriaki Koizumi, Norihiko Inoue, Ko Watanabe, Asiya Hapaer, Tadashi Miyamoto, Jun-ya Hagiwara, Kasumi Otsubo</td>
</tr>
</tbody>
</table>

2. Purpose of Education

Prof. Tanaka is in charge of the education of medical informatics and bioinformatics. For undergraduates, he teaches “Clinical Informatics”, “Statistics for Health Sciences”, “Practice in Clinical Informatics II”, “Project Research”, and “Basics of Clinical Informatics”. For graduate students, he teaches “Computational Biology”, “Systems Pathology”, “Clinical Informatics”, “Integrated Bioinformatics”, “Integrated Translational Research”, and “Statistics for Nurses”. He supervises 31 students in total (18 PhD and one Master course students in Graduate School of Medical and Dental Sciences and six PhD and six Master course students in Biomedical Science PhD Program).

3. Research Subjects

Our mission is “system-level understanding of biological systems” in molecular biology and evolution (systems evolution) and medicine (omics-based medicine, systems pathology). Recently, the whole genome sequences of diverse organisms have become available. Moreover, various “omics” information such as a proteome, transcriptome, and metabolome are currently accumulating. Our goal is to establish a grand-theory of biological sciences from the viewpoint of “evolving networks composed of biological molecules” by integrating omics information. Genomic and omics data are also utilized in the field of medicine. It has been revealed that most diseases are caused by the interaction among abnormalities of multiple genes, those at the tissue level, and environments. It is therefore possible to consider diseases as a system. From this standpoint, we try to establish the omics-based medicine and systems pathology.

1) Analysis of disease mechanism using omics-based approaches

Recent advances in analysis techniques in molecular biology have led to the investigation of genome-wide data such as genome, transcriptome and proteome. In order to reveal the underlying biological mechanisms from such a large amount of “omics” data, integration of biomedical knowledge with multivariate statistical analysis or machine learning methods is one of the most crucial tasks for bioinformatics research. We have been performing collaborative research with our university hospital and other institutes mainly based on transcriptome analysis using DNA microarray, including the
following topics: 1) identification of diagnosis marker for prognosis prediction in hepatocellular carcinoma patients, 2) development of predictive marker for metastatic relapse in colorectal cancer, and 3) analysis of spinocerebellar ataxia and hepatocellular carcinoma using next generation sequencing technologies.

2) Systems pathology analyses on disease progression of cancer, metastasis, and Alzheimer’s disease

Our mission is systems pathology studies on cancer, metastasis (epithelial-mesenchymal transition: EMT), and neurodegenerative disease (Alzheimer’s disease) using large-scale molecular biology data, so-called omics data. We inferred transcriptional, gene regulatory, and protein interaction networks of disease progression, and then explored master regulator, that is key molecule in their networks. We then estimated an attractor for each cellular state based on gene regulatory network for disease progression, cellular transformation (EMT), and cellular differentiation (iPSC/ESC) processes, showing transition of attractors along with these processes. For omics data analyses, data integration is necessary. We worked on integration of incurable diseases data using Linked Data technology.

3) i2b2: A novel technology of clinical databases as an infrastructure of translational informatics

Translational informatics is an emerging research field of computational technology for facilitating translation of genome information into the clinical application. It targets collection and computation of clinical and genomic information on the basis of mathematical models for diseases. It is a part of promoted researches after the completion of human genome sequencing, which includes industry and academia partnership in drug development and patient-centered translational research. Among the ongoing projects, the i2b2 provides an ontology-based object-oriented database system for integration of clinical information dispersed in different laboratories and different hospitals. Due to its highly flexible data-schema, the i2b2 enables persons without expert knowledge of database to collect clinical information into a database. We constructed i2b2 database with 392 clinical patients’ data collected in the university hospital of Tokyo Medical and Dental University. The patients’ data includes biomedical and clinicopathological information extracted from carcinoma and non-carcinoma specimens of cancer patients recorded in ‘Integrated Clinical Omics Database’ (iCOD). We transferred 8,580 English and 54,579 Japanese descriptions into i2b2. We employed Japanese NLP technologies in order to extract clinical terms from doctors’ comments in Japanese free texts. We built a pipeline for extraction of clinical terms and translation of the extracted terms into English computationally.

4) Analyses of the human protein-protein interaction networks and their applications to drug discovery

Since proteins exert their functions through interaction to other proteins, networks of protein-protein interactions are inevitable to discover novel drug-target genes. To discover novel targets, it is of use to understand topological and statistical characteristics of protein-protein interaction networks (PINs), and how the target-genes are distributed over the PINs. To uncover the topological features of PINs, we developed a novel method to decompose a very large complex network into simple sub-networks. Our method decomposed the genome-wide human PIN into several small simple sub-networks, and mapped target-genes on to the sub-networks. Among the sub-networks, a sub-network contains almost 60% of target-genes of small molecule drugs (e.g., kinase inhibitors) for cancerous diseases. Further, pathway enrichment analyses revealed that genes in the sub-network are involved in cancer-related signaling pathways (e.g., vascular endothelial growth factor signaling pathway). These results indicate that the listing of genes and interactions in the sub-network may help drug companies to search more efficiently for mechanisms of drug action and novel target genes for cancerous diseases.

5) Diversity of olfactory receptor gene repertoires among mammals

Olfaction, the sense of smell, is essential for the survival of animals. Odor molecules in the environment are detected by olfactory receptors (ORs) encoded by a large multigene family. To investigate the diversity of OR gene repertoires among mammals, I extensively identified the OR genes from the draft genome sequences of 38 diverse mammals. The results demonstrated that the estimated numbers of functional OR genes are extremely variable, ranging from only ~10 in dolphins to ~2,000 in elephants. However, the number of functional OR genes is not correlated with the fractions of pseudogenes. Identification of orthologous gene sets among 13 eutherian mammals with the genome of deep coverage (>6x) revealed that hundreds of gene gains and losses have occurred during eutherian evolution, suggesting dynamic changes of OR gene repertoires depending on each species’ living environment. I also examined OR genes from two turtle species with the whole genome sequences, showing drastic class I OR gene expansion, which is characteristic to turtles among amniotes.

6) Omics Research about mechanism of liver cancer progression

The complete sequencing of the human genome has ushered in a new era of systems biology referred to as Omics. The “Omics” refers to the comprehensive analysis of biological systems. Likewise, the field of bioinformatics has grown in parallel and with the help of rapid data analysis and information exchange is now possible. We have been collecting clinical and Omics (Genomics, Transcriptomics, Proteomics, Epigenomics, etc.) data. This includes both comprehensive molecular
Omics information and comprehensive clinical information from almost 400 patients who has liver cancer at TMD-Hospital. In case of liver cancer, it is very difficult to find the related gene for subtypes of liver cancer, but could find possible relation using data cleaning and integrated analysis, along with molecular biological analysis. Omics will not only have an impact on our understanding of biological processes, but the prospect of more accurately diagnosing and treating disease will soon become a reality.

7) Inferring evolutionary dynamics of SIV/HIV-1 Vpu and its co-evolution with Nef and Tetherin

Many studies on the function changes of SIV/HIV-1 Vpu after cross-species transmission have been reported in recent years, but little is known about the evolutionary history of this accessory gene. To elucidate possible evolutionary mechanisms responsible for the functional change, we conducted a computational analysis to investigate the evolutionary dynamics of Vpu and also its co-evolution with SIV/HIV Nef and primate Tetherin. Eighty-seven Vpu genes, 108 Nef genes and 35 primate Tetherin genes were retrieved from public databases. The reconstructed phylogenetic tree of Vpu was consistent with those reported in previous studies. The positive selection detection revealed that both Vpu and Tetherin had experienced adaptive evolution. Importantly, the tRMCA of SIVcpz Nef was estimated to be more ancient than that of SIVcpz Vpu, suggesting that the recombination event might be an evolutionary force driving the function loss of Vpu in SIVcpz. Notably, two SIVcpz sub-clades, SIVcpzPtt and SIVcpzPts, showed very different features at both molecular and structural levels. These results provided important information on the Vpu evolution and its co-evolution with Nef and Tetherin, which would give a new insight into the studies of SIV/HIV Vpu in future.

4. Publications

Original Article


Reviews

Book chapters