Health Intelligence Center

Division of Health Medical Data Science
健康医療データサイエンス分野

Professor	Seiya Imoto, Ph.D	教　授　博士(数理学)	井　元　清　哉
Assistant Professor	Takanori Hasegawa, Ph.D	助　教　博士(情報学)	長谷川　嵩　矩

Our mission is to utilize genomic big data and time series health medical data to realize methods for prediction and prevention of diseases and keeping/improving our health. For this purpose, we develop novel computational data analysis technologies by integrating Bayesian statistical theory and high performance computing on supercomputer system.

1. Development of Computational Platform for Clinical Sequence and Interpretation

Shimizu E, Yui H, Kasajima R, Yamaguchi K, Yokoyama K, Komura M, Saito A, Kobayashi M, Yuji K, Shibuya T, Hasegawa T, Niida A, Miyagi Y, Muto K, Tojo A, Furukawa Y, Imoto S, Miyano S, Yamaguchi R

From April 2015, Medical Genomics Research Initiative The University of Tokyo is launched. For implementing clinical sequence in the Institute of Medical Science, we formed a team of researchers and technicians who have various academic backgrounds including medicine, biology, pharmacology, genetics, statistics, computer science, ethics, etc. A highly secure infrastructure for analyzing personal genome was constructed; in the space, next generation sequencers are directly connected to a part (disconnect to internet) of supercomputer system in Human Genome Center and, for keeping traceability, laboratory information management system (LIMS) is installed to record all logs of wet experiments and computational analyses. Together with genome analysis in clinical sequence, we now intensively focus on a method for interpreting personal genome information. In July 2015, we started to use IBM Watson for cancer research to interpret the results of genome analyses. The results of genome sequence analysis including the interpretation of IBM Watson are evaluated and discussed in biweekly sequence board meeting. In 2016, we analyzed around sequence data of 100 cancer patients (more than 250 sequencing samples) with whole genome, exome, target deep sequencings. Also, multi-omics data including genome, transcriptome and epigenome were measured for integrative analysis that has the potential to achieve highly precise interpretation. This research is also performed as a part of the University of Tokyo's Center of Innovation (COI) project "Self-Managing Healthy Society".

2. Health Medical Big Data Analysis

a. Integration of the records of health examination, microbiome and genomic data for predicting disease risks

Hasegawa T, Kakuta M, Niida A, Yamaguchi R, Imoto S

To predict disease risk, we investigate a computational method to integrate the data of health examination, microbiome and genome. We collaborate with the research project at Hirosaki University

COI program "The Center of Healthy Aging Innovation" who collects those data of more than 20 thousand participants in total over 10 years. Our IMSUT team analyzes their data on the supercomputer system Shirokane3.

3. Computational Methods in Systems Biology and Immunology

a. Biomarker detection from time series gene expression data. Approximate Bayesian computation for evaluating biological simulation models

Kayano M[1], Matsui H[2], Yamaguchi R, Imoto S, Miyano S: [1]Department of Animal and Food Hygiene, Obihiro University of Agriculture and Venterinary Medicine, [2]Graduate School of Mathematics, Kyushu University

High-throughput time course expression profiles have been available in the last decade due to developments in measurement techniques and devices. Functional data analysis, which treats smoothed curves instead of originally observed discrete data, is effective for the time course expression profiles in terms of dimension reduction, robustness, and applicability to data measured at small and irregularly spaced time points. However, the statistical method of differential analysis for time course expression profiles has not been well established. We propose a functional logistic model based on elastic net regularization (F-Logistic) in order to identify the genes with dynamic alterations in case/control study. We employ a mixed model as a smoothing method to obtain functional data; then F-Logistic is applied to time course profiles measured at small and irregularly spaced time points. We evaluate the performance of F-Logistic in comparison with another functional data approach, i.e. functional ANOVA test (F-ANOVA), by applying the methods to real and synthetic time course data sets. The real data sets consist of the time course gene expression profiles for long-term effects of recombinant interferon beta on disease progression in multiple sclerosis. F-Logistic distinguishes dynamic alterations, which cannot be found by competitive approaches such as F-ANOVA, in case/control study based on time course expression profiles. F-Logistic is effective for time-dependent biomarker detection, diagnosis, and therapy.

b. Bayesian model for analyzing human leukocyte antigen regions

Hayashi S, Yamaguchi R, Mizuno S[3], Komura M, Miyano S, Nakagawa H[4], Imoto S: [3]Center for Advanced Medical Innovation, Kyushu University, [4]RIKEN Center for Integrative Medical Sciences

Although human leukocyte antigen (HLA) genotyping based on amplicon, whole exome, 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 versatile method for in- depth analysis of HLA genes. We developed ALPHLARD, which is a Bayesian model 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 6-digit resolution. Furthermore, ALPHLARD can detect rare germline variants not stored in HLA databases, reconstruct unknown non-coding sequences, and call somatic mutations from paired normal and tumor WGS data. We illustrate the capability of ALPHLARD using 25 WGS data sets from an Illumina platform. By comparing the results of HLA genotyping from SBT and amplicon sequencing methods, ALPHLARD achieved 98.1% accuracy at 6-digit resolution. Moreover, we could reconstruct 99.92% of unknown non-coding sequences of HLA-A*26:03:01. We also detected three somatic point mutations and one case of loss of heterozygosity in the HLA genes of the samples. This research is performed as a part of International Cancer Genome Consortium PanCancer Analysis Project.

c. An *in silico* automated pipeline to identify tumor specific neoantigens from next generation sequencing data

Hasegawa T, Hayashi S, Shimizu E, Mizuno S, Yamaguchi R, Miyano S, Nakagawa S, Imoto S:

Recent progress of massive parallel sequencing technology enables us to detect somatic mutations in each of cancer patients. It is known that some mutated peptides produced from missense mutations binds to the major histocompatibility complex (MHC). Since MHC presents mutated peptides to anti-tumor T cells, understanding this process is important in cancer immunotherapy. In this paper, we introduce a computational pipeline to predict binding affinity between mutated peptides and MHC molecules to detect neoantigens. We have implemented this pipeline on our supercomputer system. With nonsynonymous substitutions, frameshift insertions and deletions detected and intron retentions from whole-genome or exome sequencing data, we utilize RNA sequencing data and annotation data to make neoantigen detection pipeline more accurate.

d. Somatic mutation detection from next generation sequencing data

Moriyama T, Shiraishi Y, Chiba K, Yamaguchi R,

Imoto S, Miyano S

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

Publications

1. Zhang Y-Z, Yamaguchi R, Imoto S, Miyano S, Sequence-specific bias correction for RNA-seq data using recurrent neural networks, *BMC Genomics*, Suppl 1: 1044, 2017.
2. Yamaguchi K, Nagayama S, Shimizu E, Komura M, Yamaguchi R, Shibuya T, Arai M, Hatakeyama S, Ikenoue T, Ueno M, Miyano S, Imoto S, Furukawa Y, Reduced expression of APC-1B but not APC-1A by the deletion of promoter 1B is responsible for familial adenomatous polyposis, *Sci Rep*. 6: 26011, 2016.
3. Chapman CG, Yamaguchi R, Tamura K, Weidner J, Imoto S, Kwon J, Fang H, Yew PY, Marino SR, Miyano S, Nakamura Y, Kiyotani K. Characterization of T-cell receptor repertoire in inflamed tissues of patients with Crohn's disease through deep sequencing, *Inflamm Bowel Dis*. 22(6): 1275-85, 2016.
4. Hasegawa T, Hayashi S, Shimizu E, Mizuno S, Yamaguchi R, Miyano S, Nakagawa H, Imoto S. An in silico automated pipeline to identify tumor specic neoantigens from whole genome and exome sequencing data. *Proc. 12th International Symposium on Bioinformatics Research and Applications*, on line, 2016.
5. Yoshino T, Katayama K, Horiba Y, Munakata K, Yamaguchi R, Imoto S, Miyano S, Mima H, Watanabe K. Predicting Japanese Kampo formulas by analyzing database of medical records: a preliminary observational study, *BMC Medical Informatics and Decision Making*, 16: 118. 2016.
6. Park H, Niida A, Imoto S, Miyano S, Interaction based feature selection for uncovering cancer driver genes via copy number driven expression level, *Journal of Computational Biology*, in press.
7. Mao Y, Tamura T, Yuki Y, Abe D, Tamada Y, Imoto S, Tanaka H, Homma H, Miyano S, Okazawa H. The hnRNP-Htt axis regulates necrotic cell death induced by transcriptional repression through impaired RNA splicing, *Cell Death & Disease*, 7: e2207, 2016.
8. Muramatsu T, Kozaki K-i, Imoto S, Yamaguchi R, Tsuda H, Kawano T, Fujiwara N, Morishita M, Miyano S, Inazawa J. The hypusine cascade promotes cancer progression and metastasis through the regulation of RhoA in squamous cell carcinoma, *Oncogene*, 35(40): 5304-5316, 2016.
9. Park H, Shiraishi Y, Imoto S, Miyano S. A novel adaptive penalized logistic regression for uncovering biomarker associated with anti-cancer drug sensitivity, *IEEE IEEE/ACM Transactions on Computational Biology and Bioinformatics*, in press.
10. Moriyama T, Shiraishi Y, Chiba K, Yamaguchi R, Imoto S, Miyano S. OVarCall: Bayesian mutation calling method utilizing overlapping paired-end reads, *Lecture Notes in Computer Science*, 9683, 40-51, 2016.
11. Kato T, Inoue H, Imoto S, Tamada Y, Miyamoto T, Matsuo Y, Nakamura Y, Park JH. Oncogenic roles of TOPK and MELK, and effective growth suppression by small molecular inhibitors in kidney cancer cells, *Oncotarget*, 7, 17652-17664, 2016.
12. Kayano M, Matsui H, Yamaguchi R, Imoto S, Miyano S. Gene set differential analysis of time course expression profiles via sparse estimation in functional logistic model with application to time-dependent biomarker detection, *Biostatistics*, 17(2): 235-248, 2016.

Health Intelligence Center

Division of Health Medical Computational Science
健康医療計算科学分野

Professor	Satoru Miyano, Ph.D.	教　授	理学博士　宮　野　　悟
Assistant Professor	Atsushi Niida, Ph.D.	助　教	博士(理学)　新井田　厚　司

The mission of this division is to develop computational science for transforming biomedical data to knowledge. By making full use of supercomputers, we are now focusing on annotation, translation and interpretation of genomic data including RNA sequences for supporting cancer research and clinical sequence.

1. Computational Science for Cancer Research

a. Integrated multiregional analysis proposing a new model of colorectal cancer evolution

Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, Sawada G, Iwaya T, Kurashige J, Shinden Y, Iguchi T, Eguchi H, Chiba K, Shiraishi Y, Nagae G, Yoshida K, Nagata Y, Haeno H, Yamamoto H, Ishii H, Doki Y, Iinuma H, Sasaki S, Nagayama S, Yamada K, Yachida S, Kato M, Shibata T, Oki E, Saeki H, Shirabe K, Oda Y, Maehara Y, Komune S, Mori M, Suzuki Y, Yamamoto K, Aburatani H, Ogawa S, Miyano S, Mimori K

Understanding intratumor heterogeneity is clinically important because it could cause therapeutic failure by fostering evolutionary adaptation. To this end, we profiled the genome and epigenome in multiple regions within each of nine colorectal tumors. Extensive intertumor heterogeneity is observed, from which we inferred the evolutionary history of the tumors. First, clonally shared alterations appeared, in which C>T transitions at CpG site and CpG island hypermethylation were relatively enriched. Correlation between mutation counts and patients' ages suggests that the early-acquired alterations resulted from aging. In the late phase, a parental clone was branched into numerous subclones. Known driver alterations were observed frequently in the early-acquired alterations, but rarely in the late-acquired alterations. Consistently, our computational simulation of the branching evolution suggests that extensive intratumor heterogeneity could be generated by neutral evolution. Collectively, we propose a new model of colorectal cancer evolution, which is useful for understanding and confronting this heterogeneous disease.

b. Somatic mutations in plasma cell-free DNA are diagnostic markers for esophageal squamous cell carcinoma recurrence

Ueda M, Iguchi T, Masuda T, Nakahara Y, Hirata H, Uchi R, Niida A, Momose K, Sakimura S, Chiba K, Eguchi H, Ito S, Sugimachi K, Yamasaki M, Suzuki Y, Miyano S, Doki Y, Mori M, Mimori K

Esophageal squamous cell carcinoma (ESCC) is one of the most aggressive malignancies owing to the high frequency of tumor recurrence. The identification of markers for early ESCC diagnosis and prediction of recurrence is expected to improve the

long-term prognosis. Therefore, we searched for associations between tumor recurrence and cell-free DNA (cfDNA) mutations in blood plasma, which contains genetic markers for various cancer types.

Genomic DNA from tumors and cfDNA from plasma were obtained from 13 patients undergoing treatment for newly diagnosed ESCC. Next-generation sequencing of cfDNA in plasma was performed to identify mutations in 53 cancer-related genes, in which recurrent mutations were previously detected in ESCC. cfDNA mutational profiles were compared before and after tumor resection in four patients. Furthermore, somatic mutations in serial plasma samples were monitored after treatment in four patients.

We identified multiple concordant somatic mutations in cfDNA and primary tumor samples from 10 patients (83.3%) and in cfDNA and metastatic tumor samples from one patient (100%). Furthermore, the allele frequency of the concordant mutations in cfDNA changed concomitantly with tumor burden and increased approximately 6 months earlier than the detection of tumor recurrences by imaging tests in two patients. Conventional biomarkers, such as SCC and p53-Ab, did not reflect tumor recurrences.

The present multigene panel, which enabled the diagnosis of tumor recurrence with greater accuracy than did using standard tumor markers or imaging methods, is expected to greatly facilitate standard, postoperative follow-up monitoring in ESCC.

c. 8q24 Polymorphisms and Diabetes Mellitus Regulate Apolipoprotein A-IV in Colorectal Carcinogenesis

Sugimachi K, Yamaguchi R, Eguchi H, Ueda M, Niida A, Sakimura S, Hirata H, Uchi R, Shinden Y, Iguchi T, Morita K, Yamamoto K, Miyano S, Mori M, Maehara Y, Mimori K

We explored the genetic interactions between diabetes and oncogenic single-nucleotide polymorphisms (SNPs) that determine colorectal cancer (CRC) morbidity.

8q24 rs6983267 polymorphism analysis and cDNA microarray were performed in 107 CRCs to identify the genes associated with diabetes and the oncogenic SNP. Then clinical significance of the gene was validated in 132 CRCs. Meta-analysis of microarray data and diabetic comorbidity was performed.

Of genes associated with a minor SNP allele at 8 q24, diabetes, and MYC overexpression, apolipoprotein A-IV (ApoA-IV) was associated with oncogenesis and poor prognosis in CRC patients. Patients with high ApoA-IV expression showed significantly poorer prognosis by univariate and multivariate analysis. Meta-analysis revealed lipid metabolism was associated with ApoA-IV-related oncogenesis in diabetic patients.

Changes in lipid metabolism associated with aberrant expression of ApoA-IV were risks for CRC oncogenesis.

d. Genomic landscape of esophageal squamous cell carcinoma in a Japanese population

Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, Shiraishi Y, Chiba K, Imoto S, Takahashi Y, Iwaya T, Sudo T, Hayashi T, Takai H, Kawasaki Y, Matsukawa T, Eguchi H, Sugimachi K, Tanaka F, Suzuki H, Yamamoto K, Ishii H, Shimizu M, Yamazaki H, Yamazaki M, Tachimori Y, Kajiyama Y, Natsugoe S, Fujita H, Mafune K, Tanaka Y, Kelsell DP, Scott CA, Tsuji S, Yachida S, Shibata T, Sugano S, Doki Y, Akiyama T, Aburatani H, Ogawa S, Miyano S, Mori M, Mimori K

Esophageal squamous cell carcinoma (ESCC) is the predominant form of esophageal cancer in Japan. Smoking and drinking alcohol are environmental risk factors for ESCC, whereas single nucleotide polymorphisms in ADH1B and ALDH2, which increase harmful intermediates produced by drinking alcohol, are genetic risk factors. We conducted a large-scale genomic analysis of ESCCs from patients in Japan to determine the mutational landscape of this cancer.

We performed whole-exome sequence analysis of tumor and nontumor esophageal tissues collected from 144 patients with ESCC who underwent surgery at 5 hospitals in Japan. We also performed single-nucleotide polymorphism array-based copy number profile and germline genotype analyses of polymorphisms in ADH1B and ALDH2. Polymorphisms in CYP2A6, which increase harmful effects of smoking, were analyzed. Functions of TET2 mutants were evaluated in KYSE410 and HEK293FT cells.

A high proportion of mutations in the 144 tumor samples were C to T substitution in CpG dinucleotides (called the CpG signature) and C to G/T substitutions with a flanking 5' thymine (called the APOBEC signature). Based on mutational signatures, patients were assigned to 3 groups, which associated with environmental (drinking and smoking) and genetic (polymorphisms in ALDH2 and CYP2A6) factors. Many tumors contained mutations in genes that regulate the cell cycle (TP53, CCND1, CDKN2A, FBXW7); epigenetic processes (MLL2, EP 300, CREBBP, TET2); and the NOTCH (NOTCH1, NOTCH3), WNT (FAT1, YAP1, AJUBA) and receptor-tyrosine kinase-phosphoinositide 3-kinase signaling pathways (PIK3CA, EGFR, ERBB2). Mutations in EP300 and TET2 correlated with shorter survival times, and mutations in ZNF750 associated

with an increased number of mutations of the APOBEC signature. Expression of mutant forms of TET2 did not increase cellular levels of 5-hydroxymethylcytosine in HEK293FT cells, whereas knockdown of TET2 increased the invasive activity of KYSE410 ESCC cells. Computational analyses associated the mutations in NFE2L2 we identified with transcriptional activation of its target genes.

We associated environmental and genetic factors with base substitution patterns of somatic mutations and provide a registry of genes and pathways that are disrupted in ESCCs. These findings might be used to design specific treatments for patients with esophageal squamous cancers.

e. Integrated Molecular Profiling of Human Gastric Cancer Identifies DDR2 as a Potential Regulator of Peritoneal Dissemination

Kurashige J, Hasegawa T, Niida A, Sugimachi K, Deng N, Mima K, Uchi R, Sawada G, Takahashi Y, Eguchi H, Inomata M, Kitano S, Fukagawa T, Sasako M, Sasaki H, Sasaki S, Mori M, Yanagihara K, Baba H, Miyano S, Tan P, Mimori K

Peritoneal dissemination is the most frequent, incurable metastasis occurring in patients with advanced gastric cancer (GC). However, molecular mechanisms driving peritoneal dissemination still remain poorly understood. Here, we aimed to provide novel insights into the molecular mechanisms that drive the peritoneal dissemination of GC. We performed combined expression analysis with in vivo-selected metastatic cell lines and samples from 200 GC patients to identify driver genes of peritoneal dissemination. The driver-gene functions associated with GC dissemination were examined using a mouse xenograft model. We identified a peritoneal dissemination-associated expression signature, whose profile correlated with those of genes related to development, focal adhesion, and the extracellular matrix. Among the genes comprising the expression signature, we identified that discoidin-domain receptor 2 (*DDR2*) as a potential regulator of peritoneal dissemination. The *DDR2* was upregulated by the loss of DNA methylation and that DDR2 knockdown reduced peritoneal metastasis in a xenograft model. Dasatinib, an inhibitor of the DDR 2 signaling pathway, effectively suppressed peritoneal dissemination. *DDR2* was identified as a driver gene for GC dissemination from the combined expression signature and can potentially serve as a novel therapeutic target for inhibiting GC peritoneal dissemination.

2. Implementation of Cancer Clinical Sequence

In collaboration with Professor Yoichi Furukawa (Division of Clinical Genome Research, Advanced Clinical Research Center), Professor Arinobu Tojo (Division of Molecular Therapy, Advanced Clinical Research Center), Research Associate Professor Koichiro Yuji (Project Division of International Advanced Medical Research), IMSUT Research Hospital, and Human Genome Center, we have been implementing cancer genomic medicine since 2011.

a. Reduced expression of APC-1B but not APC-1A by the deletion of promoter 1B is responsible for familial adenomatous polyposis

Yamaguchi K, Nagayama S, Shimizu E, Komura M, Yamaguchi R, Shibuya T, Arai M, Hatakeyama S, Ikenoue T, Ueno M, Miyano S, Imoto S, Furukawa Y

Germline mutations in the tumor suppressor gene APC are associated with familial adenomatous polyposis (FAP). Here we applied whole-genome sequencing (WGS) to the DNA of a sporadic FAP patient in which we did not find any pathological APC mutations by direct sequencing. WGS identified a promoter deletion of approximately 10?kb encompassing promoter 1B and exon1B of APC. Additional allele-specific expression analysis by deep cDNA sequencing revealed that the deletion reduced the expression of the mutated APC allele to as low as 11.2% in the total APC transcripts, suggesting that the residual mutant transcripts were driven by other promoter(s). Furthermore, cap analysis of gene expression (CAGE) demonstrated that the deleted promoter 1B region is responsible for the great majority of APC transcription in many tissues except the brain. The deletion decreased the transcripts of APC-1B to 39-45% in the patient compared to the healthy controls, but it did not decrease those of APC-1A. Different deletions including promoter 1B have been reported in FAP patients. Taken together, our results strengthen the evidence that analysis of structural variations in promoter 1B should be considered for the FAP patients whose pathological mutations are not identified by conventional direct sequencing.

3. Computational Kampo Medicine

a. Predicting Japanese Kampo formulas by analyzing database of medical records: a preliminary observational study

Yoshino T, Katayama K, Horiba Y, Munakata K, Yamaguchi R, Imoto S, Miyano S, Mima H, Watanabe K

Approximately 90% of physicians in Japan use Kampo medicine in daily practice. However, it is a challenge for physicians who do not specialize in Kampo medicine to select a proper Kampo formula

out of the 148 officially approved formulas, as the decision relies on traditional measurements and traditional medicine pattern diagnoses. The present study tries to evaluate the feasibility of a decision support system for frequently used Kampo formulas.

Our study included 393 patients who visited the Kampo Clinic at Keio University Hospital for the first time between May 2008 and March 2013. We collected medical records through a browser-based questionnaire system and applied random forests to predict commonly prescribed Kampo formulas.

The discriminant rate was the highest (87.0%) when we tried to predict a Kampo formula from two candidates using age, sex, body mass index, subjective symptoms, and the two essential and predictable traditional medicine pattern diagnoses (excess-deficiency and heat-cold) as predictor variables. The discriminant rate decreased as the candidate Kampo formulas increased, with the greatest drop occurring between three (76.7%) and four (47.5%) candidates. Age, body mass index, and traditional medicine pattern diagnoses had higher importance according to the characteristics of each Kampo formula when we utilized the prediction model, which predicted a Kampo formula from among three candidates.

These results suggest that our decision support system for non-specialist physicians works well in selecting appropriate Kampo formulas from among two or three candidates. Additional studies are required to integrate the present statistical analysis in clinical practice.

b. The Difference between the Two Representative Kampo Formulas for Treating Dysmenorrhea: An Observational Study

Yoshino T, Katayama K, Horiba Y, Munakata K, Yamaguchi R, Imoto S, Miyano S, Mima H, Watanabe K, Mimura M

In Kampo medicine, two different formulas are effective for treating dysmenorrhea—tokishakuyakusan and keishibukuryogan; however, the criteria by which specialists select the appropriate formula for each patient are not clear. We compared patients treated with tokishakuyakusan and those with keishibukuryogan and proposed a predictive model. The study included 168 primary and secondary dysmenorrhea patients who visited the Kampo Clinic at Keio University Hospital. We collected clinical data from 128 dysmenorrhea patients, compared the two patient groups and selected significantly different factors as potential predictors, and used logistic regression to establish a model. An external validation was performed using 40 dysmenorrhea patients. Lightheadedness, BMI$<$18.5, and a weak abdomen were significantly more frequent in the tokishakuyakusan group; tendency to sweat, heat intolerance, leg numbness, a cold sensation in the lower back, a strong abdomen, and paraumbilical tenderness and resistance were more frequent in the keishibukuryogan group. The final model fitted the data well. Internally estimated accuracy was 81.2%, and a leave-one-out cross-validation estimate of accuracy was 80.5%. External validation accuracy was 85.0%. We proposed a model for predicting the use of two Kampo formulas for dysmenorrhea, which should be validated in prospective trials.

Publications

1. Hasegawa T, Niida A, Mori T, Shimamura T, Yamaguchi R, Miyano S, Akutsu T, Imoto S. A likelihood-free filtering method via approximate Bayesian computation in evaluating biological simulation models. *Computational Statistics & Data Analysis.* 94: 63-74, 2016.
2. Kurashige J, Hasegawa T, Niida A, Sugimachi K, Deng N, Mima K, Uchi R, Sawada G, Takahashi Y, Eguchi H, Inomata M, Kitano S, Fukagawa T, Sasako M, Sasaki H, Sasaki S, Mori M, Yanagihara K, Baba H, Miyano S, Tan P, Mimori K. Integrated Molecular Profiling of Human Gastric Cancer Identifies DDR2 as a Potential Regulator of Peritoneal Dissemination. *Sci Rep.* 6: 22371, 2016.
3. Park H, Niida A, Imoto S, Miyano S. Interaction-Based Feature Selection for uncovering cancer driver genes through copy number-driven expression level. *J Comput Biol.* 2016 Oct 19. [Epub ahead of print] PubMed PMID: 27759426.
4. Sawada G, Niida A, Uchi R, Hirata H, Shimamura T, Suzuki Y, Shiraishi Y, Chiba K, Imoto S, Takahashi Y, Iwaya T, Sudo T, Hayashi T, Takai H, Kawasaki Y, Matsukawa T, Eguchi H, Sugimachi K, Tanaka F, Suzuki H, Yamamoto K, Ishii H, Shimizu M, Yamazaki H, Yamazaki M, Tachimori Y, Kajiyama Y, Natsugoe S, Fujita H, Mafune K, Tanaka Y, Kelsell DP, Scott CA, Tsuji S, Yachida S, Shibata T, Sugano S, Doki Y, Akiyama T, Aburatani H, Ogawa S, Miyano S, Mori M, Mimori K. Genomic landscape of esophageal squamous cell carcinoma in?a Japanese population. *Gastroenterology.* 150(5): 1171-1182, 2016.
5. Sugimachi K, Yamaguchi R, Eguchi H, Ueda M,

Niida A, Sakimura S, Hirata H, Uchi R, Shinden Y, Iguchi T, Morita K, Yamamoto K, Miyano S, Mori M, Maehara Y, Mimori K. 8q24 Polymorphisms and Diabetes Mellitus Regulate Apolipoprotein A-IV in Colorectal Carcinogenesis. *Ann Surg Oncol.* 23 (Suppl 4): 546-551, 2016.

6. Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, Sawada G, Iwaya T, Kurashige J, Shinden Y, Iguchi T, Eguchi H, Chiba K, Shiraishi Y, Nagae G, Yoshida K, Nagata Y, Haeno H, Yamamoto H, Ishii H, Doki Y, Iinuma H, Sasaki S, Nagayama S, Yamada K, Yachida S, Kato M, Shibata T, Oki E, Saeki H, Shirabe K, Oda Y, Maehara Y, Komune S, Mori M, Suzuki Y, Yamamoto K, Aburatani H, Ogawa S, Miyano S, Mimori K. Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution. *PLoS Genet.* 12(2): e1005778, 2016.
7. Ueda M, Iguchi T, Masuda T, Nakahara Y, Hirata H, Uchi R, Niida A, Momose K, Sakimura S, Chiba K, Eguchi H, Ito S, Sugimachi K, Yamasaki M, Suzuki Y, Miyano S, Doki Y, Mori M, Mimori K. Somatic mutations in plasma cell-free DNA are diagnostic markers for esophageal squamous cell carcinoma recurrence. *Oncotarget.* 7(38): 62280-62291, 2016.
8. Yamaguchi K, Nagayama S, Shimizu E, Komura M, Yamaguchi R, Shibuya T, Arai M, Hatakeyama S, Ikenoue T, Ueno M, Miyano S, Imoto S, Furukawa Y. Reduced expression of APC-1B but not APC-1A by the deletion of promoter 1B is responsible for familial adenomatous polyposis. *Sci Rep.* 6: 26011, 2016.
9. Yoshino T, Katayama K, Horiba Y, Munakata K, Yamaguchi R, Imoto S, Miyano S, Mima H, Watanabe K. Predicting Japanese Kampo formulas by analyzing database of medical records: a preliminary observational study. *BMC Med Inform Decis Mak.* 16: 118, 2016.
10. Yoshino T, Katayama K, Horiba Y, Munakata K, Yamaguchi R, Imoto S, Miyano S, Mima H, Watanabe K, Mimura M. The Difference between the Two Representative Kampo Formulas for Treating Dysmenorrhea: An Observational Study. *Evid Based Complement Alternat Med.* 2016: 3159617, 2016.