Human Genome Center

Laboratory of DNA Information Analysis
Laboratory of Sequence Data Analysis
Laboratory of Genome Database

DNA情報解析分野
シークエンスデータ情報処理分野
ゲノムデータベース分野

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	Yuichi Shiraishi, Ph.D.	特任助教	博士(統計科学)	白　石　友	一
Associate Professor	Tetsuo Shibuya, Ph.D.	准教授	博士(理学)	渋　谷　哲	朗
Senior Assistant Professor	Rui Yamaguchi, Ph.D.	講　師	博士(理学)	山　口	類

We are facing with biomedical big data comprising of ultra-high dimensional ultra-heterogeneous data. Our current mission is to develop computational/informatics strategy for medical informatics to implement personalized genomic medicine through genomics, systems biology and supercomputer.

1. Systems Cancer Research and Systems Biology

a. An empirical Bayesian framework for somatic mutation detection from cancer

Shiraishi Y, Sato Y[1], Chiba K, Okuno Y[1], Nagata Y[1], Yoshida K[1], Shiba N[1], Hayashi Y[2], Kume H[1], Homma Y[1], Sanada M[1,7], Ogawa S[1,7], Miyano S: [1]Graduate School of Medicine, The University of Tokyo, [2]Gunma Children's Medical Center

High-throughput sequencing technologies have enabled comprehensive dissections of the cancer genome clarifying a large number of somatic mutations in a wide variety of cancer types. Various methods have been suggested for mutation calling based on a large amount of sequencing data, which is accomplished in most cases by tactical statistically evaluating the differences in the observed allele frequencies of possible single nucleotide variants between tumours and paired normal samples. However, more accurate detection of mutations challenge for data with low sequencing depths or tumour contents. To overcome this problem, we devised a novel statistical method, Empirical Bayesian mutation Calling (https://github.com/friend1ws/EBCall), for detecting somatic mutations. Ogawa

Lab contributed to the experimental validation of our method. Unlike previous methods, our method discriminates somatic mutations from sequencing errors based on an empirical Bayesian framework, where the model parameters are estimated using sequencing data from multiple non-paired normal samples. Using 13 whole-exome sequencing data with 87.5-206.3 mean sequencing depths, we demonstrate that our method not only outperforms several existing methods in the calling of mutations with moderate allele frequencies but also enables accurate calling of mutations with low allele frequencies ($\leq 10\%$) harboured within a minor tumour subpopulation, thus allowing for the deciphering of fine substructures within a tumour specimen.

b. Multi-omics approach for estimating metabolic networks using low-order partial correlations

Kayano M[3], Imoto S, Yamaguchi R, Miyano S: [3]Obihiro University of Agriculture and Veterinary Medicine

Kayano is a member of Systems Cancer Research and this is a result of this project. There are two typical purposes of metabolome analysis. One is to estimate metabolic pathways and the other is to understand the regulatory systems underlying the metabolism. As the source of information for these analyses, we assume a set of multi-omics data for RNA, proteins, and metabolites. However, integrated methods that analyze multi-omics data simultaneously and unravel the systems behind metabolisms have not been well established. Our statistics-based method is the first successful method whose idea is the use of low-order partial correlations with a robust correlation coefficient for estimating metabolic networks from metabolome, proteome, and transcriptome data. Our method is defined by the maximum of low-order, particularly first-order, partial correlations (MF-PCor) in order to assign a correct edge with the highest correlation and to detect the factors that strongly affect the correlation coefficient. First, through numerical experiments with real and synthetic data, we showed that the use of protein and transcript data of enzymes improved the accuracy of the estimated metabolic networks in MF-PCor. In these experiments, the effectiveness of the proposed method was also demonstrated by comparison with a correlation network (Cor) and a Gaussian graphical model (GGM). Our theoretical investigation confirmed that the performance of MF-PCor could be superior to that of the competing methods. In addition, in the real data analysis, we investigated the role of metabolites, enzymes, and enzyme genes that were identified as important factors in the network established by MF-PCor. We then found that some of them corresponded to specific reactions between metabolites mediated by catalytic enzymes that were difficult to be identified by analysis based on metabolite data alone.

c. Multilayer cluster heat map visualizing biological tensor data

Niida A, Tremmel G, Imoto S, Miyano S

Biological visualization is getting more and more important since high-throughput technologies are generating a broad range of omics data. Facing a torrent of massive biological data, visual data mining can be considered an intuitive and powerful approach for hypothesis generation. The cluster heat map approach has been popularly used to visualize the matrix types of biological data. We extended the use of the cluster heat map to reveal informative patterns hidden in third-order tensor-type biological data. By applying the extended method, a multilayer cluster heat map, to trans-omics and network tensor data, we successfully demonstrated the proof-of-concept of our approach. Our new visual data mining method will be a useful tool for increasing the amount of biological tensor data, especially, cancer data. The software and the tensor data studied are available from http://www.hgc.jp/~niiyan/MCHM.

d. The tumor-suppressive miR-497-195 cluster targets multiple cell-cycle regulators in hepatocellular carcinoma

Furuta M[4], Kozaki K[4], Tanimoto K[4], Tanaka S[4], Arii S[4], Shimamura T, Niida A, Miyano S, Inazawa J[4]: [4]Tokyo Medical and Dental University

We contributed to this study for systems biology approach and its data analysis. MicroRNAs are known to play key post-transcriptional regulators of gene expression and commonly deregulated in carcinogenesis. The aim of this study is to explore functionally crucial tumor-suppressive (TS)-miRNAs in hepatocellular carcinoma (HCC). We designed and performed integrative function- and expression-based screenings of TS-miRNAs in six HCC cell lines. The screenings identified seven miRNAs, which showed growth-suppressive activities through the overexpression of each miRNA and were endogenously downregulated in HCC cell lines. Further expression analyses using a large panel of HCC cell lines and primary tumors demonstrated four miRNAs, miR-101, -195, -378 and -497, as candidate TS-miRNAs frequently silenced in HCCs. Among them, two clustered miRNAs miR-195 and miR-497 showed significant growth-suppressive activity with induction of G1 arrest.

Comprehensive exploration of their targets using Argonute 2 - immunoprecipitation - deep - sequencing (Ago2-IP-seq) and genome-wide expression profiling after their overexpression followed by pathway analysis, revealed a significant enrichment of cell cycle regulators. Among the candidates, we successfully identified CCNE1, CDC25A, CCND3, CDK4, and BTRC as direct targets for miR-497 and miR-195. Moreover, target genes frequently upregulated in HCC in a tumor-specific manner, such as CDK6, CCNE1, CDC25A and CDK4, showed an inverse correlation in the expression of miR-195 and miR-497, and their targets. These results suggest the molecular pathway regulating cell cycle progression to be integrally altered by downregulation of miR-195 and miR-497 expression, leading to the aberrant cell proliferation in hepatocarcinogenesis.

e. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms

Kon A[1], Shih LY[5], Minamino M[6], Sanada M[7], Shiraishi Y, Nagata Y[1], Yoshida K[1], Okuno Y[1], Bando M[6], Nakato R[6], Ishikawa S[4], Sato-Otsubo A[1], Nagae G[8], Nishimoto A, Haferlach C[9], Nowak D[10], Sato Y[1], Alpermann T[9], Nagasaki M[11], Shimamura T, Tanaka H, Chiba K, Yamamoto R[12], Yamaguchi T[12], Otsu M[13], Obara N[14], Sakata-Yanagimoto M[14], Nakamaki T[15], Ishiyama K[16], Nolte F[10], Hofmann WK[10], Miyawaki S[16], Chiba S[14], Mori H[15], Nakauchi H[12], Koeffler HP[17], Aburatani H[8], Haferlach T[9], Shirahige K[6], Miyano S, Ogawa S[7]: [5]Chang Gung University, [6]Institute of Molecular and Cellular Biosciences, The University of Tokyo, [7]Graduate School of Medicine, Kyoto University, [8]Research Center for Advanced Science and Technology, The University of Tokyo, [9]Munich Leukemia Laboratory, [10]University Hospital Mannheim, [11]Tohoku Medical Megabank Organization, Tohoku University, [12]Division of Stem Cell Therapy, Institute of Medical Science, [13]Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, [14]Faculty of Medicine, University of Tsukuba, [15]Showa University School of Medicine, [16]Tokyo Metropolitan Ohtsuka Hospital, [17]Cedars-Sinai Medical Center

We contributed to this study for improving the NGS data analysis pipeline and large-scale data analysis for detecting mutations and deletions using the supercomputer of Human Genome Center. Ogawa Lab took the initiative of this study. Cohesin is a multimeric protein complex that is involved in the cohesion of sister chromatids, post-replicative DNA repair and transcriptional regulation. We found recurrent mutations and deletions involving multiple components of the cohesin complex, including STAG2, RAD21, SMC1A and SMC3, in different myeloid neoplasms. These mutations and deletions were mostly mutually exclusive and occurred in 12.1% (19/157) of acute myeloid leukemia, 8.0% (18/224) of myelodysplastic syndromes, 10.2% (9/88) of chronic myelomonocytic leukemia, 6.3% (4/64) of chronic myelogenous leukemia and 1.3% (1/77) of classical myeloproliferative neoplasms. Cohesin-mutated leukemic cells showed reduced amounts of chromatin-bound cohesin components, suggesting a substantial loss of cohesin binding sites on chromatin. The growth of leukemic cell lines harboring a mutation in RAD21 (Kasumi-1 cells) or having severely reduced expression of RAD21 and STAG2 (MOLM-13 cells) was suppressed by forced expression of wild-type RAD21 and wild-type RAD21 and STAG2, respectively. These findings suggest a role for compromised cohesin functions in myeloid leukemogenesis.

f. Integrated molecular analysis of clear-cell renal cell carcinoma

Sato Y[1], Yoshizato T[1], Shiraishi Y, Maekawa S[1], Okuno Y[1], Kamura T[18], Shimamura T, Sato-Otsubo A[1], Nagae G[8], Suzuki H[1], Nagata Y[1], Yoshida K[1], Kon A[1], Suzuki Y[19], Chiba K, Tanaka H, Niida A, Fujimoto A[20], Tsunoda T[20], Morikawa T[1], Maeda D[1], Kume H, Sugano S[19], Fukayama M[1], Aburatani H[8], Sanada M[1,7], Miyano S, Homma Y[1], Ogawa S[1,7]: [18]Graduate School of Science, Nagoya University, [19]Graduate School of Frontier Sciences, The University of Tokyo, [20]Center for Genomic Medicine, RIKEN

We contributed to this study for improving the NGS data analysis pipeline, development of data analysis method and large-scale omics-data analysis by using the supercomputer of Human Genome Center. Ogawa Lab coordinated this study. Clear-cell renal cell carcinoma (ccRCC) is the most prevalent kidney cancer and its molecular pathogenesis is incompletely understood. Here we report an integrated molecular study of ccRCC in which $\geq$100 ccRCC cases were fully analyzed by whole-genome and/or whole-exome and RNA sequencing as well as by array-based gene expression, copy number and/or methylation analyses. We identified a full spectrum of genetic lesions and analyzed gene expression and DNA methylation signatures and determined their impact on tumor behavior. Defective VHL-mediated proteolysis was a common feature of ccRCC, which was caused not only by VHL inactivation but also by new hotspot TCEB1 mutations, which abolished Elongin C-VHL binding, leading to HIF accumulation. Other newly identified pathways and components recurrently mutated in ccRCC included PI3K-AKT-mTOR signaling, the KEAP1-NRF 2-CUL3 apparatus, DNA methylation, p53-related pathways and mRNA processing. This integrated molecular analysis unmasked new correlations be-

tween DNA methylation, gene mutation and/or gene expression and copy number profiles, enabling the stratification of clinical risks for patients with ccRCC.

g. A restricted level of PQBP1 is needed for the best longevity of Drosophila

Tamura T[4], Sone M[26], Nakamura Y[4], Shimamura T, Imoto S, Miyano S, Okazawa H[4]: [26]Department of Biomolecular Science, Faculty of Science, Toho University

A number of neurological diseases are caused by mutations of RNA metabolism-related genes. A complicating issue is that whether under- or overfunction of such genes is responsible for the phenotype. Polyglutamine tract binding protein-1, a causative gene for X-linked mental retardation, is also involved in RNA metabolism, and both mutation and duplication of the gene were reported in human patients. In this study, we first report a novel phenotype of dPQBP1 (drosophila homolog of Polyglutamine tract binding protein-1)-mutant flies, lifespan shortening. We next address the gene dose-phenotype relationship in lifespan shortening and in learning disability, a previously described phenotype. The 2 phenotypes are rescued by dPQBP1 but in different dose-phenotype relationships. Either insufficient or excessive expression of dPQBP1 does not recover lifespan, while excessive expression recovers learning ability. We finally address the mechanism of lifespan shortening. Tissue-specific expression of dPQBP1-RNA interference construct reveals both neural and nonneural dPQBP 1 contribute to the lifespan, while the latter has a dominant effect. Gene expression profiling suggested retinophilin/MORN repeat containing 4, a gene promoting axonal degeneration, to contribute to lifespan shortening by neural dPQBP1. Systems biology analysis of the gene expression profiles revealed indirect influence of dPQBP1 on insulin-like growth factor 1, insulin receptor, and peroxisome proliferator-activated receptorα/γ signaling pathways in nonneural tissues. Collectively, given that dPQBP1 affects multiple pathways in different dose-dependent and tissue-specific manners, dPQBP1 at a restricted expression level is needed for the best longevity.

2. Statistical/Algorithmic Data Analysis Methods

a. Principal component analysis using QR decomposition

Sharma A, Paliwal KK[21], Imoto S, Miyano S: [21]School of Engineering, Griffith University

We developed QR based principal component analysis (PCA) method. Similar to the singular value decomposition (SVD) based PCA method this method is numerically stable. We have carried out analytical comparison as well as numerical comparison (on Matlab software) to investigate the performance (in terms of computational complexity) of our method. The computational complexity of SVD based PCA is around $14dn^2$ flops (where d is the dimensionality of feature space and n is the number of training feature vectors); whereas the computational complexity of QR based PCA is around $2dn^2 + 2dth$ flops (where t is the rank of data covariance matrix and h is the dimensionality of reduced feature space). It is observed that the QR based PCA is more efficient in terms of computational complexity.

b. A strategy to select suitable physicochemical attributes of amino acids for protein fold recognition

Sharma A, Paliwal KK[21], Dehzangi A[21], Lyons J[21], Imoto S, Miyano S

Assigning a protein into one of its folds is a transitional step for discovering three dimensional protein structure, which is a challenging task in bimolecular (biological) science. The present research focuses on: 1) the development of classifiers, and 2) the development of feature extraction techniques based on syntactic and/or physicochemical properties. Apart from the above two main categories of research, we have shown that the selection of physicochemical attributes of the amino acids is an important step in protein fold recognition and has not been explored adequately. We have presented a multi-dimensional successive feature selection (MD-SFS) approach to systematically select attributes. The proposed method is applied on protein sequence data and an improvement of around 24% in fold recognition has been noted when selecting attributes appropriately. The MD-SFS has been applied successfully in selecting physicochemical attributes of the amino acids. The selected attributes show improved protein fold recognition performance.

c. Detecting superbubbles in assembly graphs

Onodera T, Sadakane K[22], Shibuya T: [22]National Institute of Informatics

A new concept of a subgraph class called a superbubble for analyzing assembly graphs is introduced, and an efficient algorithm for detecting it was developed. Most assembly algorithms utilize assembly graphs like the de Bruijn graph or the overlap graph constructed from reads. From these graphs, many assembly algorithms first detect sim-

ple local graph structures (motifs), such as tips and bubbles, mainly to find sequencing errors. These motifs are easy to detect, but they are sometimes too simple to deal with more complex errors. The superbubble is an extension of the bubble, which is also important for analyzing assembly graphs. Though superbubbles are much more complex than ordinary bubbles, we show that they can be efficiently enumerated. We propose an average-case linear time algorithm (i.e., $O(m+n)$ for a graph with n vertices and m edges) for graphs with a reasonable model, though the worst-case time complexity of our algorithm is quadratic (i.e., $O(n(n+m))$). Moreover, the algorithm is practically very fast: Our experiments show that our algorithm runs in reasonable time with a single CPU core even against a very large graph of a whole human genome.

d. The gapped spectrum kernel for support vector machines

Onodera T, Shibuya T

The problem of classifying string data faster and more accurately arises in various fields that involve the analysis of huge amount of strings such as computational biology. A solution to this problem was found. We devised a new string kernel, called gapped spectrum kernel, that yields a kind of sequence of kernels interpolating faster and less accurate string kernels such as the spectrum kernel and slower and more accurate ones such as the wildcard kernel. As a result, we devised an algorithm to compute the wildcard kernel that is provably faster than the state-of-the-art method. The recently introduced b-suffix array data structure plays an important role in this algorithm. Another result is a better trade-off between the speed and accuracy of classification, which we demonstrate by protein classification experiment.

3. Bioinformatics and Alternative Medicine

a. Analysis of questionnaire for traditional medical and develop decision support system

Katayama K, Yamaguchi R, Imoto S, Matsuura K[23], Watanabe K[23,24], Miyano S: [23]Center for Kampo Medicine, Keio University School of Medicine, [24]Faculty of Environment and Information Study, Keio University

Kampo medicine is the Japanese adaptation of traditional medicine. In Kampo medicine, "medical interview" plays an important role. "Medical interview" in Japanese traditional medicine includes not only chief complaint but also a questionnaire that asked about the patient's lifestyle and subjective symptoms. The diagnosis by Kampo is called "Sho" and determined by completely different view from Western medicine. Specialists gather all available information and decide "Sho." This is one of the reasons why non-Kampo specialists without technical knowledge have difficulties to use traditional medicine. We analyzed "medical interview" data to establish an indicator for non-Kampo specialist without technical knowledge to perform suitable traditional medicine. We predicted "Sho" by using random forests algorithm which is powerful algorithm for classification. First, we use all the 2,830 first-visit patients' data. The discriminant ratio of training data was perfect but that of test data is only 67.0%. Second, to achieve high prediction power for practical use, we did data cleaning, and discriminant ratio of test data was 72.4%. Third, we added body mass index (BMI) data to "medical interview" data and discriminant ratio of test data is 91.2%. Originally, deficiency and excess category means that patient is strongly built or poorly built. We notice that the most important variable for classification is BMI.

b. Statistical analysis of hie (cold sensation) and hiesho (cold disorder) in kampo clinic

Yoshino T[23], Katayama K, Munakata K[24], Horiba Y[25], Yamaguchi R, Imoto S, Miyano S, Watanabe K[23,24]: [25]Center for Preventive Medicine, Keio University School of Medicine

A cold sensation (hie) is common in Japanese women and is an important treatment target in Kampo medicine. Physicians diagnose patients as having hiesho (cold disorder) when hie disturbs their daily activity. However, differences between hie and hiesho in men and women are not well described. Hie can be of three types depending on body part where patients feel hie. We aimed to clarify the characteristics of patients with hie and hiesho by analyzing data from new patients seen at the Kampo Clinic at Keio University Hospital during 2008 and 2013. We collected information about patients' subjective symptoms and their severity using visual analogue scales. Of 4,016 new patients, 2,344 complained about hie and 524 of those were diagnosed with hiesho. Hie was most common in legs/feet and combined with hands or lower back, rather than the whole body. Almost 30% of patients with hie felt upper body heat symptoms like hot flushes. Cold sensation was stronger in hiesho than non-hiesho patients. Patients with hie had more complaints. Men with hiesho had the same distribution of hie and had symptoms similar to women. The results of our study may increase awareness of hiesho and help doctors treat hie and other symptoms.

Publications

1. Barclay SS, Tamura T, Ito H, Fujita K, Tagawa K, Shimamura T, Katsuta A, Shiwaku H, Sone M, Imoto S, Miyano S, Okazawa H. Systems biology analysis of Drosophila in vivo screen data elucidates core networks for DNA damage repair in SCA1. Hum Mol Genet. 23(5): 1345-1364, 2014.
2. Becker H, Yoshida K, Blagitko-Dorfs N, Claus R, Pantic M, Abdelkarim M, Niemöller C, Greil C, Hackanson B, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Döhner K, Schnittger S, Henneke P, Niemeyer CM, Flotho C, Pfeifer D, Ogawa S, Lübbert M. Tracing the development of acute myeloid leukemia in CBL syndrome. Blood. 123 (12): 1883-1886, 2014.
3. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, Schnittger S, Sanada M, Kon A, Alpermann T, Yoshida K, Roller A, Nadarajah N, Shiraishi Y, Shiozawa Y, Chiba K, Tanaka H, Koeffler HP, Klein HU, Dugas M, Aburatani H, Kohlmann A, Miyano S, Haferlach C, Kern W, Ogawa S. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 28(2): 241-247, 2014.
4. Hasegawa S, Imai K, Yoshida K, Okuno Y, Muramatsu H, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Kojima S, Ogawa S, Morio T, Mizutani S, Takagi M. Whole-exome sequence analysis of ataxia telangiectasia-like phenotype. J Neurol Sci. S0022-510X(14)00138-5, 2014.
5. Hosono N, Makishima H, Jerez A, Yoshida K, Przychodzen B, McMahon S, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Sanada M, Gómez-Seguí I, Verma AK, McDevitt MA, Sekeres MA, Ogawa S, Maciejewski JP. Recurrent genetic defects on chromosome 7q in myeloid neoplasms. Leukemia. In press.
6. Katayama K, Yamaguchi R, Imoto S, Matsuura K, Watanabe K, Miyano S. Analysis of questionnaire for Traditional Medical and develop decision support system. Evid Based Complement Alternat Med. 2014: 974139, 2014.
7. Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K, Chen F, Asou N, Ohtake S, Miyawaki S, Miyazaki Y, Sakura T, Ozawa Y, Usui N, Kanamori H, Kiguchi T, Imai K, Uike N, Kimura F, Kitamura K, Nakaseko C, Onizuka M, Takeshita A, Ishida F, Suzushima H, Kato Y, Miwa H, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Naoe T. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. Leukemia. In press.
8. Lin TL, Nagata Y, Kao HW, Sanada M, Okuno Y, Huang CF, Liang DC, Kuo MC, Lai CL, Lee EH, Shih YS, Tanaka H, Shiraishi Y, Chiba K, Lin TH, Wu JH, Miyano S, Ogawa S, Shih LY. Clonal leukemic evolution in myelodysplastic syndromes with TET2 and IDH1/2 mutations. Haematologica. 99(1): 28-36, 2014.
9. Oki T, Nishimura K, Kitaura J, Togami K, Maehara A, Izawa K, Sakaue-Sawano A, Niida A, Miyano S, Aburatani H, Kiyonari H, Miyawaki A, Kitamura T. A novel cell-cycle-indicator, mVenus-p27K(−), identifies quiescent cells and visualizes G0-G1 transition. Sci Rep. 4: 4012, 2014.
10. Sakata-Yanagimoto M, Enami T, Yoshida K, Shiraishi Y, Ishii R, Miyake Y, Muto H, Tsuyama N, Sato-Otsubo A, Okuno Y, Sakata S, Kamada Y, Nakamoto-Matsubara R, Tran NB, Izutsu K, Sato Y, Ohta Y, Furuta J, Shimizu S, Komeno T, Sato Y, Ito T, Noguchi M, Noguchi E, Sanada M, Chiba K, Tanaka H, Suzukawa K, Nanmoku T, Hasegawa Y, Nureki O, Miyano S, Nakamura N, Takeuchi K, Ogawa S, Chiba S. Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. Nature Genet. 46(2): 171-175, 2014.
11. Sharma A, Paliwal KK, Imoto S, Miyano S, Sharma V, Ananthanarayanan R. A feature selection method using fixed-point algorithm for DNA microarray gene expression data. International Journal of Knowledge Based and Intelligent Engineering Systems. In press.
12. Sharma A, Dehzangi A, Lyons J, Imoto S, Miyano S, Nakai K, Patil A. Evaluation of sequence features from intrinsically disordered regions for the estimation of protein function. PLoS One. 9(2): e89890, 2014.
13. Sharma A, Paliwal KK, Imoto S, Miyano S. A feature selection method using improved regularized linear discriminant analysis. Mach Vis Appl. 25(3): 775-786, 2014.
14. Sugimachi K, Niida A, Yamamoto K, Shimamura T, Imoto S, Iinuma H, Shinden Y, Eguchi H, Sudo T, Watanabe M, Tanaka J, Kudo S, Hase K, Kusunoki M, Yamada K, Shimada Y, Sugihara K, Maehara Y, Miyano S, Mori M, Mimori K. Allelic Imbalance at an 8q24 Oncogenic SNP is involved in activating MYC in human colorectal cancer. Ann Surg Oncol. In press.
15. Yamaguchi K, Yamaguchi R, Takahashi N, Ikenoue T, Fujii T, Shinozaki M, Tsurita G, Hata K, Niida A, Imoto S, Miyano S, Nakamura Y, Furukawa Y. Overexpression of cohesion establishment factor DSCC1 through E2F in colorectal cancer. PLoS One. 9(1): e85750, 2014.
16. Affara M, Sanders D, Araki H, Tamada Y, Dunmore B, Humphreys S, Imoto S, Savoie C, Miyano S, Kuhara S, Print C, Charnock-Jones DS. Vasohibin-1 is identified as a master-regulator of endothelial cell apoptosis using gene network analysis. BMC Genomics. 14(1) 23, 2013.

17. Damm F, Chesnais V, Nagata Y, Yoshida K, Scourzic L, Okuno Y, Itzykson R, Sanada M, Shiraishi Y, Gelsi-Boyer V, Renneville A, Miyano S, Mori H, Shih LY, Park S, Dreyfus F, Guerci-Bresler A, Solary E, Rose C, Cheze S, Prébet T, Vey N, Legentil M, Duffourd Y, de Botton S, Preudhomme C, Birnbaum D, Bernard OA, Ogawa S, Fontenay M, Kosmider O. BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders. Blood. 122(18): 3169-3177, 2013.
18. Furuta M, Kozaki K, Tanimoto K, Tanaka S, Arii S, Shimamura T, Niida A, Miyano S, Inazawa J. The tumor-suppressive miR-497-195 cluster targets multiple cell-cycle regulators in hepatocellular carcinoma. PLoS One. 8(3): e 60155, 2013.
19. Gómez-Seguí I, Makishima H, Jerez A, Yoshida K, Przychodzen B, Miyano S, Shiraishi Y, Husseinzadeh HD, Guinta K, Clemente M, Hosono N, McDevitt MA, Moliterno AR, Sekeres MA, Ogawa S, Maciejewski JP. Novel recurrent mutations in the RAS-like GTP-binding gene RIT1 in myeloid malignancies. Leukemia. 27(9): 1943-1946, 2013.
20. Hira A, Yabe H, Yoshida K, Okuno Y, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Nakamura J, Kojima S, Ogawa S, Matsuo K, Takata M, Yabe M. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. Blood. 122 (18): 3206-3209, 2013.
21. Ishikawa T, Shimizu T, Ueki A, Yamaguchi SI, Onishi N, Sugihara E, Kuninaka S, Miyamoto T, Morioka H, Nakayama R, Kobayashi E, Toyama Y, Mabuchi Y, Matsuzaki Y, Yamaguchi R, Miyano S, Saya H. Twist2 functions as a tumor suppressor in murine osteosarcoma cells. Cancer Sci. 104(7): 880-888, 2013.
22. Katayama K, Yoshino T, Munakata K, Yamaguchi R, Imoto S, Miyano S, Watanabe K. Prescription of kampo drugs in the Japanese health care insurance program. Evid Based Complement Alternat Med. 2013: 576973, 2013.
23. Kayano M, Imoto S, Yamaguchi R, Miyano S. Multi-omics approach for estimating metabolic networks using low-order partial correlations. J Comput Biol. 20(8): 571-582, 2013.
24. Kitamura K, Yoshida K, Shiraishi Y, Chiba K, Tanaka H, Furukawa K, Miyano S, Ogawa S, Kunishima S. Normal neutrophil myosin IIA localization in an immunofluorescence analysis can rule out MYH9 disorders. J Thromb Haemost. 11(11): 2071-2073, 2013.
25. Komatsu M, Yoshimaru T, Matsuo T, Kiyotani K, Miyoshi Y, Tanahashi T, Rokutan K, Yamaguchi R, Saito A, Imoto S, Miyano S, Nakamura Y, Sasa M, Shimada M, Katagiri T. Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. Int J Oncol. 42(2): 478-506, 2013.
26. Kon A, Shih LY, Minamino M, Sanada M, Shiraishi Y, Nagata Y, Yoshida K, Okuno Y, Bando M, Nakato R, Ishikawa S, Sato-Otsubo A, Nagae G, Nishimoto A, Haferlach C, Nowak D, Sato Y, Alpermann T, Nagasaki M, Shimamura T, Tanaka H, Chiba K, Yamamoto R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Nakamaki T, Ishiyama K, Nolte F, Hofmann WK, Miyawaki S, Chiba S, Mori H, Nakauchi H, Koeffler HP, Aburatani H, Haferlach T, Shirahige K, Miyano S, Ogawa S. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nature Genet. 45(10): 1232-1237, 2013.
27. Kunishima S, Okuno Y, Yoshida K, Shiraishi Y, Sanada M, Muramatsu H, Chiba K, Tanaka H, Miyazaki K, Sakai M, Ohtake M, Kobayashi R, Iguchi A, Takahashi Y, Miyano S, Saito H, Kojima S, Ogawa S. ACTN1 mutations cause congenital macrothrombocytopenia. Am J Hum Genet. 92(3): 431-438, 2013.
28. Makishima H, Yoshida K, Nguyen N, Przychodzen B, Sanada M, Okuno Y, Ng KP, Gudmundsson KO, Vishwakarma BA, Jerez A, Gomez-Segui I, Takahashi M, Shiraishi Y, Nagata Y, Guinta K, Mori H, Sekeres MA, Chiba K, Tanaka H, Muramatsu H, Sakaguchi H, Paquette RL, McDevitt MA, Kojima S, Sauntharajah Y, Miyano S, Shih LY, Du Y, Ogawa S, Maciejewski JP. Somatic SETBP1 mutations in myeloid malignancies. Nature Genet. 45(8): 942-946, 2013.
29. Niida A, Tremmel G, Imoto S, Miyano S. Multilayer cluster heat map visualizing biological tensor data. Lecture Notes in Bioinformatics. 8213: 116-125, 2013.
30. Nagasaki M, Fujita A, Sekiya Y, Saito A, Ikeda E, Li C, Miyano S. XiP: a computational environment to create, extend and share workflows. Bioinformatics. 29(1): 137-139, 2013.
31. Onodera T, Sadakane K, Shibuya T. Detecting superbubbles in assembly graphs. Lecture Notes in Computer Science. 8126: 338-348, 2013.
32. Onodera T, Shibuya T. The gapped spectrum kernel for support vector machines. Lecture Notes in Computer Science. 7988: 1-15, 2013.
33. Osawa T, Tsuchida R, Muramatsu M, Shimamura T, Wang F, Suehiro JI, Kanki Y, Wada Y, Yuasa Y, Aburatani H, Miyano S, Minami T, Kodama T, Shibuya M. Inhibition of histone demethylase JMJD1A improves anti-angiogenic therapy and reduces tumor associated macrophages. Cancer Res. 73(10): 3019-3028, 2013.
34. Saida S, Watanabe K, Sato-Otsubo A, Terui K, Yoshida K, Okuno Y, Toki T, Wang RN, Shiraishi Y, Miyano S, Kato I, Morishima T, Fujino H, Umeda K, Hiramatsu H, Adachi S, Ito

E, Ogawa S, Ito M, Nakahata T, Heike T. Clonal selection in xenografted TAM recapitulates the evolutionary process of myeloid leukemia in Down syndrome. Blood. 121(21): 4377-4387, 2013.
35. Saito MM, Imoto S, Yamaguchi R, Tsubokura M, Kami M, Nakada H, Sato H, Miyano S, Higuchi T. Enhancement of collective immunity in Tokyo Metropolitan area by selective vaccination against an emerging influenza pandemic. PLoS One. 8(9): e72866, 2013.
36. Saito MM, Imoto S, Yamaguchi R, Sato H, Nakada H, Kami M, Miyano S, Higuchi T. Extension and verification of the SEIR model on the 2009 influenza A (H1N1) pandemic in Japan. Math Biosci. 246(1): 47-54, 2013.
37. Sakaguchi H, Okuno Y, Muramatsu H, Yoshida K, Shiraishi Y, Takahashi M, Kon A, Sanada M, Chiba K, Tanaka H, Makishima H, Wang X, Xu Y, Doisaki S, Hama A, Nakanishi K, Takahashi Y, Yoshida N, Maciejewski JP, Miyano S, Ogawa S, Kojima S. Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia. Nature Genet. 45(8): 937-941, 2013.
38. Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, Shimamura T, Sato-Otsubo A, Nagae G, Suzuki H, Nagata Y, Yoshida K, Kon A, Suzuki Y, Chiba K, Tanaka H, Niida A, Fujimoto A, Tsunoda T, Morikawa T, Maeda D, Kume H, Sugano S, Fukayama M, Aburatani H, Sanada M, Miyano S, Homma Y, Ogawa S. Integrated molecular analysis of clear-cell renal cell carcinoma. Nature Genet. 45(8): 860-867, 2013.
39. Sharma A, Paliwal KK, Dehzangi A, Lyons J, Imoto S, Miyano S. A strategy to select suitable physicochemical attributes of amino acids for protein fold recognition. BMC Bioinformatics. 14(1): 233, 2013.
40. Sharma A, Paliwal KK, Imoto S, Miyano S. Principal component analysis using QR decomposition. International J Machine Learning and Cybernetics. 4(6): 679-683, 2013.
41. Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, Shiba N, Hayashi Y, Kume H, Homma Y, Sanada M, Ogawa S, Miyano S. An empirical Bayesian framework for somatic mutation detection from cancer. Nucleic Acids Res. 41(7): e89, 2013.
42. Takatsuno Y, Mimori K, Yamamoto K, Sato T, Niida A, Inoue H, Imoto S, Kawano S, Yamaguchi R, Toh H, Iinuma H, Ishimaru S, Ishii H, Suzuki S, Tokudome S, Watanabe M, Tanaka JI, Kudo SE, Mochizuki H, Kusunoki M, Yamada K, Shimada Y, Moriya Y, Miyano S, Sugihara K, Mori M. The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. Ann Surg Oncol. 20(4): 1395-1402, 2013.
43. Tamura T, Sone M, Nakamura Y, Shimamura T, Imoto S, Miyano S, Okazawa H. A restricted level of PQBP1 is needed for the best longevity of Drosophila. Neurobiol Aging. 34(1): 356.e11-20, 2013.
44. Yamaguchi R, Imoto S, Kami M, Watanabe K, Miyano S, Yuji K. Does Twitter trigger bursts in signature collections? PLoS One. 8(3): e58252, 2013.
45. Yokobori T, Iinuma H, Shimamura T, Imoto S, Sugimachi K, Ishii H, Iwatsuki M, Ota D, Ohkuma M, Iwaya T, Nishida N, Kogo R, Sudo T, Tanaka F, Shibata K, Toh H, Sato T, Barnard GF, Fukagawa T, Yamamoto S, Nakanishi H, Sasaki S, Miyano S, Watanabe T, Kuwano H, Mimori K, Pantel K, Mori M. Plastin3 is a novel marker for circulating tumor cells undergoing the epithelial-mesenchymal transition and is associated with colorectal cancer prognosis. Cancer Res. 73(7): 2059-2069, 2013.
46. Yoshida K, Toki T, Okuno Y, Kanezaki R, Shiraishi Y, Sato-Otsubo A, Sanada M, Park MJ, Terui K, Suzuki H, Kon A, Nagata Y, Sato Y, Wang R, Shiba N, Chiba K, Tanaka H, Hama A, Muramatsu H, Hasegawa D, Nakamura K, Kanegane H, Tsukamoto K, Adachi S, Kawakami K, Kato K, Nishimura R, Izraeli S, Hayashi Y, Miyano S, Kojima S, Ito E, Ogawa S. The landscape of somatic mutations in Down syndrome-related myeloid disorders. Nature Genet. 45: 1293-1299, 2013. Erratum in: Nature Genet. 45(12): 1516, 2013.
47. Yoshimaru T, Komatsu M, Matsuo T, Chen YA, Murakami Y, Mizuguchi K, Mizohata E, Inoue T, Akiyama M, Yamaguchi R, Imoto S, Miyano S, Miyoshi Y, Sasa M, Nakamura Y, Katagiri T. Targeting BIG3-PHB2 interaction to overcome tamoxifen resistance in breast cancer cells. Nature Commun. 4: 2443, 2013.
48. Yoshino T, Katayama K, Munakata K, Horiba Y, Yamaguchi R, Imoto S, Miyano S, Watanabe K. Statistical analysis of hie (cold sensation) and hiesho (cold disorder) in kampo clinic. Evid Based Complement Alternat Med. 2013: 398458, 2013.

Human Genome Center

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

Associate Professor Koichi Matsuda, M.D., Ph.D.
Assistant Professor Chizu Tanikawa, Ph.D.

准教授 松田浩一
助教 谷川千津

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.

Koichi Matsuda, Chizu Tanikawa, Jiaying Lin, Zhenzhong Deng, Paulisally Hau Yi Lo, Yuki Funauchi, Masami Tanaka, Kang Daechun, Takashi Fujitomo, Jinichi Mori.

1. Genes playing significant roles in human cancer

(1) Epigenetics

The histone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis

Histone lysine methylation plays a fundamental role in chromatin organization. Although a set of histone methyltransferases have been identified and biochemically characterized, the pathological roles of their dysfunction in human cancers are still not well understood. In this study, we demonstrate important roles of WHSC1L1 in human carcinogenesis. Expression levels of WHSC1L1 transcript were significantly elevated in various human cancers including bladder carcinoma. Immunohistochemical analysis of bladder, lung, and liver cancers confirmed overexpression of WHSC1L1. WHSC1L1-specific small interfering RNAs significantly knocked down its expression and resulted in suppression of proliferation of bladder and lung cancer cell lines. WHSC1L1 knockdown induced cell cycle arrest at the G(2)/M phase followed by multinucleation of cancer cells. Expression profile analysis using Affymetrix GeneChip(®) showed that WHSC1L 1 affected the expression of a number of genes including CCNG1 and NEK7, which are known to play crucial roles in the cell cycle progression at mitosis. As WHSC1L1 expression is significantly low in various normal tissues including vital organs, WHSC1L1 could be a good candidate molecule for development of novel treatment for various types of cancer.

(2) Kidney cancer

Downregulation of the tumor suppressor HSPB7, involved in the p53 pathway, in renal cell carcinoma by hypermethylation.

In order to identify genes involved in renal carcinogenesis, we analyzed the expression profile of renal cell carcinomas (RCCs) using microarrays consisting of 27,648 cDNA or ESTs, and found a small heat shock protein, HSPB7, to be significantly and commonly downregulated in RCC. Subsequent quantitative PCR (qPCR) and immunohistochemical (IHC) analyses confirmed the downregulation of HSPB7 in RCC tissues and cancer cell lines in both transcriptional and protein levels. Bisulfite sequencing of a genomic region of HSPB7 detected DNA hypermethylation of some segments of HSPB7 in RCC cells and concordantly 5-aza-2'-deoxycytidine (5-Aza-dC) treatment of cancer cells restored HSPB 7 expression significantly. Ectopic introduction of HSPB7 in five RCC cell lines remarkably suppressed cancer cell growth. Interestingly, we found that HSPB7 expression could be induced by p53 in a dose-dependent manner, indicating that this gene functions in the p53 pathway. Our results imply that HSBP7 is likely to be a tumor suppressor gene regulated by p53 and its downregulation by hypermethylation may play a critical role in renal carcinogenesis.

(3) Liver cancer

Identification of a functional variant in the MICA promoter which regulates MICA expression and increases HCV-related hepatocellular carcinoma risk.

Hepatitis C virus (HCV) infection is the major cause of hepatocellular carcinoma (HCC) in Japan. We previously identified the association of SNP rs 2596542 in the 5' flanking region of the MHC class I polypeptide-related sequence A (MICA) gene with the risk of HCV-induced HCC. In the current study, we performed detailed functional analysis of 12 candidate SNPs in the promoter region and found that a SNP rs2596538 located at 2.8 kb upstream of the MICA gene affected the binding of a nuclear protein(s) to the genomic segment including this SNP. By electrophoretic mobility shift assay (EMSA) and chromatin immunoprecipitation (ChIP) assay, we identified that transcription factor Specificity Protein 1 (SP1) can bind to the protective G allele, but not to the risk A allele. In addition, reporter construct containing the G allele was found to exhibit higher transcriptional activity than that containing the A allele. Moreover, SNP rs2596538 showed stronger association with HCV-induced HCC ($P=1.82\times10(-5)$ and OR=1.34) than the previously identified SNP rs2596542. We also found significantly higher serum level of soluble MICA (sMICA) in HCV-induced HCC patients carrying the G allele than those carrying the A allele (P= 0.00616). In summary, we have identified a functional SNP that is associated with the expression of MICA and the risk for HCV-induced HCC

2. Pharmacogenetics

Impact of four loci on serum tamsulosin hydrochloride concentration.

Tamsulosin hydrochloride is one of the most potent drugs for treatment of benign prostatic hyperplasia (BPH), however, the efficacy of tamsulosin hydrochloride varies among individuals. In this study, we measured the maximum serum concentration (Cmax) of tamsulosin hydrochloride in 182 of BPH patients and found remarkable individual variability. To investigate the genetic factors that regulate pharmacokinetics of tamsulosin hydrochloride, we conducted a genome-wide association study in these 182 BPH patients. As a result, rs 16902947 on chromosome 5p13.2, rs7779057 on 7q 22.3, rs35681285 on 7p21.2 and rs2122469 on 8p21.3 indicated possible associations with Cmax of tamsulosin hydrochloride ($P=1.29\times10(-7)$, $2.15\times10(-7)$, $4.35\times10(-7)$ and $7.03\times10(-7)$, respectively), although these single-nucleotide polymorphisms (SNPs) did not reach the genome-wide significance threshold after Bonferroni correction. As these associated SNPs showed additive effects on serum tamsulosin hydrochloride concentration, we defined the `Cmax prediction index' based on genotypes of these SNPs. This index clearly associated with Cmax values ($P=4.5\times10(-6)$), indicating the possible roles of these four variants in tamsulosin hydrochloride pharmacokinetics. Our findings would partially explain the variability of the response to the tamsulosin hydrochloride treatment.

3. Genome-wide association study

A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region.

BACKGROUND & AIMS:

We performed a genome-wide association study (GWAS) of hepatitis C virus (HCV)-induced liver cirrhosis (LC) to identify predictive biomarkers for the risk of LC in patients with chronic hepatitis C (CHC).

METHODS:

A total of 682 HCV-induced LC cases and 1045 CHC patients of Japanese origin were genotyped by Illumina Human Hap 610-Quad bead Chip.

RESULTS:

Eight SNPs which showed possible associations

($p<1.0\times10(-5)$) at the GWAS stage were further genotyped using 936 LC cases and 3809 CHC patients. We found that two SNPs within the major histocompatibility complex (MHC) region on chromosome 6p21, rs910049 and rs3135363, were significantly associated with the progression from CHC to LC (pcombined=$9.15\times10(-11)$ and $1.45\times10(-10)$, odds ratio (OR)=1.46 and 1.37, respectively). We also found that HLA-DQA1(*)0601 and HLA-DRB1(*)0405 were associated with the progression from CHC to LC ($p=4.53\times10(-4)$ and $1.54\times10(-4)$ with OR=2.80 and 1.45, respectively). Multiple logistic regression analysis revealed that rs3135363, rs910049, and HLA-DQA1(*)0601 were independently associated with the risk of HCV-induced LC. In addition, individuals with four or more risk alleles for these three loci have a 2.83-fold higher risk for LC than those with no risk allele, indicating the cumulative effects of these variations.
CONCLUSIONS:
Our findings elucidated the crucial roles of multiple genetic variations within the MHC region as prognostic/predictive biomarkers for CHC patients.

A replication study for three nephrolithiasis loci at 5q35.3, 7p14.3 and 13q14.1 in the Japanese population.

A previous genome-wide association study (GWAS) reported three novel nephrolithiasis-susceptibility loci at 5q35.3, 7p14.3 and 13q14.1. Here, we investigated the association of these loci with nephrolithiasis by using an independent Japanese sample set. We performed case-control association analysis using 601 patients with nephrolithiasis and 201 control subjects. We selected seven single-nucleotide polymorphisms (SNPs): rs12654812 and rs 11746443 from 5q35.3 (RGS14-SLC34A1-PFN3-F12); rs12669187 and rs1000597 from 7p14.3 (INMT-FAM 188B-AQP1); and rs7981733, rs1170155, and rs 4142110 from 13q14.1 (DGKH (diacylglycerol kinase)), which were previously reported to be significantly associated with nephrolithiasis. rs 12654812, rs12669187 and rs7981733 were significantly associated with nephrolithiasis after Bonferroni's correction ($P=3.12\times10(-3)$, odds ratio (OR)= 1.43; $P=6.40\times10(-3)$, OR=1.57; and $P=5.00\times10(-3)$, OR=1.41, respectively). Meta-analysis of current and previous GWAS results indicated a significant association with nephrolithiasis ($P=7.65\times10(-15)$, $7.86\times10(-14)$ and $1.06\times10(-9)$, respectively). We observed a cumulative effect with these three SNPs; individuals with three or more risk alleles had a 5.9-fold higher risk for nephrolithiasis development than those with only one risk allele. Our findings elucidated the significance of genetic variation at these three loci in nephrolithiasis in the Japanese population.

Impact of PSCA variation on gastric ulcer susceptibility.

Peptic ulcer is one of the most common gastrointestinal disorders with complex etiology. Recently we conducted the genome wide association study for duodenal ulcer and identified disease susceptibility variations at two genetic loci corresponding to the Prostate stem cell antigen (PSCA) gene and the ABO blood group (ABO) gene. Here we investigated the association of these variations with gastric ulcer in two Japanese case-control sample sets, a total of 4,291 gastric ulcer cases and 22,665 controls. As a result, a C-allele of rs2294008 at PSCA increased the risk of gastric ulcer with odds ratio (OR) of 1.13 (P value of $5.85\times10(-7)$) in an additive model. On the other hand, SNP rs505922 on ABO exhibited inconsistent result between two cohorts. Our finding implies presence of the common genetic variant in the pathogenesis of gastric and duodenal ulcers.

Genome wide association study of age at menarche in the Japanese population.

Age at menarche (AAM) is a complex trait involving both genetic and environmental factors. To identify the genetic factors associated with AAM, we conducted a large-scale meta-analysis of genome-wide association studies using more than 15,000 Japanese female samples. Here, we identified an association between SNP (single nucleotide polymorphism) rs364663 at the LIN28B locus and AAM, with a P-value of $5.49\times10(-7)$ and an effect size of 0.089 (year). We also evaluated 33 SNPs that were previously reported to be associated with AAM in women of European ancestry. Among them, two SNPs rs4452860 and rs7028916 in TMEM38B indicated significant association with AAM in the same directions as reported in previous studies (P?=? 0.0013 with an effect size of 0.051) even after Bonferroni correction for the 33 SNPs. In addition, six loci in or near CCDC85A, LOC100421670, CA10, ZNF483, ARNTL, and RXRG exhibited suggestive association with AAM ($P<0.05$). Our findings elucidated the impact of genetic variations on AAM in the Japanese population.

References

1. D. Kang, H.S. Cho, G. Toyokawa, M. Kogure, Y. Yamane, Y. Iwai, S. Hayami, T. Tsunoda, H. I. Field, K. Matsuda, D.E. Neal, B.A. Ponder, Y. Maehara, Y. Nakamura, R. Hamamoto, The his-

tone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis. Genes Chromosomes Cancer 52 (2013) 126-139.
2. Y. Urabe, H. Ochi, N. Kato, V. Kumar, A. Takahashi, R. Muroyama, N. Hosono, M. Otsuka, R. Tateishi, P.H. Lo, C. Tanikawa, M. Omata, K. Koike, D. Miki, H. Abe, N. Kamatani, J. Toyota, H. Kumada, M. Kubo, K. Chayama, Y. Nakamura, K. Matsuda, A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. Journal of hepatology (2013).
3. R. Takata, K. Matsuda, J. Sugimura, W. Obara, T. Fujioka, K. Okihara, N. Takaha, T. Miki, S. Ashida, K. Inoue, C. Tanikawa, T. Shuin, S. Sasaki, Y. Kojima, K. Kohri, M. Kubo, M. Yamaguchi, Y. Ohnishi, Y. Nakamura, Impact of four loci on serum tamsulosin hydrochloride concentration. J Hum Genet 58 (2013) 21-26.
4. A. Aarnink, H.J. Garchon, Y. Okada, A. Takahashi, K. Matsuda, M. Kubo, Y. Nakamura, A. Blancher, Comparative analysis in cynomolgus macaque identifies a novel human MHC locus controlling platelet blood counts independently of BAK1. Journal of thrombosis and haemostasis: JTH 11 (2013) 384-386.
5. J. Wang, L.G. Carvajal-Carmona, J.H. Chu, A.G. Zauber, A.T. Collaborators, M. Kubo, K. Matsuda, M. Dunlop, R.S. Houlston, O.M. Sieber, L. Lipton, P. Gibbs, N.G. Martin, G.W. Montgomery, J.P. Young, P.N. Baird, M.J. Ratain, Y. Nakamura, S. Weiss, I. Tomlinson, M.M. Bertagnolli, Germline Variants and Advanced Colorectal Adenomas: Adenoma Prevention with Celecoxib Trial Genomewide Association Study. Clin Cancer Res (2013).
6. C.M. O'Seaghdha, H. Wu, Q. Yang, K. Kapur, I. Guessous, A.M. Zuber, A. Kottgen, C. Stoudmann, A. Teumer, Z. Kutalik, M. Mangino, A. Dehghan, W. Zhang, G. Eiriksdottir, G. Li, T. Tanaka, L. Portas, L.M. Lopez, C. Hayward, K. Lohman, K. Matsuda, S. Padmanabhan, D. Firsov, R. Sorice, S. Ulivi, A.C. Brockhaus, M.E. Kleber, A. Mahajan, F.D. Ernst, V. Gudnason, L. J. Launer, A. Mace, E. Boerwinckle, D.E. Arking, C. Tanikawa, Y. Nakamura, M.J. Brown, J. M. Gaspoz, J.M. Theler, D.S. Siscovick, B.M. Psaty, S. Bergmann, P. Vollenweider, V. Vitart, A.F. Wright, T. Zemunik, M. Boban, I. Kolcic, P. Navarro, E.M. Brown, K. Estrada, J. Ding, T.B. Harris, S. Bandinelli, D. Hernandez, A.B. Singleton, G. Girotto, D. Ruggiero, A.P. d'Adamo, A. Robino, T. Meitinger, C. Meisinger, G. Davies, J.M. Starr, J.C. Chambers, B.O. Boehm, B.R. Winkelmann, J. Huang, F. Murgia, S.H. Wild, H. Campbell, A.P. Morris, O.H. Franco, A. Hofman, A.G. Uitterlinden, F. Rivadeneira, U. Volker, A. Hannemann, R. Biffar, W. Hoffmann, S.Y. Shin, P. Lescuyer, H. Henry, C. Schurmann, P.B. Munroe, P. Gasparini, N. Pirastu, M. Ciullo, C. Gieger, W. Marz, L. Lind, T. D. Spector, A.V. Smith, I. Rudan, J.F. Wilson, O. Polasek, I.J. Deary, M. Pirastu, L. Ferrucci, Y. Liu, B. Kestenbaum, J.S. Kooner, J.C. Witteman, M. Nauck, W.H. Kao, H. Wallaschofski, O. Bonny, C.S. Fox, M. Bochud, Meta-analysis of genome-wide association studies identifies six new Loci for serum calcium concentrations. PLoS Genet 9 (2013) e1003796.
7. T. Yasui, A. Okada, Y. Urabe, M. Usami, K. Mizuno, Y. Kubota, K. Tozawa, S. Sasaki, Y. Higashi, Y. Sato, M. Kubo, Y. Nakamura, K. Matsuda, K. Kohri, A replication study for three nephrolithiasis loci at 5q35.3, 7p14.3 and 13q 14.1 in the Japanese population. J Hum Genet 58 (2013) 588-593.
8. C. Tanikawa, K. Matsuo, M. Kubo, A. Takahashi, H. Ito, H. Tanaka, Y. Yatabe, K. Yamao, N. Kamatani, K. Tajima, Y. Nakamura, K. Matsuda, Impact of PSCA variation on gastric ulcer susceptibility. PloS one 8 (2013) e63698.
9. C. Tanikawa, Y. Okada, A. Takahashi, K. Oda, N. Kamatani, M. Kubo, Y. Nakamura, K. Matsuda, Genome wide association study of age at menarche in the Japanese population. PloS one 8 (2013) e63821.
10. P.H. Lo, Y. Urabe, V. Kumar, C. Tanikawa, K. Koike, N. Kato, D. Miki, K. Chayama, M. Kubo, Y. Nakamura, K. Matsuda, Identification of a functional variant in the MICA promoter which regulates MICA expression and increases HCV-related hepatocellular carcinoma risk. PloS one 8 (2013) e61279.
11. S.W. Yee, J.A. Mefford, N. Singh, M.E. Percival, A. Stecula, K. Yang, J.S. Witte, A. Takahashi, M. Kubo, K. Matsuda, K.M. Giacomini, C. Andreadis, Impact of polymorphisms in drug pathway genes on disease-free survival in adults with acute myeloid leukemia. J Hum Genet 58 (2013) 353-361.
12. B. Zhang, W.H. Jia, K. Matsuo, A. Shin, Y.B. Xiang, K. Matsuda, S.H. Jee, D.H. Kim, P.Y. Cheah, Z. Ren, Q. Cai, J. Long, J. Shi, W. Wen, G. Yang, B.T. Ji, Z.Z. Pan, F. Matsuda, Y.T. Gao, J.H. Oh, Y.O. Ahn, M. Kubo, L.F. Thean, E.J. Park, H.L. Li, J.W. Park, J. Jo, J.Y. Jeong, S. Hosono, Y. Nakamura, X.O. Shu, Y.X. Zeng, W. Zheng, Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians. International journal of cancer. (2014).
13. T. Kashiyama, K. Oda, Y. Ikeda, Y. Shiose, Y. Hirota, K. Inaba, C. Makii, R. Kurikawa, A. Miyasaka, T. Koso, T. Fukuda, M. Tanikawa, K. Shoji, K. Sone, T. Arimoto, O. Wada-Hiraike, K. Kawana, S. Nakagawa, K. Matsuda, F. McCor-

mick, H. Aburatani, T. Yano, Y. Osuga, T. Fujii, Antitumor Activity and Induction of TP53-Dependent Apoptosis toward Ovarian Clear Cell Adenocarcinoma by the Dual PI3K/mTOR Inhibitor DS-7423. PloS one 9 (2014) e87220.

14. J. Lin, Z. Deng, C. Tanikawa, T. Shuin, T. Miki, K. Matsuda, Y. Nakamura, Downregulation of the tumor suppressor HSPB7, involved in the p 53 pathway, in renal cell carcinoma by hypermethylation. Int J Oncol (2014).

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. Promoter structure modeling of co-expressed genes

Yosvany López 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 factors binding sites (TFBSs) seems to be responsible for the complexity of gene regulation in every living organism. Based on the assumption that regulatory regions of genes showing similar expression profiles should share some common structural characteristics, we are attempting to explain how the binding of transcription factors is carried out in promoters of genes expressed in specific biological tissues or physiological conditions. We have found new structural patterns of motif combinations capable of describing the promoters of co-expressed genes in five different plant structures. In addition, we have also proposed a set of features including relative position between pairs of motifs, positioning from transcription start sites, order and orientation of motifs in promoters of *Drosophila melanogaster's* co-expressed genes. We expect to improve the feature set with additional characteristics to help describe the distinct promoters.

2. Analysis of changes in transcription start site distribution by a classification approach

Kuo-Ching Liang, Yutaro Kumagai[1], Yutaka Suzuki[2] and Kenta Nakai: [1]WPI-iFReC, Osaka University, [2]Graduate School of Frontier Sciences

Change in transcription start site (TSS) usage is an important mechanism for the control of transcription process, and has significant effect on the isoforms being transcribed. One of the goals in the study of TSS is the understanding of how and why their usage differs in different tissues or under different conditions. In light of recent efforts in the mapping of transcription start site landscape using high-throughput sequencing approaches, a quantitative and automated method is needed to process all the data that are being produced. We propose a statistical approach that will classify changes in TSS distribution between different samples into several categories of changes that may have biological significance. Genes selected by the classifiers can then

be analyzed together with additional supporting data to determine their biological significance. We use a set of time-course TSS data from mouse dendritic cells stimulated with lipopolysaccharide (LPS) to demonstrate the usefulness of our method. Using the proposed classifier, we show that genes involved in only one of MyD88-dependent or TRIF-dependent part of the TLR4 signaling pathway, and genes that are involved in both show different patterns of TSS usage after LPS stimulation.

3. Inferring the choreography of parental genomes during fertilization from ultra-large-scale whole-transcriptome analysis

Sung-Joon Park, Makiko Komata[3], Fukashi Inoue[4], Kaori Yamada[4], Kenta Nakai, Miho Ohsugi[4], Katsuhiko Shirahige[3]: [3]IMCB, U. Tokyo, [4]Division of Oncology

Fertilization precisely choreographs parental genomes by utilizing gamete-derived cellular factors and by activating genome regulatory programs. However, the mechanism remains elusive owing to the technical difficulties of preparing large numbers of high-quality preimplantation cells. Here, we collected more than 14×10^4 high-quality mouse metaphase II oocytes and used these to establish detailed transcriptional profiles for four early embryo stages and parthenogenetic development. By combining these profiles with other public resources, we found evidence that gene silencing appeared to be mediated in part by non-coding RNAs and that this was a prerequisite for post-fertilization development. Notably, we identified 817 genes that were differentially expressed in embryos after fertilization compared to parthenotes. The regulation of these genes were distinctly different to those expressed in parthenotes, suggesting functional specialization of particular transcription factors prior to first cell cleavage. We identified five transcription factors that were potentially necessary for developmental progression: Foxd1, Nkx2-5, Sox18, Myod1, and Runx1. Our very large scale whole-transcriptome profile of early mouse embryos has yielded a novel and valuable resource for studies in developmental biology and stem cell research. The database is available at http://dbtmee.hgc.jp/.

4. Computational promoter modeling identifies the modes of transcriptional regulation in hematopoietic stem cells

Sung-Joon Park, Terumasa Umemoto[5], Mihiko Saito-Adachi, Yoshiko Shiratsuchi[5], Masayuki Yamamoto[5], Kenta Nakai: [5]Institute of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University

Extrinsic and intrinsic regulators are responsible for the tight control of hematopoietic stem cells (HSCs), which differentiate into all blood cell lineages. To understand the fundamental basis of HSC biology, we focused on differentially expressed genes (DEGs) in long-term and short-term HSCs, which are closely related in terms of cell development but substantially differ in their stem cell capacity. To analyze the transcriptional regulation of the DEGs identified in the novel transcriptome profiles obtained by our RNA-seq analysis, we developed a computational method to model the linear relationship between gene expression and the features of putative regulatory elements. The transcriptional regulation modes characterized here suggest the importance of transcription factors (TFs) that are expressed at steady state or at low levels. Remarkably, we found that 24 differentially expressed TFs targeting 21 putative TF-binding sites contributed significantly to transcriptional regulation. These TFs tended to be modulated by other nondifferentially expressed TFs, suggesting that HSCs can achieve flexible and rapid responses via the control of nondifferentially expressed TFs through a highly complex regulatory network. Our novel transcriptome profiles and new method are powerful tools for studying the mechanistic basis of cell fate decisions.

5. Genome-wide analysis of transcription start sites and promoters in *Ciona intestinalis*

Rui Yokomori, Kotaro Shimai[6], Koki Nishitsuji[7], Yutaka Suzuki[2], Takehiro Kusakabe[6] and Kenta Nakai : [6]Fac. of Sci. and Engineering, Konan Univ, [7]Grad. Schl. of Life Sci., Univ. of Hyogo

The high-throughput sequencing method that generates millions of 5′ sequences derived from 5′ capped RNAs has enabled us to identify transcription start sites (TSSs) on a genome-wide scale and has helped identify promoters and study their characteristics. *Ciona intestinalis*, one of the urochordates, has compact gene regulatory regions compared to human. Moreover, it has recently been shown that the urochordate is an invertebrate that is evolutionarily closest to vertebrates. These indicate that *C. intestinalis* is a useful and important model organism in the study of *cis*-regulatory regions. However, genome-wide promoter analysis has not been done in *C. intestinalis* and the characteristics of the promoters are unknown. Therefore, we identified TSSs of *C. intestinalis* on a genome-wide scale and obtained the corresponding promoters. We found many unknown promoters in intergenic and intron regions in addition to known promoters. We further characterized promoters based on TSS distributions and promoter motifs and compared them with human promoters. While the char-

acteristics and the function of TCT motifs and TATA boxes were conserved, TSS usage and the association with CpG islands were different between the two species. These results help in our understanding of the evolution of promoters.

6. A study of the innate immunity interactome dynamics

Asmaa Elzawahry, Ashwini Patil, Yutaro Kumagai[1], Yutaka Suzuki[2], Kenta Nakai

Immune system of host is responsible for defense against invading pathogens. Protein-protein Interactions (PPIs) are essential components of the immune response. So far, protein-protein interactions have been curated as a static interaction map. However, interactions involved in the immune response are dynamic. This study focuses on interactome dynamics during immune response by combining time-series gene expression data with protein-protein interactions followed by protein complex identification. We identified differences in the interactome during immune response by constructing differential networks and identifying protein modules which were up/down regulated for each stage. In addition, we identified core interactions that are conserved in interactome throughout immune response. We defined Interaction protein ratio (IPR) and Pair wise differential matrix (PWDM) to assess differences between network maps. To get a comprehensive view of the TLR interactome network we investigated the TLR subnetwork and found that S100A8 is down regulated in dendritic cells after LPS stimulation. Combining time-series gene expression data with protein-protein interactions data revealed interactome dynamics during immune response. Identified protein complexes during interactome dynamics have a crucial role not only in innate immunity but also in other biological processes and pathways like pathways in cancer, circadian rhythm and p53 pathway.

7. Alterations in rRNA-mRNA interaction during plastid evolution

Kyungtaek Lim, Ichizo Kobayashi[2], and Kenta Nakai

Translation initiation depends on the recognition of mRNA by a ribosome. For this to occur, prokaryotes primarily use the Shine-Dalgarno (SD) interaction, where the 3′ tail of small subunit rRNA (core motif: 3′CCUCC) forms base pairs with a complementary signal sequence in the 5′ untranslated region of mRNA. Here we examined what happened to SD interactions during the evolution of a cyanobacterial endosymbiont into modern plastids (including chloroplasts). Our analysis of available complete plastid genome sequences revealed that the majority of plastids retained SD interactions but with varying levels of usage. Parallel losses of SD interactions took place in plastids of Chlorophyta, Euglenophyta, and Chromerida/Apicomplexa lineages, presumably related to their extensive reductive evolution. Interestingly, we discovered that the classical SD interaction (3′CCUCC/5′GGAGG (rRNA/mRNA)) was replaced by an altered SD interaction (3′CCCU/5′GGGA or 3′CUUCC/5′GAAGG) through coordinated changes in the sequences of the core rRNA motif and its paired mRNA signal. These changes in plastids of Chlorophyta and Euglenophyta proceeded through intermediate steps that allowed both the classical and altered SD interactions. This coevolution between the rRNA motif and the mRNA signal demonstrates unexpected plasticity in the translation initiation machinery.

8. Evaluation of Sequence Features from Intrinsically Disordered Regions for the Estimation of Protein Function

Alok Sharma[8,9], Abdollah Dehzangi[9,10], James Lyons[11], Seiya Imoto, Satoru Miyano, Kenta Nakai, Ashwini Patil: [8]School of Engg. and Physics, The Univ. of the South Pacific, Suva, Fiji, [9]IIIS, Griffith Univ., [10]NICTA, [11]School of Engg., Griffith Univ., Brisbane, Australia

With the exponential increase in the number of sequenced organisms, automated annotation of proteins is becoming increasingly important. Intrinsically disordered regions are known to play a significant role in protein function. Despite their abundance, especially in eukaryotes, they are rarely used to inform function prediction systems. In this study, we extracted seven sequence features in intrinsically disordered regions and developed a scheme to use them to predict Gene Ontology Slim terms associated with proteins. We evaluated the function prediction performance of each feature. Our results indicate that the residue composition based features have the highest precision while bigram probabilities, based on sequence profiles of intrinsically disordered regions obtained from PSIBlast, have the highest recall. Amino acid bigrams and features based on secondary structure show an intermediate level of precision and recall. Almost all features showed a high prediction performance for GO Slim terms related to extracellular matrix, nucleus, RNA and DNA binding. However, feature performance varied significantly for different GO Slim terms emphasizing the need for a unique classifier optimized for the prediction of each functional term. These findings provide a first comprehensive and quantitative evaluation of sequence features in intrinsically disordered regions

and will help in the development of a more informative protein function predictor.

9. Linking Transcriptional Changes over Time in Stimulated Dendritic Cells to Identify Gene Networks Activated during the Innate Immune Response

Ashwini Patil, Kuo-Ching Liang, Yutaro Kumagai[1], Yutaka Suzuki[2], and Kenta Nakai

The innate immune response is primarily mediated by the Toll-like receptors functioning through the MyD88-dependent and TRIF-dependent pathways. Despite being widely studied, it is not yet completely understood and systems-level analyses have been lacking. In this study, we identified a high-probability network of genes activated during the innate immune response using a novel approach to analyze time-course gene expression profiles of activated immune cells in combination with a large gene regulatory and protein-protein interaction network. We classified the immune response into three consecutive time-dependent stages and identified the most probable paths between genes showing a significant change in expression at each stage. The resultant network contained several novel and known regulators of the innate immune response, many of which did not show any observable change in expression at the sampled time points. The response network shows the dominance of genes from specific functional classes during different stages of the immune response. It also suggests a role for the protein phosphatase 2a catalytic subunit a in the regulation of the immunoproteasome during the late phase of the response. In order to clarify the differences between the MyD88-dependent and TRIF-dependent pathways in the innate immune response, time-course gene expression profiles from MyD88-knockout and TRIF-knockout dendritic cells were analyzed. Their response networks suggest the dominance of the MyD88-dependent pathway in the innate immune response, and an association of the circadian regulators and immunoproteasomal degradation with the TRIF-dependent pathway. The response network presented here provides the most probable associations between genes expressed in the early and the late phases of the innate immune response, while taking into account the intermediate regulators. We propose that the method described here can also be used in the identification of time-dependent gene sub-networks in other biological systems.

10. Project for constructing an IT infrastructure for accelerating the clinical application of regenerative medicine technologies

Kenta Nakai, Wataru Fujibuchi[12], Kozo Imai[13], Tatsutoshi Nakahata[14], Norio Nakatsuji[15], Kohji Nishida[16], Hideyuki Okano[17], Masayo Takahashi[18], Akihiro Umezawa[19], Masayuki Yamato[18], Mihoko Adachi, Emi Ikeda, Takashi Kurosawa, Sung-Joon Park, and Yuki Saito: [12]CiRA, Kyoto U., [13]Res. Hospital, IMSUT, [14]IFMS, Kyoto U., [15]Grad. Sch. Med., Osaka U., [16]Sch. Med., Keio U., [17]CDB, RIKEN, [18]Nat. Res. Inst. Child Health and Development, [19]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. For this purpose, the basis of an "open innovation" environment allowing research institutes to continuously create innovative technologies is under development. This project is supported by the Ministry of Health, Labor and Welfare as an inter-ministry "Highway Construction Project for Realizing Regenerative Medicine".

Publications

Kusakabe, R., Tani, T., Nishitsuji, K., Shindo, M., Okamura, K., Miyamoto, Y., Nakai, K., Suzuki, Y., Kusakabe, T. G. and Inoue, K. Characterization of the compact bicistronic microRNA precursor, miR-1/miR-133, expressed specifically in *Ciona* muscle tissues, *Gene Expression Patterns*, 13 (1-2): 43-50, 2013.

Makita, Y. and Nakai, K. Chapter 8: *Bacillus subtilis* transcriptional network. In (M. Madan Babu ed.) *Bacterial Gene Regulation and Transcriptional Networks*. Edited by Babu, M.M. (Caister Academic Press). pp. 155-165, 2013.

Sasakura,Y., Sierro, N., Nakai, K., Inaba, K. and Kusakabe, T.G. Chapter 4: Genome structures, functional genomics, and proteomics in ascidians. In *Genome mapping and genomics in laboratory animals*. Edited by Denny, P. & Kole, C. (Springer) pp. 87-132, 2013.

López, Y., Patil, A., Nakai, K. Identification of novel motif patterns to decipher the promoter architecture of co-expressed genes in Arabidopsis thaliana. BMC Syst. Biol. 7(3): dS10, 2013.

Park, S.J., Komata, M., Inoue, F., Yamada, K., Nakai, K., Ohsugi, M., Shirahige, K. Inferring the choreography of parental genomes during fertilization from ultralarge-scale whole-transcriptome analysis. Genes Dev. 27(24): 2736-2748, 2013.

Patil, A., Liang, K., Kumagai, Y., Suzuki, Y., Nakai,

K. Linking Transcriptional Changes over Time in Stimulated Dendritic Cells to Identify Gene Networks Activated during the Innate Immune Response. PLoS Comput Biol 9(11): e1003323, 2013.
Elzawahry, A., Patil, A., Kumagai, Y., Suzuki, Y., Nakai, K. Innate immunity interactome dynamics. Gene Reg. and Systems Biol. 8: 1-15, 2014.
Liang, K., Suzuki, Y., Kumagai, Y., and Nakai, K. Analysis of changes in transcription start site distribution by a classification approach. Gene 537 (1): 29-40, 2014.
Sharma, A., Dehzangi, A., Lyons, J., Imoto, S., Miyano, S., Nakai, K., Patil, A. Evaluation of Sequence Features from Intrinsically Disordered Regions for the Estimation of Protein Function. PLoS ONE 9(2): e89890, 2014.
中井謙太. ENCODEプロジェクトの衝撃. 生物物理 53(5)：272–273, 2013.

Human Genome Center

Department of Public Policy
公共政策研究分野

Professor	Kaori Muto, Ph.D.	教　授	博士(保健学)	武　藤　香　織
Project Associate Professor	Ayako Kamisato, Ph.D.	特任准教授	博士(法学)	神　里　彩　子
Assistant Professor	Yusuke Inoue, Ph.D.	助　教	博士(社会医学)	井　上　悠　輔
Project Assistant Professor	Hyunsoo Hong, Ph.D.	特任助教	博士(学術)	洪　　賢　秀
Project Assistant Professor	Yuichi Maru, M.A.	特任助教	修士(法学)	丸　　祐　一

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 the construction of a human biological substances collection as well as to vaccination policy.

1. Research ethics consultation and studies on ethical, legal, and social implications on genomic medicine

We have been commissioned to provide research ethics consultation to several large projects promoting genomic medicine, including The Biobank Japan (BBJ) Project (BBJP) and Project for Development of Innovative Research on Cancer Therapeutics (P-DIRECT). BBJP is a disease-focused biobanking project started in 2003. BBJ consists of donated DNA, sera, and clinical information from 200,000 patients of 66 hospitals in Japan, and continues to collect new DNA and clinical information. Informed consent, which ensures the autonomous decisions of participants, is believed to be practically impossible for the biobanking project in general. Since 2008, we have issued semiannual newsletters for sample donors for transparency and information.

Figure 1. Examples of BBJP brochures we developed for sample donors and their physicians

P-DIRECT promotes strategic research and development (R & D) of the basic compounds (seeds) that contribute to development of next-generation innovative diagnostic techniques and new therapeutic agents incorporating basic research results. We have provided research ethics consultation for 34 research projects at 64 designated institutions. We have addressed a basic research ethics policy to use stored and newly acquired samples from cancer patients for this research project. We have also developed a research ethics management system and collected all protocols and consent forms. We have checked all documents and made sure that all institutional review board (IRB) approvals were acceptable. We held "forums for research ethics consultants" in February and September of 2013. To comply with the 2013 revision of Japanese government's "ethical guidelines on human genetic/genomic research," we have checked basic research ethics policies of both projects and developed new policies on incidental findings.

2. Research ethics consultation and studies on ethical, legal, and social implications of stem cell research

We were commissioned to provide research ethics consultation programming called "research on the ethical, legal, and social implications related to regenerative medicine" in 2012. In order to make regenerative medicine more concrete, it is essential to promote research development with a definite focus on clinical applications and to establish a framework for clinical research at an early stage. We needed to provide researchers at each institution with ethical support from the very beginning of their basic research and to have IRBs examine it thoroughly to make sure that clinical research was

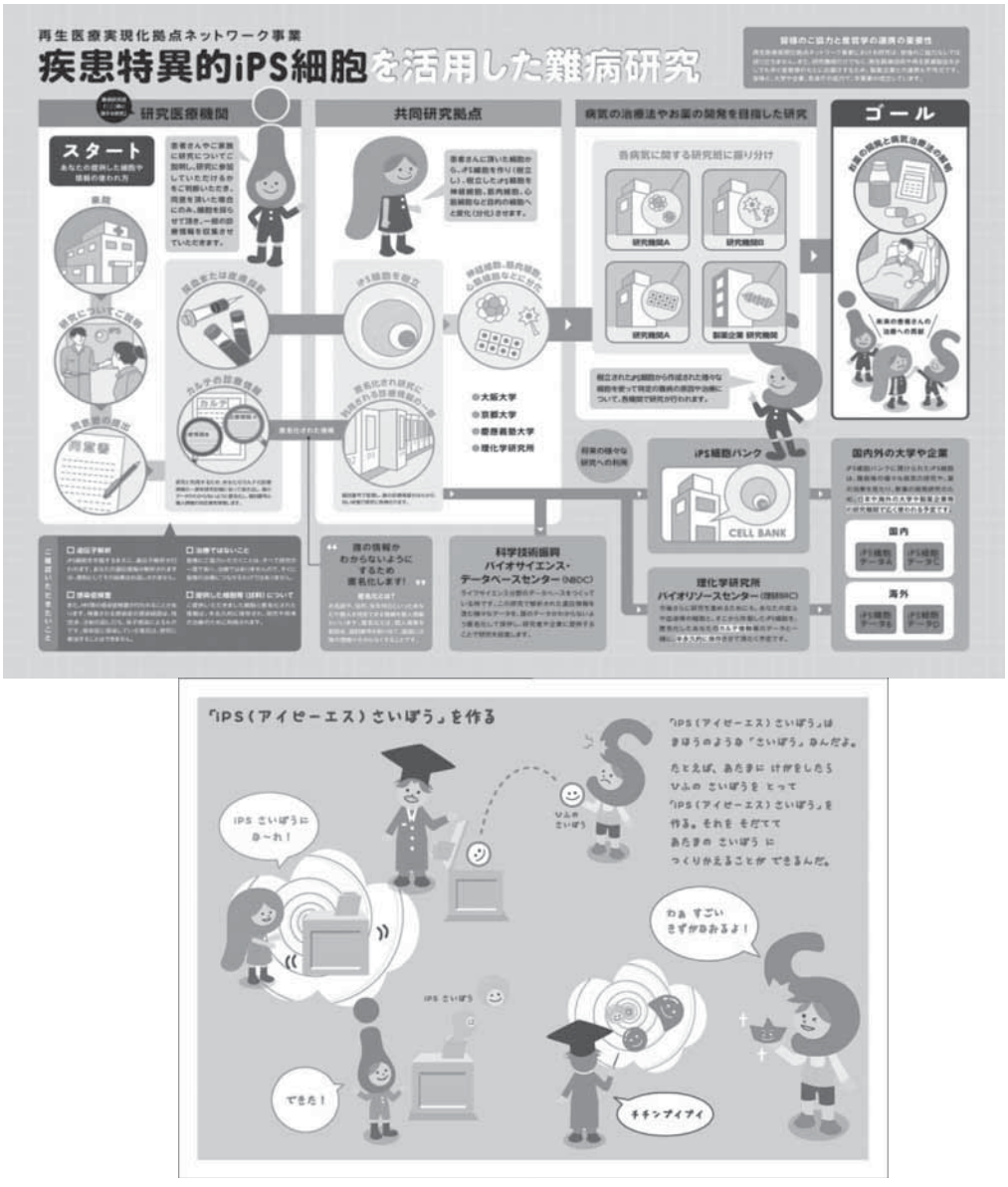

Figure 2. Examples of brochures and informed consent tools of the program for intractable diseases research using disease-specific induced pluripotent stem (iPS) cells

implemented in an appropriate manner. On the other hand, there are considerable differences among Japanese research institutions in terms of the quality of the ethical support and review systems; this is quite far from a standardized system. Moreover, there may be ethical issues unique to the field of regenerative medicine, but we are not sure of the nature of ethical support and review that is specific to this field should be. This program then offered concrete ethical support to those researchers, research facilities, and IRBs. Furthermore, we organized interdisciplinary research groups to address the ethical, legal, and social implications (ELSI) related to regenerative medicine in a comprehensive manner, with a view to establishing a framework for ethical support and review of regenerative medicine.

3. Patient perspectives and experiences of clinical trials in Japan

Clinical trials are at the heart of all medical advances. Achieving target sample sizes for clinical trials might be enhanced by understanding what is likely to motivate patients to participate, but therapeutic misconceptions must be avoided for ethical reasons. Substantial previous research suggested that people take part in clinical trials mostly for altruistic reasons, and that deriving personal benefit was a secondary consideration. However, interview studies of clinical trial participants in the UK (Locock and Smith 2010) and in Brazil (Nappo et al. 2013) showed that gaining some personal benefit has emerged as an important primary motivation, whereas altruistic considerations appeared to be largely subsidiary. Our study aimed to clarify the motivation for clinical trial participation in Japanese participants.

Recruitment was based on purposive maximum variation sampling. Our target sample size was 50. Sixteen patients (four male) have thus far participated in our in-depth, individual, narrative interviews in 2013. We explored patients' experiences of clinical trials, including informed consent, communication with healthcare professionals, and their thoughts, impressions, and motivations regarding participation. They were interviewed at home or in quiet meeting rooms. Interviews were either video or audio recorded. Before coding and thematic analysis, transcripts were returned to interviewees for corrections.

4. 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.
- Analysis of the roles of research coordinators for better recruitment and building trust with participants.
- Ethical, legal, and social implications of stem cell studies, including animal-human chimeric embryos and iPS cell banking.
- Ethics issues in collecting human biological substances for constructing a research infrastructure.

Publications

1. Sugimoto A, Nomura S, Tsubokura M, Matsumura T, Muto K, Sato M, Gilmour S: The relationship between media consumption and health-related anxieties after the Fukushima Daiichi nuclear disaster. PLOS ONE, 8(8): e65331, 2013. doi:10.1371/journal.pone.0065331
2. Forsberg J, Inoue Y. Beware side effects of research ethics revision. Science, 341(6152), 1341-1342, 2013.
3. Yasuda T, Ojima T, Nakamura M, Nagai A, Tanaka T, Kondo N, Suzuki K, Yamagata Z: Postpartum smoking relapse among women who quit during pregnancy: Cross-sectional study in Japan. The Journal of Obstetrics and Gynecology Research, 39(11): 1505-1512, 2013.
4. Tujimura T, Inoue Y, Yoshida K. Organ retention and communication of research use following medico-legal autopsy: A pilot survey of university of forensic medicine departments in Japan. Journal of Medical Ethics, in press, 2014.
5. 吉田邦広，大畑尚子，武藤香織，土屋敦，澤田甚一，狭間敬憲，池田修一，戸田達史：神経内科専門医の遺伝子診断に対する意識調査．臨床神経，53：337−344，2013.
6. 神里彩子．生命科学研究における成果発表の意義とその規制の許容性についての一考察命科学研究―2012年H5N1型インフルエンザウイルス研究論文問題を題材として―イ生命倫理　23(1)：105−114，2013.
7. 玉腰暁子・武藤香織．多施設共同疫学研究における中央事務局業務―実態の類型化と今後の標準化にむけて．日本公衆衛生学会誌　60(10)：631−638，2013.
8. 孫大鵬，安藤大輔，佐藤美理，鈴木孝太，田中太一郎，永井亜貴子，山縣然太朗：男子小中学生の踵骨の定量的超音波指標と体格・生活習慣因子の関連．山梨医科学雑誌　28(1)：39−47，2013.
9. 荒内貴子，井上悠輔，礒部太一，武藤香織「ゲノム解析技術の進展と課題：巨大化する医学・生命科学分野の技術」社会技術研究論文集，11，2014.