Human Genome Center

Laboratory of DNA Information Analysis
Laboratory of Sequence 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, PhD	教　授	理学博士	宮　野	悟
Associate Professor	Rui Yamaguchi, PhD	准教授	博士(理学)	山　口	類
Assistant Professor	Yuichi Shiraishi, PhD	助　教	博士(統計科学)	白　石　友	一
Associate Professor	Tetsuo Shibuya, PhD	准教授	博士(理学)	渋　谷　哲	朗
Assistant Professor	Kotoe Katayama, PhD	助　教	博士(工学)	片　山　琴	絵
Project Senior Assistant Professor	Yoshinori Tamada	特任講師	博士(情報学)	玉　田　嘉	紀
Project Assistant Professor	Taku Onodera	特任助教	博士(情報理工学)	小野寺	拓

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. Interaction-based feature selection for uncovering cancer driver genes through copy number-driven expression level

Park H[1], Niida A[2], Imoto S[2], Miyano S: [1]Yamaguchi University, [2]Health Intelligence Center

Driver gene selection is crucial to understand the heterogeneous system of cancer. To identity cancer driver genes, various statistical strategies have been proposed, especially the L_1-type regularization methods have drawn a large amount of attention. However, the statistical approaches have been developed purely from algorithmic and statistical point, and the existing studies have applied the statistical approaches to genomic data analysis without consideration of biological knowledge. We consider a statistical strategy incorporating biological knowledge to identify cancer driver gene. The alterations of copy number have been considered to driver cancer pathogenesis processes, and the region of strong interaction of copy number alterations and expression levels was known as a tumor-related symptom. We incorporate the influence of copy number alterations on expression levels to

cancer driver gene-selection processes. To quantify the dependence of copy number alterations on expression levels, we consider *cis-* and *trans-* effects of copy number alterations on expression levels of genes, and incorporate the symptom of tumor pathogenesis to gene-selection procedures. We then proposed an interaction-based feature-selection strategy based on the adaptive L_1-type regularization and random lasso procedures. The proposed method imposes a large amount of penalty on genes corresponding to a low dependency of the two features, thus the coefficients of the genes are estimated to be small or exactly 0. It implies that the proposed method can provide biologically relevant results in cancer driver gene selection. Monte Carlo simulations and analysis of the Cancer Genome Atlas (TCGA) data show that the proposed strategy is effective for high-dimensional genomic data analysis. Furthermore, the proposed method provides reliable and biologically relevant results for cancer driver gene selection in TCGA data analysis.

b. A novel adaptive penalized logistic regression for uncovering biomarker associated with anti-cancer drug sensitivity

Park H[1], Shiraishi Y, Imoto S[2], Miyano S

We developed a novel adaptive penalized logistic regression modeling strategy based on Wilcoxon rank sum test (WRST) to effectively uncover driver genes in classification. In order to incorporate significance of gene in classification, we first measure significance of each gene by gene ranking method based on WRST, and then the adaptive L_1-type penalty is discriminately imposed on each gene depending on the measured importance degree of gene. The incorporating significance of genes into adaptive logistic regression enables us to impose a large amount of penalty on low ranking genes, and thus noise genes are easily deleted from the model and we can effectively identify driver genes. Monte Carlo experiments and real world example are conducted to investigate effectiveness of the proposed approach. In Sanger data analysis, we introduce a strategy to identify expression modules indicating gene regulatory mechanisms via the principal component analysis (PCA), and perform logistic regression modeling based on not a single gene but gene expression modules. We can see through Monte Carlo experiments and real world example that the proposed adaptive penalized logistic regression outperforms feature selection and classification compared with existing L_1-type regularization. The discriminately imposed penalty based on WRST effectively performs crucial gene selection, and thus our method can improve classification accuracy without interruption of noise genes. Furthermore, it can be seen through Sanger data analysis that the method for gene expression modules based on principal components and their loading scores provides interpretable results in biological viewpoints.

c. OVarCall: Bayesian mutation calling method utilizing overlapping paired-end reads

Moriyama T, Shiraishi Y, Chiba K, Yamaguchi R, Imoto S[2], Miyano S

Detection of somatic mutations from tumor and matched normal sequencing data has become a standard approach in cancer research. Although a number of mutation callers are developed, it is still difficult to detect mutations with low allele frequency even in exome sequencing. We expect that overlapping paired-end read information is effective for this purpose, but no mutation caller has modeled overlapping information statistically in a proper form in exome sequence data. Here, we develop a Bayesian hierarchical method, OVarCall, where overlapping paired-end read information improves the accuracy of low allele frequency mutation detection. Firstly, we construct two generative models: one is for reads with somatic variants generated from tumor cells and the other is for reads that does not have somatic variants but potentially includes sequence errors. Secondly, we calculate marginal likelihood for each model using a variational Bayesian algorithm to compute Bayes factor for the detection of somatic mutations. We empirically evaluated the performance of OVarCall and confirmed its better performance than other existing methods.

d. D[3]M: detection of differential distributions of methylation levels

Matsui Y[3], Mizuta M[3], Ito S[3], Miyano S, Shimamura T[3]: [3]Nagoya University

DNA methylation is an important epigenetic modification related to a variety of diseases including cancers. We focus on the methylation data from Illumina's Infinium HumanMethylation450 BeadChip. One of the key issues of methylation analysis is to detect the differential methylation sites between case and control groups. Previous approaches describe data with simple summary statistics or kernel function, and then use statistical tests to determine the difference. However, a summary statistics-based approach cannot capture complicated underlying structure, and a kernel function-based approach lacks interpretability of results.

We propose a novel method D[3]M, for detection of differential distribution of methylation, based on distribution-valued data. Our method can detect the differences in high-order moments, such as

shapes of underlying distributions in methylation profiles, based on the Wasserstein metric. We test the significance of the difference between case and control groups and provide an interpretable summary of the results. The simulation results show that the proposed method achieves promising accuracy and shows favorable results compared with previous methods. Glioblastoma multiforme and lower grade glioma data from The Cancer Genome Atlas show that our method supports recent biological advances and suggests new insights.

An R implemented code is freely available from https://github.com/ymatts/D3M/.

e. A likelihood-free filtering method via approximate Bayesian computation in evaluating biological simulation models

Hasegawa T[2], Niida A[2], Mori T[4], Shimamura T[3], Yamaguchi R, Miyano S, Akutsu T[4], Imoto S[2]: [4]Kyoto University

For the evaluation of the dynamic behavior of biological processes, e.g., gene regulatory sequences, we typically utilize nonlinear differential equations within a state space model in the context of genomic data assimilation. For the estimation of the parameter values for such systems, the particle filter can be a strong approach in terms of obtaining their theoretically exact posterior distributions of the parameter values. However, it has some drawbacks for dealing with biological processes in practice: (i) the number of unique particles decreases rapidly since the dimension of the parameter vector and the number of observed time points are higher than its capability, (ii) it cannot be applied when the likelihood function is analytically intractable, and (iii) the prior distributions of the parameter values are often arbitrary determined. To address these problems, we propose a novel method that utilizes the approximate Bayesian computation in filtering the data and self-organizing ensemble Kalman filter in constructing the prior distributions of the parameter values. Simulation studies show that the proposed method can overcome these problems; thus, it can estimate the posterior distributions of the parameter values with automatically setting prior distributions even for the cases that the particle filter cannot perform good results. Finally, we apply the method to real observation data in rat circadian oscillation and demonstrate the usefulness in practical situations.

f. Tuning and performance modeling of a human gene analysis program for large-scale supercomputers

Suzuki S[5], Ito S, Ikeda N[6], Balaza G[5], Miyoshi I[6], Maruyama I[5], Takizawa S, Murase Y, Ishikawa Y[5], Miyano S: [5]RIKEN AICS, [6]Fujitsu Limited

This paper presents a performance analysis and tuning of a human gene analysis program called Genomon-fusion, one of the priority applications for the post-K supercomputer. We first optimize the analysis workflow running on the K computer using a single RNA sequence sample, which then allows us to build a performance model of multi-sample analyses on parallel computing systems in a streaming way. We present the estimated analysis throughput when the whole system of the K computer is employed, and also discuss its expected performance when applied to whole genome analyses in 2020. This paper received the Best Paper Award of HPCS2016 (High-Performance Computing and Computational Science 2016).

g. Addiction to the IGF2-ID1-IGF2 circuit for maintenance of the breast cancer stem-like cells

Tominaga K[7], Shimamura T, Kimura N[7], Murayama T[7], Matsubara D[8], Kanauchi H[9], Niida A, Shimizu S[9], Nishioka K[10], Tsuji EI[10], Yano M[11], Sugano S[12], Shimono Y[13], Ishii H[14], Saya H[15], Mori M[14], Akashi K[16], Tada KI[10], Ogawa T[10], Tojo A[7], Miyano S, Gotoh N[7,17]: [7]Division of Molecular Therapy, [8]Division of Molecular Pathology, [9]Showa General Hospital, [10]Graduate School of Medicine, University of Tokyo, [11]Minami Machida Hospital, [12]Graduate School of Frontier Science, University of Tokyo, [13]Kobe University, [14]Graduate School of Medicine, Osaka University, [15]Graduate School of Medicine, Keio University, [16]Kyushu University Graduate School of Medicine, [17]Cancer Research Institute, Kanazawa University

In this research we contributed to systems biology analysis.

The transcription factor nuclear factor-κB (NF-κB) has important roles for tumorigenesis, but how it regulates cancer stem cells (CSCs) remains largely unclear. We identified insulin-like growth factor 2 (IGF2) is a key target of NF-κB activated by HER2/HER3 signaling to form tumor spheres in breast cancer cells. The IGF2 receptor, IGF1R, was expressed at high levels in CSC-enriched populations in primary breast cancer cells. Moreover, IGF2-PI3K (IGF2-phosphatidyl inositol 3 kinase) signaling induced expression of a stemness transcription factor, inhibitor of DNA-binding 1 (ID1), and IGF2 itself. ID1 knockdown greatly reduced IGF2 expression, and tumor sphere formation. Finally, treatment with anti-IGF1/2 antibodies blocked tumorigenesis derived from the IGF1Rhigh CSC-enriched population in a patient-derived xenograft model. Thus, NF-κB may trigger IGF2-ID1-IGF2-positive feedback circuits that allow cancer stem-like cells to appear.

Then, they may become addicted to the circuits. As the circuits are the Achilles' heels of CSCs, it will be critical to break them for eradication of CSCs.

2. Cancer Genomics

a. Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers

Kataoka K[18], Shiraishi Y, Takeda Y[19], Sakata S[20], Matsumoto M[20], Nagano S[21], Maeda T[21], Nagata Y[18], Kitanaka A[22], Mizuno S[23], Tanaka H, Chiba K, Ito S, Watatani Y[18], Kakiuchi N[18], Suzuki H[18], Yoshizato T[18], Yoshida K[18], Sanada M[24], Itonaga H[25], Imaizumi Y[26], Totoki Y[27], Munakata W[28], Nakamura H[27], Hama N[27], Shide K[22], Kubuki Y[22], Hidaka T[22], Kameda T[22], Masuda K[21], Minato N[29], Kashiwase K[30], Izutsu K[31], Takaori-Kondo A[32], Miyazaki Y[26], Takahashi S[23], Shibata T[27,33], Kawamoto H[21], Akatsuka Y[34,35], Shimoda K[22], Takeuchi K[20], Seya T[19], Miyano S, Ogawa S[18]: [18]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, [19]Graduate School of Medicine, Hokkaido University, [20]Cancer Institute, Japanese Foundation for Cancer Research, [21]Institute for Frontier Medical Science, Kyoto University, [22]Faculty of Medicine, University of Miyazaki, [23]Laboratory Animal Resource Center and Faculty of Medicine, University of Tsukuba, [24]Nagoya Medical Center, [25]Sasebo City General Hospital, [26]Atomic Bomb Disease Institute, Nagasaki University, [27]National Cancer Center Research Institute, [28]National Cancer Center Hospital, [29]Department of Immunology and Cell Biology, Graduate School of Medicine, Kyoto University, [30]Japanese Red Cross Kanto-Koshinetsu Block Blood Center, [31]Department of Hematology, Toranomon Hospital, [32]Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University, [33]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, [34]Department of Hematology, Fujita Health University School of Medicine, [35]Division of Immunology, Aichi Cancer Center Research Institute

Successful treatment of many patients with advanced cancer using antibodies against programmed cell death 1 (PD-1; also known as PDCD1) and its ligand (PD-L1; also known as CD274) has highlighted the critical importance of PD-1/PD-L1-mediated immune escape in cancer development. However, the genetic basis for the immune escape has not been fully elucidated, with the exception of elevated PD-L1 expression by gene amplification and utilization of an ectopic promoter by translocation, as reported in Hodgkin and other B-cell lymphomas, as well as stomach adenocarcinoma. Here we show a unique genetic mechanism of immune escape caused by structural variations (SVs) commonly disrupting the 3' region of the PD-L1 gene. Widely affecting multiple common human cancer types, including adult T-cell leukaemia/lymphoma (27%), diffuse large B-cell lymphoma (8%), and stomach adenocarcinoma (2%), these SVs invariably lead to a marked elevation of aberrant PD-L1 transcripts that are stabilized by truncation of the 3'-untranslated region (UTR). Disruption of the Pd-l1 3'-UTR in mice enables immune evasion of EG7-OVA tumour cells with elevated Pd-l1 expression in vivo, which is effectively inhibited by Pd-1/Pd-l1 blockade, supporting the role of relevant SVs in clonal selection through immune evasion. Our findings not only unmask a novel regulatory mechanism of PD-L1 expression, but also suggest that PD-L1 3'-UTR disruption could serve as a genetic marker to identify cancers that actively evade anti-tumour immunity through PD-L1 overexpression. We contributed to bioinformatics methodology development and analysis and supercomputing.

b. Dynamics of clonal evolution in myelodysplastic syndromes

Makishima H[36], Miyano S, Ogawa S[18], et al.: [36]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic

To elucidate differential roles of mutations in myelodysplastic syndromes (MDS), we investigated clonal dynamics using whole-exome and/or targeted sequencing of 699 patients, of whom 122 were analyzed longitudinally. Including the results from previous reports, we assessed a total of 2,250 patients for mutational enrichment patterns. During progression, the number of mutations, their diversity and clone sizes increased, with alterations frequently present in dominant clones with or without their sweeping previous clones. Enriched in secondary acute myeloid leukemia (sAML; in comparison to high-risk MDS), FLT3, PTPN11, WT1, IDH1, NPM1, IDH2 and NRAS mutations (type 1) tended to be newly acquired, and were associated with faster sAML progression and a shorter overall survival time. Significantly enriched in high-risk MDS (in comparison to low-risk MDS), TP53, GATA2, KRAS, RUNX1, STAG2, ASXL1, ZRSR2 and TET2 mutations (type 2) had a weaker impact on sAML progression and overall survival than type-1 mutations. The distinct roles of type-1 and type-2 mutations suggest their potential utility in disease monitoring.

c. Other Applications of Genomon for Cancer Genomics

All laboratory members and many collaborators

We have been developing an omics analysis pipeline Genomon for analyzing genome sequence data including RNA sequences. By collaborations with many cancer researchers, we contributed to data analyses using the supercomputer at Human Genome Center and K computer at AICS, RIKEN. Due to the limit of space, we list up our contributed papers: 1, 4, 8-12, 14-15, 17-18, 21, 25-26, 28, 30, 34, 36, 38, 40, 43, 47.

d. International Cancer Genome Consortium

All laboratory members and many collaborators

(ア) Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer.

Liver cancer, which is most often associated with virus infection, is prevalent worldwide, and its underlying etiology and genomic structure are heterogeneous. Here we provide a whole-genome landscape of somatic alterations in 300 liver cancers from Japanese individuals. Our comprehensive analysis identified point mutations, structural variations (STVs), and virus integrations, in noncoding and coding regions. We discovered mutational signatures related to liver carcinogenesis and recurrently mutated coding and noncoding regions, such as long intergenic noncoding RNA genes (NEAT1 and MALAT1), promoters, CTCF-binding sites, and regulatory regions. STV analysis found a significant association with replication timing and identified known (CDKN2A, CCND1, APC, and TERT) and new (ASH1L, NCOR1, and MACROD2) cancer-related genes that were recurrently affected by STVs, leading to altered expression. These results emphasize the value of whole-genome sequencing analysis in discovering cancer driver mutations and understanding comprehensive molecular profiles of liver cancer, especially with regard to STVs and noncoding mutations.

(イ) Whole genome sequencing discriminates hepatocellular carcinoma with intrahepatic metastasis from multi-centric tumors

Patients with hepatocellular carcinoma (HCC) have a high-risk of multi-centric (MC) tumor occurrence due to a strong carcinogenic background in the liver. In addition, they have a high risk of intrahepatic metastasis (IM). Liver tumors with IM or MC are profoundly different in their development and clinical outcome. However, clinically or pathologically discriminating between IM and MC can be challenging. This study investigated whether IM or MC could be diagnosed at the molecular level.

We performed whole genome and RNA sequencing analyses of 49 tumors including two extra-hepatic metastases, and one nodule-in-nodule tumor from 23 HCC patients.

Sequencing-based molecular diagnosis using somatic single nucleotide variation information showed higher sensitivity compared to previous techniques due to the inclusion of a larger number of mutation events. This proved useful in cases, which showed inconsistent clinical diagnoses. In addition, whole genome sequencing offered advantages in profiling of other genetic alterations, such as structural variations, copy number alterations, and variant allele frequencies, and helped to confirm the IM/MCdiagnosis. Divergent alterations between IM tumors with sorafenib treatment, long time-intervals, or tumor-in-tumor nodules indicated high intra-tumor heterogeneity, evolution, and clonal switching of liver cancers.

It is important to analyze the differences between IM tumors, in addition to IM/MC diagnosis, before selecting a therapeutic strategy for multiple tumors in the liver.

3. Oncoimmunology

a. Characterization of T-cell receptor repertoire in inflamed tissues of patients with crohn's disease through deep sequencing

Chapman CG[36], Yamaguchi R, Tamura K[36], Weidner J[36], Imoto S, Kwon J[36], Fang H[36], Yew PY[36], Marino SR[36], Miyano S, Nakamura Y[36], Kiyotani K[36]: [36]Department of Medicine, The University of Chicago

Intestinal tissues of patients with Crohn's disease (CD) contain expanded populations of T cells which are believed to mediate inflammation. We performed a detailed characterization of these T-cell repertoires.

We obtained biopsies from the neoterminal ileum of 12 patients undergoing evaluation for postoperative recurrent CD and 4 individuals with normal terminal ileum and no history of inflammatory bowel disease (controls). Samples of diseased terminal ileum were obtained from 5 patients undergoing surgery for stricturing or penetrating CD. Total RNA was extracted from tissues and peripheral blood mononuclear cells, and cDNAs were generated. We used next-generation sequencing to characterize T-cell receptor (TCR)-α and TCR-β cDNAs in ileal mucosal tissue and matched peripheral blood mononuclear cells of 17 patients with CD to identify oligoclonal expansions of T-cell populations associated with CD.

TCR diversity in mucosal tissue was significantly lower than that of matched peripheral blood mono-

nuclear cells, indicating expansion of certain T-cell populations in inflamed intestinal tissue. A single TCR-β clonotype, CASSWTNGEQYF (TRBV10-1-TRBJ2-7), was enriched at a frequency of 7.0% to 28.9% in the neoterminal ileum of 4 of 12 patients with recurrent CD. The abundance of this clonotype significantly correlated with the severity of disease recurrence, based on Rutgeerts score (P = 0.015).

Specific populations of T cells are expanded in the inflamed intestinal mucosa of patients with CD; their abundance correlates with severity of disease recurrence. Studies of these T cells could provide information about mechanisms of CD pathogenesis. Deep TCR sequencing is a powerful tool that rapidly provides in-depth, real-time assessment of the T-cell repertoire.

b. Characterization of the T cell repertoire by deep T cell receptor sequencing in tissues and blood from patients with advanced colorectal cancer

Tamura K[36], Hazama S[36], Yamaguchi R, Imoto S, Takenouchi H[36], Inoue Y[36], Kanekiyo S[36], Shindo Y[36], Miyano S, Nakamura Y[36], Kiyotani K[36]

The aim of the present study was to characterize infiltrated T cell clones that define the tumor immune environment and are important in the response to treatment in patients with advanced colorectal cancer (CRC). In order to explore predictive biomarkers for the efficacy of immunochemotherapies, T cell receptor (TCR) repertoire analysis was performed using blood samples and tumor tissues obtained from patients with advanced CRC that had been treated with a combination of five-cancer peptide vaccines and oxaliplatin-based chemotherapy. The TCR-α/β complementary DNAs (cDNAs), prepared from the messenger RNAs (mRNAs) obtained from 17 tumor tissues and 39 peripheral blood mononuclear cells of 9 CRC patients at various time points, were sequenced. The oligoclonal enrichment of certain TCR sequences was identified in tumor tissues and blood samples; however, only a few TCR sequences with a frequency of >0.1% were commonly detected in pre- and post-treatment tumor tissues, or in post-treatment blood and tissue samples. The average correlation coefficients of the TCR-α and TCR-β clonotype frequencies between the post-treatment tumor tissues and blood samples were 0.023 and 0.035, respectively, and were much lower compared with the correlation coefficients of the TCR-α and TCR-β clonotype frequencies between pre- and post-treatment blood samples (0.430 and 0.370, respectively), suggesting that T cell populations in tumor tissues vary from those in blood. Although the sample size was small, a tendency for the TCR diversity in tumor tissues to drastically decrease during the treatment was indicated in two patients, who exhibited a longer progression-free survival time. The results of the present study suggest that TCR diversity scores in tissues may be a useful predictive biomarker for the therapeutic effect of immunochemotherapy for patients with advanced CRC.

4. Algorithms for Bioinformatics

a. Fast classification of protein structures by an alignment-free kernel

Onodera T, Shibuya T

Alignment is the most fundamental algorithm that has been widely used in numerous research in bioinformatics, but its computation cost becomes too expensive in various modern problems because of the recent explosive data growth. Hence the development of alignment-free algorithms, i.e., alternative algorithms that avoid the computationally expensive alignment, has become one of the recent hot topics in algorithmic bioinformatics.

Analysis of protein structures is a very important problem in bioinformatics. We focus on the problem of predicting functions of proteins from their structures, as the functions of proteins are the keys of everything in the understandings of any organisms and moreover these functions are said to be determined by their structures. But the previous best-known (i.e., the most accurate) method for this problem utilizes alignment-based kernel method, which suffers from the high computation cost of alignments.

For the problem, we developed a new kernel method that does not employ alignments. Instead of alignments, we apply the two-dimensional suffix tree and the contact map graph to reduce kernel-related computation cost dramatically. Experiments show that, compared to the previous best algorithm, our new method runs about 16 times faster in training and about 37 times faster in prediction while preserving comparatively high accuracy.

Publications

1. Arashiki N, Takakuwa Y, Mohandas N, Hale J, Yoshida K, Ogura H, Utsugisawa T, Ohga S, Miyano S, Ogawa S, Kojima S, Kanno H. ATP11C is a major flippase in human erythrocytes and its defect causes congenital hemolytic anemia. *Haematologica*. 101(5): 559-565, 2016.
2. Ayada E, Hasegawa T, Niida A, Miyano S, Imoto S. Binary contingency table method for

analysing gene mutation in cancer genome. *International Journal of Bioinformatics Research and Applications*. 12(3): 211-226, 2016.

3. Chapman CG, Yamaguchi R, Tamura K, Weidner J, Imoto S, Kwon J, Fang H, Yew PY, Marino SR, Miyano S, Nakamura Y, Kiyotani K. Characterization of T-cell Receptor Repertoire in Inflamed Tissues of Patients with Crohn's Disease Through Deep Sequencing. *Inflamm Bowel Dis*. 22(6): 1275-1285, 2016.
4. Ding L, Sun QY, Tan KT, Chien W, Thippeswamy AM, Eng Juh AY, Kawamata N, Nagata Y, Xiao JF, Loh XY, Lin DC, Garg M, Lim SL, Liu LZ, Madan V, Jiang YY, Xu L, Sanada M, Fernández LT, Preethi H, Lill M, Kantarjian H, Kornblau SM, Miyano S, Ogawa S, Liang DC, Shih LY, Yang H, Koeffler HP. Mutational landscape of pediatric acute lymphoblastic leukemia. *Cancer Res*. 77(2): 390-400, 2017.
5. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, Tanaka H, Taniguchi H, Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Wardell CP, Hayami S, Nakamura T, Aikata H, Arihiro K, Boroevich KA, Abe T, Nakano K, Maejima K, Sasaki-Oku A, Ohsawa A, Shibuya T, Nakamura H, Hama N, Hosoda F, Arai Y, Ohashi S, Urushidate T, Nagae G, Yamamoto S, Ueda H, Tatsuno K, Ojima H, Hiraoka N, Okusaka T, Kubo M, Marubashi S, Yamada T, Hirano S, Yamamoto M, Ohdan H, Shimada K, Ishikawa O, Yamaue H, Chayama K, Miyano S, Aburatani H, Shibata T, Nakagawa H. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet*. 48(5):500-509, 2016. *Erratum*: Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet*. 48(6): 700, 2016.
6. Furuta M, Ueno M, Fujimoto A, Hayami S, Yasukawa S, Kojima F, Arihiro K, Kawakami Y, Wardell CP, Shiraishi Y, Tanaka H, Nakano K, Maejima K, Sasaki-Oku A, Tokunaga N, Boroevich KA, Abe T, Aikata H, Ohdan H, Gotoh K, Kubo M, Tsunoda T, Miyano S, Chayama K, Yamaue H, Nakagawa H. Whole genome sequencing discriminates hepatocellular carcinoma with intrahepatic metastasis from multi-centric tumors. *J Hepatol*. 66(2): 363-373, 2017.
7. Hasegawa T, Niida A, Mori T, Shimamura T, Yamaguchi R, Miyano S, Akutsu T, Imoto S. A likelihood-free filtering method via approximate Bayesian computation in evaluating biological simulation models. *Computational Statistics & Data Analysis*. 94: 63-74, 2016.
8. Hoshino A, Okada S, Yoshida K, Nishida N, Okuno Y, Ueno H, Yamashita M, Okano T, Tsumura M, Nishimura S, Sakata S, Kobayashi M, Nakamura H, Kamizono J, Mitsui-Sekinaka K, Ichimura T, Ohga S, Nakazawa Y, Takagi M, Imai K, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Kojima S, Nonoyama S, Morio T, Kanegane H. Abnormal hematopoiesis and autoimmunity in human subjects with germline IKZF1 mutations. *J Allergy Clin Immunol*. 2016 Dec 1. pii: S0091-6749(16)31273-8. doi: 10.1016/j.jaci.2016.09.029. [Epub ahead of print] PMID: 27939403
9. Hosono N, Makishima H, Mahfouz R, Przychodzen B, Yoshida K, Jerez A, LaFramboise T, Polprasert C, Clemente MJ, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Sanada M, Cui E, Verma AK, McDevitt MA, List AF, Sauntharajah Y, Sekeres MA, Boultwood J, Ogawa S, Maciejewski JP. Recurrent genetic defects on chromosome 5q in myeloid neoplasms. *Oncotarget*. 2016 Dec 23. doi: 10.18632/oncotarget.14130. [Epub ahead of print] PMID: 28031539
10. Ichimura T, Yoshida K, Okuno Y, Yujiri T, Nagai K, Nishi M, Shiraishi Y, Ueno H, Toki T, Chiba K, Tanaka H, Muramatsu H, Hara T, Kanno H, Kojima S, Miyano S, Ito E, Ogawa S, Ohga S. Diagnostic challenge of Diamond-Blackfan anemia in mothers and children by whole-exome sequencing. *Int J Hematol*. 2016 Nov 23. [Epub ahead of print] PubMed PMID: 27882484.
11. Ikeda F, Yoshida K, Toki T, Uechi T, Ishida S, Nakajima Y, Sasahara Y, Okuno Y, Kanezaki R, Terui K, Kamio T, Kobayashi A, Fujita T, Sato-Otsubo A, Shiraishi Y, Tanaka H, Chiba K, Muramatsu H, Kanno H, Ohga S, Ohara A, Kojima S, Kenmochi N, Miyano S, Ogawa S, Ito E. Exome sequencing identified RPS15A as a novel causative gene for Diamond-Blackfan anemia. *Haematologica*. 2016 Dec 1. pii: haematol.2016.153932. [Epub ahead of print] PubMed PMID: 27909223.
12. Imashuku S, Muramatsu H, Sugihara T, Okuno Y, Wang X, Yoshida K, Kato A, Kato K, Tatsumi Y, Hattori A, Kita S, Oe K, Sueyoshi A, Usui T, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Kojima S, Kanno H. PIEZO1 gene mutation in a Japanese family with hereditary high phosphatidylcholine hemolytic anemia and hemochromatosis-induced diabetes mellitus. *Int J Hematol*. 104(1): 125-129, 2016.
13. Kataoka K, Shiraishi Y, Takeda Y, Sakata S, Matsumoto M, Nagano S, Maeda T, Nagata Y, Kitanaka A, Mizuno S, Tanaka H, Chiba K, Ito S, Watatani Y, Kakiuchi N, Suzuki H, Yoshizato T, Yoshida K, Sanada M, Itonaga H, Imaizumi Y, Totoki Y, Munakata W, Nakamura H, Hama N, Shide K, Kubuki Y, Hidaka T, Kameda T, Masuda K, Minato N, Kashiwase K, Izutsu K, Takaori-Kondo A, Miyazaki Y, Takahashi S, Shibata T, Kawamoto H, Akatsuka Y, Shimoda

K, Takeuchi K, Seya T, Miyano S, Ogawa S. Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers. *Nature*. 534(7607): 402-406, 2016.

14. Kato M, Ishimaru S, Seki M, Yoshida K, Shiraishi Y, Chiba K, Kakiuchi N, Sato Y, Ueno H, Tanaka H, Inukai T, Tomizawa D, Hasegawa D, Osumi T, Arakawa Y, Aoki T, Okuya M, Kaizu K, Kato K, Taneyama Y, Goto H, Taki T, Takagi M, Sanada M, Koh K, Takita J, Miyano S, Ogawa S, Ohara A, Tsuchida M, Manabe A. Long-term outcome of 6-month maintenance chemotherapy for acute lymphoblastic leukemia in children. *Leukemia*. 2016 Oct 25. doi: 10.1038/leu.2016.274. [Epub ahead of print] PubMed PMID: 27698447.
15. Kato M, Seki M, Yoshida K, Sato Y, Oyama R, Arakawa Y, Kishimoto H, Taki T, Akiyama M, Shiraishi Y, Chiba K, Tanaka H, Mitsuiki N, Kajiwara M, Mizutani S, Sanada M, Miyano S, Ogawa S, Koh K, Takita J. Genomic analysis of clonal origin of Langerhans cell histiocytosis following acute lymphoblastic leukaemia. *Br J Haematol*. 175(1): 169-172, 2016.
16. Kayano M, Matsui H, Yamaguchi R, Imoto S, Miyano S. Gene set differential analysis of time course expression profiles via sparse estimation in functional logistic model with application to time-dependent biomarker detection. *Biostatistics*. 17(2): 235-248, 2016.
17. Ki Kim S, Ueda Y, Hatano E, Kakiuchi N, Takeda H, Goto T, Shimizu T, Yoshida K, Ikura Y, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Uemoto S, Chiba T, Ogawa S, Marusawa H. TERT promoter mutations and chromosome 8p loss are characteristic of nonalcoholic fatty liver disease-related hepatocellular carcinoma. *Int J Cancer*. 139(11): 2512-2518, 2016.
18. Kitamura K, Okuno Y, Yoshida K, Sanada M, Shiraishi Y, Muramatsu H, Kobayashi R, Furukawa K, Miyano S, Kojima S, Ogawa S, Kunishima S. Functional characterization of a novel GFI1B mutation causing congenital macrothrombocytopenia. *J Thromb Haemost*. 14(7): 1462-1469, 2016.
19. Koso H, Yi H, Sheridan P, Miyano S, Ino Y, Todo T, Watanabe S. Identification of RNA-Binding Protein LARP4B as a Tumor Suppressor in Glioma. *Cancer Res*. 76(8): 2254-2264, 2016.
20. Kurashige J, Hasegawa T, Niida A, Sugimachi K, Deng N, Mima K, Uchi R, Sawada G, Takahashi Y, Eguchi H, Inomata M, Kitano S, Fukagawa T, Sasako M, Sasaki H, Sasaki S, Mori M, Yanagihara K, Baba H, Miyano S, Tan P, Mimori K. Integrated Molecular Profiling of Human Gastric Cancer Identifies DDR2 as a Potential Regulator of Peritoneal Dissemination. *Sci Rep*. 6: 22371, 2016.
21. Madan V, Shyamsunder P, Han L, Mayakonda A, Nagata Y, Sundaresan J, Kanojia D, Yoshida K, Ganesan S, Hattori N, Fulton N, Tan KT, Alpermann T, Kuo MC, Rostami S, Matthews J, Sanada M, Liu LZ, Shiraishi Y, Miyano S, Chendamarai E, Hou HA, Malnassy G, Ma T, Garg M, Ding LW, Sun QY, Chien W, Ikezoe T, Lill M, Biondi A, Larson RA, Powell BL, Lübbert M, Chng WJ, Tien HF, Heuser M, Ganser A, Koren-Michowitz M, Kornblau SM, Kantarjian HM, Nowak D, Hofmann WK, Yang H, Stock W, Ghavamzadeh A, Alimoghaddam K, Haferlach T, Ogawa S, Shih LY, Mathews V, Koeffler HP. Comprehensive mutational analysis of primary and relapse acute promyelocytic leukemia. *Leukemia*. 30(8): 1672-1681, 2016.
22. Mao Y, Tamura T, Yuki Y, Abe D, Tamada Y, Imoto S, Tanaka H, Homma H, Tagawa K, Miyano S, Okazawa H. The hnRNP-Htt axis regulates necrotic cell death induced by transcriptional repression through impaired RNA splicing. *Cell Death Dis*. 7: e2207, 2016.
23. Makishima H, Yoshizato T, Yoshida K, Sekeres MA, Radivoyevitch T, Suzuki H, Przychodzen B, Nagata Y, Meggendorfer M, Sanada M, Okuno Y, Hirsch C, Kuzmanovic T, Sato Y, Sato-Otsubo A, LaFramboise T, Hosono N, Shiraishi Y, Chiba K, Haferlach C, Kern W, Tanaka H, Shiozawa Y, Gómez-Seguí I, Husseinzadeh HD, Thota S, Guinta KM, Dienes B, Nakamaki T, Miyawaki S, Sauntharajah Y, Chiba S, Miyano S, Shih LY, Haferlach T, Ogawa S, Maciejewski JP. Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet*. 2016 Dec 19. doi: 10.1038/ng.3742. [Epub ahead of print] PMID: 27992414
24. Matsui Y, Mizuta M, Ito S, Miyano S, Shimamura T. D^3M: detection of differential distributions of methylation levels. *Bioinformatics*. 32 (15): 2248-2255, 2016.
25. Mitani K, Nagata Y, Sasaki K, Yoshida K, Chiba K, Tanaka H, Shiraishi Y, Miyano S, Makishima H, Nakamura Y, Nakamura Y, Ichikawa M, Ogawa S. Somatic mosaicism in chronic myeloid leukemia in remission. *Blood*. 128(24): 2863-2866, 2016.
26. Mori T, Nagata Y, Makishima H, Sanada M, Shiozawa Y, Kon A, Yoshizato T, Sato-Otsubo A, Kataoka K, Shiraishi Y, Chiba K, Tanaka H, Ishiyama K, Miyawaki S, Mori H, Nakamaki T, Kihara R, Kiyoi H, Koeffler HP, Shih LY, Miyano S, Naoe T, Haferlach C, Kern W, Haferlach T, Ogawa S, Yoshida K. Somatic PHF6 mutations in 1760 cases with various myeloid neoplasms. *Leukemia*. 30(11): 2270-2273, 2016.
27. Moriyama T, Shiraishi Y, Chiba K, Yamaguchi R, Imoto S, Miyano S. OVarCall: Bayesian Mutation Calling Method Utilizing Overlapping Paired-End Reads. *Lecture Notes in Computer Science*. 9683: 40-51, 2016.

28. Muramatsu H, Okuno Y, Yoshida K, Shiraishi Y, Doisaki S, Narita A, Sakaguchi H, Kawashima N, Wang X, Xu Y, Chiba K, Tanaka H, Hama A, Sanada M, Takahashi Y, Kanno H, Yamaguchi H, Ohga S, Manabe A, Harigae H, Kunishima S, Ishii E, Kobayashi M, Koike K, Watanabe K, Ito E, Takata M, Yabe M, Ogawa S, Miyano S, Kojima S. Clinical utility of next-generation sequencing for inherited bone marrow failure syndromes. *Genet Med*. 2017 Jan 19. doi: 10.1038/gim.2016.197. [Epub ahead of print] PMID: 28102861
29. Muramatsu T, Kozaki KI, Imoto S, Yamaguchi R, Tsuda H, Kawano T, Fujiwara N, Morishita M, Miyano S, Inazawa J. The hypusine cascade promotes cancer progression and metastasis through the regulation of RhoA in squamous cell carcinoma. *Oncogene*. 35(40): 5304-5316, 2016.
30. Nagata Y, Kontani K, Enami T, Kataoka K, Ishii R, Totoki Y, Kataoka TR, Hirata M, Aoki K, Nakano K, Kitanaka A, Sakata-Yanagimoto M, Egami S, Shiraishi Y, Chiba K, Tanaka H, Shiozawa Y, Yoshizato T, Suzuki H, Kon A, Yoshida K, Sato Y, Sato-Otsubo A, Sanada M, Munakata W, Nakamura H, Hama N, Miyano S, Nureki O, Shibata T, Haga H, Shimoda K, Katada T, Chiba S, Watanabe T, Ogawa S. Variegated RHOA mutations in adult T-cell leukemia/lymphoma. *Blood*. 127(5): 596-604, 2016.
31. Onodera T, Shibuya T. Fast classification of protein structures by an alignment-free kernel. *Lecture Notes in Computer Science*. 9954: 68-79, 2016.
32. Park H, Niida A, Imoto S, Miyano S. Interaction-Based Feature Selection for uncovering cancer driver genes through copy number-driven expression level. *J Comput Biol*. 2016 Oct 19. [Epub ahead of print] PubMed PMID: 27759426.
33. Park H, Shiraishi Y, Imoto S, Miyano S. A novel adaptive penalized logistic regression for uncovering biomarker associated with anti-cancer drug sensitivity. *IEEE/ACM Trans Comput Biol Bioinform*. 2016 May 3. [Epub ahead of print] PubMed PMID: 27164605.
34. Sakurai M, Kasahara H, Yoshida K, Yoshimi A, Kunimoto H, Watanabe N, Shiraishi Y, Chiba K, Tanaka H, Harada Y, Harada H, Kawakita T, Kurokawa M, Miyano S, Takahashi S, Ogawa S, Okamoto S, Nakajima H. Genetic basis of myeloid transformation in familial platelet disorder/acute myeloid leukemia patients with haploinsufficient RUNX1 allele. *Blood Cancer J*. 6: e392, 2016.
35. Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, Shiraishi Y, Chiba K, Imoto S, Takahashi Y, Iwaya T, Sudo T, Hayashi T, Takai H, Kawasaki Y, Matsukawa T, Eguchi H, Sugimachi K, Tanaka F, Suzuki H, Yamamoto K, Ishii H, Shimizu M, Yamazaki H, Yamazaki M, Tachimori Y, Kajiyama Y, Natsugoe S, Fujita H, Mafune K, Tanaka Y, Kelsell DP, Scott CA, Tsuji S, Yachida S, Shibata T, Sugano S, Doki Y, Akiyama T, Aburatani H, Ogawa S, Miyano S, Mori M, Mimori K. Genomic landscape of esophageal squamous cell carcinoma in a Japanese population. *Gastroenterology*. 150(5): 1171-1182, 2016.
36. Shiba N, Yoshida K, Shiraishi Y, Okuno Y, Yamato G, Hara Y, Nagata Y, Chiba K, Tanaka H, Terui K, Kato M, Park MJ, Ohki K, Shimada A, Takita J, Tomizawa D, Kudo K, Arakawa H, Adachi S, Taga T, Tawa A, Ito E, Horibe K, Sanada M, Miyano S, Ogawa S, Hayashi Y. Whole-exome sequencing reveals the spectrum of gene mutations and the clonal evolution patterns in paediatric acute myeloid leukaemia. *Br J Haematol*. 175(3): 476-489, 2016.
37. Sugimachi K, Yamaguchi R, Eguchi H, Ueda M, Niida A, Sakimura S, Hirata H, Uchi R, Shinden Y, Iguchi T, Morita K, Yamamoto K, Miyano S, Mori M, Maehara Y, Mimori K. 8q24 Polymorphisms and Diabetes Mellitus Regulate Apolipoprotein A-IV in Colorectal Carcinogenesis. *Ann Surg Oncol*. 23(Suppl 4): 546-551, 2016.
38. Sun QY, Ding LW, Tan KT, Chien W, Mayakonda A, Lin DC, Loh XY, Xiao JF, Meggendorfer M, Alpermann T, Garg M, Lim SL, Madan V, Hattori N, Nagata Y, Miyano S, Yeoh AE, Hou HA, Jiang YY, Takao S, Liu LZ, Tan SZ, Lill M, Hayashi M, Kinoshita A, Kantarjian HM, Kornblau SM, Ogawa S, Haferlach T, Yang H, Koeffler HP. Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). Leukemia. 31(1): 1-10, 2017.
39. Suzuki S, Ito S, Ikeda N, Balaza G, Miyoshi I, Maruyama I, Takizawa S, Murase Y, Ishikawa Y, Miyano S. Tuning and performance modeling of a human gene analysis program for large-scale supercomputers. High-Performance Computing and Computational Science 2016 (HPCS2016), 15-26, 2016. (Best Paper Award)
40. Takagi M, Ogata S, Ueno H, Yoshida K, Yeh T, Hoshino A, Piao J, Yamashita M, Nanya M, Okano T, Kajiwara M, Kanegane H, Muramatsu H, Okuno Y, Shiraishi Y, Chiba K, Tanaka H, Bando Y, Kato M, Hayashi Y, Miyano S, Imai K, Ogawa S, Kojima S, Morio T. Haploinsufficiency of TNFAIP3 (A20) by germline mutation is involved in autoimmune lymphoproliferative syndrome. *J Allergy Clin Immunol*. 2016 Nov 11. pii: S0091-6749(16)31286-6. doi: 10.1016/j.jaci.2016.09.038. [Epub ahead of print] PubMed PMID: 27845235.
41. Tamura K, Hazama S, Yamaguchi R, Imoto S, Takenouchi H, Inoue Y, Kanekiyo S, Shindo Y, Miyano S, Nakamura Y, Kiyotani K. Characterization of the T cell repertoire by deep T cell re-

ceptor sequencing in tissues and blood from patients with advanced colorectal cancer. *Oncol Lett*. 11(6): 3643-3649, 2016.

42. Tominaga K, Shimamura T, Kimura N, Murayama T, Matsubara D, Kanauchi H, Niida A, Shimizu S, Nishioka K, Tsuji EI, Yano M, Sugano S, Shimono Y, Ishii H, Saya H, Mori M, Akashi K, Tada KI, Ogawa T, Tojo A, Miyano S, Gotoh N. Addiction to the IGF2-ID1-IGF2 circuit for maintenance of the breast cancer stem-like cells. *Oncogene*. 2016 Aug 22. doi: 10.1038/onc.2016.293. [Epub ahead of print] PubMed PMID: 27546618.
43. Tsujita Y, Mitsui-Sekinaka K, Imai K, Yeh TW, Mitsuiki N, Asano T, Ohnishi H, Kato Z, Sekinaka Y, Zaha K, Kato T, Okano T, Takashima T, Kobayashi K, Kimura M, Kunitsu T, Maruo Y, Kanegane H, Takagi M, Yoshida K, Okuno Y, Muramatsu H, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Kojima S, Ogawa S, Ohara O, Okada S, Kobayashi M, Morio T, Nonoyama S. Phosphatase and tensin homolog (PTEN) mutation can cause activated phosphatidylinositol 3-kinase δ syndrome-like immunodeficiency. *J Allergy Clin Immunol*. 138(6): 1672-1680, 2016.
44. Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, Sawada G, Iwaya T, Kurashige J, Shinden Y, Iguchi T, Eguchi H, Chiba K, Shiraishi Y, Nagae G, Yoshida K, Nagata Y, Haeno H, Yamamoto H, Ishii H, Doki Y, Iinuma H, Sasaki S, Nagayama S, Yamada K, Yachida S, Kato M, Shibata T, Oki E, Saeki H, Shirabe K, Oda Y, Maehara Y, Komune S, Mori M, Suzuki Y, Yamamoto K, Aburatani H, Ogawa S, Miyano S, Mimori K. Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution. *PLoS Genet*. 12(2): e1005778, 2016.
45. Ueda M, Iguchi T, Masuda T, Nakahara Y, Hirata H, Uchi R, Niida A, Momose K, Sakimura S, Chiba K, Eguchi H, Ito S, Sugimachi K, Yamasaki M, Suzuki Y, Miyano S, Doki Y, Mori M, Mimori K. Somatic mutations in plasma cell-free DNA are diagnostic markers for esophageal squamous cell carcinoma recurrence. *Oncotarget*. 7(38): 62280-62291, 2016.
46. Yamaguchi K, Nagayama S, Shimizu E, Komura M, Yamaguchi R, Shibuya T, Arai M, Hatakeyama S, Ikenoue T, Ueno M, Miyano S, Imoto S, Furukawa Y. Reduced expression of APC-1B but not APC-1A by the deletion of promoter 1B is responsible for familial adenomatous polyposis. *Sci Rep*. 6: 26011, 2016.
47. Yamato G, Shiba N, Yoshida K, Shiraishi Y, Hara Y, Ohki K, Okubo J, Okuno H, Chiba K, Tanaka H, Kinoshita A, Moritake H, Kiyokawa N, Tomizawa D, Park MJ, Sotomatsu M, Taga T, Adachi S, Tawa A, Horibe K, Arakawa H, Miyano S, Ogawa S, Hayashi Y. ASXL2 mutations are frequently found in pediatric AML patients with t(8;21)/RUNX1-RUNX1T1 and associated with a better prognosis. *Genes Chromosomes Cancer*. 2017 Jan 7. doi: 10.1002/gcc.22443. [Epub ahead of print] PMID: 28063196
48. Yoshino T, Katayama K, Horiba Y, Munakata K, Yamaguchi R, Imoto S, Miyano S, Mima H, Watanabe K. Predicting Japanese Kampo formulas by analyzing database of medical records: a preliminary observational study. *BMC Med Inform Decis Mak*. 16: 118, 2016.
49. Yoshino T, Katayama K, Horiba Y, Munakata K, Yamaguchi R, Imoto S, Miyano S, Mima H, Watanabe K, Mimura M. The Difference between the Two Representative Kampo Formulas for Treating Dysmenorrhea: An Observational Study. *Evid Based Complement Alternat Med*. 2016:3159617, 2016.

Human Genome Center

Laboratory of Molecular Medicine
ゲノム医科学分野

Professor	Tatsuhiro Shibata, M.D., Ph.D.	教 授	柴 田 龍 弘
Assistant Professor	Satoshi Yamasaki, Ph.D.	助 教	山 﨑 智

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. Mutational signatures associated with tobacco smoking in human cancer.

Ludmil B. Alexandrov[1,2,3], Young Seok Ju[4], Kerstin Haase[5], Peter Van Loo[5,6], Iñigo Martincorena[7], Serena Nik-Zainal[7,8], Yasushi Totoki[9], Akihiro Fujimoto[10,11], Hidewaki Nakagawa[10], Tatsuhiro Shibata[9,12], Peter J. Campbell[7,13], Paolo Vineis[14,15], David H. Phillips[16], Michael R. Stratton[7]: [1]Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, USA, [2]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, USA, [3]University of New Mexico Comprehensive Cancer Center, Albuquerque, USA, [4]Graduate School of Medical Science and Engineering (GSMSE), Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea, [5]The Francis Crick Institute, Lincoln's Inn Fields Laboratory, London, UK, [6]Human Genome Laboratory, Department of Human Genetics, University of Leuven, Belgium, [7]Cancer Genome Project, Wellcome Trust Sanger Institute, Cambridgeshire, UK, [8]Department of Medical Genetics, Addenbrooke's Hospital National Health Service (NHS) Trust, Cambridge, UK, [9]Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan, [10]Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan, [11]Department of Drug Discovery Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan, [12]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [13]Department of Haematology, University of Cambridge, Cambridge, UK, [14]Human Genetics Foundation (HuGeF), Torino, Italy, [15]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK, [16]King's College London, MRC-PHE Centre for Environment and Health, Analytical and Environmental Sciences Division, London, UK.

Tobacco smoking increases the risk of at least 17 classes of cancer. We analyzed somatic mutations and DNA methylation in more than 5,000 cancers of types for which tobacco smoking confers an elevated risk. Smoking is associated with increased mutation burdens of multiple distinct mutational signatures, which contribute to different extents in different cancers. One of these signatures, mainly found in cancers derived from tissues directly exposed to tobacco smoke, is attributable to misreplication of DNA damage caused by tobacco carcinogens. Others likely reflect indirect activation of

DNA editing by APOBEC cytidinedeaminases and of an endogenous clock-like mutational process. The results are consistent with the proposition that smoking increases cancer risk by increasing the somatic mutation load, although direct evidence for this mechanism is lacking in some smoking-related cancer types.

2. Whole genome mutational landscape and characterization of non-coding and structural mutations in liver cancer.

Akihiro Fujimoto[1], Mayuko Furuta[1], Yasushi Totoki[2], Tatsuhiko Tsunoda[3], Mamoru Kato[4], Yuichi Shiraishi[5], Hiroko Tanaka[5], Hiroaki Taniguchi[1], Yoshiiku Kawakami[5], Masaki Ueno[6], Kunihito Gotoh[7], Shun-ichi Ariizumi[8], Christopher P Wardell[1], Shinya Hayami[6], Nakamura Toru[9], Hiroshi Aikata[5], Koji Arihiro[10], Keith A Boroevich[3], Tetsuo Abe[3], Kaoru Nakano[1], Kazuhiro Maejima[1], Aya Sasaki-Oku[1], Tetsuo Shibuya[11], Hiromi Nakamura[2], Natsuko Hama[2], Fumie Hosoda[2], Yasuhito Arai[2], Tomoko Urushidate[12], Shoko Ohashi[2], Genta Nagae[13], Shogo Yamamoto[13], Hiroki Ueda[13], Kenji Tatsuno[13], Hidenori Ojima[14], Nobuyuki Hiraoka[15], Takuji Okusaka[16], Shigeru Marubashi[7], Akimasa Yamada[7], Michiaki Kubo[17], Satoshi Hirano[9], Masakazu Yamamoto[8], Hideki Ohdan[18], Kazuaki Shimada[15], Osamu Ishikawa[7], Hiroki Yamaue[6], Kazuki Chayama[5,19], Satoru Miyano[5,11], Hiroyuki Aburatani[13], Tatsuhiro Shibata[2,12], Hidewaki Nakagawa[1]: [1]Laboratory for Genome Sequencing Analysis, [3]Laboratory for Medical Science Mathematics, [17]Laboratory for Genotyping Development, [19]Laboratory for Digestive Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [2]Division of Cancer Genomics, [4]Department of Bioinformatics, National Cancer Center Research Institute, Tokyo, Japan, [5]Laboratory of DNA Information Analysis, [11]Laboratory of Sequence Analysis, [12]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [5]Department of Medicine & Molecular Science, [10]Department of Anatomical Pathology, [18]Department of Gastroenterological Surgery, Hiroshima University School of Medicine, Hiroshima, Japan, [6]Second Department of Surgery, Wakayama Medical University, Wakayama, Japan, [7]Department of Surgery, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan, [8]Department of Gastroenterological Surgery, Tokyo Women's Medical University, Tokyo, Japan, [9]Department of Gastroenterological Surgery II, Hokkaido University Graduate School of Medicine, Sapporo, Japan, [13]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan, [14]Department of Pathology, Keio University School of Medicine, Tokyo, Japan, [15]Division of Pathology and Clinical Laboratories, [16]Hepatobiliary and Pancreatic Surgery Division, [17]Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital, Tokyo, Japan.

Liver cancer is prevalent worldwide, mainly associated with virus infection, and its underlying etiology and genomic structure are heterogeneous. We have unrevealed a whole-genome landscape of somatic alterations in 300 Japanese liver cancers. Our comprehensive analysis elucidated point mutations in non-coding regions, structural variants (SVs), and virus integrations, in addition to coding mutations, and demonstrated new mutational signatures related with liver carcinogenesis, novel recurrently mutated coding and non-coding regions such as lincRNA (*NEAT1*/*MALAT1*), promoters, CTCF binding sites, and regulatory regions. SV analysis found a significant association with replication timing, and revealed that known cancer genes (*CDKN2A*, *CCND1*, *APC*, and *TERT*) and new cancer genes (*ASH1L*, *NCOR1*, *MACROD2*) were recurrently affected by SVs, leading to altered expression of nearby genes. These results emphasize the value of whole genome sequencing analysis to discover cancer driver mutations and to understand comprehensive molecular profiles of liver cancer, especially for SVs and non-coding mutations.

3. Genomic sequencing identifies ELF3 as a driver of ampullary carcinoma.

Shinichi Yachida[1], Laura D. Wood[2], Masami Suzuki[1], Erina Takai[1], Yasushi Totoki[1], Mamoru Kato[1], Claudio Luchini[2], Yasuhito Arai[1], Hiromi Nakamura[1], Natsuko Hama[1], Asmaa Elzawahry[1], Fumie Hosoda[1], Tomoki Shirota[1], Nobuhiko Morimoto[3], Kunio Hori[3], Jun Funazaki[3], Hikaru Tanaka[4], Chigusa Morizane[5], Takuji Okusaka[5], Satoshi Nara[6], Kazuaki Shimada[6], Nobuyoshi Hiraoka[7], Hirokazu Taniguchi[7], Ryota Higuchi[8], Minoru Oshima[9], Keiichi Okano[9], Seiko Hirono[10], Masamichi Mizuma[11], Koji Arihiro[12], Masakazu Yamamoto[8], Michiaki Unno[11], Hiroki Yamaue[10], Matthew J. Weiss[13], Christopher L. Wolfgang[13], Toru Furukawa[14], Hitoshi Nakagama[15], Bert Vogelstein[2,16], Tohru Kiyono[17], Ralph H. Hruban[2,18], Tatsuhiro Shibata[1,19]: [1]Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan, [2]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, USA, [3]Division of Medical Elemental Technology Development, Department of Advanced Analysis Technology, R&D Group, Olympus Corporation, Tokyo, Japan, [4]Division of Virology, National Cancer Center Research Institute, Tokyo, Japan, [5]Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital, Tokyo, Japan,

[6]Hepatobiliary and Pancreatic Surgery Division, National Cancer Center Hospital, Tokyo, Japan, [7]Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan, [8]Department of Surgery, Institute of Gastroenterology, Tokyo Women's Medical University, Tokyo, Japan, [9]Department of Gastroenterological Surgery, Kagawa University, Kagawa, Japan, [10]Second Department of Surgery, Wakayama Medical University, Wakayama, Japan, [11]Department of Surgery, Tohoku University Graduate School of Medicine, Sendai Miyagi, Japan, [12]Department of Anatomical Pathology, Hiroshima University Hospital, Hiroshima, Japan, [13]Department of Surgery, The Johns Hopkins University School of Medicine, Baltimore, USA, [14]Institute for Integrated Medical Science, Tokyo Women's Medical University, Tokyo, Japan, [15]National Cancer Center Research Institute, Tokyo, Japan, [16]The Ludwig Center and The Howard Hughes Medical Institute at Johns Hopkins Kimmel Cancer Center, Baltimore, USA, [17]Division of Carcinogenesis and Cancer Prevention, National Cancer Center Research Institute, Tokyo, Japan, [18]Department of Oncology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, USA, [19]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

Ampullary carcinomas are highly malignant neoplasms that can have either intestinal or pancreatobiliary differentiation. To characterize somatic alterations in ampullary carcinomas, we performed whole-exome sequencing and DNA copy number analysis on 60 ampullary carcinomas resected from clinically well-characterized Japanese and American patients. We next selected 92 genes and performed targeted-sequencing to validate significantly mutated genes in additional 112 cancers. The prevalence of driver gene mutations in carcinomas with the intestinal phenotype is different from those with the pancreatobiliary phenotype. We identified previously known driver genes (TP53, KRAS, APC and others) and a novel significantly mutated driver gene (ELF3). Functional studies demonstrated that ELF3 silencing in normal human epithelial cells enhances their motility and invasion.

Publications

1. Alexandrov, LB, Ju YS, Haase K, Loo PV, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto A, Nakagawa H, Shibata T, Campbell PJ, Vineis P, Phillips DH, Stratton MR. Mutational signatures associated with tobacco smoking in human cancer. Science, 2016, 354: 618-622.
2. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, Tanaka H, Taniguchi H, Kawakami Y, Ueno M, Gotoh K, Ariizumi SI, Wardell CP, Hayami S, Nakamura T, Aikata H, Arihiro K, Boroevich KA, Abe T, Nakano K, Maejima K, Sasaki-Oku A, Ohsawa A, Shibuya T, Nakamura H, Hama N, Hosoda F, Arai Y, Ohashi S, Urushidate T, Nagae G, Yamamoto S, Ueda H, Tatsuno K, Ojima H, Hiraoka N, Okusaka T, Kubo M, Marubashi S, Yamada T, Hirano S, Yamamoto M, Ohdan H, Shimada K, Ishikawa O, Yamaue H, Chayama K, Miyano S, Aburatani H, Shibata T, Nakagawa H. Whole genome mutational landscape and characterization of non-coding and structural mutations in liver cancer. Nat Genet, 2016, 48: 500-9.
3. Yachida S, Wood LD, Suzuki M, Takai E, Totoki Y, Kato M, Luchini C, Arai Y, Nakamura H, Hama N, Elzawahry A, Hosoda F, Shirota T, Morimoto N, Hori K, Funazaki J, Tanaka H, Morizane C, Okusaka T, Nara S, Shimada K, Hiraoka N, Taniguchi H, Higuchi R, Oshima M, Okano K, Hirono S, Mizuma M, Arihiro K, Yamamoto M, Unno M, Yamaue H, Weiss MJ, Wolfgang CL, Furukawa T, Nakagama H, Vogelstein B, Kiyono T, Hruban RH, Shibata T. Genomic sequencing identifies ELF3 as a driver of ampullary carcinoma. Cancer Cell, 2016, 29: 229-40.

Human Genome Center

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

Professor	Satoru Miyano, Ph.D.	教　授	宮　野　　悟(兼)
Professor	Koichi Matsuda, M.D., Ph.D.	連携教授	松　田　浩　一(新領域創成科学研究科)
Assistant Professor	Chizu Tanikawa, Ph.D.	助　教	谷　川　千　津
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.

1. Genome-wide association study

Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors.

High rates of APOBEC-signature mutations are found in many tumors, but factors affecting this mutation pattern are not well understood. Here we explored the contribution of two common germline variants in the APOBEC3 region. SNP rs1014971 was associated with bladder cancer risk, increased APOBEC3B expression, and enrichment with APOBEC-signature mutations in bladder tumors. In contrast, a 30-kb deletion that eliminates APOBEC3B and creates an APOBEC3A-APOBEC3B chimera was not important in bladder cancer, whereas it was associated with breast cancer risk and enrichment with APOBEC-signature mutations in breast tumors. In vitro, APOBEC3B expression was predominantly induced by treatment with a DNA-damaging drug in bladder cancer cell lines, and APOBEC3A expression was induced as part of the antiviral interferon-stimulated response in breast cancer cell lines. These findings suggest a tissue-specific role of environmental oncogenic triggers, particularly in individuals with germline APOBEC3 risk variants.

Identification of Susceptibility Loci and Genes for Colorectal Cancer Risk.

BACKGROUND & AIMS: Known genetic factors explain only a small fraction of genetic variation in colorectal cancer (CRC). We conducted a genome-wide association study to identify risk loci for CRC.

METHODS: This discovery stage included 8027 cases and 22,577 controls of East-Asian ancestry. Promising variants were evaluated in studies including as many as 11,044 cases and 12,047 controls. Tumor-adjacent normal tissues from 188 patients were analyzed to evaluate correlations of risk variants with expression levels of nearby genes. Potential functionality of risk variants were evaluated using public genomic and epigenomic databases.

RESULTS: We identified 4 loci associated with CRC risk; P values for the most significant variant in each locus ranged from 3.92×10^{-8} to 1.24×10^{-12}:

6p21.1 (rs4711689), 8q23.3 (rs2450115, rs6469656), 10q24.3 (rs4919687), and 12p13.3 (rs11064437). We also identified 2 risk variants at loci previously associated with CRC: 10q25.2 (rs10506868) and 20q13.3 (rs6061231). These risk variants, conferring an approximate 10%-18% increase in risk per allele, are located either inside or near protein-coding genes that include transcription factor EB (lysosome biogenesis and autophagy), eukaryotic translation initiation factor 3, subunit H (initiation of translation), cytochrome P450, family 17, subfamily A, polypeptide 1 (steroidogenesis), splA/ryanodine receptor domain and SOCS box containing 2 (proteasome degradation), and ribosomal protein S2 (ribosome biogenesis). Gene expression analyses showed a significant association ($P < .05$) for rs4711689 with transcription factor EB, rs6469656 with eukaryotic translation initiation factor 3, subunit H, rs11064437 with splA/ryanodine receptor domain and SOCS box containing 2, and rs6061231 with ribosomal protein S2.

CONCLUSIONS: We identified susceptibility loci and genes associated with CRC risk, linking CRC predisposition to steroid hormone, protein synthesis and degradation, and autophagy pathways and providing added insight into the mechanism of CRC pathogenesis.

Contribution of a Non-classical HLA Gene, HLA-DOA, to the Risk of Rheumatoid Arthritis.

Despite the progress in human leukocyte antigen (HLA) causal variant mapping, independent localization of major histocompatibility complex (MHC) risk from classical HLA genes is challenging. Here, we conducted a large-scale MHC fine-mapping analysis of rheumatoid arthritis (RA) in a Japanese population (6,244 RA cases and 23,731 controls) population by using HLA imputation, followed by a multi-ethnic validation study including east Asian and European populations ($n = 7,097$ and 23,149, respectively). Our study identified an independent risk of a synonymous mutation at HLA-DOA, a non-classical HLA gene, on anti-citrullinated protein autoantibody (ACPA)-positive RA risk ($p = 1.4 \times 10^{-9}$), which demonstrated a cis-expression quantitative trait loci (cis-eQTL) effect on HLA-DOA expression. Trans-ethnic comparison revealed different linkage disequilibrium (LD) patterns in HLA-DOA and HLA-DRB1, explaining the observed HLA-DOA variant risk heterogeneity among ethnicities, which was most evident in the Japanese population. Although previous HLA fine-mapping studies have identified amino acid polymorphisms of the classical HLA genes as driving genetic susceptibility to disease, our study additionally identifies the dosage contribution of a non-classical HLA gene to disease etiology. Our study contributes to the understanding of HLA immunology in human diseases and suggests the value of incorporating additional ancestry in MHC fine-mapping.

2. Epidemiological analysis of various diseases

Overview of BioBank Japan Follow-up Data in 32 Diseases

Background: We established a patient-oriented biobank cohort, BioBank Japan with cooperation of about 200,000 patients, suffering from any of 47 common diseases. Among 47 diseases, we focused on 32 diseases for follow-up survey which may encompass poor vital prognosis of patients and collected their survival information including cause of death. We conducted survival analysis for all subjects to get an overview of BioBank Japan follow-up data.

Methods: 141,612 participants were included for follow-up survey. The survival data were last updated in 2014. Kaplan-Meier survival analysis was performed after categorizing the subjects according to sex, age-group, and disease status. Relative survival rates were estimated by using a survival-rate table of Japanese general population.

Results: Of 141,612 subjects (56.48% male) with 1,087,434 person-years, 35,482 deceased for the follow-up duration with a 97.0% follow-up rate. Mean age at the enrollment was 64.24 years in males and 63.98 in females. 5-year and 10-year relative survival rates of the all subjects were 0.944 and 0.911, respectively with median follow-up of 8.40 years. Survival analysis showed subjects with pancreatic cancer had worst prognosis (0.184 of 10-year relative survival), while those with dyslipidemia had most favorable prognosis (1.013). The most common cause of death was malignant neoplasms, though a number of subjects died from diseases other than their registered disease(s).

Conclusions: To our knowledge, this is the first report to perform follow-up survival analysis across various common diseases. Further studies, using detailed clinical and genomic information, shall identify predictors of mortality in patients with common diseases, contributing to implementation of personalized medicine.

Cross-sectional analysis of BioBank Japan Clinical Data: A Large Cohort of 200,000 Patients with 47 Common Diseases

Background: To implement personalized medicine, we established a large-scale patient cohort, BioBank Japan in 2003. BioBank Japan contains DNA and serum derived from about 200,000 patients with 47 diseases as well as their clinical information. Serum and clinical information were collected annually until 2012.

Methods: We analyzed baseline clinical informa-

tion at the enrollment including age, sex, BMI, hypertension, smoking and drinking status across 47 diseases, and compared the results with Japanese national public database; Patient Survey and National Health and Nutrition Survey. We conducted multivariate logistic-regression models adjusted for sex and age, to assess the association of family history with disease development.

Results: Analysis of clinical information indicated high association of smoking with COPD, drinking with esophageal cancer, high BMI with metabolic diseases, and hypertension with cardiovascular diseases. Comparison with the public database identified almost comparable distribution in sex and age, life-style and physical status. The logistic-regression analysis showed that individuals with family history of keloid exhibited quite high odds ratio compared with those without family history, indicating the strong impact of host genetic factor(s) on disease onset.

Conclusions: Cross-sectional analysis of clinical information at the enrollment unwrapped characteristics of the present cohort, which were mostly consistent with those in public database. Analysis of family history revealed the impact of host genetic factors in each disease. BioBank Japan, distributing the clinical information as well as DNA and serum samples publicly, could be a fundamental infrastructure for the implementation of personalized medicine.

3. Genes playing significant roles in human cancers

EPSIN 3, a novel p53 target, regulates the apoptotic pathway and gastric carcinogenesis

BACKGROUND & AIM: p53 activation by cellular stresses induces the transcription of hundreds of its target genes. To elucidate the entire picture of its downstream pathway, we screened a cDNA microarray dataset of adriamycin-treated HCT116 $p53^{-/-}$ or $p53^{+/+}$ cells and identified EPSIN 3 as a novel p53 target. *METHODS:* Potential p53 binding sequences in the *EPSIN 3* locus were evaluated by reporter and CHIP assays. To investigate the role of EPSIN 3 in the p53 downstream pathway, we assessed DNA damage-induced apoptosis in EPSIN 3-knockdown HCT116 cells or *Epsin 3*-deficient mice. In addition, we evaluated *EPSIN 3* expression levels in various tissues, including gastric adenocarcinoma, human gastric mucosa with or without *Helicobacter pylori* infection, and mouse acute gastritis tissues induced by indomethacin. *RESULTS:* In response to DNA damage, p53 induced the expression of EPSIN 3 through the p53 binding elements in the *EPSIN 3* promoter and the first intron. Knockdown of EPSIN 3 resulted in resistance to DNA damage-induced apoptosis both in vitro and in vivo. *EPSIN 3* expression was downregulated in gastric cancer tissues compared with normal tissues. In addition, *Helicobacter pylori* infection and indomethacin-induced acute gastritis repressed *EPSIN 3* expression in gastric mucosa. *CONCLUSIONS:* EPSIN 3 is a novel p53 target and a key mediator of apoptosis. Chronic or acute mucosal inflammation as well as p53 inactivation induced downregulation of *EPSIN 3* and subsequently caused apoptosis resistance, which is a hallmark of cancer cells.

Regulation of myo-inositol biosynthesis by p53-ISYNA1 pathway.

In response to various cellular stresses, p53 exerts its tumor suppressive effects such as apoptosis, cell cycle arrest, and senescence through the induction of its target genes. Recently, p53 was shown to control cellular homeostasis by regulating energy metabolism, glycolysis, antioxidant effect, and autophagy. However, its function in inositol synthesis was not reported. Through a microarray screening, we found that five genes related with myo-inositol metabolism were induced by p53. DNA damage enhanced intracellular myo-inositol content in HCT116 $p53^{+/+}$ cells, but not in HCT116 $p53^{-/-}$ cells. We also indicated that inositol 3-phosphate synthase (ISYNA1) which encodes an enzyme essential for myo-inositol biosynthesis as a direct target of p53. Activated p53 regulated ISYNA1 expression through p53 response element in the seventh exon. Ectopic ISYNA1 expression increased myo-inositol levels in the cells and suppressed tumor cell growth. Knockdown of ISYNA1 caused resistance to adriamycin treatment, demonstrating the role of ISYNA1 in p53-mediated growth suppression. Furthermore, ISYNA1 expression was significantly associated with p53 mutation in bladder, breast cancer, head and neck squamous cell carcinoma, lung squamous cell carcinoma, and pancreatic adenocarcinoma. Our findings revealed a novel role of p53 in myo-inositol biosynthesis which could be a potential therapeutic target.

Cystatin C as a p53-inducible apoptotic mediator that regulates cathepsin L activity.

In response to various cellular stresses, p53 is activated and inhibits malignant transformation through the transcriptional regulation of its target genes. However, the full picture of the p53 downstream pathway still remains to be elucidated. Here we identified cystatin C, a major inhibitor of cathepsins, as a novel p53 target. In response to DNA damage, activated p53 induced cystatin C expression through p53 binding sequence in the first intron. We showed that cathepsin L activity was decreased in HCT116 $p53^{+/+}$ cells after adriamycin

treatment, but not in HCT116 p53 $p53^{-/-}$ cells. We also found that knockdown of cystatin C reduced adriamycin-induced caspase-3 activation. Cystatin C expression was significantly downregulated in breast cancer cells with p53 mutations, and decreased cystatin C expression was associated with poor prognosis of breast cancer. Our findings revealed an important role of the p53-cystatin C pathway in human carcinogenesis.

Publications

[1] Y. Shiwa, T. Hachiya, R. Furukawa, H. Ohmomo, K. Ono, H. Kudo, J. Hata, A. Hozawa, M. Iwasaki, K. Matsuda, N. Minegishi, M. Satoh, K. Tanno, T. Yamaji, K. Wakai, J. Hitomi, Y. Kiyohara, M. Kubo, H. Tanaka, S. Tsugane, M. Yamamoto, K. Sobue, A. Shimizu, Adjustment of Cell-Type Composition Minimizes Systematic Bias in Blood DNA Methylation Profiles Derived by DNA Collection Protocols, PloS one, 11 (2016) e0147519.

[2] J. Mori, C. Tanikawa, Y. Funauchi, P.H. Lo, Y. Nakamura, K. Matsuda, Cystatin C as a p53-inducible apoptotic mediator which regulates Cathepsin L activity, Cancer Sci, (2016).

[3] S.K. Low, K. Fukunaga, A. Takahashi, K. Matsuda, F. Hongo, H. Nakanishi, H. Kitamura, T. Inoue, Y. Kato, Y. Tomita, S. Fukasawa, T. Tanaka, K. Nishimura, H. Uemura, I. Hara, M. Fujisawa, H. Matsuyama, K. Hashine, K. Tatsugami, H. Enokida, M. Kubo, T. Miki, T. Mushiroda, Association Study of a Functional Variant on ABCG2 Gene with Sunitinib-Induced Severe Adverse Drug Reaction, PloS one, 11 (2016) e0148177.

[4] T. Koguchi, C. Tanikawa, J. Mori, Y. Kojima, K. Matsuda, Regulation of myo-inositol biosynthesis by p53-ISYNA1 pathway, Int J Oncol, 48 (2016) 2415-2424.

[5] C. Zeng, K. Matsuda, W.H. Jia, J. Chang, S.S. Kweon, Y.B. Xiang, A. Shin, S.H. Jee, D.H. Kim, B. Zhang, Q. Cai, X. Guo, J. Long, N. Wang, R. Courtney, Z.Z. Pan, C. Wu, A. Takahashi, M.H. Shin, K. Matsuo, F. Matsuda, Y.T. Gao, J.H. Oh, S. Kim, K.J. Jung, Y.O. Ahn, Z. Ren, H.L. Li, J. Wu, J. Shi, W. Wen, G. Yang, B. Li, B.T. Ji, Genetics, C. Epidemiology of Colorectal Cancer, H. Brenner, R.E. Schoen, S. Kury, S. Colorectal Transdisciplinary, S.B. Gruber, F.R. Schumacher, S.L. Stenzel, R. Colon Cancer Family, G. Casey, J.L. Hopper, M. A. Jenkins, H.R. Kim, J.Y. Jeong, J.W. Park, K. Tajima, S.H. Cho, M. Kubo, X.O. Shu, D. Lin, Y.X. Zeng, W. Zheng, Identification of Susceptibility Loci and Genes for Colorectal Cancer Risk, Gastroenterology, 150 (2016) 1633-1645.

[6] A. Mahajan, A.R. Rodan, T.H. Le, K.J. Gaulton, J. Haessler, A.M. Stilp, Y. Kamatani, G. Zhu, T. Sofer, S. Puri, J.N. Schellinger, P.L. Chu, S. Cechova, N. van Zuydam, S. Consortium, P. BioBank Japan, J. Arnlov, M.F. Flessner, V. Giedraitis, A.C. Heath, M. Kubo, A. Larsson, C.M. Lindgren, P.A. Madden, G.W. Montgomery, G.J. Papanicolaou, A.P. Reiner, J. Sundstrom, T.A. Thornton, L. Lind, E. Ingelsson, J. Cai, N.G. Martin, C. Kooperberg, K. Matsuda, J.B. Whitfield, Y. Okada, C.C. Laurie, A.P. Morris, N. Franceschini, Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity, American journal of human genetics, 99 (2016) 636-646.

[7] Y. Okada, A. Suzuki, K. Ikari, C. Terao, Y. Kochi, K. Ohmura, K. Higasa, M. Akiyama, K. Ashikawa, M. Kanai, J. Hirata, N. Suita, Y.Y. Teo, H. Xu, S.C. Bae, A. Takahashi, Y. Momozawa, K. Matsuda, S. Momohara, A. Taniguchi, R. Yamada, T. Mimori, M. Kubo, M.A. Brown, S. Raychaudhuri, F. Matsuda, H. Yamanaka, Y. Kamatani, K. Yamamoto, Contribution of a Non-classical HLA Gene, HLA-DOA, to the Risk of Rheumatoid Arthritis, American journal of human genetics, 99 (2016) 366-374.

[8] M.R. Han, J. Long, J.Y. Choi, S.K. Low, S.S. Kweon, Y. Zheng, Q. Cai, J. Shi, X. Guo, K. Matsuo, M. Iwasaki, C.Y. Shen, M.K. Kim, W. Wen, B. Li, A. Takahashi, M.H. Shin, Y.B. Xiang, H. Ito, Y. Kasuga, D.Y. Noh, K. Matsuda, M.H. Park, Y.T. Gao, H. Iwata, S. Tsugane, S.K. Park, M. Kubo, X.O. Shu, D. Kang, W. Zheng, Genome-wide association study in East Asians identifies two novel breast cancer susceptibility loci, Hum Mol Genet, (2016).

[9] C.D. Middlebrooks, A.R. Banday, K. Matsuda, K.I. Udquim, O.O. Onabajo, A. Paquin, J.D. Figueroa, B. Zhu, S. Koutros, M. Kubo, T. Shuin, N.D. Freedman, M. Kogevinas, N. Malats, S.J. Chanock, M. Garcia-Closas, D.T. Silverman, N. Rothman, L. Prokunina-Olsson, Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors, Nat Genet, 48 (2016) 1330-1338.

[10] S. Hirabayashi, K. Ohki, K. Nakabayashi, H. Ichikawa, Y. Momozawa, K. Okamura, A. Yaguchi, K. Terada, Y. Saito, A. Yoshimi, H. Ogata-Kawata, H. Sakamoto, M. Kato, J. Fujimura, M. Hino, A. Kinoshita, H. Kakuda, H. Kurosawa, K. Kato, R. Kajiwara, K. Moriwaki, T. Morimoto, K. Nakamura, Y. Noguchi, T. Osumi, K. Sakashita, J. Takita, Y. Yuza, K.

Matsuda, T. Yoshida, K. Matsumoto, K. Hata, M. Kubo, Y. Matsubara, T. Fukushima, K. Koh, A. Manabe, A. Ohara, N. Kiyokawa, G. Tokyo Children's Cancer Study, ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype, Haematologica, 102 (2017) 118-129.

[11] T. Hachiya, Y. Kamatani, A. Takahashi, J. Hata, R. Furukawa, Y. Shiwa, T. Yamaji, M. Hara, K. Tanno, H. Ohmomo, K. Ono, N. Takashima, K. Matsuda, K. Wakai, N. Sawada, M. Iwasaki, K. Yamagishi, T. Ago, T. Ninomiya, A. Fukushima, A. Hozawa, N. Minegishi, M. Satoh, R. Endo, M. Sasaki, K. Sakata, S. Kobayashi, K. Ogasawara, M. Nakamura, J. Hitomi, Y. Kita, K. Tanaka, H. Iso, T. Kitazono, M. Kubo, H. Tanaka, S. Tsugane, Y. Kiyohara, M. Yamamoto, K. Sobue, A. Shimizu, Genetic Predisposition to Ischemic Stroke: A Polygenic Risk Score, Stroke, (2016).

[12] F.J. van Rooij, R. Qayyum, A.V. Smith, Y. Zhou, S. Trompet, T. Tanaka, M.F. Keller, L.C. Chang, H. Schmidt, M.L. Yang, M.H. Chen, J. Hayes, A.D. Johnson, L.R. Yanek, C. Mueller, L. Lange, J.S. Floyd, M. Ghanbari, A.B. Zonderman, J.W. Jukema, A. Hofman, C.M. van Duijn, K.C. Desch, Y. Saba, A.B. Ozel, B.M. Snively, J.Y. Wu, R. Schmidt, M. Fornage, R.J. Klein, C.S. Fox, K. Matsuda, N. Kamatani, P.S. Wild, D.J. Stott, I. Ford, P.E. Slagboom, J. Yang, A.Y. Chu, A.J. Lambert, A.G. Uitterlinden, O.H. Franco, E. Hofer, D. Ginsburg, B. Hu, B. Keating, U.M. Schick, J.A. Brody, J.Z. Li, Z. Chen, T. Zeller, J.M. Guralnik, D.I. Chasman, L.L. Peters, M. Kubo, D.M. Becker, J. Li, G. Eiriksdottir, J.I. Rotter, D. Levy, V. Grossmann, K.V. Patel, C.H. Chen, P. BioBank Japan, P.M. Ridker, H. Tang, L.J. Launer, K.M. Rice, R. Li-Gao, L. Ferrucci, M.K. Evans, A. Choudhuri, E. Trompouki, B.J. Abraham, S. Yang, A. Takahashi, Y. Kamatani, C. Kooperberg, T.B. Harris, S.H. Jee, J. Coresh, F.J. Tsai, D.L. Longo, Y.T. Chen, J.F. Felix, Q. Yang, B. M. Psaty, E. Boerwinkle, L.C. Becker, D.O. Mook-Kanamori, J.G. Wilson, V. Gudnason, C. J. O'Donnell, A. Dehghan, L.A. Cupples, M.A. Nalls, A.P. Morris, Y. Okada, A.P. Reiner, L.I. Zon, S.K. Ganesh, Genome-wide Trans-ethnic Meta-analysis Identifies Seven Genetic Loci Influencing Erythrocyte Traits and a Role for RBPMS in Erythropoiesis, American journal of human genetics, 100 (2017) 51-63.

[13] Jinichi Mori, Naomi Ohnishi, Yuki Funauchi, Osamu Toyoshima, Koji Ueda, Koichi Matsuda, EPSIN 3, a novel p53 target, regulates the apoptotic pathway and gastric carcinogenesis, Neoplasia, (2017).

[14] Makoto Hirata, Yoichiro Kamatani, Toshiharu Ninomiya, Akiko Tamakoshi, Zentaro Yamagata, Michiaki Kubo, Kaori Muto, Yutaka Kiyohara, Taisei Mushiroda, Yoshinori Murakami, Koichiro Yuji, Yoichi Furukawa, Hitoshi Zembutsu, Toshihiro Tanaka, Yozo Ohnishi, Yusuke Nakamura, BioBank Japan Cooperative Hospital Group, Koichi Matsuda, Overview of BioBank Japan Follow-up Data in 32 Diseases, Journal of epidemiology, (2017).

[15] Makoto Hirata, Akiko Nagai, Yutaka Kiyohara, Toshiharu Ninomiya, Akiko Tamakoshi, Zentaro Yamagata, Michiaki Kubo, Kaori Muto, Taisei Mushiroda, Yoshinori Murakami, Koichiro Yuji, Yoichi Furukawa, Hitoshi Zembutsu, Toshihiro Tanaka, Yozo Ohnishi, Yusuke Nakamura, BioBank Japan Cooperative Hospital Group, Koichi Matsuda, Cross-sectional analysis of BioBank Japan Clinical Data: A Large Cohort of 200,000 Patients with 47 Common Diseases, Journal of epidemiology, (2017).

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. Identification of cis-regulatory elements that control spliced-leader trans-splicing in the primitive chordate, *Ciona intestinalis*

Rui Yokomori, Kotaro Shimai[1], Koki Nishitsuji[2], Yutaka Suzuki[3], Takehiro G. Kusakabe[1] and Kenta Nakai: [1]Faculty of Science and Engineering, Konan Univ, [2]Graduate School of Life Sci., Univ. of Hyogo, [3]Graduate School of Frontier Sciences, U. Tokyo

In the primitive chordate, *Ciona intestinalis*, about half of mRNAs undergo specialized trans-splicing, in which a small non-coding RNA, called spliced-leader (SL) RNA, is replaced by the non-coding 5'-end region of pre-mRNA, called outron. However, its mechanism is still poorly understood. In this study, we identify cis-regulatory elements that play an important role in the spliced-leader trans-splicing. We found that some GT-rich elements are enriched in the first 50bp of outron regions. Interestingly, these enriched elements can be binding motifs of the 3'-end sites of the SL RNA, suggesting the possibility that the SL RNA is recruited by binding to the GT-rich elements in outron regions. This is similar to the binding of the U1 snRNA to the donor site in cis-splicing. Moreover, we found the peak of the putative binding sites of SR proteins, which is one of splicing enhancer proteins and enhance the U1 snRNA binding, at upstream regions of GT-rich elements. Taken together, it seems that the mechanism of spliced-leader trans-splicing is basically similar to that of cis-splicing and an alternative splicing that utilizes specialized small non-coding RNA and results in transcripts with different non-coding regions.

2. Cell type specific features of non-CpG methylation

Jong-Hun Lee and Kenta Nakai

DNA methylation is one of the most important epigenetic markers that regulate gene expression. Although most of the methylation occurs in the CpG context, recent studies have found that the methylation in the non-CpG context is also functional in embryonic stem cell (ESC) and neurons. Interestingly, the non-CpG methylation in these two cell types shows differential distribution and function. In this study, we attempted to uncover the reason for cell type specific non-CpG methylation pattern by integrating whole genome bisulfite sequencing data (WGBS) with hundreds of tran-

scriptomes, as well as DNA methyltransferase 3a and 3b (DNMT3a, and 3b, respectively) -knock out methylomes. We found that the differential activity of DNMT3-family induces the distinct features of non-CpG methylation in ESC and neuron, such as interacting DNA motifs or correlation with gene expression. These results give insights into the mechanism of cell type specific regulation by non-CpG methylation.

3. Application of topic models to motif finding algorithms

Josep Basha Gutierrez and Kenta Nakai

Topic models are statistical models which try to discover the topics contained in a set of documents to reveal their hidden structure. We analogize this problem to the discovery of the structure of the motifs contained in a set of biological sequences. To study if this approach is possible, we developed two methods. First, an algorithm relying solely on topic models, and, in addition, a previously developed method which we enhanced with the addition of topic models. The first approach showed a remarkable performance, and the second approach noticeably improved its accuracy and reached outstanding levels in sensitivity and in overall performance at site level. We conclude that the application of topic models has proven to be a valid method for motif finding and that it will be useful in future studies.

4. Modeling the cis-regulatory modules of genes expressed in developmental stages of *Drosophila melanogaster*

Yosvany López, Alexis Vandenbon[4], Akinao Nose[5] and Kenta Nakai: [4]WPI-iFReC, Osaka University; [5]Department of Complexity Science and Engineering, Graduate School of Frontier Sciences

Since transcription is the first step in the regulation of gene expression, our understanding of how transcription factors bind to their DNA binding motifs has become absolutely necessary. It has been shown that the promoters of genes with similar expression profiles share common structural patterns. We performed an extensive study of the regulatory regions of genes expressed in 22 developmental stages of *Drosophila melanogaster*. We used the combination of four types of structural features, such as positioning of individual motifs relative to the transcription start site, orientation, pairwise distance between motifs, and presence of motifs anywhere in the promoter to identify models of gene regulation. RNA-sequencing data was utilized to create and validate the 22 models, resulting in 19 (86.4%) statistically significant models. Each model yielded a set of highly informative features, which were used to search for genes with similar biological functions.

5. Stable feature selection based on the ensemble L_1-norm support vector machine for biomarker discovery

Myungjin Moon and Kenta Nakai

Biomarker discovery has become one of the most significant research areas in the biomedical field. Owing to the presence of high-throughput technologies, genomic data have become widely available. However, they tend to be noisy with a large number of genes (high-dimensional features) and a small number of samples; thus, conventional feature selection approaches for biomarker discovery might be problematic. We propose a stable feature selection method for high-dimensional datasets. We apply an ensemble L_1-norm support vector machine to efficiently reduce irrelevant features, considering the stability of features. The proposed methodology is evaluated by classifying the stage of renal clear cell carcinoma with RNA-seq data. A comparison with established algorithms enabled us to prove the superior performance of our method. The proposed approach is expected to be applicable to many other studies aimed at biomarker discovery.

6. Biomarker discovery in cancer cell panel by an unsupervised approach

Naoya Fujita, Katsuhiko Murakami[6] and Kenta Nakai: [6]School of Bioscience and Biotechnology, Tokyo University of Technology

Precision medicine for cancer patients with molecular targeted drugs and predictive biomarkers will lead to paradigm shift from one-size-fits-all medicine to responder enrichment medicine. Our research objective is to detect promising biomarker candidates from multi-dimensional genomic and pharmacologic data in a collection of cell lines. An unsupervised approach could detect previously reported relationships between drugs and biomarkers. Our approach clustered a well-known pair of FDA approved drug and predictive biomarker (BRAF inhibitor and BRAF mutation). Furthermore, new biomarker candidates have also been prioritized.

7. Assessing the impact of bacterial and viral contamination on human cell profiles

Sung-Joon Park, Josep Basha Gutierrez, Myungjin Moon, Kenta Nakai

During the course of sampling and culturing

cells, bacterial and viral contamination and infection can occur and lead to misinterpretations. Particularly in the field of allogenic cell transplantation therapy, eliminating contamination is a crucial issue because of its potential contribution to the development of various diseases. In this study, we implemented a bioinformatics strategy that identifies sequenced reads derived from multiple sources. In the case of RNA-seq data, our strategy simultaneously quantifies the expression variability of host genes due to the contamination. We applied it to over 300 public human RNA-seq data sets, and found 103-105 potential contaminant reads per one million human mapped reads. Remarkably, specific gene expression levels were associated with the level of contamination, which suggests the existence of functional interplay between host transcription and contamination. These results illustrate that the proposed pipeline can be used for routinely assessing the quality of NGS experiments and the safety of cultured cells.

8. Developing a Repository System for High-dimensional Chromatin Structure Information

Sung-Joon Park and Kenta Nakai

The advent of high-throughput sequencing technology for the profiling of long-range chromatin interactions, such as Hi-C and ChIA-PET, has greatly enhanced our ability to capture the functional importance of structural domains remotely positioned on the one-dimensional genome sequence. In this study, to manage and integrate the chromatin structure information, we are developing a web-based repository system (https://openlooper.hgc.jp/) as a sub project of the research project "Chromosome orchestration system" (http://www.chromosomeos.com)". This system allows users to share their deposited data along with other heterogeneous public data (e.g. TF ChIP-seq data), which may promote various downstream studies, such as the functional characterization of enhancers. Our system offers valuable tools and resources in the new era of genetic and epigenetic research.

9. ZBTB16 as a Downstream Target Gene of Osterix Regulates Osteoblastogenesis of Human Multipotent Mesenchymal Stromal Cells

Satoru Onizuka[7,8], Takanori Iwata[7], Sung-Joon Park, Kenta Nakai, Masayuki Yamato[8], Teruo Okano[8] and Yuichi Izumi[7]: [7]Department of Periodontology, Graduate School of Medical Dental Sciences, Tokyo Medical Dental University, [8]Institute of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University

Human mesenchymal stem cells (hMSCs) possess the ability to differentiate into osteoblasts, and they can be utilized as a source for bone regenerative therapy. Osteoinductive pretreatment that induces the osteoblastic differentiation of hMSCs in vitro is widely used for bone tissue engineering. However, the molecular basis of osteoblastic differentiation induced by osteoinductive medium (OIM) is still unknown. In this study, we used a next-generation sequencer to investigate the changes in gene expression during the osteoblastic differentiation of hMSCs. Our whole-transcriptome analysis revealed that the expression of zinc finger and BTB domain containing 16 (ZBTB16) is significantly increased during the osteoblastogenesis. In addition, the siRNA-mediated ZBTB16 silencing experiment evidenced the functional impact of ZBTB16 on the activity of alkaline phosphatase (ALP) and osteogenic gene expression including Osterix (Osx). These findings suggest that ZBTB16 acts as a downstream transcriptional regulator of Osx and can be useful as a late marker of osteoblastic differentiation.

10. Whole-genome sequencing data analysis for the identification of driver mutations in Esophageal Squamous Cell Carcinoma

Munmee Dutta, Masashi Fujita[9], Ashwini Patil, Kenta Nakai and Hidewaki Nakagawa[9]: [9]RIKEN Center for Integrative Medical Sciences, Laboratory for Genome Sequencing Analysis

Esophageal squamous cell carcinoma (ESCC) is the eighth most common cancer type and sixth most common cause of cancer-related death worldwide. However, a complete mutational landscape of ESCC is yet to be established for the discovery of more candidate driver genes for better treatment options. In this study, we have attempted to uncover the driver mutations and genes in coding and non-coding regions of ESCC genome and link them to future therapeutic use. We used 16 pairs of tumor and normal whole-genome sequencing (WGS) data from patients with ESCC from a Japanese cohort. The structural variations (SVs) of these 16 pairs of tumor and normal data were analyzed using VarScan2 and Genomon2 tools. Furthermore, we analyzed the somatic copy number variations with these 16 sample pairs using VarScan2. We have identified some of the novel cancer driver genes affected by SVs in ESCC, some of which are reported to be cancer-causing in other cancer-types. This study will help understand the comprehensive mutational landscape of ESCC with special reference to the non-coding and structural variation mutations.

11. Molecular characterization of TCR clones using 3D protein structure modelling of the TCR/pMHC complex with hgp100 antigen

Yasuo Ouchi[10], Ashwini Patil, Yusuke Tamura[11], Hiroshi Nishimasu[12], Naoki Takemura[10,11], Takeshi Sato[13], Yasumasa Kimura[13], Osamu Nureki[12], Kenta Nakai, Hiroshi Kiyono[12,14,15], Satoshi Uematsu[10,11]: [10]Dept of Mucosal Immunology, Chiba University, [11]Division of Innate Immune Regulation, IMS, U. Tokyo, [12]Dept of Biol Sci, U. Tokyo, [13]Division of Systems Immunology, IMS, U. Tokyo, [14]Division of Mucosal Immunology, IMS, U. Tokyo, [15]Dept of Immunology, Chiba University

Adoptive immunotherapy with genetically engineered T-cells has emerged as a promising novel strategy for cancer treatment. We selected the human (h) gp100 melanoma-associated tumour antigen as a model system, and cloned hgp100-specific high-avidity CTLs and their TCR sequences from hgp100-immunized mice. To obtain structural insights into the recognition of hgp100 by the TCR, we predicted the 3D structures of the TCR-MHC-hgp100 complex using semi-automatic modelling and docking. Consistent with a theoretical binding mode, our model indicated that the hgp100-specific TCR precisely binds to the surface H2-Db residues adjacent to the bound hgp100 peptide. Our computational analysis showed the structural significance of the IFN-γ high-expressing TCR clone for its antitumour activity.

12. Organism-level analysis of vaccination reveals networks of protection across tissues

Motohiko Kadoki[16], Ashwini Patil, Cornelius C. Thaiss[16], Donald J. Brooks[16], Surya Pandey[16], Deeksha Deep[16], David Alvarez[18], Ulrich H. von Andrian[18], Amy J. Wagers[17], Kenta Nakai, Tarjei S. Mikkelsen[17], Magali Soumillon[17], Nicolas Chevrier[16]: [16]Faculty of Arts & Sciences Center for Systems Biology, [17]Dept of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard Univ., [18]Department of Microbiology and Immunobiology, Harvard Medical School

A fundamental challenge in immunology is to decipher the principles governing immune responses at the whole-organism scale. Immune signal propagation was observed within and between organs to obtain a dynamic map of immune processes at the organismal level. By analyzing ligand-receptor connectivity across tissues, type I IFNs were found to trigger a whole-body antiviral state within hours upon skin vaccination. Combining parabiosis and single-cell analyses, a multi-organ network of tissue-resident memory T cells was observed that functionally adapt to their environment so as to stop viral particles as they progress from one tissue to the next.

13. Predicting MoRFs in protein sequences using HMM Profiles

Ronesh Sharma[19,20], Shiu Kumar[19,20], Tatsuhiko Tsunoda[21,22,23], Ashwini Patil[§], Alok Sharma[20,21,22,23§]: [19]School of Electrical and Electronics Engg, Fiji National University, Suva, Fiji, [20]School of Engg. and Physics, The Univ. of the South Pacific, Suva, Fiji, [21]CREST, JST, [22]RIKEN Center for Integrative Medical Science, [23]Medical Research Institute, Tokyo Medical and Dental University

We introduced a method using hidden Markov model (HMM) profiles to accurately identify the location of binding regions (MoRFs) in disordered protein sequences. Using windowing technique, HMM profiles were utilised to extract features from protein sequences and support vector machines (SVM) were used to calculate a propensity score for each residue. The performance of the method was compared to that of other MoRF predictors; MoRFpred and ANCHOR. The results showed that the proposed method outperforms these two predictors.

14. Multidisciplinary insight into clonal expansion of HTLV-1-infected cells in adult T-cell leukemia via modeling by deterministic finite automata coupled with high-throughput sequencing

Amir Farmanbar[*], Sanaz Firouzi[24*§], Sung-Joon Park, Kenta Nakai, Kaoru Uchimaru[24,25], Toshiki Watanabe[24,26§]: [24]Dept of Comp Biol and Medical Sciences, Grad School of Frontier Sciences, U. Tokyo, [25]Hematology/Oncology, Research Hospital, IMS, U. Tokyo, [26]Dept of Advanced Medical Innovation, St. Marianna Univ.

Clonal expansion of leukemic cells leads to onset of adult T-cell leukemia (ATL), an aggressive lymphoid malignancy with a very poor prognosis. Infection with human T-cell leukemia virus type-1 (HTLV-1) is the direct cause of ATL onset, and integration of HTLV-1 into the human genome is essential for clonal expansion of leukemic cells. We analyzed clinical samples from HTLV-1-infected individuals with a broad range of proviral loads using a high-throughput methodology that enables isolation of HTLV-1 integration sites and accurate measurement of the size of infected clones. We categorized clones into four size groups, "very small", "small", "big", and "very big", based on the patterns of clonal growth and observed clone sizes and developed a model of clonal expansion based on observed clonality data. Our model provides a basic understanding as well as a unique perspective for clarifying the mechanisms of clonal expansion in ATL.

15. Genome Informatics Analysis of Regulation between Long Non-coding RNA Genes and Transcription Factor Genes

Taiju Yamada, Izuho Hatada[27] and Kenta Nakai: [27]Institute for Molecular and Cellular Regulation, Gunma Univ.

We examined the prevalence of long non-coding RNAs in the vicinity of genes encoding transcription factors in the genomes of various eukaryotes. We checked the distribution of the distance between lncRNA and gene sequences along with co-expression correlation and evolutionary conversation. We found that long non-coding RNA loci are present in close proximity to genes encoding transcription factors at statistically significant levels and are potentially co-expressing. This indicates a potential role for the long non-coding RNAs in the regulation of expression of transcription factors.

Publications

Basha Gutierrez, J., Nakai, K. A study on the application of topic models to motif finding algorithms. *BMC Bioinformatics,* 17: 129-138, 2016.

Farmanbar, A., Firouzi, S., Park, S.J., Nakai, K., Uchimaru, K., and Watanabe, T. Multidisciplinary insight into clonal expansion of HTLV-1-infected cells in adult T-cell leukemia via modeling by deterministic finite automata coupled with high-throughput sequencing. *BMC Medical Genomics*, in press.

Fukumura, K., Wakabayashi, S., Kataoka, N., Sakamoto, H., Suzuki, Y., Nakai, K., Mayeda, A., Inoue, K. The Exon Junction Complex Controls the Efficient and Faithful Splicing of a Subset of Transcripts Involved in Mitotic Cell-Cycle Progression. *Int J Mol Sci,* 17, 2016.

Moon, M., Nakai, K. Stable feature selection based on the ensemble L 1 -norm support vector machine for biomarker discovery. *BMC Genomics,* 17: 65-74, 2016.

Onizuka, S., Iwata, T., Park, S.J., Nakai, K., Yamato, M., Okano, T., Izumi, Y. ZBTB16 as a Downstream Target Gene of Osterix Regulates Osteoblastogenesis of Human Multipotent Mesenchymal Stromal Cells. *J Cell Biochem,* 117: 2423-34, 2016.

Park, S.J., Komiyama, Y., Suemori, H., Umezawa, A., Nakai, K. OpenTein: a database of digital whole-slide images of stem cell-derived teratomas. *Nucleic Acids Res,* 44: D1000-4, 2016.

Park, S.J., Saito-Adachi, M., Komiyama, Y., Nakai, K. Advances, practice, and clinical perspectives in high-throughput sequencing. *Oral Dis,* 22: 353-64, 2016.

Park, S.J., Nakai, K. Chapter 7: Computational Inference of Gene Regulation from Whole-Transcriptome Analysis of Early Embryos, in (I. Ivanov, X. Qian and R. Pal ed.) Emerging Research in the Analysis and Modeling of Gene Regulatory Networks, 241-279, IGI Global, 2016

Sasakura, Y., Ogura, Y., Treen, N., Yokomori, R., Park, S.J., Nakai, K., Saiga, H., Sakuma, T., Yamamoto, T., Fujiwara, S., Yoshida, K. Transcriptional regulation of a horizontally transferred gene from bacterium to chordate. *Proc Biol Sci,* 283, 2016.

Sasamoto, Y., Hayashi, R., Park, S.J., Saito-Adachi, M., Suzuki, Y., Kawasaki, S., Quantock, A.J., Nakai, K., Tsujikawa, M., Nishida, K. PAX6 Isoforms, along with Reprogramming Factors, Differentially Regulate the Induction of Cornea-specific Genes. *Sci Rep,* 6: 20807, 2016.

Sharma, R., Kumar, S., Tsunoda, T., Patil, A., Sharma, A. Predicting MoRFs in protein sequences using HMM profiles. *BMC Bioinformatics,* 17: 251-258, 2016.

Yokomori, R., Shimai, K., Nishitsuji, K., Suzuki, Y., Kusakabe, T.G., Nakai, K. Genome-wide identification and characterization of transcription start sites and promoters in the tunicate Ciona intestinalis. *Genome Res,* 26: 140-50, 2016.

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
Associate Professor	Yusuke Inoue, Ph.D.	准教授	博士(社会医学)	井　上　悠　輔
Project Associate Professor	Ayako Kamisato, Ph.D.	特任准教授	博士(法学)	神　里　彩　子
Project Assistant Professor	Hyunsoo Hong, Ph.D.	特任助教	博士(学術)	洪　　賢　秀

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

Japan Agency for Medical Research and Development (AMED) has commissioned to provide research ethics consultation to several large projects promoting genomic medicine, including The Biobank Japan (BBJ) Project (BBJP), Project for Development of Innovative Research on Cancer Therapeutics (P-DIRECT) and Project for Cancer Research and Therapeutic Evolution (P-CREATE). 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 issued semiannual newsletters for sample donors for transparency and information. Since 2014, BBJ started to store samples from National Hospital Organization (NHO), Japan Clinical Oncology Group (JCOG), and Japan Children's Cancer Group (JCCG). We supported to establish ethical policies for collecting these samples.

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 and advised those principal investigators to comply with ethical guidelines and provided sample consent forms. Since April of 2016, the new initiative called P-CREATE started. We have provided seminars for researchers about revision of Act on the Protection of Personal Information Act.

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

Japan Agency for Medical Research and Development (AMED) has also commissioned us to provide research ethics consultation to stem cell research since 2012. The program is called "research on the ethical, legal, and social implications related to regenerative medicine". In order to make regenerative medicine more concrete, it is essential to promote research development with a definite focus on clini-

cal applications and to establish a framework for clinical research at an early stage. In 2015, we provided more than 50 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 and governance of iPSC banking.

Based on our training workshops informed consent (IC) for members of stem-cell research teams and an educational workshop for a certain patients' group, we published a paper to discuss necessary aspects in promoting and conducting IC appropriately. Additionally, it emphasizes the importance of continuous education for improving the IC skills and competency of people responsible for obtaining IC.

We also 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. Current public support for human-animal chimera research in Japan is limited, despite high levels of scientific approval

In Japan, the national guidelines based on the Act on Regulation of Human Cloning Techniques limited the handling of human-animal chimeric embryos to "basic research" and prohibited implanting these embryos in animal uteruses or growing them for longer than 14 days. However, policy recommendations in this arena have undergone a remarkable shift in recent years, largely due to scientific achievements by Japanese researchers (e.g., Kobayashi et al., 2010) and concerns that the prohibitions would encourage researchers to leave Japan.

To assess public attitudes toward this potential policy shift, we conducted a pair of surveys before the aforementioned recommendations by the Expert Panel on Bioethics (2012) and 3 years later (2015) about their support for human-animal chimera research and regenerative medicine in general. We received about 5,000 responses from the general public and compared their answers to those from mem-

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

bers of JSRM, the Japanese Society for Regenerative Medicine.

The survey results indicate that the general population has considerably less support for human-animal chimera research than research scientists in Japan, despite overall high levels of support for regenerative medicine research. Even among those who supported making human-animal chimeras, the general public expressed several specific views regarding which animals should not be used in this kind of research. A larger proportion of the researchers supported the use of any species (42.4%). These answers differ from the findings of past surveys conducted overseas.

4. The point of decision-making and informed consent: patients' perspective and experiences of clinical trials in Japan

Clinical trials are a crucial step in the development of new medical technologies. Obtaining informed consent (IC) is a key component of clinical trials that ensures that potential participants have the necessary information for decision-making. While patients' involvement is essential to improve the systems and circumstances of clinical trials, their perspectives and experiences are rarely shared with others. This study aims to investigate patients' perspectives and experiences of IC in clinical trials in Japan. A mixed method approach was adopted, including an internet-based survey of 21,502 patients and in-depth interviews of 41 patients. In the survey, 12,506 responses were analyzed, focusing on their understanding of IC. Among the 12,506 patients, 2,330 had some experience of clinical trials. Among the 888 patients who remembered receiving an explanation from the medical staff, 93.6% responded that it was easy to understand. Most respondents (84.5%) who had received an IC form read it again and 54.5% did not spend much time thinking before deciding whether to participate. Through the in-depth interviews, we found the following: the point of decision-making and the meaning of IC for the patients. The patients made their "informal" decision prior to the "formal" IC process. They confirmed their "informal" decision during the IC process by collecting important information regarding the health risk of participation. From the patients' perspective, the point of decision-making is outside the IC process. While patients make an informal decision, both medical staff and patients should recognize that written consent is the formal decision.

Publications

1. Akiko Nagai, Makoto Hirata, Yoichiro Kamatani, Kaori Muto, Koichi Matsuda, Yutaka Kiyohara, Toshiharu Ninomiya, Akiko Tamakoshi, Zentaro Yamagata, Taisei Mushiroda, Yoshinori Murakami, Koichiro Yuji, Yoichi Furukawa, Hitoshi Zembutsu, Toshihiro Tanaka, Yozo Ohnishi, Yusuke Nakamura, BioBank Japan Cooperative Hospital Group, and Michiaki Kubo. Overview of the BioBank Japan Project: Study Design and Profile. Journal of Epidemiology, in press.
2. Makoto Hirata, Yoichiro Kamatani, Akiko Nagai, Yutaka Kiyohara, Toshiharu Ninomiya, Akiko Tamakoshi, Zentaro Yamagata, Michiaki Kubo, Kaori Muto, Taisei Mushiroda, Yoshinori Murakami, Koichiro Yuji, Yoichi Furukawa, Hitoshi Zembutsu, Toshihiro Tanaka, Yozo Ohnishi, Yusuke Nakamura, BioBank Japan Cooperative Hospital Group, and Koichi Matsuda. Cross-sectional analysis of BioBank Japan Clinical Data: A Large Cohort of 200,000 Patients with 47 Common Diseases. Journal of Epidemiology, in press.
3. Makoto Hirata, Akiko Nagai, Yoichiro Kamatani, Toshiharu Ninomiya, Akiko Tamakoshi, Zentaro Yamagata, Michiaki Kubo, Kaori Muto, Yutaka Kiyohara, Taisei Mushiroda, Yoshinori Murakami, Koichiro Yuji, Yoichi Furukawa, Hitoshi Zembutsu, Toshihiro Tanaka, Yozo Ohnishi, Yusuke Nakamura, BioBank Japan Cooperative Hospital Group, and Koichi Matsuda. Overview of BioBank Japan Follow-up Data in 32 Diseases. Journal of Epidemiology, in press.
4. Emiko Okada, Shigekazu Ukawa, Koshi Nakamura, Makoto Hirata, Akiko Nagai, Yoichiro Kamatani, Zentaro Yamagata, Yutaka Kiyohara, Toshiharu Ninomiya, Koichi Matsuda, Kaori Muto, Michiaki Kubo, Yusuke Nakamura, the BioBank Japan Cooperative Hospital Group, and Akiko Tamakoshi. Demographic and lifestyle factors and survival among patients with esophageal and gastric cancer: The Biobank Japan Project. Journal of Epidemiology, in press.
5. Akiko Tamakoshi, Koshi Nakamura, Shigekazu Ukawa, Emiko Okada, Makoto Hirata, Akiko Nagai, Koichi Matsuda, Kaori Muto, Yutaka Kiyohara, Zentaro Yamagata Toshiharu Ninomiya, Michiaki Kubo, Yusuke Nakamura, and BioBank Japan Cooperative Hospital Group. Characteristics and prognosis of Japanese colorectal cancer patients: the BioBank Japan Project. Journal of Epidemiology, in press.
6. Shigekazu Ukawa, Emiko Okada, Koshi Nakamura, Makoto Hirata, Akiko Nagai, Yoichiro Kamatani, Zentaro Yamagata, Yutaka Kiyohara, Toshiharu Ninomiya, Koichi Matsuda, Kaori

Muto, Michiaki Kubo, Yusuke Nakamura, the BioBank Japan Cooperative Hospital Group, and Akiko Tamakoshi. Characteristics of patients with liver cancer in the BioBank Japan Project. Journal of Epidemiology, in press.
7. Koshi Nakamura, Shigekazu Ukawa, Emiko Okada, Makoto Hirata, Akiko Nagai, Yoichiro Kamatani, Zentaro Yamagata, Yutaka Kiyohara, Toshiharu Ninomiya, Koichi Matsuda, Kaori Muto, Michiaki Kubo, Yusuke Nakamura,, the BioBank Japan Cooperative Hospital Group, and Akiko Tamakoshi. Characteristics and prognosis of Japanese male and female lung cancer patients: the BioBank Japan Project. Journal of Epidemiology, in press.
8. Koshi Nakamura, Emiko Okada, Shigekazu Ukawa, Makoto Hirata, Akiko Nagai, Yoichiro Kamatani, Zentaro Yamagata, Yutaka Kiyohara, Toshiharu Ninomiya, Koichi Matsuda,, Kaori Muto, Michiaki Kubo, Yusuke Nakamura,, the BioBank Japan Cooperative Hospital Group, and Akiko Tamakoshi. Characteristics and prognosis of Japanese female breast cancer patients: the BioBank Japan Project. Journal of Epidemiology, in press.
9. Shigekazu Ukawa, Koshi Nakamura, Emiko Okada, Makoto Hirata, Akiko Nagai, Yoichiro Kamatani, Zentaro Yamagata, Yutaka Kiyohara, Toshiharu Ninomiya, Koichi Matsuda, Kaori Muto, Michiaki Kubo, Yusuke Nakamura, the BioBank Japan Cooperative Hospital Group, and Akiko Tamakoshi. Clinical and Histopathological Characteristics of Patients with Prostate Cancer in the BioBank Japan Project. Journal of Epidemiology, in press.
10. Jun Hata, Akiko Nagai, Makoto Hirata, Yoichiro Kamatani, Akiko Tamakoshi, Zentaro Yamagata, Kaori Muto, Koichi Matsuda, Michiaki Kubo, Yusuke Nakamura, Biobank Japan Cooperative Hospital Group, Yutaka Kiyohara, and Toshiharu Ninomiya. Risk prediction models for mortality in patients with cardiovascular disease: the BioBank Japan Project. Journal of Epidemiology, in press.
11. Hiroshi Yokomichi, Hokuto Noda, Akiko Nagai, Makoto Hirata, Akiko Tamakoshi, Yoichiro Kamatani, Yutaka Kiyohara, Koichi Matsuda, Kaori Muto, Toshiharu Ninomiya, Michiaki Kubo, BioBank Japan Collaboration Group, and Zentaro Yamagata. Cholesterol Levels of Japanese Dyslipidaemic Patients with Various Comorbidities: BioBank Japan. Journal of Epidemiology, in press.
12. Hiroshi Yokomichi, Akiko Nagai, Makoto Hirata, Akiko Tamakoshi, Yutaka Kiyohara, Yoichiro Kamatani, Kaori Muto, Toshiharu Ninomiya, Koichi Matsuda, Michiaki Kubo, Yusuke Nakamura, BioBank Japan Cooperative Hospital Group, and Zentaro Yamagata. Statin Use and All-Cause and Cancer Mortality: BioBank Japan Cohort. Journal of Epidemiology, in press.
13. Hiroshi Yokomichi, Akiko Nagai, Makoto Hirata, Yutaka Kiyohara, Kaori Muto, Toshiharu Ninomiya, Koichi Matsudai, Yoichiro Kamatani, Akiko Tamakoshi, Michiaki Kubo, Yusuke Nakamura, BioBank Japan Collaboration Group, and Zentaro Yamagata. Serum Glucose, Cholesterol and Blood Pressure Levels in Japanese Patients with Type 1 and 2 Diabetes: BioBank Japan Cohort. Journal of Epidemiology, in press.
14. Hiroshi Yokomichi, Akiko Nagai, Makoto Hirata, Yutaka Kiyohara, Kaori Muto, Toshiharu Ninomiya, Koichi Matsuda, Akiko Tamakoshi, Yoichiro Kamatani, Michiaki Kubo, Yusuke Nakamura, BioBank Japan Cooperative Hospital Group, and Zentaro Yamagata. Survival of Macrovascular Disease, Chronic Kidney Disease, Chronic Respiratory Disease, Cancer and Smoking in Patients with Type 2 Diabetes: BioBank Japan Cohort. Journal of Epidemiology, in press.
15. Inoue Y, Shineha R, Yashiro Y. Current public support for human-animal chimera research in Japan is limited, despite high levels of scientific approval. Cell Stem Cell 19(2): 152-3, 2016.
16. 中田はる佳，吉田幸恵，有田悦子，武藤香織．患者の経験からみる臨床試験への参加判断とインフォームドコンセントの意義．臨床薬理．（印刷中）
17. 神里彩子．医療の質を確保・向上させるために医師組織はどうあるべきか―医師会と弁護士会の歴史から学ぶこと―． 医療の質・安全学会誌．11(3): 306-314, 2016.
18. 佐藤桃子．出生前診断における国家の役割―羊水検査，母体血清マーカー，NIPTに対する国家の関与の分析．生物学史研究8，2016.
19. 李怡然，武藤香織．配偶子提供で生まれる子への真実告知とインフォームド・アセント―不妊カウンセラーへのインタビューより―．保健医療社会学論集．26(2): 74-84, 2016.