# Bioinformatics

## 1. Staffs and Students

| | | |
|---|---|---|
| Professor: | Hiroshi Tanaka | |
| Associate Professor: | Yoshihito Niimura | |
| Assistant Professor: | Soichi Ogishima | |
| Visiting Professor: | Hiroki Nogawa (-July), | Hiroshi Mizushima (Oct-) |
| Visiting Assistant Professor: | Isao Yamaguchi (-Oct), | Yasen Mahmut (Oct-) |
| Project Associate Professor: | Jun Nakaya, | Fengrong Ren, |
| | Takako Takai | |
| Project Lecturer: | Kazuro Shimokawa, | Kanae Oda |
| Project Assistant Professor: | Takeshi Hase, | Naoki Hasegawa, |
| | Kaei Hiroi, | Keisuke Ido, |
| | Kaoru Mogushi, | Satoshi Shoji, |
| | Masaki Morioka | |
| Technical Staff: | Masaya Itoda, | Ken Miyaguchi, |
| | Shota Nemoto | |
| Graduate Students: | Isao Okada, | Takayuki Ohnishi, |
| | Emilio Campos, | Yuki Katayama, |
| | Yoshitomo Tanaka, | Shun-ya Takahashi, |
| | Hironobu Yamaguchi, | Daisuke Yamakata, |
| | Yoshiyuki Kaneko, | Ryosuke Ishiwata, |
| | Hideaki Takata, | Masakuni Shibata, |
| | Arihito Endo, | Todd Johnson, |
| | Satoshi Nagaie, | Junko Shibata, |
| | Yoshimi Naruo, | Yasuko Takahashi, |
| | Akiko Hatano, | Taika Muto, |
| | Mitsuteru Hoshina, | Yuki Tanaka, |
| | Izumi Nakahara, | Hiromi Matsumae, |
| | Izumi Yoshida, | Hyeryung Kim, |
| | Ken Miyaguchi, | Kumiko Iijima, |
| | Kyaw Tun, | Satoru Suzuki, |
| | Eiichi Ueno, | Naoko Kasahara, |
| | Tadashi Urashima, | Sakiko Ohta, |
| | Masataka Kikuchi, | Wanping Aw, |
| | Yasuha Tanaka, | Hajime Sawai, |
| | Asami Suzuki, | Kasuo Iida, |
| | Taro Kishimoto, | Tadashi Ebara, |
| | Yasunori Ohto, | Afsaneh Eslami, |
| | Kae Suzuki, | Risa Iijima, |
| | Akihide Ohka, | Reiko Yamaguchi, |
| | Satoshi Mizuno, | Norio Tanaka, |
| | Yukari Watanabe | |

## 2. Purpose of Education

Prof. Tanaka is charged with education of interdisciplinary medical informatics and bioinformatics. For undergraduate classes he educates “Clinical Informatics”, “Statistics for Health Science”, “Practice in Clinical Informatics II”, “Project Research”, and “Basics of Clinical Informatics”. For graduate classes he educates “Computational Biology”, “Bioinformatics Computation”, “Systems Pathology”, “Statistical Genetics and Medical Statistics”, “English Debate”, and “Practice in Global Linkage between University and Industry", “Integrated Bioinformatics”, “Applied Biological Chemistry”, and “Integrated Translational Research”. He supervises 31 students of PhD course and 2 students in Master course in Graduate School of Medical and Dental Sciences and 6 students of PhD course and 12 students in Master course in Biomedical Science PhD Program. He is a principle investigator of “Global Linkage Program between University and

Industry" granted by Support Program for Improving Graduate School Education. This program provides students with internship opportunities at international business firms to see real-world examples and global trends so as to envisage future needs. This program also provides students with specialist consultations which support them to define their career objectives. He is also a principle investigator of "International Educational Program for Interdisciplinary Disease Science" granted by Program for Accelerating Internationalization of Higher (University) Education. This program will form a global alliance of higher education institutes in Europe, the United States, and Asia and develop international cooperation in education. In this year this program established university alliances with Shanghai Center for Bioinformatics Technology, Freiburg University, and Ecole Nomale Supérieure de Lyon, ending up in 11 international university alliances sharing the philosophy of interdisciplinary disease science. Joint-degree education has started with a graduate student of Heidelberg University. Cooperative teaching courses have also started with allied universities in the three courses of translational research, bioinformatics, and biological chemistry. He is also a principle investigator of "Educational Program for Biomedical Omics Information Scientists" granted by Special Coordination Funds for Promoting Science and Technology. This program offers study opportunities to clinical doctors and medical technologists in learning about integration of life science and informational science into practical applications in medicine. This program also educates bioinformatians who have been active in their field and are planning to diversify their activities into medical science, offering them basic and practical knowledge in clinical medicine and drug discoveries. 67 people have graduated this program in total five years of financial support.

## 3. Research Subjects

In our laboratory, we conduct biological and medical researches from the viewpoint of Systems Biology.

Biological sciences: Recently, the whole genome sequences of diverse organisms have become available. Moreover, various "omix" 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 omix information.

Medical sciences – Genomic and omix 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 omix-based medicine.

### [1] Differences in degree dependent gene duplicability cause overall structure of protein-protein interaction networks

Protein-protein interaction networks (PINs) were believed to be disassortative networks. In such networks, hub-hub interactions are suppressed. It was postulated that disassortative structure minimizes unfavorable cross-talks between hub-centric modules and thus such structure might have been positively selected in evolution. However, in this study, we investigated several PINs from various eukaryotes and showed that disassortative structures are not common features among eukaryotes. By examining network growth model based on gene duplication and divergence, we found that a preferential duplication of low- and high-degree nodes can generate disassortative and non-disassortative networks, respectively. Moreover, we compared 55 proteomes in eukaryotes and revealed that if genes with low (or high) degrees have been preferentially duplicated, PINs become disassortative (or non-disassortative). Therefore, disassortative structures observed in PINs can be a byproduct of preferential duplications of low-degree genes and it is unnecessary to assume any selective forces on the overall structures in the PINs.

### [2] Omics-based study of disease mechanisms

Due to recent advances in life science research, comprehensive data such as genome, transcriptome, and proteome can be routinely obtained. In order to interpret such genome-wide data in clinical research, we need to apply bioinformatics analysis such as data mining, statistical analysis and machine learning in combination with existing biological and medical knowledge.

We focus on development and application of bioinformatics methodology and have been conducting collaborative works with several research laboratories including following topics: (1) identification of gene sets and their interaction networks associated with phenotypes and prognosis of hepatocellular carcinoma (HCC) patients, (2) expression analysis of Aurora kinase B and alternative variant forms in HCC, (3) analysis of HCV-associated gene expression and cell signaling pathways, (4) identification of IQGAP1 and vimentin as a key regulator genes in naturally occurring hepatotumorigenesis induced by oxidative stress, and (5) identification of MUC12 as a prognosis marker in colorectal cancer.

### [3] Evolution of olfactory receptor gene families

Olfaction is essential for the survival of animals. Versatile odor molecules in the environments are received by olfactory receptors (ORs), which form the largest multigene family in vertebrates. Identification of the entire repertories of OR genes from the whole genome sequences revealed that the numbers of OR genes vary enormously, ranging from ~1,200 in rats and ~400 in humans to ~150 in zebrafish and ~15 in pufferfish. Extensive phylogenetic analyses suggested that the numbers of gene gains and losses are extremely large in the OR gene family. It appears that OR gene repertoires dynamically changed depending on each organism's living environment. For example, higher primates equipped with a well-developed vision system have lost a large number of OR genes. Moreover, two groups of OR genes for detecting airborne odorants have greatly expanded after the time of terrestrial adaption in the tetrapod lineage, whereas fishes retain diverse repertoires of genes that were present in aquatic ancestral species. The origin of vertebrate OR genes can be traced back to the common ancestor of all chordate species, but insects, nematodes, or echinoderms utilize distinctive families of chemoreceptors, suggesting that chemoreceptor genes had evolved many times independently in animal evolution.

### [4] Systems evolutionary biology

Our mission is to understand both (1) evolution and (2) dynamics of biological systems based on omics data from the point of view of "systems evolutionary biology".

(1) Evolutionary studies on biological systems are to understand evolution of life not only as gene evolution but also as systems evolution. We are focusing on evolution of both transcriptional networks of development and large-scale protein interaction networks. The former is to analyze evolution of Hox transcriptional networks reconstructed by our novel promoter analysis, while the latter is to reveal functional modularity in protein network evolution.

(2) Dynamical studies of biological systems are studies for revealing mechanism of transcriptional regulation by developing novel algorithm for trend analysis on time-series microarray data and by developing novel 3D visualization application of hierarchical molecular network based on the central dogma.

### [5] Transdisease Omics analysis of cancer by SAGE

Recently, comprehensive information on various biomolecules such as genes and proteins (Omics information) can be obtained easily by rapid advance of the molecular biological experimental technique. Therefore, drawing out of clinical useful information is becoming possible by comparing these molecular information of various human diseases and revaluing the similarity between the diseases. We have analyzed comprehensive gene expression data of 11 human diseases obtained from GEO (Gene Expression Omnibus). In this research, SAGE(serial analysis of gene expression) data was chosen to analyze gene expression data, and the comparison analysis was performed among several cancer samples. As a result, we have found that the combination of gene expression pattern of Breast cancer and Prostate cancer are similar in the 11 diseases samples. We were able to find that there is a similar character for the malignant alteration, and it has a similar treatment method, when we focus on the common feature of these carcinoma.

## 4. Publications

### [Original Papers]


1. Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H: Structures of protein interaction network and their implications on drug design, PLoS Compt Biol, 5(10) e1000550, 2009
2. Hasegawa N, Sugiura W, Shibata J, Matsuda M, Ren F, Tanaka H: Inferring within-patient HIV-1 evolutionary dynamics under anti-HIV therapy using serial virus samples with vSPA, BMC Bioinformatics, 10:360, 2009
3. Niimura Y: On the origin and evolution of vertebrate olfactory receptor genes, Comparative genome analysis among 23 chordate species, Genome Biol. Evol, 1: 34-44, 2009.
4. Nagashima T, Ushikoshi-Nakayam R, Suenaga A, Ide K, Yumoto N, Naruo Y, Takahashi K, Saeki Y, Taiji M, Tanaka H, Tasai SF, Hatakeyama M: Mutation of epidermal growth factor receptor is associated with MIG6 expression, FEBS Journal, 276:5239-5251, 2009
5. Iwami S, Takeuchi Y, Iwamoto K, Naruo Y, Yasukawa M: A mathematical design of vector vaccine against autoimmune disease, J Theor Biol, 256:382-92, 2009.
6. Ishiwata RR, Morioka MS, Ogishima S, Tanaka H: BioCichlid: central dogma-based 3D visualization system of time-course microarray data on a hierarchical biological network. Bioinformatics, 15 February, 25:543-544, 2009
7. Yasen M, Mizushima H, Mogushi K, Obulhasim G, Miyaguchi K, Inoue K, Makahara I, Ohta T, Aihara A, Tanaka S, Arii S, Tanaka H: Expression of Aurora B and their Alternative Variant Forms in Hepatocellular Carcinoma and

the Adjacent Tissue, Cancer Science, 100:472-480, 2009
8. Ota MS, Kaneko Y, Kondo K, Ogishima S, Tanaka H, Eto K, Kondo T: In silico and in vitro analyses reveal role of Hes1 in taste cell differentiation, PLoS Genetics, 5: e1000443, 2009
9. Tanaka S, Mogushi K, Yasen M, Noguchi N, Kubo A, Kurokawa T, Nakamura N, Inazawa J, Tanaka H, Arii S: Surgical contribution to recurrence-free survival in patients with macrovascular invasionnegative hepatocellular carcinoma, Journal of the American College of Surgeons, 208:368-374, 2009
10. Ren F, Tanaka H, Yang Z: A likelihood look at the supermatrix-supertree controversy, Gene, 441:119-125, 2009
11. Iwatani Y, Chan D S, Liu L, Yoshii H, Shibata J, Yamamoto N, Levin J G, Gronenborn A M, Sugiura W: HIV-1 Vif-mediated ubiquitination/degradation of APOBEC3G involves four critical lysine residues in its C-terminal domain, Proc Natl Acad Sci U S A, 106(46): 19539-44, Nov 17, 2009
12. Okamoto E, Fujii H, Tanaka H, Yamakata D, Nobutomo K, Nagata H: Development of an IT infrastructure under Japan's Health Care Reform 2008: a potential for regional health information networks, Jpn J Med Inf, 28:93-98, 2009
13. Tun K, Rao RK, Samavedham L , Tanaka H, Dhar PK: Rich can get poor: conversion of hub to non-hub proteins, Systems and Synthetic Biology, DOI 10.1007/s 11693-009-9024-9, 2009
14. Ohashi W, Tanaka H: Benefits of pharmacogenomics in drug development - earlier launch of drugs and less adverse events, Journal of Medical Systems, DOI 10.1007/s10916-009-9284-7, 2009
15. Suzuki A, Takai-igarashi T, Numabe Y, Tanak H: Development of a database and ontology for pathogenic pathways in periodontitis, In Silico Biol, 9:1-11, 2009
16. Watanabe K, Kurihara Y, Tanaka H: Ubiquitous Health Monitoring at Home-Sensing of Human Biosignals on Flooring, on Tatami Mat, in the Bathtub, and in the Lavatory, IEEE SENSORS JOURNAL, 9:1847-1855, 2009
17. Fujibuchi W, Kim H, Okada Y, Taniguchi T, Sone H: High-performance gene expression module analysis tool and its application to chemical toxicity data, Methods Mol Biol, 577:55-65, 2009

[Reviews]

1. Niimura Y: Evolutionary dynamics of olfactory receptor genes in chordates, Interaction between environments and genomic contents, Human Genomics 4: 107-118, 2009