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	Yao-zhong Zhang, Ph.D.	助　教	博士(情報理工学)	張	燿中
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 Assistant Professor	Taku Onodera, Ph.D.	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. Robust Sample-Specific Stability Selection with Effective Error Control

Park H[1], Yamada M[2], Imoto S[3], Miyano S: [1]Faculty of Global and Science Studies, Yamaguchi University, [2]Kyoto University, Graduate School of Informatics, [3]Health Intelligence Center

Identifying individual characteristics is a crucial issue in personalized genome research. To effectively identify sample-specific characteristics, we propose a novel strategy called robust sample-specific stability selection. Although stability selection shows effective feature selection results and has attractive theoretical property (i.e., per-family error rate control), the method's results are sensitive to the value of the regularization parameter because the method performs feature selection based only on the particular parameter value that maximizes the selection probability. To resolve this issue, we propose robust stability selection and show that our method provides an effective theoretical property (i.e., effective per-family error rate control). We also propose a sample-specific random lasso based on the kernel-based L_1-type regularization and weighted random sampling. The proposed robust sample-specific stability selection estimates the se-

lection probabilities of variables using the sample-specific random lasso and then selects variables based on robust stability selection. Our method controls the effect of samples on sample-specific analysis by the two-stage strategy (i.e., the weighted random sampling and the kernel-based L_1-type approach in the random lasso), and thus we can effectively perform sample-specific analysis without disturbances of samples having characteristics different from those of the target sample. We observe from the numerical studies that our strategies can effectively perform sample-specific analysis and provide biologically reliable results in gene selection.

b. A comprehensive characterization of cis-acting splicing-associated variants in human cancer

Shiraishi Y, Kataoka K[11,19], Chiba K, Okada A, Kogure Y[11], Tanaka H, Ogawa S[11], Miyano S: [19]National Cancer Center Research Institute

Although many driver mutations are thought to promote carcinogenesis via abnormal splicing, the landscape of splicing-associated variants (SAVs) remains unknown due to the complexity of splicing abnormalities. Here, we developed a statistical framework to systematically identify SAVs disrupting or newly creating splice site motifs and applied it to matched whole-exome and transcriptome sequencing data from 8976 samples across 31 cancer types, generating a catalog of 14,438 SAVs. Such a large collection of SAVs enabled us to characterize their genomic features, underlying mutational processes, and influence on cancer driver genes. In fact, ~50% of SAVs identified were those disrupting noncanonical splice sites (non-GT-AG dinucleotides), including the third and fifth intronic bases of donor sites, or newly creating splice sites. Mutation signature analysis revealed that tobacco smoking is more strongly associated with SAVs, whereas ultraviolet exposure has less impact. SAVs showed remarkable enrichment of cancer-related genes, and as many as 14.7% of samples harbored at least one SAVs affecting them, particularly in tumor suppressors. In addition to intron retention, whose association with tumor suppressor inactivation has been previously reported, exon skipping and alternative splice site usage caused by SAVs frequently affected tumor suppressors. Finally, we described high-resolution distributions of SAVs along the gene and their splicing outcomes in commonly disrupted genes, including *TP53*, *PIK3R1*, *GATA3*, and *Nak*, which offers genetic clues for understanding their functional properties. Collectively, our findings delineate a comprehensive portrait of SAVs, novel insights into transcriptional de-regulation in cancer.

c. Virtual Grid Engine: Accelerating thousands of omics sample analyses using large-scale supercomputers

Ito S, Yadome M, Nishiki T[4], Ishiduki S[4], Inoue H[4], Yamaguchi R, Miyano S: [4]Frontier Computing Center, Fujitsu Limited

DNA analyses of rare variant diseases are estimated to require tens of thousands of sample analyses. The use of supercomputers in bioinformatics has become common with the increase in the amount of analysis data. However, only a few studies have utilized massively parallel supercomputers ranked in TOP500. This is because Grid Engine (GE) services, such as Sun Grid Engine and Univa Grid Engine, are not provided on them. Most software and programs that run on supercomputers are paralleled using Message Passing Interface (MPI) [6], wherein all subprocesses work synchronously. On the other hand, array jobs, automatically paralleled subprocesses of software pipelines by GE are asynchronous. Therefore, the GE conflicts with MPI-based systems from the perspective of the job-filling factor. Here, the MPI parallelization of software pipelines requires expert knowledge and experience. It is necessary for the MPI parallelization of software pipelines to use C or Fortran language wrapper programs or to commission High Performance Computing (HPC) experts to overwrite them fundamentally, which will be difficult for users. Recently, Cloud-base systems, such as Amazon Web Services (AWS), have been popular in NGS data analysis. Cloud computing services are very useful for small laboratories and companies that do not have computational resources. However, they still require significant costs for large-scale analyses. In addition, there are still several problems to be overcome, such as data transfer time, data corruption checking, and data security management. From the perspective of HPC, DRAGEN achieved drastic acceleration of the GATK pipeline. The hardware implementation of all the processes in GATK using FPGA caused this great acceleration. This approach is the ultimate software-tuning technique. On the other hand, it makes it quite difficult to improve the implemented workflows. GATK is one of the gold-standard pipelines for somatic mutation calling, so this tuning is extremely efficient for it. However, there is a great variety of target variants for NGS data analyses for each study, and it is inevitable for algorithms and pipelines to be designed for the study. Therefore, general software pipelines still have merits in many studies and massively paralleled supercomputers are useful for accelerating their analyses. In this study, we developed MPI-based middleware named Virtual Grid Engine (VGE) that enables software pipelines based on GE system to run on massively parallel supercomput-

ers. This is a collaboration study with Frontier Computing Center of Fujitsu. VGE is freely available from https://github.com/SatoshiITO/VGE.

d. ALPHLARD: a Bayesian method for analyzing HLA genes from whole genome sequence data

Hayashi S, Yamaguchi R, Mizuno S[5], Komura M, Miyano S, Nakagawa H[6], Imoto S[3]: [5]Center for Advanced Medical Innovation, Kyushu University, [6]RIKEN Center for Integrative Medical Sciences

Regulation of transcription factor Although human leukocyte antigen (HLA) genotyping based on amplicon, whole exome sequence (WES), and RNA sequence data has been achieved in recent years, accurate genotyping from whole genome sequence (WGS) data remains a challenge due to the low depth. Furthermore, there is no method to identify the sequences of unknown HLA types not registered in HLA databases. We developed a Bayesian model, called ALPHLARD, that collects reads potentially generated from HLA genes and accurately determines a pair of HLA types for each of HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, and -DRB1 genes at 3rd field resolution. Furthermore, ALPHLARD can detect rare germline variants not stored in HLA databases and call somatic mutations from paired normal and tumor sequence data. We illustrate the capability of ALPHLARD using 253 WES data and 25 WGS data from Illumina platforms. By comparing the results of HLA genotyping from SBT and amplicon sequencing methods, ALPHLARD achieved 98.8% for WES data and 98.5% for WGS data at 2nd field resolution. We also detected three somatic point mutations and one case of loss of heterozygosity in the HLA genes from the WGS data. ALPHLARD showed good performance for HLA genotyping even from low-coverage data. It also has a potential to detect rare germline variants and somatic mutations in HLA genes. It would help to fill in the current gaps in HLA reference databases and unveil the immunological significance of somatic mutations identified in HLA genes.

e. Targeting Tyro3 ameliorates a model of PGRN-mutant FTLD-TDP via tau-mediated synaptic pathology

Fujita K[7], Chen X[7], Homma H[7], Tagawa K[7], Amano M[8], Saito A, Imoto S[3], Akatsu H[9], Hashizume Y[10], Kaibuchi K[8], Miyano S, Okazawa H[7]: [7]Tokyo Medical and Dental University, [8]Graduate School of Medicine, Nagoya University, [9]Nagoya City University, [10]Aichi Medical University

Mutations in the progranulin (PGRN) gene cause a tau pathology-negative and TDP43 pathology-positive form of frontotemporal lobar degeneration (FTLD-TDP). We generated a knock-in mouse harboring the R504X mutation (PGRN-KI). Phosphoproteomic analysis of this model revealed activation of signaling pathways connecting PKC and MAPK to tau prior to TDP43 aggregation and cognitive impairments, and identified PKCα as the kinase responsible for the early-stage tau phosphorylation at Ser203. Disinhibition of Gas6 binding to Tyro3 due to PGRN , PKCα via PLCγ, inducing tau phosphorylation at Ser203, mislocalization of tau to dendritic spines, and spine loss. Administration of a PKC inhibitor, B-Raf inhibitor, or knockdown of molecules in the Gas6-Tyro3-tau axis rescues spine loss and cognitive impairment of PGRN-KI mice. Collectively, these results suggest that targeting of early-stage and aggregation-independent tau signaling represents a promising therapeutic strategy for this disease.

2. Cancer Genomics

a. Age-related remodelling of oesophageal epithelia by mutated cancer drivers

Yokoyama A[11], Kakiuchi N[11], Yoshizato T[11], Nannya Y[11], Suzuki H[11], Takeuchi Y[11], Shiozawa Y[11], Sato Y[11], Aoki K[11], Kim SK[11], Fujii Y[11], Yoshida K[11], Kataoka K[11], Nakagawa MM[11], Inoue Y[11], Hirano T[11], Shiraishi Y, Chiba K, Tanaka H, Sanada M[12], Nishikawa Y[11], Amanuma Y[11], Ohashi S[11], Aoyama I[11], Horimatsu T[11], Miyamoto S[11], Tsunoda S[11], Sakai Y[11], Narahara M[13], Brown JB[11], Sato Y[14], Sawada G[15], Mimori K[15], Minamiguchi S[11], Haga H[11], Seno H[11], Miyano S, Makishima H[11], Muto M[11], Ogawa S[11,16]: [11]Kyoto University, [12]Nagoya Medical Center, [13]McGill University, [14]Sato Clinic, [15]Kyushu University, [16]Karolinska Institute

Clonal expansion in aged normal tissues has been implicated in the development of cancer. However, the chronology and risk dependence of the expansion are poorly understood. Here we intensively sequence 682 micro-scale oesophageal samples and show, in physiologically normal oesophageal epithelia, the progressive age-related expansion of clones that carry mutations in driver genes (predominantly *NOTCH1*), which is substantially accelerated by alcohol consumption and by smoking. Driver-mutated clones emerge multifocally from early childhood and increase their number and size with ageing, and ultimately replace almost the entire oesophageal epithelium in the extremely elderly. Compared with mutations in oesophageal cancer, there is a marked overrepresentation of *NOTCH1* and *PPM1D* mutations in physiologically normal oesophageal epithelia; these mutations can be acquired before late adolescence (as early as

early infancy) and significantly increase in number with heavy smoking and drinking. The remodelling of the oesophageal epithelium by driver-mutated clones is an inevitable consequence of normal ageing, which-depending on lifestyle risks-may affect cancer development.

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. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia

Shiozawa Y[17,11], Malcovati L[18], Gallì A[18], Sato-Otsubo A[11], Kataoka K[11], Sato Y[11], Watatani Y[11], Suzuki H[11], Yoshizato T[11], Yoshida K[11], Sanada M[12], Makishima H[11], Shiraishi Y, Chiba K, Hellström-Lindberg E[16], Miyano S, Ogawa S[11,16], Cazzola M[18]: [17]Department of Pediatrics, The University of Tokyo, [18]University of Pavia

Spliceosome mutations are frequently found in myelodysplasia. Splicing alterations induced by these mutations, their precise targets, and the effect at the transcript level have not been fully elucidated. Here we report transcriptomic analyses of 265 bone marrow samples from myelodysplasia patients, followed by a validation using CRISPR/Cas9-mediated gene editing and an assessment of nonsense-mediated decay susceptibility. Small but widespread reduction of intron-retaining isoforms is the most frequent splicing alteration in *SF3B1*-mutated samples. *SF3B1* mutation is also associated with 3' splice site alterations, leading to the most pronounced reduction of canonical transcripts. Target genes include tumor suppressors and genes of mitochondrial iron metabolism or heme biosynthesis. Alternative exon usage is predominant in *SRSF2*- and *U2AF1*-mutated samples. Usage of an *EZH2* cryptic exon harboring a premature termination codon is increased in both *SRSF2*- and *U2AF1*-mutated samples. Our study reveals a landscape of splicing alterations and precise targets of various spliceosome mutations.

c. Applications of Genomon for Cancer Genomics and Cancer Clinical Sequence

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: 1-3, 5-7, 9-11, 15-18, 20, 22-23, 26, 28, 31, 33, 39-41, 43, 46-47. With Genomon on HGC supercomputer, we have also contributed to cancer clinical sequence at IMSUT Research Hospital: 21, 29, 32, 42, 48.

3. Contributions by System for Cancer Clinical Sequencing

We have developed a system for cancer clinical sequencing using HGC supercomputer, and have been contributing cancer genomic medicine at IMSUT Research Hospital.

a. Circulating tumor DNA dynamically predicts response and/or relapse in patients with hematological malignancies

Nakamura S[20], Yokoyama K[21], Yusa N[21], Ogawa M[20], Takei T[20], Kobayashi A[20], Ito M[20], Shimizu E, Kasajima R, Wada Y, Yamaguchi R, Imoto S[3], Nagamura-Inoue T[21], Miyano S, Tojo A[20,21]: [20]Advanced Clinical Research Center, The Institute of Medical Science, The University of Tokyo, [21]Research Hospital, The Institute of Medical Science, The University of Tokyo

A growing body of evidence suggests that tumor-derived fragmentary DNA, known as circulating tumor DNA (ctDNA), has the potential to serve as a non-invasive biomarker for disease monitoring. However, in the setting of hematological malignancy, few published studies support the utility of ctDNA. We retrospectively investigated ctDNA levels of 17 patients with various hematological malignancies who had achieved remission after first-line therapy. We identified somatic driver mutations by next-generation sequencing, and designed droplet digital PCR assays for each mutation to measure ctDNA. Variant allele frequencies of ctDNA changed in association with clinical response in all patients. Eight patients clinically relapsed after a median of 297 days post-first-line therapy (termed, "relapsed group"); the remaining nine patients remained disease-free for a median of 332 days (termed, "remission group"). Among patients in the relapsed group, ctDNA levels increased more than twofold at paired serial time points. In marked contrast, ctDNA levels of all patients in the remission group remained undetectable or stable during clinical remission. Notably, ctDNA-based molecular relapse demonstrated a median 30-day lead time over clinical relapse. In summary, ctDNA monitoring may help identify hematologic cancer patients at risk for relapse in advance of established clinical parameters.

b. Cell-lineage level-targeted sequencing to identify acute myeloid leukemia with myelodysplasia-related changes

Yokoyama K[21], Shimizu E, Yokoyama N[21], Nakamura S[20], Kasajima R[3], Ogawa M[20], Takei T[20], Ito M[20], Kobayashi A[20], Yamaguchi R, Imoto S[3], Miyano S, Tojo A[20,21]

Acute myeloid leukemia (AML) is a clonal myeloid neoplasm that typically arises de novo; however, some cases evolve from a preleukemic state, such as myelodysplastic syndrome (MDS). Such secondary AMLs and those with typical MDS-related clinical features are known as AMLs with myelodysplasia-related changes (AML-MRC). Because patients with AML-MRC have poor prognosis, more accurate diagnostic approaches are required. In this study, we performed targeted sequencing of 54 genes in 3 cell populations (granulocyte, blast, and T-cell fractions) using samples from 13 patients with MDS, 16 patients with clinically diagnosed AML-MRC, 4 patients with suspected AML-MRC but clinically diagnosed as AML not otherwise specified (AML-NOS), and 11 patients with de novo AML. We found that overlapping mutations, defined as those shared at least by the blast and granulocyte fractions, were significantly enriched in patients with MDS and AML-MRC, including those with suspected AML-MRC, indicating a substantial history of clonal hematopoiesis. In contrast, blast-specific nonoverlapping mutations were significantly enriched in patients with de novo AML. Furthermore, the presence of overlapping mutations, excluding *DNMT3A*, *TET2*, and *ASXL1*, effectively segregated patients with MDS and AML-MRC or suspected AML-MRC from patients with de novo AML. Additionally, the presence of ≥3 mutations in the blast fraction was useful for distinguishing patients with AML-MRC from those with MDS. In conclusion, our approach is useful for classifying clinically diagnosable AML-MRC and identifying clinically diagnosed AML-NOS as latent AML-MRC. Additional prospective studies are needed to confirm the utility of this approach.

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 preleukemic clones in therapy-related myeloid neoplasms following autologous SCT. *Blood*. 131(16): 1846-1857, 2018.
3. Cardinez C, Miraghazadeh B, Tanita K, da Silva E, Hoshino A, Okada S, Chand R, Asano T, Tsumura M, Yoshida K, Ohnishi H, Kato Z, Yamazaki M, Okuno Y, Miyano S, Kojima S, Ogawa S, Andrews TD, Field MA, Burgio G, Morio T, Vinuesa CG, Kanegane H, Cook MC. Gain-of-function IKBKB mutation causes human combined immune deficiency. *J Exp Med*. 215 (11): 2715-2724, 2018.
4. Fujita K, Chen X, Homma H, Tagawa K, Amano M, Saito A, Imoto S, Akatsu H, Hashizume Y, Kaibuchi K, Miyano S, Okazawa H. Targeting Tyro3 ameliorates a model of PGRN-mutant FTLD-TDP via tau-mediated synaptic pathology. *Nat Commun*. 9(1): 433, 2018.
5. Fujita M, Matsubara N, Matsuda I, Maejima K, Oosawa A, Yamano T, Fujimoto A, Furuta M, Nakano K, Oku-Sasaki A, Tanaka H, Shiraishi Y, Mateos RN, Nakai K, Miyano S, Tomita N, Hirota S, Ikeuchi H, Nakagawa H. Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling. *Oncotarget*. 9(1): 969-981, 2018.
6. Furuta M, Tanaka H, Shiraishi Y, Unida T, Imamura M, Fujimoto A, Fujita M, Sasaki-Oku A, Maejima K, Nakano K, Kawakami Y, Arihiro K, Aikata H, Ueno M, Hayami S, Ariizumi SI, Yamamoto M, Gotoh K, Ohdan H, Yamaue H, Miyano S, Chayama K, Nakagawa H. Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. *Oncotarget*. 9(38): 25075-25088, 2018.
7. Hamada M, Doisaki S, Okuno Y, Muramatsu H, Hama A, Kawashima N, Narita A, Nishio N, Yoshida K, Kanno H, Manabe A, Taga T, Takahashi Y, Miyano S, Ogawa S, Kojima S. Whole-exome analysis to detect congenital hemolytic anemia mimicking congenital dyserythropoietic anemia. *Int J Hematol*. 108(3), 306-311, 2018.
8. Hayashi S, Yamaguchi R, Mizuno S, Komura M, Miyano S, Nakagawa H, Imoto S. ALPHLARD: a Bayesian method for analyzing HLA genes from whole genome sequence data. *BMC Genomics*. 19(1): 790, 2018.
9. 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*. 53(2): 225-227, 2018.

10. Hoshino A, Yang X, Tanita K, Yoshida K, Ono T, Nishida N, Okuno Y, Kanzaki T, Goi K, Fujino H, Ohshima K, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Kojima S, Morio T, Kanegane H. Modification of cellular and humoral immunity by somatically reverted T cells in X-linked lymphoproliferative syndrome type 1. *J Allergy Clin Immunol*. 143(1): 421-424.e11, 2019.
11. Hoshino A, Takashima T, Yoshida K, Morimoto A, Kawahara Y, Yeh TW, Okano T, Yamashita M, Mitsuiki N, Imai K, Sakatani T, Nakazawa A, Okuno Y, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Kojima S, Morio T, Kanegane H. Dysregulation of Epstein-Barr Virus Infection in Hypomorphic ZAP70 Mutation. *J Infect Dis*. 218(5): 825-834, 2018.
12. Inoue D, Fujino T, Sheridan P, Zhang YZ, Nagase R, Horikawa S, Li Z, Matsui H, Kanai A, Saika M, Yamaguchi R, Kozuka-Hata H, Kawabata KC, Yokoyama A, Goyama S, Inaba T, Imoto S, Miyano S, Xu M, Yang FC, Oyama M, Kitamura T. A novel ASXL1-OGT axis plays roles in H3K4 methylation and tumor suppression in myeloid malignancies. *Leukemia*. 32(6): 1327-1337, 2018.
13. 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*. 78(4): 865-876, 2018.
14. Ito S, Yadome M, Nishiki T, Ishiduki S, Inoue H, Yamaguchi R, Miyano S. Virtual Grid Engine: Accelerating thousands of omics sample analyses using large-scale supercomputers. *IEEE BIBM 2018*, 2018. In press.
15. 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.
16. 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*. 59(6): 1490-1493, 2018.
17. 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.
18. Kimura S, Seki M, Yoshida K, Shiraishi Y, Akiyama M, Koh K, Imamura T, Manabe A, Hayashi Y, Kobayashi M, Oka A, Miyano S, Ogawa S, Takita J. NOTCH1 pathway activating mutations and clonal evolution in pediatric T-cell acute lymphoblastic leukemia. *Cancer Sci*. 2018 Nov 2. doi: 10.1111/cas.13859.
19. 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*. 63(2): 239-248, 2018.
20. Kobayashi K, Mizuta S, Yamane N, Ueno H, Yoshida K, Kato I, Umeda K, Hiramatsu H, Suehiro M, Maihara T, Usami I, Shiraishi Y, Chiba K, Miyano S, Adachi S, Ogawa S, Kiyokawa N, Heike T. Paraneoplastic hypereosinophilic syndrome associated with IL3-IgH positive acute lymphoblastic leukemia. *Pediatr Blood Cancer*. 66(1): e27449, 2019.
21. 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.
22. Kotani S, Yoda A, Kon A, Kataoka K, Ochi Y, Shiozawa Y, Hirsch C, Takeda J, Ueno H, Yoshizato T, Yoshida K, Nakagawa MM, Nannya Y, Kakiuchi N, Yamauchi T, Aoki K, Shiraishi Y, Miyano S, Maeda T, Maciejewski JP, Takaori-Kondo A, Ogawa S, Makishima H. Molecular pathogenesis of disease progression in MLL-rearranged AML. *Leukemia*. 2018 Sep 12. doi: 10.1038/s41375-018-0253-3.
23. Matsuo H, Yoshida K, Fukumura K, Nakatani K, Noguchi Y, Takasaki S, Noura M, Shiozawa Y, Shiraishi Y, Chiba K, Tanaka H, Okada A, Nannya Y, Takeda J, Ueno H, Shiba N, Yamato G, Handa H, Ono Y, Hiramoto N, Ishikawa T,

Usuki K, Ishiyama K, Miyawaki S, Itonaga H, Miyazaki Y, Kawamura M, Yamaguchi H, Kiyokawa N, Tomizawa D, Taga T, Tawa A, Hayashi Y, Mano H, Miyano S, Kamikubo Y, Ogawa S, Adachi S. Recurrent CCND3 mutations in MLL-rearranged acute myeloid leukemia. *Blood Adv*. 2(21): 2879-2889, 2018.

24. Mimori K, Saito T, Niida A, Miyano S. Cancer evolution and heterogeneity. *Ann Gastroenterol Surg*. 2(5): 332-338, 2018.
25. Miyano S. Artificial Intelligence for Cancer Genomic Medicine: Understanding Cancer is Beyond Human Ability. *Brain Nerve*. 71(1): 25-32, 2019.
26. Murakami N, Okuno Y, Yoshida K, Shiraishi Y, Nagae G, Suzuki K, Narita A, Sakaguchi H, Kawashima N, Wang X, Xu Y, Chiba K, Tanaka H, Hama A, Sanada M, Ito M, Hirayama M, Watanabe A, Ueno T, Kojima S, Aburatani H, Mano H, Miyano S, Ogawa S, Takahashi Y, Muramatsu H. Integrated molecular profiling of juvenile myelomonocytic leukemia. *Blood*. 131 (14): 1576-1586, 2018.
27. Muraoka D, Seo N, Hayashi T, Tahara Y, Fujii K, Tawara I, Miyahara Y, Okamori K, Yagita H, Imoto S, Yamaguchi R, Komura M, Miyano S, Goto M, Sawada SI, Asai A, Ikeda H, Akiyoshi K, Harada N, Shiku H. Antigen delivery targeted to tumor-associated macrophages overcomes tumor immune resistance. *J Clin Invest*. 2019 Jan 10. pii: 97642. doi: 10.1172/JCI97642.
28. Nagao Y, Mimura N, Takeda J, Yoshida K, Shiozawa Y, Oshima M, Aoyama K, Saraya A, Koide S, Rizq O, Hasegawa Y, Shiraishi Y, Chiba K, Tanaka H, Nishijima D, Isshiki Y, Kayamori K, Kawajiri-Manako C, Oshima-Hasegawa N, Tsukamoto S, Mitsukawa S, Takeda Y, Ohwada C, Takeuchi M, Iseki T, Misawa S, Miyano S, Ohara O, Yokote K, Sakaida E, Kuwabara S, Sanada M, Iwama A, Ogawa S, Nakaseko C. Genetic and transcriptional landscape of plasma cells in POEMS syndrome. *Leukemia*. 2019 Jan 11. doi: 10.1038/s41375-018-0348-x.
29. Nakamura S, Yokoyama K, Yusa N, Ogawa M, Takei T, Kobayashi A, Ito M, Shimizu E, Kasajima R, Wada Y, Yamaguchi R, Imoto S, Nagamura-Inoue T, Miyano S, Tojo A. Circulating tumor DNA dynamically predicts response and/or relapse in patients with hematological malignancies. *Int J Hematol*. 108(4): 402-410, 2018.
30. Niida A, Nagayama S, Miyano S, Mimori K. Understanding intratumor heterogeneity by combining genome analysis and mathematical modeling. *Cancer Sci*. 109(4): 884-892, 2018.
31. Ochi Y, Hiramoto N, Yoshizato T, Ono Y, Takeda J, Shiozawa Y, Yoshida K, Kakiuchi N, Shiraishi Y, Tanaka H, Chiba K, Kazuma Y, Tabata S, Yonetani N, Uehara K, Yamashita D, Imai Y, Nagafuji K, Yamakawa M, Miyano S, Takaori-Kondo A, Ogawa S, Ishikawa T. Clonally related diffuse large B-cell lymphoma and interdigitating dendritic cell sarcoma sharing MYC translocation. *Haematologica*. 103(11): e553-e556, 2018.
32. 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*. 183(5): 842-845, 2018.
33. Ono S, Matsuda J, Watanabe E, Akaike H, Teranishi H, Miyata I, Otomo T, Sadahira Y, Mizuochi T, Kusano H, Kage M, Ueno H, Yoshida K, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Ogawa S, Hayashi Y, Kanegane H, Ouchi K. Novel neuroblastoma amplified sequence (NBAS) mutations in a Japanese boy with fever-triggered recurrent acute liver failure. *Hum Genome Var*. 6: 2, 2019.
34. Onodera T, Shibuya T. Succinct Oblivious RAM. *Theoretical Analysis of Computer Science*. 96 (52): 1-16, 2018.
35. Park H, Yamada M, Imoto S, Miyano S. Robust Sample-Specific Stability Selection with Effective Error Control. *J Comput Biol*. 2019 Jan 14. doi: 10.1089/cmb.2018.0180.
36. Park H, Shimamura T, Imoto S, Miyano S. Adaptive NetworkProfiler for Identifying Cancer Characteristic-Specific Gene Regulatory Networks. *J Comput Biol*. 25(2): 130-145, 2018.
37. Saito T, Niida A, Uchi R, Hirata H, Komatsu H, Sakimura S, Hayashi S, Nambara S, Kuroda Y, Ito S, Eguchi H, Masuda T, Sugimachi K, Tobo T, Nishida H, Daa T, Chiba K, Shiraishi Y, Yoshizato T, Kodama M, Okimoto T, Mizukami K, Ogawa R, Okamoto K, Shuto M, Fukuda K, Matsui Y, Shimamura T, Hasegawa T, Doki Y, Nagayama S, Yamada K, Kato M, Shibata T, Mori M, Aburatani H, Murakami K, Suzuki Y, Ogawa S, Miyano S, Mimori K. A temporal shift of the evolutionary principle shaping intratumor heterogeneity in colorectal cancer. *Nat Commun*. 9(1): 2884, 2018.
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*. 32 (3): 839-843, 2018.
39. Shiozawa Y, Malcovati L, Gallì A, Sato-Otsubo A, Kataoka K, Sato Y, Watatani Y, Suzuki H,

Yoshizato T, Yoshida K, Sanada M, Makishima H, Shiraishi Y, Chiba K, Hellström-Lindberg E, Miyano S, Ogawa S, Cazzola M. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. *Nat Commun*. 9(1): 3649, 2018.

40. Shiraishi Y, Kataoka K, Chiba K, Okada A, Kogure Y, Tanaka H, Ogawa S, Miyano S. A comprehensive characterization of cis-acting splicing-associated variants in human cancer. *Genome Res*. 28(8): 1111-1125, 2018.
41. 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). doi: 10.1002/pbc.26831, 2018.
42. Takei T, Yokoyama K, Shimizu E, Konuma T, Takahashi S, Yamaguchi R, Imoto S, Miyano S, Tojo A. Azacitidine effectively reduces TP53-mutant leukemic cell burden in secondary acute myeloid leukemia after cord blood transplantation. *Leuk Lymphoma*. 59(11): 2755-2756, 2018.
43. Toki T, Yoshida K, Wang R, Nakamura S, Maekawa T, Goi K, Katoh MC, Mizuno S, Sugiyama F, Kanezaki R, Uechi T, Nakajima Y, Sato Y, Okuno Y, Sato-Otsubo A, Shiozawa Y, Kataoka K, Shiraishi Y, Sanada M, Chiba K, Tanaka H, Terui K, Sato T, Kamio T, Sakaguchi H, Ohga S, Kuramitsu M, Hamaguchi I, Ohara A, Kanno H, Miyano S, Kojima S, Ishiguro A, Sugita K, Kenmochi N, Takahashi S, Eto K, Ogawa S, Ito E. De Novo Mutations Activating Germline TP53 in an Inherited Bone-Marrow-Failure Syndrome. *Am J Hum Genet*. 103(3): 440-447, 2018.
44. VanderWeele DJ, Finney R, Katayama K, Gillard M, Paner G, Imoto S, Yamaguchi R, Wheeler D, Lack J, Cam M, Pontier A, Nguyen YTM, Maejima K, Sasaki-Oku A, Nakano K, Tanaka H, Vander Griend D, Kubo M, Ratain MJ, Miyano S, Nakagawa H. Genomic Heterogeneity Within Individual Prostate Cancer Foci Impacts Predictive Biomarkers of Targeted Therapy. *Eur Urol Focus*. pii: S2405-4569(18) 30007-5, 2018. doi: 10.1016/j.euf.2018.01.006.
45. Wardell CP, Fujita M, Yamada T, Simbolo M, Fassan M, Karlic R, Polak P, Kim J, Hatanaka Y, Maejima K, Lawlor RT, Nakanishi Y, Mitsuhashi T, Fujimoto A, Furuta M, Ruzzenente A, Conci S, Oosawa A, Sasaki-Oku A, Nakano K, Tanaka H, Yamamoto Y, Michiaki K, Kawakami Y, Aikata H, Ueno M, Hayami S, Gotoh K, Ariizumi SI, Yamamoto M, Yamaue H, Chayama K, Miyano S, Getz G, Scarpa A, Hirano S, Nakamura T, Nakagawa H. Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. *J Hepatol*. 68(5): 959-969, 2018.
46. Yamato G, Shiba N, Yoshida K, Hara Y, Shiraishi Y, Ohki K, Okubo J, Park MJ, Sotomatsu M, Arakawa H, Kiyokawa N, Tomizawa D, Adachi S, Taga T, Horibe K, Miyano S, Ogawa S, Hayashi Y. RUNX1 mutations in pediatric acute myeloid leukemia are associated with distinct genetic features and an inferior prognosis. *Blood*. 131(20): 2266-2270, 2018.
47. Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, Shiozawa Y, Sato Y, Aoki K, Kim SK, Fujii Y, Yoshida K, Kataoka K, Nakagawa MM, Inoue Y, Hirano T, Shiraishi Y, Chiba K, Tanaka H, Sanada M, Nishikawa Y, Amanuma Y, Ohashi S, Aoyama I, Horimatsu T, Miyamoto S, Tsunoda S, Sakai Y, Narahara M, Brown JB, Sato Y, Sawada G, Mimori K, Minamiguchi S, Haga H, Seno H, Miyano S, Makishima H, Muto M, Ogawa S. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature*. 565(7739): 312-317, 2019.
48. Yokoyama K, Shimizu E, Yokoyama N, Nakamura S, Kasajima R, Ogawa M, Takei T, Ito M, Kobayashi A, Yamaguchi R, Imoto S, Miyano S, Tojo A. Cell-lineage level-targeted sequencing to identify acute myeloid leukemia with myelodysplasia-related changes. *Blood Adv*. 2(19): 2513-2521, 2018.

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. Epigenetic landscape influences the liver cancer genome architecture.

Hama N[1], Totoki Y[1], Miura F[2], Tatsuno K[3], Saito-Adachi M[1], Nakamura H[1], Arai Y[1], Hosoda F[1], Urushidate T[4], Ohashi S[1], Mukai W[1], Hiraoka N[5], Aburatani H[3], Ito T[2], Shibata T[6,7]; [1]Division of Cancer Genomics, National Cancer Center Research Institute, Japan. [2]Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Japan. [3]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Japan. [4]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Japan. [5]Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Japan. [6]Division of Cancer Genomics, National Cancer Center Research Institute, Japan. [7]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Japan.

The accumulations of different types of genetic alterations such as nucleotide substitutions, structural rearrangements and viral genome integrations and epigenetic alterations contribute to carcinogenesis. Here, we report correlation between the occurrence of epigenetic features and genetic aberrations by whole-genome bisulfite, whole-genome shotgun, long-read, and virus capture sequencing of 373 liver cancers. Somatic substitutions and rearrangement breakpoints are enriched in tumor-specific hypo-methylated regions with inactive chromatin marks and actively transcribed highly methylated regions in the cancer genome. Individual mutation signatures depend on chromatin status, especially, signatures with a higher transcriptional strand bias occur within active chromatic areas. Hepatitis B virus (HBV) integration sites are frequently detected within inactive chromatin regions in cancer cells, as a consequence of negative selection for integrations in active chromatin regions. Ultra-high structural instability and preserved unmethylation of integrated HBV genomes are observed. We conclude that both precancerous and somatic epigenetic features contribute to the cancer genome architecture.

2. Current state of therapeutic development for rare cancers in Japan, and proposals for improvement.

Kawai A[1,2], Goto T[1,3], Shibata T[1,4], Tani K[1,5], Mizutani S[1,6], Nishikawa A[1,7], Shibata T[1,8], Matsumoto S[1,9], Nagata K[1,10], Narukawa M[1,11], Matsui S[1,12], Ando M[1,13], Toguchida J[1,14], Monden M[1,15], Heike

T[1,16], Kimura S[1,17], Ueda R[1,18]; [1]Subcommittee on Rare Cancers, The Science Board to the Pharmaceuticals and Medical Devices Agency, Tokyo, Japan. [2]Department of Musculoskeletal Oncology and Rehabilitation, Rare Cancer Center, National Cancer Center Hospital, Tokyo, Japan. [3]Program for Drug Discovery and Medical Technology Platforms, RIKEN, Tsukuba, Japan. [4]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [5]Project Division of ALA Advanced Medical Research, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [6]Tokyo Medical and Dental University, Tokyo, Japan. [7]Biological Safety Research Center, National Institute of Health Sciences, Tokyo, Japan. [8]Biostatistics Division, Center for Research Administration and Support, National Cancer Center, Tokyo, Japan. [9]Sarcoma Center, The Cancer Institute Hospital of JFCR, Tokyo, Japan. [10]University of Tsukuba, Tsukuba, Japan. [11]Department of Clinical Medicine (Pharmaceutical Medicine), School of Pharmacy, Kitasato University, Sagamihara, Japan. [12]Department of Biostatistics, Nagoya University Graduate School of Medicine, Nagoya, Japan. [13]Department of Clinical Oncology, Aichi Cancer Center Hospital, Nagakute, Japan. [14]Institute for Frontier Life and Medical Sciences/Center for iPS Cell Research and Application, Kyoto University, Tokyo, Japan. [15]Sakai City Hospital Organization, Sakai, Japan. [16]Hyogo Prefectural Amagasaki General Medical Center, Amagasaki, Japan. [17]Division of Hematology, Respiratory Medicine and Oncology, Department of Internal Medicine, Faculty of Medicine, Saga University, Saga, Japan. [18]Department of Tumor Immunology, Aichi Medical University School of Medicine, Nagakute, Japan.

In order to develop novel clinical options for rare cancers, which tend to remain left out of novel therapeutic development because of their paucity, efficient recruitment of eligible patients, who tend to be widely dispersed across the country and treated at different centers, is necessary. For this purpose, it is important to establish rare cancer registries that are linked with clinical studies, to organize a central pathological diagnosis system and biobanks for rare cancers, and to consolidate patients with rare cancers to facilities that can conduct clinical studies meeting international standards. Establishing an all-Japan cooperative network is essential. Clinical studies of rare cancers have considerable limitations in study design and sample size as a result of paucity of eligible patients and, as a result, the level of confirmation of the efficacy and safety shown by the studies is relatively low. Therefore, measures to alleviate these weaknesses inherent to external conditions need to be explored. It is also important to reform the current research environment in order to develop world-leading treatment for rare cancers, including promotion of basic research, collaboration between industry and academia, and improvement of the infrastructure for clinical studies. Collaboration among a wide range of stakeholders is required to promote the clinical development of treatment for rare cancers under a nationwide consensus.

3. Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors.

Anderson ND[1,2], de Borja R[1], Young MD[3], Fuligni F[1], Rosic A[1], Roberts ND[3], Hajjar S[1], Layeghifard M[1], Novokmet A[1], Kowalski PE[1], Anaka M[1], Davidson S[4], Zarrei M[5], Id Said B[1], Schreiner LC[1], Marchand R[1], Sitter J[1], Gokgoz N[6], Brunga L[1], Graham GT[7], Fullam A[3], Pillay N[8,9], Toretsky JA[7], Yoshida A[10], Shibata T[11,12], Metzler M[13], Somers GR[2,14], Scherer SW[1,5,15,16], Flanagan AM[9,10], Campbell PJ[3,17], Schiffman JD[18], Shago M[2,4], Alexandrov LB[19], Wunder JS[20,21], Andrulis IL[6,15], Malkin D[22,23,24], Behjati S[25,26], Shlien A[22,2,4]; [1]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. [2]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. [3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK. [4]Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada. [5]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada. [6]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada. [7]Department of Oncology and Pediatrics, Georgetown University, Washington, DC, USA. [8]University College London Cancer Institute, Huntley Street, London, UK. [9]Histopathology, Royal National Orthopaedic Hospital NHS Trust, Stanmore, Middlesex, UK. [10]Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan. [11]Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan. [12]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Sciences, The University of Tokyo, Japan. [13]Department of Pediatrics and Adolescent Medicine, University Hospital Erlangen, Erlangen, Germany. [14]Department of Pathology, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. [15]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [16]The McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada. [17]Department of Haematology, University of Cambridge, Cambridge, UK. [18]Departments of Pediatrics and Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA. [19]Department of Cellular and Molecular

Medicine and Department of Bioengineering and Moores Cancer Center, University of California, La Jolla, San Diego, CA, USA. [20]University Musculoskeletal Oncology Unit, Mount Sinai Hospital, Toronto, Ontario, Canada. [21]Division of Orthopaedic Surgery, Department of Surgery, University of Toronto, Toronto, Ontario, Canada. [22]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. [23]Division of Hematology-Oncology, The Hospital for Sick Children, Toronto, Ontario, Canada. [24]Department of Pediatrics, University of Toronto, Ontario, Canada. [25]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK. [26]Department of Paediatrics, University of Cambridge, Cambridge, UK.

Sarcomas are cancers of the bone and soft tissue often defined by gene fusions. Ewing sarcoma involves fusions between EWSR1, a gene encoding an RNA binding protein, and E26 transformation-specific (ETS) transcription factors. We explored how and when EWSR1-ETS fusions arise by studying the whole genomes of Ewing sarcomas. In 52 of 124 (42%) of tumors, the fusion gene arises by a sudden burst of complex, loop-like rearrangements, a process called chromoplexy, rather than by simple reciprocal translocations. These loops always contained the disease-defining fusion at the center, but they disrupted multiple additional genes. The loops occurred preferentially in early replicating and transcriptionally active genomic regions. Similar loops forming canonical fusions were found in three other sarcoma types. Chromoplexy-generated fusions appear to be associated with an aggressive form of Ewing sarcoma. These loops arise early, giving rise to both primary and relapse Ewing sarcoma tumors, which can continue to evolve in parallel.

4. Changes in immune-related gene expression profiles of gastric cancer patients before and after chemotherapy.

Satoshi Yamasaki[1], Tomoko Urushidate[1], Tatsuhiro Shibata[1,2]: [1]Laboratory of Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo. [2]Division of Cancer Genomics, National Cancer Center Research Institute.

Programmed death 1(PD1) and PDL1, one of ligands of PD1, are well known molecules as regulators of immune response system. Tumor cells can escape from immune response using such immune checkpoint pathways, which inhibit proliferation or activation of antigen specific T cells. Many studies have shown increased expression of PD-L1 in various types of cancers, but the factor which induces such increase is still unclear. Identification of precursive change will be useful for immunotherapy. In this study we performed transcriptome analysis of 90 gastric cancer patients before and after chemotherapy, and investigated changes in immune related gene expression profile. We found differences in gene expression profile of checkpoints, interleukins and cytokines in patients, but no correlation with clinical status of patient was observed. We will further explore molecular mechanisms underlising increased expression of immunregulators in cancers.

5. Prediction of RNA tertiary structures and RNA-RNA/Protein interactions.

Satoshi Yamasaki[1,2], Takayuki Amemiya[1], Yukimitsu Yabuki[1,3], Katsuhisa Horimoto[1] ,Kazuhiko Fukui[1]: [1]Molecular Profiling for Drug Discovery Research Center (molprof), National Institute of Advanced Industrial Science and Technology (AIST). [2]The Institute of Medical Science, University of Tokyo (IMSUT). [3]IMSBIO Co., Ltd.

Recent progress in molecular biology has revealed that many non-coding RNAs regulate gene expression or catalyze biochemical reactions in tumors, viruses and other diseases. The tertiary structures of RNA molecules and RNA-RNA/protein interaction sites are of increasing importance as potential targets for new therapies that could treat a broad array of human diseases. Current RNA drugs are split into two groups: antisense RNA molecules and aptamers. In this study, we present a novel workflow to predict RNA tertiary structures and RNA-RNA/protein interactions using the KNIME environment, which enables us to assemble a combination of RNA-related analytical tools and databases. In this workflow, three analytical workflows for comprehensive structural analysis of RNA are available: (1) prediction of the tertiary structure of RNA; (2) prediction of the structures of RNA-RNA complexes and analysis of their interactions; and (3) prediction of the structures of RNA-protein complexes and analysis of their interactions. We demonstrated that the tertiary structure prediction of several RNA aptamer drugs, and performed docking calculations of the aptamer and its target proteins using a fragment of the interaction site of the aptamer. The affinity of aptamer-protein complex was evaluated using MMGB/SA method. The results provide valuable information for designing novel features of aptamer-protein complexes. This work is a collaborative research with Molecular Profiling for Drug Discovery Research Center, AIST.

Publications

1. Hama N, Totoki Y, Miura F, Tatsuno K, Saito-Adachi M, Nakamura H, Arai Y, Hosoda F, Urushidate T, Ohashi S, Mukai W, Hiraoka N, Aburatani H, Ito T, Shibata T; Epigenetic landscape influences the liver cancer genome architecture. Nat Commun. 9: 1643, 2018.
2. Anderson ND, de Borja R, Young MD, Fuligni F, Rosic A, Roberts ND, Hajjar S, Layeghifard M, Novokmet A, Kowalski PE, Anaka M, Davidson S, Zarrei M, Id Said B, Schreiner LC, Marchand R, Sitter J, Gokgoz N, Brunga L, Graham GT, Fullam A, Pillay N, Toretsky JA, Yoshida A, Shibata T, Metzler M, Somers GR, Scherer SW, Flanagan AM, Campbell PJ, Schiffman JD, Shago M, Alexandrov LB, Wunder JS, Andrulis IL, Malkin D, Behjati S, Shlien A; Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. Science. 361 (6405), 2018.
3. Kawai A, Goto T, Shibata T, Tani K, Mizutani S, Nishikawa A, Shibata T, Matsumoto S, Nagata K, Narukawa M, Matsui S, Ando M, Toguchida J, Monden M, Heike T, Kimura S, Ueda R; Current state of therapeutic development for rare cancers in Japan, and proposals for improvement. Cancer Sci. 109: 1731-1737, 2018.

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	
Project Senior Assistant Professor	Makoto Hirata, M.D., Ph.D.	特任講師	博士（医学）	平　田	真
Assistant Professor	Chizu Tanikawa, Ph.D.	助　教	博士（医学）	谷　川　千　津	

The major goal of our group is to identify genes of medical importance, and to develop new diagnostic and therapeutic tools. We have been attempting to isolate genes involving in carcinogenesis and also those causing or predisposing to various diseases as well as those related to drug efficacies and adverse reactions. By means of technologies developed through the genome project including a high-resolution SNP map, a large-scale DNA sequencing, and the cDNA microarray method, we have isolated a number of biologically and/or medically important genes, and are developing novel diagnostic and therapeutic tools.

1. Genome-wide association study

Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases.

Clinical measurements can be viewed as useful intermediate phenotypes to promote understanding of complex human diseases. To acquire comprehensive insights into the underlying genetics, here we conducted a genome-wide association study (GWAS) of 58 quantitative traits in 162,255 Japanese individuals. Overall, we identified 1,407 trait-associated loci ($P<5.0\times10-8$), 679 of which were novel. By incorporating 32 additional GWAS results for complex diseases and traits in Japanese individuals, we further highlighted pleiotropy, genetic correlations, and cell-type specificity across quantitative traits and diseases, which substantially expands the current understanding of the associated genetics and biology. This study identified both shared polygenic effects and cell-type specificity, represented by the genetic links among clinical measurements, complex diseases, and relevant cell types. Our findings demonstrate that even without prior biological knowledge of cross-phenotype relationships, genetics corresponding to clinical measurements successfully recapture those measurements' relevance to diseases, and thus can contribute to the elucidation of unknown etiology and pathogenesis.

GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12.

Colorectal cancer (CRC) is the fourth leading cause of cancer mortality worldwide. Genome-wide association studies (GWAS) identified more than 50 CRC loci. However, most of the previous studies were conducted in European population, and host genetic factors among Japanese population are largely remained to be identified. To identify novel loci in the Japanese population, here, we performed a large-scale GWAS using 6692 cases and 27 178 controls followed by a replication analysis using more than 11000 case-control samples. We found the significant association of 10 loci ($P<5\times10-8$), including 2 novel loci on 16q24.1 (IRF8-FOXF1, rs

847208, P = 3.15 × 10 − 9 and odds ratio = 1.107 with 95% confidence interval (CI) of 1.071-1.145) and 20q 13.12 (TOX2, rs6065668, P = 4.47 × 10 − 11 and odds ratio = 0.897 with 95% CI of 0.868-0.926). Moreover, 35 previously reported single nucleotide polymorphisms (SNPs) in 24 regions were validated in the Japanese population (P<0.05) with the same risk allele as in the previous studies. SNP rs6065668 was significantly associated with TOX2 expression in the sigmoid colon. In addition, nucleotide substitutions in the regulatory region of TOX2 were predicted to alter the binding of several transcription factors, including KLF5. Our findings elucidate the important role of genetic variations in the development of CRC in the Japanese population.

Large-scale Genome-wide Association Study of East Asians Identifies Loci Associated With Risk for Colorectal Cancer.

BACKGROUND&AIMS: Genome-wide association studies (GWASs) have associated approximately 50 loci with risk of colorectal cancer (CRC)-nearly one-third of these loci were initially associated with CRC in studies conducted in East Asian populations. We conducted a GWAS of East Asians to identify CRC risk loci and evaluate the generalizability of findings from GWAS of European populations to Asian populations METHODS: We analyzed genetic data from 22,775 patients with CRC (cases) and 47,731 individuals without cancer (controls) from 14 studies in the Asia Colorectal Cancer Consortium. First, we performed a meta-analysis of 7 GWAS (10,625 cases and 34,595 controls) and identified 46,554 promising risk variants for replication by adding them to the Multi-Ethnic Global Array (MEGA) for genotype analysis in 6445 cases and 7175 controls. These data were analyzed, along with data from additional 5705 cases and 5961 controls genotyped using the ONCOARRAY: We also obtained data from 57,976 cases and 67,242 controls of European descent. Variants at identified risk loci were functionally annotated and evaluated in correlation with gene expression levels.
RESULTS: A meta-analyses of all samples from people of Asian descent identified 13 loci and 1 new variant at a known locus (10q24.2) associated with risk of CRC at the genome-wide significance level of P<5 × 10 − 8. We did not perform experiments to replicate these associations in additional individuals of Asian ancestry. However, the lead risk variant in 6 of these loci was also significantly associated with risk of CRC in European descendants. A strong association (44%-75% increase in risk per allele) was found for 2 low-frequency variants: rs201395236 at 1q44 (minor allele frequency, 1.34%) and rs77969132 at 12p11.21 (minor allele frequency, 1.53%). For 8 of the 13 associated loci, the variants with the highest levels of significant association were located inside or near the protein-coding genes L1TD1, EFCAB2, PPP1R21, SLCO2A1, HLA-G, NOTCH4, DENND5B, and GNAS. For other inter-genic loci, we provided evidence for the possible involvement of the genes ALDH7A1, PRICKLE1, KLF5, WWOX, and GLP2R. We replicated findings for 41 of 52 previously reported risk loci.
CONCLUSIONS: We showed that most of the risk loci previously associated with CRC risk in individuals of European descent were also associated with CRC risk in East Asians. Furthermore, we identified 13 loci significantly associated with risk for CRC in Asians. Many of these loci contained genes that regulate the immune response, Wnt signaling to beta-catenin, prostaglandin E2 catabolism, and cell pluripotency and proliferation. Further analyses of these genes and their variants is warranted-particularly for the 8 loci for which the lead CRC risk variants were not replicated in persons of European descent.

Genome-wide association study (GWAS) of ovarian cancer in Japanese predicted regulatory variants in 22q13.1.

Genome-wide association studies (GWAS) have identified greater than 30 variants associated with ovarian cancer, but most of these variants were investigated in European populations. Here, we integrated GWAS and subsequent functional analyses to identify the genetic variants with potential regulatory effects. We conducted GWAS for ovarian cancer using 681 Japanese cases and 17,492 controls and found that rs137672 on 22q13.1 exhibited a strong association with a P-value of 1.05 × 10 − 7 and an odds ratio of 0.573 with a 95% confidence interval of 0.466-0.703. In addition, three previously reported SNPs, i.e., rs10088218, rs9870207 and rs 1400482, were validated in the Japanese population (P<0.05) with the same risk allele as noted in previous studies. Functional studies including regulatory feature analysis and electrophoretic mobility shift assay (EMSA) revealed two regulatory SNPs in 22q 13.1, rs2072872 and rs6509, that affect the binding affinity to some nuclear proteins in ovarian cancer cells. The plausible regulatory proteins whose motifs could be affected by the allele changes of these two SNPs were also proposed. Moreover, the protective G allele of rs6509 was associated with a decreased SYNGR1 expression level in normal ovarian tissues. Our findings elucidated the regulatory variants in 22q13.1 that are associated with ovarian cancer risk.

Genome-wide association study identifies gastric cancer susceptibility loci at 12q24.11-12 and 20q11.21.

Gastric cancer is the third leading cause of cancer mortality in Japan and worldwide. Although previous studies identify various genetic variations associated with gastric cancer, host genetic factors are largely unidentified. To identify novel gastric cancer loci in the Japanese population, herein, we carried out a large-scale genome-wide association study using 6171 cases and 27 178 controls followed by three replication analyses. Analysis using a total of 11 507 cases and 38 904 controls identified two novel loci on 12q24.11-12 (rs6490061, $P = 3.20 \times 10-8$ with an odds ratio [OR] of 0.905) and 20q 11.21 (rs2376549, $P = 8.11 \times 10-10$ with an OR of 1.109). rs6490061 is located at intron 19 of the CUX2 gene, and its expression was suppressed by Helicobacter pylori infection. rs2376549 is included within the gene cluster of DEFB families that encode antibacterial peptides. We also found a significant association of rs7849280 in the ABO gene locus on 9q 34.2 ($P = 2.64 \times 10-13$ with an OR of 1.148). CUX2 and ABO expression in gastric mucosal tissues was significantly associated with rs6490061 and rs 7849280 ($P = 0.0153$ and $8.00 \times 10-11$), respectively. Our findings show the crucial roles of genetic variations in the pathogenesis of gastric cancer.

2. Genetic analysis of various diseases

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.

Single Nucleotide Polymorphisms of HAAO and IRX6 Genes as Risk Factors for Hypospadias.

PURPOSE: We evaluated the association of hypospadias and 17 susceptibility loci previously identified by a European genome-wide association study in a cohort of Japanese patients. We also examined the expression of candidate genes in male mouse embryos to determine the possible underlying mechanisms of this disease.
MATERIALS AND METHODS: We enrolled 169 Japanese patients (mean age at surgery 3.7 years) who underwent repair of hypospadias. Genotyping of 17 single nucleotide polymorphisms was performed using a multiplex polymerase chain reaction invader assay. We also performed in situ hybridization to determine whether candidate genes were expressed in the male genital tubercle during embryonic development of the external genitalia in mice.
RESULTS: Single nucleotide polymorphism rs 3816183 of HAAO was significantly associated with susceptibility to hypospadias in general ($p = 0.0019$) and to anterior/middle hypospadias ($p = 0.0283$) and posterior hypospadias ($p = 0.0226$), while single nucleotide polymorphism rs6499755 of IRX6 showed an association with susceptibility to anterior/middle hypospadias ($p = 0.0472$). In mouse embryos there was no significant upregulation of Haao expression in the developing male external genitalia. Irx3 and Irx5, which are linked to Irx6 within the IrxB cluster, were expressed in the mesenchyme remote from the urethral plate epithelium during the critical embryonic period for masculinization. Irx6 was expressed in the ectodermal epithelium, demonstrating prominent dorsal ectodermal expression without expression in the ventral ectoderm adjacent to the urethral plate during the same period.
CONCLUSIONS: Genetic variations of HAAO and IRX6 influence susceptibility to hypospadias in the Japanese population. Further research is needed to clarify the mechanism by which variations in these genes contribute to the pathogenesis of hypospadias.

Decrease in PSCA expression caused by Helicobacter pylori infection may promote progression to severe gastritis.

SNP rs2294008 in Prostate Stem Cell Antigen (PSCA) and decreased PSCA expression are associated with gastric cancer. The objective of this study is to investigate the role of rs2294008 and PSCA expression in the gastritis-gastric cancer carcinogenic pathway. We conducted a case-control association study of H. pylori-infected gastritis and gastric cancer. rs2294008 was associated with the progression to chronic active gastritis ($P = 9.4 \times 10-5$; odds ratio $= 3.88$, TT + TC vs CC genotype), but not with H. pylori infection per se nor with the progression from active gastritis to gastric cancer. We also assessed the association of rs2294008 with PSCA

mRNA expression in the gastric mucosa at various disease stages and found that rs2294008 was associated with PSCA expression ($P = 1.3 \times 10 - 12$). H. pylori infection ($P = 5.1 \times 10 - 8$) and eradication therapy ($P<1 \times 10 - 11$) resulted in the reduced and increased PSCA expression, respectively, indicating negative regulation of PSCA expression by H. pylori infection. PSCA expression was decreased in severe gastritis compared with mild gastritis only among T allele carriers. Our findings revealed the regulation of PSCA expression by host genetic variation and bacterial infection might contribute to gastritis progression after H. pylori infection.

3. Genes playing significant roles in human cancers

Frequent mutations of genes encoding vacuolar H+ -ATPase components in granular cell tumors.

Granular cell tumors (GCTs) are rare mesenchymal tumors that exhibit a characteristic morphology and a finely granular cytoplasm. The genetic alterations responsible for GCT tumorigenesis had been unknown until recently, when loss-of-function mutations of ATP6AP1 and ATP6AP2 were described. Thus, we performed whole-exome sequencing, RNA sequencing, and targeted sequencing of 51 GCT samples. From these genomic analyses, we identified mutations in genes encoding vacuolar H+ -ATPase (V-ATPase) components, including ATP6AP1 and ATP6AP2, in 33 (65%) GCTs. ATP6 AP1 and ATP6AP2 mutations were found in 23 (45 %) and 2 (4%) samples, respectively, and all were truncating or splice site mutations. In addition, seven other genes encoding V-ATPase components were also mutated, and three mutations in ATP6V0C occurred on the same amino acid (isoleucine 136). These V-ATPase component gene mutations were mutually exclusive, with one exception. These results suggest that V-ATPase function is impaired in GCTs not only by loss-of-function mutations of ATP6AP1 and ATP6AP2 but also through mutations of other subunits. Our findings provide additional support for the hypothesis that V-ATPase dysfunction promotes GCT tumorigenesis.

Publications

[1] O. Toyoshima, C. Tanikawa, R. Yamamoto, H. Watanabe, H. Yamashita, K. Sakitani, S. Yoshida, M. Kubo, K. Matsuo, H. Ito, K. Koike, Y. Seto, K. Matsuda, Decrease in PSCA expression caused by Helicobacter pylori infection may promote progression to severe gastritis, Oncotarget, 9 (2018) 3936-3945.

[2] M. Kanai, M. Akiyama, A. Takahashi, N. Matoba, Y. Momozawa, M. Ikeda, N. Iwata, S. Ikegawa, M. Hirata, K. Matsuda, M. Kubo, Y. Okada, Y. Kamatani, Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases, Nat Genet, (2018).

[3] C. Tanikawa, K. Ueda, A. Suzuki, A. Iida, R. Nakamura, N. Atsuta, G. Tohnai, G. Sobue, N. Saichi, Y. Momozawa, Y. Kamatani, M. Kubo, K. Yamamoto, Y. Nakamura, K. Matsuda, Citrullination of RGG Motifs in FET Proteins by PAD4 Regulates Protein Aggregation and ALS Susceptibility, Cell Rep, 22 (2018) 1473-1483.

[4] C. Tanikawa, Y. Kamatani, A. Takahashi, Y. Momozawa, K. Leveque, S. Nagayama, K. Mimori, M. Mori, H. Ishii, J. Inazawa, J. Yasuda, A. Tsuboi, A. Shimizu, M. Sasaki, T. Yamaji, N. Sawada, M. Iwasaki, S. Tsugane, M. Naito, K. Wakai, T. Koyama, T. Takezaki, K. Yuji, Y. Murakami, Y. Nakamura, M. Kubo, K. Matsuda, GWAS Identifies Two Novel Colorectal Cancer Loci at 16q24.1 and 20q13.12, Carcinogenesis, (2018).

[5] Y. Shiga, M. Akiyama, K.M. Nishiguchi, K. Sato, N. Shimozawa, A. Takahashi, Y. Momozawa, M. Hirata, K. Matsuda, T. Yamaji, M. Iwasaki, S. Tsugane, I. Oze, H. Mikami, M. Naito, K. Wakai, M. Yoshikawa, M. Miyake, K. Yamashiro, G. Japan Glaucoma Society Omics, K. Kashiwagi, T. Iwata, F. Mabuchi, M. Takamoto, M. Ozaki, K. Kawase, M. Aihara, M. Araie, T. Yamamoto, Y. Kiuchi, M. Nakamura, Y. Ikeda, K.H. Sonoda, T. Ishibashi, K. Nitta, A. Iwase, S. Shirato, Y. Oka, M. Satoh, M. Sasaki, N. Fuse, Y. Suzuki, C.Y. Cheng, C. C. Khor, M. Baskaran, S. Perera, T. Aung, E.N. Vithana, J.N. Cooke Bailey, J.H. Kang, L.R. Pasquale, J.L. Haines, N. Consortium, J.L. Wiggs, K.P. Burdon, P. Gharahkhani, A.W. Hewitt, D.A. Mackey, S. MacGregor, J.E. Craig, R.R. Allingham, M. Hauser, A. Ashaye, D.L. Budenz, S. Akafo, S.E.I. Williams, Y. Kamatani, T. Nakazawa, M. Kubo, Genome-wide association study identifies seven novel susceptibility loci for primary open-angle glaucoma, Hum Mol Genet, 27 (2018) 1486-1496.

[6] Y. Okada, Y. Momozawa, S. Sakaue, M. Kanai, K. Ishigaki, M. Akiyama, T. Kishikawa, Y. Arai, T. Sasaki, K. Kosaki, M. Suematsu, K. Matsuda, K. Yamamoto, M. Kubo, N. Hirose, Y. Kamatani, Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese, Nature

communications, 9 (2018) 1631.
[7] R. Malik, G. Chauhan, M. Traylor, M. Sargurupremraj, Y. Okada, A. Mishra, L. Rutten-Jacobs, A.K. Giese, S.W. van der Laan, S. Gretarsdottir, C.D. Anderson, M. Chong, H.H.H. Adams, T. Ago, P. Almgren, P. Amouyel, H. Ay, T.M. Bartz, O.R. Benavente, S. Bevan, G.B. Boncoraglio, R.D. Brown, Jr., A.S. Butterworth, C. Carrera, C.L. Carty, D.I. Chasman, W.M. Chen, J.W. Cole, A. Correa, I. Cotlarciuc, C. Cruchaga, J. Danesh, P.I.W. de Bakker, A.L. DeStefano, M. den Hoed, Q. Duan, S.T. Engelter, G.J. Falcone, R.F. Gottesman, R.P. Grewal, V. Gudnason, S. Gustafsson, J. Haessler, T.B. Harris, A. Hassan, A.S. Havulinna, S.R. Heckbert, E.G. Holliday, G. Howard, F.C. Hsu, H.I. Hyacinth, M.A. Ikram, E. Ingelsson, M.R. Irvin, X. Jian, J. Jimenez-Conde, J.A. Johnson, J.W. Jukema, M. Kanai, K.L. Keene, B.M. Kissela, D. O. Kleindorfer, C. Kooperberg, M. Kubo, L.A. Lange, C.D. Langefeld, C. Langenberg, L.J. Launer, J.M. Lee, R. Lemmens, D. Leys, C.M. Lewis, W.Y. Lin, A.G. Lindgren, E. Lorentzen, P.K. Magnusson, J. Maguire, A. Manichaikul, P.F. McArdle, J.F. Meschia, B.D. Mitchell, T.H. Mosley, M.A. Nalls, T. Ninomiya, M.J. O'Donnell, B.M. Psaty, S.L. Pulit, K. Rannikmae, A.P. Reiner, K.M. Rexrode, K. Rice, S.S. Rich, P.M. Ridker, N.S. Rost, P.M. Rothwell, J.I. Rotter, T. Rundek, R.L. Sacco, S. Sakaue, M.M. Sale, V. Salomaa, B.R. Sapkota, R. Schmidt, C.O. Schmidt, U. Schminke, P. Sharma, A. Slowik, C.L.M. Sudlow, C. Tanislav, T. Tatlisumak, K. D. Taylor, V.N.S. Thijs, G. Thorleifsson, U. Thorsteinsdottir, S. Tiedt, S. Trompet, C. Tzourio, C.M. van Duijn, M. Walters, N.J. Wareham, S. Wassertheil-Smoller, J.G. Wilson, K.L. Wiggins, Q. Yang, S. Yusuf, J.C. Bis, T. Pastinen, A. Ruusalepp, E.E. Schadt, S. Koplev, J.L.M. Bjorkegren, V. Codoni, M. Civelek, N.L. Smith, D.A. Tregouet, I.E. Christophersen, C. Roselli, S.A. Lubitz, P.T. Ellinor, E.S. Tai, J.S. Kooner, N. Kato, J. He, P. van der Harst, P. Elliott, J.C. Chambers, F. Takeuchi, A.D. Johnson, D.K. Sanghera, O. Melander, C. Jern, D. Strbian, I. Fernandez-Cadenas, W.T. Longstreth, Jr., A. Rolfs, J. Hata, D. Woo, J. Rosand, G. Pare, J.C. Hopewell, D. Saleheen, K. Stefansson, B.B. Worrall, S.J. Kittner, S. Seshadri, M. Fornage, H.S. Markus, J.M.M. Howson, Y. Kamatani, S. Debette, M. Dichgans, R. Malik, G. Chauhan, M. Traylor, M. Sargurupremraj, Y. Okada, A. Mishra, L. Rutten-Jacobs, A.K. Giese, S.W. van der Laan, S. Gretarsdottir, C.D. Anderson, M. Chong, H.H.H. Adams, T. Ago, P. Almgren, P. Amouyel, H. Ay, T.M. Bartz, O. R. Benavente, S. Bevan, G.B. Boncoraglio, R.D. Brown, Jr., A.S. Butterworth, C. Carrera, C.L. Carty, D.I. Chasman, W.M. Chen, J.W. Cole, A. Correa, I. Cotlarciuc, C. Cruchaga, J. Danesh, P.I.W. de Bakker, A.L. DeStefano, M.D. Hoed, Q. Duan, S.T. Engelter, G.J. Falcone, R.F. Gottesman, R.P. Grewal, V. Gudnason, S. Gustafsson, J. Haessler, T.B. Harris, A. Hassan, A. S. Havulinna, S.R. Heckbert, E.G. Holliday, G. Howard, F.C. Hsu, H.I. Hyacinth, M.A. Ikram, E. Ingelsson, M.R. Irvin, X. Jian, J. Jimenez-Conde, J.A. Johnson, J.W. Jukema, M. Kanai, K. L. Keene, B.M. Kissela, D.O. Kleindorfer, C. Kooperberg, M. Kubo, L.A. Lange, C.D. Langefeld, C. Langenberg, L.J. Launer, J.M. Lee, R. Lemmens, D. Leys, C.M. Lewis, W.Y. Lin, A.G. Lindgren, E. Lorentzen, P.K. Magnusson, J. Maguire, A. Manichaikul, P.F. McArdle, J.F. Meschia, B.D. Mitchell, T.H. Mosley, M.A. Nalls, T. Ninomiya, M.J. O'Donnell, B.M. Psaty, S.L. Pulit, K. Rannikmae, A.P. Reiner, K. M. Rexrode, K. Rice, S.S. Rich, P.M. Ridker, N. S. Rost, P.M. Rothwell, J.I. Rotter, T. Rundek, R.L. Sacco, S. Sakaue, M.M. Sale, V. Salomaa, B.R. Sapkota, R. Schmidt, C.O. Schmidt, U. Schminke, P. Sharma, A. Slowik, C.L.M. Sudlow, C. Tanislav, T. Tatlisumak, K.D. Taylor, V.N.S. Thijs, G. Thorleifsson, U. Thorsteinsdottir, S. Tiedt, S. Trompet, C. Tzourio, C.M. van Duijn, M. Walters, N.J. Wareham, S. Wassertheil-Smoller, J.G. Wilson, K.L. Wiggins, Q. Yang, S. Yusuf, N. Amin, H.S. Aparicio, D.K. Arnett, J. Attia, A.S. Beiser, C. Berr, J.E. Buring, M. Bustamante, V. Caso, Y.C. Cheng, S.H. Choi, A. Chowhan, N. Cullell, J.F. Dartigues, H. Delavaran, P. Delgado, M. Dorr, G. Engstrom, I. Ford, W.S. Gurpreet, A. Hamsten, L. Heitsch, A. Hozawa, L. Ibanez, A. Ilinca, M. Ingelsson, M. Iwasaki, R.D. Jackson, K. Jood, P. Jousilahti, S. Kaffashian, L. Kalra, M. Kamouchi, T. Kitazono, O. Kjartansson, M. Kloss, P.J. Koudstaal, J. Krupinski, D.L. Labovitz, C.C. Laurie, C.R. Levi, L. Li, L. Lind, C.M. Lindgren, V. Lioutas, Y.M. Liu, O.L. Lopez, H. Makoto, N. Martinez-Majander, K. Matsuda, N. Minegishi, J. Montaner, A.P. Morris, E. Muino, M. Muller-Nurasyid, B. Norrving, S. Ogishima, E.A. Parati, L.R. Peddareddygari, N. L. Pedersen, J. Pera, M. Perola, A. Pezzini, S. Pileggi, R. Rabionet, I. Riba-Llena, M. Ribases, J.R. Romero, J. Roquer, A.G. Rudd, A.P. Sarin, R. Sarju, C. Sarnowski, M. Sasaki, C.L. Satizabal, M. Satoh, N. Sattar, N. Sawada, G. Sibolt, A. Sigurdsson, A. Smith, K. Sobue, C. Soriano-Tarraga, T. Stanne, O.C. Stine, D.J. Stott, K. Strauch, T. Takai, H. Tanaka, K. Tanno, A. Teumer, L. Tomppo, N.P. Torres-Aguila, E. Touze, S. Tsugane, A.G. Uitterlinden, E.M. Valdimarsson, S.J. van der Lee, H. Volzke, K. Wakai, D. Weir, S.R. Williams, C.D.A. Wolfe, Q. Wong, H. Xu, T. Yamaji, D.K. Sanghera, O. Melander, C. Jern, D. Strbian, I. Fernandez-Ca-

denas, W.T. Longstreth, Jr., A. Rolfs, J. Hata, D. Woo, J. Rosand, G. Pare, J.C. Hopewell, D. Saleheen, K. Stefansson, B.B. Worrall, S.J. Kittner, S. Seshadri, M. Fornage, H.S. Markus, J.M.M. Howson, Y. Kamatani, S. Debette, M. Dichgans, A.F. Consortium, H. Cohorts for, C. Aging Research in Genomic Epidemiology, C. International Genomics of Blood Pressure, I. Consortium, Starnet, G. BioBank Japan Cooperative Hospital, C. Consortium, E.-C. Consortium, E.P.-I. Consortium, C. International Stroke Genetics, M. Consortium, C.C. Neurology Working Group of the, N.S.G. Network, U.K.Y.L.D. Study, M. Consortium, M. Consortium, Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes, Nat Genet, 50 (2018) 524-537.

[8] S.L. Schmit, C.K. Edlund, F.R. Schumacher, J. Gong, T.A. Harrison, J.R. Huyghe, C. Qu, M. Melas, D.J. Van Den Berg, H. Wang, S. Tring, S.J. Plummer, D. Albanes, M.H. Alonso, C.I. Amos, K. Anton, A.K. Aragaki, V. Arndt, E.L. Barry, S.I. Berndt, S. Bezieau, S. Bien, A. Bloomer, J. Boehm, M.C. Boutron-Ruault, H. Brenner, S. Brezina, D.D. Buchanan, K. Butterbach, B.J. Caan, P.T. Campbell, C.S. Carlson, J. E. Castelao, A.T. Chan, J. Chang-Claude, S.J. Chanock, I. Cheng, Y.W. Cheng, L.S. Chin, J.M. Church, T. Church, G.A. Coetzee, M. Cotterchio, M. Cruz Correa, K.R. Curtis, D. Duggan, D.F. Easton, D. English, E.J.M. Feskens, R. Fischer, L.M. FitzGerald, B.K. Fortini, L.G. Fritsche, C.S. Fuchs, M. Gago-Dominguez, M. Gala, S.J. Gallinger, W.J. Gauderman, G.G. Giles, E.L. Giovannucci, S.M. Gogarten, C. Gonzalez-Villalpando, E.M. Gonzalez-Villalpando, W.M. Grady, J.K. Greenson, A. Gsur, M. Gunter, C.A. Haiman, J. Hampe, S. Harlid, J.F. Harju, R.B. Hayes, P. Hofer, M. Hoffmeister, J.L. Hopper, S.C. Huang, J.M. Huerta, T.J. Hudson, D.J. Hunter, G.E. Idos, M. Iwasaki, R. D. Jackson, E.J. Jacobs, S.H. Jee, M.A. Jenkins, W.H. Jia, S. Jiao, A.D. Joshi, L.N. Kolonel, S. Kono, C. Kooperberg, V. Krogh, T. Kuehn, S. Kury, A. LaCroix, C.A. Laurie, F. Lejbkowicz, M. Lemire, H.J. Lenz, D. Levine, C.I. Li, L. Li, W. Lieb, Y. Lin, N.M. Lindor, Y.R. Liu, F. Loupakis, Y. Lu, F. Luh, J. Ma, C. Mancao, F.J. Manion, S.D. Markowitz, V. Martin, K. Matsuda, K. Matsuo, K.J. McDonnell, C.E. McNeil, R. Milne, A.J. Molina, B. Mukherjee, N. Murphy, P.A. Newcomb, K. Offit, H. Omichessan, D. Palli, J.P.P. Cotore, J. Perez-Mayoral, P.D. Pharoah, J.D. Potter, C. Qu, L. Raskin, G. Rennert, H.S. Rennert, B.M. Riggs, C. Schafmayer, R.E. Schoen, T.A. Sellers, D. Seminara, G. Severi, W. Shi, D. Shibata, X.O. Shu, E.M. Siegel, M.L. Slattery, M. Southey, Z. K. Stadler, M.C. Stern, S. Stintzing, D. Taverna, S.N. Thibodeau, D.C. Thomas, A. Trichopoulou, S. Tsugane, C.M. Ulrich, F.J.B. van Duijnhoven, B. van Guelpan, J. Vijai, J. Virtamo, S.J. Weinstein, E. White, A.K. Win, A. Wolk, M. Woods, A.H. Wu, K. Wu, Y.B. Xiang, Y. Yen, B.W. Zanke, Y.X. Zeng, B. Zhang, N. Zubair, S.S. Kweon, J.C. Figueiredo, W. Zheng, L.L. Marchand, A. Lindblom, V. Moreno, U. Peters, G. Casey, L. Hsu, D.V. Conti, S.B. Gruber, Novel Common Genetic Susceptibility Loci for Colorectal Cancer, J Natl Cancer Inst, (2018).

[9] M. Miyazaki, R. Otomo, Y. Matsushima-Hibiya, H. Suzuki, A. Nakajima, N. Abe, A. Tomiyama, K. Ichimura, K. Matsuda, T. Watanabe, T. Ochiya, H. Nakagama, R. Sakai, M. Enari, The p53 activator overcomes resistance to ALK inhibitors by regulating p53-target selectivity in ALK-driven neuroblastomas, Cell Death Discov, 4 (2018) 56.

[10] Y. Kojima, T. Koguchi, K. Mizuno, Y. Sato, S. Hoshi, J. Hata, H. Nishio, D. Hashimoto, S. Matsushita, K. Suzuki, S. Miyagawa, C.C. Hui, C. Tanikawa, Y. Murakami, G. Yamada, Y. Hayashi, K. Matsuda, Single nucleotide polymorphisms of HAAO and IRX 6 genes as risk factors for hypospadias, J Urol, (2018).

[11] Y. Momozawa, Y. Iwasaki, M.T. Parsons, Y. Kamatani, A. Takahashi, C. Tamura, T. Katagiri, T. Yoshida, S. Nakamura, K. Sugano, Y. Miki, M. Hirata, K. Matsuda, A.B. Spurdle, M. Kubo, Germline pathogenic variants of 11 breast cancer genes in 7,051 Japanese patients and 11,241 controls, Nature communications, 9 (2018) 4083.

[12] C. Tanikawa, Y. Kamatani, O. Toyoshima, H. Sakamoto, H. Ito, A. Takahashi, Y. Momozawa, M. Hirata, N. Fuse, T. Takai-Igarashi, A. Shimizu, M. Sasaki, T. Yamaji, N. Sawada, M. Iwasaki, S. Tsugane, M. Naito, A. Hishida, K. Wakai, N. Furusyo, Y. Murakami, Y. Nakamura, I. Imoto, J. Inazawa, I. Oze, N. Sato, F. Tanioka, H. Sugimura, H. Hirose, T. Yoshida, K. Matsuo, K. Michiaki, K. Matsuda, A GWAS identifies gastric cancer susceptibility loci at 12 q24.11-12 and 20q11.21, Cancer Sci, (2018).

[13] K. Ohki, N. Kiyokawa, Y. Saito, S. Hirabayashi, K. Nakabayashi, H. Ichikawa, Y. Momozawa, K. Okamura, A. Yoshimi, H. Ogata-Kawata, H. Sakamoto, M. Kato, K. Fukushima, D. Hasegawa, H. Fukushima, M. Imai, R. Kajiwara, T. Koike, I. Komori, A. Matsui, M. Mori, K. Moriwaki, Y. Noguchi, M.J. Park, T. Ueda, S. Yamamoto, K. Matsuda, T. Yoshida, K. Matsumoto, K. Hata, M. Kubo, Y. Matsubara, H. Takahashi, T. Fukushima, Y. Hayashi, K. Koh, A. Manabe, A. Ohara, Clinical and molecular characteristics of MEF2D fusion-positive pre-

cursor B-cell acute lymphoblastic leukemia in childhood, including a novel translocation resulting in MEF2D-HNRNPH1 gene fusion, Haematologica, (2018).

[14] F. Takeuchi, M. Akiyama, N. Matoba, T. Katsuya, M. Nakatochi, Y. Tabara, A. Narita, W.Y. Saw, S. Moon, C.N. Spracklen, J.F. Chai, Y.J. Kim, L. Zhang, C. Wang, H. Li, H. Li, J.Y. Wu, R. Dorajoo, J.L. Nierenberg, Y.X. Wang, J. He, D.A. Bennett, A. Takahashi, Y. Momozawa, M. Hirata, K. Matsuda, H. Rakugi, E. Nakashima, M. Isono, M. Shirota, A. Hozawa, S. Ichihara, T. Matsubara, K. Yamamoto, K. Kohara, M. Igase, S. Han, P. Gordon-Larsen, W. Huang, N.R. Lee, L.S. Adair, M.Y. Hwang, J. Lee, M.L. Chee, C. Sabanayagam, W. Zhao, J. Liu, D.F. Reilly, L. Sun, S. Huo, T.L. Edwards, J. Long, L.C. Chang, C.H. Chen, J.M. Yuan, W. P. Koh, Y. Friedlander, T.N. Kelly, W. Bin Wei, L. Xu, H. Cai, Y.B. Xiang, K. Lin, R. Clarke, R. G. Walters, I.Y. Millwood, L. Li, J.C. Chambers, J.S. Kooner, P. Elliott, P. van der Harst, C. International Genomics of Blood Pressure, Z. Chen, M. Sasaki, X.O. Shu, J.B. Jonas, J. He, C. K. Heng, Y.T. Chen, W. Zheng, X. Lin, Y.Y. Teo, E.S. Tai, C.Y. Cheng, T.Y. Wong, X. Sim, K.L. Mohlke, M. Yamamoto, B.J. Kim, T. Miki, T. Nabika, M. Yokota, Y. Kamatani, M. Kubo, N. Kato, Interethnic analyses of blood pressure loci in populations of East Asian and European descent, Nature communications, 9 (2018) 5052.

[15] A. Teumer, L. Chaker, S. Groeneweg, Y. Li, C. Di Munno, C. Barbieri, U.T. Schultheiss, M. Traglia, T.S. Ahluwalia, M. Akiyama, E.V.R. Appel, D.E. Arking, A. Arnold, A. Astrup, M. Beekman, J.P. Beilby, S. Bekaert, E. Boerwinkle, S.J. Brown, M. De Buyzere, P.J. Campbell, G. Ceresini, C. Cerqueira, F. Cucca, I.J. Deary, J. Deelen, K.U. Eckardt, A.B. Ekici, J.G. Eriksson, L. Ferrrucci, T. Fiers, E. Fiorillo, I. Ford, C.S. Fox, C. Fuchsberger, T.E. Galesloot, C. Gieger, M. Gogele, A. De Grandi, N. Grarup, K.H. Greiser, K. Haljas, T. Hansen, S.E. Harris, D. van Heemst, M. den Heijer, A.A. Hicks, W. den Hollander, G. Homuth, J. Hui, M.A. Ikram, T. Ittermann, R.A. Jensen, J. Jing, J.W. Jukema, E. Kajantie, Y. Kamatani, E. Kasbohm, J.M. Kaufman, L.A. Kiemeney, M. Kloppenburg, F. Kronenberg, M. Kubo, J. Lahti, B. Lapauw, S. Li, D.C.M. Liewald, S. Lifelines Cohort, E.M. Lim, A. Linneberg, M. Marina, D. Mascalzoni, K. Matsuda, D. Medenwald, C. Meisinger, I. Meulenbelt, T. De Meyer, H.E. Meyer Zu Schwabedissen, R. Mikolajczyk, M. Moed, R.T. Netea-Maier, I.M. Nolte, Y. Okada, M. Pala, C. Pattaro, O. Pedersen, A. Petersmann, E. Porcu, I. Postmus, P.P. Pramstaller, B. M. Psaty, Y.F.M. Ramos, R. Rawal, P. Redmond, J.B. Richards, E.R. Rietzschel, F. Rivadeneira, G. Roef, J.I. Rotter, C.F. Sala, D. Schlessinger, E. Selvin, P.E. Slagboom, N. Soranzo, T. I.A. Sorensen, T.D. Spector, J.M. Starr, D.J. Stott, Y. Taes, D. Taliun, T. Tanaka, B. Thuesen, D. Tiller, D. Toniolo, A.G. Uitterlinden, W. E. Visser, J.P. Walsh, S.G. Wilson, B.H.R. Wolffenbuttel, Q. Yang, H.F. Zheng, A. Cappola, R. P. Peeters, S. Naitza, H. Volzke, S. Sanna, A. Kottgen, T.J. Visser, M. Medici, Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation, Nature communications, 9 (2018) 4455.

[16] V. Yodsurang, Y. Tang, Y. Takahashi, C. Tanikawa, Y. Kamatani, A. Takahashi, Y. Momozawa, N. Fuse, J. Sugawara, A. Shimizu, A. Fukushima, A. Hishida, N. Furusyo, M. Naito, K. Wakai, T. Yamaji, N. Sawada, M. Iwasaki, S. Tsugane, M. Hirata, Y. Murakami, M. Kubo, K. Matsuda, Genome-wide association study (GWAS) of ovarian cancer in Japanese predicted regulatory variants in 22q13.1, PloS one, 13 (2018) e0209096.

[17] Y. Lu, S.S. Kweon, C. Tanikawa, W.H. Jia, Y.B. Xiang, Q. Cai, C. Zeng, S.L. Schmit, A. Shin, K. Matsuo, S.H. Jee, D.H. Kim, J. Kim, W. Wen, J. Shi, X. Guo, B. Li, N. Wang, B. Zhang, X. Li, M.H. Shin, H.L. Li, Z. Ren, J.H. Oh, I. Oze, Y. O. Ahn, K.J. Jung, D.V. Conti, F.R. Schumacher, G. Rennert, M.A. Jenkins, P.T. Campbell, M. Hoffmeister, G. Casey, S.B. Gruber, J. Gao, Y.T. Gao, Z.Z. Pan, Y. Kamatani, Y.X. Zeng, X.O. Shu, J. Long, K. Matsuda, W. Zheng, Large-scale Genome-wide Association Study of East Asians Identifies Loci Associated With Risk for Colorectal Cancer, Gastroenterology, (2018).

[18] Suzuki K, Akiyama M, Ishigaki K, Kanai M, Hosoe J, Shojima N, Hozawa A, Kadota A, Kuriki K, Naito M, Tanno K, Ishigaki Y, Hirata M, Matsuda K, Iwata N, Ikeda M, Sawada N, Yamaji T, Iwasaki M, Ikegawa S, Maeda S, Murakami Y, Wakai K, Tsugane S, Sasaki M, Yamamoto M, Okada Y, Kubo M, Kamatani Y, Horikoshi M, Yamauchi T, Kadowaki T. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. Nat Genet 27.125. 2019 Feb 4. doi: 10.1038/s41588-018-0332-4.

[19] Hirata J, Hosomichi K, Sakaue S, Kanai M, Nakaoka H, Ishigaki K, Suzuki K, Akiyama M, Kishikawa T, Ogawa K, Masuda T, Yamamoto K, Hirata M, Matsuda K, Momozawa Y, Inoue I, Kubo M, Kamatani Y, Okada Y. Genetic and phenotypic landscape of the major histocompatibilty complex region in the Japanese popu lation. Nat Genet doi: 10.1038/s41588-018-0336-0.

[20] Saito M, Okumura K, Isogai E, Araki K, Tanikawa C, Matsuda K, Kamijo T, Kominami R, Wakabayashi Y. A polymorphic variant in p 19Arf confers resistance to chemically-induced

skin tumors by activating the p53 pathway. J Invest Dermatol 6.448. doi: 10.1016/j.jid.2018.12.027.

[21] Okazaki S, Morimoto T, Kamatani Y, Kamimura T, Kobayashi H, Harada K, Tomita T, Higashiyama A, Takahashi JC, Nakagawara J, Koga M, Toyoda K, Washida K, Saito S, Takahashi A, Hirata M, Matsuda K, Mochizuki H, Chong M, Paré G, O'Donnell M, Ago T, Hata J, Ninomiya T, Dichgans M, Debette S, Kubo M, Koizumi A, Ihara M. Moyamoya Disease Susceptibility Variant RNF213 p.R4810K Increases the Risk of Ischemic Stroke Attributable to Large-Artery Atherosclerosis. Circulation doi: 10.1161/CIRCULATIONAHA.118.038439.

[22] Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, Smith GD, Mahajan A, Gaulton KJ, Nadkarni GN, Valladares-Salgado A, Wacher-Rodarte N, Mychaleckyj JC, Dueker ND, Guo X, Hai Y, Haessler J, Kamatani Y, Stilp AM, Zhu G, Cook JP, Ärnlöv J, Blanton SH, de Borst MH, Bottinger EP, Buchanan TA, Cechova S, Charchar FJ, Chu PL, Damman J, Eales J, Gharavi AG, Giedraitis V, Heath AC, Ipp E, Kiryluk K, Kramer HJ, Kubo M, Larsson A, Lindgren CM, Lu Y, Madden PAF, Montgomery GW, Papanicolaou GJ, Raffel LJ, Sacco RL, Sanchez E, Stark H, Sundstrom J, Taylor KD, Xiang AH, Zivkovic A, Lind L, Ingelsson E, Martin NG, Whitfield JB, Cai J, Laurie CC, Okada Y, Matsuda K, Kooperberg C, Chen YI, Rundek T, Rich SS, Loos RJF, Parra EJ, Cruz M, Rotter JI, Snieder H, Tomaszewski M, Humphreys BD, Franceschini N. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. Nat Commun. doi: 10.1038/s41467-018-07867-7.

[23] Sekimizu M, Yoshida A, Mitani S, Asano N, Hirata M, Kubo T, Yamazaki F, Sakamoto H, Kato M, Makise N, Mori T, Yamazaki N, Sekine S, Oda I, Watanabe SI, Hiraga H, Yonemoto T, Kawamoto T, Naka N, Funauchi Y, Nishida Y, Honoki K, Kawano H, Tsuchiya H, Kunisada T, Matsuda K, Inagaki K, Kawai A, Ichikawa H. Frequent mutations of genes encoding vacuolar H+ -ATPase components in granular cell tumors. Genes Chromosomes Cancer. (2018) doi: 10.1002/gcc.22727.

[24] M. Horikoshi, F.R. Day, M. Akiyama, M. Hirata, Y. Kamatani, K. Matsuda, K. Ishigaki, M. Kanai, H. Wright, C.A. Toro, S.R. Ojeda, A. Lomniczi, M. Kubo, K.K. Ong, J.R.B. Perry, Elucidating the genetic architecture of reproductive ageing in the Japanese population, Nature communications, 9 (2018) 1977.

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. Comprehensive characterization of spliced chimeric RNAs: insights into the mechanism of *trans*-splicing

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

Trans-splicing is a post-transcriptional processing event that joins exons from separate RNAs to produce a chimeric RNA. Although *trans*-splicing is thought to be a rare event in vertebrates, recent studies have suggested the existence of an unexpected number of *trans*-spliced RNAs in various human tissues and cells, increasing the potential importance of *trans*-splicing. Moreover, the principle of *trans*-splicing has been applied to gene therapy for human genetic diseases in the last two decades. However, the detailed mechanism of *trans*-splicing remains poorly understood. Here we extensively characterize *trans*-spliced genes and provide insights into the mechanism of *trans*-splicing. To study the *trans*-splicing mechanism, we mainly use *Ciona,* the closest invertebrate relative to humans, in which *trans*-splicing frequently occurs. Our *in silico* analysis revealed several characteristics of *Ciona trans*-spliced genes: i) the preferential location of *trans*-splice acceptor sites (TASs) at the first functional acceptor site, ii) the weak 5' splice sites, and iii) AU- and GU-rich regions upstream of TASs. Interestingly, these characteristics appear to be conserved between *Ciona* and humans, raising the possibility that they share the same basic mechanism of *trans*-splicing and highlighting the potential usefulness of *Ciona* as a model organism for *trans*-splicing studies. Moreover, motif analysis suggested that RNA-binding proteins including hnRNP and SR proteins may promote *trans*-splicing to produce chimeric RNAs. Our results will not only help us better understand the mechanism of *trans*-splicing in *Ciona* and humans, but also may lead to the development of more efficient *trans*-splicing-based gene therapy for human genetic diseases.

2. Cell specific change of DNA hydroxymethylation and function analysis about 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.

3. Analyzing the 3D chromatin organization coordinating with gene expression regulation in B-cell lymphoma

Luis Augusto Eijy Nagai, Sung-Joon Park and Kenta Nakai

Eukaryotes compact chromosomes densely and non-randomly, forming three-dimensional structures. Alterations of the chromatin structures are often associated with diseases. In particular, aggressive cancer development from the disruption of the humoral immune system presents abnormal gene regulation which is accompanied by chromatin reorganizations. How the chromatin structures orchestrate the gene expression regulation is still poorly understood. Herein, we focus on chromatin dynamics in normal and abnormal B cell lymphocytes, and investigate its functional impact on the regulation of gene expression. We conducted an integrative analysis using publicly available multi-omics data that include Hi-C, RNA-seq and ChIP-seq experiments with normal B cells, lymphoma and ES cells. We processed and re-analyzed the data exhaustively and combined different scales of genome structures with transcriptomic and epigenetic features. We found that the chromatin organizations are highly preserved among the cells. 5.2% of genes at the specific repressive compartment in normal pro-B cells were switched to the permissive compartment in lymphoma along with increased gene expression. The genes are involved in B-cell related biological processes. Remarkably, the boundaries of topologically associating domains were not enriched by CTCF motif, but significantly enriched with Prdm1 motif that is known to be the key factor of B-cell dysfunction in aggressive lymphoma. This study shows evidence of a complex relationship between chromatin reorganization and gene regulation. However, an unknown mechanism may exist to restrict the structural and functional changes of genomic regions and cognate genes in a specific manner. Our findings suggest the presence of an intricate crosstalk between the higher-order chromatin structure and cancer development.

4. Genomic Analysis of Pancreatic Juice DNA Assesses Malignant Risk of Intraductal Papillary Mucinous Neoplasm of Pancreas

Raúl Nicolás Mateos, Hidewaki Nakagawa[2], Seiko Hirono[3], Shinichi Takano[4], Mitsuharu Fukasawa[4], Akio Yanagisawa[5], Satoru Yasukawa[5], Kazuhiro Maejima[2], Aya Oku-Sasaki[2], Kaoru Nakano[2], Munmee Dutta, Hiroko Tanaka[5], Satoru Miyano[6], Nobuyuki Enomoto[4], Hiroki Yamaue[3], Kenta Nakai, and Masashi Fujita[2]: [2]RIKEN Center for Integrative Medical Sciences, [3]Second Dept. of Surgery, Wakayama Medical Univ., [4]First Dept. of Internal Medicine, Univ. of Yamanashi, [5]Dept. of Surgical Pathology, Kyoto Pref. Univ. of Medicine, [6]Lab. of DNA Information Analysis, IMS, Univ. of Tokyo

Intraductal papillary mucinous neoplasm (IPMN) of pancreas has a high risk to develop into invasive cancer or co-occur with malignant lesion. For this reason, it is important to assess its malignant risk by non-invasive approach. An ideal material for this purpose would be Pancreatic juice cell-free DNA (PJD), but genetic biomarkers for predicting malignant risk from PJD are not yet established. Here, I performed deep exome sequencing analysis of PJD from 40 IPMN patients with or without malignant lesion. In order to evaluate their potential as a malignancy marker I compared the somatic alterations and copy number alterations detected in PJD with the histologic grade of IPMN. Somatic mutations of KRAS, GNAS, TP53, and RNF43 were commonly detected in PJD of IPMNs, but no association with the histological grades of IPMN was found. Instead, histologic grade was positively correlated with mutation burden ($r = 0.417$, $P = 0.018$). I was also able to observe frequent copy number deletions in 17p13 (TP53) and amplifications in 7q21 and 8q24 (MYC) in PJDs. The amplifications in 7q21 and 8q24 were positively correlated with the histologic grade and most prevalent in the cases of grade 3 ($P = 0.012$ and 7/11; $P = 0.011$ and 6/11, re-

spectively). Mutation burden and copy number alterations detected in PJD have potential to assess the malignant progression risk of IPMNs. These findings showed clinical usefulness of genomic profiling of PJD for IPMN.

5. Whole Genome sequencing analysis of Esophageal Squamous cell carcinoma in Japanese population

Munmee Dutta, Masashi Fujita[2], Tadashi Yasuda[7], Raúl Nicolás Mateos, Ashwini Patil, Kenta Nakai and Hidewaki Nakagawa[2]: [7]Dept. of Surgery, Kinki Univ. School of Medicine, Osaka

Esophageal cancer is one of the most aggressive cancer, ranked 6th worldwide in terms of cancer-related death. While the Esophageal Squamous Cell Carcinoma (ESCC) is predominant in Asian region such as Japan, China and India, the other type, Esophageal Adenocarcinoma (EAC) is common in western countries. However, the therapeutic and diagnostic options are still poor. Despite the advances in recent technology, the survival rate of the ESCC is still low, and in some cases tumor recur within short period of post-surgery. This implies that the molecular carcinogenesis and the progression of ESCC is still not clear. In this study, we attempted to elucidate the underlying genomic alterations affected by different mutational events in ESCC. We performed whole genome sequencing (WGS) of 20 pairs of matched normal and tumor samples with ESCC. We identified frequent somatic mutations in genes such as TP53, TTN, AHNAK2 and F5 in ESCC. We also detected 6 mutation signatures, one of which has association with smoking. In addition, our copy number alterations (CNAs) analysis identified 10 amplified and deleted regions in ESCC, such as 3q26.33 and 9p21.3. Identification of recurrent gene mutation, mutational signatures and CNAs improves our understanding of the ESCC development at molecular level and provides future therapeutic targets for precision medicine.

6. Genome-wide prediction and analysis of cell-specific enhancers and their strength integrating sequence-based features with functional genomic datasets

Leyi Wei and Kenta Nakai

Enhancers are DNA regulatory elements that bind transcription factors (TFs) to boost the expression of distal target genes such as promoters. Understanding enhancers is currently an area of great interest, since their importance is not only in developmental gene expression but also in evolution and disease. To fully understand the functional mechanisms of enhancers, it is critical to identify the sequence elements that have enhancer activity on a genome-wide scale. Recently, machine learning methods have emerged as effective and promising approaches in the enhancer prediction field. However, feature representation for enhancers is still quite a challenging task for accurate prediction of enhancers. In this study, we propose a novel machine learning method for enhancer prediction using sequence only. In this method, we used multiple sequence-based feature extraction algorithms from different aspects, such as nucleotide composition, DNA motif, and DNA physical properties. Specifically, to improve the feature representation ability, we introduce a supervised feature representation learning scheme that automatically incorporates the class information to original low-level feature space and learn high-latent features via multiple machine learning models. T-SNE visualization results showed that the true enhancers and non-enhancers in our learnt feature space are distributed more clearly in two clusters as compared to the original feature space, demonstrating that the class information is complementary to sequential information to correctly distinguish enhancers from non-enhancers. Moreover, benchmarking results with the state-of-art methods that use sequence information only showed that our proposed method significantly outperforms existing methods in overall performance, suggesting that our method is more effective to identify the regulatory enhancers.

7. Whole genome sequence enhancer annotation without epigenetic information

Vincent Berthier and Kenta Nakai

Most enhancer identification tools use epigenetic information (such as histone modifications) to identify enhancer sequences. While exhibiting good results, getting the required information is both expensive and time consuming, which is why some new techniques, relying only on the sequence themselves have been developed. Those techniques however have two problems: first they consider each sequence isolated from the genome, second they only aim at identifying if a given, very specific sequence, is an enhancer or not. In this study, given the whole sequence of a chromosome or organism we aim to identify where the enhancers are likely to be found, which would guide researchers to finding the enhancers and not simply checking if a sequence is one or not.

8. Detecting Microbial Contaminants in Next Generation Sequencing Data

Sung-Joon Park, Satoru Onizuka[8,9], Takanori Iwata[8], Kenta Nakai: [8]Inst. of Advanced Biomedical Engg. and Science, Tokyo Women's Medical Univ.,

[9]Dept of Oral Function, Kyushu Dental Univ.

Quality assurance is becoming an increasingly important issue in various research settings. In this regard, computational approaches using next-generation sequencing (NGS) data offer promising diagnostics to assess the presence of contaminants in biological and biomedical research. Since the biological resources contaminated by multiple microorganisms impedes successful research, bioinformatics requires careful attention to intra- and interspecies sequence similarities that is indispensable for preventing misinterpretations. Here we tackle this issue by proposing a novel method that estimates the likelihood of reads mapped to a specific microbial genus and incorporates the reads mapped to multiple microbial genera. Through the analysis of large-scale NGS samples including microbe spike-in datasets, we found that an RNA-seq assay includes 1,000-100,000 contaminant reads when one million host reads are sequenced, and that the contaminant reads consist of multiple genera with 10-10,000 reads. In addition, Cutibacterium was determined as one of predominant contaminants in public NGS data, and we identified 13 host genes from Mycoplasma-infected mesenchymal stem cells as biomarkers. We believe that our approach facilitates the profiling of contamination landscape with reduced detection uncertainty.

9. Computational Modeling of Gene Regulation with 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, as a sub project of the research project "Chromosome orchestration system" (http://www.chromosomeos.com)", we are trying to build a computational model for explaining the importance of high-dimensional chromatin structures underlying the regulation of gene expression. By incorporating the histone modification and RNA-seq data in the web-based repository system (https://openlooper.hgc.jp/) we launched, we particularly analyzed the gene expression regulation in B-cell lymphoma through a linear regression modeling approach that successfully predicted the differentially expressed genes in lymphoma. The model demonstrated that intra-chromosome interactions found within 5k-bp upstream regions from TSSs greatly impact on lymphoma development. This observation suggests the existence of complex cross-talk among genetic and epigenetic regulatory elements.

10. Analyzing Chromatin Accessibility Dynamics in Male Germ Cell Development

Luis Augusto Eijy Nagai, Sung-Joon Park, Soichiro Yamanaka[10], Haruhiko Shiomi[10], Kenta Nakai: [10]Dept. of Molecular Biology, Keio Univ. 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 DNA demethylation and pervasive transcriptomic activity. To unveil the dynamics of chromatin conformational changes in the transition from primordial germ cell to spermatogonial stem cells, we here analyzed ATAC-seq and HiC-seq datasets 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 highly enriched with transposable elements potentially include regulatory elements for genes located at distal intra- or inter-chromosome loci. In addition, the process of DNA demethylation and re-establishment cooperates with the state of chromatin accessibility. These observations suggest the global reprogramming during the transition is strictly controlled by various epigenetic regulatory elements.

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

Yasuo Ouchi[11], Ashwini Patil, Yusuke Tamura[12], Hiroshi Nishimasu[13], Naoki Takemura[11,12], Takeshi Sato[14], Yasumasa Kimura[14], Osamu Nureki[13], Kenta Nakai, Hiroshi Kiyono[13,15,16], Satoshi Uematsu[11,12]: [11]Dept of Mucosal Immunology, Chiba Univ., [12]Division of Innate Immune Regulation, IMS, Univ. of Tokyo, [13]Dept of Biol Sci, Univ. of Tokyo, [14]Division of Systems Immunology, IMS, Univ. of Tokyo, [15]Division of Mucosal Immunology, IMS, Univ. of Tokyo, [16]Dept of Immunology, Chiba Univ.

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.

12. Prediction of MoRF regions in intrinsically disordered protein sequences

Ronesh Sharma[17,18], Gaurav Raicar[17], Maitsetseg Bayarjargal[18], Tatsuhiko Tsunoda[2,19], Ashwini Patil[§], Alok Sharma[2,18,19§]: [17]Fiji National Univ., Suva, Fiji, [18]The Univ. of South Pacific, Suva, Fiji, [19]Medical Research Institute, Tokyo Med. and Dental Univ.

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, MoRFPred-plus, OPAL and OPAL+, to identify MoRFs in disordered protein sequences. All 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. OPAL+ uses a length-specific strategy and performs equally well.

13. 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 evaluated in multiple species and compared with other similar algorithms. It has been shown to reconstruct known pathways in KEGG in the mammalian immune system1 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.

14. Waves of chromatin modifications in mouse dendritic cells in response to LPS stimulation

Alexis Vandenbon[19], Yutaro Kumagai[20], Mengjie Lin[19], Yutaka Suzuki[21], and Kenta Nakai: [19]Inst. Front. Life Med. Sci., Kyoto Univ., [20]IFReC, Osaka Univ., [21]Grad. Sch. Front. Sci., Univ. of Tokyo

The importance of transcription factors (TFs) and epigenetic modifications in the control of gene expression is widely accepted. However, causal relationships between changes in TF binding, histone modifications, and gene expression during the response to extracellular stimuli are not well understood. Here, we analyze the ordering of these events on a genome-wide scale in dendritic cells in response to lipopolysaccharide (LPS) stimulation. Using a ChIP-seq time series dataset, we find that the LPS-induced accumulation of different histone modifications follows clearly distinct patterns. Increases in H3K4me3 appear to coincide with transcriptional activation. In contrast, H3K9K14ac accumulates early after stimulation, and H3K36me3 at later time points. Integrative analysis with TF binding data reveals potential links between TF activation and dynamics in histone modifications. Especially, LPS-induced increases in H3K9K14ac and H3 K4me3 are associated with binding by STAT1/2 and were severely impaired in Stat1$^{-/-}$ cells. While the timing of short-term changes of some histone modifications coincides with changes in transcriptional activity, this is not the case for others. In the latter case, dynamics in modifications more likely reflect strict regulation by stimulus-induced TFs and their interactions with chromatin modifiers.

Publications

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., Le-

maire, 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.

Farmanbar, A., Firouzi, S., Makałowski, W., Kneller, R., Iwanaga, M., Utsunomiya, A., Nakai, K., Watanabe, T. Mutational intratumor heterogeneity is a complex and early event in the development of adult T-cell leukemia/lymphoma. *Neoplasia*, 20: 883-893, 2018.

Fujita, M., Matsubara, N., Matsuda, I., Maejima, K., Oosawa, A., Yamano, T., Fujimoto, A., Furuta, M., Nakano, K., Oku-Sasaki, A., Tanaka, H., Shiraishi, Y., Mateos, R.N., Nakai, K., Miyano, S., Tomita, N., Hirota, S., Ikeuchi, H., Nakagawa, H. Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling. *Oncotarget,* 9: 969-981, 2018.

Fujita, N., Mizuarai, S., Murakami, K., Nakai, K. Biomarker discovery by integrated joint non-negative matrix factorization and pathway signature analyses. *Sci Rep,* 8: 9743, 2018.

Moon, M., Nakai, K. Integrative analysis of gene expression and DNA methylation using unsupervised feature extraction for detecting candidate cancer biomarkers. *J Bioinform Comput Biol,* 16: 1850006, 2018.

Nagai, L.A.E., Park, S.J., Nakai, K. Analyzing the 3D chromatin organization coordinating with gene expression regulation in B-cell lymphoma. *BMC Med Genomics,* in press, 2019.

Nakai, K 2019. Prediction of protein-binding sites in DNA sequences. *In:* RANGANATHAN, S., NAKAI, K., GRIBSKOV, M. & SCHÖNBACH, C. (eds.) *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier.

Nakai, K., Imai, K. 2019. Prediction of protein localization. *In:* RANGANATHAN, S., NAKAI, K., GRIBSKOV, M. & SCHÖNBACH, C. (eds.) *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier.

Ouchi, Y., Patil, A., Tamura, Y., Nishimasu, H., Negishi, A., Paul, S.K., Takemura, N., Satoh, T., Kimura, Y., Kurachi, M., Nureki, O., Nakai, K., Kiyono, H., Uematsu, S. Generation of tumor antigen-specific murine CD8+ T cells with enhanced anti-tumor activity via highly efficient CRISPR/Cas9 genome editing. *Int Immunol,* 30: 141-154, 2018.

Park, S.J. 2019. Genome-Wide Scanning of Gene Expression. *In:* RANGANATHAN, S., NAKAI, K., GRIBSKOV, M. & SCHÖNBACH, C. (eds.) *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier.

Patil, A. 2019. Protein-protein interaction databases. *In:* RANGANATHAN, S., NAKAI, K., GRIBSKOV, M. & SCHÖNBACH, C. (eds.) *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier.

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.

Sharma, R., Raicar, G., Tsunoda, T., Patil, A.*, Sharma, A.* OPAL: prediction of MoRF regions in intrinsically disordered protein sequences. *Bioinformatics,* 34: 1850-1858, 2018.

Sharma, R., Sharma, A., Raicar, G., Tsunoda, T., Patil, A. OPAL+: Length-Specific MoRF Prediction in Intrinsically Disordered Protein Sequences. *Proteomics*: e1800058, 2018.

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.

Tan, P.L., Lopez, Y., Nakai, K., Patil, A. TimeXNet Web: identifying cellular response networks from diverse omics time-course data. *Bioinformatics,* 34: 3764-3765, 2018.

Vandenbon, A., Kumagai, Y., Lin, M., Suzuki, Y., Nakai, K. Waves of chromatin modifications in mouse dendritic cells in response to LPS stimulation. *Genome Biol,* 19: 138, 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	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 of stem cell research

Japan Agency for Medical Research and Development (AMED) has commissioned us to provide research ethics consultation to stem cell research since 2012. The program is called "research on the ethical, legal, and social implications related to regenerative medicine". In order to make regenerative medicine more concrete, it is essential to promote research development with a definite focus on 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, research ethics committees, return of research results, inclusion criterion of participants of first-in-human trials and governance of iPSC banking. 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.

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

In order to create next-generation cancer therapies, this research program promotes research aimed at elucidating the biological properties of cancer, research based on patients' clinical data, and research combining both aspects. Through this process, the program accelerates the development and practical application of diagnostic biomarkers for the prevention and early detection of cancer, and innovative drugs for cancer treatment. We provided a model form for information sheet, consent form and leaflet for research participants.

3. Research ethics consultation and studies on ethical, legal, and social implications of prospective cohort study of elderly people

Japan Prospective Studies Collaboration for Aging and Dementia (JPSC-AD) study is a collaborative prospective cohort study of approximately 10,000 elderly people from 8 newly-established community-based dementia cohort studies in Japan, in which the data is prospectively collected by using the pre-specified standardized protocol. Ap-

proximately 10,000 community-dwelling individuals aged 65 years or older at 8 sites of Japan will be recruited in the baseline survey from 2016 to 2018, and followed up to detect incident cases of dementia for at least 5 years. In the baseline survey, the following data has been collected: lifestyle information (smoking habits, alcohol intakes, diet, physical activity, etc.), medical history, physical examination, blood test, and brain magnetic resonance imaging (MRI). Frozen samples of serum and plasma (−80℃) have been also stored. We provided several suggestions for protecting vulnerable research participants and created a short movie to remind them that their data has been collected and analyzed for this cohort study.

4. Survey on the perception of germline genome editing among the general public in Japan

Genome editing of human embryos could become a fundamental treatment approach for genetic diseases; however, a few technical and ethical issues need to be resolved before its application in clinical settings. Presently, the Japanese government has issued a statement prohibiting human germline editing and emphasizing the need for discussions that include a wide range of perspectives. However, current discussions tend to exclude the general public. Therefore, we conducted a survey of 10,881 general adults and 1044 patients in Japan who indicated that their disease conditions are related to their genetic makeup, and clarified their attitude toward this technology. The results clearly indicated that the Japanese people generally accepted the use of genome editing for disease-related genes, but many were concerned about the risks. In addition, many Japanese people did not understand the technology well. To improve awareness and understanding about genome editing, it is important that scientists and science communicators create opportunities for the public to participate in relevant discussions without harming vulnerable participants. It is also important to continuously track changes in the acceptance of genome editing by the public.

5. Ethical concerns on sharing genomic data including patients' family members

Platforms for sharing genomic and phenotype data have been developed to promote genomic research, while maximizing the utility of existing datasets and minimizing the burden on participants. The value of genomic analysis of trios or family members has increased, especially in rare diseases and cancers. However, current data-sharing policies have no specific safeguards or provisions for familial data sharing. A quantitative survey conducted on 10,881 general adults in Japan indicated that they expected stronger protection mechanisms when their family members' clinical and/or genomic data were shared together, as compared to when only their data were shared. A framework that respects decision-making and the right of withdrawal of participants, including family members, along with ensuring usefulness and security of data is needed. To enable this, we propose recommendations on ancillary safeguards for familial data sharing according to the stakeholders, namely, initial researchers, genomic researchers, data submitters, database operators, institutional review boards, and the public and participants. Families have played significant roles in genetic research, and its value is re-illuminated in the era of genomic medicine. It is important to make progress in data sharing while simultaneously protecting the privacy and interests of patients and families, and return its benefits to them.

6. Attitudes toward genomic tumor profiling tests in Japan

Genomic tumor profiling tests (GTPTs) to find molecular targeted drugs for patients with advanced cancer are being introduced into clinical settings, which may result in secondary germline findings. We conducted anonymous surveys with 757 cancer patients (CPs), 763 family members (FMs), and 3697 general adults (GAs) in Japan. Awareness of GTPTs was low in all groups, however, both CPs and FMs showed a higher degree of recognition in the benefits of GTPTs. FMs wanted information on germline findings to be shared more than the CPs. Since advanced CPs may have psychological burdens that make it difficult to express their opinions on their therapeutic options and sharing germline findings, GTPTs should be offered with advanced care planning for patients.

7. Patient deliberation time during informal decision-making in clinical trials

Informed consent is an essential part of an ethical clinical trial; to this end, researchers have developed several interventions to promote participants' full understanding of trials and thereby improve the consent process. However, few empirical studies have examined how patients make the decision of whether to give consent. The objective of this study, therefore, is to analyze patients' decision-making process when participating in clinical trials. We conducted an internet survey (n=2,045) and interview data analysis (n=40) with patients and categorize respondents into three types of participants: active, passive, and non-participation. Our results show that patients often make informal and quick decisions before medical staff provide them with relevant information during the informed consent

process. For example, 55.9% of patients received initial information on clinical trials from an online article or web advertising, and 54.5% consulted no one about whether to participate in the clinical trial before making a decision. Only 20.7% of respondents subjectively spent time making the decision whether to participate; 43.0% of patients who said that they "spent time" coming to a decision took four or more days to reach a decision, while 8.3% of people who "did not spend time" making a decision took this among of time. Based on these results, we were able to break patients' decision-making process into four steps: first contact, informal decision making, relevant information, and formal decision making. Our results show that patients are most likely to make a decision based on the first information they receive on the clinical trial, whatever the source. To this end, having a list of questions for potential participants to ask researchers would be useful in helping better collecting information of clinical trials. In addition, research teams should give patients more than four days to decide between providing them with relevant information and obtaining written consent, even if the patient seems to make a quick decision.

Publications

1. Nakada H, Yoshida S, Muto K. "Tell me what you suggest, and let's do that, doctor": Patient deliberation time during informal decision-making in clinical trials. PLoS ONE 14(1): e0211338. 2019.
2. Takashima K, Maru Y, Mori S, Mano H, Noda T, Muto K. Ethical concerns on sharing genomic data including patients' family members. BMC Medical Ethics 19: 61, 2018.
3. Nagai A, Ri I, Muto K. Attitudes toward genomic tumor profiling tests in Japan: patients, family members, and the public. J Hum Genet. 2019 Jan 10. (Epub ahead of print)
4. Uchiyama M, Nagai A, Muto K. Survey on the perception of germline genome editing among the general public in Japan. J Hum Genet. 2018 JUN; 63(6): 745-748.
5. 永井亜貴子，武藤香織，井上悠輔．本人通知制度の実態と住民票を用いた予後調査への影響の検討．65(5)：223–232，2018
6. 稲野彰洋，内田英二，井上悠輔，楠岡英雄，田代志門，長谷川純一，森下典子，笹栗俊之．多施設共同治験における倫理審査集約化の現状と課題，臨床薬理．49(4)：159–167，2018
7. 船橋亜希子．過失犯における行為者個人の能力の取扱いに関する序論的考察．明大論集．49：1–18，2018
8. 李怡然・武藤香織，ゲノム医療時代における「知らないでいる権利」，保健医療社会学論集，29(1)，72–82，2018