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, Ph.D.	教　授	理学博士	宮　野	悟
Associate Professor	Rui Yamaguchi, Ph.D.	准教授	博士(理学)	山　口	類
Assistant Professor	Yuichi Shiraishi, Ph.D.	助　教	博士(統計科学)	白　石　友	一
Associate Professor	Tetsuo Shibuya, Ph.D.	准教授	博士(理学)	渋　谷　哲	朗
Assistant Professor	Kotoe Katayama, Ph.D.	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. Adaptive NetworkProfiler for identifying cancer characteristic-specific gene regulatory networks

Park H[1], Shimamura T[2], Imoto S[3], Miyano S: [2]Nagoya University School of Medicine, [3]Health Intelligence Center

There is currently much discussion about sample (patient)-specific gene regulatory network identification, since the efficiently constructed sample-specific gene networks lead to effective personalized cancer therapy. Although statistical approaches have been proposed for inferring gene regulatory networks, the methods cannot reveal sample-specific characteristics because the existing methods, such as an L_1-type regularization, provide averaged results for all samples. Thus, we cannot reveal sample-specific characteristics in transcriptional regulatory networks. To settle on this issue, the NetworkProfiler was proposed based on the kernel-based L_1-type regularization. The NetworkProfiler imposes a weight on each sample based on the Gaussian kernal function for controlling effect of samples on modeling a target sample, where the amount of

weight depends on similarity of cancer characteristics between samples. The method, however, cannot perform gene regulatory network identification well for a target sample in a sparse region (i.e., for a target sample, there are only a few samples having a similar characteristic of the target sample, where the characteristic is considered as a modulator in sample-specific gene network construction), since a constant bandwidth in the Gaussian kernel function cannot effectively group samples for modeling a target sample in sparse region. The cancer characteristics, such as an anti-cancer drug sensitivity, are usually nonuniformly distributed, and thus modeling for samples in a sparse region is also a crucial issue. We propose a novel kernel-based L_1-type regularization method based on a modified *k*-nearest neighbor (KNN)-Gaussian kernel function, called an adaptive NetworkProfiler. By using the modified KNN-Gaussian kernel function, our method provides robust results against the distribution of modulators, and properly groups samples according to a cancer characteristic for sample-specific analysis. Furthermore, we propose a sample-specific generalized cross-validation for choosing the sample-specific tuning parameters in the kernel-based L_1-type regularization method. Numerical studies demonstrate that the proposed adaptive NetworkProfiler effectively performs sample-specific gene network construction. We apply the proposed statistical strategy to the publicly available Sanger Genomic data analysis, and extract anti-cancer drug sensitivity-specific gene regulatory networks.

b. Sequence-specific bias correction for RNA-seq data using recurrent neural networks

Zhang YZ, Yamaguchi R, Imoto S[3], Miyano S

The recent success of deep learning techniques in machine learning and artificial intelligence has stimulated a great deal of interest among bioinformaticians, who now wish to bring the power of deep learning to bare on a host of bioinformatical problems. Deep learning is ideally suited for biological problems that require automatic or hierarchical feature representation for biological data when prior knowledge is limited. In this work, we address the sequence-specific bias correction problem for RNA-seq data redusing Recurrent Neural Networks (RNNs) to model nucleotide sequences without pre-determining sequence structures. The sequence-specific bias of a read is then calculated based on the sequence probabilities estimated by RNNs, and used in the estimation of gene abundance. We explored the application of two popular RNN recurrent units for this task and demonstrate that RNN-based approaches provide a flexible way to model nucleotide sequences without knowledge of predetermined sequence structures. Our experiments show that training a RNN-based nucleotide sequence model is efficient and RNN-based bias correction methods compare well with the-state-of-the-art sequence-specific bias correction method on the commonly used MAQC-III data set. As a conclusion, RNNs provides an alternative and flexible way to calculate sequence-specific bias without explicitly pre-determining sequence structures.

c. GIMLET: Identifying biological modulators in context-specific gene regulation using local energy statistics

Shimamura T[2], Matsui Y[2], Kajino T[2], Ito S, Takahashi T[2], Miyano S

Regulation of transcription factor activity is dynamically changed across cellular conditions and disease subtypes. The identification of biological modulators contributing to context-specific gene regulation is one of the challenging tasks in systems biology, in order to understand and control cellular responses across different genetic backgrounds and environmental conditions. Previous approaches for the identification of biological modulators from gene expression data are restricted to the capturing of a particular type of a three-way dependency between a regulator, its target gene, and a modulator, and these methods cannot describe complex regulation structure, such as where multiple regulators, their target genes, and modulators are functionally related. Here, we propose a statistical method for the identification of biological modulators by capturing multivariate local dependencies, based on energy statistics, which is a class of statistics based on distances. Subsequently, out method assigns a measure of statistical significance to each candidate modulator by permutation test. We compared our approach with a leading competitor for the identification of modulators, and illustrated its performance both through the simulation and real data analysis. GIMLET is implemented with R (≥3.2.2) and is available from github (https://github.com/tshimam/GIMLET).

d. Tumor subclonal progression model for cancer hallmark acquisition

Matsui Y[2], Miyano S, Shimamura T[2]

Recent advances in the methodologies of reconstructing cancer evolutionary trajectories opened the horizon for deciphering the subclonal populations and their evolutionary architectures under the cancer ecosystems. An important challenge of the cancer evolution studies is connecting genetic aberrations in subclones to clinically interpretable and actionable target of subclones for individual pa-

tients. In this paper, we present a novel method for constructing tumor subclonal progression model for cancer hallmark acquisition using multi-regional sequencing data. We prepare a subclonal evolutionary tree inferred from variant allele frequencies and estimate the pathway alternation probabilities from large scale cohort genomic data. We then construct an evolutionary tree of pathway alternation that takes account of selectivity of pathway alternations by the notion of probabilistic causality. We show the effectiveness of our method using a dataset of clear cell renal cell carcinomas.

e. Identification of a p53-repressed gene module in breast cancer cells

Miyamoto T[4], Tanikawa C[4], Yodsurang V[5], Zhang YZ, Imoto S[3], Yamaguchi R, Miyano S, Nakagawa H[6], Matsuda K[5]: [4]Laboratory of Genome Technology, Human Genome Center, [5]Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Toky, [6]RIKEN Center for Integrative Medical Sciences

The p53 protein is a sophisticated transcription factor that regulates dozens of target genes simultaneously in accordance with the cellular circumstances. Although considerable efforts have been made to elucidate the functions of p53-induced genes, a holistic understanding of the orchestrated signaling network repressed by p53 remains elusive. Here, we performed a systematic analysis to identify simultaneously regulated p53-repressed genes in breast cancer cells. Consequently, 28 genes were designated as the p53-repressed gene module, whose gene components were simultaneously suppressed in breast cancer cells treated with Adriamycin. A ChIP-seq database showed that p53 does not preferably bind to the region around the transcription start site of the p53-repressed gene module elements compared with that of p53-induced genes. Furthermore, we demonstrated that p21/CDKN1A plays a pivotal role in the suppression of the p53-repressed gene module in breast cancer cells. Finally, we showed that appropriate suppression of some genes belonging to the p53-repressed gene module contributed to a better prognosis of breast cancer patients. Taken together, these findings disentangle the gene regulatory network underlying the built-in p53-mediated tumor suppression system.

f. The transcriptional landscape of p53 signalling pathway

Tanikawa C[4], Zhang YZ, Yamamoto R[4], Tsuda Y[4], Tanaka M[4], Funauchi Y[4], Mori J[8], Imoto S[3], Yamaguchi R, Nakamura Y[7], Miyano S, Nakagawa H[6], Matsuda K[5]: [7]University of Chicago, [8]Laboratory of Molecular Medicine, Human Genome Center

Although recent cancer genomics studies have identified a large number of genes that were mutated in human cancers, *p53* remains as the most frequently mutated gene. To further elucidate the p53-signalling network, we performed transcriptome analysis on 24 tissues in $p53^{+/+}$ or $p53^{-/-}$ mice after whole-body X-ray irradiation. Here we found transactivation of a total of 3551 genes in one or more of the 24 tissues only in $p53^{+/+}$ mice, while 2576 genes were downregulated. p53 mRNA expression level in each tissue was significantly associated with the number of genes upregulated by irradiation. Annotation using TCGA (The Cancer Genome Atlas) database revealed that p53 negatively regulated mRNA expression of several cancer therapeutic targets or pathways such as *BTK, SYK,* and *CTLA4* in breast cancer tissues. In addition, stomach exhibited the induction of Krt6, Krt16, and Krt17 as well as loricrin, an epidermal differentiation marker, after the X-ray irradiation only in $p53^{+/+}$ mice, implying a mechanism to protect damaged tissues by rapid induction of differentiation. Our comprehensive transcriptome analysis elucidated tissue specific roles of p53 and its signalling networks in DNA-damage response that will enhance our understanding of cancer biology.

g. Genome-wide screening of DNA methylation associated with lymph node metastasis in esophageal squamous cell carcinoma

Nagata H[9], Kozaki KI[9], Muramatsu T[9], Hiramoto H[9], Tanimoto K[10], Fujiwara N[9], Imoto S[3], Ichikawa D[10], Otsuji E[10], Miyano S, Kawano T[9], Inazawa J[9]: [9]Tokyo Medical and Dental University, [10]Kyoto Prefectural University of Medicine

Lymph node metastasis (LNM) of esophageal squamous cell carcinoma (ESCC) is well-known to be an early event associated with poor prognosis in patients with ESCC. Recently, tumor-specific aberrant DNA methylation of CpG islands around the promoter regions of tumor-related genes has been investigated as a possible biomarker for use in early diagnosis and prediction of prognosis. However, there are few DNA methylation markers able to predict the presence of LNM in ESCC. To identify DNA methylation markers associated with LNM of ESCC, we performed a genome-wide screening of DNA methylation status in a discovery cohort of 67 primary ESCC tissues and their paired normal esophageal tissues using the Illumina Infinium HumanMethylation450 BeadChip. In this screening, we focused on differentially methylated

regions (DMRs) that were associated with LNM of ESCC, as prime candidates for DNA methylation markers. We extracted three genes, *HOXB2*, *SLC15A3*, and *SEPT9*, as candidates predicting LNM of ESCC, using pyrosequencing and several statistical analyses in the discovery cohort. We confirmed that *HOXB2* and *SEPT9* were highly methylated in LNM-positive tumors in 59 ESCC validation samples. These results suggested that *HOXB2* and *SEPT9* may be useful epigenetic biomarkers for the prediction of the presence of LNM in ESCC.

h. Circulating exosomal microRNA-203 is associated with metastasis possibly via inducing tumor-associated macrophages in colorectal cancer

Takano Y[19], Masuda T, Iinuma H[20], Yamaguchi R, Sato K[19], Tobo T[19], Hirata H[19], Kuroda Y[19], Nambara S[19], Hayashi N[19], Iguchi T[19], Ito S[19], Eguchi H[19], Ochiya T[21], Yanaga K[22], Miyano S, Mimori K[19]: [19]Kyushu University Beppu Hospital, [20]Teikyo University, [21]National Cancer Center Research Institute, [22]Jikei University School of Medicine

A primary tumor can create a premetastatic niche in distant organs to facilitate the development of metastasis. The mechanism by which tumor cells communicate with host cells to develop premetastatic niches is unclear. We focused on the role of microRNA (miR) signaling in promoting metastasis. Here, we identified *miR-203* as a signaling molecule between tumors and monocytes in metastatic colorectal cancer (CRC) patients. Notably, high expression of serum exosomal *miR-203*, a major form in circulation, was associated with distant metastasis and an independent poor prognostic factor, whereas low expression in tumor tissues was a poor prognostic factor in CRC patients. We also found that exosomes carrying miR-203 from CRC cells were incorporated into monocytes and *miR-203* could promote the expression of M2 markers *in vitro*, suggesting *miR-203* promoted the differentiation of monocytes to M2-tumor-associated macrophages (TAMs). In a xenograft mouse model, *miR-203*-transfected CRC cells developed more liver metastasis compared to control cells. In conclusion, serum exosomal miR-203 expression is a novel biomarker for predicting metastasis, possibly via promoting the differentiation of monocytes to M2-TAMs in CRC. Furthermore, we propose the concept of site-dependent functions for *miR-203* in tumor progression.

2. Cancer Genomics

a. Clonal evolution in myelodysplastic syndromes

da Silva-Coelho P[11], Kroeze LI[11], Yoshida K[12], Koorenhof-Scheele TN[11], Knops R[11], van de Locht LT[11], de Graaf AO[11], Massop M[11], Sandmann S[13], Dugas M[13], Stevens-Kroef MJ[11], Cermak J[14], Shiraishi Y, Chiba K, Tanaka H, Miyano S, de Witte T[11], Blijlevens NMA[11], Muus P[11], Huls G[11], van der Reijden BA[11], Ogawa S[12], Jansen JH[11]: [11]Radboud University Medical Center, [12]Kyoto University School of Medicine, [13]University of Münster, [14]Institute of Hematology and Blood Transfusion

Cancer development is a dynamic process during which the successive accumulation of mutations results in cells with increasingly malignant characteristics. Here, we show the clonal evolution pattern in myelodysplastic syndrome (MDS) patients receiving supportive care, with or without lenalidomide (follow-up 2.5-11 years). Whole-exome and targeted deep sequencing at multiple time points during the disease course reveals that both linear and branched evolutionary patterns occur with and without disease-modifying treatment. The application of disease-modifying therapy may create an evolutionary bottleneck after which more complex MDS, but also unrelated clones of haematopoietic cells, may emerge. In addition, subclones that acquired an additional mutation associated with treatment resistance (TP53) or disease progression (NRAS, KRAS) may be detected months before clinical changes become apparent. Monitoring the genetic landscape during the disease may help to guide treatment decisions. Our DNA and RNA-sequence analysis pipe line Genomon (https://github.com/Genomon-Project) on HGC supercomputer SHIROKANE played an important role in this study. We contributed to sequence data analysis and statistical methodology development using HGC supercomputer SHIROKANE.

b. 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, 7-12, 14-20, 22, 24, 29, 32-33, 38, 40-41, 43-47, 51, 55-58.

3. Oncoimmulogy

a. Identification of an immunogenic neo-epitope encoded by mouse sarcoma using CXCR3 ligand mRNAs as sensors

Fujii K[15], Miyahara Y[15], Harada N[15], Muraoka D[15], Komura M, Yamaguchi R, Yagita H[16], Nakamura J[15], Sugino S[15], Okumura S[15], Imoto S[3], Miyano S, Shiku H[15]: [15]Mie University Graduate School of Medicine, [16]Juntendo University School of Medicine

The CXCR3 ligands CXCL9, 10, and 11 play critical roles in the amplification of immune responses by recruiting $CXCR3^+$ immune effector cells to the tumor site. Taking advantage of this property of CXCR3 ligands, we aimed to establish a novel approach to identify immunogenic mutated-antigens. We examined the feasibility of using CXCR3 ligand mRNAs as sensors for detection of specific immune responses in human and murine systems. We further investigated whether this approach is applicable for the identification of immunogenic mutated-antigens by using murine sarcoma lines. Rapid synthesis of CXCR3 ligand mRNAs occurred shortly after specific immune responses in both human and murine immune systems. Particularly, in CMS5 tumor-bearing mice, we detected specific immune responses to mutated mitogen-activated protein kinase 2 (ERK2), which has previously been identified as an immunogenic mutated-antigen. Furthermore, by combining this approach with whole-exome and transcriptome sequencing analyses, we identified an immunogenic neo-epitope derived from mutated staphylococcal nuclease domain-containing protein 1 (Snd1) in CMS7 tumor-bearing mice. Most importantly, we successfully detected the specific immune response to this neo-epitope even without co-administration of anti-cytotoxic T-lymphocyte protein-4 (CTLA-4), anti-programmed cell death-1 (PD-1) and anti-glucocorticoid-induced TNFR-related protein (GITR) antibodies, which vigorously augmented the immune response and consequently enabled us to detect the specific immune response to this neo-epitope by conventional IFNγ intracellular staining method. Our data indicate the potential usefulness of this strategy for the identification of immunogenic mutated-antigens. We propose that this approach would be of great help for the development of personalized cancer vaccine therapies in future. We contributed to bioinformatics tool development and analysis in this study.

b. Clinical significance of T cell clonality and expression levels of immune-related genes in endometrial cancer

Ikeda Y[7], Kiyotani K[7], Yew PY[7], Sato S[17], Imai Y[17], Yamaguchi R, Miyano S, Fujiwara K[17], Hasegawa K[17], Nakamura Y[7]: [17]Saitama Medical University International Medical Center

Immune microenvironment characterized by T cell clonality as well as expression signatures of immune-related genes in endometrial cancer tissues may play significant roles in clinical outcome of patients. We aimed to investigate the clinical significance of immune-related gene expression and TCR repertoire in endometrial cancer. Using total RNAs extracted from 32 endometrioid endometrial cancer cases, we performed quantitative real-time PCR to measure mRNA expression levels of immune-related genes including *TRB*, *CD8*, *GZMA*, *HLA-A*, *CD11c* and *PD-L1*. Higher mRNA expression levels of CD8 ($P = 0.039$) and *CD11c* ($P = 0.046$) in the 32 tissue samples were significantly associated with longer progression-free survival (PFS). Expression levels of *CD8* ($P < 0.001$) and *CD11c* ($P = 0.048$) were also significantly associated with longer PFS in 540 cases in TCGA database. We also performed T cell receptor β (TCRβ) sequencing of tumor-infiltrating lymphocytes (TILs) on an Illumina MiSeq platform. To evaluate clonal expansion of TCRβ clonotypes, we adjusted the number of abundant TCRβ clonotypes by *TRB* mRNA expression levels and examined TCR clonality with the expression levels of immune-related genes and clinicopathological factors. The cases with high clonal T cell expansion along with high *PD-L1* expression in cancer tissues was related to higher mRNA expression levels of *CD8* ($P < 0.001$), *GZMA* ($P < 0.001$) and *HLA-A* ($P = 0.027$), showed a significantly longer PFS ($P = 0.015$), indicating a possibility that these parameters may serve as faborable prognostic factors. Considering clinical stage, mRNA expression of *CD8* ($P = 0.037$), *GZMA* ($P = 0.027$) and HLA-A ($P = 0.022$) was significantly higher in tumors at an early stage. Thus, we identified clinical and prognostic significance of immune microenvironment including the T cell clonality of TILs as well as *PD-L1* and *CD11c* mRNA expression levels in endometrial cancer tissues. We contributed to sequence data analysis and statistical methodology using HGC supercomputer SHIROKANE.

c. Characterization of the B-cell receptor repertoires in peanut allergic subjects undergoing oral immunotherapy

Kiyotani K[7], Mai TH[7], Yamaguchi R, Yew PY, Kulis M[18], Orgel K[18], Imoto S[3], Miyano S, Burks AW[18], Nakamura Y[7]: [18]University of North Carolina

B-cell receptors (BCRs) play a critical role in adaptive immunity as they generate highly diverse immunoglobulin repertoires to recognize a wide variety of antigens. To better understand immune responses, it is critically important to establish a

quantitative and rapid method to analyze BCR repertoire comprehensively. Here, we developed "Bcrip", a novel approach to characterize BCR repertoire by sequencing millions of BCR cDNA using next-generation sequencer. Using this method and quantitative real-time PCR, we analyzed expression levels and repertoires of BCRs in a total of 17 peanut allergic subjects' peripheral blood samples before and after receiving oral immunotherapy (OIT) or placebo. By our methods, we successfully identified all of variable (V), joining (J), and constant (C) regions, in an average of 79.1% of total reads and 99.6% of these VJC-mapped reads contained the C region corresponding to the isotypes that we aimed to analyze. In the 17 peanut allergic subjects' peripheral blood samples, we observed an oligoclonal enrichment of certain immunoglobulin heavy chain alpha (IGHA) and IGH gamma (IGHG) clones ($P = 0.034$ and $P = 0.027$, respectively) in peanut allergic subjects after OIT. This newly developed BCR sequencing and analysis method can be applied to investigate B-cell repertoires in various research areas, including food allergies as well as autoimmune and infectious diseases. We contributed to sequence data analysis and statistical methodology using HGC supercomputer SHIROKANE.

Publications

1. Aoki K, Nakamura H, Suzuki H, Matsuo K, Kataoka K, Shimamura T, Motomura K, Ohka F, Shiina S, Yamamoto T, Nagata Y, Yoshizato T, Mizoguchi M, Abe T, Momii Y, Muragaki Y, Watanabe R, Ito I, Sanada M, Yajima H, Morita N, Takeuchi I, Miyano S, Wakabayashi T, Ogawa S, Natsume A. Prognostic relevance of genetic alterations in diffuse lower-grade gliomas. *Neuro Oncol*. 20(1): 66-77, 2018.
2. Berger G, Kroeze LI, Koorenhof-Scheele TN, de Graaf AO, Yoshida K, Ueno H, Shiraishi Y, Miyano S, van den Berg E, Schepers H, van der Reijden BA, Ogawa S, Vellenga E, Jansen JH. Early detection and evolution of pre-leukemic clones in therapy-related myeloid neoplasms following autologous SCT. *Blood*. 2018 Jan 8. pii: blood-2017-09-805879. doi: 10.1182/blood-2017-09-805879.
3. da Silva-Coelho P, Kroeze LI, Yoshida K, Koorenhof-Scheele TN, Knops R, van de Locht LT, de Graaf AO, Massop M, Sandmann S, Dugas M, Stevens-Kroef MJ, Cermak J, Shiraishi Y, Chiba K, Tanaka H, Miyano S, de Witte T, Blijlevens NMA, Muus P, Huls G, van der Reijden BA, Ogawa S, Jansen JH. Clonal evolution in myelodysplastic syndromes. *Nat Commun*. 8: 15099, 2017.
4. Ding LW, Sun QY, Tan KT, Chien W, Thippeswamy AM, Eng Juh Yeoh A, Kawamata N, Nagata Y, Xiao JF, Loh XY, Lin DC, Garg M, Jiang YY, Xu L, Lim SL, Liu LZ, Madan V, Sanada M, Fernández LT, Preethi H, Lill M, Kantarjian HM, Kornblau SM, Miyano S, Liang DC, Ogawa S, Shih LY, Yang H, Koeffler HP. Mutational landscape of pediatric acute lymphoblastic leukemia. *Cancer Res*. 77(2): 390-400, 2017. Erratum in: *Cancer Res*. 77(8): 2174, 2017.
5. Fujii K, Miyahara Y, Harada N, Muraoka D, Komura M, Yamaguchi R, Yagita H, Nakamura J, Sugino S, Okumura S, Imoto S, Miyano S, Shiku H. Identification of an immunogenic neoepitope encoded by mouse sarcoma using CXCR3 ligand mRNAs as sensors. *Oncoimmunology*. 6(5): e1306617, 2017.
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 multicentric tumors. *J Hepatol*. 66(2): 363-373, 2017.
7. Hirabayashi S, Seki M, Hasegawa D, Kato M, Hyakuna N, Shuo T, Kimura S, Yoshida K, Kataoka K, Fujii Y, Shiraishi Y, Chiba K, Tanaka H, Kiyokawa N, Miyano S, Ogawa S, Takita J, Manabe A. Constitutional abnormalities of IDH1 combined with secondary mutations predispose a patient with Maffucci syndrome to acute lymphoblastic leukemia. Pediatr *Blood Cancer*. 64(12), 2017.
8. Hiramoto N, Takeda J, Yoshida K, Ono Y, Yoshioka S, Yamauchi N, Fujimoto A, Maruoka H, Shiraishi Y, Tanaka H, Chiba K, Imai Y, Miyano S, Ogawa S, Ishikawa T. Donor cell-derived transient abnormal myelopoiesis as a specific complication of umbilical cord blood transplantation. *Bone Marrow Transplant*. 2017 Oct 9. doi: 10.1038/bmt.2017.226.
9. Hiwatari M, Seki M, Akahoshi S, Yoshida K, Miyano S, Shiraishi Y, Tanaka H, Chiba K, Ogawa S, Takita J. Molecular studies reveal MLL-MLLT10/AF10 and ARID5B-MLL gene fusions displaced in a case of infantile acute lymphoblastic leukemia with complex karyotype. *Oncol Lett*. 14(2): 2295-2299, 2017.
10. 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*. 8(4): 6483-6495, 2017.

11. 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*. 140(1): 223-231, 2017.
12. 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*. 105(4): 515-520, 2017.
13. Ikeda Y, Kiyotani K, Yew PY, Sato S, Imai Y, Yamaguchi R, Miyano S, Fujiwara K, Hasegawa K, Nakamura Y. Clinical significance of T cell clonality and expression levels of immune-related genes in endometrial cancer. *Oncol Rep*. 37 (5): 2603-2610, 2017.
14. 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*. 2102(3): e93-e96, 2017.
15. Isobe T, Seki M, Yoshida K, Sekiguchi M, Shiozawa Y, Shiraishi Y, Kimura S, Yoshida M, Inoue Y, Yokoyama A, Kakiuchi N, Suzuki H, Kataoka K, Sato Y, Kawai T, Chiba K, Tanaka H, Shimamura T, Kato M, Iguchi A, Hama A, Taguchi T, Akiyama M, Fujimura J, Inoue A, Ito T, Deguchi T, Kiyotani C, Iehara T, Hosoi H, Oka A, Sanada M, Tanaka Y, Hata K, Miyano S, Ogawa S, Takita J. Integrated molecular characterization of the lethal pediatric cancer pancreatoblastoma. *Cancer Res*. 2017 Dec 12. pii: canres.2581.2017. doi: 10.1158/0008-5472.CAN-17-2581.
16. Kamijo R, Itonaga H, Kihara R, Nagata Y, Hata T, Asou N, Ohtake S, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Naoe T, Kiyoi H, Miyazaki Y. Distinct gene alterations with a high percentage of myeloperoxidase-positive leukemic blasts in de novo acute myeloid leukemia. *Leuk Res*. 65: 34-41, 2018.
17. Katagiri S, Umezu T, Azuma K, Asano M, Akahane D, Makishima H, Yoshida K, Watatani Y, Chiba K, Miyano S, Ogawa S, Ohyashiki JH, Ohyashiki K. Hidden FLT3-D835Y clone in FLT3-ITD-positive acute myeloid leukemia that evolved into very late relapse with T-lymphoblastic leukemia. *Leuk Lymphoma*. 2017 Oct 3: 1-4. doi: 10.1080/10428194.2017.1382696.
18. Kataoka K, Iwanaga M, Yasunaga JI, Nagata Y, Kitanaka A, Kameda T, Yoshimitsu M, Shiraishi Y, Sato-Otsubo A, Sanada M, Chiba K, Tanaka H, Ochi Y, Aoki K, Suzuki H, Shiozawa Y, Yoshizato T, Sato Y, Yoshida K, Nosaka K, Hishizawa M, Itonaga H, Imaizumi Y, Munakata W, Shide K, Kubuki Y, Hidaka T, Nakamaki T, Ishiyama K, Miyawaki S, Ishii R, Nureki O, Tobinai K, Miyazaki Y, Takaori-Kondo A, Shibata T, Miyano S, Ishitsuka K, Utsunomiya A, Shimoda K, Matsuoka M, Watanabe T, Ogawa S. Prognostic relevance of integrated genetic profiling in adult T-cell leukemia/lymphoma. *Blood*. 131(2): 215-225, 2018.
19. 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*. 31(3): 580-584, 2017.
20. Kato I, Nishinaka Y, Nakamura M, Akarca AU, Niwa A, Ozawa H, Yoshida K, Mori M, Wang D, Morita M, Ueno H, Shiozawa Y, Shiraishi Y, Miyano S, Gupta R, Umeda K, Watanabe K, Koh K, Adachi S, Heike T, Saito MK, Sanada M, Ogawa S, Marafioti T, Watanabe A, Nakahata T, Enver T. Hypoxic adaptation of leukemic cells infiltrating the CNS affords a therapeutic strategy targeting VEGFA. *Blood*. 129(23): 3126-3129, 2017.
21. Kiyotani K, Mai TH, Yamaguchi R, Yew PY, Kulis M, Orgel K, Imoto S, Miyano S, Burks AW, Nakamura Y. Characterization of the B-cell receptor repertoires in peanut allergic subjects undergoing oral immunotherapy. *J Hum Genet*. 2017 Nov 30. doi: 10.1038/s10038-017-0364-0.
22. Kobayashi M, Yokoyama K, Shimizu E, Yusa N, Ito M, Yamaguchi R, Imoto S, Miyano S, Tojo A. Phenotype-based gene analysis allowed successful diagnosis of X-linked neutropenia associated with a novel WASp mutation. *Ann Hematol*. 97(2): 367-369, 2018.
23. Lyons E, Sheridan P, Tremmel G, Miyano S, Sugano S. Large-scale DNA barcode library generation for biomolecule identification in high-throughput screens. *Sci Rep*. 7(1): 13899, 2017.
24. 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*. 49(2): 204-212, 2017.

25. Matsui Y, Miyano S, Shimamura T. Tumor subclonal progression model for cancer hallmark acquisition. *LNBI*, 2018. In press.
26. Matsui Y, Niida A, Uchi R, Mimori K, Miyano S, Shimamura T. phyC: Clustering cancer evolutionary trees. *PLoS Comput Biol*. 13(5): e1005509, 2017.
27. Miyamoto T, Tanikawa C, Yodsurang V, Zhang YZ, Imoto S, Yamaguchi R, Miyano S, Nakagawa H, Matsuda K. Identification of a p53-repressed gene module in breast cancer cells. *Oncotarget*. 8(34): 55821-55836, 2017.
28. Moriyama T, Shiraishi Y, Chiba K, Yamaguchi R, Imoto S, Miyano S. OVarCall: Bayesian mutation calling method utilizing overlapping paired-end reads. *IEEE Trans Nanobioscience*. 16 (2): 116-122, 2017.
29. 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*. 19(7): 796-802, 2017.
30. Nagata H, Kozaki KI, Muramatsu T, Hiramoto H, Tanimoto K, Fujiwara N, Imoto S, Ichikawa D, Otsuji E, Miyano S, Kawano T, Inazawa J. Genome-wide screening of DNA methylation associated with lymph node metastasis in esophageal squamous cell carcinoma. *Oncotarget*. 8(23): 37740-37750, 2017.
31. Niida A, Nagayama S, Miyano S, Mimori K. Understanding intratumor heterogeneity by combining genome analysis and mathematical modeling. *Cancer Sci*. 2018 Jan 20. doi: 10.1111/cas.13510.
32. Nguyen TB, Sakata-Yanagimoto M, Asabe Y, Matsubara D, Kano J, Yoshida K, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Izutsu K, Nakamura N, Takeuchi K, Miyoshi H, Ohshima K, Minowa T, Ogawa S, Noguchi M, Chiba S. Identification of cell-type-specific mutations in nodal T-cell lymphomas. *Blood Cancer J*. 7(1): e516, 2017.
33. Ogawa M, Yokoyama K, Hirano M, Jimbo K, Ochi K, Kawamata T, Ohno N, Shimizu E, Yokoyama N, Yamaguchi R, Imoto S, Uchimaru K, Takahashi N, Miyano S, Imai Y, Tojo A. Different clonal dynamics of chronic myeloid leukaemia between bone marrow and the central nervous system. *Br J Haematol*. 2017 Dec 19. doi: 10.1111/bjh.15065.
34. Onuki R, Yamaguchi R, Shibuya T, Kanehisa M, Goto S. Revealing phenotype-associated functional differences by genome-wide scan of ancient haplotype blocks. *PLoS One*. 12(4): e0176530, 2017.
35. 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*. 24(2): 138-152, 2017.
36. Park H, Shimamura T, Imoto S, Miyano S. Adaptive NetworkProfiler for identifying cancer characteristic-specific gene regulatory networks. *J Comput Biol*. 2017 Oct 20. doi: 10.1089/cmb.2017.0120.
37. 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*. 14(4): 771-782, 2017.
38. Sakai H, Hosono N, Nakazawa H, Przychodzen B, Polprasert C, Carraway HE, Sekeres MA, Radivoyevitch T, Yoshida K, Sanada M, Yoshizato T, Kataoka K, Nakagawa MM, Ueno H, Nannya Y, Kon A, Shiozawa Y, Takeda J, Shiraishi Y, Chiba K, Miyano S, Singh J, Padgett RA, Ogawa S, Maciejewski JP, Makishima H. A novel genetic and morphologic phenotype of ARID2-mediated myelodysplasia. *Leukemia*. 2017 Nov 3. doi: 10.1038/leu.2017.319.
39. Sato R, Shibata T, Tanaka Y, Kato C, Yamaguchi K, Furukawa Y, Shimizu E, Yamaguchi R, Imoto S, Miyano S, Miyake K. Requirement of glycosylation machinery in TLR responses revealed by CRISPR/Cas9 screening. *Int Immunol*. 29(8): 347-355, 2017.
40. Seki M, Kimura S, Isobe T, Yoshida K, Ueno H, Nakajima-Takagi Y, Wang C, Lin L, Kon A, Suzuki H, Shiozawa Y, Kataoka K, Fujii Y, Shiraishi Y, Chiba K, Tanaka H, Shimamura T, Masuda K, Kawamoto H, Ohki K, Kato M, Arakawa Y, Koh K, Hanada R, Moritake H, Akiyama M, Kobayashi R, Deguchi T, Hashii Y, Imamura T, Sato A, Kiyokawa N, Oka A, Hayashi Y, Takagi M, Manabe A, Ohara A, Horibe K, Sanada M, Iwama A, Mano H, Miyano S, Ogawa S, Takita J. Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. *Nat Genet*. 49(8): 1274-1281, 2017.
41. Sekinaka Y, Mitsuiki N, Imai K, Yabe M, Yabe

H, Mitsui-Sekinaka K, Honma K, Takagi M, Arai A, Yoshida K, Okuno Y, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Muramatsu H, Kojima S, Hira A, Takata M, Ohara O, Ogawa S, Morio T, Nonoyama S. Common variable immunodeficiency caused by FANC mutations. *J Clin Immunol*. 2 37(5): 434-444, 2017.

42. Shimamura T, Matsui Y, Kajino T, Ito S, Takahashi T and Miyano S. GIMLET: Identifying biological modulators in context-specific gene regulation using local energy statistics. *LNBI*, 2018. In press.
43. Shiozawa Y, Malcovati L, Gallì A, Pellagatti A, Karimi M, Sato-Otsubo A, Sato Y, Suzuki H, Yoshizato T, Yoshida K, Shiraishi Y, Chiba K, Makishima H, Boultwood J, Hellström-Lindberg E, Miyano S, Cazzola M, Ogawa S. Gene expression and risk of leukemic transformation in myelodysplasia. *Blood*. 130(24): 2642-2653, 2017.
44. 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.
45. Takagi M, Hoshino A, Yoshida K, Ueno H, Imai K, Piao J, Kanegane H, Yamashita M, Okano T, Muramatsu H, Okuno Y, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Hayashi Y, Kojima S, Morio T. Genetic heterogeneity of uncharacterized childhood autoimmune diseases with lymphoproliferation. *Pediatr Blood Cancer*. 65(2): e26831, 2018.
46. 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*. 139(6): 1914-1922, 2017.
47. Takagi M, Yoshida M, Nemoto Y, Tamaichi H, Tsuchida R, Seki M, Uryu K, Nishii R, Miyamoto S, Saito M, Hanada R, Kaneko H, Miyano S, Kataoka K, Yoshida K, Ohira M, Hayashi Y, Nakagawara A, Ogawa S, Mizutani S, Takita J. Loss of DNA damage response in neuroblastoma and utility of a PARP inhibitor. *J Natl Cancer Inst*. 2017 Nov 1; 109(11). doi: 10.1093/jnci/djx062.
48. Takahashi Y, Sugimachi K, Yamamoto K, Niida A, Shimamura T, Sato T, Watanabe M, Tanaka J, Kudo S, Sugihara K, Hase K, Kusunoki M, Yamada K, Shimada Y, Moriya Y, Suzuki Y, Miyano S, Mori M, Mimori K. Japanese genome-wide association study identifies a significant colorectal cancer susceptibility locus at chromosome 10p14. *Cancer Sci*. 108(11): 2239-2247, 2017.
49. Takano Y, Masuda T, Iinuma H, Yamaguchi R, Sato K, Tobo T, Hirata H, Kuroda Y, Nambara S, Hayashi N, Iguchi T, Ito S, Eguchi H, Ochiya T, Yanaga K, Miyano S, Mimori K. Circulating exosomal microRNA-203 is associated with metastasis possibly via inducing tumor-associated macrophages in colorectal cancer. *Oncotarget*. 8 (45): 78598-78613, 2017.
50. Tanikawa C, Zhang YZ, Yamamoto R, Tsuda Y, Tanaka M, Funauchi Y, Mori J, Imoto S, Yamaguchi R, Nakamura Y, Miyano S, Nakagawa H, Matsuda K. The transcriptional landscape of p53 signalling pathway. *EBioMedicine*. 20: 109-119, 2017.
51. Togasaki E, Takeda J, Yoshida K, Shiozawa Y, Takeuchi M, Oshima M, Saraya A, Iwama A, Yokote K, Sakaida E, Hirase C, Takeshita A, Imai K, Okumura H, Morishita Y, Usui N, Takahashi N, Fujisawa S, Shiraishi Y, Chiba K, Tanaka H, Kiyoi H, Ohnishi K, Ohtake S, Asou N, Kobayashi Y, Miyazaki Y, Miyano S, Ogawa S, Matsumura I, Nakaseko C, Naoe T. Frequent somatic mutations in epigenetic regulators in newly diagnosed chronic myeloid leukemia. *Blood Cancer J*. 7(4): e559, 2017.
52. 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*. 36(9): 1276-1286, 2017.
53. Tsuda Y, Tanikawa C, Miyamoto T, Hirata M, Yodsurang V, Zhang YZ, Imoto S, Yamaguchi R, Miyano S, Takayanagi H, Kawano H, Nakagawa H, Tanaka S, Matsuda K. Identification of a p53 target, CD137L, that mediates growth suppression and immune response of osteosarcoma cells. *Sci Rep*. 7(1): 10739, 2017.
54. 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. Correction: Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution.

PLoS Genet. 13(5): e1006798, 2017.

55. Uryu K, Nishimura R, Kataoka K, Sato Y, Nakazawa A, Suzuki H, Yoshida K, Seki M, Hiwatari M, Isobe T, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Koh K, Hanada R, Oka A, Hayashi Y, Ohira M, Kamijo T, Nagase H, Takimoto T, Tajiri T, Nakagawara A, Ogawa S, Takita J. Identification of the genetic and clinical characteristics of neuroblastomas using genome-wide analysis. *Oncotarget*. 8(64): 107513-107529, 2017.
56. 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*. 56(5): 382-393, 2017.
57. Yoshizato T, Nannya Y, Atsuta Y, Shiozawa Y, Iijima-Yamashita Y, Yoshida K, Shiraishi Y, Suzuki H, Nagata Y, Sato Y, Kakiuchi N, Matsuo K, Onizuka M, Kataoka K, Chiba K, Tanaka H, Ueno H, Nakagawa MM, Przychodzen B, Haferlach C, Kern W, Aoki K, Itonaga H, Kanda Y, Sekeres MA, Maciejewski JP, Haferlach T, Miyazaki Y, Horibe K, Sanada M, Miyano S, Makishima H, Ogawa S. Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. *Blood*. 129(17): 2347-2358, 2017.
58. Zhang YZ, Yamaguchi R, Imoto S, Miyano S. Sequence-specific bias correction for RNA-seq data using recurrent neural networks. *BMC Genomics*. 18(Suppl 1): 1044, 2017.

Human Genome Center

Laboratory of Molecular Medicine
ゲノム医科学分野

Professor	Tatsuhiro Shibata, M.D., Ph.D.	教　授　医学博士	柴　田　龍　弘
Assistant Professor	Satoshi Yamasaki, Ph.D.	助　教　博士(農学)	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. Common Molecular Subtypes among Asian Hepatocellular and Cholangiocarcinoma.

Chaisaingmongkol J[1], Budhu A[2], Dang H[2], Rabibhadana S[3], Pupacdi B[3], Kwon SM[2], Forgues M[2], Pomyen Y[4], Bhudhisawasdi V[5], Lertprasertsuke N[6], Chotirosniramit A[6], Pairojkul C[5], Auewarakul CU[7], Sricharunrat T[8], Phornphutkul K[9], Sangrajrang S[10], Cam M[11], He P[12], Hewitt SM[13], Ylaya K[13], Wu X[14], Andersen JB[15], Thorgeirsson SS[2], Waterfall JJ[16], Zhu YJ[16], Walling J[16], Stevenson HS[16], Edelman D[16], Meltzer PS[16], Loffredo CA[17], Hama N[18], Shibata T[19], Wiltrout RH[20], Harris CC[2], Mahidol C[21], Ruchirawat M[22], Wang XW[23]; TIGER-LC Consortium: [1]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, USA; Laboratory of Chemical Carcinogenesis, Chulabhorn Research Institute, Thailand; Center of Excellence on Environmental Health and Toxicology, Office of Higher Education Commission, Ministry of Education, Thailand. [2]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, USA. [3]Laboratory of Chemical Carcinogenesis, Chulabhorn Research Institute, Thailand. [4]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, USA; Laboratory of Chemical Carcinogenesis, Chulabhorn Research Institute, Thailand. [5]Faculty of Medicine, Khon Kaen University, Thailand. [6]Faculty of Medicine, Chiang Mai University, Thailand. [7]HRH Princess Chulabhorn College of Medical Science, Thailand. [8]Chulabhorn Hospital, Thailand. [9]Rajavej Hospital, Thailand. [10]National Cancer Institute, Thailand. [11]Center for Cancer Research, National Cancer Institute, USA. [12]FDA, USA. [13]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, USA. [14]Frederick National Laboratory for Cancer Research, USA. [15]Biotech Research and Innovation Centre (BRIC), Department of Health and Medical Sciences, University of Copenhagen, Denmark. [16]Genetics Branch, Center for Cancer Research, National Cancer Institute, USA. [17]Georgetown University Medical Center, USA. [18,19]Division of Cancer Genomics, National Cancer Center Research Institute, Japan; Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Japan. [20]Cancer Inflammation Program, Center for Cancer Research, National Cancer Institute, USA. [21]Laboratory of Chemical Carcinogenesis, Chulabhorn Research Institute, Thailand; HRH Princess Chulabhorn College of Medical Science, Thailand. [22]Laboratory of Chemical Carcinogenesis, Chulabhorn Research Institute, Thailand; Center of Excellence on Environmental Health and

Toxicology, Office of Higher Education Commission, Ministry of Education, Thailand. [23]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, USA.

Intrahepatic cholangiocarcinoma (ICC) and hepatocellular carcinoma (HCC) are clinically disparate primary liver cancers with etiological and biological heterogeneity. We identified common molecular subtypes linked to similar prognosis among 199 Thai ICC and HCC patients through systems integration of genomics, transcriptomics, and metabolomics. While ICC and HCC share recurrently mutated genes, including TP53, ARID1A, and ARID 2, mitotic checkpoint anomalies distinguish the C1 subtype with key drivers PLK1 and ECT2, whereas the C2 subtype is linked to obesity, T cell infiltration, and bile acid metabolism. These molecular subtypes are found in 582 Asian, but less so in 265 Caucasian patients. Thus, Asian ICC and HCC, while clinically treated as separate entities, share common molecular subtypes with similar actionable drivers to improve precision therapy.

2. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma.

Cancer Genome Atlas Research Network.

Liver cancer has the second highest worldwide cancer mortality rate and has limited therapeutic options. We analyzed 363 hepatocellular carcinoma (HCC) cases by whole-exome sequencing and DNA copy number analyses, and we analyzed 196 HCC cases by DNA methylation, RNA, miRNA, and proteomic expression also. DNA sequencing and mutation analysis identified significantly mutated genes, including LZTR1, EEF1A1, SF3B1, and SMARCA4. Significant alterations by mutation or downregulation by hypermethylation in genes likely to result in HCC metabolic reprogramming (ALB, APOB, and CPS1) were observed. Integrative molecular HCC subtyping incorporating unsupervised clustering of five data platforms identified three subtypes, one of which was associated with poorer prognosis in three HCC cohorts. Integrated analyses enabled development of a p53 target gene expression signature correlating with poor survival. Potential therapeutic targets for which inhibitors exist include WNT signaling, MDM4, MET, VEGFA, MCL1, IDH 1, TERT, and immune checkpoint proteins CTLA-4, PD-1, and PD-L1.

3. Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma.

Jusakul A[1,2,3], Cutcutache I[4], Yong CH[1,4], Lim JQ[2,5], Huang MN[4], Padmanabhan N[1], Nellore V[6], Kongpetch S[2,7,8], Ng AWT[9], Ng LM[10], Choo SP[11], Myint SS[2], Thanan R[12], Nagarajan S[2], Lim WK[1,2], Ng CCY[2], Boot A[1,4], Liu M[1,4], Ong CK[5], Rajasegaran V[2], Lie S[2,13], Lim AST[14], Lim TH[14], Tan J[2], Loh JL[2], McPherson JR[4], Khuntikeo N[7,15], Bhudhisawasdi V[15], Yongvanit P[7], Wongkham S[12], Totoki Y[16], Nakamura H[16], Arai Y[16], Yamasaki S[17], Chow PK[18], Chung AYF[19], Ooi LLPJ[19], Lim KH[20], Dima S[21], Duda DG[22], Popescu I[21], Broet P[23], Hsieh SY[24], Yu MC[25], Scarpa A[26], Lai J[27], Luo DX[28], Carvalho AL[29], Vettore AL[30], Rhee H[31], Park YN[31], Alexandrov LB[32], Gordân R[33,34], Rozen SG[35,4,36], Shibata T[37,17], Pairojkul C[38], Teh BT[35,2,10,36,39], Tan P[35,10,36,40]: [1]Program in Cancer and Stem Cell Biology, Duke-NUS Medical School, Singapore. [2]Laboratory of Cancer Epigenome, Division of Medical Science, National Cancer Centre Singapore, Singapore. [3]The Centre for Research and Development of Medical Diagnostic Laboratories and Department of Clinical Immunology and Transfusion Sciences, Faculty of Associated Medical Sciences, Khon Kaen University, Thailand. [4]Centre for Computational Biology, Duke-NUS Medical School, Singapore. [5]Lymphoma Genomic Translational Research Laboratory, National Cancer Centre Singapore, Division of Medical Oncology, Singapore. [6]Department of Biostatistics and Bioinformatics, Center for Genomic and Computational Biology, Duke University, USA. [7]Cholangiocarcinoma Screening and Care Program and Liver Fluke and Cholangiocarcinoma Research Centre, Faculty of Medicine, Khon Kaen University, Thailand. [8]Department of Pharmacology, Faculty of Medicine, Khon Kaen University, Thailand. [9]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore. [10]Cancer Science Institute of Singapore, National University of Singapore, Singapore. [11]Division of Medical Oncology, National Cancer Centre Singapore, Singapore. [12]Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Thailand. [13]Division of Radiation Oncology, National Cancer Centre Singapore, Singapore. [14]Cytogenetics Laboratory, Department of Molecular Pathology, Singapore General Hospital, Singapore. [15]Department of Surgery, Faculty of Medicine, Khon Kaen University, Thailand. [16]Division of Cancer Genomics, National Cancer Center Research Institute, Japan. [17]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Japan. [18]Division of Surgical Oncology, National Cancer Center Singapore and Office of Clinical Sciences, Duke-NUS Medical School, Singapore. [19]Department of Hepatopancreatobiliary/Transplant Surgery, Singapore General Hospital, Singapore. [20]Department of Anatomical Pathology, Singapore General Hospital, Singapore. [21]Center of Digestive Diseases and Liver Transplantation, Fundeni Clinical Institute, Roma-

nia. [22]Edwin L. Steele Laboratories for Tumor Biology, Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, USA. [23]DHU Hepatinov, Hôpital Paul Brousse, AP-HP, France. [24]Department of Gastroenterology and Hepatology, Chang Gung Memorial Hospital and Chang Gung University, Taiwan. [25]Department of General Surgery, Chang Gung Memorial Hospital and Chang Gung University, Taiwan. [26]Applied Research on Cancer Centre (ARC-Net), University and Hospital Trust of Verona, Italy. [27]Department of Hepatobiliary Surgery, the First Affiliated Hospital of Sun Yat-sen University, P. R. China. [28]National and Local Joint Engineering Laboratory of High-through Molecular Diagnostic Technology, the First People's Hospital of Chenzhou, Southern Medical University, P. R. China. [29]Barretos Cancer Hospital, Brazil. [30]Laboratory of Cancer Molecular Biology, Department of Biological Sciences, Federal University of São Paulo, Brazil. [31]Department of Pathology, Brain Korea 21 PLUS Project for Medical Science, Integrated Genomic Research Center for Metabolic Regulation, Yonsei University College of Medicine, Korea. [32]Theoretical Biology and Biophysics, Los Alamos National Laboratory, and Center for Nonlinear Studies, Los Alamos National Laboratory, USA. [33]Department of Biostatistics and Bioinformatics, Center for Genomic and Computational Biology, Duke University, USA. [34]Department of Computer Science, Duke University, USA. [35]Program in Cancer and Stem Cell Biology, Duke-NUS Medical School, Singapore. [36]SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre, Singapore. [37]Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan. [38]Department of Pathology, Faculty of Medicine, Khon Kaen University, Thailand. [39]Institute of Molecular and Cell Biology, Singapore. [40]Genome Institute of Singapore, Singapore.

Cholangiocarcinoma (CCA) is a hepatobiliary malignancy exhibiting high incidence in countries with endemic liver-fluke infection. We analyzed 489 CCAs from 10 countries, combining whole-genome (71 cases), targeted/exome, copy-number, gene expression, and DNA methylation information. Integrative clustering defined 4 CCA clusters-fluke-positive CCAs (clusters 1/2) are enriched in ERBB2 amplifications and TP53 mutations; conversely, fluke-negative CCAs (clusters 3/4) exhibit high copy-number alterations and PD-1/PD-L2 expression, or epigenetic mutations (IDH1/2, BAP1) and FGFR/PRKA-related gene rearrangements. Whole-genome analysis highlighted FGFR2 3' untranslated region deletion as a mechanism of FGFR2 upregulation. Integration of noncoding promoter mutations with protein-DNA binding profiles demonstrates pervasive modulation of H3K27me3-associated sites in CCA. Clusters 1 and 4 exhibit distinct DNA hypermethylation patterns targeting either CpG islands or shores-mutation signature and subclonality analysis suggests that these reflect different mutational pathways. Our results exemplify how genetics, epigenetics, and environmental carcinogens can interplay across different geographies to generate distinct molecular subtypes of cancer.

Publications

1. Chaisaingmongkol J, Budhu A, Dang H, Rabibhadana S, Pupacdi B, Kwon SM, Forgues M, Pomyen Y, Bhudhisawasdi V, Lertprasertsuke N, Chotirosniramit A, Pairojkul C, Auewarakul CU, Sricharunrat T, Phornphutkul K, Sangrajrang S, Cam M, He P, Hewitt SM, Ylaya K, Wu X, Andersen JB, Thorgeirsson SS, Waterfall JJ, Zhu YJ, Walling J, Stevenson HS, Edelman D, Meltzer PS, Loffredo CA, Hama N, Shibata T, Wiltrout RH, Harris CC, Mahidol C, Ruchirawat M, Wang XW; TIGER-LC Consortium. Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. Cancer Cell. 32: 57-70, 2017.
2. Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 169: 1327-1341, 2017.
3. Jusakul A, Cutcutache I, Yong CH, Lim JQ, Huang MN, Padmanabhan N, Nellore V, Kongpetch S, Ng AWT, Ng LM, Choo SP, Myint SS, Thanan R, Nagarajan S, Lim WK, Ng CCY, Boot A, Liu M, Ong CK, Rajasegaran V, Lie S, Lim AST, Lim TH, Tan J, Loh JL, McPherson JR, Khuntikeo N, Bhudhisawasdi V, Yongvanit P, Wongkham S, Totoki Y, Nakamura H, Arai Y, Yamasaki S, Chow PK, Chung AYF, Ooi LLPJ, Lim KH, Dima S, Duda DG, Popescu I, Broet P, Hsieh SY, Yu MC, Scarpa A, Lai J, Luo DX, Carvalho AL, Vettore AL, Rhee H, Park YN, Alexandrov LB, Gordân R, Rozen SG, Shibata T, Pairojkul C, Teh BT, Tan P. Whole-Genome and Epigenomic Landscapes of Etiologically Distinct Subtypes of Cholangiocarcinoma. Cancer Discov. 7: 1116-1135, 2017.

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（兼務）	理学博士	宮　野　　悟
Professor	Koichi Matsuda, M.D., Ph.D.	連携教授（新領域創成科学研究科）	博士（医学）	松　田　浩　一
Assistant Professor	Chizu Tanikawa, Ph.D.	助　教	博士（医学）	谷　川　千　津
Project Assistant Professor	Takafumi Miyamoto, Ph.D.	特任助教	博士（医学）	宮　本　崇　史
Project Assistant Professor	Makoto Hirata, M.D., Ph.D.	特任助教	博士（医学）	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

Genome-wide association study identifies 112 new loci for body mass index in the Japanese population.

Obesity is a risk factor for a wide variety of health problems. In a genome-wide association study (GWAS) of body mass index (BMI) in Japanese people (n＝173,430), we found 85 loci significantly associated with obesity ($P < 5.0 \times 10^{-8}$), of which 51 were previously unknown. We conducted trans-ancestral meta-analyses by integrating these results with the results from a GWAS of Europeans and identified 61 additional new loci. In total, this study identifies 112 novel loci, doubling the number of previously known BMI-associated loci. By annotating associated variants with cell-type-specific regulatory marks, we found enrichment of variants in CD19＋ cells. We also found significant genetic correlations between BMI and lymphocyte count ($P = 6.46 \times 10^{-5}$, rg＝0.18) and between BMI and multiple complex diseases. These findings provide genetic evidence that lymphocytes are relevant to body weight regulation and offer insights into the pathogenesis of obesity.

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 sur-

vival 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 information 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

Citrullination of RGG motifs in FET proteins by PAD regulates protein aggregation and ALS susceptibility.

Recent proteome analyses have provided a comprehensive overview of various posttranscriptional modifications (PTMs); however, PTMs involving protein citrullination remain unclear. We performed a proteomic analysis of citrullinated proteins and identified more than 100 PAD4 (peptidyl arginine deiminase 4) substrates. Approximately one-fifth of the PAD4 substrates contained an RG/RGG motif, and PAD4 competitively inhibited the methylation of the RGG motif in FET proteins (FUS, EWS, and TAF15) and hnRNPA1, which are causative genes for ALS (Amyotrophic lateral sclerosis). PAD4-mediated citrullination significantly inhibited the aggregation of FET proteins, a frequently observed feature in neurodegenerative diseases. FUS protein levels in arsenic-induced stress granules were significantly increased in Padi4$-/-$ MEF. Moreover, rs2240335 was associated with low expression of PADI4 in the brain and a high risk of ALS ($P = 0.0381$ and odds ratio of 1.072). Our findings suggest that PAD4-mediated RGG citrullination plays a key role in protein solubility and ALS pathogenesis.

Argininosuccinate synthase 1 is an intrinsic Akt repressor transactivated by p53.

The transcription factor p53 is at the core of a built-in tumor suppression system that responds to varying degrees of stress input and is deregulated in most human cancers. Befitting its role in maintaining cellular fitness and fidelity, p53 regulates an appropriate set of target genes in response to cellular stresses. However, a comprehensive understanding of this scheme has not been accomplished. We show that argininosuccinate synthase 1 (ASS1), a citrulline-aspartate ligase in de novo arginine synthesis pathway, was directly transactivated by p53 in response to genotoxic stress, resulting in the rearrangement of arginine metabolism. Furthermore, we found that x-ray irradiation promoted the systemic induction of Ass1 and concomitantly increased plasma arginine levels in p53$+/+$ mice but not in p53$-/-$ mice. Notably, Ass1$+/-$ mice exhibited hypersensitivity to whole-body irradiation

owing to increased apoptosis in the small intestinal crypts. Analyses of ASS1-deficient cells generated using the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 (CRISPR-associated 9) system revealed that ASS1 plays a pivotal role in limiting Akt phosphorylation. In addition, aberrant activation of Akt resulting from ASS1 loss disrupted Akt-mediated cell survival signaling activity under genotoxic stress. Building on these results, we demonstrated that p53 induced an intrinsic Akt repressor, ASS1, and the perturbation of ASS1 expression rendered cells susceptible to genotoxic stress. Our findings uncover a new function of p53 in the regulation of Akt signaling and reveal how p 53, ASS1, and Akt are interrelated to each other.

Identification of a p53-repressed gene module in breast cancer cells.

The p53 protein is a sophisticated transcription factor that regulates dozens of target genes simultaneously in accordance with the cellular circumstances. Although considerable efforts have been made to elucidate the functions of p53-induced genes, a holistic understanding of the orchestrated signaling network repressed by p53 remains elusive. Here, we performed a systematic analysis to identify simultaneously regulated p53-repressed genes in breast cancer cells. Consequently, 28 genes were designated as the p53-repressed gene module, whose gene components were simultaneously suppressed in breast cancer cells treated with Adriamycin. A ChIP-seq database showed that p53 does not preferably bind to the region around the transcription start site of the p53-repressed gene module elements compared with that of p53-induced genes. Furthermore, we demonstrated that p21/CDKN1A plays a pivotal role in the suppression of the p53-repressed gene module in breast cancer cells. Finally, we showed that appropriate suppression of some genes belonging to the p53-repressed gene module contributed to a better prognosis of breast cancer patients. Taken together, these findings disentangle the gene regulatory network underlying the built-in p53-mediated tumor suppression system.

The Transcriptional Landscape of p53 Signalling Pathway.

Although recent cancer genomics studies have identified a large number of genes that were mutated in human cancers, p53 remains as the most frequently mutated gene. To further elucidate the p 53-signalling network, we performed transcriptome analysis on 24 tissues in p53+/+ or p53−/− mice after whole-body X-ray irradiation. Here we found transactivation of a total of 3551 genes in one or more of the 24 tissues only in p53+/+ mice, while 2576 genes were downregulated. p53 mRNA expression level in each tissue was significantly associated with the number of genes upregulated by irradiation. Annotation using TCGA (The Cancer Genome Atlas) database revealed that p53 negatively regulated mRNA expression of several cancer therapeutic targets or pathways such as BTK, SYK, and CTLA4 in breast cancer tissues. In addition, stomach exhibited the induction of Krt6, Krt16, and Krt17 as well as loricrin, an epidermal differentiation marker, after the X-ray irradiation only in p53 +/+ mice, implying a mechanism to protect damaged tissues by rapid induction of differentiation. Our comprehensive transcriptome analysis elucidated tissue specific roles of p53 and its signalling networks in DNA-damage response that will enhance our understanding of cancer biology.

Identification of a p53 target, CD137L, that mediates growth suppression and immune response of osteosarcoma cells.

p53 encodes a transcription factor that transactivates downstream target genes involved in tumour suppression. Although osteosarcoma frequently has p53 mutations, the role of p53 in osteosarcomagenesis is not fully understood. To explore p53-target genes comprehensively in calvarial bone and find out novel druggable p53 target genes for osteosarcoma, we performed RNA sequencing using the calvarial bone and 23 other tissues from p53 +/+ and p53 −/− mice after radiation exposure. Of 23,813 genes, 69 genes were induced more than two-fold in irradiated p53 +/+ calvarial bone, and 127 genes were repressed. Pathway analysis of the p53-induced genes showed that genes associated with cytokine-cytokine receptor interactions were enriched. Three genes, CD137L, CDC42 binding protein kinase gamma and Follistatin, were identified as novel direct p53 target genes that exhibited growth-suppressive effects on osteosarcoma cell lines. Of the three genes, costimulatory molecule Cd137l was induced only in calvarial bone among the 24 tissues tested. CD137L-expressing cells exhibited growth-suppressive effects in vivo. In addition, recombinant Fc-fusion Cd137l protein activated the immune response in vitro and suppressed osteosarcoma cell growth in vivo. We clarified the role of CD137L in osteosarcomagenesis and its potential therapeutic application. Our transcriptome analysis also indicated the regulation of the immune response through p53.

Identification of a novel p53 target, COL17A1, that inhibits breast cancer cell migration and invasion.

p53 mutation is a marker of poor prognosis in breast cancers. To identify downstream targets of p

53, we screened two transcriptome datasets, including cDNA microarrays of MCF10A breast epithelial cells with wild-type p53 or p53-null background, and RNA sequence analysis of breast invasive carcinoma. Here, we unveil ten novel p53 target candidates that are up-regulated after the induction of p 53 in wild-type cells. Their expressions are also high in breast invasive carcinoma tissues with wild-type p53. The GO analysis identified epidermis development and ectoderm development, which COL 17A1 participates, as significantly up-regulated by wild-type p53. The COL17A1 expressions increased in a p53-dependent manner in human breast cells and mouse mammary tissues. Reporter assay and ChIP assay identified intronic p53-binding sequences in the COL17A1 gene. The MDA-MB-231 cells that genetically over-express COL17A1 gene product exhibited reduced migration and invasion in vitro. Similarly, COL17A1 expression was decreased in metastatic tumors compared to primary tumors and normal tissues, even from the same patients. Moreover, high COL17A1 expression was associated with longer survival of patients with invasive breast carcinoma. In conclusion, we revealed that COL17A1 is a novel p53 transcriptional target in breast tissues that inhibits cell migration and invasion and is associated with better prognosis.

Publications

1. Tanikawa, C., Ueda, K., Suzuki, A., Iida, A., Nakamura, R., Atsuta, N., Tohnai, G., Sobue, G., Saichi, N., Momozawa, Y., Kamatani, Y., Kubo, M., Yamamoto, K., Nakamura, Y., Matsuda, K, Citrullination of RGG motifs in FET proteins by PAD regulates protein aggregation and ALS susceptibility. *Cell reports* in press
2. Toyoshima, O., Tanikawa, C, Yamamoto, R., Watanabe, H., Yamashita, H., Sakitani, K., Yoshida, S., Kubo, M., Matsuo, K., Ito, H., Koike, K., Seto, Y., Matsuda, K Decrease in PSCA expression caused by Helicobacter pylori infection may promote progression to severe gastritis. *Oncotarget* in press
3. Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, M., Ikegawa, S., Hirata, M., Matsuda, K, Kubo, M., Okada, Y., Kamatani, Y. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature Genetics* in press
4. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., Iwasaki, M., Yamaji, T., Sawada, N., Hachiya, T., Tanno, K., Shimizu, A., Hozawa, A., Minegishi, N., Tsugane, S., Yamamoto, M., Kubo, M., Kamatani, Y. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat Genet *49*, 1458-1467.
5. Hachiya, T., Kamatani, Y., Takahashi, A., Hata, J., Furukawa, R., Shiwa, Y., Yamaji, T., Hara, M., Tanno, K., Ohmomo, H., *et al.* (2017). Genetic Predisposition to Ischemic Stroke: A Polygenic Risk Score. Stroke *48*, 253-258.
6. Hata, J., Nagai, A., Hirata, M., Kamatani, Y., Tamakoshi, A., Yamagata, Z., Muto, Matsuda, K, K., Kubo, M., Nakamura, Y., *et al.* (2017). Risk prediction models for mortality in patients with cardiovascular disease: The BioBank Japan project. J Epidemiol *27*, S71-S76.
7. Hirabayashi, S., Ohki, K., Nakabayashi, K., Ichikawa, H., Momozawa, Y., Okamura, K., Yaguchi, A., Terada, K., Saito, Y., Yoshimi, A., *et al.* (2017). ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype. Haematologica *102*, 118-129.
8. Hirata, M., Kamatani, Y., Nagai, A., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Kubo, M., Muto, K., Mushiroda, T., *et al.* (2017 a). Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases. J Epidemiol *27*, S9-S21.
9. Hirata, M., Nagai, A., Kamatani, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Kubo, M., Muto, K., Kiyohara, Y., Mushiroda, T., *et al.* (2017b). Overview of BioBank Japan follow-up data in 32 diseases. J Epidemiol *27*, S22-S28.
10. Ikeda, M., Takahashi, A., Kamatani, Y., Okahisa, Y., Kunugi, H., Mori, N., Sasaki, T., Ohmori, T., Okamoto, Y., Kawasaki, H., *et al.* (2017). A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. Mol Psychiatry.
11. Lin, J., Chung, S., Ueda, K, Matsuda, K., Nakamura, Y., and Park, J.H. (2017). GALNT6 Stabilizes GRP78 Protein by O-glycosylation and Enhances its Activity to Suppress Apoptosis Under Stress Condition. Neoplasia *19*, 43-53.
12. Miyamoto, T., Lo, P.H.Y., Saichi, N., Ueda, K., Hirata, M., Tanikawa. C., and Matsuda, K. (2017a). Argininosuccinate synthase 1 is an intrinsic Akt repressor transactivated by p53. Sci Adv *3*, e1603204.
13. Miyamoto, T., Tanikawa. C., Yodsurang, V., Zhang, Y.Z., Imoto, S., Yamaguchi, R., Miyano, S., Nakagawa, H., and Matsuda, K. (2017b). Identification of a p53-repressed gene module

in breast cancer cells. Oncotarget *8*, 55821-55836.
14. Mori, J., Tanikawa. C., Ohnishi, N., Funauchi, Y., Toyoshima, O., Ueda, K., and Matsuda, K. (2017). EPSIN 3, a novel p53 target, regulates the apoptotic pathway and gastric carcinogenesis. Neoplasia *19*, 185-195.
15. Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., *et al.* (2017). Overview of the BioBank Japan Project: Study design and profile. J Epidemiol *27*, S2-S8.
16. Nakamura, K., Okada, E., Ukawa, S., Hirata, M., Nagai, A., Yamagata, Z., Kiyohara, Y., Muto, K., Kamatani, Y., Ninomiya, T., *et al.* (2017a). Characteristics and prognosis of Japanese female breast cancer patients: The BioBank Japan project. J Epidemiol *27*, S58-S64.
17. Nakamura, K., Ukawa, S., Okada, E., Hirata, M., Nagai, A., Yamagata, Z., Ninomiya, T., Muto, K., Kiyohara, Y., Matsuda, K., *et al.* (2017 b). Characteristics and prognosis of Japanese male and female lung cancer patients: The BioBank Japan Project. J Epidemiol *27*, S49-S57.
18. Okada, E., Ukawa, S., Nakamura, K., Hirata, M., Nagai, A., Matsuda, K., Ninomiya, T., Kiyohara, Y., Muto, K., Kamatani, Y., *et al.* (2017). Demographic and lifestyle factors and survival among patients with esophageal and gastric cancer: The Biobank Japan Project. J Epidemiol *27*, S29-S35.
19. Sapkota, Y., Steinthorsdottir, V., Morris, A.P., Fassbender, A., Rahmioglu, N., De Vivo, I., Buring, J.E., Zhang, F., Edwards, T.L., Jones, S., *et al.* (2017). Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. Nature communications *8*, 15539.
20. Takahashi, Y., Tanikawa. C., Miyamoto, T., Hirata, M., Wang, G., Ueda, K., Komatsu, T., and Matsuda, K. (2017). Regulation of tubular recycling endosome biogenesis by the p53-MICALL1 pathway. Int J Oncol *51*, 724-736.
21. Tamakoshi, A., Nakamura, K., Ukawa, S., Okada, E., Hirata, M., Nagai, A., Matsuda, K., Kamatani, Y., Muto, K., Kiyohara, Y., *et al.* (2017). Characteristics and prognosis of Japanese colorectal cancer patients: The BioBank Japan Project. J Epidemiol *27*, S36-S42.
22. Tanikawa. C., Zhang, Y.Z., Yamamoto, R., Tsuda, Y., Tanaka, M., Funauchi, Y., Mori, J., Imoto, S., Yamaguchi, R., Nakamura, Y., Miyano, S., Nakagawa, H., Matsuda, K. (2017). The Transcriptional Landscape of p53 Signalling Pathway. EBioMedicine *20*, 109-119.
23. Tsuda, Y., Tanikawa. C., Miyamoto, T., Hirata, M., Yodsurang, V., Zhang, Y.Z., Imoto, S., Yamaguchi, R., Miyano, S., Takayanagi, H., Kawano, H., Nakagawa, H., Tanaka, S., Matsuda, K. (2017). Identification of a p53 target, CD137L, that mediates growth suppression and immune response of osteosarcoma cells. Scientific reports *7*, 10739.
24. Ukawa, S., Nakamura, K., Okada, E., Hirata, M., Nagai, A., Yamagata, Z., Muto, Matsuda, K, K., Ninomiya, T., Kiyohara, Y., *et al.* (2017a). Clinical and histopathological characteristics of patients with prostate cancer in the BioBank Japan project. J Epidemiol *27*, S65-S70.
25. Ukawa, S., Okada, E., Nakamura, K., Hirata, M., Nagai, A., Matsuda, K., Yamagata, Z., Kamatani, Y., Ninomiya, T., Kiyohara, Y., *et al.* (2017b). Characteristics of patients with liver cancer in the BioBank Japan project. J Epidemiol *27*, S43-S48.
26. van Rooij, F.J., Qayyum, R., Smith, A.V., Zhou, Y., Trompet, S., Tanaka, T., Keller, M.F., Chang, L.C., Schmidt, H., Yang, M.L., *et al.* (2017). 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*, 51-63.
27. Yodsurang, V., Tanikawa. C., Miyamoto, T., Lo, P.H.Y., Hirata, M., and Matsuda, K. (2017). Identification of a novel p53 target, COL17A1, that inhibits breast cancer cell migration and invasion. Oncotarget *8*, 55790-55803.
28. Yokomichi, H., Nagai, A., Hirata, M., Kiyohara, Y., Muto, K., Ninomiya, T., Matsuda, K., Kamatani, Y., Tamakoshi, A., Kubo, M., *et al.* (2017a). Serum glucose, cholesterol and blood pressure levels in Japanese type 1 and 2 diabetic patients: BioBank Japan. J Epidemiol *27*, S92-S97.
29. Yokomichi, H., Nagai, A., Hirata, M., Kiyohara, Y., Muto, K., Ninomiya, T., Matsuda, K., Kamatani, Y., Tamakoshi, A., Kubo, M., *et al.* (2017b). Survival of macrovascular disease, chronic kidney disease, chronic respiratory disease, cancer and smoking in patients with type 2 diabetes: BioBank Japan cohort. J Epidemiol *27*, S98-S106.
30. Yokomichi, H., Nagai, A., Hirata, M., Tamakoshi, A., Kiyohara, Y., Kamatani, Y., Muto, K., Ninomiya, T., Matsuda, K., Kubo, M., *et al.* (2017c). Statin use and all-cause and cancer mortality: BioBank Japan cohort. J Epidemiol *27*, S84-S91.
31. Yokomichi, H., Noda, H., Nagai, A., Hirata, M., Tamakoshi, A., Kamatani, Y., Kiyohara, Y., Matsuda, K., Muto, K., Ninomiya, T., *et al.* (2017d). Cholesterol levels of Japanese dyslipidaemic patients with various comorbidities: BioBank Japan. J Epidemiol *27*, S77-S83.

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. *In silico* analysis of trans-splicing mechanism in chordates

Rui Yokomori, Kotaro Shimai[1], Takehiro G. Kusakabe1 and Kenta Nakai: [1]Fac. of Sci. and Engineering, Konan Univ

Trans-splicing is a post-transcriptional event, which joins exons from two separate transcripts, while *cis*-splicing joins exons from an identical transcript. It has been reported that *trans*-splicing occurs in various organisms, such as bacteria, virus, and animals. For example, in human, chimeric RNAs expressed in cancer cells were shown to be expressed also in normal cells by *trans*-splicing. Also a recent study has shown that *trans*-spliced long-non-coding RNA, tsRMST, plays an important role in maintenance of pluripotency in human ES cells. However, the mechanism of *trans*-splicing is still poorly understood. In addition, *trans*-splicing event is quite rare in vertebrates, making it difficult to study its global characteristics. In this study, we use *Ciona intestinalis*, the closest invertebrate relative of vertebrates, as a model organism to study *trans*-splicing mechanism. In *Ciona*, approximately half the genes are thought to undergo trans-splicing, called spliced-leader (SL) *trans*-splicing which joins the 5'-exon of a small RNA and 3'-exons of a mRNA. We found that 5'end regions of trans-spliced genes were GU-rich, and several known factors, which are involved in alternative splicing, were significantly enriched in those regions compared to non-trans-spliced genes. Also we found that the strength of the first 5'-splice sites of *trans*-spliced genes was significantly lower than that of non-trans-spliced genes. Interestingly, this characteristic was conserved between *Ciona* and human. Taken together, our results suggest that *trans*-splicing is regulated by some splicing factor, and its basic mechanism is conserved in chordates including vertebrates.

2. Cell type specific features of non-CpG methylation

Jong-Hun Lee and Kenta Nakai

The methylated non-CpG sites (i.e. mCpHs) are emerging as key epigenetic marks in mammalian cells. They are abundant in pluripotent stem cells and brain tissue, regulating cell-type specific processes such as cell differentiation and neurogenesis. Interestingly, in the two cell types, the mCpHs are differently distributed, and involved in gene ex-

pression. However, underlying mechanism of the cell type-specific regulation of the mCpHs is still unknown. In this year, we focused on understanding the mechanism of cell-type specific regulation of mCpHs. To do this, we analyzed public whole genome bisulfite sequencing (WGBS) data, RNA sequencing data, and ChIP sequencing data of pluripotent stem cells and brain tissues. Primary results are as below. First, we found that the cell-type specific distribution of mCpHs results from different activity of DNMT3a and DNMT3b. In addition, the DNMT3b preferentially interacts with H3K36me3 marks, resulting in hypermethylation on highly expressed gene-body regions. Second, we found that even though DNMT3a and DNMT3b tend to methylate CpGs and CpHs simultaneously, in some genomic regions, the distribution of methylated CpGs (i.e. mCpGs) and mCpHs are greatly different. We systematically detected those regions using hidden markov model (HMM), and found that the regions are much broad in brain tissues than in other tissues. In addition, we found that the regions are largely covered by H3K36me3 and H3K9me3 marks, indicating that cell type-specific CpH methylation mechanism is linked to histone modification. Altogether, this study gives insights on understanding cell-type specific regulation of mCpHs.

3. Cell specific change of DNA hydroxymethylation and functional analysis of pluripotent stem cell and somatic cell

Yasuhisa Ishikawa and Kenta Nakai

Epigenetic Factors like DNA methylation or histone modification are well known for influencing various biological phenomena (gene expression, cell differentiation, cell reprogramming and so on). In recent years, among these epigenetic factors DNA hydroxymethylation and the product 5hmC (5-hydroxymethylated-cytosine) is attracting researcher's attention. 5hmC is found to be widespread in many tissues and cell types at different levels. In particular, 5hmC is abundant in the central nervous system and ESCs. Therefore, understanding the dynamic 5hmC changes during reprogramming will provide further insight into somatic cell reprogramming mechanisms. More than ten years ago the well-known reprogramming factors (Oct3/4, Sox2, Klf4, Myc) were identified based on the difference of gene expression between ESC and somatic cell. Investigating the change of DNA methylation and hydroxy methylation may lead to finding new reprogramming factors, terms, mechanism, and other clues instead of gene expression change. So, we investigated the distribution of these epigenetic factors in ESC and somatic cells, and checked the differences around whole genes. As a result, remarkable differences are observed around genebody. Next, we investigated the relationship between gene expression and the previous differences of distribution. Certain distribution patterns were correlated with gene expression. From these results and further analysis, there is a possibility that research of combining these epigenetic factors and gene expression can bring more information about reprogramming.

4. Profiling expressed genes identified by 3D chromosomal conformation analyses in mouse development

Luis AE Nagai and Kenta Nakai

In eukaryotes, long genomic DNA strands are divided into chromosomes and densely compacted within small nuclear compartments. Consequently, each chromosome is non-randomly organized into 3 D structures, which functions in diverse biological processes. The 3D structure includes highly self-interactive genomic regions, defined as compartments and sub-compartments called topologically associating domains (TADs). Recently, many studies show that TADs play an important role in the regulation of gene expression. However, the full scope of how gene expression is regulated by the higher-order chromatin structure remains to be clarified. Here, by integrating heterogeneous NGS data of mouse cells (Hi-C, RNA-seq and ChIP-seq), we investigate the impact of TADs on gene expression regulation. Using RNA-seq data, we first identified differentially expressed genes (DEGs) in each cell compared with ES cell; 2572, 1728 and 4383 genes for B cell, cortex and sperm, respectively. Next, we defined A/B compartments and identified ~3000 TADs ranging from 80 kb to 3 Mb with average of 880 kb in length. We found that 61% of DEG promoters are included in TADs we identified. A compartments tend to include histone enrichment of H3K4me1, H 3K4me3, H3K36me3, and H3K27ac. Preliminary results suggest that Hi-C analysis from different sources can be integrated with histone modification and RNA-Seq data to provide insights into chromatin conformation function. TADs and compartments provide different level for profiling DEGs.

5. Integrative analysis of gene expression and DNA methylation using unsupervised feature extraction for detecting candidate cancer biomarkers

Myungjin Moon and Kenta Nakai

Currently, cancer biomarker discovery is one of the important research topics worldwide. In particular, detecting significant genes related to cancer is an important task for early diagnosis and treatment of cancer. Conventional studies mostly focus

on genes that are differentially expressed in different states of cancer; however, noise in gene expression datasets and insufficient information in limited datasets impede precise analysis of novel candidate biomarkers. In this study, we propose an integrative analysis of gene expression and DNA methylation using normalization and unsupervised feature extractions to identify candidate biomarkers of cancer using renal cell carcinoma RNA-seq datasets. Gene expression and DNA methylation datasets are normalized and integrated into a one-dimensional dataset that retains the major characteristics of the original datasets by unsupervised feature extraction methods, and differentially expressed genes are selected from the integrated dataset. Use of the integrated dataset demonstrated improved performance as compared with conventional approaches that utilize gene expression or DNA methylation datasets alone. Validation based on the literature showed that a considerable number of top-ranked genes from the integrated dataset have known relationships with cancer, implying that novel candidate biomarkers can also be acquired from the proposed analysis method. Furthermore, we expect that the proposed method can be expanded for applications involving various types of multi-omics datasets.

6. Biomarker discovery by integrated joint non-negative matrix factorization and pathway signature analyses

Naoya Fujita, Katsuhiko Murakami[2] and Kenta Nakai: [2]School of Bioscience and Biotechnology, Tokyo University of Technology

Predictive biomarkers are important for selecting appropriate patients for particular treatments. Comprehensive genomic, transcriptomic, and pharmacological data provide clues for understanding relationships between biomarkers and drugs. However, it is still difficult to mine biologically meaningful biomarkers from multi-omics data. Here, we developed an approach for mining multi-omics cell line data by integrating joint non-negative matrix factorization (JNMF) and pathway signature analyses to identify candidate biomarkers. The JNMF detected known associations between biomarkers and drugs such as BRAF mutation with PLX4720 and HER2 amplification with lapatinib. Furthermore, new candidate biomarkers have also been prioritized. Our biomarker discovery scheme represents an integration of JNMF multi-omics clustering and multi-layer interpretation based on pathway gene signature analyses. This approach is also expected to be useful for establishing drug development strategies, identifying pharmacodynamic biomarkers, in mode of action analysis, as well as for mining drug response data in a clinical setting.

7. Cell-free DNA exome sequencing of pancreatic juice from intraductal papillary mucinous tumors of pancreas

Raúl Nicolás Mateos, Masashi Fujita[3], Seiko Hirono, Shinichi Takano, Mitsuharu Fukazawa, Satoru Yasukawa, Munmee Dutta, Nobuyuki Enomoto, Yuki Yamaue, Hidewaki Nakagawa[3] and Kenta Nakai: [3]RIKEN Center for Integrative Medical Sciences

Intraductal papillary mucinous neoplasm (IPMN), despite being an indolent neoplasm of pancreas, has high risk to develop to invasive cancer or co-occur with malignant lesion. Hence, it is important to assess its malignant risk by non-invasive way. In order to evaluate, establish the potential of liquid biopsy of pancreatic juice (PJ) from IPMN as a reliable source for genomic analysis and find potential markers for malignancy, we performed deep exome sequencing analysis of cell-free DNAs obtained from pancreatic juice and blood from 42 patients with IPMN with or without malignant lesion. After filtering low-quality sample data, we detected somatic mutations of KRAS, GNAS, TP 53, and RNF43 among others. We also analyzed copy number alterations (CNAs) from these exome data, and compared those between IPMNs with and without malignant lesions. We observed loss of 7q22.1 band, and gain of multiple bands, two of them significantly, which could be related with the presence of malignant lesions in IPMNs. These findings indicate that cell-free sequencing analysis of PJ and detecting mutations and CNAs of driver genes would have a high potential to assess the malignant progression risk of IPMNs.

8. Whole genome sequencing analysis of esophageal cancers responding or non-responding to neo-adjuvant chemotherapy

Munmee Dutta, Masashi Fujita[3], Tadashi Yasuda[4], Raúl Nicolás Mateos, Ashwini Patil, Kenta Nakai and Hidewaki Nakagawa[3]: [4]Department of Surgery, Kinki University School of Medicine, Osaka

Esophageal squamous cell carcinoma (ESCC) is one of the most aggressive types of cancer predominant in Asian countries, including Japan. However, it is less common in western countries, where Esophageal Adenocarcinoma (EAC) is prevalent. In Japan, neoadjuvant chemotherapy has been used as primary therapeutic strategy with or without radiation followed by surgical resection, though the survival rate is still poor. To better understand the underlying genetic mechanisms related with its carcinogenesis and response to chemotherapy, we performed whole genome sequencing (WGS) analysis on biopsy specimens from 16 ESCC patients be-

fore chemo-(radiation) therapy. Among them, 7 patients responded very well (Complete Response (CR) or almost CR) and 9 responded poorly (Stable Disease (SD) or Partial Disease (PD)). Overall, WGS analysis detected non-silent mutations of TP53, TTN and NOVA1, and recurrent structural variations (SVs) of TP63, LRP1B, SHANK2, TTC28 and WWOX. Furthermore, WGS analysis found recurrent amplified and deleted regions, such as 11q13.3, 9p21.3 and 4q35.2. These genomic alterations detected by WGS may have a potential to predict the response to chemotherapy for ESCC.

9. Identification of *cis*-regulatory modules of promoters in *Drosophila* embryo from single-cell RNA-seq transcriptome data.

Katsuhiko Murakami and Kenta Nakai

Single-cell RNA-seq is becoming an emerging established technology. To date, the data of special pattern of gene expression at the single-cell level in Drosophila embryo have become available. The information allows us to explore the difference of gene expression levels in different cells. Our study aims to examine the relationships among cis-regulatory modules (CRM) and gene expression from the viewpoint of special (cell) differences to better explain the behavior of gene expression in Drosophila embryo. To this end, firstly we obtained new gene expression data with 8,924 genes and 3,039 estimated cell positions at the single-cell level in stage 6 embryo from DVEX dataset. Then we applied several clustering methods, and found dozens of clusters of genes and cells, which might be the basic modules of complicated patterns. The resultant clusters will be compared with the biological knowledge with some genes known to regulate cell developments.

10. Functional profiling of microbial contaminants found from human cultured cells by next generation sequencing data

Sung-Joon Park, Satoru Onizuka[5], Takanori Iwata[5], Kenta Nakai,: [5]Institute of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University

Stem cell therapy and transplantation hold promise for treating diseases difficult to cure. To ensure safety and efficacy of the cell therapy, protecting the cell resources from bacterial and viral infections is a crucial issue. Particularly in the field of allogenic cell therapy, the diagnosis of intrinsic infection of donors is imperative to prevent the disease transmission. Here, we developed an analysis pipeline to identify contaminant DNA and RNA molecules from host NGS data. Unlike other methods, our pipeline simultaneously performs the contaminant profiling and the usual NGS analyses (e.g. whole-genome and transcriptome analyses). Using simulated data, we confirmed that our pipeline is able to correctly identify 99% of contaminants in genera level. In species level, this performance was reduced to 65% due to the higher degree of sequence homology among contaminant species in a same genus. We applied the method to over 300 public human RNA-seq datasets, and found the prevalence of microbial contaminants including potential pathogens. Interestingly, specific gene expression levels were associated with the level of contamination, which suggests the existence of functional interplay between host transcription and contamination.

11. OpenLooper: 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 have developed 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 NGS data, which promotes 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 researches.

12. Analyzing chromatin accessibility dynamics in male germ cell development

Sung-Joon Park, Soichiro Yamanaka[6], Haruhiko Siomi[6], Kenta Nakai: [6]Department of Molecular Biology, Keio University School of Medicine

One of the most dramatic phenomena in mammalian life cycle is that the chromatin architecture of primordial germ cells undergoes remodeling during differentiation and spermatogenesis, accompanied with drastic DNA demethylation and pervasive transcriptomic activity. To clarify the dynamics of chromatin conformational changes in male germ cell development, here we analyzed ATAC-seq (assay for transposase-accessible chromatin using sequencing) data examined with mouse EGFP-positive gonocytes at seven developmental stages ranging from embryonic day E13.5 to postnatal P6. By

incorporating public RNA-seq and ChIP-seq data, we found that specific genomic domains, spanning more than mega bases, exhibit open/close chromatin states with positive or negative correlation of gene expression changes during development. These domains potentially include regulatory elements for certain gene groups that are located at distal intra- or different inter-chromosome loci. In addition, the process of DNA demethylation and re-establishment cooperates with the state of chromatin accessibility. These observations suggest the existence of complex cross-talk among various genetic and epigenetic regulatory elements.

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

Yasuo Ouchi[7], Ashwini Patil, Yusuke Tamura[8], Hiroshi Nishimasu[9], Naoki Takemura[7,8], Takeshi Sato[10], Yasumasa Kimura[10], Osamu Nureki[9], Kenta Nakai, Hiroshi Kiyono[9,11,12], Satoshi Uematsu[7,8]: [7]Dept of Mucosal Immunology, Chiba University, [8]Division of Innate Immune Regulation, IMS, U. Tokyo, [9]Dept of Biol Sci, U. Tokyo, [10]Division of Systems Immunology, IMS, U. Tokyo, [11]Division of Mucosal Immunology, IMS, U. Tokyo,[12]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 the 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 anti-tumour activity.

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

Motohiko Kadoki[13], Ashwini Patil, Cornelius C. Thaiss[13], Donald J. Brooks[13], Surya Pandey[13], Deeksha Deep[13], David Alvarez[15], Ulrich H. von Andrian[15], Amy J. Wagers[14], Kenta Nakai, Tarjei S. Mikkelsen[14], Magali Soumillon[14], Nicolas Chevrier[13]: [13]Faculty of Arts & Sciences Center for Systems Biology, [14]Dept of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard Univ., [15]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.

15. Prediction of MoRF regions in intrinsically disordered protein sequences

Ronesh Sharma[16,17], Gaurav Raicar[16], Maitsetseg Bayarjargal[17], Tatsuhiko Tsunoda[18,19], Ashwini Patil[§], Alok Sharma[17,18,19§]: [16]Fiji National University, Suva, Fiji, [17]The University of South Pacific, Suva, Fiji, [18]RIKEN Center for Integrative Medical Science, [19]Medical Research Institute, Tokyo Medical and Dental University

Intrinsically disordered proteins lack stable 3-dimensional structure and play a crucial role in performing various biological functions. Key to their biological function are the molecular recognition features (MoRFs) located within long disordered protein sequences. Computationally identifying these MoRFs is a challenging task. In this study, we created new MoRF predictors, MorfPredplus and OPAL, to identify MoRFs in disordered protein sequences. Both predictors were evaluated using multiple test sets that have been previously used to evaluate MoRF predictors. The results demonstrate that OPAL outperforms all the available MoRF predictors and is the most accurate predictor available for MoRF prediction.

16. TimeXNet Web: Active gene networks and pathways using time-course biological data

Phit Ling Tan, Yosvany López[19], Kenta Nakai, Ashwini Patil

TimeXNet Web implements an algorithm to identify cellular response networks using time-course transcriptomic, proteomic or phospho-proteomic data and a molecular interaction network. It uses minimum cost flow optimization to find the most probable paths connecting genes/proteins activated at successive time points within the interaction network. It is implemented in Java and uses the GNU Linear Programming Kit. TimeXNet has been evalu-

ated in multiple species and compared with other similar algorithms. It has been shown to reconstruct known pathways in KEGG in the mammalian immune system and the yeast osmotic stress response. It is the only tool providing a web interface for the identification of cellular response networks using multiple types of biological time-course data.

Publications

Farmanbar, A., Firouzi, S., Makalowski, W., Iwanaga, M., Uchimaru, K., Utsunomiya, A., Watanabe, T., Nakai, K. Inferring clonal structure in HTLV-1-infected individuals: towards bridging the gap between analysis and visualization. *Hum Genomics,* 11: 15, 2017.

Farmanbar, A., Firouzi, S., Park, S.J., Nakai, K., Uchimaru, K., 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 Med Genomics,* 10: 4, 2017.

Firouzi, S., Farmanbar, A., Nakai, K., Iwanaga, M., Uchimaru, K., Utsunomiya, A., Suzuki, Y., Watanabe, T. Clonality of HTLV-1-infected T cells as a risk indicator for development and progression of adult T-cell leukemia. *Blood Adv,* 1: 1195-1205, 2017.

Kadoki, M., Patil, A., Thaiss, C.C., Brooks, D.J., Pandey, S., Deep, D., Alvarez, D., Von Andrian, U.H., Wagers, A.J., Nakai, K., Mikkelsen, T.S., Soumillon, M., Chevrier, N. Organism-Level Analysis of Vaccination Reveals Networks of Protection across Tissues. *Cell,* 171: 398-413 e21, 2017.

Lee, J.H., Park, S.J., Nakai, K. Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. *Sci Rep,* 7: 11295, 2017.

Lopez, Y., Vandenbon, A., Nose, A., Nakai, K. Modeling the cis-regulatory modules of genes expressed in developmental stages of Drosophila melanogaster. *PeerJ,* 5: e3389, 2017.

Suzuki, A., Kawano, S., Mitsuyama, T., Suyama, M., Kanai, Y., Shirahige, K., Sasaki, H., Tokunaga, K., Tsuchihara, K., Sugano, S., Nakai, K., Suzuki, Y. DBTSS/DBKERO for integrated analysis of transcriptional regulation. *Nucleic Acids Res,* 46: D229-D238, 2018.

Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K.R., Piette, J., Reeves, W., Scornavacca, C., Simion, P., Vincentelli, R., Bellec, M., Aicha, S.B., Fagotto, M., Gueroult-Bellone, M., Haeussler, M., Jacox, E., Lowe, E.K., Mendez, M., Roberge, A., Stolfi, A., Yokomori, R., Brown, C.T., Cambillau, C., Christiaen, L., Delsuc, F., Douzery, E., Dumollard, R., Kusakabe, T., Nakai, K., Nishida, H., Satou, Y., Swalla, B., Veeman, M., Volff, J.N., Lemaire, P. ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Res,* 46: D718-D725, 2018.

Sharma, R., Bayarjargal, M., Tsunoda, T., Patil, A., Sharma, A. MoRFPred-plus: Computational Identification of MoRFs in Protein Sequences using Physicochemical Properties and HMM profiles. *J. Theor Biol .*, 437: 9-16, 2018.

Patil, A. Protein-protein interaction databases. *Encyclopedia of Bioinformatics and Computational Biology*, accepted, 2018.

Sharma, R., Raicar, G., Tsunoda, T., Patil, A.*, Sharma, A*. OPAL: Prediction of MoRF regions in intrinsically disordered protein sequences. *Bioinformatics*, accepted, 2018.

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.		准助教	博士(社会医学)	井　上　悠　輔
Project Assistant Professor	Hyunsoo Hong, Ph.D.		特任助教	博士(学術)	洪　　賢　秀
Project Assistant Professor	Akiko Nagai, Ph.D.		特任助教	博士(医科学)	永　井　亜貴子

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) 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 (the 1st cohort) and 66,000 patients (the 2nd cohort). 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-CREATE 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 and opportunities for ethical training based on ethical guidelines.

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

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" (Figure 1). In order to make regenerative medicine more concrete, it is essential to promote research development with a definite focus on clinical applications and to establish a framework for clinical research at an early stage. We provided more than 70 consultations for stem cell researchers per year. Topics of those consultations include research design, informed consent, Institutional Review Boards (IRBs), return of research results, inclusion criterion of participants of first-in-human trials and governance of iPSC banking.

Figure 1. Examples of brochures and informed consent tools of the program for intractable diseases research using disease-specific induced pluripotent stem (iPS) cells

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. Japanese Public Attitudes toward the creation and utilization of human-animal chimeras

Ongoing research on making "human-animal chimeras" or "animals containing human material" (ACHM) to develop regenerative medicine and to solve the shortage of organs available for transplantation has raised many ethical issues regarding the creation and utilization of such constructs, including cultural views regarding the status of those creations. A pilot study was conducted to explore Japanese public attitudes toward human-animal chimeras or ACHM. The February 2012 study consisted of focus group interviews (FGIs) with citizens from the Greater Tokyo Area, aged between 20 and 54. The 24 participants were divided into four groups. Transcripts of the interviews were analyzed and participants' attitudes were categorized. Five categories of participant attitudes were identified: (1) resistance to the unnatural, (2) concerns about animal welfare, (3) concerns about controlling human-animal chimeras, (4) concerns about the possible birth of intermediate entities, and (5) resistance to creating and utilizing animals containing my material or my child's material. Our FGI results showed a broader and greater variety of public concerns than those reported in previous studies. While researchers have tried to establish new methods to avoid creating intermediate entities, our participants expressed concerns about not only intermediate entities but also animals containing their own material or their child's material. Based upon their responses in the interviews, we are introducing a new ethical concern: "animals containing my material/my child's material."

4. Lessons for reviewing clinical trials using induced pluripotent stem cells: examining the case of a first-in-human trial for age-related macular degeneration

The iPSC-FIH trial was approved in 2014. To identify future lessons for the committees reviewing cutting-edge FIH trials using novel stem cells, the minutes of the iPSC-FIH trial review committee meetings were examined. The protocol of the iPSC-FIH trial was reviewed in 14 meetings held by three IRBs and two National Research Couucils (NRCs) from 5 July 2012 to 29 July 2013; the Ministry granted its approval on 19 July 2013. These meeting minutes were obtained for analysis from the relevant RIKEN and Ministry of Health, Labor and Welfare websites. The minutes from the remaining two IRBs were not available online; thus, these were requested directly from them. The meetings covered a wide range of subjects, including the quality control and risks of the trial (e.g., the risks of iPSCs especially the risks of tumorigenicity and genetic aberrations, contamination with harmful viruses and bacteria during the process of establishing iPSCs, and surgical complications), protection of research participants (e.g., the suitability of the patient information sheet in securing informed consent, preventing therapeutic misconception in which the research participant fails to distinguish between participating in the trial and receiving standard treatment [6], compensation for research-related health damages, and psychological care for participants), the background of the trial and target disease, and the appropriateness of the review process. We highlighted how the review committees evaluated the risks and benefits of the iPSC-FIH trial. Moreover, conversations regarding how information on the potential direct benefits could be provided to participants, in particular relating to how to prevent therapeutic misconception and how to assist the participants in making reasonable decisions, are of special focus in the present report.

5. Comparative analysis of communication on stem cell research and regenerative medicine between the public and the scientific community.

For effective communication on issues concerning stem cell research (SCR) and regenerative medicine (RM), it is essential to understand the hurdles, motivation, and other issues affecting scientists' active participation in science communication to bridge the gap between science and society. Our study analyzed 1,115 responses of the Japanese Society for Regenerative Medicine regarding their attitudes toward science communication through a questionnaire focusing on the field of SCR and RM. As a result, we found that scientists face systemic issues such as lack of funding, time, opportunities, and evaluation systems for science communication. At the same time, there is a disparity of attitudes toward media discourse between scientists and the public.

Furthermore, we compared with the responses of a large-scale survey with 2,160 citizens. Results showed that the public is more interested in the post-realization aspects of RM, such as cost of care, countermeasures for risks and accidents, and clarification of responsibility and liability, than in the scientific aspects; the latter is of greater interest only to scientists. Our data indicate that an increased awareness about RM-associated social responsibility and regulatory framework is required among scientists, such as those regarding its benefits, potential accidents, abuse, and other social consequences. Awareness regarding the importance of communication and education for scientists are critical to bridge the gaps in the interests of the public and scientists.

6. Informed assent in birth cohort studies

One of the ethical issues surrounding birth cohort studies is how to obtain informed assent from children as they grow up. What and how parents tell their children affects children's future choices about the study, yet few studies have focused on parents' influence on children. Our study examines parents' attitudes towards telling their children about their participation in a specific birth cohort study. We conducted surveys and in-depth interviews with the parents of children who participated in the "Japan Environment and Children's Study" (JECS), which follows children from the foetal stage to age 13. Forty-four mothers and 23 fathers answered the survey, and 11 mothers and 3 fathers participated in in-depth interviews. Parents' attitudes towards "telling" were categorized into 3 communication styles depending on their perception of the risk/benefits for their children. Most parents predicted that the study would benefit their children and preferred "directive telling," which we divided into "empowered telling" (provides children with a positive identity as participants) and "persuasive telling" (attempts to persuade children even if they express reluctance as they grow). A few parents, weighing the study's potential risk, preferred "non-directive telling," which respects children's choices even if that means withdrawing from the study. While "directive telling" may lead children to have positive associations with the study, children should also be told about the risks. Investigators can provide materials that support parents and give children age-appropriate information about their participation, as well as ensure opportunities for children to express their feelings.

Publications

1. Shineha R, Inoue Y, Ikka T, Kishimoto A, Yashiro Y. A comparative analysis of attitudes on communication toward stem cell research and regenerative medicine between the public and the scientific community. Stem Cells Transl Med. 2018 Jan 26. doi: 10.1002/sctm.17-0184.
2. Shineha R, Inoue Y, Ikka T, Kishimoto A, Yashiro Y. Science communication in regenerative medicine: Implications for the role of academic society and science policy. Regenerative Therapy. 7: 89-97, 2017.
3. Takashima K, Inoue Y, Tashiro S, Muto K. Lessons for reviewing clinical trials using induced pluripotent stem cells: examining the case of a first-in-human trial for age-related macular degeneration. Regenerative Medicine. 2017 Dec 6. doi: 10.2217/rme-2017-0130.
4. Ri I, Suda E, Yamagata Z, Nitta H, Muto K. "Telling" and assent: Parents' attitudes towards children's participation in a birth cohort study. Health Expectations. 21 (1): 358-366, 2018.
5. 吉田幸恵，中田はる佳，武藤香織．臨床試験に関与した，がん患者の語り―「治療」と「研究」を区別することの困難さに関する考察．生命倫理．28: 122−131, 2017.
6. 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．48(2): 31−39, 2017.
7. 標葉隆馬，井上悠輔，八代嘉美．ヒト動物キメラを巡る意識の多様性―一般モニター調査の分析から．成城文藝．240: 398－416, 2017.