Human Genome Center

Laboratory of Genome Database
ゲノムデータベース分野

Professor	Minoru Kanehisa, Ph.D.	教　授	理学博士	金　久　　實
Research Associate	Toshiaki Katayama, M.Sc.	助　教	理学修士	片　山　俊　明
Research Associate	Shuichi Kawashima, M.Sc.	助　教	理学修士	川　島　秀　一

Since the completion of the Human Genome Project, high-throughput experimental projects have been initiated for uncovering genomic information in an extended sense, including transcriptomics, proteomics metabolomics, glycomics, and chemical genomics. We are developing a new generation of databases and computational technologies, beyond the traditional genome databases and sequence analysis tools, for making full use of these divergent and ever-increasing amounts of data, especially for medical and pharmaceutical applications.

1. KEGG DRUG and KEGG DISEASE

Minoru Kanehisa

KEGG is a database of biological systems that integrates genomic, chemical, and systemic functional information. It is widely used as a reference knowledge base for understanding higher-order functions and utilities of the cell or the organism from genomic information. Although the basic components of the KEGG resource are developed in Kyoto University, this Laboratory in the Human Genome Center is responsible for the applied areas of KEGG, especially in medical and pharmaceutical sciences. We consider diseases as perturbed states of the molecular system that operates the cell and the organism, and drugs as perturbants to the molecular system. We develop a new disease information resource, KEGG DISEASE (http://www.genome.jp/kegg/disease/), which is intended for use by computational analysis rather than just for humans to read and understand. When the detail of the molecular system is relatively well characterized, we draw KEGG pathway maps. When the detail is not known but disease genes are identified, we create KEGG DISEASE entries, each of which contains a list of known disease genes and other relevant molecules including environmental factors, diagnostic markers, and therapeutic drugs. The list simply defines the membership to the underlying molecular system, but is still useful for computational analysis. We also develop a comprehensive drug information resource, KEGG DRUG (http://www.genome.jp/kegg/drug/), containing chemical structures and/or chemical components of all prescription and OTC drugs in Japan, and most prescription drugs in the USA and Europe. In KEGG DRUG we also capture knowledge on two types of molecular networks. One is the interaction network of drugs with target molecules, metabolizing enzymes, transporters, other drugs, and the pathways involving all these molecules. The other is the chemical structure transformation network in the history of drug development where drug structures have been continuously modified by medicinal chemists. Furthermore, crude drug information in KEGG DRUG is being linked to metabolic pathways and metabolic compounds, especially in plants.

2. KEGG OC: Automatic assignments of orthologs and paralogs in complete genomes

Toshiaki Katayama, Shuichi Kawashima, Akihiro Nakaya[1], Minoru Kanehisa: [1]Graduate School of Frontier Sciences, The University of Tokyo

The increase in the number of complete genomes has provided clues to gain useful insights to understand the evolution of the gene universe. Among the KEGG suites of databases, the GENES database contains more than 4.2 million genes from over 1,100 organisms as of January 2010. Sequence similarities among these genes are calculated by all-against-all SSEARCH comparison and stored in the SSDB database. Based on those databases, the ORTHOLOGY database has been manually constructed to store the relationships among the genes sharing the same biological function. However, in this strategy, only the well known functions can be used for annotation of newly added genes, thus the number of annotated genes is limited. To overcome this situation, we have developed a fully automated procedure to find candidate orthologous clusters including those without any functional annotation. The method is based on a graph analysis of the SSDB database, treating genes as nodes and the Smith-Waterman sequence similarity scores as edge weights. The cluster is found by our heuristic method for finding quasi-cliques, but the SSDB graph is too large to perform quasi-clique finding at a time. Therefore, we introduce a hierarchy (evolutionary relationship) of organisms and treat the SSDB graph as a nested graph. The automatic decomposition of the SSDB graph into a set of quasi-cliques results in the KEGG OC (Ortholog Cluster) database. We have built a system that performs automatic update of KEGG OC, which can be run on a weekly basis. As a result, we obtained 830,320 clusters including 496,924 singleton clusters from 4,958,080 protein coding genes. Among them, only 5,489 clusters were shared across kingdoms and other clusters were kingdom specific. The automatic classification of our ortholog clusters is largely consistent with the manually curated ORTHOLOGY database. A web interface to search and browse genes in clusters is made available at http://oc.kegg.jp/.

3. EGENES: A database for expressed sequence tag indices of plant species

Shuichi Kawashima, Yuki Moriya[1], Toshiaki Tokimatsu[1], Susumu Goto[1] and Minoru Kanehisa: [1]Institute for Chemical Science, Kyoto University

EGENES is a knowledge-based database for efficient analysis of plant expressed sequence tags (ESTs), which was recently been added to KEGG PLANT. It links plant genomic information to higher order functional information in KEGG. The genomic information in EGENES is a collection of EST contigs constructed from assembled plant ESTs by using EGassembler. The EST indices are automatically annotated with the KEGG Orthology identifiers (K numbers) by KEGG Automatics Annotation Server (KAAS). Currently, EGENES contains 3,197,938 sequence catalogues in 78 plants, among which 25% have assigned K numbers. EGENES is available at http://www.genome.jp/kegg/catalog/org_list2.html

4. KEGG API: SOAP/WSDL interface for the KEGG system

Shuichi Kawashima, Toshiaki Katayama and Minoru Kanehisa

KEGG is a suite of databases and associated software, integrating our current knowledge of molecular interaction/reaction pathways and other systemic functions (PATHWAY and BRITE databases), the information about the genomic space (GENES database), and information about the chemical space (LIGAND databases). To facilitate large-scale applications of the KEGG system programmatically, we have been developing and maintaining the KEGG API as a stable SOAP/WSDL based web service. The KEGG API is available at http://www.genome.jp/kegg/soap/.

5. KEGG DAS: Comprehensive repository for community genome annotation

Toshiaki Katayama, Mari Watanabe and Minoru Kanehisa

KEGG DAS is an advanced genome database system providing DAS (Distributed Annotation System) service for all bacterial organisms in the GENOME database in KEGG. Currently, KEGG DAS contains over 10 million annotations assigned to the genome sequences of 1019 organisms (increased from 817 organisms in last year). The KEGG DAS server provides gene annotations linked to the KEGG PATHWAY and LIGAND databases. In addition to the coding genes, information of non-coding RNAs predicted using Rfam database is also provided to fill the annotation of the intergenic regions of the genomes. The KEGG DAS service is avail-

able at http://kegg das.hgc.jp/.

6. Semantic integration of biological databases

Toshiaki Katayama, Mitsuteru Nakao[1], Tatsuya Nishizawa[2], Shuichi Kawashima: [1]Database Center for Life Science, ROIS, [2]Information and Mathematical Science Laboratory Inc.

Web service takes an important role in virtual integration of various bioinformatics resources. However, as web services become widely used in the bioinformatics analysis, incompatibilities in interfaces and data types among services come into critical problems which prevent user from fully utilizing these services in combination. Therefore, we developed the TogoWS service (http://togows.dbcls.jp/) to provide an integrated interface with advanced features. In the TogoWS REST API, we introduced a unified access method to major database resources by intuitive URIs that are mapped to search, retrieve, parse and convert the database entries. Meanwhile, the TogoWS SOAP API is developed to resolve compatibility issues found in the server and client side of the SOAP implementations. We also monitor the availability of major web service providers including KEGG, PDBj and DDBJ so that user can track temporal down of these services. Recently, the Semantic Web is getting considered as a key technology to integrate heterogeneous datasets as the Linked Data. Some biological databases including UniProt are already provided in the RDF format which requires information on the web explicitly encoded in a machine-readable syntax. To explore the feasibility of these technologies in bioinformatics, we organized an international workshop, BioHackathon (http://hackathon3.dbcls.jp/), gathering major database providers and software developers. As a consequence, we started (1) to extend our TogoDB service (http://togodb.dbcls.jp/), which hosts various biological databases deposited by researchers, to provide its contents also in the RDF format, and (2) to provide converters to RDF for databases supported in the TogoWS service. This makes both major databases and private databases can be seamlessly combined to discover hidden relations.

7. Draft genome sequence of *Rammazzottius cf. varieornatus*

Toshiaki Katayama, Kazuharu Arakawa[1], Takekazu Kunieda[2], Atsushi Toyoda[3], Tadasu Shin-I[3], Daiki D. Horikawa[4], Hirokazu Kuwahara (University of Tokyo, Japan), Kazuko Ohishi[3], Ayuko Motoyama[3], Tomoyuki Aizu[3], Minoru Kanehisa, Yuji Kohara[3], Asao Fujiyama[5]: [1]Institute of Advanced Bioscience, Keio University, [2]Graduate School of Science, The University of Tokyo, [3]National Institute of Genetics, ROIS, [4]Ames Research Center, NASA, [5]National Institute of Informatics, ROIS

Tardigrades are extremotolerant animals that can enter an ametabolic dry state called anhydrobiosis and have high tolerance to a variety of extreme environmental conditions, such as high (151°C) and low (−273°C) temperatures, high (7.5GPa) and low (open space vacuum) pressures, and radiations X-rays (10000Gy) and Gamma-rays (6000Gy), particularly while in anhydrobiosis. To understand the molecular basis of these tolerance, we sequenced the genome of an extremotolerant tardigrade, *Rammazzottius cf. varieornatus*, YOKOZUNA-1 strain. The draft genome sequence is assembled from 772,854 whole genome shotgun reads and additional 42,592 paired-end reads of fosmid clones. As a result, ~3000 super-contigs are obtained, in which 95% of the sequences are covered by the top 30 super-contigs. Total amount of the assembled sequence is about 58.3Mb covering approximately to >98% of the estimated 60Mb genome size. The average GC-content is 47.5% which is relatively higher than that of D. melanogaster (41%) and C. elegans (35%). Based on the assembled sequences, we then applied sequence similarity search against UniProt and KEGG databases to estimate conserved protein sequences. Additionally, rRNA and tRNA sequences are predicted by RNAmmer and tRNAscan SE programs respectively. We predicted protein coding genes using *ab initio* method implemented in SNAP whereby we obtained about 15,000 candidate genes. To examine the difference of gene expressions between the active state and the dry state, we also performed transcriptome analysis by the Illumina sequencer and obtained 160 million reads. By mapping these short read sequences to the assembled genome, we estimates that expressions of ~7,000 genes and >10,000 transcripts are captured.

8. Characterization of alpha-phosphoglucomutase isozymes from *Toxoplasma gondii*.

Mihoko Imada[1], Shuichi Kawashima, Minoru Kanehisa, Tsutomu Takeuchi[1], Takashi Asai[1]: [1]School of Medicine, Keio University

The *Toxoplasma gondii* genome project has revealed two putative isoforms (TgPGM-I and TgPGM-II) of alpha-phosphoglucomutase (EC

5.4.2.2). We obtained recombinant proteins of these isoforms from the Beverley strain of *T. gondii* and their properties were characterized, particularly the kinetic properties of these isoforms. The specific activities of TgPGM-I and TgPGM-II for alpha-d-glucose 1-phosphate were 338+/−9 and 84+/−6mumol/min/mg protein, respectively, at 37 degrees C under optimal conditions. The Kcat and Km values of TgPGM-I were 398+/−11/s and 0.19+/−0.03mM and those for TgPGM-II were 93+/−7/s and 3.53+/−0.91mM, respectively, for alpha-d-glucose 1-phosphate. Magnesium ions were the most effective divalent cations for both the enzyme activities. The maximum activities of both the enzymes were obtained in the presence of more than 0.2 mM alpha-d-glucose 1,6-bisphosphate. Although both enzymes were attached to the alpha-phosphohexomutase superfamily, amino acid sequence homology between TgPGM-I and TgPGM-II showed very low overall identity (25%). No alpha-phosphomannomutase (EC 5.4.2.8) activity was detected for either enzyme. The data indicated that TgPGM-I, but not TgPGM-II, may play an important role in alpha-d-glucose 6-phosphate production.

9. EST analysis of genes that are expressed in the larva of tsutsugamushi mite, *Leptotrombidium scutellare*

Shuichi Kawashima, Junichi Watanabe[1], Minoru Kanehisa Atsushi Yamashita[2], Masahira Hattori[3], Susumu Yamamoto[4], Shinichi Noda[5]: [1]Division of Molecular Cell Signaling, Institute of Medical Science, The University of Tokyo, [2]Kitasato Institute for Life Science, Kitasato University, [3]Graduate School of Frontier Sciences, The University of Tokyo, [4]Faculty of Medicine, Kagoshima University, [5]Research Center for the Pacific Island, Kagoshima University

Chigger mites are vectors of Orientia tsutsugamushi which is the causative agent of tsutsugamushi disease (scrub typhus). Only the first-stage larvae (newly hatched chigger mites) attach to various animals including human and sometimes transmit O. tsutsugamushi to the hosts. To begin to decipher the molecular mechanisms in the larvae, that underlie feeding and infection behavior on the host, we sequenced ESTs of *Leptotrombidium scutellare*, one of the well-studied chigger mites in Japan. A cDNA library was constructed from *L. scutellare* larvae collected at Kagoshima prefecture. Sequencing of 5,214 cDNA clones resulted in 3,011 unique sequences consisting of 548 contings and 2,463 singletons. Approximately 55 % of the unique sequences were potentially homologous to proteins of *Ixodes scapularis*, which is currently only species whose proteome information is available among the Acari. Likewise, BLAST search to UniProt revealed 2,908 significant hits, that is 103 sequences are likely to be unique for L. scutellare. We assigned KEGG ORHOLOGY IDs (KO) to 986 sequences (634 unique KOs), which includes 342 KOs related to metabolic pathways. These results will be available in the FullMite database (http://fullmite.hgc.jp/).

10. HiGet and SSS: Search engines for the large-scale biological databases

Toshiaki Katayama, Shuichi Kawashima, Kazuhiro Ohi[1], Kenta Nakai[2] Minoru Kanehisa: [1]Hitatchi Ltd., [2]Laboratory of Functional Analysis In Silico, Institute of Medical Science, The University of Tokyo

Recently, the number of entries in biological databases is exponentially increasing year by year. For example, there were 10,106,023 entries in the GenBank database in the year 2000, which has now grown to 120,023,234 (Release 175+ daily updates). In order for such a vast amount of data to be searched at a high speed, we have developed a high performance database entry retrieval system, named HiGet. For this purpose, the system is constructed on the HiRDB, a commercial ORDBMS (Object-oriented Relational Database Management System) developed by Hitachi, Ltd. HiGet can perform full text search on various biological databases including GenBank, RefSeq, UniProt, Prosite, OMIM and PDB. Additional advantage of the HiGet system is the capability of a field specific search, which enables users to narrow down the number of results, especially useful for collecting sequences of their specific needs. We have also developed a sequence similarity search (SSS) service to find homologous sequences with various algorithms including BLAST, FASTA, SSEARCH, TRANS, and EXONERATE. This variety of options is unique among the public services and users can select an appropriate method to search similar sequences according to their query. Because algorithms such as TRANS and EXONERATE are highly time consuming, the SSS service is backended by the distributed computing environment with the Sun Grid Engine in our super computer system. HiGet and SSS services are available at http://higet.hgc.jp/ and http://sss.hgc.jp/ respectively.

Publications

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research* 38: D355-D360, 2010

Shigemizu, D., Araki, M., Okuda, S., Goto, S., Kanehisa, M. Extraction and analysis of chemical modification patterns in drug development. *J. Chem. Inf. Model*. 49: 1122-1129, 2009

Yamanishi, Y., Hattori, M., Kotera, M., Goto, S., Kanehisa, M. E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. *Bioinformatics* 25: i179-i186, 2009

Wheelock, C.E., Wheelock, A.M., Kawashima, S., Diez, D., Kanehisa, M., van Erk, M., Kleemann, R., Haeggström, J.Z., Goto, S. Systems biology approaches and pathway tools for investigating cardiovascular disease. *Mol. Biosyst*. 5: 588-602, 2009

Kawashima, T., Kawashima, S., Tanaka, C., Murai, M., Yoneda, M., Putnam, N.H., Rokhsar, D.S., Kanehisa, M., Satoh, N., Wada, H. Domain shuffling and the evolution of vertebrates. *Genome Research* 19: 1393-403, 2009

Imada, M., Kawashima, S., Kanehisa, M., Takeuchi, T., Asai, T. Characterization of alpha-phosphoglucomutase isozymes from Toxoplasma gondii. *Parasitology International*, 2010

Onuki, R., Shibuya, T., Kanehisa, M. New kernel methods for phenotype prediction from genotype data. *Genome Informatics*, 22: 132-141, 2009

Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., and Kanehisa, M. Characterization and classification of adverse drug interactions. *Genome Informatics*, 22: 167-175, 2009

Tanaka, M., Moriya, Y., Goto, S., and Kanehisa, M. Analysis of a lipid biosynthesis protein family and phospholipid structural variations. *Genome Informatics*, 22: 191-201, 2009

Human Genome Center

Laboratory of DNA Information Analysis
Laboratory of Sequence Data Analysis
DNA情報解析分野
シークエンスデータ情報処理分野

Professor	Satoru Miyano, Ph.D.	教　授	理学博士	宮　野	悟
Associate Professor	Seiya Imoto, Ph.D.	准教授	博士(数理学)	井　元　清	哉
Assistant Professor	Masao Nagasaki, Ph.D.	助　教	博士(理学)	長　崎　正	朗
Project Assistant Professor	Yoshinori Tamada, Ph.D.	特任助教	博士(情報学)	玉　田　嘉	紀
Associate Professor	Tetsuo Shibuya	准教授	博士(理学)	渋　谷　哲	朗
Lecturer	Rui Yamaguchi, Ph.D.	講　師	博士(理学)	山　口	類

The recent advances in biomedical research have been producing large-scale, ultra-high dimensional, ultra-heterogeneous data. Due to these post-genomic research progresses, our current mission is to create computational strategy for systems biology and medicine towards translational bioinformatics. With this mission, we have been developing computational methods for understanding life as system and applying them to practical issues in medicine and biology.

1. Systems Biology

a. Unraveling dynamic activities of autocrine pathways that control drug-response transcriptome networks

Yoshinori Tamada, Hiromitsu Araki[1], Seiya Imoto, Masao Nagasaki, Atsushi Doi[1], Yukiko Nakanishi[1], Yuki Tomiyasu[1], Kaori Yasuda[1], Ben Dunmore[2], Deborah Sanders[2], Sally Humphries[2], Cristin Print[3], D. Stephen Charnock-Jones[2], Kousuke Tashiro[4], Satoru Kuhara[4], Satoru Miyano: [1]GNI Ltd., [2]University of Cambridge, [3]University of Auckland, [4]Kyushu University

Some drugs affect secretion of secreted proteins (e.g. cytokines) released from target cells, but it remains unclear whether these proteins have direct effects to target cells, like autocrine loops, or not. In this study, we propose a computational method for validating a biological hypothesis: there exist autocrine signaling pathways that are dynamically induced by drug response transcriptome networks and control them simultaneously. If such pathways are identified, they could be very useful for revealing drug mode-of-action and identifying novel drug targets. By the node-set separation method newly proposed, dynamic structural changes can be embedded in transcriptome networks that enable us to find master-regulator genes or critical paths at each observed time. We then combine the protein-protein interaction network with the estimated dynamic transcriptome network to discover drug-affected autocrine pathways if they exist. The statistical significance (pvalues) of the pathways are evaluated by the meta-analysis technique. The dynamics of the interactions between the transcriptome networks

and the signaling pathways will be shown in this framework. We illustrate our strategy by an application using anti-hyperlipidemia drug, Fenofibrate. From over one million protein-protein interaction pathways, we extracted significant 22 autocrine-like pathways with the Bonferroni correction, including VEGF-NRP1-GIPC1-PRKCA-PPAR, that is the most significant and contains PPAR, a target of Fenofibrate.

b. Analysis of PPARα-dependent and PPARα-independent transcript regulation following fenofibrate treatment of human endothelial cells

Hiromitsu Araki[1], Yoshinori Tamada, Seiya Imoto, Ben Dunmore[2], Deborah Sanders[2], Sally Humphrey[2], Masao Nagasaki, Atsushi Doi[1], Yukiko Nakanishi[1], Kaori Yasuda[1], Yuki Tomiyasu[1], Kousuke Tashiro[4], Cristin Print[3], D. Stephen Charnock-Jones[2], Satoru Kuhara[4], Satoru Miyano

Fenofibrate is a synthetic ligand for the nuclear receptor peroxisome proliferator-activated receptor (PPAR) alpha and has been widely used in the treatment of metabolic disorders, especially hyperlipemia, due to its lipid-lowering effect. The molecular mechanism of lipid-lowering is relatively well defined: an activated PPARα forms a PPAR-RXR heterodimer and this regulates the transcription of genes involved in energy metabolism by binding to PPAR response elements in their promoter regions, so-called "trans-activation". In addition, fenofibrate also has anti-inflammatory and anti-athrogenic effects in vascular endothelial and smooth muscle cells. We have limited information about the anti-inflammatory mechanism of fenofibrate; however, "trans-repression" which suppresses production of inflammatory cytokines and adhesion molecules probably contributes to this mechanism. Furthermore, there are reports that fenofibrate affects endothelial cells in a PPARα-independent manner. In order to identify PPARα-dependently and PPARα-independently regulated transcripts, we generated microarray data from human endothelial cells treated with fenofibrate, and with and without siRNA-mediated knock-down of PPARα. We also constructed dynamic Bayesian transcriptome networks to reveal PPARα-dependent and -independent pathways. Our transcriptome network analysis identified growth differentiation factor 15 (GDF15) as a hub gene having PPARα-independently regulated transcripts as its direct downstream children. This result suggests that GDF15 may be PPARα-independent master-regulator of fenofibrate action in human endothelial cells.

c. Simulation-based model checking approach to cell fate specification during *Caenorhabditis elegans* vulval development by hybrid functional Petri net with extension.

Chen Li, Masao Nagasaki, Kazuko Ueno, Satoru Miyano

Model checking approaches were applied to biological pathway validations around 2003. Recently, Fisher et al. have proved the importance of model checking approach by inferring new regulation of signaling crosstalk in *C. elegans* and confirming the regulation with biological experiments. They took a discrete and state-based approach to explore all possible states of the system underlying vulval precursor cell (VPC) fate specification for desired properties. However, since both discrete and continuous features appear to be an indispensable part of biological processes, it is more appropriate to use quantitative models to capture the dynamics of biological systems. Our key motivation of this research is to establish a quantitative methodology to model and analyze *in silico* models incorporating the use of model checking approach. A novel method of modeling and simulating biological systems with the use of model checking approach is proposed based on hybrid functional Petri net with extension (HFPNe) as the framework dealing with both discrete and continuous events. Firstly, we construct a quantitative VPC fate model with 1761 components by using HFPNe. Secondly, we employ two major biological fate determination rules-Rule I and Rule II-to VPC fate model. We then conduct 10,000 simulations for each of 48 sets of different genotypes, investigate variations of cell fate patterns under each genotype, and validate the two rules by comparing three simulation targets consisting of fate patterns obtained from *in silico* and in vivo experiments. In particular, an evaluation was successfully done by using our VPC fate model to investigate one target derived from biological experiments involving hybrid lineage observations. However, the understandings of hybrid lineages are hard to make on a discrete model because the hybrid lineage occurs when the system comes close to certain thresholds as discussed by Sternberg and Horvitz in 1986. Our simulation results suggest that: Rule I that cannot be applied with qualitative based model checking, is more reasonable than Rule II owing to the high coverage of predicted fate patterns (except for the genotype of lin-15 ko; lin-12ko double mutants). More insights are

also suggested. The quantitative simulation-based model checking approach is a useful means to provide us valuable biological insights and better understandings of biological systems and observation data that may be hard to capture with the qualitative one.

d. A novel meta-analysis approach of cancer transcriptomes reveals prevailing transcriptional networks in cancer cells

Atsushi Niida, Seiya Imoto, Masao Nagasaki, Rui Yamaguchi, Satoru Miyano

Although microarray technology has revealed transcriptomic diversities underlining various cancer phenotypes, transcriptional programs controlling them have not been well elucidated. To decode transcriptional programs governing cancer transcriptomes, we have recently developed a computational method termed EEM, which searches for expression modules from prescribed gene sets defined by prior biological knowledge like TF binding motifs. In this paper, we extend our EEM approach to predict cancer transcriptional networks. Starting from functional TF binding motifs and expression modules identified by EEM, we predict cancer transcriptional networks containing regulatory TFs, associated GO terms, and interactions between TF binding motifs. To systematically analyze transcriptional programs in broad types of cancer, we applied our EEM-based network prediction method to 122 microarray datasets collected from public databases. The data sets contain about 15000 experiments for tumor samples of various tissue origins including breast, colon, lung etc. This EEM based meta-analysis successfully revealed a prevailing cancer transcriptional network which functions in a large fraction of cancer transcriptomes; they include cell-cycle and immune related sub-networks. This study demonstrates broad applicability of EEM, and opens a way to comprehensive understanding of transcriptional networks in cancer cells.

e. Recursive regularization for inferring gene networks from time-course gene expression profiles

Teppei Shimamura, Seiya Imoto, Rui Yamaguchi, André Fujita, Masao Nagasaki, Satoru Miyano

Inferring gene networks from time-course microarray experiments with vector autoregressive (VAR) model is the process of identifying functional associations between genes through multivariate time series. This problem can be cast as a variable selection problem in Statistics. One of the promising methods for variable selection is the elastic net proposed by Zou and Hastie (2005). However, VAR modeling with the elastic net succeeds in increasing the number of true positives while it also results in increasing the number of false positives. By incorporating relative importance of the VAR coefficients into the elastic net, we propose a new class of regularization, called recursive elastic net, to increase the capability of the elastic net and estimate gene networks based on the VAR model. The recursive elastic net can reduce the number of false positives gradually by updating the importance. Numerical simulations and comparisons demonstrate that the proposed method succeeds in reducing the number of false positives drastically while keeping the high number of true positives in the network inference and achieves two or more times higher true discovery rate (the proportion of true positives among the selected edges) than the competing methods even when the number of time points is small. We also compared our method with various reverse-engineering algorithms on experimental data of MCF-7 breast cancer cells stimulated with two ErbB ligands, EGF and HRG. The recursive elastic net is a powerful tool for inferring gene networks from time-course gene expression profiles.

f. A state space representation of VAR models with sparse learning for dynamic gene networks

Kaname Kojima, Rui Yamaguchi, Seiya Imoto, Mai Yamauchi[5], Masao Nagasaki, Ryo Yoshida[6], Teppei Shimamura, Kazuko Ueno, Tomoyuki, Higuchi[6], Noriko Gotoh[5], Satoru Miyano: [5]Division of Systems Medical Technolog, Institute of Medical Science, University of Tokyo, [6]Institute of Statistical Mathematics

We propose a state space representation of vector autoregressive model and its sparse learning based on L1 regularization to achieve efficient estimation of dynamic gene networks based on time course microarray data. The proposed method can overcome drawbacks of the vector autoregressive model and state space model; the assumption of equal time interval and lack of separation ability of observation and systems noises in the former method and the assumption of modularity of network structure in the latter method. However, in a simple implementation the proposed model requires the calculation of large inverse matrices in a large number of times during parameter estimation process based on EM algorithm. This limits the

applicability of the proposed method to a relatively small gene set. We thus introduce a new calculation technique for EM algorithm that does not require the calculation of inverse matrices. The proposed method is applied to time course microarray data of lung cells treated by stimulating EGF receptors and dosing an anticancer drug, Gefitinib. By comparing the estimated network with the control network estimated using non-treated lung cells, perturbed genes by the anticancer drug could be found, whose up- and down-stream genes in the estimated networks may be related to side effects of the anticancer drug.

g. Comparing Pearson, Spearman and Hoeffding's D measure for gene expression association analysis

André Fujita, João Ricardo Sato[7], Marcos Angelo Almeida Demasi[7], Mari Cleide Sogayar[7], Carlow Eduardo Ferreira[7], Satoru Miyano: [7]University of São Paulo

DNA microarrays have become a powerful tool to describe gene expression profiles associated with different cellular states, various phenotypes and responses to drugs and other extra- or intra-cellular perturbations. In order to cluster co-expressed genes and/or to construct regulatory networks, definition of distance or similarity between measured gene expression data is usually required, the most common choices being Pearson's and Spearman's correlations. Here, we evaluate these two methods and also compare them with a third one, namely Hoeffding's D measure, which is used to infer nonlinear and non-monotonic associations, i.e. independence in a general sense. By comparing three different variable association approaches, namely Pearson's correlation, Spearman's correlation and Hoeffding's D measure, we aimed at assessing the most approppriate one for each purpose. Using simulations, we demonstrate that the Hoeffding's D measure outperforms Pearson's and Spearman's approaches in identifying nonlinear associations. Our results demonstrate that Hoeffding's D measure is less sensitive to outliers and is a more powerful tool to identify nonlinear and non-monotonic associations. We have also applied Hoeffding's D measure in order to identify new putative genes associated with tp53. Therefore, we propose the Hoeffding's D measure to identify nonlinear associations between gene expression profiles.

h. The impact of measurement error in the identification of regulatory networks

André Fujita, Alexandre G. Patriota[7], João Ricardo Sato[7], Satoru Miyano

There are several studies in the literature depicting measurement error in gene expression data and also, several others about regulatory network models. However, only a little fraction describes a combination of measurement error in mathematical regulatory networks and shows how to identify these networks under different rates of noise. This research investigates the effects of measurement error on the estimation of the parameters in regulatory networks. Simulation studies indicate that, in both time series (dependent) and non-time series (independent) data, the measurement error strongly affects the estimated parameters of the regulatory network models, biasing them as predicted by the theory. Moreover, when testing the parameters of the regulatory network models, p-values computed by ignoring the measurement error are not reliable, since the rate of false positives are not controlled under the null hypothesis. In order to overcome these problems, we present an improved version of the Ordinary Least Square estimator in independent (regression models) and dependent (autoregressive models) data when the variables are subject to noises. Moreover, measurement error estimation procedures for microarrays are also described. Simulation results also show that both corrected methods perform better than the standard ones (i.e., ignoring measurement error). The proposed methodologies are illustrated using microarray data from lung cancer patients and mouse liver time series data. Measurement error dangerously affects the identification of regulatory network models, thus, they must be reduced or taken into account in order to avoid erroneous conclusions. This could be one of the reasons for high biological false positive rates identified in actual regulatory network models.

i. Quality control and reproducibility in DNA microarray experiments

André Fujita, João Ricardo Sato[7], Fernando H..L. da Silva[7], Maria C. Galvao[8], Mari Cleide Sogayar[7], Satoru Miyano: [8]University Methodist of São Paulo

Biological experiments are usually set up in technical replicates (duplicates or triplicates) in order to ensure reproducibility and, to assess any significant error introduced during the experimental process. The first step in biological data analysis is to check the technical replicates and to confirm that the error of measure is small enough to be of no concern. However, lit-

tle attention has been paid to this part of analysis. Here, we propose a general process to estimate the error of measure and consequently, to provide an interpretable and objective way to ensure the technical replicates' quality. Particularly, we illustrate our application in a DNA microarray dataset set up in technical duplicates.

j. Exploring temporal transcription regulation structure of *Aspergillus fumigatus* in heat shock by state space model

Jin Hwan Do, Rui Yamaguchi, Satoru Miyano

The thermotolerance of *Aspergillus fumigatus* plays a critical role in mammalian and avian infections. Thus, the identification of its adaptation mechanism to higher temperature is very important for an efficient anti-fungal drug development as well as fundamental understanding of its pathogenesis. We explored the temporal transcription regulation structure of this pathogenic fungus under heat shock conditions using the time series microarray data reported by Nierman *et al*. (*Nature* 2005, 438:1151-1156). The estimated transcription regulation structure of *A. fumigatus* shows that the heat shock proteins are strongly negatively associated with central metabolic pathway genes such as the tricarboxylic acid cycle (TCA cycle) and carbohydrate metabolism. It was 60 min and 120 min, respectively, after the growth temperature changes from 30℃ (corresponding to environments of tropical soil) to 37℃ and 48℃ (corresponding to temperatures in the human body and compost, respectively) that some of genes in TCA cycle were started to be upregulated. In these points, most of heat shock proteins showed lowest expression level after heat shocks. Among the heat shock proteins, the HSP30 (AFU6G06470), a single integral plasma membrane heat shock protein, presented most active role in transcription regulation structure in both heat shock conditions of 37℃ and 48℃. The metabolic genes associated with multiple genes in the gene regulation network showed a tendency to have opposite expression patterns of heat shock proteins. The role of those metabolic genes was second regulator in the coherent feed-forward loop type of regulation structure having heat shock protein as its first regulator. This type of regulation structure might be very advantageous for the thermal adaptation of *A. fumigatus* under heat shock because a small amount of heat shock proteins can rapidly magnify their regulation effect on target genes. However, the coherent feed-forward loop type of regulation of heat shock proteins with metabolic genes became less frequent with increasing temperature. This might be the reason for dramatic increase in the expression of heat shock proteins and the number of heat shock response genes at heat shock of 48℃. We systemically analysed the thermal adaption mechanism of *A. fumigatus* by state space model with times series microarray data in terms of transcription regulation structure. We suggest for the first time that heat shock proteins might efficiently regulate metabolic genes using the coherent feed-forward loop type of regulation structure. This type of regulation structure would also be efficient for adjustment to the other stresses requiring rapid change of metabolic mode as well as thermal adaptation.

k. Parameter estimation of *in silico* biological pathways with particle filtering towards a petascale computing

Kazuyuki Nakamura[6], Ryo Yoshida[6], Masao Nagasaki, Satoru Miyano, Tomoyuki Higuchi[6]

The aim of this research is to demonstrate the potential power of large-scale particle filtering for the parameter estimations of *in silico* biological pathways where time course measurements of biochemical reactions are observable. The method of particle filtering has been a popular technique in the field of statistical science, which approximates posterior distributions of model parameters of dynamic system by using sequentially-generated Monte Carlo samples. In order to apply the particle filtering to system identifications of biological pathways, it is often needed to explore the posterior distributions which are defined over an exceedingly high-dimensional parameter space. It is then essential to use a fairly large amount of Monte Carlo samples to obtain an approximation with a high-degree of accuracy. In this paper, we address some implementation issues on large-scale particle filtering, and then, indicate the importance of large-scale computing for parameter learning of *in silico* biological pathways. We have tested the ability of the particle filtering with 10(8) Monte Carlo samples on the transcription circuit of circadian clock that contains 45 unknown kinetic parameters. The proposed approach could reveal clearly the shape of the posterior distributions over the 45 dimensional parameter space.

l. Hybrid Petri net based modeling for biological pathway simulation

Hiroshi Matsuno[9], Masao Nagasaki, Satoru Miyano.: [9]Yamaguchi University

Hybrid Petri net (HPN) is an extension of the Petri net formalism, which enables us to handle continuous information in addition to discrete information. Firstly, this paper demonstrates how biological pathways can be modeled by the integration of discrete and continuous elements, with an example of the λ phage genetic switch system including induction and retroregulation mechanisms. Although HPN allows intuitive modeling of biological pathways, some fundamental biological processes such as complex formation cannot be represented with HPN. Thus, this paper next provides the formal definition of hybrid functional Petri net with extension (HFPNe), which has high potential for modeling various kinds of biological processes. Cell Illustrator is a software tool developed on the basis of the definition of HFPNe. Hypothesis creation by Cell Illustrator is demonstrated with the example of the cyanobacterial circadian gene clock system. Finally, our ongoing tasks, which include the development of a computational platform for systems biology, are shown.

2. Bioinformatics for Kampo Medicine

a. Orengedokuto and berberine improve indomethacin-induced small intestinal injury via adenosine

Yoko Watanabe-Fukuda[10], Masahiro Yamamoto[10], Naoko Miura[10], Masato Fukutake[11], Atsushi Ishige[10], Rui Yamaguchi, Masao Nagasaki, Ayumu Saito, Seiya Imoto, Satoru Miyano, Junzo Takeda[10], Kenji Watanabe[10]: [10]Keio University, [11]Tsumura & Co.

Recent endoscopic technology has revealed that small intestinal injury is a serious threat to patients receiving nonsteroidal anti-inflammatory drugs (NSAIDs). We previously showed that Japanese herbal medicine, Orengedokuto (OGT; Huang-Lian-Jie-Du-Tang in Chinese), protects mice from lethal indomethacin (IND)-induced enteropathy. To elucidate the mechanism of the protective effect of OGT, we performed microarray analyses and high power statistical analyses of microarray data using new bioinformatics tools. Methods Female BALB/c mice were subcutaneously injected with IND (20 mg/kg) once a day for 2 days. OGT-treated mice received a diet containing OGT from the first IND injection until the end of the experiment. Gene expression signals of small intestine were obtained with GeneChip. Analyses for overrepresentation of Gene Ontology categories were conducted using MetaGene Profiler (MGP) and the changes were visualized by Cell Illustrator Online (CIO). Furthermore, active ingredients of OGT were investigated. MGP and CIO suggested a critical role for the adenosine system, especially adenosine deaminase (ADA), a key enzyme of adenosine catabolism. Quantitative real time RT-PCR and in situ hybridization showed that OGT decreased the expression of ADA, which possibly resulted in the elevation of the anti-inflammatory nucleoside adenosine. Blockade of the adenosine A2a receptor abrogated the protective effect of OGT. Berberine, a major ingredient of OGT, suppressed ADA expression and reduced the incidence of lethality. OGT may prevent IND-induced enteropathy by decreasing ADA which results in the elevation of adenosine. Modulation of the adenosine system may be an efficient therapeutic strategy for NSAID-induced enteropathy.

3. Computational Knowledge Discovery

a. Computational predictions for functional proteins working after cleaved in apoptotic pathway.

Chigusa Miyakawa[9], Manabu Sugii[9] Hiroshi Matsuno[9], Satoru Miyano

Protein cleavage by a caspase enzyme is observed in many pathways including apoptosis pathway, which induces protein modifications such as N-myristoylation. By exhaustive search of NCBI GenBank database based on the information in PeptideCutter website, we have identified a protein very likely to control apoptosis after N-myristoylation which is caused by the cleavage by caspase-8. To find more rules for sequences cleaved by caspase enzymes, we conducted computational experiments using the machine learning system BONSAI, and extracted some characteristic sequences involving serine-threonine kinase motif. This suggests the new possibility that the machine learning technique finds the significant sequence such as the recognition site of the enzyme.

4. Algorithms for Bioinformatics

a. BFL: a node and edge betweenness based fast layout algorithm for large scale networks

Tatsunori B. Hashimoto[12], Masao Nagasaki, Kaname Kojima, Satoru Miyano: [12]Harvard University

Network visualization would serve as a useful first step for analysis. However, current graph layout algorithms for biological pathways are insensitive to biologically important information,

e.g. subcellular localization, biological node and graph attributes, or/and not available for large scale networks, e.g. more than 10000 elements. To overcome these problems, we propose the use of a biologically important graph metric, betweenness, a measure of network flow. This metric is highly correlated with many biological phenomena such as lethality and clusters. We devise a new fast parallel algorithm calculating betweenness to minimize the preprocessing cost. Using this metric, we also invent a node and edge betweenness based fast layout algorithm (BFL). BFL places the high-betweenness nodes to optimal positions and allows the low-betweenness nodes to reach suboptimal positions. Furthermore, BFL reduces the runtime by combining a sequential insertion algorithm with betweenness. For a graph with n nodes, this approach reduces the expected runtime of the algorithm to $O(n^2)$ when considering edge crossings, and to $O(n \log n)$ when considering only density and edge lengths. Our BFL algorithm is compared against fast graph layout algorithms and approaches requiring intensive optimizations. For gene networks, we show that our algorithm is faster than all layout algorithms tested while providing readability on par with intensive optimization algorithms. We achieve a 1.4 second runtime for a graph with 4000 nodes and 12000 edges on a standard desktop computer.

b. Linear-time protein 3-D structure searching with insertions and deletions

Tetsuo Shibuya, Jesper Jansson[13], Kunihiko Sadakane[4]: [13]Ochanomizu University

It becomes more and more important to search for similar structures from molecular 3-D structure databases in the structural biology of the post genomic era. Two molecules are said to be similar if the RMSD (root mean square deviation) of the two molecules is less than or equal to some given constant bound. In this paper, we consider an important, fundamental problem of finding all the similar substructures from 3-D structure databases of chain molecules (such as proteins), with consideration of indels (*i.e.*, insertions and deletions). The problem has been believed to be very difficult, but its computational difficulty has not been well known. In this paper, we first show that the same problem in arbitrary dimension is NP-hard. Moreover, we also propose a new algorithm that dramatically improves the average-case time complexity for the problem, in case the number of indels k is bounded by some constant. Our algorithm solves the above problem in average $O(N)$ time, while the time complexity of the best known algorithm was $O(Nm^{k+1})$, for a query of size m and a database of size N.

c. Better decomposition heuristics for the maximum-weight connected graph problem using betweenness centrality

Takanori Yamamoto[4], Hideo Bannai[4], Masao Nagasaki, Satoru Miyano

We developed new decomposition heuristics for finding the optimal solution for the maximum-weight connected graph problem, which is known to be NP-hard. Previous optimal algorithms for solving the problem decompose the input graph into subgraphs using heuristics based on node degree. We propose new heuristics based on betweenness centrality measures, and show through computational experiments that our new heuristics tend to reduce the number of subgraphs in the decomposition, and therefore could lead to the reduction in computational time for finding the optimal solution. The method is further applied to analysis of biological pathway data.

Publications

1. Araki, H., Tamada, Y., Imoto, S., Dunmore, B., Sanders, D., Humphrey, S., Nagasaki, M., Doi, A., Nakanishi, Y., Yasuda, K., Tomiyasu, Y., Tashiro, K., Print, C., Charnock-Jones, D.S., Kuhara, S., Miyano, S. Analysis of PPARalpha-dependent and PPARalpha-independent transcript regulation following fenofibrate treatment of human endothelial cells. Angiogenesis. 12(3): 221-229, 2009.
2. DeLisi, C., Kanehisa, M., Miyano, S., Mohr, S., Wallach, I. (Eds.) Genome Informatics. 22. Imperial College Press, London, 2009.
3. Do, J.H., Miyano, S., Choi, D.K. Statistical approaches to genome-wide biological networks. Biochip J. 3(3): 190-202, 2009.
4. Do, J.H., Yamaguchi, R., Satoru Miyano, S. Exploring temporal transcription regulation structure of *Aspergillus fumigatus* in heat shock by state space model. BMC Genomics. 10:306, 2009.
5. Fujita, A., Patriota, A.G., Sato, J.R., Miyano, S. The impact of measurement error in the identification of regulatory networks. BMC Bioinformatics. 10:412, 2009.

6. Fujita, A., Sato, J.R., da Silva, F.H.L., Galvao, M.C., Sogayar, M.C., Miyano, S. Quality control and reproducibility in DNA microarray experiments. Genome Informatics. 23: 21-31, 2009.
7. Fujita, A., Sato, J.R., Demasi, M.A.A, Sogayar, M.C., Ferreira, C.E., Miyano, S. Comparing Pearson , Spearman and Hoeffding's D measure for gene expression association analysis. J. Bioinformatics and Computational Biology. 7(4): 663-684, 2009.
8. Hachimori, B., Shibuya, T. Lazy suffix array: the data structure for online construction and pattern searching, IEICE Transactions on Fundamentals. E92-A(8):1750-1756, 2009.
9. Hashimoto, T.B., Nagasaki, M., Kojima, K., Miyano, S. BFL: a node and edge betweenness based fast layout algorithm for large scale networks. BMC Bioinformatics. 10:19, 2009.
10. Kojima, K., Yamaguchi, R., Imoto, S., Yamauchi, M., Nagasaki, M., Yoshida, R., Shimamura, T., Ueno, K., Higuchi, T., Gotoh, N., Miyano, S. A state space representation of VAR models with sparse learning for dynamic gene networks. Genome Informatics. 22:56-68, 2009.
11. Li, C., Nagasaki, M., Ueno, K., Miyano, S. Simulation-based model checking approach to cell fate specification during Caenorhabditis elegans vulval development by hybrid functional Petri net with extension. BMC Systems Biology. 3:42, 2009.
12. Matsuno, H., Nagasaki, M., Miyano, S., Hybrid Petri net based modeling for biological pathway simulation. Nat. Comput. DOI 10.1007/s11047-009-9164-6.
13. Miyakawa, C., Sugii, M., Matsuno, H., Miyano, S. Computational predictions for functional proteins working after cleaved in apoptotic pathway. Proc. 2009 International Conference on Complex, Intelligent and Software Intensive Systems, IEEE Computer Society Press, 807-812, 2009.
14. Miyano, S., Yamaguchi, R., Tamada, Y., Nagasaki, M., Imoto, S. Gene networks viewed through two models. Lecture Notes in Computer Science. 5462: 54-66, 2009.
15. Nagasaki, M., Saito, A., Doi, A., Matsuno, H., Miyano, S. Foundations of Systems Biology-Using Cell Illustrator and Pathway Databases. Springer, 2009.
16. Nakamura, K., Yoshida, R., Nagasaki, M., Miyano, S., Higuchi, T. Parameter estimation of in silico biological pathways with particle filtering towards a petascale computing. Pacific Symposium on Biocomputing. 14. 4: 227-238, 2009.
17. Niida, A., Imoto, S., Nagasaki, M., Yamaguchi, R., Miyano, S. A novel meta-analysis approach of cancer transcriptomes reveals prevailing transcriptional networks in cancer cells. Genome Informatics. 22:121-131, 2009.
18. Niida, A., Smith, A.D., Imoto, S., Aburatani, H., Zhang, M.Q., Akiyama, T. Gene set-based module discovery in the breast cancer transcriptome. BMC Bioinformatics. 10:71, 2009.
19. Onuki, R., Shibuya, T., Kanehisa, M. New kernel methods for phenotype prediction from genotype data. Genome Informatics. 22:132-141, 2009.
20. Shibuya, T. Searching protein 3-D structures in linear time. Lecture Notes in Bioinformatics. 5541: 1-15, 2009.
21. Shibuya, T., Jansson, J., Sadakane, K.. Linear-time protein 3-D structure searching with insertions and deletions. Lecture Notes in Computer Science. 5724: 310-320, 2009.
22. Shimamura, T., Imoto, S., Yamaguchi, R., Fujita, A., Nagasaki, M., Miyano, S. Recursive regularization for inferring gene networks from time-course gene expression profiles. BMC Systems Biology. 3:41, 2009.
23. Tamada, Y., Araki, H., Imoto, S., Nagasaki, M., Doi, A., Nakanishi, Y., Tomiyasu, Y., Yasuda, K., Dunmore, B., Sanders, D., Humphries, S., Print, C., Charnock-Jones, D.S., Tashiro, K., Kuhara, S., Miyano, S. Unraveling dynamic activities of autocrine pathways that control drug-response transcriptome networks. Pacific Symposium on Biocomputing. 14: 251-263, 2009.
24. Watanabe-Fukuda, Y., Yamamoto, M., Miura, N., Fukutake, M., Ishige, A., Yamaguchi, R., Nagasaki, M., Saito, A., Imoto, S., Miyano, S., Takeda, J., Watanabe, K. Orengedokuto and berberine improve indomethacin-induced small intestinal injury via adenosine. J. Gastroenterology. 44(5):380-389, 2009.
25. Yamamoto, T., Bannai, H., Nagasaki, M., Miyano, S. Better decomposition heuristics for the maximum-weight connected graph problem using betweenness centrality. Lecture Notes in Artificial Intelligence. 5808:465-472, 2009.
26. Yashiro, Y., Bannai, H., Minowa, T., Yabiku, T., Miyano, S., Osawa, M., Iwama, A., Nakauchi, H. Transcriptional profiling of hematopoietic stem cells by high-throughput sequencing. I.J. Hematology. 89(1):24-33, 2009.
27. Yoshikawa, N., Nagasaki, M., Sano, M., Tokudome, S., Ueno, K., Shimizu, N., Imoto, S., Miyano, S., Suematsu, M., Fukuda, K., Morimoto, C., Tanaka, H. Ligand-based gene expression profiling reveals novel roles

of glucocorticoid receptor in cardiac metabolism. Am J Physiol Endocrinol Metab. 296(6): E1363-E1373, 2009.

Human Genome Center

Laboratory of Molecular Medicine
Laboratory of Genome Technology
ゲノムシークエンス解析分野
シークエンス技術開発分野

Professor	Yusuke Nakamura, M.D., Ph.D.	教　授	医学博士	中　村　祐　輔
Associate Professor	Yataro Daigo, M.D., Ph.D.	准教授	医学博士	醍　醐　弥太郎
Associate Professor	Koichi Matsuda, M.D., Ph.D.	准教授	医学博士	松　田　浩　一
Assistant Professor	Ryuji Hamamoto, Ph.D.	助　教	理学博士	浜　本　隆　二
Assistant Professor	Hitoshi Zembutsu, M.D., Ph.D.	助　教	医学博士	前　佛　　均

The major goal of our group is to identify genes of medical importance, and to develop new diagnostic and therapeutic tools. We have been attempting to isolate genes involving in carcinogenesis and also those causing or predisposing to various diseases as well as those related to drug efficacies and adverse reactions. By means of technologies developed through the genome project including a high-resolution SNP map, a large-scale DNA sequencing, and the cDNA microarray method, we have isolated a number of biologically and/or medically important genes, and are developing novel diagnostic and therapeutic tools.

1. Genes playing significant roles in human cancer

Yataro Daigo, Koichi Matsuda, Toyomasa Katagiri, Hidewaki Nakagawa, Ryuji Hamamoto, Hitoshi Zembutsu, Tomomi Ueki, Chikako Fukukawa, Meng-Lay Lin, Miki Akiyama, Jae-Hyun Park, Yosuke Harada, Satoshi Nagayama, Toshihiko Nishidate, Arata Shimo, Masahiko Ajiro, Jung-Won Kim, Koji Ueda, Atsushi Takano, Nobuhisa Ishikawa, Nagato Sato, Minh-Hue, Nguyen Ryohei Nishino, Junkichi Koinuma, Daiki Miki, Ken Masuda, Masato Aragaki, Takashi Fujitomo, Hideto Oshita, Satoko Uno, Yoichiro Kato, Kenji Tamura, Shingo Ashida, Su-Youn Chung, Motohide Uemura, Lianhua Piao, Kazuhito Morioka, Chizu Tanikawa, Motoko Unoki, Masanori Yoshimatsu, Shinya Hayami, Hyun-Soo Cho, Goji Toyokawa, Tadashi Takawa, and Yusuke Nakamura

(1) Lung cancer

Nectin-4

Gene expression profile analysis of lung cancers revealed the transactivation of an immunoglobulin-like molecule Nectin-4 in the majority of non-small cell lung cancers (NSCLC). Immunohistochemical staining of 422 NSCLCs showed that a high level of Nectin-4 expression was associated with poor prognosis for NSCLC patients ($P < 0.0001$), and multivariate analysis confirmed its independent prognostic value ($P < 0.0001$). We established an ELISA to measure serum Nectin-4 and found that serum Nectin-4 levels were significantly higher in NSCLC patients than in healthy volunteers. The proportion of the serum Nectin-4-positive cases

was 88 of 164 (53.7%) NSCLCs, whereas only 3 of 131 (2.3%) healthy volunteers were falsely diagnosed as positive, which was superior to carcinoembryonic antigen (CEA) and cytokeratin 19-fragment (CYFRA21-1) in sensitivity and specificity. A combined ELISA for both Nectin-4 and CEA increased sensitivity and classified 65.0% of lung adenocarcinomas as positive with false-positive rate of 4.6%. The use of both Nectin-4 and CYFRA21-1 classified 68.3% of lung squamous cell carcinomas as positive with false-positive rate of 6.1%. Treatment of lung cancer cells with small interfering RNAs against Nectin-4 suppressed its expression and cell growth. In addition, exogenous expression of Nectin-4 increased the lamellipodia formation and the invasive ability of mammalian cells through activation of small GTPase Rac1. Nectin-4 might play a significant role in lung carcinogenesis, and it should be a new candidate serum and tissue biomarker, as well as a therapeutic target.

WDHD1 (WD repeat and high-mobility group box DNA binding protein 1)

To identify novel biomarkers and therapeutic targets for lung and esophageal cancers, we screened for genes that were overexpressed in a large proportion of lung and esophageal carcinomas using a cDNA microarray representing 27,648 genes or expressed sequence tags. A gene encoding *WDHD1*, a WD repeat and high-mobility group box DNA binding protein 1, was selected as a candidate. Tumor tissue microarray analyses covering 267 archival non-small cell lung cancers and 283 esophageal squamous cell carcinomas (ESCC) revealed that positive WDHD1 immunostaining was associated with a poor prognosis for patients with non-small cell lung cancer ($P = 0.0403$) as well as ESCC ($P = 0.0426$). Multivariate analysis indicated it to be an independent prognostic factor for ESCC ($P = 0.0104$). Suppression of WDHD1 expression with small interfering RNAs effectively suppressed lung and esophageal cancer cell growth. In addition, induction of the exogenous expression of WDHD1 promoted the growth of mammalian cells. AKT1 kinase seemed to phosphorylate and stabilize the WDHD1 protein in cancer cells. WDHD1 expression is likely to play an important role in lung and esophageal carcinogenesis as a cell cycle regulator and a downstream molecule in the phosphoinositide 3-kinase/AKT pathway, and that WDHD1 is a candidate biomarker and a promising therapeutic target for cancer.

TBC1D7 (TBC1 domain family, member 7)

We screened molecules that were highly expressed in lung cancers by means of cDNA microarray analysis and found an elevated expression of TBC1 domain family, member 7 (*TBC1D7*) in the majority of lung cancers. Northern-blot analysis using mRNAs from 16 normal tissues detected its expression only in testis. Immunohistochemical staining using tumor tissue microarrays consisting of 261 archived non-small cell lung cancer (NSCLC) specimens suggested an association of TBC1D7 expression with poor prognosis for NSCLC patients ($P = 0.0063$). Treatment of lung cancer cells using siRNA against TBC1D7, suppressed its expression and resulted in inhibition of the cell growth. Furthermore, the induction of exogenous expression of TBC1D7 conferred growth-promoting activity at *in vitro* and *in vivo* conditions. We also identified TBC1D7 to interact with TSC1 protein in lung cancer cells. TSC1 introduction into cells increased the level of TBC1D7 protein, whereas knockdown of TSC1 expression decreased the level of TBC1D7 protein, suggesting that TBC1D7 is stabilized probably through interaction with TSC1. In addition, inhibition of the binding between TBC1D7 and TSC1 by a TBC1D7-derived 20-amino acid cell-permeable peptide (11R-TBC1D7(152-171)), which corresponded to the binding domain to TSC1, effectively suppressed growth of lung cancer cells. Selective suppression of TBC1D7 and/or inhibition of the TBC1D7-TSC1 complex formation could be promising therapeutic strategies for lung cancer therapy.

(2) Breast Cancer

GPATCH2 (G-patch domain containing 2)

Through analysis of the detailed genome-wide gene expression profiles of 81 breast tumors, we identified a novel gene, G-patch domain containing 2 (*GPATCH2*), that was overexpressed in the great majority of breast cancer cases. Treatment of breast cancer cells MCF-7 and T47D with siRNA against GPATCH2 effectively suppressed its expression, and resulted in the growth suppression of cancer cells, suggesting its essential role in breast cancer cell growth. We found an interaction of GPATCH2 protein with hPrp43, an RNA-dependent ATPase. Their interaction could significantly enhance the ATPase activity of hPrp43 and induce a growth-promoting effect on mammalian cells. Because northern blot analyses of normal human organs implied GPATCH2 to be a novel cancer/testis antigen, targeting GPATCH2 or inhibition of the

interaction between GPATCH2 and hPrp43 could be a promising novel therapeutic strategy of breast cancer.

BIG 3 (brefeldin A-inhibited guanine nucleotide-exchange protein 3)

Breast cancer is known to be a hormone-dependent disease, and estrogens through an interaction with estrogen receptor (ER) enhance the proliferative and metastatic activity of breast tumor cells. Here we show a critical role of transactivation of BIG3, brefeldin A-inhibited guanine nucleotide-exchange protein 3, in activation of the estrogen/ER signaling in breast cancer cells. Knocking-down of BIG3 expression with small-interfering RNA (siRNA) drastically suppressed the growth of breast cancer cells. Subsequent coimmunoprecipitation and immunoblotting assays revealed an interaction of BIG3 with prohibitin 2/repressor of estrogen receptor activity (PHB2/REA). When BIG3 was absent, stimulation of estradiol caused the translocation of PHB2/REA to the nucleus, enhanced the interaction of PHB2/REA and ERalpha, and resulted in suppression of the ERalpha transcriptional activity. On the other hand, when BIG3 was present, BIG3 trapped PHB2/REA in the cytoplasm and inhibited its nuclear translocation, and caused enhancement of ERalpha transcriptional activity. Our results imply that BIG3 overexpression is one of the important mechanisms causing the activation of the estrogen / ERalpha signaling pathway in the hormone-related growth of breast cancer cells.

RQCD1 (required for cell differentiation 1 homolog)

We identified required for cell differentiation 1 homolog (RQCD1) as a potential therapeutic target for breast cancer. Gene-expression profiling analysis of breast cancer cells, semi-quantitative RT-PCR, Northern blotting and Western blotting confirmed RQCD1 to be frequently up-regulated in breast cancer specimens and breast cancer cell lines. On the other hand, its expression was very weak or hardly detectable in normal human tissues except testis, indicating this molecule to be a novel cancer-testis antigen. Treatment of breast cancer cell lines with siRNA targeting RQCD1 drastically suppressed cell proliferation. Concordantly, introduction of exogenous RQCD1 into HEK293 cells significantly enhanced cell growth, implying RQCD1 to have an oncogenic activity. Co-immunoprecipitation experiments and immunocytochemical staining revealed an interaction of RQCD1 protein with Grb10 interacting GYF protein 1 (GIGYF1) and 2 (GIGYF2) proteins, involved in regulation of Akt activation, in breast cancer cells. Interestingly, knockdown of either of RQCD1, GIGYF1 or GIGYF2 resulted in significant reduction of the phosphorylation of Akt at Ser 473 in breast cancer cell lines. Our findings suggest that RQCD1 is a potential molecular target for treatment of breast cancer.

UBE2T (ubiquitin-conjugating enzyme E2T)

We found a critical role of ubiquitin-conjugating enzyme E2T (UBE2T), an E2 ubiquitin-conjugating enzyme, in mammary carcinogenesis. Immunocytochemical staining and in vitro binding assay revealed that UBE2T interacted and colocalized with the BRCA1/BRCA 1-associated RING domain protein (BARD1) complex. Knocking down of UBE2T expression with small interfering RNA drastically suppressed the growth of breast cancer cells. Interestingly, *in vivo* ubiquitination assay indicated BRCA1 to be polyubiquitinated by incubation with wild-type UBE2T protein, but not with C86A-UBE2T protein, an E2 activity-dead mutant, in which the 86th residue of cysteine was replaced with alanine. Furthermore, knocking down of UBE2T protein induced upregulation of BRCA1 protein in breast cancer cells, whereas its overexpression caused the decrease of the BRCA1 protein. Our data imply a critical role of UBE2T in development and/or progression of breast cancer through the interaction with and the regulation of the BRCA1/BARD1 complex.

TOPK (T-LAK cell-originated protein kinase)

We previously reported up-regulation of T-LAK cell-originated protein kinase (TOPK), a novel mitotic kinase, in the great majority of breast cancers. We found that protein phosphatase 1 alpha (PP1alpha) inactivation by cyclin-dependent kinase 1 (CDK1)/cyclin B1 caused enhancement of autophosphorylation of TOPK and resulted in its activation at an early stage of mitosis. Then TOPK interacted with and phosphorylated p97, a member of the AAA+ family of ATPase proteins, through an interaction with p47 protein as an adaptor protein. Interestingly, knockdown of TOPK or p97 in breast cancer cells caused the mitotic failures in the abscission process. This mitotic failure could be rescued by additional exogenous introduction of wild-type TOPK protein, but not by that of its kinase-dead form. Our findings suggest that TOPK is indispensable for cancer cell cytokinesis throughout phosphorylation on p97.

(3) Prostate cancer

PKIB (cAMP-dependent protein kinase inhibitor-beta)

Prostate cancer (PC) is the most common malignancy in males. Despite high response rates and clinical benefits, androgen-ablation therapy is ineffective for advanced or relapsed PC because of the emergence of aggressive castration-resistant prostate cancer (CRPC). Through our genome-wide gene expression analysis of PC cells purified from clinical CRPC tissues, we here identified a novel molecular target, *PKIB* (cAMP-dependent protein kinase inhibitor-beta), which was overexpressed specifically in CRPCs and aggressive PCs. Immunohistochemical analysis confirmed its overexpression in CRPCs and its strong correlation with high Gleason scores of PCs. Knockdown of PKIB by siRNA resulted in drastic growth suppression of PC cells, and, concordantly, exogenous introduction of PKIB into PC cells enhanced their growth and mobility. We found the direct interaction between PKIB and cAMP-dependent protein kinase A catalytic subunit (PKA-C), and showed that knockdown of PKIB in PC cells diminished the nuclear translocation of PKA-C. Knockdown of PKIB also decreased the phosphorylation level of Akt at Ser473 in PC cells, and exogenous PKIB introduction enhanced Akt phosphorylation in PC cells by incorporating with endogenous PKA-C kinase. *In vitro* kinase assay validated the recombinant PKIB enhanced phosphorylation of Akt at Ser473 by PKA-C kinase. These findings show that PKIB and PKA-C kinase can have critical functions of aggressive phenotype of PCs through Akt phosphorylation and that they should be a promising molecular target for PC treatment.

STYK1 (serine/threonine/tyrosine kinase 1)

Despite high response rates and clinical benefits, androgen ablation often fails to cure advanced or relapsed prostate cancer because castration-resistant prostate cancer (CRPC) cells inevitably emerge. CRPC cells not only grow under castration, but also behave more aggressively, indicating that a number of malignant signaling pathways are activated in CRPC cells as well as androgen receptor signaling. Based on information from the gene expression profiles of clinical CRPC cells, we here identified one overexpressed gene , serine / threonine / tyrosine kinase 1 (*STYK1*), encoding a potential kinase, as a molecular target for CRPC. RNA and immunohistochemical analyses validated the overexpression of STYK1 in prostate cancer cells, and its expression was distinct in CRPC cells. Knockdown of STYK1 by siRNA resulted in drastic suppression of prostate cancer cell growth and, concordantly, enforced expression of STYK1 promoted cell proliferation, whereas ectopic expression of a kinase-dead mutant STYK1 did not. An *in vitro* kinase assay using recombinant STYK1 demonstrated that STYK1 could have some potential as a kinase, although its specific substrates are unknown. These findings suggest that STYK1 could be a possible molecular target for CRPC, and small molecules specifically inhibiting STYK1 kinase could be a possible approach for the development of novel CRPC therapies.

ELOVL7 (long-chain fatty acid elongase 7)

A number of epidemiologic studies have indicated a strong association between dietary fat intake and prostate cancer development, suggesting that lipid metabolism plays some important roles in prostate carcinogenesis and its progression. In this study, through our genome-wide gene expression analysis of clinical prostate cancer cells, we identified a novel lipogenic gene, *ELOVL7*, coding a possible long-chain fatty acid elongase, as overexpressed in prostate cancer cells. ELOVL7 expression is regulated by the androgen pathway through SREBP1, as well as other lipogenic enzymes. Knockdown of ELOVL7 resulted in drastic attenuation of prostate cancer cell growth, and it is notable that high-fat diet promoted the growth of *in vivo* tumors of ELOVL7-expressed prostate cancer. *In vitro* fatty acid elongation assay and fatty acid composition analysis indicated that ELOVL7 was preferentially involved in fatty acid elongation of saturated very-long-chain fatty acids (SVLFA, C20:0 approximately). Lipid profiles showed that knockdown of ELOVL7 in prostate cancer cells affected SVLFAs in the phospholipids and the neutral lipids, such as cholesterol ester. Focusing on cholesterol ester as a source of de novo steroid synthesis, we show that ELOVL7 affected de novo androgen synthesis in prostate cancer cells. These findings suggest that EVOLV7 could be involved in prostate cancer growth and survival through the metabolism of SVLFAs and their derivatives, could be a key molecule to elucidate the association between fat dietary intake and prostate carcinogenesis, and could also be a promising molecular target for development of new therapeutic or preventive strategies for prostate cancers.

(4) Osteosarcoma

ROR2 (receptor tyrosine kinase-like orphan receptor 2)

Osteosarcoma is the most prevalent bone malignant tumor in children and adolescents, and displays heterogeneous histology and high propensity for distant metastasis. Although adjuvant chemotherapy remarkably improved treatment outcome over the past few decades, prognosis for osteosarcoma patients with pulmonary metastasis is still unsatisfactory. To identify novel therapeutic targets for osteosarcoma, we investigated the gene expression profile of osteosarcomas by cDNA microarray analysis and found transactivation of receptor tyrosine kinase-like orphan receptor 2 (ROR2) expression in the majority of osteosarcoma samples. Treatment of osteosarcoma cell lines with siRNA against ROR2 significantly inhibited cell proliferation and migration. We also identified wingless-type MMTV integration site family, member 5B (WNT5B) as a putative ROR2 ligand and that the physiological interaction of WNT5B and ROR2 could enhance cell migration, indicating the possible roles of ROR2 and WNT5B in the metastatic property of osteosarcoma cells. Taken together, our findings revealed that the WNT5B/ROR2 signaling pathway is a promising therapeutic target for osteosarcoma.

(5) Bladder cancer.

UHRF1 (ubiquitin-like with PHD and ring-finger domains 1)

Previous reports have shown that UHRF1 (ubiquitin-like with PHD and ring-finger domains 1) is essential for cellular proliferation. In this study, we examined whether *UHRF1* can be a novel molecular marker of bladder cancer. We performed real-time TaqMan quantitative reverse transcription-PCR and immunohistochemistry to examine expression levels of *UHRF1* in bladder and kidney cancers. Significant overexpression of UHRF1 was observed in bladder cancer. The overexpression was correlated with the stage and grade of the cancer. Although UHRF1 expression in muscle-invasive cancer was greater than in non-invasive (pTa) or superficially invasive (pT1) cancers, UHRF1 could still be detected by immunohistochemistry in these early-stage cancers. Overexpression of UHRF1 in bladder cancer was associated with increased risk of progression after transurethral resection. High expression of UHRF1 in kidney cancer was also observed. But the increased levels of UHRF1 in kidney cancer were less significant compared with those in bladder cancer. Our result indicates that an immunohistochemistry-based UHRF1 detection in urine sediment or surgical specimens can be a sensitive and cancer-specific diagnostic and/or prognosis method, and may greatly improve the current diagnosis based on cytology.

(6) p53 target genes

XEDAR (X-linked ectodermal dysplasia receptor)

Colorectal cancers with mutations in the *p53* gene have an invasive property, but its underlying mechanism is not fully understood. Through the screening of two data sets of the genome-wide expression profile, one for p53-introduced cells and the other for the numbers of cancer tissues, we report here X-linked ectodermal dysplasia receptor (XEDAR), a member of the TNFR superfamily, as a novel p53 target that has a crucial role in colorectal carcinogenesis. p53 upregulated XEDAR expression through two p53-binding sites within intron 1 of the XEDAR gene. We also found a significant correlation between decreased XEDAR expressions and *p53* gene mutations in breast and lung cancer cell lines ($P=0.0043$ and $P=0.0122$, respectively). Furthermore, promoter hypermethylation of the *XEDAR* gene was detected in 20 of 20 colorectal cancer cell lines (100%) and in 6 of 12 colorectal cancer tissues (50%), respectively. Thus, the XEDAR expression was suppressed to $<25\%$ of surrounding normal tissues in 12 of 18 colorectal cancer tissues (66.7%) due to either its epigenetic alterations and/or p53 mutations. We also found that XEDAR interacted with and subsequently caused the accumulation of FAS protein, another member of p53-inducible TNFR. Moreover, XEDAR negatively regulated FAK, a central component of focal adhesion. As a result, inactivation of XEDAR resulted in the enhancement of cell adhesion and spreading, as well as resistance to p53-induced apoptosis. Taken together, our findings showed that XEDAR is a putative tumor suppressor that could prevent malignant transformation and tumor progression by regulating apoptosis and anoikis.

PADI4 (peptidylarginine deiminase type 4)

Upon a wide range of cellular stresses, p53 is activated and inhibits malignant transformation through the transcriptional regulation of its target genes related to apoptosis, cell cycle arrest, and DNA repair. However, its involvement in posttranslational modifications of proteins has not yet been well characterized. Here, we report the novel role of p53 in the regulation of protein citrullination. p53 transactivated peptidy-

larginine deiminase type 4 (PADI4) through an intronic p53-binding site. The PADI4 gene encodes an enzyme catalyzing the citrullination of arginine residues in proteins, and ectopic expression of p53 or PADI4 induced protein citrullination. In addition, various proteins were citrullinated in response to DNA damage, but knockdown of PADI4 or p53 remarkably inhibited their citrullination, indicating the regulation of protein citrullination in a p53/PADI4-dependent manner. We found that PADI4 citrullinated the histone chaperone protein, nucleophosmin (NPM1), at the arginine 197 residue *in vivo* under physiologic conditions. Citrullination of NPM1 by PADI4 resulted in its translocation from the nucleoli to the nucleoplasm, whereas PADI4 did not alter the localization of mutant NPM1 (R197K). Furthermore, ectopic expression of PADI4 inhibited tumor cell growth, and concordantly, the knockdown of PADI4 attenuated p53-mediated growth-inhibitory activity, demonstrating the significance of PADI4-mediated protein citrullination in the p53 signaling pathway

2. Cancer vaccine

Phase I clinical trial using peptide vaccine for novel cancer-testis antigens (TTK, LY6K, IMP-3) for patients with advanced esophageal cancer

Koji Kono[1], Yoshiki Mizukami[1], Yataro Daigo[2], Atsushi Takano[2], Ken Masuda[2], Koji Yoshida[2], Takuya Tsunoda[2], Yoshihiko Kawaguchi[1], Yusuke Nakamura[2], and Hideki Fujii[1]: [1]First Department of Surgery, University of Yamanashi; [2]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo

We previously identified three novel HLA-A24-restricted epitope peptides, which were derived from three cancer-testis antigens, TTK protein kinase (TTK), lymphocyte antigen 6 complex locus K (LY6K), and insulin-like growth factor (IGF)-II mRNA binding protein 3 (IMP-3), as targets for cancer vaccination against esophageal squamous cell carcinoma (ESCC). To examine the safety, immunogenicity, and antitumor effect of vaccine treatment using a combination of these three peptides, 10 HLA-A2402-positive advanced ESCC patients who failed to standard therapy were enrolled in a phase I clinical trial. Each of the three peptides (1 mg each) was intradermally administered with 1 ml of incomplete Freund's adjuvant to the neck in three separate regions weekly for 5 weeks. The cancer vaccination therapy was well tolerated without any treatment-associated adverse events of grade 3 or 4. The TTK-, LY6K-, and/or IMP-3-specific T-cell immune responses were observed by enzyme-linked immunospot assay in peripheral blood lymphocytes obtained from nine of the 10 ESCC patients after their vaccination. The median survival time after the vaccination was 6.6 months. The vaccination could induce good clinical responses in 50% of the 10 patients. One patient experienced a complete response in hepatic metastasis lasting 7 months, one showed objective responses in all lung metastasis lesions, and three patients revealed a stable disease condition for at least 2.5 months. The cancer vaccine therapy using these three peptides demonstrated satisfactory safety and good immunogenicity as well as promising disease control rate, and therefore warrants further clinical studies.

Phase I clinical trial using peptide vaccine for human vascular endothelial growth factor receptor 2 (VEGFR2) in combination with gemcitabine for patients with advanced pancreatic cancer.

Motoki Miyazawa[1], Ryuji Ohsawa[1,2], Takuya Tsunoda[1,2,3], Seiko Hirono[1], Manabu Kawai[1], Masaji Tani[1], Yusuke Nakamura[3], and Hiroki Yamaue[1]: [1]Second Department of Surgery, Wakayama Medical University; [2]Research Department, OncoTherapy Science Inc.; [3]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo

Vascular endothelial growth factor receptor 2 (VEGFR2) is an essential factor in tumor angiogenesis and in the growth of pancreatic cancer. Immunotherapy using epitope peptide for VEGFR2 (VEGFR2-169) that we identified previously is expected to improve the clinical outcome. Therefore, a phase I clinical trial combining of VEGFR2-169 with gemcitabine was conducted for patients with advanced pancreatic cancer. Patients with metastatic and unresectable pancreatic cancer were eligible for the trial. Gemcitabine was administered at a dose of 1000 mg/m^2 on days 1, 8, and 15 in a 28-day cycle. The VEGFR2-169 peptide was subcutaneously injected weekly in a dose-escalation manner (doses of 0.5, 1, and 2 mg/body, six patients/one cohort). Safety and immunological parameters were assessed. No severe adverse effect of grade 4 or higher was observed. Of the 18 patients who completed at least one course of the treatment, 15 (83%) developed immunological reactions at the injection sites. Specific cytotoxic T lymphocytes (CTL) reacting to the VEGFR2-169 peptide were induced in 11 (61%)

of the 18 patients. The disease control rate was 67%, and the median overall survival time was 8.7 months. This combination therapy for pancreatic cancer patients was tolerable at all doses. Peptide-specific CTL could be induced by the VEGFR2-169 peptide vaccine at a high rate, even in combination with gemcitabine. From an immunological point of view, the optimal dose for further clinical trials might be 2 mg/body or higher.

3. Pharmacogenetics

Polymorphisms in *CYP2D6* and *ABCC2*, and clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients

Kazuma Kiyotani, Taisei Mushiroda, Chiyo K. Imamura, Naoya Hosono, Tatsuhiko Tsunoda, Michiaki Kubo, Yusuke Tanigawara, David A. Flockhart, Zeruesenay Desta, Todd C. Skaar, Fuminori Aki, Koichi Hirata, Yuichi Takatsuka, Minoru Okazaki, Shozo Ohsumi, Takashi Yamakawa, Mitsunori Sasa, Yusuke Nakamura, and Hitoshi Zembutsu

From the Laboratories for Pharmacogenetics, Genotyping Development, and Medical Informatics, RIKEN Center for Genomic Medicine; Department of Clinical Pharmacokinetics and Pharmacodynamics, School of Medicine, Keio University; Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo; Department of Surgery, Itoh Surgery and Breast Clinic; Department of Surgery, Yamakawa Breast Clinic; First Department of Surgery, Sapporo Medical University; Department of Surgery, Sapporo Breast Surgical Clinic; Department of Breast Surgery, Kansai Rosai Hospotal; Department of Breast Oncology, Shikoku Cancer Center; Department of Surgery, Tokushima Breast Care Clinic; Division of Clinical Pharmacology, Department of Medicine, Indiana University School of Medicine.

The clinical efficacy of tamoxifen is suspected to be influenced by the activity of drug-metabolizing enzymes and transporters involved in the formation, metabolism, and elimination of its active forms. We investigated relationships of polymorphisms in transporter genes and *CYP2D6* to clinical outcome of patients receiving tamoxifen. We studied 282 patients with hormone receptor-positive, invasive breast cancer receiving tamoxifen monotherapy, including 67 patients who have been previously reported. We investigated the effects of allelic variants of *CYP2D6* and haplotype-tagging single nucleotide polymorphisms (tag-SNPs) of *ABCB1*, *ABCC2*, and *ABCG2* on recurrence-free survival using the Kaplan-Meier method and Cox regression analysis. Plasma concentrations of tamoxifen metabolites were measured in 98 patients receiving tamoxifen 20 mg/d. *CYP2D6* variants were significantly associated with shorter recurrence-free survival ($P = 0.000036$; hazard ratio [HR] = 9.52; 95% CI, 2.79 to 32.45 in patients with two variant alleles v patients without variant alleles). Among 51 tag-SNPs in transporter genes, a significant association was found at rs3740065 in *ABCC2* ($P = 0.00017$; HR = 10.64; 95% CI, 1.44 to 78.88 in patients with AA *v* GG genotypes). The number of risk alleles of *CYP2D6* and *ABCC2* showed cumulative effects on recurrence-free survival ($P = 0.000000055$). Patients carrying four risk alleles had 45.25-fold higher risk compared with patients with one risk allele. *CYP2D6* variants were associated with lower plasma levels of endoxifen and 4-hydroxytamoxifen ($P = 0.0000043$ and 0.00052), whereas no significant difference was found among *ABCC2* genotype groups. Our results suggest that polymorphisms in *CYP2D6* and *ABCC2* are important predictors for the prognosis of patients with breast cancer treated with tamoxifen.

4. Genome-wide association study

Functional variants in *ADH1B* and *ALDH2* coupled with alcohol and smoking and esophageal cancer risk

Ri Cui[1], Yoichro Kamatani[1], Atsushi Takahashi[2], Masayuki Usami[1], Naoya Hosono[3], Takahisa Kawaguchi[4], Tatsuhiko Tsunoda[4], Naoyuki Kamatani[2], Michiaki Kubo[3], Yusuke Nakamura[1,5], and Koichi Matsuda[1]: [1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo; [2]Laboratory for Statistical Analysis, [3]Laboratory for Genotyping Development, [4]Laboratory for Medical Informatics, and [5]Laboratory for International Alliance, Center for Genomic Medicine, Institute of Physical and Chemical Research (RIKEN)

Esophageal squamous cell carcinoma (ESCC) is prevalent among Asian populations, with marked regional variations in incidence and mortality. Patients with ESCC have a very poor prognosis, but detection of ESCC at earlier stages could improve clinical outcome. Therefore, identification of epidemiologic factors that influence the development of ESCC would facilitate prevention and/or early detection of the disease. We performed a 2-step genome-wide

association study with subsequent replication analysis using a total of 1070 Japanese ESCC cases and 2836 controls. We also used logistic regression analysis to estimate the effect of gene-gene and gene-environmental interactions. We identified the significant associations of ESCC with 4q21-23 and 12q24 regions, which include nonsynonymous single nucleotide polymorphisms (SNP) in *ADH1B* (rs1229984, $P = 6.76 \times 10^{-35}$) and *ALDH2* (rs671, $P = 3.68 \times 10^{-68}$) that were previously shown to be associated with ESCC susceptibility. Multiple logistic regression analysis revealed SNP rs671, rs1229984, alcohol drinking, and smoking as the independent risk factors for ESCC (odds ratios of 1.66, 1.85, 1.92, and 1.79, respectively). Moreover, individuals who had both genetic and lifestyle-related risk factors had a nearly 190 times higher risk of ESCC than those who had neither of these. We found 2 known functional variants involved in the metabolism of alcohol and tobacco by-products as the most significant risk factors for the development of ESCC in a Japanese population. The individuals carrying both risk genotypes have a higher baseline risk of ESCC that is substantially increased by 2 lifestyle risk factors.

References

1. M. Harada, T. Hirota, A.I. Jodo, S. Doi, M. Kameda, K. Fujita, A. Miyatake, T. Enomoto, E. Noguchi, S. Yoshihara, M. Ebisawa, H. Saito, K. Matsumoto, Y. Nakamura, S.F. Ziegler, and M. Tamari: Functional analysis of the Thymic Stromal Lymphopoietin Variants in Human Bronchial Epithelial Cells. Am J Respir Cell Mol Biol, 40: 368-374, 2009
2. S. Chantarangsu, T. Mushiroda, S. Mahasirimongkol, S. Kiertiburanakul, S. Sungkanuparph, W. Manosuthi, W. Tantisiriwat, A. Charoenyingwattana, T. Sura, W. Chantratita, and Y. Nakamura: HLA-B*3505 allele is a strong predictor for nevirapine-induced skin adverse drug reactions in Thai HIV-infected patients. Pharmacogenetics and Genomics, 19: 139-146, 2009
3. D.N. Nikolova, N. Doganov, R. Dimitrov, K. Angelov, S.-K. Low, I. Dimova, D. Toncheva, Y. Nakamura, and H. Zembutsu: Genome-wide gene expression profiles of ovarian carcinoma: identification of molecular targets for the treatment of ovarian carcinoma. Molecular Medicine Reports, 2: 365-384, 2009
4. K. Yamazaki, A. Takahashi, M Takazoe, M. Kubo, Y. Onouchi, A. Fujino, N. Kamatani, Y. Nakamura, and A. Hata: Positive association of genetic variants in the upstream region of NXT2-3 with Crohn's disease in Japanese patients. Gut, 58: 228-232, 2009
5. D. Hirata, T. Yamabuki, D. Miki, T. Ito, E. Tsuchiya, M. Fujita, M. Hosokawa, K. Chayama, Y. Nakamura, and Y. Daigo: Involvement of epithelial cell transforming sequence-2 oncoantigen in lung and esophageal cancer progression. Clinical Cancer Research, 15: 256-266, 2009
6. S. Dobashi, T. Katagiri, E. Hirota, S. Ashida, Y. Daigo, T. Shuin, T. Fujioka, T. Miki, and Y. Nakamura: Involvement of TMEM22 overexpression in the growth of renal cell carcinoma cells. Oncology Reports, 21: 305-312, 2009
7. H. Zembutsu, Y. Suzuki, A. Sasaki, T. Tsunoda, M. Okazaki, M. Yoshimoto, T. Hasegawa, K. Hirata, and Y. Nakamura: Predicting response to Docetaxel neoadjuvant chemotherapy for advanced breast cancers through genome-wide gene expression profiling. International Journal of Oncology, 34: 361-370, 2009
8. Y. Nakamura:, DNA variations in human and medical genetics: 25 years of my experience. (review) Journal of Human Genetics, 54: 1-8, 2009
9. K. Ozaki, H. Sato, K. Inoue, T. Tsunoda, Y. Sakata, H. Mizuno, T.-H. Lin, Y. Miyamoto, A. Aoki, Y. Onouchi, S.-H. Sheu, S. Ikegawa, K. Odashiro, M. Nobuyoshi, S.-H. H. Juo, M. Hori, Y. Nakamura, and T. Tanaka: SNPs in BRAP associated with risk of myocardial infarction in Asian populations. Nature Genetics, 41: 329-333, 2009
10. K. Kashiwaya, M. Hosokawa, H. Eguchi, H. Ohigashi, O. Ishikawa, Y. Shinomura, Y. Nakamura, and H. Nakagawa: Identification of C2orf18, Termed ANT2BP (ANT2-binding protein), as one of key molecules involved in pancreatic carcinogenesis. Cancer Science, 100: 457-464, 2009
11. S. Nagayama, E. Yamada, Y. Kohno, T. Aoyama, C. Fukukawa, H. Kubo, G. Watanabe, T. Katagiri, Y. Nakamura, Y. Sakai, and J. Toguchida: Inverse correlation of the upregulation of FZD10 expression and the activation of β-catenin in synchronous colorectal tumors. Cancer Science, 100: 405-412, 2009
12. K. Ueda, Y. Fukase, T. Katagiri, N. Ishikawa, S. Irie, T. Sato, H. Ito, H. Nakayama, Y. Miyagi, E. Tsuchiya, N. Kohno, M. Shiwa, Y.

Nakamura, and Y. Daigo: Targeted serum glycoproteomics for the discovery of lung cancer-associated glycosylation disorders using lectin-coupled ProteinChip arrays. Proteomocs, 9: 2182-2192, 2009

13. The International Warfarin Pharmacogenetics Consortium:, Improved warfarin dosing with a global pharmacogenetic algorithm. New Eng. J. Med., in press, 2009
14. E.T. Betcheva, T. Mushiroda, A. Takahashi, M. Kubo, S.K. Karachanak, I.T. Zaharieva, R. V. Vazharova, I.I. Dimova, V.K. Milanova, T. Tolev, G. Kirov, M.J. Owen, M.C. O'Donovan, N. Kamatani, Y. Nakamura, and D.I. Toncheva: Case-control association study of 59 candidate genes reveals the DRD2 SNP rs 6277 (C957T) as the only susceptibility factor for schizophrenia in Bulgarian population. Journal of Human Genetics, in press, 2009
15. C. Fukukawa, S. Nagayama, T. Tsunoda, J. Toguchida, Y. Nakamura, and T. Katagiri: Activation of the non-canonical Dvl-Rac1-JNK pathway by Frizzled homologue 10 in human synovial sarcoma. Oncogene, 28: 1110-1120, 2009

    A. Yosifova, T. Mushiroda, D. Stoianov, R. Vazharova, I. Dimova, S. Karachanak, I. Zaharieva, V. Milanova, N. Madjirova, I. Gerdjikov, T. Tolev, S. Velkova, G. Kirov, M.J. Owen, M.C. O'Donovan, D. Toncheva, and Y. Nakamura: Case-control association study of 65 candidate genes revealed a possible association of a SNP of HTR5A to be a factor susceptible to bipolar disease in Bulgarian population. Journal of Affective Disorders, in press, 2009
16. Y. Kamatani, S. Wattanapokayakit, H. Ochi, T. Kawaguchi, A. Takahashi, N. Hosono, M. Kubo, T. Tsunoda, N. Kamatani, H. Kumada, A. Puseenam, T. Sura, Y. Daigo, K. Chayama, W. Chantratita, Y. Nakamura, and K. Matsuda: A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nature Genetics, 41: 591-595, 2009
17. K. Tamura, M. Furihata, S. Chung, M. Uemura, H. Yoshioka, T. Iiyama, S. Ashida, Y. Nasu, T. Fujioka, T. Shuin, Y. Nakamura, and H. Nakagawa: Stanniocalcin 2 (STC2) over-expression in castration-resistant prostate cancer and aggressive prostate cancer. Cancer Science, 100: 914-919, 2009
18. H. Tsukada, H. Ochi, T. Maekawa, H. Abe, Y. Fujimoto, M. Tsuge, H. Takahashi, H. Kumada, N. Kamatani, Y. Nakamura, and K. Chayama: A Polymorphism in MAPKAPK3 affects response to interferon therapy for chronic hepatitis C. Gastroenterology, in press, 2009
19. E.M. Dunleavy, D. Roche, H. Tagami, N. Lacoste, D. Ray-Gallet, Y. Nakamura, Y. Daigo, Y. Nakatani, and G. Almouzni-Pettinotti: HJURP, a key CENP-A-partner for maintenance and deposition of CENP-A at centromeres at late telophase/G1. Cell, 137: 485-497, 2009
20. M.-L. Lin, C. Fukukawa, J.-H. Park, K. Naito, K. Kijima, M. Ajiro, A. Shimo, T. Nishidate, Y. Nakamura, and T. Katagiri: Involvement of GPATCH2 in breast carcinogenesis though its interaction with DDX15/hPrp43, an RNA-dependent ATPase. Cancer Science, in press, 2009
21. K. Morioka, C. Tanikawa, K. Ochi, Y. Daigo, T. Katagiri, H. Kawano, H. Kawaguchi, A. Myoui, H. Yoshikawa, N. Naka, N. Araki, I. Kudawara, M. Ieguchi, K. Nakamura, Y. Nakamura, and K. Matsuda: The orphan receptor tyrosine kinase ROR2 as a potential therapeutic target for osteosarcoma. Cancer Science, in press, 2009
22. K. Nishimoto, Y. Kochi, K. Ikari, K. Yamamoto, A. Suzuki, K. Shimane, Y. Nakamura, K. Yano, N. Iikuni, S. Tsukahara, N. Kamatani, H. Okamoto, H. Kaneko, Y. Kawaguchi, M. Hara, Y. Toyama, T. Horiuchi, K. Tao, K. Yasumoto, D. Hamada, N. Yasui, H. Inoue, M. Itakura, H. Yamanaka, and S. Momohara:
    Association study of TRAF1-C5 polymorphisms with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in Japanese.
    Annals of the Rheumatic Diseases, in press, 2009
23. S. Chung, M. Furihata, K. Tamura, M. Uemura, Y. Daigo, Y. Nasu, T. Miki, T. Shuin, T. Fujioka, Y. Nakamura, and H. Nakagawa: Overexpressing PKIB in prostate cancer promotes its aggressiveness by linking between PKA and Akt pathways. Oncogene, 282849-2859, 2009
24. C.C. Ng, P.Y. Yew, S.M. Puah, G. Krishnan, L.F. Yap, S.H. Teo, P.V.H. Lim, S. Govindaraju, K. Ratnavelu, C.K. Sam, A. Takahashi, M. Kubo, N. Kamatani, Y. Nakamura, and T. Mushiroda: A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. Journal of Human Genetics, in press, 2009
25. J.-W. Kim, M. Akiyama, J.-H. Park, M.-L. Lin, A. Shimo, T. Ueki, T. Nishidate, Y. Nakamura, and T. Katagiri: Up-regulation of an estrogen/estrogen receptor (ER) signaling pathway by over-expression of brefeldin A-inhibited guanine nucleotide-exchange protein 3 (BIG3) in breast cancer. Cancer Science, in press, 2009

26. M. Unoki, J.D. Kelly, D.E. Neal, B.A.J. Ponder, Y. Nakamura, and R. Hamamoto: UHRF1 is a novel molecular marker for diagnosis and the prognosis of bladder cancer. British Journal of Cancer, in press, 2009
27. C. Tanikawa, Y. Furukawa, N. Yoshida, H. Arakawa, Y. Nakamura, and K. Matsuda: XEDAR as a putative colorectal tumor suppressor that mediates p53-regulated anoikis pathway. Oncogene, 28: 3081-3092, 2009
28. K. Hotta, T. Nakamura, J. Takasaki, H. Takahashi, A. Takahashi, Y. Nakata, S. Kamohara, K. Kotani, R. Komatsu, N. Itoh, I. Mineo, J. Wada, H. Masuzaki, M. Yoneda, A. Nakajima, T. Funahashi, S. Miyazaki, K. Tokunaga, K. Hamaguchi, K. Tanaka, K. Yamada, T. Hanafusa, S. Oikawa, H. Yoshimatsu, K. Nakao, T. Sakata, Y. Matsuzawa, N. Kamatani, and Y. Nakamura: Screening of 336 single-nucleotide polymorphisms in 85 obesity-related genes revealed McKusick-Kaufman syndrome gene variants are associated with metabolic syndrome. Journal of Human Genetics, 54: 230-235, 2009
29. K. Kono, Y. Mizukami, Y. Daigo, A. Takano, K. Masuda, K. Yoshida, T. Tsunoda, Y. Kawaguchi, Y. Nakamura, and H. Fujii: Vaccination with multiple peptides derived from novel cancer-testis antigens can induce specific T-cell responses and clinical responses in advanced esophageal cancer. Cancer Science, in press, 2009
30. N. Hosono, M. Kato, K. Kiyotani, T. Mushiroda, S. Takata, H. Sato, H. Amitan, Y. Tsuchiya, K. Yamazaki, T. Tsunoda, H. Zenbutsu, Y. Nakamura, and M. Kubo: New CYP2D6 genotyping method for functional gene dosage analysis by allele copy number detection. Clinical Chemistry, in press, 2009
A. Takano, N. Ishikawa, R. Nishino, K. Masuda, W. Yasui, K. Inai, H. Nishimura, H. Ito, H. Nakayama, Y. Miyagi, E. Tsuchiya, N. Kohno, Y. Nakamura, and Y. Daigo: Identification of Nectin-4 oncoprotein as a diagnostic and therapeutic target for lung cancer. Cancer Research, in press, 2009
31. H. Abe, H. Ochi, T. Maekawa, T. Hatakeyama, M. Tsuge, S. Kitamura, T. Kimura, D. Miki, F. Mitsui, N. Hiraga, M. Imamura, Y. Fujimoto, S. Takahashi, Y. Nakamura, H. Kumada, and K. Chayama: Effects of structural variations of APOBEC3 A and APOBEC3B genes in chronic hepatitis B virus infection. Hepatology Research, in press, 2009
32. K. Moriwaki, K. Noda, Y. Furukawa, K. Ohshima, A. Uchiyama, T. Nakagawa, N. Taniguchi, Y. Daigo, Y. Nakamura, N. Hayashi, and E. Miyoshi: Deficiency of GMDS leads to escape from NK cell-mediated tumor surveillance through modulation of TRAIL signaling. Gastroenterology, 137: 188-198, 2009
33. S. Chung, K. Tamura, M. Furihata, M. Uemura, Y. Daigo, Y. Nasu, T. Miki, T. Shuin, T. Fujioka, Y. Nakamura, and H. Nakagawa: Potential kinase, STYK1 (Serine/Threonine/Tyrosine Kinase 1), over-expressing in castration-resistant prostate cancer. Cancer Science, in press, 2009
34. S.-K. Low, K. Kiyotani, T. Mushiroda, Y. Daigo, Y. Nakamura, and H. Zembutsu: Association study of genetic polymorphism in ABCC4 with cyclophosphamide-induced adverse drug reactions in breast cancer patients. Journal of Human Genetics, 54: 564-571, 2009
35. P.-C. Cha, T. Mushiroda, H. Zembutsu, H. Harada, N. Shinoda, S. Kawamoto, R. Shimoyama, T Nishidate, T. Furuhata, K. Sasaki, K. Hirata, and Y. Nakamura: Single nucleotide polymorphism in ABCG2 is associated with irinotecan-induced severe myelosuppression. Journal of Human Genetics, 54: 572-580, 2009
36. M. Ajiro, T. Katagiri, K. Ueda, H. Nakagawa, C. Fukukawa, M.-L. Lin, J.-H. Park, T. Nishidate, Y. Daigo, and Y. Nakamura: Involvement of RQCD1 overexpression, a novel cancer-testis antigen, in the Akt pathway in breast cancer cells. International Journal of Oncology, 35: 673-681, 2009
37. Y. Okada, R. Yamada, A. Suzuki, Y. Kochi, K. Shimane, K. Myouzen, M. Kubo, Y. Nakamura, and K. Yamamoto: Contribution of a Haplotype in the HLA region to anticyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA-DRB1. Arthritis & Rheumatism, 603582-3590, 2009
38. R. Cui, Y. Kamatani, A. Takahashi, M. Usami, N. Hosono, T. Kawaguchi, T. Tsunoda, N. Kamatani, M. Kubo, Y. Nakamura, and K. Matsuda: Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. Gastroenterology, in press, 2009
39. K. Tamura, A. Makino, F. Hullin-Masuda, T. Kobayashi, M. Furihata, S. Chung, S. Ashida, T. Miki, T. Fujioka, T. Shuin, Y. Nakamura, and H. Nakagawa: Novel lipogenic enzyme ELOVL7 involves in prostate cancer growth through saturated long-chain fatty acid metabolism. Cancer Research, 698133-8140, 2009
40. T. Ueki, J.-H. Park, T. Nishidate, K. Kijima, K. Hirata, Y. Nakamura, and T. Katagiri:

Ubiquitination and downregulation of BRCA1 by ubiquitin-conjugating enzyme E2T overexpression in human breast cancer cells. Cancer Research, 698752-8760, 2009
41. K. Yokomine, S. Senju, T. Nakatsura, A. Irie, Y. Hayashida, Y. Ikuta, M. Harao, K. Imai, H. Baba, H. Iwase, H. Nomori, K. Takahashi, Y. Daigo, T. Tsunoda, Y. Nakamura, Y. Sasaki, and Y. Nishimura: The forkhead box M1 transcription factor as a candidate of target for anti-cancer immunotherapy. International Journal of Cancer, in press, 2009
42. K. Asano, T. Matsushita, J. Umeno, N. Hosono, A. Takahashi, T. Kawaguchi, T. Matsumoto, T. Matsui, Y. Kakuta, Y. Kinouchi, T. Shimosegawa, M. Hosokawa, Y. Arimura, Y. Shinomura, Y. Kiyohara, T. Tsunoda, N. Kamatani, M. Iida, Y. Nakamura, and M. Kubo: A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. Nature Genetics, 411325-1331, 2009
43. C. Tanikawa, K. Ueda, H. Nakagawa, N. Yoshida, Y. Nakamura, and K. Matsuda: Regulation of protein citrullination through p53/PADI4 network in DNA damage response. Cancer Research, 698761-8769, 2009
44. N. Sato, J. Koinuma, M. Fujita, M. Hosokawa, T. Ito, E. Tsuchiya, S. Kondo, Y. Nakamura, and Y. Daigo: Activation of WD repeat and HMG-box DNA binding protein 1 in pulmonary and esophageal carcinogenesis. Clinical Cancer Research, in press, 2009
45. J.-H. Park, T. Nishidate, Y. Nakamura, and T. Katagiri: Critical roles of T-LAK cell-originated protein kinase in cytokinesis. Cancer Science, in press, 2009
46. K. Kiyotani, T. Mushiroda, C.K. Imamura, N. Hosono, T. Tsunoda, M. Kubo, Y. Tanigawara, D.A. Flockhart, Z. Desta, T.C. Skaar, F. Aki, K. Hirata, Y. Takatsuka, M. Okazaki, S. Ohsumi, T. Yamakawa, M. Sasa, Y. Nakamura, and H. Zembutsu: Significant effect of polymorphisms in CYP2D6 and ABCC2 on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. J. Clin. Oncology, in press, 2009
47. S. Omori, Y. Tanaka, M. Horikoshi, A. Takahashi, K. Hara, H. Hirose, A. Kashiwagi, K. Kaku, R. Kawamori, T. Kadowaki, Y. Nakamura, and S. Maeda: Replication study for the association of new meta-analysis-derived risk loci with susceptibility to type 2 diabetes in 6,244 Japanese individuals. Diabetologia, 521554-1560, 2009
48. T. Matsushita, M. Kubo, K. Yonemoto, T. Ninomiya, K. Ashikawa, B. Liang, J. Hata, Y. Doi, T. Kitazono, S. Ibayashi, M. Iida, Y. Kiyoharam and Y. Nakamura: Lack of association between variations of PDE4D and ischemic stroke in the Japanese population. Stroke, 40: 1245-1251, 2009
49. K. Shimane, Y. Kochi, R. Yamada, Y. Okada, A. Suzuki, A. Miyake, M. Kubo, Y. Nakamura, and K. Yamamoto: A single nucleotide polymorphism in the IRF5 promoter region is associated with susceptibility to rheumatoid arthritis in the Japanese patients. Annals of the Rheumatic Diseases, 68: 377-383, 2009
50. M. Tamari, M. Harada, T. Hirota, and Y. Nakamura: Host molecular defence mechanisms against Chlamydophilia pneumoniae and genetic studies of immune-response-related genes in asthma. Recent Pat Inflamm Allergy Drug Discovory, 3: 17-25, 2009
51. M. Harada, K. Obara, T. Hirota, T. Yoshimoto, Y. Hitomi, M. Sakashita, S. Doi, A. Miyatake, K. Fujita, T. Enomoto, M. Taniguchi, N. Higashi, Y. Fukutomi, K. Nakanishi, Y. Nakamura, and M. Tamari: A functional polymorphism in IL-18 is associated with severity of bronchimal asthma. Am J Respir Cell Mol Biol, in press, 2009
52. Y. Hitomi, M. Ebisawa, M. Tomikawa, T. Imai, T. Kamata, T. Hirota, M. Harada, M. Sakashita, Y. Suzuki, N. Shomojo, Y. Kohno, K. Fujita, A. Miyatake, S. Doi, T. Enomoto, M. Taniguchi, N. Higashi, Y. Nakamura, and M. Tamari: Association of functional NLRP3 polymorphisms with susceptibility to food-induced anaphylaxis and aspirin-induced asthma. Journal of Allergy and Clinical Immunology, 124779-785, 2009
53. H. Naruse, N. Ikawa, K. Yamaguchi, Y Nakamura, M. Arai, C. Ishioka, K. Sugano, K. Tamura, N. Tomita, N. Matsubara, T. Yoshida, Y. Moriya, and Y. Furukawa: Determination of splice-site mutations in Lynch syndrome (hereditary non-polyposis colorectal cancer) patients using functional splicing assay. Familial Cancer, 8: 509-517, 2009
54. H. Eguchi, O. Ishikawa, H. Ohigashi, H. Takahashi, M. Yano, K. Nishiyama, Y. Tomita, R. Uehara, A. Takehara, Y. Nakamura, and H. Nakagawa: Serum REG4 level is a predictive biomarker for the response to preoperative chemoradiotherapy in patients with pancreatic cancer. Pancreas, 38: 791-798, 2009
55. K. Hotta, M. Nakamura, T. Nakamura, T. Matuso, Y. Nakata, S. Kamohara, N. Miyatake, K. Kotani, R. Komatsu, N. Itoh, I. Mineo, J. Wada, H. Masuzaki, M. Yoneda, A. Nakajima, T. Funahashi, S. Miyazaki, K. Tokunaga, M. Kawamoto, T. Ueno, K. Hamaguchi, K. Tanaka, K. Yamada, T. Ha-

nafusa, S. Oikawa, H. Yoshimatsu, K. Nakao, T. Sakata, Y. Matsuzawa, N. Kamatani, and Y. Nakamura: Association between obesity and polymorphisms in SEC16 B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population. Journal of Human Genetics, 54: 727-731, 2009
56. T. Nagao, M. Kira, M. Takahashi, J. Honda, T. Hirose, A. Tangoku, H. Zembutsu, Y. Nakamura, and M. Sasa: Serum estradiol should be monitored not only during the peri-menopausal period but also the post-menopausal period at the time of aromatase inhibitor administration. World Journal of Surgical Oncology, 7: 88-92, 2009
57. W. Satake, Y. Nakabayashi, I. Mizuta, Y. Hirota, C. Ito, M. Kubo, T. Kawaguchi, T. Tsunoda, M. Watanabe, A. Takeda, H. Tomiyama, K. Nakashima, K. Hasegawa, F. Obata, T. Yoshikawa, H. Kawakami, S. Sakoda, M. Yamamoto, N. Hattori, M. Murata, Y. Nakamura, and T. Toda: Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nature Genetics, 41: 1303-1307, 2009
58. W. Wangsomboonsiri, S. Mahasirimongkol, S. Chantarangsu, S. Kiertiburanakul, A. Charoenyingwattana, S. Komindr, C. Thongnak, T. Mushiroda, Y. Nakamura, W. Chantratita, and S. Sungkanuparph: Association between HLA-B*4001 and lipodystrophy among HIV-infected patients from Thailand who received a stavudine-containing antiretroviral regimen. Clinical Infectious Diseases, in press:, 2010
59. M. Miyazawa, R. Ohsawa, T. Tsunoda, S. Hirono, M. Kawai, M. Tani, Y. Nakamura, and H. Yamaue: Phase I clinical trial using peptide vaccine (VEGFR2-169) for human vascular endothelial growth factor receptor 2 in combination with gemcitabine for patients with advanced pancreatic cancer. Cancer Science, in press, 2010
60. M. Nuinoon, W. Makarasara, T. Mushiroda, I. Setianingsih, P.A. Wahidiyat, O. Sripichai, N. Kumasaka, A. Takahashi, S. Svasti, T. Munkongdee, S. Mahasirimongkol, C. Peerapittayamongkol, V. Viprakasit, N. Kamatani, P. Winichagoon, M. Kubo, Y. Nakamura, and S. Fucharoen: A genome-wide association identified the common genetic variants influence disease severity in β0-thalassemia/hemoglobin E. Human Genetics, in press, 2010
61. H. Nakahara, K. Sekiguchi, N. Hosono, M. Kubo, A. Takahashi, Y. Nakamura, and K. Kasai: Criterion values for multiplex SNP genotyping by the Invader assay. Forensic Science International: Genetics, in press, 2010
62. H. Nakahara, N. Hosono, T. Kitayama, K. Sekiguchi, M. Kubo, A. Takahashi, Y. Nakamura, Y. Yamano, and K. Kai: Automated SNPs typing system based on the Invader assay. Legal Medicine, in press, 2010
63. N. Sato, J. Koinuma, T. Ito, E. Tsuchiya, S. Kondo, Y. Nakamura, and Y. Daogo: Activation of an oncogenic TBC1D7 (TBC1 Domain Family, Member 7) protein in pulmonary carcinogenesis. Genes Chromosomes and Cancer, in press, 2010
64. M. Kato, T. Kawaguchi, S. Ishikawa, T. Umeda, R. Nakamichi, M.H. Shapero, K.W. Jones, Y. Nakamura, H. Aburatani, and T. Tsunoda: Population-genetic nature of copy number variations in the human genome. Human Molecular Genetics, in press, 2010
65. Y. Kamatani, K. Matsuda, M. Kubo, N. Hosono, Y. Daigo, Y. Nakamura, and N. Kamatani: A Japanese genome-wide association study of multiple hematological and biochemical traits reveals identification of 46 novel loci and replication of the majority of reported European associations. Nature Genetics, in press, 2010
66. M. Inoue, S. Senju, S. Hirata, Y. Ikuta, Y. Hayashida, A. Irie, M. Harao, K. Imai, Y. Tomita, T. Tsunoda, Y. Furukawa, T. Ito, Y. Nakamura, H. Baba, and Y. Nishimura: Identification of SPARC as a candidate target antigen for immunotherapy of various cancers. International Journal of Cancer, in press, 2010
67. S. Maeda, M. Kobayashi, S. Araki, T. Babazono, B.I. Freedman, M.A. Bostrom, J.N. Cooke, M. Toyoda, T. Umezono, L. Tarnow, T. Hansen, P. Gaede, A. Jorsal, D.P K Ng, M. Ikeda, T. Yanagimoto, T. Tsunoda, H. Unoki, K. Kawai, M. Imanishi, D. Suzuki, H. D. Shin, K.S. Park, A. Kashiwagi, Y. Iwamoto, K. Kaku, R. Kawamori, H.-H. Parving, D.W. Bowden, O. Pedersen, and Y. Nakamura: A single nucleotide polymorphism within the acetyl-coenzyme A carboxylase beta gene is associated with proteinuria in patients with type 2 Diabetes. PLoS Genetics, in press, 2010
68. J.-H. Park, T. Nishidate, K. Kijima, T. Ohashi, K. Takegawa, T. Fujikane, K. Hirata, Y. Nakamura, and T. Katagiri: Critical roles of MUC1 glycosylation by transactivated polypeptide N-acetylgalactosaminyltransferase 6 (GALNT6) in mammary carcinogenesis. Cancer Research, in press, 2010
69. H. Mototani, A. Iida, Y. Nakamura, and S. Ikegawa: Identification of sequence polymorphisms in CALM2 and association

analyses with hip osteoarthritis in a Japanese population. Journal of Bone and Mineral Metabolism, in press, 2010
70. N.J. Prescott, K.M. Dominy, M. Kubo, C.M. Lewis, S.A. Fisher, R. Redon, N. Huang, B.E. Stranger, K. Blaszczyk, B. Hudspith, G. Parkes, N. Hosono, K. Yamazaki, C.M. Onnie, A. Forbes, E.T. Dermitzakis, Y. Nakamura, J.C. Mansfield, J. Sanderson, M.E. Hurles, R.G. Roberts, and C.G. Mathew: Independent and population-specific association of risk variants at the IRGM locus with Crohn's disease. Human Molecular Genetics, in press, 2010
71. S. Hayami, M. Yoshimatsu, A. Veerakumarasivam, M. Unoki, Y. Iwai, T. Tsunoda, H.I. Field, J.D. Kelly, D.E. Neal, H. Yamaue, B.A. J. Ponder, Y. Nakamura, and R. Hamamoto: Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/RB pathway. Molecular Cancer, in press, 2010
72. R. Takayama, H. Nakagawa, A. Sawaki, N. Mizuno, H. Kawai, M. Tajika, Y. Yatabe, K. Matsuo, R. Uehara, K. Ono, Y. Nakamura, and K. Yamao: Serum tumor antigen REG4 a diagnostic biomarker in pancreatic ductal adenocarcinoma. Journal of Gastroenterol, 45: 52-59, 2010

Human Genome Center

Laboratory of Functional Analysis *In Silico*
機能解析イン・シリコ分野

Professor	Kenta Nakai, Ph.D.	教　授	理学博士　中　井　謙　太
Associate Professor	Kengo Kinoshita, Ph.D.	准教授	理学博士　木　下　賢　吾(10月まで)
Assistant Professor	Ashwini Patil, Ph.D.	助　教	理学博士　パティル、アシュウィニ（2010年１月から）

The mission of our laboratory is to conduct computational ("in silico") studies on the functional aspects of genome information. Roughly speaking, genome information represents what kind of proteins/RNAs are synthesized on what conditions. Thus, our study includes the structural analysis of molecular function of each gene product as well as the analysis of its regulatory information, which will lead us to the understanding of its cellular role represented by the networks of inter-gene interaction.

1. Modeling tissue-specific structural patterns in human and mouse promoters

Alexis Vandenbon and Kenta Nakai

Sets of genes expressed in the same tissue are believed to be under the regulation of a similar set of transcription factors, and can thus be assumed to contain similar structural patterns in their regulatory regions. Here we present a study of the structural patterns in promoters of genes expressed specifically in 26 human and 34 mouse tissues. For each tissue we constructed promoter structure models, taking into account the presence of motifs, their positioning to the transcription start site, and pairwise positioning of motifs. We found that 35 out of 60 models (58%) were able to distinguish positive test promoter sequences from control promoter sequences with statistical significance. Models with high performance include those for liver, skeletal muscle, kidney and tongue. Many of the important structural patterns in these models involve transcription factors of known importance in the tissues in question and structural patterns tend to be conserved between human and mouse. In addition to that, promoter models for related tissues tend to have high inter-tissue performance, indicating that their promoters share common structural patterns. Together, these results illustrate the validity of our models, but also indicate that the promoter structures for some tissues are easier to model than those of others.

2. Gene regulatory networks in the early development of *Ciona intestinalis*

Xuyang Yuan, Atsushi Kubo[1], Yutaka Satou[1] and Kenta Nakai: [1]Kyoto University

The ascidian *Ciona intestinalis* shares basic gene repertories and many other physical and developmental characteristics with vertebrates making it an idea model system for elucidating the functional interaction among genes responsible for the early body plan of chordates. ChIP-chip experiment is an innovative technique for genome-wide identification of transcription factor binding sites. By doing ChIP-chip experiments for 11 transcription factors essential for *Ciona intestinalis* early development, we deter-

mined the structure of core networks at the *cis*-regulatory level in the early embryo of *Ciona intestinalis.* In order to prove the quality of our ChIP-chip data and thus the consequent analysis, we carefully checked the enrichment of motifs in the bound regions detected through ChIP-chip experiments. For each bound region of a transcription factor, a corresponding background region with the same length as the bound region is randomly selected from the whole genome. The sampled region was adopted only if it satisfied two conditions. First, the 10 kb region around its centre position does not overlap with any bound region of the TF in question. Second, its GC content was above a given threshold which was initially set to a value slightly lower than the average GC content of all bound regions. The threshold was then properly adjusted and the sampling was redone until the average GC content of all sampled background regions was almost the same as the average GC content of all bound regions. The 10 kb region around the peak position of each bound region and the 10 kb region around the centre position of each sampled background ground region were used for GC content comparison and motif searching. By comparing the results of ChIP-chip experiment to the previous gene knockdown experiment results, we also found that many of the previous detected regulations are direct.

3. Gradual transition from mosaic to global DNA methylation patterns during deuterostome evolution

Kazuaki A. Matsumoto[2], Kohji Okamura, and Kenta Nakai: [2]Waseda University

DNA methylation by the Dnmt family occurs in vertebrates and invertebrates, including ascidians, and is thought to play important roles in gene regulation and genome stability, especially in vertebrates. However, the global methylation patterns of vertebrates and invertebrates are distinctive. Whereas almost all CpG sites are methylated in vertebrates, with the exception of those in CpG islands, the ascidian genome contains approximately equal amounts of methylated and unmethylated regions. Curiously, methylation status can be reliably estimated from the local frequency of CpG dinucleotides in the ascidian genome. Methylated and unmethylated regions tend to have few and many CpG sites, respectively, consistent with our knowledge of the methylation status of CpG islands and other regions in mammals. However, DNA methylation patterns and levels in vertebrates and invertebrates have not been analyzed in the same way. Using a new computational methodology based on the decomposition of the bimodal distributions of methylated and unmethylated regions, we estimated the extent of the global methylation patterns in a wide range of animals. We then examined the epigenetic changes *in silico* along the phylogenetic tree. We observed a gradual transition from fractional to global patterns of methylation in deuterostomes, rather than a clear demarcation between vertebrates and invertebrates. When we applied this methodology to six piscine genomes, some of them showed features similar to those of invertebrates. The mammalian global DNA methylation pattern was probably not acquired at an early stage of vertebrate evolution, but gradually expanded from that of a more ancient organism.

4. An Assessment of Prediction Algorithms for Nucleosome Positioning

Yoshiaki Tanaka and Kenta Nakai

Nucleosome configuration in eukaryotic genomes is an important clue to clarify the mechanisms of regulation for various nuclear events. In the past few years, numerous computational tools have been developed for the prediction of nucleosome positioning, but there is no third-party benchmark about their performance. Here we present a performance evaluation using genome-scale in vivo nucleosome maps of two vertebrates and three invertebrates. In our measurement, two recently updated versions of Segal's model and Gupta's SVM with the RBF kernel, which was not implemented originally, showed higher prediction accuracy although their performances differ significantly in the prediction of medaka fish and candida yeast. The cross-species prediction results using Gupta's SVM also suggested rather specific characters of nucleosomal DNAs in medaka and budding yeast. With the analyses for over- and under-representation of DNA oligomers, we found both general and species-specific motifs in nucleosomal and linker DNAs. The oligomers commonly enriched in all five eukaryotes were only CA/TG and AC/GT. Thus, to achieve relatively high performance for a species, it is desirable to prepare the training data from the same species.

5. Positional variations among heterogeneous nucleosome maps give dynamical information on chromatin

Yoshiaki Tanaka, Itsuki Yoshimura[3], and Kenta Nakai: [3]Faculty of Medicine

Although nucleosome remodeling is essential to transcriptional regulation in eukaryotes, little is known about its genome-wide behavior. Since a number of nucleosome positioning maps in vivo have been recently determined, we examined if their comparisons might be used for obtaining a genome-wide profile of nucleosome remodeling. Using seven yeast maps, the local variability of nucleosomes, measured by the entropy, was significantly higher in a set of reported unstable nucleosomes. The binding sites of four transcription factors, known as the remodeling factors, were distinctively high both in entropy and linker ratio, whereas those of Yhp1, their potential inhibitor, showed the lowest values in both of them. Taken together, our map shows the general information of nucleosome dynamics reasonably well. The 'nucleosome dynamics' map provides the new significant correlation with the degree of expression variety instead of their intensity. Furthermore, the associations with gene function and histone modification were also discussed here.

6. Genome-wide Characterization of Transcriptional Start Sites based on the huge number of sequences

Riu Yamashita[4], Nuankanya P. Sathira[5], Akinori Kanai[5], Kousuke Tanimoto[5], Takako Arauchi[5], Soutaro Kanematsu[5], Sumio Sugano[5], Yutaka Suzuki[5], and Kenta Nakai: [4]Frontier Research Initiative, IMS, [5]Graduate School of Frontier Sciences

Using DBTSS information, we characterized 140 million transcriptional start sites (TSSs) in 12 human cell types. Despite the large number of TSS clusters (TSCs), the TSCs with significant expression levels were rare (around 3%). We also observed significant characteristics difference in major TSCs: namely, highly ordered nucleosome structures, strong RNA polymerase II binding signals, more frequent translation. RNA Seq analysis of polysome-incorporated RNAs yielded direct evidence that those transcript products are actually used for protein translation. These result suggest that integrative transcriptome analysis provides a powerful tool to discriminate TSCs having clear biological significance from the other possible noise level transcriptions.

7. Classification of heterodimer interfaces using docking models and construction of scoring functions for the complex structure prediction

Yuko Tsuchiya, Eiji Kanamori[6], Haruki Nakamura[7], and Kengo Kinoshita: [6]Japan Biological Informatics Consortium, [7]Osaka University

The precise prediction of complex structures is required because the number of protein-protein complex structures is still smaller than the known protein-protein interactions. In a protein-protein docking simulation that is one of the popular approaches, scoring functions are often used to select putative models, which are designed to select the complex model with the most abundant interaction mode found among the known complexes, as the correct model. However, because the formation schemes of heterodimers are diverse, a single scoring function may be not sufficient to select the putative models other than the most abundant mode. In this study, we first constructed the classification method of heterodimers into individual interaction modes, based on the discriminative characters between near-native and decoy models in terms of the interface complementarities for the hydrophobicity, the electrostatic potential and the shape. We found four heterodimer types, and constructed the four scoring functions, each of which was optimized for each type.

8. Analyses on hydrophobicity and attractiveness of all-atom distance-dependent potentials

M. Shirota, T. Ishida and K. Kinoshita

Statistical all-atom distance-dependent potentials are widely used for the model evaluation in protein structure prediction. They are derived from the comparison between the observed atomic distance frequencies and those in reference states, where all of the atomic interactions are turned off. In this study, seven all-atom distance dependent potentials with different reference states were examined. The variations in the atom pair composition and those in the distance distributions produced systematic changes in the hydrophobic and attractive characteristics of the potentials, respectively. The preference of hydrophobic interactions improved the correlation of the energy to the structural quality, but decreased the selectivity of the native structure. The attractive characteristic was beneficial for template-based modeling targets, but the benefit was smaller in free modeling targets. These results demonstrated that the performance of the potentials are more dependent on their characteristics than on the accuracy of the reference states.

9. Development of a new meta-score for protein structure prediction using support vector regression.

M. Shirota, T. Ishida and K. Kinoshita

The development of an accurate scoring function of protein structures is crucial for protein structure prediction. Such a function should be able to recognize the native structure among many decoys (model structures) and to correlate with the quality of the model structures. We developed a meta-score for protein structures by combining seven all-atom distance-dependent potentials to fulfill these two requirements. The meta-score was designed to predict the GDT_TS score of the structure from the score of the component methods using support vector regression (SVR). The decoys in the CASP6 (Critical Assessment of Techniques for Protein Structure Prediction) experiment were used for training the meta-score and the decoys in the CASP7 experiments were used to assess the component potentials and the meta-score. As results, the meta-score achieved as good performances as the best of the component methods in both of the criteria. Our result would suggest the benefit of combining various evaluation techniques in model evaluation.

10. Assessing coexpression measures for a prediction of gene function.

T. Obayashi and K. Kinoshita

The information of gene coexpression is useful to predict gene function. Several databases have been constructed for gene coexpression in model organisms based on a large amount of publicly available gene expression data measured by GeneChip platforms. In these databases, Pearson's correlation coefficients (PCCs) of gene expression patterns are widely used as a measure of gene coexpression. Although coexpression measure or GeneChip summarization method would affect the performance of the constructing gene coexpression data, previous studies for these calculation procedures were tested with a small number of samples and a particular species. To effectively construct gene coexpression database, assessments with large-scale high-quality microarray data are required. We first examined characteristics of PCC and found some problems such that the optimal PCC threshold to retrieve functionally related genes was changed by constructing method of gene expression data and a focused gene function. This problem indicated that we could not directly compare the coexpression data provided by several coexpression databases or studies using the PCC values. In addition, we found all the problems we examined could be overcome when we used correlation ranks instead of correlation values, which were assessed by the performance of prediction of Gene Ontology annotation with large-scale gene expression data for four species; Arabidopsis, human, mouse and rat. The rank-based coexpression measures can make it possible to compare any coexpression data for any species directory.

11. Multi-dimensional correlations for gene coexpression and application to the large-scale data of *Arabidopsis*

K. Kinoshita and T. Obayashi

Biological functions are realized through the interactions of many gene products. Therefore, co-existence of gene product is a key to construct and to understand the complex biological networks, because co-existence is a necessary condition for the interactions. Recent improvements of DNA microarray technique, a large variety of gene expression data are available in public database, and they can be used to evaluate the strength of gene coexpression by calculating the correlation of expression levels among the genes over many experiments. However, gene expression levels are very different in each tissues in higher organisms, and in each cellular location in eukaryote in different cell state, thus the usual correlation measure can only evaluate the difference of tissues or cellular localizations, and cannot well elucidate the functional relationship from the coexpression of genes. To overcome this difficulty, we propose a new measure of coexpressions by expanding the usual correlation into multidimensional indexes. We used principal component analyses to identify the major factor of correlation of gene expressions at first, and then re-calculate the correlation by extracting the major components. The extraction of the major components repeatedly to get a set of correlation values for each pair of genes. In this study, we observed the correlation changes when the first ten principal components were extracted step-by-step in the large scale *Arabidopsis* expression data, and found that our new indexes can be a good major to know the functional relationship of the genes by examining a few examples and higher performance of GO term prediction by using support vector machine and the multidimensional index.

12. Study of the domain distribution and intrinsic disorder in highly interacting proteins (hubs) in human protein-protein interaction networks

A. Patil, K. Kinoshita, and H. Nakamura[7]

Hubs are proteins with a large number of interactions in protein-protein interaction networks. By virtue of their ability to interact with multiple proteins, they are central parts of the interaction network defining its function and stability. Intrinsic disorder, or the lack of a stable structure in vivo, and high surface charge have been previously identified as the causes of multi-specific interactions in hubs. We extended this work by looking at the prevalence of multi-domain architectures and the functional nature of the domains found in hubs and non-hubs. We found that hubs have a greater tendency towards a multi-domain architecture i.e. have 4 or more domains, as compared to non-hubs. Hubs are also more likely to have kinase and adaptor domains, like SH2 or SH3, both of which are involved in promiscuous binding. Non-hubs are more likely to have DNA-binding domains or catalysis domains. Intrinsic disorder levels are higher in single domain hubs than in multi-domain hubs hinting at a complementary relationship between the two characteristics in promoting promiscuous binding.

Publications

Higurashi, M,. Ishida, T. and Kinoshita, K. PiSite: a database of protein interaction sites using multiple binding states in the PDB. Nucleic Acids Res. 37: D360-364, 2009.

Kinoshita, K. and Obayashi, T. Multi-dimensional correlations for gene coexpression and application to the large-scale data of Arabidopsis. Bioinformatics. 25: 2677-2684, 2009.

Maeda, M. and Kinoshita, K. Development of new indices to evaluate protein-protein interfaces: Assembling space volume, assembling space distance, and global shape descriptor. J. Mol. Graph. Mod. 27: 706-711, 2009.

Maruyama, F,. Kobata, M,. Kurokawa, K,. Nishida, K,. Sakurai, A,. Nakano, K,. Nomura, R,. Kawabata, S,. Oshima, T,. Nakai, K,. Hattori, M,. Hamada, S. and Nakagawa, I. Comparative genomic analyses of Streptococcus mutans provide insights into chromosomal shuffling and species-specific content. BMC Genomics. 10: 358, 2009.

Mizuno, M,. Kitada, K,. Nakai, K. and Sarai, A. PrognoScan: a new database for meta-analysis of the prognositic value of genes. BMC Med. Genomics. 2: 18, 2009.

Nishida, K,. Frith, M. and Nakai, K. Pseudocounts for transcription factor binding sites. Nucl. Acids Res. 37(3): 939-944, 2009.

Obayashi, T,. Hayashi, S,. Saeki, M,. Ohta, H. and Kinoshita, K. ATTED-II provides coexpressed gene networks for Arabidopsis. Nucleic Acids Res. 37: D987-991, 2009.

Obayashi, T. and Kinoshita, K. Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. DNA Res. 16: 249-60, 2009.

Sathira, N,. Yamashita, R,. Tanimoto, K,. Kanai, A,. Arauchi, T,. Kanematsu, S,. Nakai, K,. Suzuki, Y. and Sugano, S. Characterization of transcription start sites of putative non-coding RNAs by multifaceted use of massively paralleled sequencer. DNA Res. in press.

Shirota, M,. Ishida, T. and Kinoshita, K. Analyses on hydrophobicity and attractiveness of all-atom distance-dependent potentials. Protein Sci. 18: 1906-1915, 2009.

Shirota, M,. Ishida, T. and Kinoshita, K. Development of a new meta-score for protein structure prediction from seven all-atom distance dependent potentials using support vector regression. Genome Informatics, 23: 149-158, 2009.

Sierro, N,. Li, S,. Suzuki, Y,. Yamashita, R. and Nakai, K. Spatial and temporal preferences for trans-splicing in Ciona intestinalis revealed by EST-based gene expression analysis. Gene. 430: 44-49, 2009.

Tanaka, Y. and Nakai, K. An assessment of prediction algorithms for nucleosome positioning. Genome Informatics. 23: 169-178, 2009.

Tanaka, Y,. Yoshimura, I. and Nakai, K. Positional variations among heterogeneous nucleosome maps give dynamical information on chromatin. Chromosoma. in press.

Tsuchihara, K,. Suzuki, Y,. Wakaguri, H,. Irie, T,. Tanimoto, K,. Hashimoto, S,. Matsushima, K,. Mizushima-Sugano, J,. Yamashita, R,. Nakai, K,. Bentley, D,. Esumi, H. and Sugano, S. Massive transcriptional start site analysis of human genes in hypoxia cells. Nucl. Acids Res. 37: 2249-2263, 2009.

Tsuchiya, Y,. Kanamori, E,. Nakamura, H. and Kinoshita, K. Classification of hetero-dimer interfaces using docking models and construction of scoring functions for the complex structure prediction. Adv. Appl. Bioinfo. Chem. 2: 79-100, 2009.

Vandenbon, A. and Nakai, K. Modeling tissue-specific structural patterns in human and mouse promoters. Nucl. Acids Res. 38(1): 17-25, 2010.

Yamashita, R,. Wakaguri, H,. Sugano, S,. Suzuki, Y. and Nakai, K. DBTSS provides a tissue specific dynamic view of Transcription Start Sites. Nucl. Acids Res. 38: D98-D104, 2010.

中井謙太．第５章　ゲノムを読み解く―バイオインフォマティクス＆付録バイオインフォマティクス理解のためのデータベース．浅島誠・黒岩常祥・小原雄治編．現代生物科学入門１　ゲノム科学の基礎．岩波書店．201−255，2009．

Human Genome Center

Department of Public Policy
公共政策研究分野

Associate Professor	Kaori Muto, Ph.D.	准教授	保健学博士	武 藤	香 織
Project Assistant Professor	Hyongoo Hong, Ph.D.	特任助教	学術博士	洪	賢 秀
Project Assistant Professor	Ayako Kamisato, M.A.	特任助教	法学修士	神 里	彩 子

Department of Public Policy works for three major missions; public policy studies on translational research, its application and impact on society; research ethics consultation for scientists to comply with ethical guidelines and to build public trust; and development of "minority-centered" scientific communication. We've published various studies including informed consent on genetic epidemiological studies, public and professional attitudes towards genetic testing, regulation on quality assurance of genetic testing in Japan, recent revision on organ transplantation act in Japan, and surrogacy issues in several countries. In overall, Japanese regulative options and approach on medicine are unique and difficult to harmonize with other countries. We held Sci/Art Café twice at the Medical Science Museum as one of the outreach activities via art.

1. Informed consent procedures in genetic epidemiological studies

We conducted an international web-based survey to obtain information about the informed consent (IC) procedures used in genetic epidemiological studies of multifactorial diseases. In total, 53 responses from 15 countries were obtained. We found that support personnel such as research coordinators took charge of a large part of the IC procedures, especially in the United States. Although most support personnel had documented and/or verbal confidentiality agreements and all had undertaken IC training, accreditation examinations were only conducted in 60% of studies. Peripheral blood specimens were reported to be a major source for DNA extraction, whereas noninvasive methods were used in some studies. To undertake high-quality genetic epidemiological studies, participation of a large number of volunteers is essential for independent sets of samples that allow confirmation of results. On the basis of the survey results obtained, support personnel dedicated to completing the IC procedures are reported to be beneficial and indispensable in alleviating the burden on medical doctors, helping participants to make autonomous decisions and promoting genome research. The establishment of a training program and accreditation system for such personnel is warranted, especially in Japan, where medical staff usually finds it difficult to conduct IC procedures in clinical settings due to heavy workloads.

2. Surveys on public and professional attitudes towards genetic testing

Recent advances in studies on human genetics have led to the use of genetic information in various applications. We conducted a survey to know the opinions of healthcare providers in Japan on new genetic testing services, such as direct-to-consumer (DTC) genetic testing. A total of 1124 general practitioners and 294 clinical geneticists replied to our questionnaire. Thirty-

eight percent of the general practitioners and 68.4% of the clinical geneticists were aware of DTC genetic testing. Some physicians had gained information on this service through their patients or commercial activities of companies providing such services. General practitioners expected that DTC genetic testing would be convenient, promote preventive medicine, provide personalized services and would enable to maintain confidentiality of information. Clinical geneticists showed greater concern with regard to the reliability of the results, provision of information/counseling and the understanding of results. Awareness of DTC genetic testing enhances general practitioners' positive opinions of it. Although the market for DTC genetic testing in Japan may still be limited, it is possible that general practitioners will play a role in the provision of DTC genetic testing services in the future. On the basis of their knowledge and experience, clinical geneticists should provide information to both healthcare providers and to the public.

3. Regulation on quality assurance of genetic testing in Japan

As one of the countries that have invested greatly in the field of bioscience, Japan is facing difficulties introducing human genetic research to the market. A key issue is how to regulate the quality of genetic testing. Since genetic testing is a part of clinical laboratory tests, the regulatory framework for these tests should cover the regulation of genetic testing. Nevertheless, the quality of clinical laboratory tests has been regulated largely by the authority of medical professionals. The fact that genetic testing can be provided without supervision of medical professionals reveals the necessity for the regulation of quality of genetic testing. While medical geneticists have publicly criticized direct-to-consumer (DTC) genetic testing, a group of industries related to DTC genetic testing have established self-regulatory guidelines on the quality control of genetic analysis, based on the OECD guidelines. We described the regulatory framework for clinical laboratory tests including genetic tests, and the gaps in regulation, which are particularly highlighted by the appearance of DTC genetic testing.

4. Recent revision on organ transplantation act in Japan and its implications

Generally, in Japan, the donor-recipient relationship is primarily contained within families, not only in living organ transplantations, but also in cadaveric transplantation. Surviving family members perform multiple roles, sometimes action against the wishes of the deceased. Nevertheless, both the definition of a family and the moral basis of their decision remain obscure. This ambiguity may have led to the restriction of cadaveric organ donation in Japan, while promoting living organ transplantation. The 2009 revision was made with the intention of increasing cadaveric organ donations. However, the revision further enhanced this ambiguity, by introducing "prioritized cadaveric donation", and by widening the role of family decision-making. We suggest an interim conclusion: the problem is not whether the opt-in system or the opt-out system as a measure to respect the will of the deceased person will work better, but whether the fair distribution or the enlarged familial determination is better for overcoming the shortage of organs for transplantation.

5. Surrogacy regulation in five countries

As of 2008, surrogacy is legal and openly practiced in various places; Japan, however, has no regulations or laws regarding surrogacy. This paper reports the situation of surrogacy in Japan and in five other regions (the USA, the UK, Taiwan, Korea and France) to clarify the pros and cons of prohibiting surrogacy, along with the problems and issues relating to surrogacy compensation. Not only in a country such as France that completely prohibits surrogacy within the country, but also in a country such as the UK that allows non-commercial surrogacy, infertile couples travel overseas for the purpose of surrogacy. In addition, some couples might seek underground surrogacy if the government prohibits surrogacy. If an intended parent couple and a surrogate make an agreement among themselves and then a problem occurs, they cannot ask for support from professionals or bring a case to court, as can be observed in South Korea and Taiwan. We also conclude that there is little difference between commercial surrogacy and non-commercial surrogacy in the absence of a clear definition of 'reasonable expenses.' In the UK, the law does not allow surrogates to receive compensation. However, in reality, there may be little difference between the amounts paid to surrogates for profit in the US and those paid to surrogates for reasonable expenses in the UK. We conclude that the issue of surrogacy demands further discussion in Japan.

6. Research in progress

We've been conducting other studies below;

- Ethical, legal and social implications of biobanking in East Asia

- Development and evaluation of communication methods with participants of Biobank Japan and other long-term studies
- Analysis of roles of research coordinators for better recruitment and building trust from participants
- Ethical, legal and social implications on stem cell studies including animal-human chimeric embryos and iPS cell banking
- Bench side research ethics consultation and quality assurance of research ethics committees
- Science communication through art and ethical challenges of biomaterial art
- Beyond Declaration of Istanbul: protection of living liver donors in Japan

Publications

1. Ohata, T., Tsuchiya, A., Watanabe, M., Sumida, T. and Takada, F. Physicians' opinion for 'new' genetic testing in Japan. Journal of Human Genetics. 54: 203-208, 2009.
2. Yamamoto, M., Funamizu, M., Muto, K. and Hata, A. International web-based survey of informed consent procedures in genetic epidemiological studies: towards the establishment of a research coordinator accreditation system. Journal of Human Genetics. 54: 713-720, 2009.
3. Inoue, Y and Hong, H. Cadaveric donation and the family: perspectives from the legal history of Japan. Asian Bioethics Review. 1 (4): 342-353, 2009.
4. Watanabe, M., Muto, K., Ohata, T., Takada, F. Problems in the regulation of genetic tests in Japan: what can we learn from DTC genetic tests? Public Health and Genomics. Published online: Oct 26, 2009.
5. Semba, Y., Chang C., Hong, H., Kamisato, A., Kokado, M., Muto, K. Surrogacy: donor conception regulation in Japan. Bioethics: 1467-8519. Published Online: Nov 30, 2009
6. 洪賢秀，井上悠輔，武藤香織．ヒト幹細胞研究における規制と倫理についての日本と世界の現状．再生医療の最前線2010．実験医学（増刊）：36–41，2010．