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	Yuichi Shiraishi, Ph.D	助　教	博士(統計科学)	白　石　友　一
Project Assistant Professor	Atsushi Niida, Ph.D	特任助教	博士(理学)	新井田　厚　司
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. Robust prediction of anti-cancer drug sensitivity and sensitivity-specific biomarker

Park H, Shimamura T[1], Miyano S, Imoto S.: [1]Graduate School of Medicine, Nagoya University

The personal genomics era has attracted a large amount of attention for anti-cancer therapy by patient-specific analysis. Patient-specific analysis enables discovery of individual genomic characteristics for each patient, and thus we can effectively predict individual genetic risk of disease and perform personalized anti-cancer therapy. Although the existing methods for patient-specific analysis have successfully uncovered crucial biomarkers, their performance takes a sudden turn for the worst in the presence of outliers, since the methods are based on non-robust manners. In practice, clinical and genomic alterations datasets usually contain outliers from various sources (e.g., experiment error, coding error, etc.) and the outliers may significantly affect the result of patient-specific analysis. We propose a robust methodology for patient-specific analysis in line with the NetwrokProfiler. In the proposed method, outliers in high dimensional gene expression levels and drug response datasets are simultaneously controlled by robust Mahalano-

bis distance in robust principal component space. Thus, we can effectively perform for predicting anti-cancer drug sensitivity and identifying sensitivity-specific biomarkers for individual patients. We observe through Monte Carlo simulations that the proposed robust method produces outstanding performances for predicting response variable in the presence of outliers. We also apply the proposed methodology to the Sanger dataset in order to uncover cancer biomarkers and predict anti-cancer drug sensitivity, and show the effectiveness of our method.

b. Sparse overlapping group lasso for integrative multi-omics analysis

Park H, Niida, A, Miyano S, Imoto S

Gene networks and graphs are crucial tools for understanding a heterogeneous system of cancer, since cancer is a disease that does not involve individual genes but combinations of genes associated with oncogenic process. A goal of genomic data analysis via gene networks is to identify both gene networks and individual genes within the selected networks. Existing methods, however, perform only network selection, and thus all genes in selected networks are included in models. This leads to overfitting when uncovering driver genes, and the results are not biologically interpretable. To accomplish both "groupwise sparsity" and "within group sparsity" for identifying driver genes based on biological knowledge (i.e., predefined overlapping groups of features), we propose a sparse overlapping group lasso via duplicated predictors in extended space. The proposed method effectively identifies driver genes and their interactions using known biological pathway information. Monte Carlo simulations and The Cancer Genome Atlas (TCGA) project data analysis indicate that the proposed method is effective for fitting a regression model (i.e., feature selection and prediction accuracy) constructed with duplicated predictors in overlapping groups. In the TCGA data analysis, we uncover potential cancer driver genes via expression modules and gene networks constructed by multi-omics data and identify that the uncovered genes have strong evidences as a cancer driver gene. The proposed method is a useful tool for identifying cancer driver genes and for integrative multi-omics analysis.

c. HapMuC: somatic mutation calling using heterozygous germline variants near candidate mutations

Usuyama N, Shiraishi Y, Sato Y[2,3], Kume H, Homma Y[3], Ogawa S[2], Miyano S, Imoto S: [2]Graduate School of Medicine, Kyoto University, [3]Graduate School of Medicine, The University of Tokyo

Identifying somatic changes from tumor and matched normal sequences has become a standard approach in cancer research. More specifically, this requires accurate detection of somatic point mutations with low allele frequencies in impure and heterogeneous cancer samples. Although haplotype phasing information derived by using heterozygous germ line variants near candidate mutations would improve accuracy, no somatic mutation caller that uses such information is currently available. We propose a Bayesian hierarchical method, termed HapMuC, in which power is increased by using available information on heterozygous germ line variants located near candidate mutations. We first constructed two generative models (the mutation model and the error model). In the generative models, we prepared candidate haplotypes, considering a heterozygous germ line variant if available, and the observed reads were realigned to the haplotypes. We then inferred the haplotype frequencies and computed the marginal likelihoods using a variational Bayesian algorithm. Finally, we derived a Bayes factor for evaluating the possibility of the existence of somatic mutations. We also demonstrated that our algorithm has superior specificity and sensitivity compared with existing methods, as determined based on a simulation, the TCGA Mutation Calling Benchmark 4 datasets and data from the COLO-829 cell line. The HapMuC source code is available from http://github.com/usuyama/hapmuc.

d. Inference of gene regulatory networks incorporating multi-source biological knowledge via a state space model with L1 regularization

Hasegawa T[4], Yamaguchi R, Nagasaki M[5], Miyano S, Imoto S: [4]Bioinformatics Center, Institute for Chemical Research, Kyoto University, [5]Tohoku Medical Megabank Organization, Tohoku University

Comprehensive understanding of gene regulatory networks (GRNs) is a major challenge in the field of systems biology. Currently, there are two main approaches in GRN analysis using time-course observation data, namely an ordinary differential equation (ODE)-based approach and a statistical model-based approach. The ODE-based approach can generate complex dynamics of GRNs according to biologically validated nonlinear models. However, it cannot be applied to ten or more genes to simultaneously estimate system dynamics and regulatory relationships due to the computational difficulties. The statistical model-based approach uses highly abstract models to simply describe bio-

logical systems and to infer relationships among several hundreds of genes from the data. However, the high abstraction generates false regulations that are not permitted biologically. Thus, when dealing with several tens of genes of which the relationships are partially known, a method that can infer regulatory relationships based on a model with low abstraction and that can emulate the dynamics of ODE-based models while incorporating prior knowledge is urgently required. To accomplish this, we propose a method for inference of GRNs using a state space representation of a vector auto-regressive (VAR) model with L1 regularization. This method can estimate the dynamic behavior of genes based on linear time-series modeling constructed from an ODE-based model and can infer the regulatory structure among several tens of genes maximizing prediction ability for the observational data. Furthermore, the method is capable of incorporating various types of existing biological knowledge, e.g., drug kinetics and literature-recorded pathways. The effectiveness of the proposed method is shown through a comparison of simulation studies with several previous methods. For an application example, we evaluated mRNA expression profiles over time upon corticosteroid stimulation in rats, thus incorporating corticosteroid kinetics/dynamics, literature-recorded pathways and transcription factor (TF) information.

e. An efficient method of exploring simulation models by assimilating literature and biological observational data

Hasegawa T[4], Nagasaki M[5], Yamaguchi R, Imoto S, Miyano S

Several biological simulation models of, e.g., gene regulatory networks and metabolic pathways, have been constructed based on existing knowledge of biomolecular reactions, e.g., DNA-protein and protein-protein interactions. However, since these do not always contain all necessary molecules and reactions, their simulation results can be inconsistent with observational data. Therefore, improvements in such simulation models are urgently required. A previously reported method created multiple candidate simulation models by partially modifying existing models. However, this approach was computationally costly and could not handle a large number of candidates that are required to find models whose simulation results are highly consistent with the data. In order to overcome the problem, we focused on the fact that the qualitative dynamics of simulation models are highly similar if they share a certain amount of regulatory structures. This indicates that better fitting candidates tend to share the basic regulatory structure of the best fitting candidate, which can best predict the data among candidates. Thus, instead of evaluating all candidates, we propose an efficient explorative method that can selectively and sequentially evaluate candidates based on the similarity of their regulatory structures. Furthermore, in estimating the parameter values of a candidate, e.g., synthesis and degradation rates of mRNA, for the data, those of the previously evaluated candidates can be utilized. The method is applied here to the pharmacogenomic pathways for corticosteroids in rats, using time-series microarray expression data. In the performance test, we succeeded in obtaining more than 80% of consistent solutions within 15% of the computational time as compared to the comprehensive evaluation. Then, we applied this approach to 142 literature-recorded simulation models of corticosteroid-induced genes, and consequently selected 134 newly constructed better models. The method described here was found to be capable of efficiently exploring candidate simulation models and obtaining better models within a short span of time. Furthermore, the results suggest that there may be room for improvement in literature recorded pathways and that they can be systematically updated using biological observational data

f. An efficient data assimilation schema for restoration and extension of gene regulatory networks using time-course observation data

Hasegawa T[4], Mori T[4], Yamaguchi R, Imoto S, Miyano S, Akutsu T[4]

Gene regulatory networks (GRNs) play a central role in sustaining complex biological systems in cells. Although we can construct GRNs by integrating biological interactions that have been recorded in literature, they can include suspicious data and a lack of information. Therefore, there has been an urgent need for an approach by which the validity of constructed networks can be evaluated; simulation-based methods have been applied in which biological observational data are assimilated. However, these methods apply nonlinear models that require high computational power to evaluate even one network consisting of only several genes. Therefore, to explore candidate networks whose simulation models can better predict the data by modifying and extending literature-based GRNs, an efficient and versatile method is urgently required. We applied a combinatorial transcription model, which can represent combinatorial regulatory effects of genes, as a biological simulation model, to reproduce the dynamic behavior of gene expressions within a state space model. Under the model, we applied the unscented Kalman filter to obtain the approximate posterior probability distribution of the hidden state to efficiently estimate parameter values maximizing prediction ability for observa-

tional data by the EM-algorithm. Utilizing the method, we propose a novel algorithm to modify GRNs reported in the literature so that their simulation models become consistent with observed data. The effectiveness of our approach was validated through comparison analysis to the previous methods using synthetic networks. Finally, as an application example, a Kyoto Encyclopedia of Genes and Genomes (KEGG)-based yeast cell cycle network was extended with additional candidate genes to better predict the real mRNA expressions data using the proposed method.

g. Lung adenocarcinoma subtypes definable by lung development-related miRNA expression profiles in association with clinicopathologic features

Arima C[1], Kajino T[1], Tamada Y, Imoto S, Shimada Y[1], Nakatochi M[1], Suzuki M[1], Isomura H[1], Yatabe Y[1], Yamaguchi T[1], Yanagisawa K[1], Miyano S, Takahashi T[1]

Accumulation of genetic and epigenetic changes alters regulation of a web of interconnected genes including microRNAs (miRNAs), which confer hallmark capabilities and characteristic cancer features. In this study, the miRNA and messenger RNA expression profiles of 126 non-small cell lung cancer specimens were analyzed, with special attention given to the diversity of lung adenocarcinomas. Of those, 76 adenocarcinomas were classified into two major subtypes, developing lung-like and adult lung-like, based on their distinctive miRNA expression profiles resembling those of either developing or adult lungs, respectively. A systems biology-based approach using a Bayesian network and nonparametric regression was employed to estimate the gene regulatory circuitry functioning in patient tumors in order to identify subnetworks enriched for genes with differential expression between the two major subtypes. miR-30d and miR-195, identified as hub genes in such subnetworks, had lower levels of expression in the developing lung-like subtype, whereas introduction of miR-30d or miR-195 into the lung cancer cell lines evoked shifts of messenger RNA expression profiles toward the adult lung-like subtype. Conversely, the influence of miR-30d and miR-195 was significantly different between the developing lung-like and adult lung-like subtypes in our analysis of the patient data set. In addition, RRM2, a child gene of the miR-30d-centered subnetwork, was found to be a direct target of miR-30d. Together, our findings reveal the existence of two miRNA expression profile-defined lung adenocarcinoma subtypes with distinctive clinicopathologic features and also suggest the usefulness of a systems biology-based approach to gain insight into the altered regulatory circuitry involved in cancer development.

h. Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers

Shiraishi Y, Fujimoto A[5], Furuta M[5], Tanaka H, Chiba K, Boroevich KA[5], Abe T[5], Kawakami Y[6], Ueno M[7], Gotoh K[8], Ariizumi S[9], Shibuya T, Nakano K[5], Sasaki A[5], Maejima K[5], Kitada R[5], Hayami S[7], Shigekawa Y[7], Marubashi S[8], Yamada T[8], Kubo M[5], Ishikawa O[8], Aikata H[6], Arihiro K[6], Ohdan H[6], Yamamoto M[9], Yamaue H[7], Chayama K[6], Tsunoda T[5], Miyano S, Nakagawa H[5]: [5]RIEKN Center for Integrative Medical Sciences, [6]Hiroshima University School of Medicine, [7]Wakayama Medical University, [8]Osaka Medical Center for Cancer and Cardiovascular Diseases, [9]Tokyo Women's Medical University

Recent studies applying high-throughput sequencing technologies have identified several recurrently mutated genes and pathways in multiple cancer genomes. However, transcriptional consequences from these genomic alterations in cancer genome remain unclear. In this study, we performed integrated and comparative analyses of whole genomes and transcriptomes of 22 hepatitis B virus (HBV)-related hepatocellular carcinomas (HCCs) and their matched controls. Comparison of whole genome sequence (WGS) and RNA-Seq revealed much evidence that various types of genomic mutations triggered diverse transcriptional changes. Not only splice-site mutations, but also silent mutations in coding regions, deep intronic mutations and structural changes caused splicing aberrations. HBV integrations generated diverse patterns of virus-human fusion transcripts depending on affected gene, such as TERT, CDK15, FN1 and MLL4. Structural variations could drive over-expression of genes such as WNT ligands, with/without creating gene fusions. Furthermore, by taking account of genomic mutations causing transcriptional aberrations, we could improve the sensitivity of deleterious mutation detection in known cancer driver genes (TP53, AXIN1, ARID2, RPS6KA3), and identified recurrent disruptions in putative cancer driver genes such as HNF4A, CPS1, TSC1 and THRAP3 in HCCs. These findings indicate genomic alterations in cancer genome have diverse transcriptomic effects, and integrated analysis of WGS and RNA-Seq can facilitate the interpretation of a large number of genomic alterations detected in cancer genome.

2. Statistical/Algorithmic Data Analysis Methods

a. Improving miRNA classification using an exhaustive set of features

ElGokhy SM[10], Shibuya T, Shoukry A[11]: [10]Egypt-Japan University of Science and Technology, [11]Alexandria University

MicroRNAs (miRNAs) are short (~22 nucleotides), endogenously-initiated non-coding RNAs that control gene expression post transcriptionally, either by the degradation of target miRNAs or by the inhibition of protein translation. The prediction of miRNA genes is a challenging problem towards the understanding of post transcriptional gene regulation. The present paper focuses on developing a computational method for the identification of miRNA precursors. We propose a machine learning algorithm based on Random Forests (RF) for miRNA prediction. The prediction algorithm relies on a set of features; compiled from known features as well as others introduced for the first time; that results in a performance that is better than most well-known miRNA classifiers. The method achieves 91.3% accuracy, 86% f-measure, 97.2% specificity, 93.4% precision and 79.6% sensitivity, when tested on real data. Our method succeeds in getting better results than MiPred (the best currently known RF algorithm in literature), Triplet-SVM and Virgo and EumiR. The obtained results indicate that Random Forests is a better alternative to Support Vector Machines (SVM) for miRNA prediction, especially from the point of view of accuracy and f-measure metrics.

b. A feature selection method using improved regularized linear discriminant analysis

Sharma A, Paliwal KK[12], Imoto S, Miyano S: [12]School of Engineering, Griffith University

Investigation of genes, using data analysis and computer-based methods, has gained widespread attention in solving human cancer classification problem. DNA microarray gene expression datasets are readily utilized for this purpose. In this paper, we propose a feature selection method using improved regularized linear discriminant analysis technique to select important genes, crucial for human cancer classification problem. The experiment is conducted on several DNA microarray gene expression datasets and promising results are obtained when compared with several other existing feature selection methods.

Publications

1. Chiba K, Shiraishi Y, Nagata Y, Yoshida K, Imoto S, Ogawa S, Miyano S. Genomon ITDetector: A tool for somatic internal tandem duplication detection from cancer genome sequencing data. *Bioinformatics*. 31(1), 116-118, 2015.
2. Fujimoto A, Furuta M, Shiraishi Y, Gotoh K, Kawakami Y, Arihiro K, Nakamura T, Ueno M, Ariizumi SI, Hai Nguyen H, Shigemizu D, Abe T, Boroevich KA, Nakano K, Sasaki A, Kitada R, Maejima K, Yamamoto Y, Tanaka H, Shibuya T, Shibata T, Ojima H, Shimada K, Hayami S, Shigekawa Y, Aikata H, Ohdan H, Marubashi S, Yamada T, Kubo M, Hirano S, Ishikawa O, Yamamoto M, Yamaue H, Chayama K, Miyano S, Tsunoda T, Nakagawa H. Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity. *Nat Commun*. 6: 6120, 2015.
3. Hoshino A, Nomura K, Hamashima T, Isobe T, Seki M, Hiwatari M, Yoshida K, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Takita J, Kanegane H. Aggressive transformation of anaplastic large cell lymphoma with increased number of ALK-translocated chromosomes. *Int J Hematol*. In press.
4. Ito H, Shiwaku H, Yoshida C, Homma H, Luo H, Chen X, Fujita K, Musante L, Fischer U, Frints SGM, Romano C, Ikeuchi Y, Shimamura T, Imoto S, Miyano S, Muramatsu S, Kawauchi T, Hoshino M, Sudol M, Arumughan A, Wanker EE, Rich T, Schwartz C, Matsuzaki F, Bonni A, Kalscheuer VM, Okazawa H. In utero gene therapy rescues microcephaly caused by Pqbp1-hypofunction in neural stem progenitor cells. *Mol Psychiatry*. In press.
5. Kawashima-Goto S, Imamura T, Seki M, Kato M, Yoshida K, Sugimoto A, Kaneda D, Fujiki A, Miyachi M, Nakatani T, Osone S, Ishida H, Taki T, Takita J, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Hosoi H. Identification of a homozygous JAK3 V674A mutation caused by acquired uniparental disomy in a relapsed early T-cell precursor ALL patient. *Int J Hematol*. In press.
6. Kurtovic-Kozaric A, Przychodzen B, Singh J, Konarska MM, Clemente MJ, Otrock ZK, Nakashima M, Hsi ED, Yoshida K, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Boultwood J, Makishima H, Maciejewski JP, Padgett RA. PRPF8 defects cause missplicing in myeloid malignancies. *Leukemia*. 29, 126-136, 2015.
7. Madan V, Kanojia D, Li J, Okamoto R, Sato-Otsubo A, Kohlmann A, Sanada M, Grossmann V, Sundaresan J, Shiraishi Y, Miyano S, Thol F, Ganser A, Yang H, Haferlach T, Ogawa S, Koef-

fler HP. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat Commun*. 6: 6042, 2015.
8. Park H, Niida, A, Miyano S, Imoto S. Sparse overlapping group lasso for integrative multi-omics analysis. *J Comput Biol*. In press.
9. Saini H, Raicar G, Lal S, Dehzangi A, Lyons J, Paliwal KK, Imoto S, Miyano S, Sharma A. Genetic algorithm for an optimized weighted voting scheme incorporating k-separated bigram transition probabilities to improve protein fold recognition. *IEEE Asia-Pacific World Congress on Computer Science and Engineering 2014*. In press.
10. Sung W-K, Sadakane K, Shibuya T, Belorkar A, Pyrogova I. An O(m log m)-time algorithm for detecting. *IEEE/ACM Trans Bioinf Comput Biol*. In press.
11. Tagawa K, Homma H., Saito A, Fujita K, Chen X, Imoto S, Oka T, Ito H, Motoki K, Yoshida C, Hatsuta H, Murayama S, Iwatsubo T, Miyano S, Okazawa H. Comprehensive phosphoproteome analysis unravels the core signaling network that initiates the earliest synapse pathology in preclinical Alzheimer's disease brain. *Hum Mol Genet*. 24(2), 540-558, 2015.
12. Takahashi Y, Sheridan P, Niida A, Sawada G, Uchi R, Mizuno H, Kurashige J, Sugimachi K, Sasaki S, Shimada Y, Hase K, Kusunoki M, Kudo S, Watanabe M, Yamada K, Sugihara K, Yamamoto H, Suzuki A, Doki Y, Miyano S, Mori M, Mimori K. The AURKA/TPX2 axis drives colon tumorigenesis cooperatively with MYC. *Ann Oncol*. In press.
13. Wang R, Yoshida K, Toki T, Sawada T, Uechi T, Okuno Y, Sato-Otsubo A, Kudo K, Kamimaki I, Kanezaki R, Shiraishi Y, Chiba K, Tanaka H, Terui K, Sato T, Iribe Y, Ohga S, Kuramitsu M, Hamaguchi I, Ohara A, Hara J, Goi K, Matsubara K, Koike K, Ishiguro A, Okamoto Y, Watanabe K, Kanno H, Kojima S, Miyano S, Kenmochi N, Ogawa S, Ito E. Loss of function mutations in RPL27 and RPS27 identified by whole-exome sequencing in Diamond-Blackfan anaemia. *Br J Haematol*. In press.
14. Yasuda T, Miyano S. Inferring the global structure of chromosomes from structural variations. *BMC Genomics*. In press.
15. Arima C, Kajino T, Tamada Y, Imoto S, Shimada Y, Nakatochi M, Suzuki M, Isomura H, Yatabe Y, Yamaguchi T, Yanagisawa K, Miyano S, Takahashi T. Lung adenocarcinoma subtypes definable by lung development-related miRNA expression profiles in association with clinicopathologic features. *Carcinogenesis*. 35(10): 2224-2231, 2014.
16. 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.
17. 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.
18. Damm F, Mylonas E, Cosson A, Yoshida K, Della Valle V, Mouly E, Diop M, Scourzic L, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Kikushige Y, Davi F, Lambert J, Gautheret D, Merle-Béral H, Sutton L, Dessen P, Solary E, Akashi K, Vainchenker W, Mercher T, Droin N, Ogawa S, Nguyen-Khac F, Bernard OA. Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov*. 4(9): 1088-1101, 2014.
19. ElGokhy SM, Shibuya T, Shoukry A. Improving miRNA classification using an exhaustive set of features. *Advances in Intelligent Systems and Computing*. 294: 31-39, 2014.
20. Fujita A, Miyano S. A tutorial to identify nonlinear associations in gene expression time series data. *Methods Mol Biol*. 1164: 87-95, 2014.
21. 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.
22. 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*. 340(1-2): 86-90, 2014.
23. Hasegawa T, Mori T, Yamaguchi R, Imoto S, Miyano S, Akutsu T. An efficient data assimilation schema for restoration and extension of gene regulatory networks using time-course observation data. *J Comput Biol*. 21(11): 785-798, 2014.
24. Hasegawa T, Nagasaki M, Yamaguchi R, Imoto S, Miyano S. An efficient method of exploring simulation models by assimilating literature and biological observational data. *Biosystems*. 121: 54-66, 2014.
25. Hasegawa T, Yamaguchi R, Nagasaki M, Miyano S, Imoto S. Inference of gene regulatory networks incorporating multi-source biological knowledge via a state space model with L1 regularization. *PLoS One*. 9(8): e105942, 2014.

26. 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*. 28(6): 1348-1351, 2014.
27. Katayama K, Yamaguchi R, Imoto S, Watanabe K, Miyano S. Analysis of questionnaire for traditional medicine and development of decision support system. *Evid Based Complement Alternat Med*. 2014: 974139, 2014.
28. 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*. 28(8): 1586-1595, 2014.
29. Li C, Nagasaki M, Ikeda E, Sekiya Y, Miyano S. CSML2SBML: a novel tool for converting quantitative biological pathway models from CSML into SBML. *Biosystems*. 121: 22-28, 2014.
30. 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.
31. Matsunawa M, Yamamoto R, Sanada M, Sato-Otsubo A, Shiozawa Y, Yoshida K, Otsu M, Shiraishi Y, Miyano S, Isono K, Koseki H, Nakauchi H, Ogawa S. Haploinsufficiency of Sf3b1 leads to compromised stem cell function but not to myelodysplasia. *Leukemia*. 28(9): 1844-1850, 2014.
32. Nishiura, H., Ejima, K., Mizumoto, K., Nakaoka, S., Inaba, H., Imoto, S., Yamaguchi, R., Saito, M.M. Cost-effective length and timing of school closure during an influenza pandemic depend on the severity. Theor Biol Med Model. 11: 5, 2014.
33. Ohtsuki T, Nariai N, Kojima K, Mimori T, Sato Y, Kawai Y, Yamaguchi-Kabata Y, Shibuya T, Nagasaki M. SVEM: a structural variant estimation method using multi-mapped reads on breakpoints. *Lecture Notes in Computer Science*. 8542: 208-219, 2014.
34. 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.
35. Park H, Shimamura T, Miyano S, Imoto S. Robust prediction of anti-cancer drug sensitivity and sensitivity-specific biomarker. *PLoS One*. 9 (1): e108990, 2014.
36. Saito MM, Imoto S, Yamaguchi R, Miyano S, Higuchi T. Parameter estimation in multi-compartment SIR model. *FUSION* 2014: 1-5, 2014.
37. 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. *Nat Genet*. 46(2): 171-175, 2014.
38. Sato Y, Maekawa S, Ishii R, Sanada M, Morikawa T, Shiraishi Y, Yoshida K, Nagata Y, Sato-Otsubo A, Yoshizato T, Suzuki H, Shiozawa Y, Kataoka K, Kon A, Aoki K, Chiba K, Tanaka H, Kume H, Miyano S, Fukayama M, Nureki O, Homma Y, Ogawa S. Recurrent somatic mutations underlie corticotropin-independent Cushing's syndrome. *Science*. 344(6186): 917-920, 2014.
39. Sawada G, Takahashi Y, Niida A, Shimamura T, Kurashige J, Matsumura T, Ueo H, Uchi R, Takano Y, Ueda M, Hirata H, Sakimura S, Shinden Y, Eguchi H, Sudo T, Sugimachi K, Miyano S, Doki Y, Mori M, Mimori K. Loss of CDCP1 expression promotes invasiveness and poor prognosis in esophageal squamous cell carcinoma. *Ann Surg Oncol*. 21 Suppl 4: 640-647, 2014.
40. Seki M, Yoshida K, Shiraishi Y, Shimamura T, Sato Y, Nishimura R, Okuno Y, Chiba K, Tanaka H, Kato K, Kato M, Hanada R, Nomura Y, Park MJ, Ishida T, Oka A, Igarashi T, Miyano S, Hayashi Y, Ogawa S, Takita J. Biallelic DICER1 mutations in sporadic pleuropulmonary blastoma. *Cancer Res*. 74(10): 2742-2749, 2014.
41. 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*. 18(1): 55-59, 2014.
42. 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.
43. Sharma A, Paliwal KK, Imoto S, Miyano S. A feature selection method using improved regu-

larized linear discriminant analysis. *Mach Vis Appl*. 25(3): 775-786, 2014.
44. Shen W, Clemente MJ, Hosono N, Yoshida K, Przychodzen B, Yoshizato T, Shiraishi Y, Miyano S, Ogawa S, Maciejewski JP, Makishima H. Deep sequencing reveals stepwise mutation acquisition in paroxysmal nocturnal hemoglobinuria. *J Clin Invest*. 124(10): 4529-4538, 2014.
45. Shiraishi Y, Fujimoto A, Furuta M, Tanaka H, Chiba K, Boroevich KA, Abe T, Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Shibuya T, Nakano K, Sasaki A, Maejima K, Kitada R, Hayami S, Shigekawa Y, Marubashi S, Yamada T, Kubo M, Ishikawa O, Aikata H, Arihiro K, Ohdan H, Yamamoto M, Yamaue H, Chayama K, Tsunoda T, Miyano S, Nakagawa H. Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. *PLoS One*. 9(12): e114263, 2014.
46. 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*. 21 Suppl 4: 515-521, 2014.
47. Takahashi R, Nagayama S, Furu M, Kajita Y, Jin Y-H, Kato T, Imoto S, Sakai Y, Toguchida J. AFAP1L1, a novel associating partner with vinculin, modulates cellular morphology and motility, and promotes the progression of colorectal cancers. *Cancer Medicine*. 3(4): 759-774, 2014.
48. Tokunaga H, Munakata K, Katayama K, Yamaguchi R, Imoto S, Miyano S, Watanabe K. Clinical data mining related to the Japanese kampo concept "hie" (oversensitivity to coldness) in men and pre- and postmenopausal women. *Evid Based Complement Alternat Med*. 2014: 832824, 2014.
49. Totoki Y, Yoshida A, Hosoda F, Nakamura H, Hama N, Ogura K, Yoshida A, Fujiwara T, Arai Y, Toguchida J, Tsuda H, Miyano S, Kawai A, Shibata T. Unique mutation portraits and frequent COL2A1 gene alteration in chondrosarcoma. *Genome Res*. 24(9): 1411-1420, 2014.
50. Usuyama N, Shiraishi Y, Sato Y, Kume H, Homma Y, Ogawa S, Miyano S, Imoto S. HapMuC: somatic mutation calling using heterozygous germline variants near candidate mutations. *Bioinformatics*. 30(23): 3302-3309, 2014.
51. 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.
52. Yoshino T, Katayama K, Munakata K, Horiba Y, Yamaguchi R, Imoto S, Miyano S, Watanabe K. Kampo traditional pattern diagnosis and the clustering analysis of patients with cold sensation. *J Altern Complement Med*. 20(5): A47, 2014.

Human Genome Center

Laboratory of Molecular Medicine
ゲノムシークエンス解析分野

Professor Tatsuhiro Shibata, M.D., Ph.D. 教 授 柴 田 龍 弘

The Laboratory of Molecular Medicine focuses on comprehensive characterization of currently-untreatable diseases including cancer on the basis of molecular genomics and aims to make "breakthroughs for human health" by identifying novel disease-related genes/pathways, including potential therapeutic or preventive targets and biomarkers, and to understand human diseases as heterogeneous but intervention-able "biological systems". This group has also organized the facility for the analysis of next-generation high-performance sequencers.

1. Establishment of the next-generation high performance sequencer platform at IMSUT

Tomoko Urushidate, Akiko Matsumoto, Azusa Sekine, Yasuko Katayama, Tatsuhiro Shibata

To realize individually tailored healthcare by enforcing the Japan BioBank Project at IMSUT, sequencing facilities including two Hiseq2500 high-performance sequencers have been set up in this laboratory. Data production and its quality from this platform have been evaluated.

2. Trans-ethnic landscape of liver cancer genomes

Yasushi Totoki[1,12], Kenji Tatsuno[2,12], Kyle R. Covington[3,12], Hiroki Ueda[2], Chad J. Creighton[3], Mamoru Kato[1], Shingo Tsuji[2], Lawrence A. Donehower[4], Betty L. Slagle[4], Hiromi Nakamura[1], Shogo Yamamoto[2], Eve Shinbrot[3], Natsuko Hama[1], Megan Lehmkuhl[3], Fumie Hosoda[1], Yasuhito Arai[1], Kim Walker[3], Mahmoud Dahdouli[3], Kengo Gotoh[2], Genta Nagae[2], Marie-Claude Gingras[3], Donna M. Muzny[3], Hidenori Ojima[5], Kazuaki Shimada[6], Yutaka Midorikawa[7], John A. Goss[8], Ronald Cotton[8], Akimasa Hayashi[2,9], Junji Shibahara[9], Shumpei Ishikawa[9], Jacfranz Guiteau[8], Mariko Tanaka[9], *Tomoko Urushidate*[1,11], Shoko Ohashi[1], Naoko Okada[1], Harsha Doddapaneni[3], Min Wang[3], Yiming Zhu[3], Huyen Dinh[3], Takuya Okusaka[10], Tomoo Kosuge[6], Tadatoshi Takayama[7], Masashi Fukayama[9], Richard A. Gibbs[3], David A. Wheeler[3], Hiroyuki Aburatani[2], Tatsuhiro Shibata[1,11]: [1]Division of Cancer Genomics, [5]Division of Molecular Pathology, National Cancer Center Research Institute, [2]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, [3]Human Genome Sequencing Center, [4]Department of Molecular Virology & Microbiology, [8]Department of Surgery, Baylor College of Medicine, [6]Hepatoboliary and Pancreatic Surgery Division, [10]Hepatoboliary and Pancreatic Oncology Division, National Cancer Center Hospital, [7]Department of Digestive Surgery, Nihon University School of Medicine, [9]Department of Pathology, Graduate School of Medicine, [11]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, the University of Tokyo.

Multiple etiological factors (hepatitis virus, alcohol, obesity etc.) are associated with the occurrence of liver cancer and their contributions diverse among ethnicity. To elucidate genetic diversities in liver cancer with regards to ethnic and epidemiological differences, we have conducted the trans-ethnic cancer genome research under the umbrella

of International Cancer Genome Consortium and The Cancer Genome Atlas. We performed whole exome sequencing and copy number analysis of 503 pairs of liver cancers, which include several ethnic populations with various etiological backgrounds. In total more than 100,000 somatic mutations (including non-coding region) were collected, and their signatures were more significantly associated with ethnic backgrounds but not with virus status. Aberration of the TERT pathway by various mechanisms (promoter/coding mutations, gene amplification and viral genome integration) was found in ~ 70% of cases, which should play a major role in liver cancer.

Publications

1. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, Tsuji S, Donehower LA, Slagle BL, Nakamura H, Yamamoto S, Shinbrot E, Hama N, Lehmkuhl M, Hosoda F, Arai Y, Walker K, Dahdouli M, Gotoh K, Nagae G, Gingras MC, Muzny DM, Ojima H, Shimada K, Midorikawa Y, Goss JA, Cotton R, Hayashi A, Shibahara J, Ishikawa S, Guiteau J, Tanaka M, Urushidate T, Ohashi S, Okada N, Doddapaneni H, Wang M, Zhu Y, Dinh H, Okusaka T, Kokudo N, Kosuge T, Takayama T, Fukayama M, Gibbs RA, Wheeler DA, Aburatani H, Shibata T. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet. 46: 1267-1273, 2014.

Human Genome Center

Laboratory of Genome Technology
シークエンス技術開発分野

Associate Professor	Koichi Matsuda, M.D., Ph.D.	准教授	松田浩一
Assistant Professor	Chizu Tanikawa, Ph.D.	助教	谷川千津
Project Assistant Professor	May Yamauchi, Ph.D.	特任助教	山内麻衣
Project Assistant Professor	Takafumi Miyamoto, Ph.D.	特任助教	宮本崇史
Project Assistant Professor	Makoto Hirata, M.D., Ph.D.	特任助教	平田真

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, Paulisally Hau Yi Lo, Mai Yamauchi, Takafumi Miyamoto, Makoto Hirata, Yuki Funauchi, Takashi Fujitomo, Jinichi Mori, Tomoyuki Koguchi, Yusuke Tsuda, Ryuta Yamamoto, Jiaying Lin, Zhenzhong Deng,.

1. Genome-wide association study

Genome-wide association study identified SNP on 15q24 associated with bladder cancer risk in Japanese population.

Through genome-wide association analysis and an independent replication study using a total of 1131 bladder cancer cases and 12 558 non-cancer controls of Japanese populations, we identified a susceptibility locus on chromosome 15q24. SNP rs 11543198 was associated with bladder cancer risk with odds ratio (OR) of 1.41 and P-value of $4.03 \times 10\text{-}9$. Subgroup analysis revealed rs11543198 to have a stronger effect in male smokers with OR of 1.66. SNP rs8041357, which is in complete linkage disequilibrium (r2=1) with rs11543198, was also associated with bladder cancer risk in Europeans (P= 0.045 for an additive and P=0.025 for a recessive model), despite much lower minor allele frequency in Europeans (3.7%) compared with the Japanese (22.2%). Imputational analysis in this region suggested CYP1A2, which metabolizes tobacco-derived carcinogen, as a causative candidate gene. We also confirmed the association of previously reported loci, namely SLC14A1, APOBEC3A, PSCA and MYC, with bladder cancer. Our finding implies the crucial roles of genetic variations on the chemically associated development of bladder cancer.

Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk.

Known genetic loci explain only a small proportion of the familial relative risk of colorectal cancer (CRC). We conducted a genome-wide association study of CRC in East Asians with 14,963 cases and 31,945 controls and identified 6 new loci associated with CRC risk ($P=3.42 \times 10(-8)$ to $9.22 \times 10(-21)$) at 10q22.3, 10q25.2, 11q12.2, 12p13.31, 17p13.3 and 19q13.2. Two of these loci map to genes (TCF7L2

and TGFB1) with established roles in colorectal tumorigenesis. Four other loci are located in or near genes involved in transcriptional regulation (ZMIZ 1), genome maintenance (FEN1), fatty acid metabolism (FADS1 and FADS2), cancer cell motility and metastasis (CD9), and cell growth and differentiation (NXN). We also found suggestive evidence for three additional loci associated with CRC risk near genome-wide significance at 8q24.11, 10q21.1 and 10q24.2. Furthermore, we replicated 22 previously reported CRC-associated loci. Our study provides insights into the genetic basis of CRC and suggests the involvement of new biological pathways.

Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1.

In a three-stage genome-wide association study among East Asian women including 22,780 cases and 24,181 controls, we identified 3 genetic loci newly associated with breast cancer risk, including rs4951011 at 1q32.1 (in intron 2 of the ZC3H11A gene; $P=8.82\times10(-9)$), rs10474352 at 5q14.3 (near the ARRDC3 gene; $P=1.67\times10(-9)$) and rs2290203 at 15q26.1 (in intron 14 of the PRC1 gene; $P=4.25\times10(-8)$). We replicated these associations in 16,003 cases and 41,335 controls of European ancestry (P= 0.030, 0.004 and 0.010, respectively). Data from the ENCODE Project suggest that variants rs4951011 and rs10474352 might be located in an enhancer region and transcription factor binding sites, respectively. This study provides additional insights into the genetics and biology of breast cancer.

Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians.

Genome-wide association studies (GWAS) of colorectal cancer (CRC) have been conducted primarily in European descendants. In a GWAS conducted in East Asians, we first analyzed approximately 1.7 million single-nucleotide polymorphisms (SNPs) in four studies with 1,773 CRC cases and 2,642 controls. We then selected 66 promising SNPs for replication and genotyped them in three independent studies with 3,612 cases and 3,523 controls. Five SNPs were further evaluated using data from four additional studies including up to 3,290 cases and 4,339 controls. SNP rs7229639 in the SMAD7 gene was found to be associated with CRC risk with an odds ratio (95% confidence interval) associated with the minor allele (A) of 1.22 (1.15 − 1.29) in the combined analysis of all 11 studies ($p=2.93\times10(-11)$). SNP rs7229639 is 2,487 bp upstream from rs4939827, a risk variant identified previously in a European-ancestry GWAS in relation to CRC risk. However, these two SNPs are not correlated in East Asians (r(2)=0.008) nor in Europeans (r(2)=0.146). The CRC association with rs7229639 remained statistically significant after adjusting for rs4939827 as well as three additional CRC risk variants (rs58920878, rs12953717 and rs4464148) reported previously in this region. SNPs rs7229639 and rs4939827 explained approximately 1% of the familial relative risk of CRC in East Asians. This study identifies a new CRC risk variant in the SMAD7 gene, further highlighting the significant role of this gene in the etiology of CRC.

2. Genes playing significant roles in human cancers

(1) p53 pathway

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.

Late Cornified Envelope Group I, a novel target of p53, regulates PRMT5 activity.

p53 is one of the most important tumor suppressor genes involved in human carcinogenesis. Although downstream targets of p53 and their biologic functions in cancer cells have been extensively investigated, it is still far from the full understanding. Here, we demonstrate that Late Cornified Envelope Group I (LCE1) genes, which are located in the LCE gene clusters encoding multiple well-conserved stratum-corneum proteins, are novel downstream targets of p53. Exogenous p53 overexpres-

sion using an adenoviral vector system significantly enhanced the expression of LCE1 cluster genes. We also observed induction of LCE1 expressions by DNA damage, which was caused by treatment with adriamycin or UV irradiation in a wild-type p53-dependent manner. Concordantly, the induction of LCE1 by DNA damage was significantly attenuated by the knockdown of p53. Among predicted p53-binding sites within the LCE1 gene cluster, we confirmed one site to be a p53-enhancer sequence by reporter assays. Furthermore, we identified LCE1 to interact with protein arginine methyltransferase 5 (PRMT5). Knockdown of LCE1 by specific small interfering RNAs significantly increased the symmetric dimethylation of histone H3 arginine 8, a substrate of PRMT5, and overexpression of LCE1F remarkably decreased its methylation level. Our data suggest that LCE1 is a novel p53 downstream target that can be directly transactivated by p53 and is likely to have tumor suppressor functions through modulation of the PRMT5 activity.

(2) lung cancer

Identification of a nuclear protein, LRRC42, involved in lung carcinogenesis.

On the basis of the gene expression profiles of 120 lung cancer cases using a cDNA microarray containing 27,648 genes or expressed sequence tags (ESTs), we identified LRRC42 (Leucine-rich repeat containing 42) to be significantly upregulated in the majority of lung cancers. Northern blot analysis demonstrated that LRRC42 was expressed only in testis among normal tissues examined. Knockdown of LRRC42 expression by siRNA against LRRC42 significantly suppressed the growth of lung cancer cells. On the other hand, stable induction of LRRC 42 expression significantly promoted cell growth. LRRC42, which was found to localize in the nucleus of mammalian cells, is likely to interact with and stabilize GATAD2B (GATA zinc finger domain-containing 2B) and MBD3 (Methyl-CpG-binding domain protein 3) proteins that could contribute to lung cancer cell proliferation partly through the regulation of p21Waf1/Cip1. Our findings suggest that LRRC42 overexpression as well as its interaction with LRRC42-GATAD2B might play essential roles in lung carcinogenesis, and be a promising molecular target for lung cancer therapy.

(3) Breast cancer

Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma.

Development of cancer is a complex process involving multiple genetic and epigenetic alterations. In our microarray analysis of 81 breast carcinoma specimens, we identified peptidase domain containing associated with muscle regeneration 1 (PAMR1) as being frequently suppressed in breast cancer tissues. PAMR1 expression was also reduced in all tested breast cancer cell lines, while PAMR1 was expressed moderately in normal breast tissues and primary mammary epithelial cells. DNA sequencing of the PAMR1 promoter after sodium bisulfite treatment revealed that CpG sites were hypermethylated in the breast cancer tissues and cell lines. PAMR1 expression was restored by 5-aza-2′ deoxycytidine treatment, demonstrating that promoter hypermethylation contributed to PAMR1 inactivation in the breast cancer cells. In addition, ectopic expression of PAMR1 markedly suppressed cancer cell growth. In summary, our study identified PAMR1 as a putative tumor suppressor which was frequently inactivated by promoter hypermethylation in breast cancer tissues.

Publications

1. 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. McCormick, 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.
2. J. Lin, Z. Deng, C. Tanikawa, T. Shuin, T. Miki, K. Matsuda, Y. Nakamura, Downregulation of the tumor suppressor HSPB7, involved in the p53 pathway, in renal cell carcinoma by hypermethylation, Int J Oncol, 44 (2014) 1490-1498.
3. Y. Yamamoto, M. Miyamoto, D. Tatsuda, M. Kubo, H. Nakagama, Y. Nakamura, H. Satoh, K. Matsuda, T. Watanabe, T. Ohta, A rare polymorphic variant of NBS1 reduces DNA repair activity and elevates chromosomal instability, Cancer Res, 74 (2014) 3707-3715.
4. 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, Interna-

tional journal of cancer. Journal international du cancer, 135 (2014) 948-955.
5. B. Zhang, W.H. Jia, K. Matsuda, S.S. Kweon, K. Matsuo, Y.B. Xiang, A. Shin, S.H. Jee, D.H. Kim, Q. Cai, J. Long, J. Shi, W. Wen, G. Yang, Y. Zhang, C. Li, B. Li, Y. Guo, Z. Ren, B.T. Ji, Z. Z. Pan, A. Takahashi, M.H. Shin, F. Matsuda, Y. T. Gao, J.H. Oh, S. Kim, Y.O. Ahn, A.T. Chan, J. Chang-Claude, M.L. Slattery, S.B. Gruber, F.R. Schumacher, S.L. Stenzel, G. Casey, H.R. Kim, J. Y. Jeong, J.W. Park, H.L. Li, S. Hosono, S.H. Cho, M. Kubo, X.O. Shu, Y.X. Zeng, W. Zheng, Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk, Nat Genet, 46 (2014) 533-542.
6. T. Fujitomo, Y. Daigo, K. Matsuda, K. Ueda, Y. Nakamura, Identification of a nuclear protein, LRRC42, involved in lung carcinogenesis, Int J Oncol, 45 (2014) 147-156.
7. M. Hirokawa, H. Morita, T. Tajima, A. Takahashi, K. Ashikawa, F. Miya, D. Shigemizu, K. Ozaki, Y. Sakata, D. Nakatani, S. Suna, Y. Imai, T. Tanaka, T. Tsunoda, K. Matsuda, T. Kadowaki, Y. Nakamura, R. Nagai, I. Komuro, M. Kubo, A genome-wide association study identifies PLCL2 and AP3D1-DOT1L-SF3A2 as new susceptibility loci for myocardial infarction in Japanese, Eur J Hum Genet, (2014).
8. Q. Cai, B. Zhang, H. Sung, S.K. Low, S.S. Kweon, W. Lu, J. Shi, J. Long, W. Wen, J.Y. Choi, D.Y. Noh, C.Y. Shen, K. Matsuo, S.H. Teo, M.K. Kim, U.S. Khoo, M. Iwasaki, M. Hartman, A. Takahashi, K. Ashikawa, K. Matsuda, M.H. Shin, M.H. Park, Y. Zheng, Y.B. Xiang, B.T. Ji, S.K. Park, P.E. Wu, C.N. Hsiung, H. Ito, Y. Kasuga, P. Kang, S. Mariapun, S.H. Ahn, H.S. Kang, K.Y. Chan, E.P. Man, H. Iwata, S. Tsugane, H. Miao, J. Liao, Y. Nakamura, M. Kubo, R.J. Delahanty, Y. Zhang, B. Li, C. Li, Y. T. Gao, X.O. Shu, D. Kang, W. Zheng, Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1, Nat Genet, 46 (2014) 886-890.
9. Z. Deng, K. Matsuda, C. Tanikawa, J. Lin, Y. Furukawa, R. Hamamoto, Y. Nakamura, Late Cornified Envelope Group I, a Novel Target of p53, Regulates PRMT5 Activity, Neoplasia, 16 (2014) 656-664.
10. M.F. Keller, A.P. Reiner, Y. Okada, F.J. van Rooij, A.D. Johnson, M.H. Chen, A.V. Smith, A. P. Morris, T. Tanaka, L. Ferrucci, A.B. Zonderman, G. Lettre, T. Harris, M. Garcia, S. Bandinelli, R. Qayyum, L.R. Yanek, D.M. Becker, L.C. Becker, C. Kooperberg, B. Keating, J. Reis, H. Tang, E. Boerwinkle, Y. Kamatani, K. Matsuda, N. Kamatani, Y. Nakamura, M. Kubo, S. Liu, A. Dehghan, J.F. Felix, A. Hofman, A.G. Uitterlinden, C.M. van Duijn, O.H. Franco, D.L. Longo, A.B. Singleton, B.M. Psaty, M.K. Evans, L.A. Cupples, J.I. Rotter, C.J. O'Donnell, A. Takahashi, J.G. Wilson, S.K. Ganesh, M.A. Nalls, C. for the Charge Hematology, G. BioBank Japan Project Working, Trans-ethnic meta-analysis of white blood cell phenotypes, Hum Mol Genet, (2014).
11. K. Matsuda, A. Takahashi, C.D. Middlebrooks, W. Obara, Y. Nasu, K. Inoue, K. Tamura, I. Yamasaki, Y. Naya, C. Tanikawa, R. Cui, J.D. Figueroa, D.T. Silverman, N. Rothman, M. Namiki, Y. Tomita, H. Nishiyama, K. Kohri, T. Deguchi, M. Nakagawa, M. Yokoyama, T. Miki, H. Kumon, T. Fujioka, L. Prokunina-Olsson, M. Kubo, Y. Nakamura, T. Shuin, Genome-wide association study identified SNP on 15q24 associated with bladder cancer risk in Japanese population, Hum Mol Genet, (2014).
12. P.H. Lo, C. Tanikawa, T. Katagiri, Y. Nakamura, K. Matsuda, Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma, Oncol Rep, 33 (2015) 267-273.

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	博士(理学)	中　井　謙　太
Senior Assistant Professor	Ashwini Patil, Ph.D.		講　師	博士(理学)	パティル, アシュウィニ
Project Senior Assistant Professor	Sung-Joon Park, Ph.D.		特任講師	博士(工学)	朴　　聖　俊

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. Modeling the promoter architecture of co-expressed genes

Yosvany López, Alexis Vandenbon[1] and Kenta Nakai: [1]WPI-iFReC, Osaka University

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. Thus, the identification and characterization of regulatory components has turned out to be valuable because the presence or absence of transcription factor binding sites (TFBSs) seems to be responsible for the complexity of gene regulation in every living organism. Holding the hypothesis that regulatory regions of genes with similar expression profiles might share common structural features, we have been attempting to explain the binding of transcription factors to promoters of genes expressed in specific tissues, cell types or physiological conditions. A collection of structural features consisting of positioning of motifs from transcription start sites, pairwise positioning, order and their orientation was proposed for promoters of antenna-expressed genes in *Drosophila melanogaster*. Such features were then used to score its entire promoter set and genes with highly scoring promoters were successfully validated with RNA-seq data from modENCODE project. Future work will focus on improving this computational approach in order to model the promoters of co-expressed genes in different developmental stages of *D. melanogaster* and *C. elegans*.

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: [2]Graduate School of Frontier Sciences, U. Tokyo

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. Discovery of differentially activated intermediary pathways in knockout experiments using sparse regression

Kuo-ching Liang, Ashwini Patil, and Kenta Nakai

Since its introduction, functional overrepresentation analysis has been an important tool for the analysis of gene sets. In overrepresentation analysis, the method for gene set construction is critical to finding biologically relevant information. Currently, these gene sets often come from clustering or differential expression analysis. However, naive application of these approaches often result in gene sets that encompass all genes with the same expression profile, or all genes that show differential expression in different conditions. We would like to propose a method to construct a more precise gene set. In particular, we want to be able to find shared neighbor genes between two pathways, and be able to compare them to detect differential intermediary pathway usage in different samples. In our proposed methodology, we use a model based on elastic-net regression to find neighborhood genes for genes in a given pathway. We compare the elastic-net regression approach to a correlation-distance-based approach using time-series gene knockout RNA-Seq data. We show gene sets constructed by the elastic-net regression approach contain genes that are more enriched with functional gene ontology terms that agree with previous research results compared to the correlation distance approach.

4. DBTMEE: a database of transcriptome in mouse early embryos

Sung-Joon Park, Katsuhiko Shirahige[3], Miho Ohsugi[4], Kenta Nakai: [3]IMCB, [4]Department of Life Sciences, U. Tokyo

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/.

5. 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.

6. Genome-wide identification and characterization of transcription start sites and promoters in the tunicate *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 tunicate *Ciona intestinalis*, an invertebrate chordate, has recently emerged as a powerful model organism for gene regulation analysis. However, few studies have been conducted to identify and characterize transcription start sites (TSSs) and promoters at the genome-wide level. In this study, we identify TSSs at the genome-wide scale using the TSS-seq method and characterize promoter regions in *C. intestinalis*. We identified TSS clusters (TSCs), which are high-density regions of TSS-seq tags. Although TSCs represent promoters in principle, we found several types of TSCs that were unlikely to represent promoters, such as AT-rich, 1-bp width TSCs. Reliable TSCs were found not only at known TSSs but also in other regions, suggesting the existence of many unknown transcription units in the genome. We identified accurate TSSs of 79 ribosomal protein genes and found that their promoters possessed a polypyrimidine initiator motif and a sharp TSS distribution, but they did not have TATA elements. In *C. intestinalis* promoters, two pyrimidine-purine dinucleotides, CA and TA, were the most frequent dinucleotides used as TSSs. Despite the absence of CpG islands, *Ciona* TATA-less promoters showed low expression specificity like CpG-associated human TATA-less promoters. Using multiple samples, TSS-seq allowed us to identify many novel candidate promoters of spliced leader *trans*-spliced genes for the first time. Furthermore, we identified many putative alternative promoters, some of which were regulated in a tissue-specific manner. Our results provide valuable information about TSSs and promoter characteristics in *C. intestinalis* and will be helpful in future analysis of transcriptional regulation in chordates.

7. 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.

8. 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 Euglen-

phyta 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.

9. A genetic algorithm for motif finding based on statistical significance

Josep Basha Gutierrez, Martin Frith[8], and Kenta Nakai: [8]Computational Biology Research Center, AIST Tokyo

Understanding of transcriptional regulation through the discovery of sequence elements that are bound by DNA-binding proteins is a fundamental problem in molecular biology research. These regulatory motifs, called transcription factor binding sites (TFBS), are short in length and can show sequence variation, which makes them especially difficult to identify. To facilitate the task of identifying these elements, different computational strategies have been developed. In this study, we propose a new computational method for motif discovery, by mixing a Genetic Algorithm structure with several statistical coefficients. The algorithm was tested with 56 data sets from four different species. The motifs obtained were compared to the known motifs for each one of the data sets, and the accuracy in this prediction compared to 14 other methods both at nucleotide and site level. The results, though did not stand out in detection of false positives, showed a remarkable performance in most of the cases in sensitivity and in overall performance at site level, generally outperforming the other methods in these three statistics, and suggesting that the algorithm can be a powerful tool to successfully predict motifs in different kinds of sets of DNA sequences.

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

Alok Sharma[9,10], Abdollah Dehzangi[10,11], James Lyons[12] Seiya Imoto, Satoru Miyano, Kenta Nakai, Ashwini Patil: [9]School of Engg. and Physics, The Univ. of the South Pacific, Suva, Fiji, [10]IIIS, Griffith Univ., [11]NICTA, [12]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. This project was funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

11. Identifying active gene sub-networks using time-course gene expression profiles using TimeXNet

Ashwini Patil and Kenta Nakai

Time-course gene expression profiles are frequently used to provide insight into the changes in cellular state over time and to infer the molecular pathways involved. When combined with large-scale molecular interaction networks, such data can provide information about the dynamics of cellular response to stimulus. However, few tools are currently available to predict a single active gene sub-network from time-course gene expression profiles. We introduce a tool, TimeXNet, which identifies active gene sub-networks with temporal paths using time-course gene expression profiles in the context of a weighted gene regulatory and protein-protein interaction network. TimeXNet uses a specialized form of the network flow optimization approach to identify the most probable paths connecting the genes with significant changes in expression at consecutive time intervals. TimeXNet has been extensively evaluated for its ability to predict novel regulators and their associated pathways within active gene sub-networks in the mouse innate immune response and the yeast osmotic stress response. Compared to other similar methods, TimeXNet identified up to 50% more novel regulators from independent experimental datasets. It predicted paths within a greater number of known pathways with longer overlaps (up to 7 consecutive edges) within these pathways. TimeXNet was also

shown to be robust in the presence of varying amounts of noise in the molecular interaction network. TimeXNet is a reliable tool that can be used to study cellular response to stimuli through the identification of time-dependent active gene subnetworks in diverse biological systems. It is significantly better than other similar tools. TimeXNet is implemented in Java as a stand-alone application and supported on Linux, MS Windows and Macintosh. The output of TimeXNet can be directly viewed in Cytoscape. This project is funded by the Japan Society for the Promotion of Science.

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

Kenta Nakai, Wataru Fujibuchi[13], Kozo Imai[14], Tatsutoshi Nakahata[13], Norio Nakatsuji[15], Kohji Nishida[16], Hideyuki Okano[17], Masayo Takahashi[18], Akihiro Umezawa[19], Masayuki Yamato[20], Mihoko Adachi, Emi Ikeda, Yusuke Komiyama, Takashi Kurosawa, Sung-Joon Park, and Yuki Saito: [13]CiRA, Kyoto U., [14]Res. Hospital, IMSUT, [15]IFMS, Kyoto U., [16]Grad. Sch. Med., Osaka U., [17]Sch. Med., Keio U., [18]CDB, RIKEN, [19]Nat. Res. Inst. Child Health and Development, [20]Tokyo Women's Med. U.

The level of research on regenerative medicine is quite high in Japan but the study is under severe international competition. One way to boost the research level is to share raw data of failed and/or unpublished experiments between top researchers in Japan. With such sharing, they may be able to avoid repeating unnecessary experiments or to find new value in unpublished data from a new perspective. In addition, applying data mining techniques across data will be likely to lead to new discoveries. Comparing data from different labs would be useful in establishing new guidelines for the quality control of stem cells and/or their derivatives. The aim of this project is, therefore, to build a collaborative platform among researchers using information and communications technology. We are developing a computer network which connects researchers securely as well as its accompanying software, while addressing various difficulties, including intellectual property-related, ethical, and information security-related ones, to achieve our goal. This project is supported by the Ministry of Health, Labor and Welfare.

Publications

López, Y., Vandenbon, A., Nakai, K. A Set of Structural Features Defines the Cis-Regulatory Modules of Antenna-Expressed Genes in *Drosophila melanogaster*. PLoS ONE 9(8): e104342, 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.

Park, S.J., Umemoto, T., Saito-Adachi, M., Shiratsuchi, Y., Yamato, M., Nakai, K. Computational Promoter Modeling Identifies the Modes of Transcriptional Regulation in Hematopoietic Stem Cells. PLoS ONE 9(4): e93853, 2014.

Elzawahry, A., Patil, A., Kumagai, Y., Suzuki, Y., Nakai, K. Innate immunity interactome dynamics. Gene Reg. and Systems Biol. 8: 1-15, 2014.

Lim, K., Kobayashi, I., and Nakai, K. Alterations in rRNA-mRNA interaction during plastid evolution, Mol. Biol. Evol., 31(7): 1728-1740, 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.

Patil, A. and Nakai, K. TimeXNet: Identifying active gene sub-networks using time-course gene expression profiles, BMC Syst. Biol., 8 (Suppl.4): S2, 2014.

Firouzi, A., López, Y., Suzuki, Y., Nakai, K., Sugano, S., Yamochi, T. and Watanabe, T. Development and validation of a new high-throughput method to investigate the clonality of HTLV-1-infected cells based on provirus integration sites, *Genome Medicine*, 6: 46, 2014.

Park, S.J., Shirahige, K., Ohsugi, M., Nakai, K. DBTMEE: a database of transcriptome in mouse early embryos. Nucl. Acids Res. 43: D772-777, 2015.

Suzuki, A., Wakaguri, H., Yamashita, R., Kawano, S., Tsuchihara, K., Sugano, S., Suzuki, Y. and Nakai, K. DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucl. Acids Res.* 43: D87-91, 2015.

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 contributes to achieve three major missions: public policy science studies of translational research 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.

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 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. We have issued semiannual newsletters for sample donors for transparency and information (Fig. 1). Since 2014, BBJ has started to store tissue samples from cancer patients and to share DNA samples from participants of several clinical trials. We have supported for its

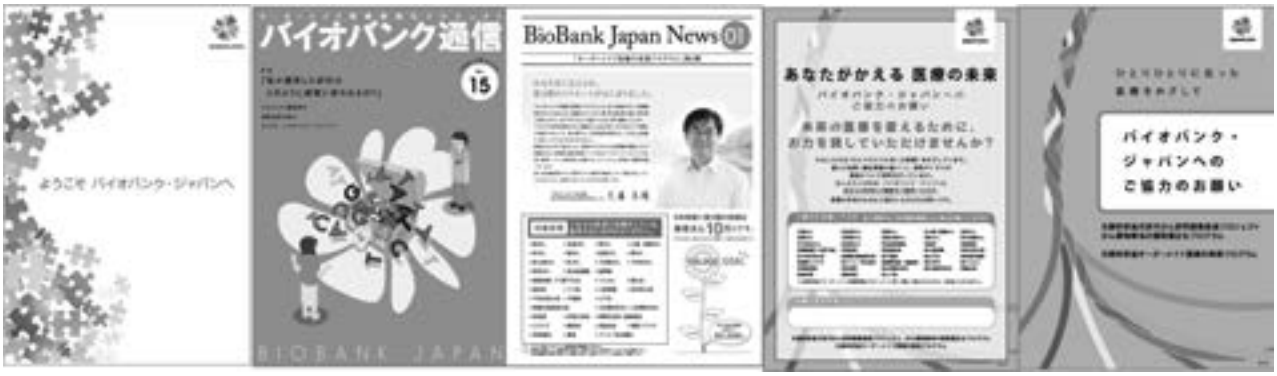

Figure 1. Examples of BBJP brochures we developed for sample donors and their physicians

launch from ethical perspectives.

On the other hands, 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 provided research ethics consultation for 34 research projects at 64 designated institutions. Since 2014, P-DIRECT welcomed new 20 research projects. We have advised those principal investigators to comply with ethical guidelines and provided sample 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 September of 2014.

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

We have been commissioned to provide research ethics consultation called "research on the ethical, legal, and social implications related to regenerative medicine" since 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. In 2014, we provided more than 60 consultations for stem cell researchers. Topics of those consultations include research design, informed consent, IRBs, return of research results, inclusion criterion of participants of first-in-human trials, governance of iPSC banking and communication with patient advocacy groups (Fig. 2).

We also provided educational opportunities for researchers who are in charge of obtaining informed consent. As identified by participants at the workshop, one of the key challenges with obtaining true informed consent was ensuring equitable subject selection, particularly for potentially vulnerable patients enrolled in clinical trials using stem cells ("stable ethics"). Other issues identified include ensuring that study volunteers understand that they can withdraw from the study at any time, and

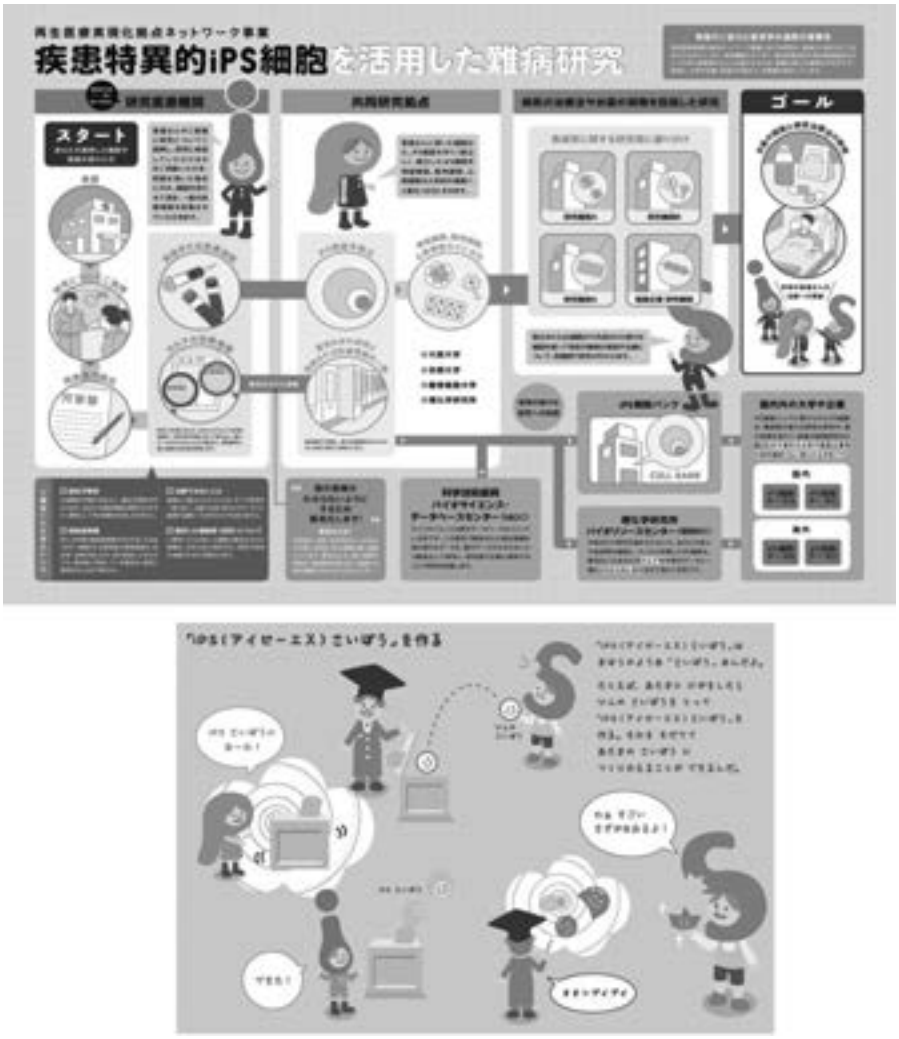

Figure 2. Examples of brochures and informed consent tools of the program for intractable diseases research using disease-specific induced pluripotent stem (iPS) cells

monitoring and ensuring the long-term safety of patients. Most important, particularly for studies involving cutting-edge interventions like the use of stem cells, researchers must understand that they cannot always adequately address patient questions and concerns during the initial recruitment and consent process, so must continue to engage study volunteers in discussions of trial design, conduct and outcomes throughout the research process.

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. Thus, we have conducted a quantitative and a qualitative study about clinical trial participants.

Our quantitative study is an online survey on clinical trials. Invitation was sent to 21,502 adult patients (> 20 year-old) in Japan. The questionnaire included items relating to knowledge, informed consent, the decision to participate, risk and benefit of clinical trials, and information after the trials. Of the 21,502, 12,506 responses were analyzed. The study periods were from March 26 to March 28, 2014. Mean age of the respondents was 52 (range, 20 to 79). Men and women were 49.7% and 50.3%, respectively. Most of them (89.5%, n = 11,197) had heard or seen the word "clinical trial" (including "CHIKEN") via television (57.3%), internet (52.7%), newspaper/magazine/advertisement (43.8%), or other information sources. "Randomization" and "placebo" were poorly understood (25.5% and 56.1 %, respectively) while most of them (87.4%) understood that clinical trials were conducted to establish evidence for future medicine. Of the 11,197, 8.7% (n = 967) had enrolled a clinical trial, and others (77.4%) had never involved any clinical trials. The major motivations for enrolling were "contributing to develop new therapies for my disease" (44.2%), "contributing to the advances in medicine" (42.7%), "potential benefit for my condition" (37.3%), and "receiving a new therapy" (32.4%). Among the enrollee who withdraw the trials according to researchers' convenience (n = 52), more enrollees were motivated by personal benefit, including "potential benefit for my condition" (55.8%) or "receiving a new therapy" (40.4%). Respondents with no experience of clinical trials (n = 8877) recognized the benefits; "potential benefit for my condition" (59.0 %), "contributing to develop new therapies for my disease" (54.0%), "contributing to the advances in medicine" (53.1%), and "receiving a new therapy" (59.0%). Fewer respondents (26.2%) recognized "payment for the burden" than experienced respondents (37.1%). Most of the experienced and non-experienced patients (54.7% and 87.7%, respectively) recognized "adverse event" as a risk of clinical trials. The continuous outreach efforts by government, hospitals, and research institutions have achieved a certain result in patients' understanding of clinical trials. Experiences may have some impacts on the patients' perspective of clinical trials; understanding of "randomization" and "placebo", and the degree of expectation of personal and therapeutic benefits for clinical trials. As financial issues on clinical trials widely differ with types of trials, it should be discussed according to each case. Patients' perspective on clinical trials would depend on their condition and trial types. Further investigation should be conducted at this point.

On the other hand, our qualitative study aims to clarify the motivation for clinical trial participation in Japanese participants. Recruitment was based on purposive maximum variation sampling. Recruitment packs (containing the participant information sheet and a leaflet about the study) were distributed through research networks, patient and public involvement groups, trial clinics and researchers, websites, and other media. Our target sample size was 50. Twenty-one patients have thus far participated in our in-depth, individual narrative interviews. We explored patients' experiences of clinical trials, including informed consent, communications with healthcare professionals, as well as 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. As shown by the previous study in the UK, we observed a variety of reasons for participating in clinical trials to gain personal benefits, including increased survival, access to new treatment, care from an expert medical team, and financial compensation. Some interviewees vaguely expressed that their participation in clinical trials might have improved their symptoms. Others believed that their physicians would not recommend

clinical trials if they did not benefit their health. Wanting to benefit others was mentioned by just two interviewees as feedback to the patient involvement group and gratitude for the hospital. Among several motivations for gaining personal benefits, we observed therapeutic misconception. Two altruistic reasons emerged for "visible" others in interviewees' personal communities, whereas participants in the previous study in the UK showed motivation for "invisible" others, that is, acknowledgement to past generations, benefiting future generations, moral duty, and furthering medical knowledge for the common good. If Japanese citizens understood the nature of clinical trials as an essential tool for the development of new drugs or therapies for future generations, the reasons people give for taking part in clinical trials for benefitting others should be deeper and more complex. Further interviews should be conducted.

4. DTC genetic testing issues

With the development of genome sequencing technology, direct-to-consumer (DTC) genetic testing services have begun to be used widely. In Japan, genetic-medicine-related societies develop guidelines for genetic testing and issue statements concerning DTC genetic testing. In this context, the Ministry of Economy, Trade and Industry (METI) considered the regulations governing DTC genetic testing and attached importance to quality control, scientific evidence, and informed consent. In particular, the METI has endeavored to set regulations, which recommend against the sale of scientifically groundless tests. In other countries, there have been reports on opinion surveys of the scientific evidence of genetic testing and the regulations of genetic testing, however the attitudes held by the Japanese public toward scientific evidence in support of genetic testing are unclear. Our aim was to clarify public attitudes toward genetic testing and related scientific evidence. In March 2014, an anonymous online survey was administrated to 24,718 men and women aged 20-69 years in Japan. The questionnaire included questions concerning genetic knowledge, attitudes toward genetic testing, the regulations governing DTC genetic testing, and technology. There were 4 response options used to indicate attitudes toward scientific evidence supporting genetic testing. The options were "even if there is little scientific evidence to support genetic testing, I would like to undergo the tests," "if there is scientific evidence in support of genetic testing for people of different races, I would like to undergo the tests," "if there is scientific evidence in support of genetic testing for people of East Asian ethnicity, I would like to undergo the tests," and "even if there is sufficient scientific evidence to support genetic testing, I would not like to undergo the tests." A total of 7,540 individuals completed surveys. The mean age of the respondents was 45.7 ± 14.0 years. With respect to questions regarding willingness to undergo 6 types of genetic testing by scientific evidence, around 50% of respondents reported that they could not decide whether to undergo any type of genetic testing. Our results suggest that members of the Japanese public have positive attitudes toward genetic testing, which provides useful information concerning the prevention and treatment of disease, and may attach less importance to scientific evidence supporting genetic testing than academic societies and the METI do. In addition, our results suggest the possibility that there is a relation between the attributes of the respondents and the attitudes toward scientific evidence of genetic testing. Further analyses are needed to clarify these associations.

5. Incidental findings

To establish appropriate measures that deal with incidental findings (IFs), the neuroscience community needs to address various ethical issues. The current state of research facilities regarding IFs and investigator attitudes as well as potentially eligible research participants must be assessed prior to future discussions and before the development of policies and guidelines. To this end, we conducted two questionnaire surveys to clarify i) how IFs are addressed at neuroimaging research facilities in Japan and ii) the views of investigators and potential research participants regarding the handling of IFs. Thirty-one principal investigators (PIs) involved in the Strategic Research Program for Brain Sciences (SRPBS), a government-funded project, were asked to fill out a questionnaire regarding ways IFs were handled at the facility. A total of 110 investigators engaged in SRPBS tasks, including 31 PIs who participated in the research facility survey and researchers conducting studies under the management of the PIs, and 500 individuals from the general public (i.e., general population) were asked to select the most appropriate way to deal with IFs in two scenarios, namely the medical school and humanities and social sciences department scenarios. More than 40% of PIs responded that they did not know or were unsure of what type of approach was employed to handle IFs at their research facilities. Nevertheless, they were willing to improve the current status if sufficient resources were provided. With regard to specialist involvement, 37.7% of investigators responded that it was appropriate to have a specialist check all images in the medical school scenario, whereas 13.3% responded that such involvement was appropriate in the humanities and social sciences department scenario. In contrast, 76.1% and 61.0% of the general population indicated that specialist involvement was appropri-

ate in the medical school and humanities and social sciences department scenarios, respectively. These results show that expectations of the general population exceed those of investigators regarding measures to address IFs. Both investigators and the general population demanded more responsibility from PIs at medical institutions, compared to PIs at non-medical institutions. Based on our preliminary results, we recommended that a licensed physician perform a screening test to appropriately examine clear abnormalities. These recommendations were implemented by the SRPBS as guidelines for handling IFs in national research projects in Japan.

Publications

1. Fujita M, Hayashi Y, Tashiro S, Takashima K, Nakazawa E, Akabayashi A. Handling incidental findings in neuroimaging research in Japan: current state of research facilities and attitudes of investigators and the general population. Health research policy and systems. 2014; 12: 58. doi: 10.1186/1478-4505-12-58.
2. Nakada H, Tsubokura M, Kishi Y, Yuji K, Matsumura T, Kami M. How do medical journalists treat cancer-related issues? Ecancermedicalscience. in press, 2014.
3. Nakada H, Yuji K, Tsubokura M, Ohsawa Y, Kami M. Development of a National Agreement on Human Papillomavirus Vaccination in Japan: An Infodemiology Study. J Med Internet Res. 16(5) e129, 2014.
4. Nakada H, Akagawa E, Tsukiya T, Mizuno T, Takewa Y, Tatsumi E. Establishment of the guideline for innovative medical devices: collaboration with academia, clinicians, industry and regulatory agency. Int J Arif Organd 2014; 37(8): 603.
5. Akagawa E, Tatsumi E, Nakada H, Ootou K, Hasegawa S, Taenaka Y. Development and implementation of evaluation tool for intellectual properties created in the study of medical devices. Int J Arif Organd 2014; 37(8): 638.
6. 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―，社会技術研究論集，11：138–148，2014.
7. 文部科学省「オーダーメイド医療の実現プログラム」編．ヒトゲノム解析研究と医療をつなぐ倫理的配慮，2014.
8. 武藤香織．臨床試験への患者・市民参画(PPI)とは何か，医薬ジャーナル，50(8)：93–98，2014.
9. 井上悠輔．臨床研究の不正と医師の「誠実さ」，年報医事法学，29：196–202，2014.
10. 井上悠輔．世界医師会のヘルシンキ宣言と2013年のフォルタレザ改訂，医薬ジャーナル，50(8)：55–62，2014.
11. 井上悠輔．ヒトゲノム解析に関する倫理指針の改正，Organ Biology，21(1)：24–32，2014.
12. 磯部哲，田代志門，井上悠輔ほか．「臨床研究に関する欧米諸国と我が国の規制・法制度の比較研究」．厚生労働省医療技術実用化総合研究事業(2014年度)
13. 井上悠輔．臨床研究の不正と医師の資質，厚生労働省地域医療基盤開発推進研究事業(医療事故におけるJust Culture(正義・公正の文化)を支える法制度の構築を目指して―医療事故の原因分析・再発防止推進のための法制度)(2014年)
14. 丸祐一．医療機器の臨床試験にかかる同意取得と撤回―研究倫理の立場から，薬理と治療，42(4)：232–242，2014.
15. 丸祐一．医学研究における遺伝情報の活用と倫理的課題，産科と婦人科，81(6)：753–758，2014.
16. 松井健志，會澤久仁子，丸祐一，栗原千絵子，齊尾武郎．論文非掲載：学術誌の公正性と倫理性をめぐる問題事例の検討，臨床評価，42(2)：513–517，2014.