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

The large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies are the basis for understanding life as a molecular system and for bringing the genomic revolution to society. We are developing bioinformatics technologies to integrate and interpret such large-scale datasets, especially for medical and pharmaceutical applications.

1. KEGG DISEASE and KEGG DRUG

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, the organism, and the ecosystem from genomic information. This Laboratory 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. The KEGG DISEASE database (http://www.genome.jp/kegg/disease/) is a collection of disease entries capturing knowledge on genetic and environmental perturbations. The database currently contains about 1,000 entries for diseases with known genetic fators and infectious diseases with known pathogen genomes. The KEGG DRUG database (http://www.genome.jp/kegg/drug/) contains about 10,000 entries for the approved drugs in Japan, USA, and Europe unified based on the chemical structures and/or the chemical components. Each entry is associated with target, metabolizing enzyme, and other molecular interaction network information, enabling understanding of the perturbants in the context of KEGG pathway maps. Furthermore, it contains another type of molecular network information, namely, knowledge about chemical structure transformation patterns during the history of drug development as represented in the KEGG DRUG structure maps. KEGG DISEASE and KEGG DRUG now constitute the core part of KEGG MEDICUS (http://www.genome.jp/kegg/medicus.html), an integrated information resource of diseases, drugs, and health-related substances, aiming to bring the genomic revolution to society.

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

Toshiaki Katayama, Shuichi Kawashima, Akihiro Nakaya[1] and Minoru Kanehisa: [1]Bioresource Science Branch, Brain Research Institute, Niigata University

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 6.6 million genes from over 1,560 organisms as of August 2011. 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 1,139,058 clusters including 715,367 singleton clusters from 6,686,868 protein coding genes. Among them, 5,704 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. The resulted table of orthologous genes is made available through the KEGG FTP site.

3. 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), 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/.

4. Whole transcriptome analysis of *Anopheles stephensi* 14 days post *Plasmodium yoelii* infection using RNA-seq

Shuichi Kawashima, Lucky R. Runtuwene[1], Yutaka Suzuki[2], Sumio Sugano[2], Kenta Nakai[3], Yuki Eshita[1]: [1]Department of Infectious Disease Control, Faculty of Medicine, Oita University, [2]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, [3]Laboratory of Functional Analysis in Silico, Institute of Medical Science, The University of Tokyo

Anopheles stephensi is one of the major vectors of human malaria in Asia. However, transcriptome of malaria-infected *An. stephensi* have not been studied enough yet compared to that of *Anopheles gambiae* which is known as the most important vector of *Plasmodium falciparum*. Thus we conducted an RNA-seq experiment against *An. stephensi* of 14 days post-plasmodium-infection and non-infection to find the mosquito genes that are affected by plasmodium infection. Illumina Genome Analyzer produced approximately 73 and 63 millions short reads for each control and target samples respectively. Because *Ae. stephensi* genome sequence is not available yet, we assembled the short reads *de novo* to generate transcript sequences as the reference for the short reads mapping. We obtained 28,074 transcripts by using the Trinity assembler and then mapped the short reads to the transcripts by using the Bowtie aligner. A statistical test was carried out by the DEGseq which is an R package in the Bioconductor. MARS method implemented in the DEGseq showed 5,563 transcripts were expressed differentially between both samples with p-value of 0.001 or less. Among them, a total of 79 and 787 genes were up- and down-regulated more than 2 folds in a natural logarithm scale, respectively. While most of them were homologous to genes annotated with "putative uncharacterized gene", we found two Trypsin homologs in the up-regulated genes and some histones in the down-regulated genes.

5. Whole transcriptome analysis of *Aedes aegypti* 14 days post-dengue infection using RNA-seq

Lucky R. Runtuwene[1], Shuichi Kawashima, Yutaka Suzuki[2], Sumio Sugano[2], Kenta Nakai[3], Ryuichiro Maeda[4], Chihiro Sugimoto[5], Tomohiko Takasaki[6], Ichiro Kurane[6] and Yuki

Eshita[1]: [1]Department of Infectious Disease Control, Faculty of Medicine, Oita University, [2]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, [3]Laboratory of Functional Analysis in Silico, Institute of Medical Science, The University of Tokyo, [4]Obihiro University of Agriculture and Veterinary Medicine, [5]Department of Collaboration and Education, Research Center for Zoonosis Control, Hokkaido University, [6]National Institute of Infectious Diseases

Dengue virus (DENV) is the causative agent of fatal disease which has no vaccine available. Breakage of transmission is still the core of prevention. The prospects of innovating new prevention techniques include genetically modified dengue-resistant transgenic mosquitoes. That innovation begins with a search of mosquito genes that are affected by dengue infection. To address the issue, we analyzed the whole transcriptome of 14 days post-dengue-infection *Aedes aegypti* by using RNA-seq technique. Illumina Genome Analyzer produced approximately 60 millions short reads for each control and target samples. These were mapped onto 14,839 and 14,394 genes respectively. A statistical test showed 9,059 genes were expressed differentially between both samples with false discovery rate (FDR) of 0.01 or less. Among them, a total of 91 and 394 genes were up- and down-regulated more than 7.39 folds, respectively. Histones were the most prominent in the up-regulated genes. In addition, five of the 22 HSP20 homologous genes were also significantly up-regulated in the DENV infected mosquito. Although HSP 20s are known to serve diverse protective functions, its detailed functions in insects are still unclear. Induced specific HSP20s we found suggest that those are responsible to the viral infection.

6. Biogem: an effective tool based approach for scaling up open source software development in bioinformatics

Bonnal, R.J.P.[1], Aerts, J.[2], Githinji, G.[3], Goto, N.[4], MacLean, D.[5], Miller C.A.[6], Mishima, H.[7], Pagani, M.[1], Ramirez-Gonzalez, R.[8], Smant G.[9], Strozzi, F.[10], Syme, R.[11], Vos, R.[12], Wennblom, T.J.[13], Woodcroft, B.J.[14], Katayama, T., Prins, P.[9]: [1]Integrative Biology Program, Istituto Nazionale Genetica Molecolare, Milan, Italy, [2]ESAT/SCD, Faculty of Engineering and IBBT Future Health Department, University of Leuven, Belgium, [3]KEMRI-Wellcome Trust Research Program, Kilifi, Kenya, [4]Research Institute for Microbial Diseases, Osaka University, Japan, [5]The Sainsbury Laboratory, Norwich, UK, [6]Biology Department, Boston College, USA, [7]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Japan, [8]The Genome Analysis Centre, Norwich, UK, [9]Laboratory of Nematology, Wageningen University, the Netherlands, [10]Parco Tecnologico Padano, Lodi, Italy, [11]Dept Environment & Agriculture, Curtin University, Australia, [12]NCB Naturalis, Leiden, the Netherlands, [13]Silicon Life Sciences, Minneapolis, USA, [14]Department of Biochemistry and Molecular Biology, University of Melbourne, Australia

Biogem provides a software development environment for the Ruby programming language, which encourages communitybased software development for bioinformatics while lowering the barrier to entry and encouraging best practices. Biogem, with its targeted modular and decentralized approach, software generator, tools, and tight web integration, is an improved general model for scaling up collaborative open source software development in bioinformatics. Availability: Biogem and modules are free and open source software. Biogem runs on all systems that support recent versions of Ruby, including Linux, Mac OS X and Windows. Further information at http://www.biogems.info. A tutorial is available at http://www.biogems.info/howto.html.

7. TogoWS: integrated SOAP and REST APIs for interoperable bioinformatics Web services

Toshiaki Katayama

Web services have become widely used in bioinformatics analysis, but there exist incompatibilities in interfaces and data types, which prevent users from making full use of a combination of these services. Therefore, we have developed the TogoWS service to provide an integrated interface with advanced features. In the TogoWS REST (REpresentative State Transfer) API (application programming interface), we introduce a unified access method for major database resources through intuitive URIs that can be used to search, retrieve, parse and convert the database entries. Recently, converters for CSV, GenBank/DDBJ/EMBL and HMMER3 into Semantic Web data formats including RDF/XML and Turtle are added. The TogoWS service is freely available at: http://togows.dbcls.jp/.

8. TogoDB: Instantly publish your research material as a public database

Toshiaki Katayama, Mitsuteru Nakao[1] and Toshihisa Takagi[2,3]: [1]Database Center for Life Science, ROIS, [2]National Bioscience Database Center, JST, [3]Graduate School of Frontier Sciences, The University of Tokyo

Supplemental materials are often provided as separate files downloadable from the publisher's site along with the publication of a journal article. However, these data are not fully utilized since they are not available in the form of regular biological databases and hard to find by the popular Web search engines. TogoDB is a simple and intuitive database system to publish tabular formatted data instantly on the Web. Users can upload their research materials to the TogoDB through the simple web interface and the data will be made available as a fully functional database in a minute or two. TogoWS is an integrated and uniformed interface for the major bioinformatics web services and also provides REST API for the contents in the TogoDB. Recently, we extended TogoDB and TogoWS to be used as a consolidated platform for the Semantic Web by adding a metadata editor and an automatic Resource Description Framework (RDF) dumper. Users can specify their choice of a Property for each column and a Class for the range of each cell from a pull-down menu for RDF, RDFS, DC, SKOS and FOAF vocabularies, or directly assign a URI from any other ontologies. This system fills the gap between user's data and major public databases to deliver effective variations in the Linked Data.

Publications

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res*. 40, D109-114, 2012

Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., Kanehisa, M. Network-based analysis and characterization of adverse drug-drug interactions. *J. Chem. Inf. Model*. 51, 2977-2985, 2011

Kotera, M., Tokimatsu, T., Kanehisa, M., Goto, S. MUCHA: multiple chemical alignment algorithm to obtain building block substructures of orphan metabolites. *BMC Bioinformatics* 12, S1, 2011

Bonnal, R.J.P., Aerts, J., Githinji, G., Goto, N., MacLean, D., Miller C.A., Mishima, H., Pagani, M., Ramirez-Gonzalez, R., Smant G., Strozzi, F., Syme, R., Vos, R., Wennblom, T.J., Woodcroft, B.J., Katayama, T., Prins, P. Biogem: an effective tool based approach for scaling up open source software development in bioinformatics. *Bioinformatics*, 2012, *in press*

Guberman, J.M., Ai, J., Arnaiz, O., Baran, J., Blake, A., Baldock, R., Chelala, C., Croft, D., Cros, A., Cutts, R.J., Di Génova, A., Forbes, S., Fujisawa, T., Gadaleta, E., Goodstein, D.M., Gundem, G., Haggarty, B., Haider, S., Hall, M., Harris, T., Haw, R., Hu, S., Hubbard, S., Hsu, J., Iyer, V., Jones, P., Katayama, T., Kinsella, R., Kong, L., Lawson, D., Liang, Y., Lopez-Bigas, N., Luo, J., Lush, M., Mason, J., Moreews, F., Ndegwa, N., Oakley, D., Perez-Llamas, C., Primig, M., Rivkin, E., Rosanoff, S., Shepherd, R., Simon R, Skarnes, B., Smedley, D., Sperling, L., Spooner, W., Stevenson, P., Stone, K., Teague, J., Wang, J., Wang, J., Whitty, B., Wong, D.T., Wong-Erasmus, M., Yao, L., Youens-Clark, K., Yung, C., Zhang, J., Kasprzyk, A. BioMart Central Portal: an open database network for the biological community. *Database*, bar041, 2011.

Katayama, T., Wilkinson, M.D., Vos, R., Kawashima, T., Kawashima, S., Nakao, M., Yamamoto, Y., Chun, H.W., Yamaguchi, A., Kawano, S., Aerts, J., Aoki-Kinoshita, K.F., Arakawa, K., Aranda, B., Bonnal, R.J., Fernández, J.M., Fujisawa, T., Gordon, P.M., Goto, N., Haider, S., Harris, T., Hatakeyama, T., Ho, I., Itoh, M., Kasprzyk, A., Kido, N., Kim, Y.J., Kinjo, A.R., Konishi, F., Kovarskaya, Y., von Kuster, G., Labarga, A., Limviphuvadh, V., McCarthy, L., Nakamura, Y., Nam, Y., Nishida, K., Nishimura, K., Nishizawa, T., Ogishima, S., Oinn, T., Okamoto, S., Okuda, S., Ono, K., Oshita, K., Park, K.J., Putnam, N., Senger, M., Severin, J., Shigemoto, Y., Sugawara, H., Taylor, J., Trelles, O., Yamasaki, C., Yamashita, R., Satoh, N., Takagi, T. The 2nd DBCLS BioHackathon: interoperable bioinformatics Web services for integrated applications. *J Biomed Semantics*. 2:4 2011

Human Genome Center

Laboratory of DNA Information Analysis
Laboratory of Sequence Data Analysis
Laboratory of Functional Genomics

DNA情報解析分野
シークエンスデータ情報処理分野
ゲノム機能解析分野

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	Teppei Shimamura, Ph.D.	助　教	博士(工学)	島　村　徹　平
Project Assistant Professor	Atsushi Niida, Ph.D.	特任助教	博士(理学)	新井田　厚　司
Project Assistant Professor	Yoshinori Tamada, Ph.D.	特任助教	博士(情報学)	玉　田　嘉　紀
Associate Professor	Tetsuo Shibuya, Ph.D.	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
Associate Professor	Masao Nagasaki, Ph.D.	准教授	博士(理学)	長　崎　正　朗

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. Gene Network Analysis and Cancer Systems Biology

a. A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition

Shimamura T, Imoto S, Shimada Y[1], Hosono Y[1], Niida A, Nagasaki M, Yamaguchi R, Takahashi T[1], Miyano S: [1]Nagoya University Graduate School of Medicine

Patient-specific analysis of molecular networks is a promising strategy for making individual risk predictions and treatment decisions in cancer therapy. Although systems biology allows the gene network of a cell to be reconstructed from clinical gene expression data, traditional methods, such as Bayesian networks, only provide an averaged network for all samples. Therefore, these methods cannot reveal patient-specific differences in molecular networks during cancer progression. In this study, we developed a novel statistical method called Network-Profiler, which infers patient-specific gene regulatory networks for a specific clinical characteris-

tic, such as cancer progression, from gene expression data of cancer patients. We applied NetworkProfiler to microarray gene expression data from 762 cancer cell lines and extracted the system changes that were related to the epithelial-mesenchymal transition (EMT). Out of 1732 possible regulators of E-cadherin, a cell adhesion molecule that modulates the EMT, NetworkProfiler, identified 25 candidate regulators, of which about half have been experimentally verified in the literature. In addition, we used NetworkProfiler to predict EMT-dependent master regulators that enhanced cell adhesion, migration, invasion, and metastasis. In order to further evaluate the performance of NetworkProfiler, we selected Krueppel-like factor 5 (KLF5) from a list of the remaining candidate regulators of E-cadherin and conducted in vitro validation experiments. As a result, we found that knockdown of KLF5 by siRNA significantly decreased E-cadherin expression and induced morphological changes characteristic of EMT. In addition, in vitro experiments of a novel candidate EMT-related microRNA, miR-100, confirmed the involvement of miR-100 in several EMT-related aspects, which was consistent with the predictions obtained by NetworkProfiler.

b. Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers

Tamada Y, Imoto S, Araki H[2], Nagasaki M, Print C[3], Charnock-Jones DS[4], Miyano S: [2]Cell Inovator, Inc. [3]University of Auckland [4]University of Cambridge

We developed a novel algorithm to estimate genome-wide gene networks consisting of more than 20 000 genes from gene expression data using nonparametric Bayesian networks. Due to the difficulty of learning Bayesian network structures, existing algorithms cannot be applied to more than a few thousand genes. Our algorithm overcomes this limitation by repeatedly estimating subnetworks in parallel for genes selected by neighbor node sampling. Through numerical simulation, we confirmed that our algorithm outperformed a heuristic algorithm in a shorter time. We applied our algorithm to microarray data from human umbilical vein endothelial cells (HUVECs) treated with siRNAs, to construct a human genome-wide gene network, which we compared to a small gene network estimated for the genes extracted using a traditional bioinformatics method. The results showed that our genome-wide gene network contains many features of the small network, as well as others that could not be captured during the small network estimation. The results also revealed master-regulator genes that are not in the small network but that control many of the genes in the small network. These analyses were impossible to realize without our proposed algorithm.

c. SiGN-SSM: open source parallel software for estimating gene networks with state space models

Tamada Y, Yamaguchi R, Imoto S, Hirose O, Yoshida R[5], Nagasaki M, Miyano S: [5]Institute of Statistical Mathematics

SiGN-SSM is an open-source gene network estimation software able to run in parallel on PCs and massively parallel supercomputers. The software estimates a state space model (SSM), that is a statistical dynamic model suitable for analyzing short time and/or replicated time series gene expression profiles. SiGN-SSM implements a novel parameter constraint effective to stabilize the estimated models. Also, by using a supercomputer, it is able to determine the gene network structure by a statistical permutation test in a practical time. SiGN-SSM is applicable not only to analyzing temporal regulatory dependencies between genes, but also to extracting the differentially regulated genes from time series expression profiles. SiGN-SSM is distributed under GNU Affero General Public Licence (GNU AGPL) version 3 and can be downloaded at http://sign.hgc.jp/signssm/. The pre-compiled binaries for some architectures are available in addition to the source code. The pre-installed binaries are also available on the Human Genome Center supercomputer system. The online manual and the supplementary information of SiGN-SSM is available on our web site.

d. Inferring contagion in regulatory networks

Fujita A[6], Sato JR[7], Demas MAA[8], Yamaguchi R, Shimamura T, Ferreira CE[8], Sogayar MC[8], Miyano S: [6]RIKEN, ISLiM [7]Universidade Federal do ABC: [8]University of São Paulo

Several gene regulatory network models containing concepts of directionality at the edges have been proposed. However, only a few reports have an interpretable definition of directionality. Here, differently from the standard causality concept defined by Pearl, we introduce the concept of contagion in order to infer directionality at the edges, i.e., asymmetries in gene expression dependences of regulatory networks. Moreover, we present a bootstrap algorithm in

order to test the contagion concept. This technique was applied in simulated data and, also, in an actual large sample of biological data. Literature review has confirmed some genes identified by contagion as actually belonging to the TP53 pathway.

e. Searching optimal Bayesian network structure on constraint search space: super-structure approach

Imoto S, Kojima K, Perrier E, Tamada Y, Miyano S

Optimal search on Bayesian network structure is known as an NP-hard problem and the applicability of existing optimal algorithms is limited in small Bayesian networks with 30 nodes or so. To learn larger Bayesian networks from observational data, some heuristic algorithms were used, but only a local optimal structure is found and its accuracy is not high in many cases. In this paper, we review optimal search algorithms in a constraint search space; The skeleton of the learned Bayesian network is a sub-graph of the given undirected graph called super-structure. The introduced optimal search algorithm can learn Bayesian networks with several hundreds of nodes when the degree of super-structure is around four. Numerical experiments indicate that constraint optimal search outperforms state-of-the-art heuristic algorithms in terms of accuracy, even if the super-structure is also learned by data.

f. Parallel algorithm for learning optimal Bayesian network structure

Tamada Y, Imoto S, Miyano S

We developed a parallel algorithm for the score-based optimal structure search of Bayesian networks. This algorithm is based on a dynamic programming (DP) algorithm having $O(n \cdot 2^n)$ time and space complexity, which is known to be the fastest algorithm for the optimal structure search of networks with n nodes. The bottleneck of the problem is the memory requirement, and therefore, the algorithm is currently applicable for up to a few tens of nodes. While the recently proposed algorithm overcomes this limitation by a space-time trade-off, our proposed algorithm realizes direct parallelization of the original DP algorithm with $O(n^\sigma)$ time and space overhead calculations, where $\sigma > 0$ controls the communication-space trade-off. The overall time and space complexity is $O(n^{\sigma+1} \cdot 2^n)$. This algorithm splits the search space so that the required communication between independent calculations is minimal. Because of this advantage, our algorithm can run on distributed memory supercomputers. Through computational experiments, we confirmed that our algorithm can run in parallel using up to 256 processors with a parallelization efficiency of 0.74, compared to the original DP algorithm with a single processor. We also demonstrate optimal structure search for a 32-node network without any constraints, which is the largest network search presented in literature.

g. Gene network inference and visualization tools for biologists: application to new human transcriptome datasets

Hurley D[3], Araki H[3], Tamada Y, Dunmore B[4], Sanders D[4], Humphreys S[4], Affara M[4], Imoto S, Yasuda K[9] Tomiyasu Y[9], Tashiro K[9], Savoie C[10], Cho V[3], Smith S[4], Kuhara S[9], Miyano S, Charnock-Jones DS[4], Crampin EJ3, Print CG[3]: [9]Kyushu University [10]GNI, Inc.

Gene regulatory networks inferred from RNA abundance data have generated significant interest, but despite this, gene network approaches are used infrequently and often require input from bioinformaticians. We have assembled a suite of tools for analysing regulatory networks, and we illustrate their use with microarray datasets generated in human endothelial cells. We infer a range of regulatory networks, and based on this analysis discuss the strengths and limitations of network inference from RNA abundance data. We welcome contact from researchers interested in using our inference and visualization tools to answer biological questions.

h. Frequent pathway mutations of splicing machinery in myelodysplasia

Yoshida K[11], Sanada M[11], Shiraishi Y, Nowak D[12], Nagata Y[11], Yamamoto R[13], Sato Y[11], Sato-Otsubo A[11], Kon A[11], Nagasaki M, Chalkidis G, Suzuki Y[14], Shiosaka M[11], Kawahata R[11], Yamaguchi T[13], Otsu M[13], Obara N[15], Sakata-Yanagimoto M[15], Ishiyama K[16], Mori H[17], Nolte F[12], Hofmann WK[12], Miyawaki S[16], Sugano S[14], Haferlach C[18], Koeffler HP[19], Shih LY[20], Haferlach T[18], Chiba S[15], Nakauchi H[13], Miyano S, Ogawa S[11]: [11]Graduate School of Medicine, University of Tokyo [12]University of Heidelberg [13]Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, University of Tokyo [14]Department of Medical Genome Sciences, University of Tokyo [15]Institute of Clinical Medicine, University of Tsukuba [16]Tokyo Metropolitan Ohtsuka Hospital [17]Showa University Fujigaoka Hospital [18]Munich Leukemia Laboratory [19]Cedars-Sinai Me-

dical Center [20]Chang Gung University

By using the supercomputer system of Human Genome Center, we contributed to data analysis of whole-exome sequencing data of 29 myelodysplasia specimens, which unexpectedly revealed novel pathway mutations involving multiple components of the RNA splicing machinery, including U2AF35, ZRSR2, SRSF2 and SF3B1.

i. Long non-coding RNA HOTAIR regulates Polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers

Kogo R[21], Shimamura T, Mimori K[22], Kawahara K[21], Imoto S, Sudo T[21], Tanaka F[21], Shibata K[21], Suzuki A[21], Komune S[23], Miyano S, Mori M[24]: [21]Medical Institute of Bioregulation, Kyushu University [22]Kyushu University Beppu Hospital [23]Graduate School of Medicine, Kyushu University [24]Gastroenterological Surgery, Osaka University

We contributed to data analysis in the study of examining the status and function of HOTAIR in stage IV colorectal cancer (CRC) patients who have liver metastases and a poor prognosis. The analysis showed that HOTAIR expression levels were higher in cancerous tissues than corresponding noncancerous tissues and high HOTAIR expression correlated tightly with the presence of liver metastasis. Moreover, patients with high HOTAIR expression had a relatively poorer prognosis. In a subset of 32 CRC speciments, gene set enrichment analysis using cDNA array data revealed a close correlation between expression of HOTAIR and members of the PRC2 complex (SUZ12, EZH2 and H 3K27me3). These findings suggest that HOTAIR expression is associated with a genome-wide reprogramming of PRC2 function not only in breast cancer but also in CRC, where upregulation of this long ncRNA may be a critical element in metastatic progression.

j. N-cadherin expression is a potential survival mechanism of gefitinib-resistant lung cancer cells

Yamauchi M[32], Yoshino I[32], Yamaguchi R, Shimamura T, Nagasaki M, Imoto S, Niida A, Koizumi F[33], Kohno T[33], Yokota J[33], Miyano S, Gotoh N[32]: [32]Division of Systems Biomedical Technology, Institute of Medical Science, University of Tokyo: [33]National Cancer Center Researh Institute

We contributed to design of measurement of gene expression data for systems understanding of gefitinib-sensitive NSCLC and gefitinib-resistant NSCLC. We used DNA microarrays to examine gene expression profiles of gefitinib-resistant PC9/ZD cells that are derived from gefitinib-sensitive PC9 cells and harbor a threonine to methionine mutation at codon 790 (T790M) in EGFR, a known mechanism of acquired resistance to gefitinib. Data analysis suggested showed that N-cadherin expression was significantly upregulated in PC9/ZD cells compared with PC9 cells. Inhibition of N-cadherin expression by siRNA or treatment with antibodies against N-cadherin induced apoptosis of PC9/ZD cells in association with reduced phosphorylation of Akt and Bad, a proapoptotic protein. Moreover, inhibition of Akt expression by siRNA or treatment with an inhibitor for phosphatidylinositol (PI)-3 kinase reduced survival of PC9/ZD cells. In addition, we found several N-cadherin-expressing lung cancer cells that showed inherent resistance to gefitinib treatment and reduced survival owing to siRNA-induced inhibition of N-cadherin expression. Thus, it appears that N-cadherin maintains the survival of the gefitinib-resistant lung cancer cells via the PI-3 kinase/Akt survival pathway. From these results, we propose that N-cadherin signaling contributes, at least in part, to the survival mechanisms of gefitinib-resistant NSCLC cells and that N-cadherin is a potential molecular target in the treatment of NSCLC. (AJCR0000074).

k. MRG-binding protein contributes to development of colorectal cancer

Yamaguchi K[31], Sakai M[32], Kim J[31], Tsunesumi S[31], Fujii T[31], Ikenoue T[31], Yamada Y[33], Akiyama Y[33], Muto Y[33], Yamaguchi R, Miyano S, Nakamura Y[32], Furukawa Y[31]: [31]Division of Clinical Genome Research, Advanced Clinical Research Center, Institute of Medical Science: [32]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science [33]Tokyo Hitachi Hospital

MRGBP (MORF4-related gene-binding protein; also known as chromosome 20 open reading frame 20) encodes a subunit of the transformation/transcription domain-associated protein (TRRAP)/tat-interacting protein 60 (TIP60)-containing histone acetyltransferase complex. To investigated the role MRGBP in colorectal carcinogenesis, we constributed to statistical data analysis of the differences between gene expression data of MRGBP-knockdown colorectral cell line and normal colorectral cell line. We applied Fisher exact test. Based gene sets based on GO

and annother functional annotation databases (KEGG, etc), we screened out significantly exressed gene sets together with functinal annotations.

2. Pathway Modeling, Simulation and Analysis

a. Comprehensive pharmacogenomic pathway screening by data assimilation

Hasegawa T, Yamaguchi R, Nagasaki M, Imoto S, Miyano S

We developed a computational method to comprehensively screen for pharmacogenomic pathway simulation models. A systematic model generation strategy is developed; candidate pharmacogenomic models are automatically generated from some prototype models constructed from existing literature. The parameters in the model are automatically estimated based on time-course observed gene expression data by data assimilation technique. The candidate simulation models are also ranked based on their prediction power measured by Bayesian information criterion. We generated 53 pharmacogenomic simulation models from five prototypes and applied the proposed method to microarray gene expression data of rat liver cells treated with corticosteroid. We found that some extended simulation models have higher prediction power for some genes than the original models.

b. Systems biology model repository for macrophage pathway simulation

Nagasaki M, Saito A, Fujita A[6], Tremmel G, Ueno K, Ikeda E, Jeong E, Miyano S

The Macrophage Pathway Knowledgebase (MACPAK) is a computational system which allows biomedical researchers to query and study the dynamic behaviors of macrophage molecular pathways. It integrates the knowledge of 230 reviews that were carefully checked by specialists for their accuracy and then converted to 230 dynamic mathematical pathway models. MACPAK comprises a total of 24,009 entities and 12,774 processes and is described in the Cell System Markup Language (CSML), an XML format that runs on the Cell Illustrator platform and can be visualized with a customized Cytoscape for further analysis. MACPAK can be accessed via an interactive website at http://macpak.csml.org. The CSML pathway models are available under the Creative Commons license.

c. MIRACH: Efficient model checker for quantitative biological pathway models

Koh CH[25], Nagasaki M, Saito A, Li C, Wong L[25], Miyano S: [25]National University of Singapore

Model checking is playing an increasingly important role in systems biology as larger and more complex biological pathways are being modeled. In this article we report the release of an efficient model checker MIRACH 1.0, which supports any model written in popular formats such as CSML and SBML. MIRACH is integrated with a Petri-net-based simulation engine, enabling efficient online (on-the-fly) checking. In our experiment, by using Levchenko et al. model, we reveal that timesaving gains by using MIRACH easily surpass 400% compared with its offline-based counterpart. MIRACH 1.0 was developed using Java and thus executable on any platform installed with JDK 6.0 (not JRE 6.0) or later. MIRACH 1.0, along with its source codes, documentation and examples are available at http://sourceforge.net/projects/mirach/ under the LGPLv3 license.

d. Parameter estimation of biological pathways using data assimilation and model checking

Li C, Kuroyanagi K, Nagasaki M, Miyano S

In this study we developed a novel method for estimating kinetic parameter of biological pathways by using observed time-series data and other knowledge that cannot be formulated in the form of time-series data. Our method utilizes data assimilation (DA) framework and model checking (MC) technique, with a quantitative modeling and simulation architecture named hybrid functional Petri net with extension (HFPNe). Proposed method is applied to an HFPNe model underlying circadian rhythm in mouse. We first translate 23 rules of biological knowledge with temporal logic for the model checking, which are not described in the time-series data. Next, we employ particle filter often applied to DA for our estimation procedure. Each particle checks whether its simulation result satisfies the rules or not, and the result of the checking is used for its resampling step. Our simulation results show that proposed method is faster and more accurate than previous method.

e. CSO validator: improving manual curation workflow for biological pathways

Jeong E, Nagasaki M, Ikeda E, Saito A, Miyano S

Manual curation and validation of large-scale biological pathways are required to obtain high-quality pathway databases. In a typical curation process, model validation and model update based on appropriate feedback are repeated and requires considerable cooperation of scientists. We have developed a CSO (Cell System Ontology) validator to reduce the repetition and time during the curation process. This tool assists in quickly obtaining agreement among curators and domain exper ts and in providing a consistent and accurate pathway database. The tool is available at http://csovalidator.csml.org.

f. Ontology-based instance data validation for high-quality curated biological pathways

Jeong E, Nagasaki M, Ueno K, Miyano S

Modeling in systems biology is vital for understanding the complexity of biological systems across scales and predicting system-level behaviors. To obtain high-quality pathway databases, it is essential to improve the efficiency of model validation and model update based on appropriate feedback. We have developed a new method to guide creating novel high-quality biological pathways, using a rule-based validation. Rules are defined to correct models against biological semantics and improve models for dynamic simulation. In this work, we have defined 40 rules which constrain event-specific participants and the related features and adding missing processes based on biological events. This approach is applied to data in Cell System Ontology which is a comprehensive ontology that represents complex biological pathways with dynamics and visualization. The experimental results show that the relatively simple rules can efficiently detect errors made during curation, such as misassignment and misuse of ontology concepts and terms in curated models. A new rule-based approach has been developed to facilitate model validation and model complementation. Our rule-based validation embedding biological semantics enables us to provide high-quality curated biological pathways. This approach can serve as a preprocessing step for model integration, exchange and extraction data, and simulation.

g. High performance hybrid functional Petri net simulations of biological pathway models on CUDA

Chalkidis G, Nagasaki M, Miyano S

Hybrid functional Petri nets are a wide-spread tool for representing and simulating biological models. Due to their potential of providing virtual drug testing environments, biological simulations have a growing impact on pharmaceutical research. Continuous research advancements in biology and medicine lead to exponentially increasing simulation times, thus raising the demand for performance accelerations by efficient and inexpensive parallel computation solutions. Recent developments in the field of general-purpose computation on graphics processing units (GPGPU) enabled the scientific community to port a variety of compute intensive algorithms onto the graphics processing unit (GPU). This work presents the first scheme for mapping biological hybrid functional Petri net models, which can handle both discrete and continuous entities, onto compute unified device architecture (CUDA) enabled GPUs. GPU accelerated simulations are observed to run up to 18 times faster than sequential implementations. Simulating the cell boundary formation by Delta-Notch signaling on a CUDA enabled GPU results in a speedup of approximately 7 times for a model containing 1,600 cells.

h. Online model checking approach based parameter estimation to a neuronal fate decision simulation model in *Caenorhabditis elegans* with hybrid functional Petri net with extension

Li C, Nagasaki M, Miyano S

Mathematical modeling and simulation studies are playing an increasingly important role in helping researchers elucidate how living organisms function in cells. In systems biology, researchers typically tune many parameters manually to achieve simulation results that are consistent with biological knowledge. This severely limits the size and complexity of simulation models built. In order to break this limitation, we propose a computational framework to automatically estimate kinetic parameters for a given network structure. We utilized an online (on-the-fly) model checking technique (which saves resources compared to the offline approach), with a quantitative modeling and simulation architecture named hybrid functional Petri net with extension (HFPNe). We demonstrate the applicability of this framework by the analysis of the underlying model for the neuronal cell fate decision model (ASE fate model) in *Caenorhabditis elegans*. First, we built a quantitative ASE fate model containing 3327 components

emulating nine genetic conditions. Then, using our developed efficient online model checker, MIRACH 1.0, together with parameter estimation, we ran 20-million simulation runs, and were able to locate 57 parameter sets for 23 parameters in the model that are consistent with 45 biological rules extracted from published biological articles without much manual intervention. To evaluate the robustness of these 57 parameter sets, we run another 20 million simulation runs using different magnitudes of noise. Our simulation results concluded that among these models, one model is the most reasonable and robust simulation model owing to the high stability against these stochastic noises. Our simulation results provide interesting biological findings which could be used for future wet-lab experiments.

3. Statistical Data Analysis Methods for Gene Expression Data, and Next-Generation Sequence Data, and Clinical Data

a. Estimating exogenous variables in data with more variables than observations

Sogawa Y[26], Shimizu S[26], Shimamura T, Hyvarinen A[27], Washio T[26], Imoto S: [26]The Institute of Scientific and Industrial Research, Osaka University [27]Department of Mathematics and Statistics, University of Helsinki

Many statistical methods have been proposed to estimate causal models in classical situations with fewer variables than observations. However, modern datasets including gene expression data increase the needs of high-dimensional causal modeling in challenging situations with orders of magnitude more variables than observations. In this paper, we propose a method to find exogenous variables in a linear non-Gaussian causal model, which requires much smaller sample sizes than conventional methods and works even under orders of magnitude more variables than observations. Exogenous variables work as triggers that activate causal chains in the model, and their identification leads to more efficient experimental designs and better understanding of the causal mechanism. We present experiments with artificial data and real-world gene expression data to evaluate the method.

b. A rank-based statistical test for measuring synergistic effects between two gene sets

Shiraishi U, Okada-Hatakeyama M[28], Miyano S: [28]RIKEN Research Center for Allergy and Immunology

Due to recent advances in high-throughput technologies, data on various types of genomic annotation have accumulated. These data will be crucially helpful for elucidating the combinatorial logic of transcription. Although several approaches have been proposed for inferring cooperativity among multiple factors, most approaches are haunted by the issues of normalization and threshold values. In this study we proposed a rank-based nonparametric statistical test for measuring the effects between two gene sets. This method is free from the issues of normalization and threshold value determination for gene expression values. Furthermore, we have proposed an efficient Markov chain Monte Carlo method for calculating an approximate significance value of synergy. We have applied this approach for detecting synergistic combinations of transcription factor binding motifs and histone modifications. C implementation of the method is available from http://www.hgc.jp/~yshira/software/rankSynergy.zip.

c. Strategy of finding optimal number of features on gene expression data

Sharma A, Koh CH, Imoto S, Miyano S

Feature selection is considered to be an important step in the analysis of transcriptomes or gene expression data. Carrying out feature selection reduces the curse of the dimensionality problem and improves the interpretability of the problem. Numerous feature selection methods have been proposed in the literature and these methods rank the genes in order of their relative importance. However, most of these methods determine the number of genes to be used in an arbitraryly or heuristic fashion. Proposed is a theoretical way to determine the optimal number of genes to be selected for a given task. This proposed strategy has been applied on a number of gene expression datasets and promising results have been obtained.

d. A top-*r* feature selection algorithm for microarray gene expression data

Sharma A, Imoto S, Miyano S

Most of the conventional feature selection algorithms have a drawback whereby a weakly ranked gene that could perform well in terms of classification accuracy with an appropriate subset of genes will be left out of the selection. Considering this shortcoming, we propose a feature selection algorithm in gene expression data analysis of sample classifications. The proposed algorithm first divides genes into subsets, the

sizes of which are relatively small (roughly of size h), then selects informative smaller subsets of genes (of size $r < h$) from a subset and merges the chosen genes with another gene subset (of size r) to update the gene subset. We repeat this process until all subsets are merged into one informative subset. We illustrate the effectiveness of the proposed algorithm by analyzing three distinct gene expression datasets. Our method shows promising classification accuracy for all the test datasets. We also show the relevance of the selected genes in terms of their biological functions.

e. ClipCrop: a tool for detecting structural variations with single-base resolution using soft-clipping information

Suzuki S, Yasuda T, Shiraishi Y, Miyano S, Nagasaki M

Structural variations (SVs) change the structure of the genome and are therefore the causes of various diseases. Next-generation sequencing allows us to obtain a multitude of sequence data, some of which can be used to infer the position of SVs. We developed a new method and implementation named ClipCrop for detecting SVs with single-base resolution using soft-clipping information. A soft-clipped sequence is an unmatched fragment in a partially mapped read. To assess the performance of ClipCrop with other SV-detecting tools, we generated various patterns of simulation data-SV lengths, read lengths, and the depth of coverage of short reads-with insertions, deletions, tandem duplications, inversions and single nucleotide alterations in a human chromosome. For comparison, we selected BreakDancer, CNVnator and Pindel, each of which adopts a different approach to detect SVs, e.g. discordant pair approach, depth of coverage approach and split read approach, respectively. Our method outperformed BreakDancer and CNVnator in both discovering rate and call accuracy in any type of SV. Pindel offered a similar performance as our method, but our method crucially outperformed for detecting small duplications. From our experiments, ClipCrop infer reliable SVs for the data set with more than 50 bases read lengths and 20x depth of coverage, both of which are reasonable values in current NGS data set. ClipCrop can detect SVs with higher discovering rate and call accuracy than any other tool in our simulation data set.

f. Symbolic hierarchical clustering for visual analogue scale data

Katayama K, Yamaguchi R, Imoto S, Tokunaga H[29], Imazu Y[29], Matuura K[29], Watanabe K[29], Miyano S: [29]Center for Kampo Medicine, Keio University School of Medicine

We proposed a hierarchical clustering in the framework of Symbolic Data Analysis (SDA). SDA was proposed by Diday at the end of the 1980s and is a new approach for analysing huge and complex data. In SDA, an observation is described by not only numerical values but also "higher-level units"; sets, intervals, distributions, etc. Most SDA works have dealt with only intervals as the descriptions. In this study, we defined "*pain distribution*" as new type data in SDA and proposed a hierarchical clustering for this new type data.

g. Clustering for visual analogue scale data in symbolic data analysis

Katayama K, Yamaguchi R, Imoto S, Matsuura K[29], Watanabe K[29], Miyano S

We proposed a hierarchical clustering for the visual analogue scale (VAS) in the framework of Symbolic Data Analysis (SDA). The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. VAS is of most value when looking at change within people, and is of less value for comparing across a group of people because they have different sense. It could be argued that a VAS is trying to produce interval/ratio data out of subjective values that are at best ordinal. Thus, some caution is required in handling VAS. We described VAS as distribution and handle it as new type data in SDA. In this study, we defined "VAS distribution" as new type data in SDA and proposed a hierarchical clustering for this new type data.

4. Algorithms and Data Structures for Bioinformatics and Cheminformatics

a. A subpath kernel for rooted unordered trees

Kimura D[30], Kuboyama T, Shibuya T, Kashima H[30]: [30]Graduate School of Information Science and Technology, University of Tokyo [31]Computer Centre, Gakushuin University

Kernel method is one of the promising approaches to learning with tree-structured data, and various efficient tree kernels have been proposed to capture informative structures in trees. In this paper, we propose a new tree kernel

function based on "subpath sets" to capture vertical structures in rooted unordered trees, since such tree-structures are often used to code hierarchical information in data. We also propose a simple and efficient algorithm for computing the kernel by extending the multikey quicksort algorithm used for sorting strings. The time complexity of the algorithm is $O((|T_1|+|T_2|)\log(|T_1|+|T_2|))$ time on average, and the space complexity is $O(|T_1|+|T_2|)$, where $|T_1|$ and $|T_2|$ are the numbers of nodes in two trees T_1 and T_2. We apply the proposed kernel to two supervised classification tasks, XML classification in web mining and glycan classification in bioinformatics. The experimental results show that the predictive performance of the proposed kernel is competitive with that of the existing efficient tree kernel for unordered trees proposed by "Vishwanathan *et al*. Fast kernels for string and tree matching. In: Advances in Neural Information Processing Systems, vol. 15, pp. 569-576 (2003)", and is also empirically faster than the existing kernel.

b. An index structure for spaced seed search

Onodera T, Shibuya T

In this study, we introduced an index structure of texts which supports fast search of patterns with "don't care"s in predetermined positions. This data structure is a generalization of the suffix array and has many applications especially for computational biology. We proposed three algorithms to construct the index. Two of them are based on a variant of radix sort but each utilizes different types of referential information to sort suffixes by multiple characters at a time. The other is for the case when "don't care"s appear periodically in patterns and can be combined with the others.

c. 2D-Qsar for 450 types of amino acid induction peptides with a novel substructure pair descriptor having wider scope

Osoda T, Miyano S

Quantitative structure-activity relationships (QSAR) analysis of peptides is helpful for designing various types of drugs such as kinase inhibitor or antigen. Capturing various properties of peptides is essential for analyzing two-dimensional QSAR. A descriptor of peptides is an important element for capturing properties. The atom pair holographic (APH) code is designed for the description of peptides and it represents peptides as the combination of thirty-six types of key atoms and their intermediate binding between two key atoms. The substructure pair descriptor (SPAD) represents peptides as the combination of forty-nine types of key substructures and the sequence of amino acid residues between two substructures. The size of the key substructures is larger and the length of the sequence is longer than traditional descriptors. Similarity searches on C5a inhibitor data set and kinase inhibitor data set showed that order of inhibitors become three times higher by representing peptides with SPAD, respectively. Comparing scope of each descriptor shows that SPAD captures different properties from APH. QSAR/QSPR for peptides is helpful for designing various types of drugs such as kinase inhibitor and antigen. SPAD is a novel and powerful descriptor for various types of peptides. Accuracy of QSAR/QSPR becomes higher by describing peptides with SPAD.

d. Noise-tolerant active learning algorithm

Osoda T, Miyano S

Selection of unlabeled instances seriously affects the efficiency of active learning. Noise in data degrades efficiency because noise causes a predictive model to be inaccurate. To analyze this effect, we introduce a noise parameter to the expected log likelihood and formulate the objective function. Then, we define the density around a labeled instance to approximate the expected log likelihood with noise more precisely. Empirical experiments performed on the UCT data set show that the accuracy of noise-tolerant active learning is always the highest among several approaches under various noise conditions. Therefore, the noise-tolerant active learning algorithm is stably effective regardless of whether the data set includes considerable noise or not.

5. Pandemic Control Simulation

a. Estimation of macroscopic parameter in agent-based pandemic simulation

Saito MM[5], Imoto S, Yamaguchi R, Miyano S, Higuchi T[5]

Simulations of epidemic spread is an appealing approach to design intervention programs against influenza epidemic. Agent-based approach is particularly useful, since it enable us to administrate and evaluate effectiveness of various intervention measures in the simulated city. On the other hand, observation data are obtained as macroscopic information. Hence, we need to extract some macroscopic information,

to examine the validity of the simulation result. The reproduction number, which indicates how many new infected persons are produced due to one infected person, is often used in epidemiology. This quantity is directly related to coefficients in differential equations in macro simulations, whereas it is non-trivial in agent-based simulations. In this paper, we demonstrate the estimation of the reproduction number from an agent-based simulation result, by assimilating the number of infected persons yielded by an agent-based simulator to the SEIR model, which is a representative macro simulation model.

b. Parallel agent-based simulator for influenza pandemic

Saito MM[5], Imoto S, Yamaguchi R, Miyano S, Higuchi T[5]

We developed a parallel agent-based influenza pandemic simulator, in order to study the influenza spread in a city. In the simulator, the city consists of several towns connected tightly by trains. Residents of the towns walk around places such as corporations and schools using trains by need, following to own schedulers. The influenza spread in these congested places is simulated as stochastic processes. We have demonstrated simulations with a realistic scale of population (an order of million) and showed that one simulation run is completed around one hour.

Publications

1. Chalkidis G, Nagasaki M, Miyano S. High performance hybrid functional Petri net simulations of biological pathway models on CUDA. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 8(6): 1545-1556, 2011.
2. Chen X-W, Miyano S (Eds.) Special Issue on The First IEEE Conference on Healthcare Informatics, Imaging, and Systems biology HISB'11. J. Bioinformatics and Computational Biology 9(5), 2011.
3. Fujita A, Sato JR, Demas MAA, Yamaguchi R, Shimamura T, Ferreira CE, Sogayar MC, Miyano S. Inferring contagion in regulatory networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 8(2): 570-576, 2011.
4. Furu M, Kajita Y, Nagayama S, Ishibe T, Shima Y, Nishijo K, Uejima D, Takahashi R, Aoyama T, Nakayama T, Nakamura T, Nakashima Y, Ikegawa M, Imoto S, Katagiri T, Nakamura Y, Toguchida J. Toguchida. Identification of AFAP1L1 as a prognostic marker for spindle cell sarcomas. Oncogene. 30(38): 4015-4025, 2011.
5. Hasegawa T, Yamaguchi R, Nagasaki M, Imoto S, Miyano S. Comprehensive pharmacogenomic pathway screening by data assimilation. Lecture Notes in Bioinformatics. 6674: 160-171, 2011.
6. Hurley D, Araki H, Tamada Y, Dunmore B, Sanders D, Humphreys S, Affara M, Imoto S, Yasuda K, Tomiyasu Y, Tashiro K, Savoie C, Cho V, Smith S, Kuhara S, Miyano S, Charnock-Jones DS, Crampin EJ, Print CG. Gene network inference and visualization tools for biologists: application to new human transcriptome datasets. Nucleic Acids Res. 2011 Dec 6. [Epub ahead of print]
7. Imoto S, Kojima K, Perrier E, Tamada Y, Miyano S. Searching optimal Bayesian network structure on constraint search space: superstructure approach. Lecture Notes in Computer Science. 6797: 210-218, 2011.
8. Imoto S, Tamada Y, Araki H, Miyano S. Computational Drug Target Pathway Discovery: A Bayesian Network Approach. Handbook of Computational Statistics: Statistical Bioinformatics, Springer. p. 501-532, 2011.
9. Jeong E, Nagasaki M, Ueno K, Miyano S. Ontology-based instance data validation for high-quality curated biological pathways. BMC Bioinformatics. 12(Suppl 1): S8, 2011.
10. Jeong E, Nagasaki M, Ikeda E, Saito A, Miyano S. CSO validator: improving manual curation workflow for biological pathways. Bioinformatics. 27(17): 2471-2472, 2011.
11. Katayama K, Yamaguchi R, Imoto S, Tokunaga H, Imazu Y, Matuura K, Watanabe K, Miyano S. Symbolic hierarchical clustering for visual analogue scale data. Smart Innovation, Systems and Technologies. 10: 799-805, 2011.
12. Katayama K, Yamaguchi R, Imoto S, Matsuura K, Watanabe K, Miyano S. Clustering for Visual analogue scale data in symbolic data analysis. Procedia Computer Science. 6: 370-374, 2011.
13. Kimura D, Kuboyama T, Shibuya T, Kashima H. A subpath kernel for rooted unordered trees. Proc. 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2011). 62-74, 2011.
14. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, Tanaka F, Shibata K, Suzuki A, Komune S, Miyano S, Mori M. Long non-coding RNA HOTAIR regulates

Polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 2011 Aug 23. [Epub ahead of print]
15. Koh CH, Nagasaki M, Saito A, Li C, Wong L, Miyano S. MIRACH: Efficient model checker for quantitative biological pathway models. Bioinformatics. 27(5): 734-735, 2011.
16. Li C, Kuroyanagi K, Nagasaki M, Miyano S. Parameter estimation of biological pathways using data assimilation and model checking. Proceedings of 2nd International Workshop on Biological Processes & Petri Nets (BioPPN2011) (http://ceur-ws.org/Vol-724). 53-70, 2011.
17. Li C, Nagasaki M, Miyano S. Online model checking approach based parameter estimation to a neuronal fate decision simulation model in *Caenorhabditis elegans* with hybrid functional Petri net with extension. Molecular BioSystems. 7(5): 1576-1592, 2011.
18. Mandoiu II, Miyano S, Przytycka TM, Rajasekaran S (Eds.) Proceedings of The IEEE First International Conference on Computational Advances in Bio and Medical Sciences. IEEE Computer Society Press, 2011.
19. Matsuno H, Nagasaki M, Miyano S. Hybrid Petri net based modeling for biological pathway simulation. Natural Computing 10(3): 1099-1120, 2011.
20. Nagasaki M, Saito A, Fujita A, Tremmel G, Ueno K, Ikeda E, Jeong E, Miyano S. Systems biology model repository for macrophage pathway simulation. Bioinformatics. 27 (11): 1591-1593, 2011.
21. Onodera T, Shibuya T. An index structure for spaced seed search. Lecture Notes in Computer Science. 7074: 764-772, 2011. (Proc. 22nd annual International Symposium on Algorithms and Computation (ISAAC 2011))
22. Osoda T, Miyano S. 2D-Qsar for 450 types of amino acid induction peptides with a novel substructure pair descriptor having wider scope. J Cheminform. 3(1): 50, 2011.
23. Osoda T, Miyano S. Noise-tolerant active learning algorithm. Proc. the 2011 International Conference on Data Mining. 10-14, 2011.
24. Saito MM, Imoto S, Yamaguchi R, Miyano S, Higuchi T. Parallel agent-based simulator for influenza pandemic. Lecture Notes in Computer Science. 7068: 361-370, 2011.
25. Saito MM, Imoto S, Yamaguchi R, Miyano S, Higuchi T. Estimation of macroscopic parameter in agent-based pandemic simulation. Proc. 13th International Conference on Information Fusion. 1-6, 2011.
26. Sharma A, Imoto S, Miyano S.A top-r feature selection algorithm for microarray gene expression data. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2011 Nov 11. [Epub ahead of print]
27. Sharma A, Koh CH, Imoto S, Miyano S. Strategy of finding optimal number of features on gene expression data. Electronic Letters. 47(8): 480-482, 2011.
28. Shimamura T, Imoto S, Shimada Y, Hosono Y, Niida A, Nagasaki M, Yamaguchi R, Takahashi T, Miyano S.A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition. PLoS ONE. 6(6): e20804, 2011.
29. Shiraishi U, Okada-Hatakeyama M, Miyano S.A rank-based statistical test for measuring synergistic effects between two gene sets. Bioinformatics. 27(17): 2399-2405, 2011.
30. Sogawa Y, Shimizu S, Shimamura T, Hyvarinen A, Washio T, Imoto S. Estimating exogenous variables in data with more variables than observations. Neural Networks. 24(8): 875-880, 2011.
31. Suzuki S, Yasuda T, Shiraishi Y, Miyano S, Nagasaki M. ClipCrop: a tool for detecting structural variations with single-base resolution using soft-clipping information. BMC Bioinformatics. 12: S7, 2011.
32. Tamada Y, Imoto S, Miyano S. Parallel algorithm for learning optimal Bayesian network structure. J Machine Learning Research. 12: 2437-2459, 2011.
33. Tamada Y, Imoto S, Araki H, Nagasaki M, Print C, Charnock-Jones DS, Miyano S. Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 8(3): 683-697, 2011.
34. Tamada Y, Shimamura T, Yamaguchi R, Imoto S, Nagasaki M, Miyano S. SiGN: large-scale gene network estimation environment for high performance computing. Genome Informatics. 25: 40-52, 2011.
35. Tamada Y, Yamaguchi R, Imoto S, Hirose O, Yoshida R, Nagasaki M, Miyano S. SiGN-SSM: open source parallel software for estimating gene networks with state space models. Bioinformatics. 27: 1172-1173, 2011.
36. Yamaguchi K, Sakai M, Kim J, Tsunesumi S, Fujii T, Ikenoue T, Yamada Y, Akiyama Y, Muto Y, Yamaguchi R, Miyano S, Nakamura Y, Furukawa Y. MRG-binding protein contributes to development of colorectal cancer. Cancer Sci. 102(8): 1486-1492, 2011.
37. Yamauchi M, Yoshino I, Yamaguchi R, Shimamura T, Nagasaki M, Imoto S, Niida A, Koizumi F, Kohno T, Yokota J, Miyano S, Gotoh N. N-cadherin expression is a poten-

tial survival mechanism of gefitinib-resistant lung cancer cells. Am J Cancer Res. 1(7): 823-833, 2011.
38. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 478 (7367): 64-69, 2011.

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	中　村　祐　輔
Associate Professor	Koichi Matsuda, M.D., Ph.D.	准教授	松　田　浩　一
Assistant Professor	Hitoshi Zembutsu, M.D., Ph.D.	助　教	前　佛　　均
Assistant Professor	Ryuji Hamamoto, Ph.D.	助　教	浜　本　隆　二

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

Yusuke Nakamura, Koichi Matsuda, Hitoshi Zembutsu, Ryuji Hamamoto, Yataro Daigo, Hidewaki Nakagawa, Chizu Tanikawa, Cui Ri, Hamdi Mbarek, Vinod Kumar, Yuji Urabe, Jiaying Lin, Zhenzhong Deng, Paulisally Hau Yi Lo, Martha Espinosa, Zhenzhong Deng, Satoko Uno, Yoichiro Kato, Mitsuko Nakashima, Motoko Unoki, Masanori Yoshimatsu, Shinya Hayami, Hyun-Soo Cho, Goji Toyokawa, Tadashi Takawa, Reem Abdelrahim Ibrahim, Kang Daechun, Lianhua Piao, Su-Youn Chung, Osman W Mohammed, Takashi Fujitomo, Seham Elgazzar, Low Siew Kee, Cha Pei Chieng, Koji Ueda, Nguyen Minh-Hue, Junkichi Koinuma, Daiki Miki, Ken Masuda, Masato Aragaki, Hideto Oshita

(1) Epigenetics

Regulation of histone modification and chromatin structure by the p53-PADI4 pathway

Histone proteins are modified in response to various external signals, however their mechanisms are still not fully understood. Citrullination is a post-transcriptional modification which converts arginine in protein into citrulline. Here we show *in vivo* and *in vitro* citrullination of arginine 3 residue of histone H4 (cit-H4R3) in response to DNA damage through the p53-PADI4 pathway. We also observed DNA damage-induced citrullination of Lamin C. Cit-H4R3 and citrullinated Lamin C are located around fragmented nuclei in apoptotic cells. Ectopic expression of PADI4 led to chromatin decondensation and promoted DNA cleavage, while *Padi4*$^{-/-}$ mice exhibited resistance to radiation-induced apoptosis in the thymus. Furthermore, the level of cit-H4R3 was negatively correlated with p53 protein expression and with tumor size in non-

small cell lung cancer tissues. Our findings reveal that cit-H4R3 would be an "apoptotic histone code" to detect damaged cells and induce nuclear fragmentation, which plays a crucial role in carcinogenesis.

Histone lysine methyltransferase Wolf-Hirschhorn syndrome candidate 1 is involved in human carcinogenesis through regulation of the Wnt pathway

A number of histone methyltransferases have been identified and biochemically characterized, but the pathologic roles of their dysfunction in human diseases like cancer are not well understood. Here, we demonstrate that Wolf-Hirschhorn syndrome candidate 1 (WHSC1) plays important roles in human carcinogenesis. Transcriptional levels of this gene are significantly elevated in various types of cancer including bladder and lung cancers. Immunohistochemical analysis using a number of clinical tissues confirmed significant up-regulation of WHSC1 expression in bladder and lung cancer cells at the protein level. Treatment of cancer cell lines with small interfering RNA targeting WHSC1 significantly knocked down its expression and resulted in the suppression of proliferation. Cell cycle analysis by flow cytometry indicated that knockdown of WHSC1 decreased the cell population of cancer cells at the S phase while increasing that at the G2/M phase. WHSC 1 interacts with some proteins related to the WNT pathway including β-catenin and transcriptionally regulates CCND1, the target gene of the β-catenin/Tcf-4 complex, through histone H3 at lysine 36 trimethylation. This is a novel mechanism for WNT pathway dysregulation in human carcinogenesis, mediated by the epigenetic regulation of histone H3. Because expression levels of WHSC1 are significantly low in most normal tissue types, it should be feasible to develop specific and selective inhibitors targeting the enzyme as antitumor agents that have a minimal risk of adverse reaction.

The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene

A number of histone demethylases have been identified and biochemically characterized, yet their biological functions largely remain uncharacterized, particularly in the context of human diseases such as cancer. In this study, we describe important roles for the histone demethylase KDM3A, also known as JMJD1A, in human carcinogensis. Expression levels of KDM3A were significantly elevated in human bladder carcinomas compared with nonneoplastic bladder tissues ($p < 0.0001$), when assessed by real-time PCR. We confirmed that some other cancers including lung cancer also overexpressed KDM3A, using cDNA microarray analysis. Treatment of cancer cell lines with small interfering RNA targeting KDM3A significantly knocked down its expression and resulted in the suppression of proliferation. Importantly, we found that KDM3 A activates transcription of the HOXA1 gene through demethylating histone H3 at lysine 9 di -methylation by binding to its promoter region. Indeed, expression levels of KDM3A and HOXA 1 in several types of cancer cell lines and bladder cancer samples were statistically correlated. We observed the down-regulation of HOXA1 as well as CCND1 after treatment with KDM3A siRNA, indicating G1 arrest of cancer cells. Together, our results suggest that elevated expression of KDM3A plays a critical role in the growth of cancer cells, and further studies may reveal a cancer therapeutic potential in KDM3A inhibition.

The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase 6

Histone methyltransferases and demethylases are known to regulate transcription by altering the epigenetic marks on histones, but the pathologic roles of their dysfunction in human diseases, such as cancer, still remain to be elucidated. Herein, we show that the histone demethylase JMJD2B is involved in human carcinogenesis. Quantitative real-time PCR showed notably elevated levels of JMJD2B expression in bladder cancers, compared with corresponding nonneoplastic tissues ($P < 0.0001$), and elevated protein expression was confirmed by immunohistochemistry. In addition, cDNA microarray analysis revealed transactivation of JMJD2B in lung cancer, and immunohistochemical analysis showed protein overexpression in lung cancer. siRNA-mediated reduction of expression of JMJD2B in bladder and lung cancer cell lines significantly suppressed the proliferation of cancer cells, and suppressing JMJD2B expression lead to a decreased population of cancer cells in S phase, with a concomitant increase of cells in G(1) phase. Furthermore, a clonogenicity assay showed that the demethylase activity of JMJD2B possesses an oncogenic activity. Microarray analysis after knockdown of JMJD2B revealed that JMJD2B could regulate multiple pathways which contribute to carcinogenesis, including the cell-cycle pathway. Of the downstream

genes, chromatin immunoprecipitation showed that CDK6 (cyclin-dependent kinase 6), essential in G1-S transition, was directly regulated by JMJD2B, via demethylation of histone H3-K9 in its promoter region. Expression levels of JMJD2B and CDK6 were significantly correlated in various types of cell lines. Deregulation of histone demethylation resulting in perturbation of the cell cycle, represents a novel mechanism for human carcinogenesis and JMJD2B is a feasible molecular target for anticancer therapy.

Enhanced expression of EHMT2 is involved in the proliferation of cancer cells through negative regulation of SIAH1

EHMT2 is a histone lysine methyltransferase localized in euchromatin regions and acting as a corepressor for specific transcription factors. Although the role of EHMT2 in transcriptional regulation has been well documented, the pathologic consequences of its dysfunction in human disease have not been well understood. Here, we describe important roles of EHMT2 in human carcinogenesis. Expression levels of EHMT2 are significantly elevated in human bladder carcinomas compared with nonneoplastic bladder tissues ($P < .0001$) in real-time polymerase chain reaction analysis. Complementary DNA microarray analysis also revealed its overexpression in various types of cancer. The reduction of EHMT2 expression by small interfering RNAs resulted in the suppression of the growth of cancer cells and possibly caused apoptotic cell death in cancer cells. Importantly, we show that EHMT2 can suppress transcription of the SIAH1 gene by binding to its promoter region (-293 to +51) and by methylating lysine 9 of histone H3. Furthermore, an EHMT2-specific inhibitor, BIX-01294, significantly suppressed the growth of cancer cells. Our results suggest that dysregulation of EHMT2 plays an important role in the growth regulation of cancer cells, and further functional studies may affirm the importance of EHMT2 as a promising therapeutic target for various types of cancer.

Minichromosome Maintenance Protein 7 is a potential therapeutic target in human cancer and a novel prognostic marker of non-small cell lung cancer

The research emphasis in anti-cancer drug discovery has always been to search for a drug with the greatest antitumor potential but fewest side effects. This can only be achieved if the drug used is against a specific target located in the tumor cells. In this study, we evaluated Minichromosome Maintenance Protein 7 (MCM 7) as a novel therapeutic target in cancer. Immunohistochemical analysis showed that MCM7 was positively stained in 196 of 331 non-small cell lung cancer (NSCLC), 21 of 29 bladder tumor and 25 of 70 liver tumor cases whereas no significant staining was observed in various normal tissues. We also found an elevated expression of MCM7 to be associated with poor prognosis for patients with NSCLC ($P = 0.0055$). qRT-PCR revealed a higher expression of MCM7 in clinical bladder cancer tissues than in corresponding non-neoplastic tissues ($P < 0.0001$), and we confirmed that a wide range of cancers also overexpressed MCM7 by cDNA microarray analysis. Suppression of MCM7 using specific siRNAs inhibited incorporation of BrdU in lung and bladder cancer cells overexpressing MCM7, and suppressed the growth of those cells more efficiently than that of normal cell strains expressing lower levels of MCM7. Since MCM7 expression was generally low in a number of normal tissues we examined, MCM7 has the characteristics of an ideal candidate for molecular targeted cancer therapy in various tumors and also as a good prognostic biomarker for NSCLC patients.

Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker

The emphasis in anticancer drug discovery has always been on finding a drug with great antitumor potential but few side-effects. This can be achieved if the drug is specific for a molecular site found only in tumor cells. Here, we find the enhancer of zeste homolog 2 (EZH2) to be highly overexpressed in lung and other cancers, and show that EZH2 is integral to proliferation in cancer cells. Quantitative real-time PCR analysis revealed higher expression of EZH 2 in clinical bladder cancer tissues than in corresponding non-neoplastic tissues ($P < 0.0001$), and we confirmed that a wide range of cancers also overexpress EZH2, using cDNA microarray analysis. Immunohistochemical analysis showed positive staining for EZH2 in 14 of 29 cases of bladder cancer, 135 of 292 cases of non-small-cell lung cancer (NSCLC), and 214 of 245 cases of colorectal cancer, whereas no significant staining was observed in various normal tissues. We found elevated expression of EZH2 to be associated with poor prognosis for patients with NSCLC ($P = 0.0239$). In lung and bladder cancer cells overexpressing EZH2, suppression of EZH2 using specific siRNAs inhibited incorporation of BrdU and resulted in significant suppression of cell growth, even though no significant effect

was observed in the normal cell strain CCD-18 Co, which has undetectable EZH2. Because EZH 2 expression was scarcely detectable in all normal tissues we examined, EZH2 shows promise as a tumor-specific therapeutic target. Furthermore, as elevated levels of EZH2 are associated with poor prognosis of patients with NSCLC, its overexpression in resected specimens could prove a useful molecular marker, indicating the necessity for a more extensive follow-up in some lung cancer patients after surgical treatment.

Demethylation of RB regulator MYPT1 by histone demethylase LSD1 promotes cell cycle progression in cancer cells

Histone demethylase LSD1 (also known as KDM1 and AOF2) is active in various cancer cells, but its biological significance in human carcinogenesis is unexplored. In this study, we explored hypothesized interactions between LSD 1 and MYPT1, a known regulator of RB1 phosphorylation. We found that MYPT1 was methylated in vitro and in vivo by histone lysine methyltransferase SETD7 and demethylated by LSD1, identifying Lys 442 of MYPT1 as a target for methylation/demethylation by these enzymes. LSD1 silencing increased MYPT1 protein levels, decreasing the steady state level of phosphorylated RB1 (Ser 807/811) and reducing E2F activity. MYPT1 methylation status influenced the affinity of MYPT1 for the ubiquitin-proteasome pathway of protein turnover. MYPT1 was unstable in murine cells deficient in SETD7, supporting the concept that MYPT1 protein stability is physiologically regulated by methylation status. LSD1 overexpression could activate RB1 phosphorylation by inducing a destabilization of MYPT1 protein. Taken together, our results comprise a novel cell cycle regulatory mechanism mediated by methylation/demethylation dynamics, and they reveal the significance of LSD1 overexpression in human carcinogenesis.

Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers

Protein arginine methylation is a novel post-translational modification regulating a diversity of cellular processes, including histone functions, but the roles of protein arginine methyltransferases (PRMTs) in human cancer are not well investigated. To address this issue, we first examined expression levels of genes belonging to the PRMT family and found significantly higher expression of PRMT1 and PRMT6, both of which are Type I PRMTs, in cancer cells of various tissues than in non-neoplastic cells. Abrogation of the expression of these genes with specific siRNAs significantly suppressed growth of bladder and lung cancer cells. Expression profile analysis using the cells transfected with the siRNAs indicated that PRMT1 and PRMT6 interplay in multiple pathways, supporting regulatory roles in the cell cycle, RNA processing and also DNA replication that are fundamentally important for cancer cell proliferation. Furthermore, we demonstrated that serum asymmetric dimethylarginine (ADMA) levels of a number of cancer cases are significantly higher than those of nontumor control cases. In summary, our results suggest that dysregulation of PRMT1 and PRMT6 can be involved in human carcinogenesis and that these Type I arginine methyltransferases are good therapeutic targets for various types of cancer.

Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers

A number of histone demethylases have been identified and biochemically characterized, but the pathological roles of their dysfunction in human disease like cancer have not been well understood. Here, we demonstrate important roles of lysine-specific demethylase 1 (LSD1) in human carcinogenesis. Expression levels of LSD1 are significantly elevated in human bladder carcinomas compared with nonneoplastic bladder tissues ($p < 0.0001$). cDNA microarray analysis also revealed its transactivation in lung and colorectal carcinomas. LSD1-specific small interfering RNAs significantly knocked down its expression and resulted in suppression of proliferation of various bladder and lung cancer cell lines. Concordantly, introduction of exogenous LSD1 expression promoted cell cycle progression of human embryonic kidney fibroblast cells. Expression profile analysis showed that LSD1 could affect the expression of genes involved in various chromatin-modifying pathways such as chromatin remodeling at centromere, centromeric heterochromatin formation and chromatin assembly, indicating its essential roles in carcinogenesis through chromatin modification.

(2) Lung cancer

Chondrolectin is a novel diagnostic biomarker and a therapeutic target for lung cancer.

PURPOSE: This study aims to identify molecules that might be useful as diagnostic/prognostic biomarkers and as targets for the devel-

opment of new molecular therapies for lung cancer.

EXPERIMENTAL DESIGN: We screened for genes that were highly transactivated in a large proportion of 120 lung cancers by means of a cDNA microarray representing 27,648 genes and found chondrolectin (CHODL) as a candidate. Tumor tissue microarray was applied to examine the expression of CHODL protein and its clinicopathologic significance in archival non-small cell lung cancer (NSCLC) tissues from 295 patients. A role of CHODL in cancer cell growth and/or survival was examined by siRNA experiments. Cellular invasive effect of CHODL on mammalian cells was examined by Matrigel assays.

RESULTS: Immunohistochemical staining revealed that strong positivity of CHODL protein was associated with shorter survival of patients with NSCLC (P＝0.0006), and multivariate analysis confirmed it to be an independent prognostic factor. Treatment of lung cancer cells with siRNAs against CHODL suppressed growth of the cancer cells. Furthermore, induction of exogenous expression of CHODL conferred growth and invasive activity of mammalian cells.

CONCLUSIONS: CHODL is likely to be a prognostic biomarker in the clinic and targeting CHODL might be a strategy for the development of anticancer drugs.

Identification of Epstein-Barr virus-induced gene 3 as a novel serum and tissue biomarker and a therapeutic target for lung cancer.

PURPOSE: This study aims to identify novel biomarkers and therapeutic targets for lung cancer.

EXPERIMENTAL DESIGN: We carried out gene expression profile analysis of 120 lung cancers to screen for genes encoding transmembrane/secretory molecules that are commonly transactivated in lung cancers. Epstein-Barr virus-induced gene 3 (EBI3), which encodes a secretory glycoprotein, was selected as a good candidate. Immunohistochemical staining using tissue microarray consisting of 414 non-small cell lung cancers was applied to examine the expression level and prognostic value of EBI3. Serum EBI3 levels in 400 individuals for training assays (274 lung cancers and 126 healthy volunteers) and those in 173 individuals for validation analysis (132 lung cancers and 41 healthy volunteers) were measured by ELISA. The role of EBI 3 in cancer cell growth was examined by siRNA and cell growth assays, using cells stably expressing exogenous EBI3.

RESULTS: Immunohistochemical staining of EBI3 using tissue microarrays revealed that a high level of EBI3 expression was associated with a poor prognosis of lung cancer (P＝0.0014) and multivariate analysis confirmed it to be an independent prognostic factor (P＝0.0439). Serum levels of EBI3 in the training set were found to be significantly higher in lung cancer patients than in healthy volunteers; this result was also observed in the validation set. Furthermore, reduction in EBI3 expression by siRNA suppressed cancer cell proliferation whereas induction of exogenous EBI3 conferred growth-promoting activity.

CONCLUSIONS: EBI3 is a potential serum and tissue biomarker as well as therapeutic target for lung cancer.

2. Pharmacogenetics

Genome-wide association study of epirubicin-induced leukopenia in Japanese patients.

OBJECTIVES: Despite long-term clinical experience with epirubicin, unpredictable severe adverse reactions remain an important determinant to limit the drug use. To identify a genetic factor(s) affecting the risk of epirubicin-induced leukopenia/neutropenia, we performed a genome-wide association study.

METHODS: We studied 270 patients consisting of 67 patients with grade 3 or 4 leukopenia/neutropenia, and 203 patients showing no toxicity (patients with grade 1 or 2 were excluded from the study) for genome-wide association study. We further examined the single nucleotide polymorphisms (SNPs) showing P values of less than 0.0001 using an additional set of 48 patients with grade 3/4 leukopenia/neutropenia.

RESULTS: The combined analysis indicated that rs2916733 in microcephalin 1 [combined PFisher min＝2.27×10, odds ratio (OR)＝2.74 with 95% confidence interval (CI)＝1.96-3.83; the nonrisk genotype as reference] was significantly associated with epirubicin-induced leukopenia/neutropenia. A subgroup analysis of patients with only breast cancer showed a similar trend of association for the marker SNP rs 2916733 (combined PFisher min＝6.76×10, OR ＝2.80 with 95% CI＝1.86-4.21). We subsequently performed haplotype analysis and found that a haplotype constructed from rs 2916733 and rs1031309, which was in linkage disequilibrium with rs2916733 (r＝0.64), showed stronger association (P＝2.20×10, OR＝2.88 with 95% CI＝2.05-4.03) than a single landmark SNP (rs2916733; P＝2.27×10, OR＝2.74 with 95 % CI＝1.96-3.83), suggesting that causative vari-

ant(s) that could influence the susceptibility of epirubicin-induced adverse drug reactions (ADRs) might exist in this haplotype.

CONCLUSION: Our findings show that genetic variants in the microcephalin 1 locus are suggestively associated with the risk of epirubicin-induced ADRs and might be applicable in development of diagnostic system for predicting the risk of the ADRs, leading to better prognosis and quality of life for patients with cancer. However, these results should be considered preliminary until replicated in adequately larger powered and controlled samples

Dose-adjustment study of tamoxifen based on CYP2D6 genotypes in Japanese breast cancer patients.

CYP2D6 is a key enzyme responsible for the metabolism of tamoxifen to active metabolites, endoxifen, and 4-hydroxytamoxifen. The breast cancer patients who are heterozygous and homozygous for decreased-function and null alleles of CYP2D6 showed lower plasma concentrations of endoxifen and 4-hydroxytamoxifen compared to patients with homozygous-wild-type allele, resulting in worse clinical outcome in tamoxifen therapy. We recruited 98 Japanese breast cancer patients, who had been taking 20 mg of tamoxifen daily as adjuvant setting. For the patients who have one or no normal allele of CYP2D6, dosages of tamoxifen were increased to 30 and 40 mg/day, respectively. The plasma concentrations of tamoxifen and its metabolites were measured at 8 weeks after dose-adjustment using liquid chromatography-tandem mass spectrometry. Association between tamoxifen dose and the incidence of adverse events during the tamoxifen treatment was investigated. In the patients with CYP2D6*1/*10 and CYP2D6*10/*10, the mean plasma endoxifen levels after dose increase were 1.4- and 1.7-fold higher, respectively, than those before the increase ($P < 0.001$). These plasma concentrations of endoxifen achieved similar level of those in the CYP2D6*1/*1 patients receiving 20 mg/day of tamoxifen. Plasma 4-hydroxytamoxifen concentrations in the patients with CYP2D6*1/*10 and CYP2D6*10/*10 were also significantly increased to the similar levels of the CYP2D6*1/*1 patients according to the increasing tamoxifen dosages ($P < 0.001$). The incidence of adverse events was not significantly different between before and after dose adjustment. This study provides the evidence that dose adjustment is useful for the patients carrying CYP2D6*10 allele to maintain the effective endoxifen level.

Pharmacogenomics of tamoxifen: roles of drug metabolizing enzymes and transporters.

Tamoxifen has been widely used for the prevention of recurrence in patients with hormone receptor-positive breast cancer. Tamoxifen requires metabolic activation by cytochrome P450 (CYP) enzymes for formation of active metabolites, 4-hydroxytamoxifen and endoxifen, which have 30- to 100-fold greater affinity to the estrogen receptor and the potency to suppress estrogen-dependent breast cancer cell proliferation. CYP2D6 is a key enzyme in this metabolic activation and it has been suggested that the genetic polymorphisms of CYP2D6 influence the plasma concentrations of active tamoxifen metabolites and clinical outcomes for breast cancer patients treated with tamoxifen. The genetic polymorphisms in the other drug-metabolizing enzymes, including other CYP isoforms, sulfotransferases and UDP-glucuronosyltransferases might contribute to individual differences in the tamoxifen metabolism and clinical outcome of tamoxifen therapy although their contributions would be small. Recently, involvement of a drug transporter in disposition of active tamoxifen metabolites was identified. The genetic polymorphisms of transporter genes have the potential to improve the prediction of clinical outcome for the treatment of hormone receptor-positive breast cancer. This review summarizes current knowledge on the roles of polymorphisms in the drug-metabolizing enzymes and transporters in tamoxifen pharmacogenomics.

A Genome-Wide Association Study of Overall Survival in Pancreatic Cancer Patients Treated with Gemcitabine in CALGB 80303.

Background and Aims: Cancer and Leukemia Group B 80303 was a randomized, phase III study in patients with advanced pancreatic cancer treated with gemcitabine plus either bevacizumab or placebo. We prospectively collected germline DNA and conducted a genome-wide association study (GWAS) using overall survival (OS) as the endpoint.

EXPERIMENTAL DESIGN: DNA from 351 patients was genotyped for more than 550,000 single-nucleotide polymorphisms (SNP). Associations between OS and SNPs were investigated using the log-linear 2-way multiplicative Cox proportional hazards model. The subset of 294 genetically European patients was used for the primary analysis.

RESULTS: A nonsynonymous SNP in interleukin (IL)17F (rs763780, H161R) and an intronic SNP in strong linkage disequilibrium (rs 7771466) were associated with OS using

genome-wide criteria (P≤10(-7)). Median OS was significantly shorter (P=2.61×10(-8)) for the rs763780 heterozygotes [3.1 months; 95% confidence interval (CI), 2.3-4.3] than for the patients without this variant (6.8 months; 95% CI, 5.8-7.3). After adjustment by stratification factors, the P value for the association was 9.51×10(-7).

CONCLUSIONS: The variant 161R form of IL-17F is a natural antagonist of the antiangiogenic effects of wild-type 161H IL-17F, and angiogenesis may play an important role in the metastatic spread of pancreatic cancer. In this preliminary study, we hypothesize that the angiogenetic potential of pancreatic cancers in patients with variant IL-17F is higher than that of tumors in patients with wild-type IL-17F, conferring worse prognosis. This exploratory GWAS may provide the foundation for testing the biology and clinical effects of novel genes and their heritable variants through mechanistic and confirmatory studies in pancreatic cancer.

A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese.

Although many association studies of polymorphisms in candidate genes with the clinical outcomes of breast cancer patients receiving adjuvant tamoxifen therapy have been reported, genetic factors determining individual response to tamoxifen are not fully understood. To identify genetic polymorphisms associated with clinical outcomes of patients with tamoxifen treatment, we conducted a genome-wide association study (GWAS). We studied 462 Japanese patients with hormone receptor-positive, invasive breast cancer receiving adjuvant tamoxifen therapy. Of them, 240 patients were analyzed by genome-wide genotyping using the Illumina Human610-Quad BeadChips, and two independent sets of 105 and 117 cases were used for replication studies. In the GWAS, we detected significant associations with recurrence-free survival at 15 single-nucleotide polymorphisms (SNPs) on nine chromosomal loci (1p31, 1q41, 5 q33, 7p11, 10q22, 12q13, 13q22, 18q12 and 19p 13) that satisfied a genome-wide significant threshold (log-rank P=2.87×10(-9)-9.41×10(-8)). Among them, rs10509373 in C10orf11 gene on 10q22 was significantly associated with recurrence-free survival in the replication study (log-rank P=2.02×10(-4)) and a combined analysis indicated a strong association of this SNP with recurrence-free survival in breast cancer patients treated with tamoxifen (log-rank P =1.26×10(-10)). Hazard ratio per C allele of rs 10509373 was 4.51 [95% confidence interval (CI), 2.72-7.51; P=6.29×10(-9)]. In a combined analysis of rs10509373 genotype with previously identified genetic makers, CYP2D6 and ABCC2, the number of risk alleles of these three genes had cumulative effects on recurrence-free survival among 345 patients receiving tamoxifen monotherapy (log-rank P=2.28×10(-12)). In conclusion, we identified a novel locus associated with recurrence-free survival in Japanese breast cancer patients receiving adjuvant tamoxifen therapy.

3. Genome-wide association study

(1) cancer susceptibility gene

Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma

To identify the genetic susceptibility factor(s) for hepatitis C virus-induced hepatocellular carcinoma (HCV-induced HCC), we conducted a genome-wide association study using 432,703 autosomal SNPs in 721 individuals with HCV-induced HCC (cases) and 2,890 HCV-negative controls of Japanese origin. Eight SNPs that showed possible association (P<1×10(-5)) in the genome-wide association study were further genotyped in 673 cases and 2,596 controls. We found a previously unidentified locus in the 5′ flanking region of MICA on 6p21.33 (rs2596542, P(combined)=4.21×10(-13), odds ratio=1.39) to be strongly associated with HCV-induced HCC. Subsequent analyses using individuals with chronic hepatitis C (CHC) indicated that this SNP is not associated with CHC susceptibility (P=0.61) but is significantly associated with progression from CHC to HCC (P=3.13×10(-8)). We also found that the risk allele of rs 2596542 was associated with lower soluble MICA protein levels in individuals with HCV-induced HCC (P=1.38×10(-13)).

Common variant in 6q26-q27 is associated with distal colon cancer in Asian population

Colorectal cancer (CRC) is a multifactorial disease with both environmental and genetic factors contributing to its development. The incidence of CRC is increasing year by year in Japan. Patients with CRC in advanced stages have a poor prognosis, but detection of CRC at earlier stages can improve clinical outcome. Therefore, identification of epidemiologic factors that influence development of CRC would facilitate the prevention or early detection of disease.

To identify loci associated with CRC risk, we

performed a genome-wide association study (GWAS) for CRC and sub-analyses by tumor location using 1,583 Japanese CRC cases and 1,898 controls. Subsequently, we conducted replication analyses using a total of 4,809 CRC cases and 2,973 controls including 225 Korean subjects with distal colon cancer and 377 controls.

We identified a novel locus on 6q26-q27 region (rs7758229 in SLC22A3, P=7.92×10-9, Odds ratio of 1.28) that was significantly associated with distal colon cancer. We also replicated the association between CRC and SNPs on 8q24 (rs6983267 and rs7837328, P=1.51×10-8 and 7.44×10-8, Odds ratios of 1.18 and 1.17, respectively). Moreover, we found cumulative effects of three genetic (rs7758229, rs6983267, and rs 4939827 in SMAD7) and one environmental factors (alcohol drinking) which appear to increase CRC risk approximately twofold.

We found a novel susceptible locus in SLC22 A3 that contributes to the risk of distal colon cancer in Asian population. These findings would further extend our understanding of the role of common genetic variants in CRC etiology.

(2) other diseases

A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population

Through a genome-wide association analysis using a total of 7,035 duodenal ulcer cases and 25,323 controls of Japanese populations, we identified two susceptibility loci at the *Prostate stem cell antigen* (*PSCA*) on 8q24 and at the *ABO blood group* (*ABO*) on 9q34. A C-allele of rs 2294008 at *PSCA* increased the risk of duodenal ulcer (odds ratio (OR) of 1.84 with *P* value of 3.92×10^{-33}) in a recessive model, while it decreased the risk of gastric cancer (OR of 0.79 with *P* value of 6.79×10^{-12}) as reported previously[1]. A T-allele of SNP rs2294008 created the upstream translational initiation codon and affects the protein localization from cytoplasm to cell surface. SNP rs505922 on *ABO* also associated with duodenal ulcer in a recessive model (OR of 1.32 with *P* value of 1.15×10^{-10}). Our finding implies the crucial roles of genetic variations on the pathogenesis of duodenal ulcer.

Common variant near the endothelin receptor type A (EDNRA) gene is associated with intracranial aneurysm risk.

The pathogenesis of intracranial aneurysm (IA) formation and rupture is complex, with significant contribution from genetic factors. We previously reported genome-wide association studies based on European discovery and Japanese replication cohorts of 5,891 cases and 14,181 controls that identified five disease-related loci. These studies were based on testing replication of genomic regions that contained SNPs with posterior probability of association (PPA) greater than 0.5 in the discovery cohort. To identify additional IA risk loci, we pursued 14 loci with PPAs in the discovery cohort between 0.1 and 0.5. Twenty-five SNPs from these loci were genotyped using two independent Japanese cohorts, and the results from discovery and replication cohorts were combined by meta-analysis. The results demonstrated significant association of IA with rs6841581 on chromosome 4q31.23, immediately 5′ of the endothelin receptor type A with P=2.2×10(-8) [odds ratio (OR) =1.22, PPA=0.986]. We also observed substantially increased evidence of association for two other regions on chromosomes 12q22 (OR=1.16, P=1.1×10(-7), PPA=0.934) and 20p12.1 (OR= 1.20, P=6.9×10(-7), PPA=0.728). Although endothelin signaling has been hypothesized to play a role in various cardiovascular disorders for over two decades, our results are unique in providing genetic evidence for a significant association with IA and suggest that manipulation of the endothelin pathway may have important implications for the prevention and treatment of IA

A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population.

Hepatitis B virus (HBV) infection is a major health issue worldwide which may lead to hepatic dysfunction, liver cirrhosis and hepatocellular carcinoma. To identify host genetic factors that are associated with chronic hepatitis B (CHB) susceptibility, we previously conducted a two-stage genome-wide association study (GWAS) and identified the association of HLA-DP variants with CHB in Asians; however, only 179 cases and 934 controls were genotyped using genome-wide single nucleotide polymorphism (SNP) arrays. Here, we performed a second GWAS of 519 747 SNPs in 458 Japanese CHB cases and 2056 controls. After adjustment with the previously identified variants in the HLA-DP locus (rs9277535), we detected strong associations at 16 loci with P-value of <5×10(-5). We analyzed these loci in three independent Japanese cohorts (2209 CHB cases and 4440 controls) and found significant association of two SNPs (rs2856718 and rs7453920) within the HLA-DQ locus (overall P-value of 5.98×10(-28) and 3.99×10(-37)). Association of CHB with

SNPs rs2856718 and rs7453920 remains significant even after stratification with rs3077 and rs 9277535, indicating independent effect of HLA-DQ variants on CHB susceptibility (P-value of $1.52\times10(-21)$- $2.38\times10(-30)$). Subsequent analyses revealed DQA1*0102-DQB1*0604 and DQA 1*0101-DQB1*0501 [odds ratios (OR) =0.16, and 0.39, respectively] as protective haplotypes and DQA1*0102-DQB1*0303 and DQA1*0301-DQB1*0601 (OR=19.03 and 5.02, respectively) as risk haplotypes. These findings indicated that variants in antigen-binding regions of HLA-DP and HLA-DQ contribute to the risk of persistent HBV infection.

A genome-wide association study of nephrolithiasis in the Japanese population identifies novel susceptible loci at 5q35.3, 7p14.3 and 13q14.1

Nephrolithiasis is a common nephrologic disorder with complex etiology. To identify the genetic factor(s) for nephrolithiasis, we conducted a three-stage genome-wide association study (GWAS) using a total of 5,892 nephrolithiasis cases and 17,809 controls of Japanese origin. Here we found three novel loci for nephrolithiasis: *RGS14-SLC34A1-PFN3-F12* on 5q35.3 (rs 11746443; $P=8.51\times10^{-12}$, odds ratio (OR)= 1.19), *INMT-FAM188B-AQP1* on 7p14.3 (rs 1000597; $P=2.16\times10^{-14}$, OR=1.22), and *DGKH* on 13q14.1 (rs4142110; $P=4.62\times10^{-9}$, OR=1.14). Subsequent analyses in 21,842 Japanese subjects revealed the association of SNP rs11746443 with the reduction of estimated glomerular filtration rate (eGFR) ($P=6.54\times10^{-8}$), suggesting the crucial role of this variation on renal function. Our findings elucidated the significance of genetic variations for the pathogenesis of nephrolithiasis.

References

1. H.-S. Cho, T. Suzuki, N. Dohmae, S. Hayami, M. Unoki, M. Yoshimatsu, G. Toyokowa, M. Takawa, T. Chen, J.K. Kurash, H.I. Field, B.A.J. Ponder, Y. Nakamura, and R. Hamamoto: Demethylation of RB regulator MYPT1 by histone demethylase LSD1 promotes cell cycle progression in cancer cells. Cancer Research, 71655-660, 2011
2. R. Cui, Y. Okada, S.G. Jang, J.L. Ku, J.G. Park, Y. Kamatani, N. Hosono, T. Tsunoda, V. Kumar, C. Tanikawa, N. Kamatani, R. Yamada, M. Kubo, Y. Nakamura, and K. Matsuda: Common variant in 6q26-q27 is associated with distal colon cancer in Asian population. Gut, 60799-805, 2011
3. K. Imai, S. Hirata, A. Irie, S. Senju, Y. Ikuta, K. Yokomine, M. Harao, M. Inoue, Y. Tomita, T. Tsunoda, H. Nakagawa, Y. Nakamura, H. Baba, and Y. Nishimura: Identification of HLA-A2-restricted CTL epitopes of a novel tumor-associated antigen, KIF20A, overexpressed in pancreatic cancer. British Journal of Cancer, 104300-307, 2011
4. A. Iida, N. Hosono, M. Sano, T. Kamei, S. Oshima, T. Tokuda, M. Kubo, Y. Nakamura, and S. Ikegawa: Optineurin mutations in Japanese amyotrophic lateral sclerosis. Journal of Neurology, Neurosurgery & Psychiatry, Jan 8. [Epub ahead of print], 2011
5. N. Akuta, F. Suzuki, M. Hirakawa, Y. Kawamura, H. Sezaki, Y. Suzuki, T. Hosaka, M. Kobayashi, M. Kobayashi, S. Saitoh, Y. Arase, K. Ikeda, K. Chayama, Y. Nakamura, and H. Kumada: Amino acid substitution in HCV core/NS5A region and genetic variation near IL28B gene affect treatment efficacy to interferon plus ribavirin combination therapy. Intervirology, Feb 16, [Epub ahead of print], 2011
6. T. Ozeki, T. Mushiroda, A. Yowang, A. Takahashi, M. Kubo, Y. Shirakata, Z. Ikezawa, M. Iijima, T. Shiohara, K. Hashimoto, N. Kamatani, and Y. Nakamura: Genome-wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. Human Molecular Genetics, 20: 1034-1041, 2011
7. Y. Fujimoto, H. Ochi, T. Maekawa, H. Abe, C.N. Hayes, H. Kumada, Y. Nakamura, and K. Chayama: A single nucleotide polymorphism in activated cdc42 associated tyrosine kinase 1 influences the interferon therapy in hepatitis C patients. J. Hepatology, 54: 629-639, 2011
8. Y. Hashimoto, H. Ochi, H. Abe, Y. Hayashida, M. Tsuge, F. Mitsui, N. Hiraga, M. Imamura, S. Takahashi, C.N. Hayes, W. Ohishi, M. Kubo, T. Tsunoda, N. Kamatani, Y. Nakamura, and K. Chayama: Prediction of response to peginterferon-alfa-2b plus ribavirin therapy in Japanese patients infected with hepatitis C virus genotype 1b. J. Med. Virol., 83981-988, 2011
9. F. Suzuki, Y. Suzuki, N. Akuta, H. Sezaki, M. Hirakawa, Y. Kawamura, T. Hosaka, M. Kobayashi, S. Saito, Y. Arase, K. Ikeda, M. Kobayashi, K. Chayama, N. Kamatani, Y. Nakamura, Y. Miyakawa, and H. Kumada:

Influence of ITPA polymorphism on decreases of hemoglobin during treatment with pegylated IFN, ribavirin and telaprevir. Hepatology, 53415-421, 2011
10. T. Kawaoka , C.N. Hayes, W. Ohishi, H. Ochi, T. Maekawa, H. Abe, M. Tsuge, F. Mitsui, N. Hiraga, M. Imamura, S. Tkakahashi, M. Kubo, T. Tsunoda, Y. Nakamura, H. Kumada, and K. Chayama: Predictive value of IL28B polymorphism on the effect of interferon therapy in chronic hepatitis C patients with genotypes 2a and 2b. J. Hepatology, 54408-414, 2011
11. A. Aoki, K. Ozaki, H. Sato, A. Takahashi, M. Kubo, Y. Sakata, Y. Onouchi, T. Kawaguchi, T.-H. Lin, H. Takano, M. Yasutake, P.-C. Hsu, S. Ikegawa, N. Kamatani, T. Tsunoda, S.-H. H. Juo, M. Hori, I. Komuro, K. Mizuno, Y. Nakamura, and T. Tanaka: SNPs on chromosome 5p15.3 associated with myocardial infarction in Japanese population. Journal of Human Genetics, 5647-51, 2011
12. M. Yoshimatsu, G. Toyokawa, S. Hayami, M. Unoki, T. Tsunoda, H.I. Field, J.D. Kelly, D.E. Neal, Y. Maehara, B.A.J. Ponder, Y. Nakamura, and R. Hamamoto: Dysregulation of PRMT1 and PRMT6, type I arginine methyltransferases, is involved in various types of human cancers. Int. J. Cancer, 128562-573, 2011
13. C.N. Hayes, M. Kobayashi, N. Akuta, F. Suzuki, H. Kumada, H. Abe, D. Miki, M. Imamura, H. Ochi, N. Kamatani, Y. Nakamura, and K. Chayama: HCV substitutions and IL28B polymorphisms on outcome of peg-interferon plus ribavirin combination therapy. Gut, 60261-267, 2011
14. P. Saetre, M. Vares, T. Werge, O.A. Andreassen, T. Arinami, H. Ishiguro, S. Nanko, E.C. Tan, D.H. Han, J. Roffman, J.-W. Muntjewerff, P.P. Jagodzinski, B. Kempisty, J. Hauser, E. Vilella, E. Betcheva, Y. Nakamura, B. Regland, I. Agartz, H. Hall, L. Terenius, and E.G. Jönsson: Methylenetetrahydrofolate reductase (MTHFR) C677T and A1298C polymorphisms and age of onset in schizophrenia: a combined analysis of independent samples. Neuropsychiatric Genetics, Jan 13. [Epub ahead of print]:, 2011
15. S. Chung, H. Nakagawa, M. Uemura, L. Piao, K. Ashikawa, N. Hosono, R. Takata, S. Akamatsu, T. Kawaguchi, T. Morizono, T. Tsunoda, Y. Daigo, K. Matsuda, N. Kamatani, Y. Nakamura, and M. Kubo: Association of a novel long non-coding RNA in 8q 24 with prostate cancer susceptibility. Cancer Science, 102245-252, 2011
16. L. Piao, H. Nakagawa, K. Ueda, S. Chung, K. Kashiwaya, H. Eguchi, H. Ohigashi, O. Ishikawa, Y. Daigo, K. Matsuda, and Y. Nakamura: C12orf48, termed PARP-1 binding protein (PARPBP), enhances poly (ADP-ribose) polymerase-1 (PARP-1) activity and protects pancreatic cancer cells from DNA damage. Genes Chromosomes and Cancer, 5013-24, 2011
17. A. Iida, Takahashi, M. Deng, Y. Zhang, J. Wang, N. Atsuta, F. Tanaka, T. Kamei, M. Sano, S. Oshima, T. Tokuda, M. Morita, C. Akimoto, M. Nakajima, M. Kubo, N. Kamatani, I. Nakano, G. Sobue, Y. Nakamura, D. Fan, and S. Ikegawa: Replication analysis of SNPs on 9p21.2 and 19p13.3 with amyotrophic lateral sclerosis in East Asians. Neurobiology of Aging, in press, 2011
18. Y. Tomita, K. Imai, S. Senju, A. Irie, M. Inoue, Y. Hayashida, K. Shiraishi, T. Mori, Y. Daigo, T. Tsunoda, T. Ito, H. Nomori, Y. Nakamura, H. Kohrogi, and Y. Nishimura: A novel tumor-associated antigen, cell division cycle 45-like can induce cytotoxic T lymphocytes reactive to tumor cells. Cancer Science, in press, 2011
19. J.-H. Park, T. Katagiri, S. Chung, K. Kijima, and Y. Nakamura: Polypeptide N-acetylgalactosaminyltransferase 6 (GALNT6) disrupts mammary acinar morphogenesis through O-glycosylation of fibronectin. Neoplasia, in press, 2011
20. Y. Kato, H. Zembutsu, R. Takata, F. Miya, T. Tsunoda, W. Obara, T. Fujioka, and Y. Nakamura: Predicting response of bladder cancers to gemcitabine and carboplatin neoadjuvant chemotherapy through genome-wide gene expression profiling. Experimental and Therapeutic Medicine, 247-56, 2011
21. H. Zembutsu, M. Sasa, K. Kiyotani, T. Mushiroda, and Y. Nakamura: Should CYP2 D6 inhibitors be administered in conjunction with tamoxifen? Expert Review of Anticancer Therapy, in press, 2011
22. S. Maeda, D. Koya, S. Araki, T. Babazono, T. Umezono, M. Toyoda, K. Kawai, M. Imanishi, T. Uzu, D. Suzuki, H. Maegawa, A. Kashiwagi, Y. Iwamoto, and Y. Nakamura: Association of single nucleotide polymorphisms within genes encoding sirtuin families with diabetic nephropathy in Japanese subjects with type 2 diabetes. Clinical Experimental Nephrology, in press, 2011
23. K. Okamoto, K. Tokunaga, K. Doi, T. Fujita, H. Suzuki, T. Katoh, T. Watanabe, N. Nishida, A. Mabuchi, A. Takahashi, M. Kubo, S. Maeda, Y. Nakamura, and E. Noiri: Common variation in GPC5 is associated with acquired nephrotic syndrome. Nature Genetics, 43459-463, 2011

24. T.A. Johnson, Y. Niimura, H. Tanaka, Y. Nakamura, and T. Tsunoda: hzAnalyzer: Detection, quantification, and visualization of contiguous homozygosity in high-density genotyping datasets. Genome Biology, in press, 2011
25. K.G. Tantisira, J. Lasky-Su, M. Harada, A. Murphy, A.A. Litonjua, B.E. Himes, C. Lange, R. Lazarus, J. Sylvia, B. Klanderman, Q.L. Duan, W. Qiu, T. Hirota, F.D. Martinez, D. Mauger, C. Sorkness, S. Szefler, S.C. Lazarus, R.F. Lemanske Jr, S.P. Peters, J.J. Lima, Y. Nakamura, M. Tamari, and S.T. Weiss: Genome-wide association of GLCCI1 with asthma steroid treatment response. New Eng. J. Med., in press, 2011
26. V. Kumar, N. Kato, K. Urabe, A. Takahashi, R. Muroyama, N. Hosono, M. Otsuka, R. Tateishi, M. Omata, H. Nakagawa, K. Koike, N. Kamatani, M. Kubo, Y. Nakamura, and K. Matsuda: Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nature Genetics, 43455-458, 2011
27. P.-C. Cha, A. Takahashi, N. Hosono, S.-K. Low, N. Kamatani, M. Kubo, and Y. Nakamura: A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. Nature Genetics, 43447-450, 2011
28. F. Nyberg, B.J. Barratt, T. Mushiroda, A. Takahashi, A. Jawaid, S. Hada, T. Umemura, M. Fukuoka, K. Nakata, Y. Ohe, H. Kato, S. Kudoh, R. March, Y. Nakamura, and N. Kamatani: Interstitial lung disease in gefitinib-treated Japanese patients with non-small-cell lung cancer: genome-wide analysis of genetic data. Pharmacogenomics, 12965-975, 2011
29. V. Kumar, K. Matsuo, A. Takahashi, N. Hosono, T. Tsunoda, N. Kamatani, S.-Y. Kong, H. Nakagawa, R. Cui, C. Tanikawa, M. Seto, Y. Morishima, M. Kubo, Y. Nakamura, and K. Matsuda: Common variants on 14q32 and 13q12 are associated with DLBCL susceptibility. Journal of Human Genetics, 56436-439, 2011
30. Y. Okada, T. Hirota, Y. Kamatani, A. Takahashi, H. Ohmiya, N. Kumasaka, K. Higasa, Y. Yamaguchi-Kabata, N. Hosono, M.A. Nalls, M.H. Chen, F.J.A. van Rooij, A.V. Smith, T. Tanaka, D.J. Couper, N.A. Zakai, L. Ferrucci, D.L. Longo, D.G. Hernandez, J. C.M. Witteman, T.B. Harris, C.J. O'Donnell, S.K. Ganesh, K. Matsuda, T. Tsunoda, T. Tanaka, M. Kubo, Y. Nakamura, M. Tamari, K. Yamamoto, and N. Kamatani: Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. PLoS Genetics, 7e1002067, 2011"
31. K. Ueda, N. Saichi, S. Takami, D. Kang, A. Toyama, Y. Daigo, N. Ishikawa, N. Kohno, K. Tamura, T. Shuin, A. Yamane, M. Ota, T. Sato, Y. Nakamura, and H. Nakagawa: Rapid and efficient peptidome profiling technology for the comprehensive identification of serum peptide biomarkers. PLoS ONE, in press, 2011
32. A. Toyama, H. Nakagawa, K. Matsuda, N. Ishikawa, N. Kohno, Y. Daigo, T. Sato, Y. Nakamura, and K. Ueda: Deglycosylation and label-free quantitative LC-MALDI MS applied to efficient serum biomarker discovery of lung cancer. Proteome Science, in press, 2011
33. M. Takawa, K. Masuda, M. Kunizaki, Y. Daigo, K. Takagi, Y. Iwai, H.-S. Cho, G. Toyokawa, Y. Yamane, K. Maejima, H.I. Field, T. Kobayashi, T. Akasu, M. Sugiyama, E. Tsuchiya, Y. Atomi, B.A.J. Ponder, Y. Nakamura, and R. Hamamoto: Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker. Cancer Science, in press, 2011
34. I. Kou, A. Takahashi, T. Urano, N. Fukui, H. Ito, K. Ozaki, T. Tanaka, T, Hosoi, M. Shiraki, S. Inoue, Y. Nakamura, N. Kamatani, M. Kubo, S. Mori, and S. Ikegawa: Common variants in a novel gene, FONG on chromosome 2q33.1 confer risk of osteoporosis in Japanese. PLoS ONE, in press, 2011
35. M.A. Nalls, D.J. Couper, T. Tanaka, F.J.A. van Rooij, M-H. Chen, A.V. Smith, D. Toniolo, N.A. Zakai, Q. Yang, A. Greinacher, A.R. Wood, M. Garcia, P. Gasparini, Y. Liu, T. Lumley, A.R. Folsom, A.P. Reiner, C. Gieger, V. Lagou, J.F. Felix, H. Völzke, N.A. Gouskova, A. Biffi, A. Döring, U. Völker, S. Chong, K.L. Wiggins, A. Rendon, A. Dehghan, M. Moore, K. Taylor, J.G. Wilson, G. Lettre, A. Hofman, J.C. Bis, N. Pirastu, C. S. Fox, C. Meisinger, J. Sambrook, S. Arepalli, M. Nauck, H. Prokisch, J. Stephens, N. L. Glazer, L.A. Cupples, Y. Okada, A. Takahashi, Y. Kamatani, K. Matsuda, T. Tsunoda, T. Tanaka, M. Kubo, Y. Nakamura, K. Yamamoto, N. Kamatani, M. Stumvoll, A. Tonjes, I. Prokopenko, T. Illig, K.V. Patel, S. F. Garner, B. Kuhnel, M. Mangino, B.A. Oostra, S.L. Thein, J. Coresh, H.-E. Wichmann, S. Menzel, J. Lin, G. Pistis, A.G. Uitterlinden, T.D. Spector, A. Teumer, G. Eriksdottir, V. Gudnason, S. Bandinelli, T. Frayling, A. Chakravarti, C.M. van Duijn, D. Melzer, W. H. Ouwehand, D. Levy, E. Boerwinkle, A.B. Singleton, D.G. Hernandez, D.L. Longo, N. Soranzo, J.C.M. Witteman, B.M. Psaty, L.

Ferrucci, T.B. Harris, C.J. O'Donnell, and S. K. Ganesh: Multiple loci are associated with white blood cell phenotypes. PLoS Genetics, 7e1002113, 2011
36. C. Terao, R. Yamada, K. Ohmura, M. Takahashi, T. Kawaguchi, Y. Kochi, Human Disease Genomics Working Group, RA Clinical and Genetic Study Consortium, Y. Okada, Y. Nakamura, K. Yamamoto, I. Melchers, M. Lathrop, T. Mimori, and F. Matsuda: The human AIRE gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. Human Molecular Genetics, in press:, 2011
37. K. Chayama, et al. IL28B but not ITPA polymorphism is predictive of response to peg-interferon, ribavirin and telaprevir triple therapy in patients with genotype 1 hepatitis C. Journal of Infectious Diseases, in press:, 2011
38. T. Azakami, C.N. Hayes, H. Sezaki, M. Kobayashi, N. Akuta, F. Suzuki, H. Kumada, H. Abe, D. Miki, M. Tsuge, M. Imamura, Y. Kawakami, S. Takahashi, H. Ochi, Y. Nakamura, N. Kamatani, and K. Chayama: Common genetic polymorphism of ITPA gene affects ribavirin-induced anemia and effect of peg-interferon plus ribavirin therapy. J. Med. Virol., 831048-1057, 2011
39. H. Ochi, T. Maekawa, H. Abe, Y. Hayashida, R. Nakano, M. Imamura, N. Hiraga, Y. Kawakami, S. Aimitsu, J.-H. Kao, M. Kubo, T. Tsunoda, H. Kumada, Y. Nakamura, C.N. Hayes, and K. Chayama: IL-28B predicts response to chronic hepatitis C therapy - fine-mapping and replication study in Asian populations. Journal or General Virology, 921071-1081, 2011
40. T. Mushiroda, Y. Nakamura: Personalizing carbamazepine therapy (Review). Genomic Medicine, in press, 2011
41. S. Chantarangsu, T. Mushiroda, S. Mahasirimongkol, S. Kiertiburanakul, S. Sungkanuparph, W. Manosuthi, W. Tantisiriwat, A. Charoenyingwattana, T. Sura, A. Takahashi, M. Kubo, N. Kamatani, W. Chantratita, and Y. Nakamura: Genome-wide association study identifies variations in 6p21.3 associated with nevirapine-induced rash. Clinical Infectious Diseases, in press, 2011
42. G. Toyokawa, M. Yoshimatsu, K. Masuda, Y. Daigo, E. Tsuchiya, H.-S. Cho, M. Takawa, S. Hayami, K. Maejima, M. Chino, H.I. Field, D.E. Neal, B.A.J. Ponder, Y. Maehara, Y. Nakamura, and R. Hamamoto: Minichromosome Maintenance Protein 7 is a potential therapeutic target in human cancer and a novel prognostic marker of non-small cell lung cancer. Molecular Cancer, in press, 2011
43. Y. Srinivasan, M. Sasa, J. Honda, A. Takahashi, S. Uno, N. Kamatani, M. Kubo, Y. Nakamura. and H. Zembutsu: Genome-wide association study of epirubicin-induced leukopenia in Japanese patients. Pharmacogenetics and Genomics, in press, 2011
44. H.-S. Cho, J.D. Kelly, S. Hayami, G. Toyokawa, M. Takawa, M. Yoshimatsu, T. Tsunoda, H.I. Field, D.E. Neal, B.A.J. Ponder, Y. Nakamura, and R. Hamamoto: Enhanced expression of EHMT2 is involved in the proliferation of cancer cells through negative regulation of SIAH. Neoplasia, 13: 676-684, 2011
45. T. Tanabe, N. Yamaguchi, K. Matsuda, K. Yamazaki, S. Takahashi, A. Tojo, M. Onizuka, Y. Eishi, H. Akiyama, J. Ishikawa, T. Mori, M. Hara, K. Koike, K. Kawa, T. Kawase, Y. Morishima, H. Amano, M. Kobayashi-Miura, T. Kakamu, Y. Nakamura, S. Asano, and Y. Fujita: Association analysis of the NOD2 gene with susceptibility to graft-versus-host disease in a Japanese population. International Journal of Hematology, DOI 10.1007/s12185-011-0860-5, 2011
46. D. Miki, H. Ochi, C.N. Hayes, H. Abe, T. Yoshima, H. Aikata, K. Ikeda, H. Kumada, J. Toyota, T. Morizono, T. Tsunoda, M. Kubo, Y. Nakamura, N. Kamatani, and K. Chayama: Variation in the DEPDC5 locus is associated with progression to hepatocellular carcinoma in chronic hepatitis C virus carriers. Nature Genetics, in press, 2011
47. T. Hirota, A. Takahashi, M. Kubo, T. Tsunoda, K. Tomita, S. Doi, K. Fujita, A. Miyatake, T. Enomoto, T. Miyagawa, M. Adachi, H. Tanaka, A. Niimi, H. Matsumoto, I. Ito, H. Masuko, T. Sakamoto, N. Hizawa, M. Taniguchi, J.J. Lima, C.G. Irvin, S.P. Peters, B.E. Himes, A.A. Litonjua, K.G. Tantisira, S. T. Weiss, N. Kamatani, Y. Nakamura, and M. Tamari: Genome-wide association study identifies five susceptibility loci for adult asthma in the Japanese population. Nature Genetics, in press:, 2011
48. S. Arakawa, A. Takahashi, K. Ashikawa, N. Hosono, T. Aoi, M. Yasuda, Y. Oshima, S. Yoshida, H. Enaida, T. Tsuchihashi, K. Mori, S. Honda, A. Negi, A. Arakawa, K. Kadonosono, Y. Kiyohara, N. Kamatani, Y. Nakamura, T. Ishibashi, and M. Kubo: Genome-wide association study identifies two susceptibility loci for exudative age-related macular degeneration in the Japanese population. Nature Genetics, in press:, 2011
49. A.P. Reiner, G. Lettre, M.A. Nalls, S.K. Ganesh, R. Mathias, M.A. Austin, E. Dean, S.

Arepalli, A. Britton, Z. Chen, D. Couper, J.D. Curb, C.B. Eaton, M. Fornage, S.F.A. Grant, T.B. Harris, D. Hernandez, N. Kamatani, B.J. Keating, M. Kubo, A. LaCroix, L.A. Lange, S. Liu, K. Lohman, Y. Meng, E.R. Mohler III, S. Musani, Y. Nakamura, C.J. O'Donnell, Y. Okada, C.D. Palmer, G.J. Papanicolaou, K.V. Patel, A.B. Singleton, A. Takahashi, H. Tang, H.A. Taylor Jr., K. Taylor, C. Thomson, L.R. Yanek, L. Yang, E. Ziv, A.B. Zonderman, A. R. Folsom, M.K. Evans, Y. Liu, D.M. Becker, B.M. Snively, and J.G. Wilson: Genome-wide association study of white blood cell count in 16,388 African Americans: the Continental Origins and Genetic Epidemiology Network (COGENT). PLoS Genetics, 7e1002108, 2011

50. Y. Okada, A. Takahashi, H. Ohmiya, N. Kumasaka, Y. kamatani, N. Hosono, T. Tsunoda, K. Matsuda, T. Tanaka, M. Kubo, Y. Nakamura, K. Yamamoto, and N. Kamatani: Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. Human Molecular Genetics, 201224-1231, 2011
51. A. Iida, A. Takahashi, M. Kubo, S. Saito, N. Hosono, Y. Ohnishi, K. Kiyotani, T. Mushiroda, M. Nakajima, K. Ozaki, T. Tanaka, T. Tsunoda, S. Oshima, M. Sano, T. Kamei, T. Tokuda, M. Aoki, K. Hasegawa, K. Mizoguchi, M. Morita , Y. Takahashi, M. Katsuno, N. Atsuta, H. Watanabe, F. Tanaka, R. Kaji, I. Nakano, N. Kamatani, S. Tsuji, G. Sobue, Y. Nakamura, and S. Ikegawa: A functional variant in ZNF512B is associated with susceptibility to amyotrophic lateral sclerosis in Japanese. Human Molecular Genetics, in press:, 2011
52. A. Yosifova, T. Mushiroda, M. Kubo, A. Takahashi, Y. Kamatani, N. Kamatani, D. Stoianov, R. Vazharova, S. Karachanak, I. Zaharieva, I. Dimova, S. Hadjidekova, V. Milanova, N. Madjirova, I. Gerdjikov, T. Tolev, N. Poryazova, G. Kirov, M. Owen, M. O'Donovan, D. Toncheva, and Y. Nakamura: Genome wide association study on bipolar disorder in the Bulgarian population. Genes, Brain and Behavior, in press, 2011
53. H. Mbarek, H. Ochi, Y. Urabe, V. Kumar, M. Kubo, N. Hosono, A. Takahashi, Y. Kamatani, D. Miki, H. Abe, T. Tsunoda, N. Kamatani, K. Chayama, Y. Nakamura, and K. Matsuda: A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. Human Molecular Genetics, doi:10.1093/hmg/ddr301, 2011
54. S. Chung, M. Nakashima, H. Zembutsu, and Y. Nakamura: Possible involvement of NEDD4 in keloid formation; its critical role in fibroblast proliferation and collagen production. Proceedings of the Japan Academy, Ser. B, in press, 2011
55. M. Aragaki, K. Takahashi, H. Akiyama, E. Tsuchiya, S. Kondo, Y. Nakamura, and Y. Daigo: Characterization of a Cleavage Stimulation Factor, 3' pre-RNA, Subunit 2, 64kDa (CSTF2) as a therapeutic target for lung cancer. Clinical Cancer Research, in press, 2011
56. R. Nishino, A. Takano, N. Ishikawa, H. Akiyama, H. Ito, H. Nakayama, Y. Miyagi, E. Tsuchiya, N. Kohno, Y. Nakamura, and Y. Daigo: Identification of Epstein-Barr virus-induced gene 3 as a novel serum and tissue biomarker and a therapeutic target for lung cancer. Clinical Cancer Research, in press, 2011
57. M. Maimbo, K. Kiyotani, T. Mushiroda, C. Masimirembwa, and Y. Nakamura: CYP2B6 genotype is a strong predictor for systemic exposure of efavirenz in HIV-infected Zimbabweans. European Journal of Clinical Pharmacology, in press, 2011
58. I.P.M. Tomlinson, M. Dunlop, H. Campbell, B. Zanke, S. Gallinger, T. Hudson, T. Koessler, P.D. Pharoah, I. Niittymäkix, S. Tuupanenx, L.A. Aaltonen, K. Hemminki, A. Lindblom, A. Försti, O. Sieber, L. Lipton, T. van Wezel, H. Morreau, J.T. Wijnen, P. Devilee, K. Matsuda, Y. Nakamura, S. CastellvÍ-Bel, C. Ruiz-Ponte, A. Castells, A. Carracedo, J.W.C. Ho, P. Sham, R.M.W. Hofstra, P. Vodicka, H. Brenner, J. Hampe, C. Schafmayer, J. Tepel, S. Schreiber, H. Völzke, M. M. Lerch, C.A. Schmidt, S. Buch, V. Moreno, C.M. Villanueva, P. Peterlongo, P. Radice, M.M. Echeverry, A. Velez, L. Carvajal-Carmona, R. Scott, S. Penegar, P. Broderick, A. Tenesa, and R.S. Houlston: COGENT (COlorectal cancer GENeTics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer. Mutagenesis, in press, 2011
59. J.C. Chambers, W. Zhang, J. Sehm, Y. Nakamura et al. Identification of genetic loci influencing markers of liver function in man. Nature Genetics, in press, 2011
60. N. Kumasaka , H. Fujisawa, N. Hosono, Y. Okada, A. Takahashi, Y. Nakamura, M. Kubo, and N. Kamatani: Platinum CNV: a bayesian gaussian mixture model for genotyping copy number polymorphisms using SNP array signal intensity data. Genetic Epidemiology, in press, 2011
61. K. Kiyotani, T. Mushiroda, and Y. Nakamura: Pharmacogenomics of anticonvulsant agents (review). BioTech International, in press, 2011

62. G. Toyokawa, H.-S. Cho, K. Masuda, M. Yoshimatsu, S. Hayami, M. Takawa, Y. Iwai, Y. Daigo, E. Tsuchiya, T. Tsunoda, H.I. Field, J.D. Kelly, D.E. Neal, Y. Maehara, B.A. J. Ponder, Y. Nakamura, and R. Hamamoto: The histone lysine methyltransferase Wolf-Hirschhorn Syndrome Candidate 1 is involved in human carcinogenesis through regulation of the Wnt pathway. Neoplasia, in press, 2011
63. J. Li, D. Yang, Y. He, M. Wang, Z. Wen, L. Liu, J. Yao, K. Matsuda, Y. Nakamura, J. Yu, X. Jiang, S. Sun, Q. Liu, X. Jiang, Q. Song, M. Chen, H. Yang, F. Tang, X. Hu, J. Wang, Y. Chang, X. He, Y. Chen, and J. Lin: Associations of HLA-DP variants with hepatitis B virus infection in Southern and Northern Han Chinese populations: A multicenter case-control study. PLoS ONE, 6e24221, 2011
64. G. Toyokawa, H.-S. Cho, Y. Iwai, M. Takawa, M. Yoshimatsu, S. Hayami, K. Maejima, N. Shimizu, H. Tanaka, T. Tsunoda, H.I. Field, J.D. Kelly, D.E. Neal, B.A.J. Ponder, Y. Maehara, Y. Nakamura, and R. Hamamoto: The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase. Cancer Prevention Research, 4:2051-2061, 2011
65. K. Kiyotani, T. Mushiroda, Y. Nakamura, and H. Zembutsu: Pharmacogenomics of tamoxifen: roles of drug metabolizing enzymes and transporters. Drug Metabolism and Pharmacokinetics, in press, 2011
66. H.-S. Cho, G. Toyokawa, Y. Daigo, S. Hayami, K. Masuda, N. Ikawa, Y. Yamane, K. Maejima, M. Yoshimatsu, T. Tsunoda, H. I. Field, J.D. Kelly, D.E. Neal, B.A.J. Ponder, Y. Maehara, Y. Nakamura, and R. Hamamoto: The JmjC domain-containing histone demethylase KDM3A Is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene. Int. J. Cancer, in press, 2011
67. M. Imamura, M. Iwata, H. Maegawa, H. Watada, H. Hirose, Y. Tanaka, K. Tobe, K. Kaku, A. Kashiwagi, R. Kawamori, Y. Nakamura, and S. Maeda: Genetic variants at CDC123/CAMK1D and SPRY2 are associated with susceptibility to type 2 diabetes in the Japanese population. Diabetologia, in press, 2011
68. T. Ohshige, M. Iwata, S. Omori, Y. Tanaka, H. Hirose, K. Kaku, H. Maegawa, H. Watada, A. Kashiwagi, R. Kawamori, K. Tobe, T. Kadowaki, Y. Nakamura, and S. Maeda: Association of new loci identified in European genome-wide association studies with susceptibility to type 2 diabetes in the Japanese. PLoS ONE, 6e26911, 2011
69. C. Gieger, A. Radhakrishnan, A. Cvejic, W. Tang, Y. Nakamura, et al.: New gene functions in megakaryopoiesis and platelet formation. Nature, in press, 2011
70. J. Koinuma, H. Akiyama, M. Fujita, M. Hosokawa, E. Tsuchiya, S. Kondo, Y. Nakamura, and Y. Daigo: Characterization of an Opa interacting protein 5 (OIP5) involved in lung and esophageal carcinogenesis. Cancer Science, in press, 2011
71. F. Innocenti, K. Owzar, N.L. Cox, P. Evans, M. Kubo, H. Zembutsu, C. Jiang, D. Hollis, T. Mushiroda, L. Li, P. Friedman, L. Wang, H. Hurwitz, K.M. Giacomini, H.L. McLeod, R.M. Goldberg, R.L. Schilsky, H.L. Kindler, Y. Nakamura, and M.J. Ratain: A genome-wide association study of overall survival in pancreatic cancer patients treated with gemcitabine in CALGB 80303. Clinical Cancer Research, in press, 2011
72. M. Furu, Y. Kajita, S. Nagayama, T. Ishibe, Y. Shima, K. Nishijo, D. Uejima, R Takahashi, T. Aoyama, T. Nakayama, T Nakamura, Y. Nakashima, M. Ikegawa, S. Imoto, T. Katagiri, Y. Nakamura, and J. Toguchida: Identification of AFAP1L1 as a prognostic marker for spindle cell sarcomas. Oncogene, 304015-4025, 2011
73. K. Masuda, A. Takano, H. Oshita, H. Akiyama, E. Tsuchiya, N. Kohno, Y. Nakamura, and Y. Daigo: Chondrolectin is a novel diagnostic biomarker and a therapeutic target for lung cancer. Clinical Cancer Research, in press, 2012
74. K. Kiyotani, S. Uno, T. Mushiroda, A. Takahashi, M. Kubo, N. Mitsuhata, S. Ina, C. Kihara, Y. Kimura, H. Yamaue, K. Hirata, Y. Nakamura, and H. Zembutsu: Genome-wide association study identifies four genetic markers for hematological toxicities in cancer patients receiving gemcitabine therapy. Pharmacogenetics and Genomics, in press, 2012
75. Y. Okada, M. Kubo, H. Ohmiya, A. Takahashi, N. Kumasaka, N. Hosono, S. Maeda, W. Wen, R. Dorajoo, M.-J. Go, W. Zheng, N. Kato, J.-Y. Wu, Q. Lu, the GIANT consortium, T. Tsunoda, K. Yamamoto, Y. Nakamura, N. Kamatani, and T. Tanaka: Common variants at CDKAL1 and KLF9 are associated with body mass index in East Asian populations. Nature Genetics, in press, 2012
76. K. Dabanaka, S.-Y. Chung, H. Nakagawa, Y. Nakamura, T. Okabayashi, T. Sugimoto, K. Hanazaki, and M. Furihata: PKIB expression strongly correlated with phosphorylated Akt expression in the breast cancers and also with triple negative breast cancer subtype.

Medical Molecular Morphology, in press, 2012
77. W. Wen, Y.S. Cho, W. Zheng, R. Dorajoo, Y. Nakamura, et al.: Meta-analysis of genome-wide association studies in East Asians identifies novel genetic variants associated with body mass index. Nature Genetics, in press, 2012
78. R. Abo, S. Hebbring, Y. Ji, H. Zhu, Z.-B. Zeng, A. Batzler, G.D. Jenkins, J. Biernacka, K. Snyder, M. Drews, O. Fiehn, B. Fridley, D. Schaid, N. Kamatani, Y. Nakamura, M. Kubo, T. Mushiroda, R. Kaddurah-Daouk, D.A. Mrazek, and R.M. Weinshilboum: Merging pharmacometabolomics with pharmacogenomics using "1000 Genomes" SNP imputation: Selective serotonin reuptake inhibitor response pharmacogenomics. Pharmacogenetics and Genomics, in press, 2012
79. H. Nakagawa, S. Akamatsu, R. Takata, A. Takahashi, M. Kubo, and Y. Nakamura: Prostate cancer genomics, biology, and risk assessment through genome-wide association studies. Cancer Science, in press, 2012
80. K. Kiyotani, T. Mushiroda, T. Tsunoda, T. Morizono, N. Hosono, M. Kubo, Y. Tanigawara, C.K. Imamura, D.A. Flockhart, F. Aki, K. Hirata, Y. Takatsuka, M. Okazaki, S. Ohsumi, T. Yamakawa, M. Sasa, Y. Nakamura, and H. Zembutsu: A Genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. Human Molecular Genetics, in press, 2012
81. H. Yoshioka, S. Yamamoto, H. Hanaoka, Y. Iida, P. Paudya, T. Higuchi, H. Tominaga, N. Oriuchi, H. Nakagawa, Y. Shiba, K. Yoshida, R. Osawa, T. Katagiri, T. Tsunoda, Y. Nakamura, and K. Endo: In vivo therapeutic effect of CDH3/P-cadherin-targeting radioimmunotherapy. Cancer Immunology, Immunotherapy, in press, 2012
82. C. Tanikawa, Y. Urabe, K. Matsuo, M. Kubo, A. Takahashi, H. Ito, K. Tajima, N. Kamatani, Y. Nakamura, and K. Matsuda: Genome wide association study identified two susceptible loci for duodenal ulcer in Japanese population. Nature Genetics, in press, 2012
83. Y. Urabe, C. Tanikawa, A. Takahashi, Y. Okada, T. Morizono, T. Tsunoda, N. Kamatani, K. Kohri, K. Chayama, M. Kubo, Y. Nakamura, and K. Matsuda: Genome-wide association study of nephrolithiasis in Japanese population identifies novel susceptible loci at 5q35.3, 7p14.3 and 13q14.1. PLoS Genetics, in press, 2012
84. S.-K. Low, A. Takahashi, P.-C. Cha, H. Zembutsu, N. Kamatani, M. Kubo, and Y. Nakamura: Genome-wide association study for intracranial aneurysm in Japanese population identifies three candidate susceptible loci and a functional genetic variant at EDNRA. Human Molecular Genetics, in press, 2012
85. C. Tanikawa, M. Espinosa, A. Suzuki, K. Masuda, K. Yamamoto, E. Tsuchiya, K. Ueda, Y. Daigo, Y. Nakamura, and K. Matsuda: Regulation of histone modification and chromatin structure by the p53-PADI4 pathway. Nature Communications, in press, 2012
86. P.-C. Cha, H. Zembutsu, A. Takahashi, M. Kubo, N. Kamatani, and Y. Nakamura A genome-wide association study (GWAS) identifies SNP in DCC is associated with gallbladder cancer (GC) in the Japanese population Journal of Human Genetics, in press, 2012
87. S. Akamatsu, R. Takata, C.A. Haiman, A. Takahashi, T. Inoue, M. Kubo, M. Furihata, N. Kamatani, J. Inazawa, G.K. Chen, L.L. Marchand, L.N. Kolonel, T. Katoh, Y. Yamano, M. Yamakado, H. Takahashi, H. Yamada, S. Egawa, T. Fujioka, B.E. Henderson, T. Habuchi, O. Ogawa, Y. Nakamura, and H. Nakagawa: Common variants at 11q 12, 10q26 and 3p11.2 are associated with prostate cancer susceptibility in Japanese. Nature Genetics, in press, 2012
88. W.-C. Chang, C.-H. Lee, T. Hirota, L.-F. Wang, S. Doi, A. Miyatake, T. Enomoto, K. Tomita, M. Sakashita, T. Yamada, S. Fujieda, K. Ebe, H. Saeki, S. Takeuchi, M. Furue, W.-C. Chen, Y.-C. Chiu, W.P. Chang, C.-H. Hong, E. Hsi, S.-H. H. Juo, H.-S. Yu, Y. Nakamura, and M. Tamari: ORAI1 genetic polymorphisms associated with the susceptibility of atopic dermatitis in Japanese and Taiwanese populations. PLoS ONE, 1e29387, 2012
89. W. Osman, Y. Okada, Y. Kamatani, M. Kubo, K. Matsuda, and Y. Nakamura: Association of common variants in TNFRSF13B, TNFSF13, and ANXA3 with serum levels of non-albumin protein and immunoglobulin isotypes in the Japanese population. PLoS ONE, in press, 2012
90. H.H. Nguyen, R. Takata, S. Akamatsu, D. Shigemizu, T. Tsunoda, M. Furihata, A. Takahashi, M. Kubo, N. Kamatani, O. Ogawa, T. Fujioka, Y Nakamura, and H. Nakagawa: IRX4 at 5p15 suppresses prostate cancer growth through the interaction with vitamin D receptor, conferring prostate cancer susceptibility. Human Molecular Genetics, in press, 2012
91. Y. Okada, C. Terao, K. Ikari, Y. Kochi, K.

Ohmura, A. Suzuki, T. Kawaguchi, E. Stahl, F. Kurreman, N. Nishida, H. Ohmiya, K. Myouzen, M. Takahashi, T. Sawada, Y. Nishioka, M. Yukioka, T. Matsubara, S. Wakitani, R. Teshima, S. Tohma, K. Takasugi, K. Shimada, A. Murasawa, S. Honjo, K. Matsuo, H. Tanaka, K. Tajima, T. Suzuki, T. Iwamoto, Y. Kawamura, H. Tanii, Y. Okazaki, T. Sasaki, P.K. Gregersen, L. Padyukov, J. Worthington, K.A. Siminovitch, M. Lathrop, A. Taniguchi, A. Takahashi, K. Tokunaga, M. Kubo, Y. Nakamura, N. Kamatani, T. Mimori, R.M. Plenge, H. Yamanaka, S. Momohara, R. Yamada, F. Matsuda, and K. Yamamoto: Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nature Genetics, in press, 2012

92. W. Jongjaroenpraser, T. Phusantisampan, S. Mahasirimongkol, T. Mushiroda, N. Hirankarn, T. Snabboon, S. Chanprasertyotin, P. Tantiwong, S. Soonthornpun, P. Rattanapichart, S. Mamanasiri, T. Himathongkam, B. Ongphiphadhanakul, A. Takahashi, N. Kamatani, M. Kubo, and Y. Nakamura: A genome-wide association study identifies novel susceptibility loci for thyrotoxic hypokalemic periodic paralysis. Human Genetics, in press, 2012
93. M. Aoki, N. Hosono, S. Kakata, Y. Nakamura, N. Kamatani, and M. Kubo: New pharmacogenetic test for detecting HLA-A* 31:01 allele using invaderPlus assay. Pharmacogenetics and Genomics, in press, 2012
94. K. Kiyotani, T. Mushiroda, C.K. Imamura, Y. Tanigawara, N. Hosono, M. Kubo, M. Sasa, Y. Nakamura, and H. Zembutsu: Dose-adjustment study of tamoxifen based on CYP2D6 genotypes in Japanese breast cancer patients. Breast Cancer Res Treat, in press, 2012
95. C. Tanikawa, H. Nakagawa, Y. Furukawa, Y. Nakamura, and K. Matsuda: Title CLCA2 as a p53-inducible senescence mediator. Neoplasia, in press, 2012

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
Assistant Professor	Ashwini Patil, Ph.D.	助　教　理学博士	パティル，アシュウィニ

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 under which 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 interactions.

1. Computational model for deciphering transcription of co-expressed genes

Yosvany Lopez and Kenta Nakai

The understanding of the mechanisms of transcriptional regulation remains a great challenge for molecular biologists in the post-genome era. At the transcriptional level, DNA-binding proteins (transcription factors) modulate the expression of genes by binding to their specific DNA regulatory elements in nearby genomic regions. Nowadays, the identification and characterization of these components is valuable because the presence or absence of transcription factor binding sites (TFBSs) is thought to be responsible for the complexity of gene regulation in every living organism. Based on the assumption that the regulatory regions of (at least a part of) those genes showing similar expression profiles should share some common structural characteristics, we are attempting to design a computational model capable of explaining how the binding of transcription factors is carried out in promoter regions of co-regulated genes in specific biological tissues. We are working on a database of co-expressed genes in *Arabidopsis thaliana* (ATTED-II) because, among other reasons, the intergenic regions of plant genes are smaller than that of higher organisms. Working on *A. thaliana* genes will facilitate our initial analysis, but once a good enough model has been achieved, it will be applied to different human tissues to shed light on the transcriptional processes behind co-regulated genes.

2. TSS-Seq time-series analysis of mice dendritic cells after LPS stimulation

Kuo-Ching Liang, Yutaro Kumagai[1], Shizuo Akira[1], and K. Nakai: [1]WPI-iFReC, Osaka University.

Dendritic cells act as intermediary between external environment and mammalian adaptive immunity mechanism by presenting foreign antigens to various types of lymphocytes. The dynamic changes in the complex web of control and interaction relationships in the dendritic cells due to immune responses are therefore of great interest in the understanding of mammalian immune system. In this work, we attempt to elucidate and characterize the temporal behavior of transcription start sites (TSS) and gene expression in mouse dendritic cell after eliciting immune response by lipopolysaccharide (LPS)

stimulation, we analyze the TSS-Seq data for mouse dendritic cell samples induced from bone-marrow cells with the existence of GM-CSF, and collected at 0hr, 0.5hr, 1hr, 2hrs, 3hrs, 4hrs, 6hrs, 8hrs, 16hrs and 24hrs from LPS stimulation. In addition to looking for changes in the TSS distribution patterns of the dendritic cell genome across the time steps, the TSS-tags for each gene are also collected as a digital count of gene expression for time-series analysis. After the appropriate normalization and re-sampling steps, genes are clustered based on their gene expression time-series, and compared to gene-ontology annotations and known protein and genomic networks to discover any functional over-representation. Cross correlation and convolutional analysis are also used to discover possible time-delayed gene regulatory relationship present in the data. Furthermore, we explore the expression time-series to discover any cyclic behavior.

3. Construction of DBTSS version 8 with epigenomic information

Riu Yamashita[2], Yutaka Suzuki[3], Sumio Sugano[3], Kenta Nakai: [2]Frontier Research Initiative, IMSUT, [3]Grad. Sch. Frontier Sciences, U. Tokyo.

We have constructed the DBTSS (DataBase of Transcriptional Start Sites), which represents exact positions of transcriptional start sites (TSSs) in the genome based on our unique experimentally validated TSS sequencing method, TSS-Seq. In this update, we included new TSS data, so that a major part of human adult and embryonic tissues are covered. DBTSS now contains 491 million TSS tag sequences for collected from a total of 20 tissues and 7 cell cultures. We also integrated our newly generated RNA-seq data of subcellular- fractionated RNAs and ChIP-Seq data of histone modifications, RNA polymerase II and several transcriptional regulatory factors in cultured cell lines. We also included recently accumulating external epigenomic data, such as chromatin map of the ENCODE project. We further associated this TSS information with public and original SNV data, in order to indentify single nucleotide variations (SNVs) in the regulatory regions. This data can be browsed in our new viewer which also supports versatile search conditions for users (http://dbtss.hgc.jp).

4. Massive-scale RNA-Seq analysis of mouse early embryo development

Sung-Joon Park, Makiko Komata[4], Fukashi Inoue[5], Kaori Yamada[5], Kenta Nakai, Miho Ohsugi[5], Katsuhiko Shirahige[4]: [4]IMCB, U. Tokyo, [5]Division of Oncology.

To understand the mechanism of biological events observed during the early mammalian embryo development, we performed transcriptome analysis on mouse oocyte, one-cell, two-cell, and four-cell embryonic stages using approximately 10,000 high-quality cells per stage, which is an unprecedented experimental scale. Over 100 million short reads sequenced by SOLiD were analyzed by Tophat and Cufflinks software coupled with a recursive mapping strategy, where unmapped reads are truncated and mapped again. Consequently, we profiled over 16,000 RefSeq coding transcripts and 37 candidates of novel genes. The profile covered almost all genes initially assayed by microarray and detected more 7,718 genes, supporting the advance of RNA-Seq assay. By clustering the mRNA abundance at each stage, we found 21 gene expression patterns. The patterns are linked to the well-known expression transitions, such as the degradation of maternal mRNAs, minor or major zygotic expression, and stage-specific transient expression. Remarkably, we discovered four novel patterns that exhibit an acute oscillatory transition of activation and repression. The analysis of gene ontology enrichment for the novel patterns suggested their importance in the development. In addition, we observed that particular ncRNAs, such as snoRNA and RNase MRP RNA, have been increasingly pooled after fertilizing. It implies that the gene regulation program switches on the gene expression required for the appropriate development and the regulation is accompanied by ncRNA activity. The transcriptome profile we established here is an important resource to help enhance our understanding of mammalian embryo development.

5. Comprehensive analysis of transcription initiation patterns, promoter architecture and function in *Ciona intestinalis*

Rui Yokomori, Kotaro Shimai[6], Koki Nishitsuji[7], Yutaka Suzuki[3], Takehiro Kusakabe[6] and Kenta Nakai: [6]Fac. Sci. and Eng., Konan U., [7]Grad. Sch. Life Sci., U. Hyogo.

Transcription is known to start from multiple positions and the distribution of transcription start sites (TSSs) can be classified into three types: NP (Narrow with Peak), BP (Broad with Peak) and WP (Weak Peak). However, the biological meaning of each type is poorly understood. In this study, we identified promoters across the *Ciona intestinalis* genome and ana-

lyzed their TSS distributions (transcription initiation patterns), promoter architectures and functions using 8 different tissues. As a result, we found that TATA-box was significantly enriched in NP type compared to WP type, although initiator motif did not show the difference across each type. Gene Ontology enrichment analysis showed that the transcripts with NP, BP and WP are associated with development, translation and transport, respectively. Also, our analysis suggested current annotation of transcription start sites is incomplete and identified 40 novel candidate promoters.

6. Analyses of relationship between Y14 RNAi knockdown and occurrences of irregular splicing

Shunichi Wakabayashi, Kazuhiro Fukumura[8], Masami Shiimori[8], Kenta Nakai, Kunio Inoue[8], and Hiroshi Sakamoto[8]: [8]Gradu. Sch. Sci., Kobe U.

RNA-binding protein Y14 (also known as RBM8A) is a member of the protein complex EJC (exon junction complex) which influences mRNA splicing and their transport to the cytoplasm. Under the condition of Y14 knockdown, it was reported that some mRNAs retained introns and occurrences of irregular splicing in cytoplasm were observed. Here we analyzed functions of Y14 for mRNA processing in worm and human cells. We identified some significantly retained introns (e.g. 6th intron of TALDO1, 4th intron of TMEM147) by Y14 RNAi using computational RNA-seq data analyses and experimental confirmations using RT-PCR. We analyzed relationships between occurrences of irregular splicing and attributes of introns like length or position in respective genes.

7. Non-repetitive protein regions encoded by nucleotide repeats have a significant impact on protein structure

Toshiyuki Tsuji, Ashwini Patil and Kenta Nakai

Nucleotide repeats in coding regions significantly affect the structure and function of proteins. While the structural properties of the resulting amino acid repeats have been widely studied, those of the encoded non-repetitive regions have so far not been investigated. In this study, we compared the structural features of non-repetitive protein regions encoded by nucleotide repeats with those of amino acid repeats and non-repeat fragments. Similar to amino acid repeat regions, non-repetitive protein regions resulting from nucleotide repeats show higher levels of intrinsic disorder and structural flexibility compared to non-repeat regions. Their amino acid propensity is also similar to that of amino acid repeat regions. However, their patterns of amino acid conservation and their greater prevalence in the interface regions, as well as their ability to form stable secondary structures bear a greater similarity to non-repeat regions. We conclude that non-repetitive protein regions encoded by nucleotide repeats are distinct from amino acid repeat and non-repeat regions, but retain certain structural properties of both. We propose that their unique combination of structural characteristics enables them to play an important role in protein function.

8. Functional annotation of proteins with intrinsically disordered regions using amino acid content similarity

Ashwini Patil, Shunsuke Teraguchi[9], Huy Dinh[9], Daron Standley[9] and Kenta Nakai: [9]Inst. Protein Res., Osaka U.

Intrinsically disordered regions in proteins are regions without a stable tertiary structure. Despite the lack of a stable structure, these regions play an important role in protein function as a result of their flexibility and adaptability. Intrinsically disordered regions are found in proteins functioning in cell signaling and transcription regulation. However, several such regions are not associated with any function and are often ignored during the functional annotation of proteins. Function prediction of proteins with large disordered regions is difficult using conventional techniques of sequence similarity because these regions evolve rapidly and are often of low complexity. In this project, we developed a web server called the IDD Navigator that, given a disordered region, identifies similar disordered regions using amino acid content and predicts the function of the query sequence using Gene Ontology terms enriched in the hit sequences. We found that this method performs better than random at predicting associated Gene Ontology terms with given disordered regions.

9. Assessing the utility of a gene co-expression stability measure in the study of protein-protein interaction networks.

Ashwini Patil, Kengo Kinoshita[10] and Kenta Nakai: [10]Tohoku U.

We assessed the utility of gene co-expression

stability, a means of measuring the bias in gene co-expression, as an additional measure to support the co-expression correlation in the analysis of protein-protein interaction networks. We studied the patterns of co-expression correlation and stability in interacting proteins with respect to their interaction promiscuity, levels of intrinsic disorder, and essentiality or disease-relatedness. Co-expression stability, along with co-expression correlation, acts as a better classifier of hub proteins in interaction networks, than co-expression correlation alone, enabling the identification of a class of hubs that are functionally distinct from the widely accepted transient (date) and obligate (party) hubs. Proteins with high levels of intrinsic disorder have low co-expression correlation and high stability with their interaction partners suggesting their involvement in transient interactions, except for a small group that have high co-expression correlation and are typically subunits of stable complexes. Similar behavior was seen for disease-related and essential genes. Interacting proteins that are both disordered have higher co -expression stability than ordered protein pairs. Using co-expression correlation and stability, we found that transient interactions are more likely to occur between an ordered and a disordered protein while obligate interactions primarily occur between proteins that are either both ordered, or disordered. We observed that co-expression stability shows distinct patterns in structurally and functionally different groups of proteins and interactions.

10. A large-scale study of the conservation and classification of intrinsically disordered regions

Ashwini Patil, Harry Amri Moesa, Shunichi Wakabayashi and Kenta Nakai

Despite the importance of intrinsically disordered regions in proteins, there is currently no classification system available for these regions. We studied the levels of conservation of known and predicted disordered regions in eukaryotes and proposed a method to classify them based on their conservation and their residue content i. e. amount of charge and hydropathy. We found that highly conserved disordered regions can be classified into distinct groups based on charge and hydropathy which are associated with distinct functions suggesting a possible functional classification of disordered regions. We also found that disordered regions showing low conservation often show high residue type content conservation (i.e. similar fraction of charged, polar and hydrophobic residues), which may be used to maintain their disorderliness in order to stay functional despite high rates of evolution.

11. Project for accelerating the clinical application of regenerative medicine technologies

Kenta Nakai, Tatsutoshi Nakahata[11], Norio Nakatsuji[12], Kohji Nishida[13], Hideyuki Okano[14], Masayo Takahashi[15], Akihiro Umezawa[16], Masayuki Yamato[17], and Shinya Yamanaka[11]: [11]CiRA, Kyoto U., [12]IFMS, Kyoto U., [13]Grad. Sch. Med., Osaka U., [14]Sch. Med. Keio U., [15]CDB, RIKEN, [16]Nat. Res. Inst. Child Health and Development, [17]Tokyo Women's Med. U.

This project aims to achieve the earliest possible clinical application of regenerative medicine technologies using safe, effective and high-quality human stem cells by building a collaborative platform among researchers through a network centric research infrastructure. Also in the project, the basis of an "open innovation" environment allowing research institutes to continuously create innovative technologies will be developed. To support this activity, an advisory committee has been formed with members from various sectors.

12. Construction of a system for analyzing the clonality of retroviruses using next generation sequencer

Sakura Aoki[3], Yutaka Suzuki[3], Kenta Nakai, and Toshiki Watanabe[3]

Human T-cell Leukemia Virus type-1 (HTLV-1) is the causative agent of Adult T-cell Leukemia (ATL). While most of the HTLV-1 infected individuals remain as asymptomatic carriers (ACs) throughout their lifetime, about 5% of them develop ATL. However, the mechanism responsible for such a significant risk variation among ACs remains to be elucidated. ACs harbor polyclonal population of HTLV-1 infected cells, whereas ATL patients show mono-clonal expansion. Thus we postulate that the onset of a clonal expansion of HTLV-1-infected cells can be a reliable risk indicator of progression into ATL among ACs. Although it has not yet been accomplished by the conventional methods, tracking the behavior of each clone in the complex population of HTLV-1 infected cells is essential to understand the clonality of HTLV-1 infected cells. In the present study, taking advantage of next-generation sequencing technology and nested-splinkerette PCR, a new high-throughput method has been developed. We expect that this technique will enable us to track the changes in

the clonality during the course of ATL development. Data obtained from ATL patients have been compared with those from ACs, including the "high-risk carriers" who subsequently developed ATL. The resulting information will not only provide us with a useful method to predict the probability of ATL onset among ACs, but also a new insight into the natural course of the disease.

Publications

Patil, A., Nakai, K., and Nakamura, H. HitPredict: a database of quality assessed protein-protein interactions in nine species, *Nucl. Acids Res*., 39, D744-D799 (2011)

Park, S.-J., and Nakai, K. A regression analysis of gene expression in ES cells reveals two gene classes that are significantly different in epigenetic patterns. BMC Bioinformatics. 12 (Suppl 1): S50, 2011.

Khare, P., Mortimer, SI., Cleto, CL,. Okamura, K., Suzuki, Y., Kusakabe, T., Nakai, K., Meedel, TH., and Hastings, KEM. Cross-validated methods for promoter/transcription start site mapping in SL trans-spliced genes, established using the *Ciona intestinalis* troponin I gene. Nucleic Acids Res. 39(7): 2638-2648 (2011).

Yamashita, R., Sathira, N.P., Kanai, A., Tanimoto, K., Arauchi, T., Tanaka, Y., Hashimoto, S., Sugano, S., Nakai, K., and Suzuki, Y. Genome-wide characterization of transcription start sites in humans by integrative transcriptome analysis. Genome Res. 21(5): 775-789 (2011).

Irie, T., Park, S.J., Yamashita, R., Seki, M., Yada, T., Sugano, S., Nakai, K., and Suzuki, Y. Predicting promoter activities of primary human DNA sequences. Nucl. Acids Res. 39(11): e75 (2011).

Ohshima, D., Qin, J., Konnno, H., Hirosawa, A., Shiraishi, T., Yanai, H., Shimo, Y., Akiyama, N., Yamashita, R., Nakai, K., and Inoue, J. RANK signaling induces interferon-stimulated genes in the fetal thymic stroma. Biochem. Biophys. Res. Comm. 408(4): 530-536, 2011.

Kimura, K., Koike, A., and Nakai, K. Seed-set construction by equi-entropy partitioning for efficient and sensitive short-read mapping. Algorithms in Bioinformatics in (T.M. Przytycka and M.-F. Sagot ed.) Lecture Notes in Computer Science. 6833: 151-162, 2011. (ISBN: 978-3-642-23037-0).

Okamura, K., Yamashita, R., Takimoto, N. Nishitsuji, K., Suzuki, Y., Kusakabe, T.G., and Nakai, K. Profiling ascidian promoters as the primodial type of vertebrate promoter. BMC Genomics 12(Suppl. 3): S7, 2011. BEST PAPER AWARD in InCoB211

Patil, A., Nakai, K., and Kinoshita, K. Assessing the utility of gene co-expression stability in combination with correlation in the analysis of protein-protein interaction networks. BMC Genomics 12(Suppl. 3): S19, 2011.

Patil, A., Teraguchi, S., Dinh, H., Nakai, K., and Standley, DM. Functional annotation of intrinsically disordered domains by their amino acid content using IDD Navigator. Pacific Symposium on Biocomputing 17: 164-175, 2012.

Yamashita, R., Sugano, S., Suzuki, Y., and Nakai, K. DBTSS: database of transcriptional start sites progress report in 2012. Nucl. Acids Res. 40(Database Issue): D150-154, 2012.

Kimura, K., Koike, A., and Nakai, K. A Bit-parallel dynamic programming algorithm suitable for DNA sequence alignment. J. Bioinformatics and Computational Biology. In press

Makita, Y. and Nakai, K. Bacillus subtilis transcriptional network. In (Babu, M. ed.) Bacterial Gene Regulation and Transcriptional Networks, Horizon Sci. Press., in press.

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
Assistant Professor	Yusuke Inoue, Ph.D.		助 教	社会医学博士	井 上 悠 輔
Project Assistant Professor	Hyunsoo 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

The Department of Public Policy works to achieve three major missions: public policy studies of translational research, its application, and its 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. By conducting qualitative and quantitative social science study and policy analysis, we facilitate discussion of challenges arising from advances in medical sciences. Furthermore, we study specific ethics issues related to construction of a human biological substances collection, and related to vaccination policy.

1. Research ethics consultation for multi-center cancer genome research

Public interest in research ethics has grown. Society increasingly makes demands in this area. Provision of a system that can support "on-site" researchers to avail themselves of immediate consultation when concerns and issues arise related to research ethics and other matters is also among those demands. Based on these demands, in recent years, the universities and research institutes providing "research ethics consultation" have become increasingly numerous in the United States. We have been commissioned by the government to provide research ethics consultation to several big projects promoting medical sciences.

We have started to provide research ethics consultation for a new and comprehensive cancer genome research project for 34 clinical seeds of 64 designated institutions, called as Project for Development of Innovative Research on Cancer Therapeutics (MEXT). We have addressed a basic research ethics policy to use stored samples for research. We have also developed a research ethics management system on the website.

2. Biobank Japan Project (BBJP) and its ethical, legal and social implications

The Biobank Japan Project (BBJP) is a disease-focused biobanking project headed by Professor Yusuke Nakamura since 2003. Biobank Japan consists of donated DNA, sera and clinical information from 200,000 patients of 66 hospitals in Japan. Informed consent, which ensures the autonomous decisions of participants, is believed to be practically impossible for the biobanking project in general.

We have conducted interview studies of research coordinators ($n=50$) since 2010 and we have just compiled our self-evaluation report on informed consent process of this project. As a means to maintain the participants' trust of the project, research coordinators who had been specially trained for the BBJP have played important roles. We have worked steadily to complete analyses of the research coordinators of the BBJP. At the beginning of the BBJP, their pri-

mary roles were recruitment. After the end of recruitment, their roles shifted to the tracing of participants to extract clinical information and to input it into the database. However, their support and encouragement of participants complemented the contents of the initial consent process and reinforced participants' incentive to continue in their role. The results of this study will contribute to improved quality control and better communication between administrators of long-term research projects and the project participants.

We also have provided research ethics consultation for conducting a follow-up study without consent, to obtain permission from 1,100 municipalities to access cause of death of 44,698 participants.

3. Rethinking voluntary egg donation

In the scandal around Korean stem cell scientist Woo-Suk Hwang, the inappropriate collection of human eggs as research material, fabricated data on ES cells obtained through somatic cell nuclear transfer, and fraudulent fundraising were condemned as legal and ethical transgressions. Among the criticisms, the donation of eggs by many women became a big issue. Some of the women were motivated by financial compensation or in-kind support, while others decided to donate their eggs without payment, being convinced that the research would bring therapy for thus far incurable patients, a promise unfulfilled. Regardless of the multiple reports published to articulate why the Hwang scandal happened in South Korea, we realized during our ethnographical fieldwork in that country that it would be meaningful to consider the ethical issues in a global context. In this paper (#3), we focus on the motivations of the South Korean women who donated their eggs voluntarily as research materials, and aim to understand it in a more general context. We point out that not only their love of family but also other altruistic motivations for donating eggs are affected by the attitudes revealed in their narratives. Finally, we argue that there is a serious bioethical issue when a social environment of sick or disabled people makes women decide to help these individuals by donating eggs.

4. Direct-to-consumer genetic testing for talent identification

The regulation of direct-to-consumer genetic testing has till now focused on identifying predispositions to specific health problems and quality issues in testing, such as analytical and clinical validity criteria. In the United States, for example, the Food and Drug Administration warned that federal regulations for medical devices will apply to commercial genetic tests for health purposes. By contrast, the regulation of commercial genetic testing for other purposes is rarely discussed. These tests are advertised as identifying intelligence, athletic ability, and artistic sensibility in children. One company offers testing for a "gold medalist gene" purportedly associated with remarkable athletic talent; in Japan, the test has been favorably featured on some television programs and is recruiting customers through the Internet. We have alerted ethical issues among these sales in our paper (# 4).

5. A cross-sectional survey of physicians' recommendations of the 2009 influenza A/H1N1 pandemic in Japan

Striking a balance between the rapid availability of a novel vaccine while ensuring its safety, quality, and efficacy is a major challenge during a pandemic. In our paper (#5), we aimed to elucidate physicians' attitudes regarding the novel vaccine during the influenza A/H1N1 pandemic of 2009, and to determine factors that affected their vaccination recommendations to patients. Of a random sample of 1,000 general practitioners (GPs) in Japan, 515 participated in the cross-sectional anonymous survey conducted just before the novel vaccine was available (between 28 September and 18 October 2009). A total of 453 GPs (88.3%) replied that they intended to receive the new vaccine themselves; however, only 177 GPs (34.6%) intended to proactively recommend it to their patients. The anticipated cost of the vaccine negatively influenced the intention to vaccinate themselves and their recommendations to patients ($P<0.001$, $\chi(2)$ test). Results of multivariate logistic regression analysis showed that physicians with experience in influenza A/H1N1 patient contacts [1-20 contacts, odds ratio (OR)=7.49 (95% confidence interval [CI]: 1.73-32.36), $P=0.007$; >20 contacts, OR=8.03 (95% CI: 1.77-36.50), $P=0.007$, compared with no contacts] were more likely to recommend the vaccine to patients, whereas those with knowledge of the fear on the causal association between Guillain-Barré syndrome (GBS) cases and the 1976 swine flu vaccination in the USA were less likely to recommend the vaccine [OR=0.66 (95% CI: 0.45-0.97), $P=0.036$].

6. The agendas for discussion on human and animal chimera for scientific research

In July 2010, the Minister of Education, Cul-

ture, Sports, Science, and Technology received the first notification of the procedure for creating human-to-animal chimeric embryos by inserting human induced pluripotent stem cells (iPS cells) to animal blastocysts. This case drew attention to the necessity of reviewing the regulations of human-to-animal chimeric embryos under the Act on Regulation of Human Cloning Techniques and the Guidelines for Specified Embryo. In Kamisato's paper (#6, in Japanese), she has succeeded to organize the concept of "human and animal chimera"-including human-to-animal chimera and animal-to-human chimera-which eventually became quite complicated. Subsequently, she has discussed the problematic issues of the existing regulations and the agendas for reviewing the regulations of creating and using human-to-animal chimeric embryos. As a result, she has clearly listed four problems of the existing regulations: for example, the duration that the human-to-animal chimeric embryos can be maintained is lacking validity. In addition, she has proposed the discussion for human and animal chimera with wider vision which includes the scientific, ethical, legal, and sociological aspect. In the end, she has mentioned three important points that need to be discussed including what kind of human and animal chimera can be allowed to create.

7. Research in progress

We have been conducting other studies as described below.

- Ethical, legal and social implications of commercial genetic/genomic testing services in eastern 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 for building trust from participants
- Ethical, legal and social implications of 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
- Ethics issues in collecting human biological substances for constructing a research infrastructure
- Vaccination policy

Publications

1. Ishiyama I, Tanzawa T, Watanabe M, Maeda T, Muto K, Tamakoshi A, Nagai A, and Yamagata Z. Public attitudes to the promotion of genomic crop studies in Japan: correlations between genomic literacy, trust, and favourable attitude. Public Understanding of Science, first published on January 17, 2012
2. Suzuki K, Sawa R, Muto K, Kosuda S, Banno K, Yamagata Z. Risk perception of pregnancy promotes disapproval of gestational surrogacy: Analysis of a nationally representative opinion survey in Japan, International Journal of Fertility and Sterility, 5(2): 78-85, 2011.
3. Tsuge A, Hong H. Reconsidering ethical issues about "voluntary egg donors" in Hwang's case in global context, New Genetics and Society. 30 (3): 241-252, 2011.
4. Inoue Y, Muto K. Children and the genetic identification of talent. The Hastings Center Report, 41, inside back cover, 2011.
5. Inoue Y, Matsui K. Physicians' recommendations to their patients concerning a novel vaccine: a cross-sectional survey on 2009 A/H1N1 vaccination in Japan. Environmental Health and Preventive Medicine, 16 (5): 320-326, 2011.
6. 神里彩子．ヒトと動物のキメラを作成する研究はどこまで認められるか？―再議論に向けた検討課題の提示―，生命倫理　21(1)：22–32，2011．
7. 武藤香織．難病をもつ地域住民への支援～市町村の役割再考．月刊「自治研」2011年７月号，19–26，2011．
8. 井上悠輔．『臨床研究のための倫理審査ハンドブック』(笹栗俊之ほか編)，(1–4–2, 2–4–1分担)，丸善，2011年．