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

Imoto S, Miyano S, Yamaguchi R, Shibuya T, Hasegawa T, Shimizu E, Komura M, Saito A, Yamaguchi K, Yokoyama K, Kobayashi M, Hamada K, Yuji K, Muto K, Tojo A, Furukawa Y

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 interpretation of the results of genome analyses of cancer patients. The results of genome sequence analysis including the interpretation of IBM Watson are evaluated and discussed in biweekly sequence board meeting. In 2015, we analyzed around sequence data of 100 cancer patients with whole genome, exome, and/or target deep sequencings. Also, multi-omics data 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. Computational Methods in Systems Biology and Immunology

a. Approximate Bayesian computation for evaluating biological simulation models

Hasegawa T, Niida A, Moria T[1], Shimamura T[2], Yamaguchi R, Miyano S, Akutsu T[1], Imoto S: [1]Bioinformatics Center, Institute for Chemical Research, Kyoto University, [2]Division of Systems Biology, Nagoya University Graduate School of Medicine

For the evaluation of the dynamic behavior of biological processes, e.g., gene regulatory sequences, we typically utilize nonlinear differential equations

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

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 genotyping of Human leukocyte antigen (HLA) based on whole exome sequencing (WES), amplicon and RNA sequencing (RNA-seq) data has been addressed in recent few years, HLA genotyping with whole genome sequencing (WGS) data is still challenging problem. We developed a computational method, termed ALPHLARD, that can collect reads potentially generated from HLA regions, and accurately determines the pair of HLA types for each of HLA-A, -B and -C genes with 4-digit or higher resolution by a novel Bayesian model. Furthermore, ALPHLARD can call rare germline variants that are not stored in the HLA database and somatic mutations of cancer patient from the pair of normal-tumor WGS data. We illustrate the performance of ALPHARD using 25 WGS data sets from an illumina platform. Using the results of HLA genotyping from SBT and amplicon sequencing methods, ALPHLARD achieved 98.0% accuracy at 6-digit resolution. Also, three somatic mutations we found were correctly validated by the Sanger sequencing. This research is performed as a part of International Cancer Genome Consortium PanCancer Analysis Project.

Publications

1. Hasegawa T, Niida A, Moria 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 and Data Analysis*, 94, 63-74 (2016).
2. 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*, first published online (September 28, 2015).

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(兼)
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. Global Genomic Medicine Collaborative (G2 MC)

a. Global implementation of genomic medicine: We are not alone

Manolio TA[1], Abramowicz M[2], Al-Mulla F[3], Anderson W[4], Balling R[5], Berger AC[6], Bleyl S[7], Chakravarti A[8], Chantratita W[9], Chisholm RL[10], Dissanayake VH[11], Dunn M[12], Dzau VJ[13], Han BG[14], Hubbard T[15], Kolbe A[16], Korf B[17], Kubo M[18], Lasko P[19], Leego E[20], Mahasirimongkol S[21], Majumdar PP[22], Matthijs G[23], McLeod HL[24], Metspalu A[20], Meulien P[25], Miyano S[26], Naparstek Y[27], O'Rourke PP[28], Patrinos GP[29], Rehm HL[30], Relling MV[31], Rennert G[32], Rodriguez LL[33], Roden DM[34], Shuldiner AR[35], Sinha S[36], Tan P[37], Ulfendahl M[38], Ward R[39], Williams MS[40], Wong JE[41], Green ED[33], Ginsburg GS[42]: [1]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-9305, USA, [2]Université Libre de Bruxelles 1070 Brussels, Belgium, [3]Genatak-Global Med Clinic, Kuwait University, Kuwait 46300, Kuwait, [4]National Health and Medical Research Council, Canberra, ACT 2601, Australia, [5]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, L-4362 Luxembourg, [6]Board on Health Sciences Policy, Institute of Medicine, Washington, DC 20001, USA, [7]Intermountain Healthcare, Salt Lake City, UT 84111, USA, [8]McKusick-Nathans Institute of Genetic Medicine, John Hopkins University School of Medicine, Baltimore, MD 21205, USA, [9]Ramathibodi Hospital, Mahidol University, Bangkok, 10400 Thailand, [10]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA, [11]Human Genetics Unit, Faculty of Medicine, University of Colombo, Colombo 00800, Sri Lanka, [12]Genetic and Molecular Sciences, The Wellcome Trust, London NW1 2 BE, UK, [13]National Academy of Medicine, Washington, DC 20001, USA, [14]Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do 363-951 Korea, [15]Department of Medical and Molecular Genetics, King's College, London SE1 9RT, and Genomics England, London EC1M 6BQ, UK, [16]National Health Committee, Auckland 1050, New Zealand, [17]Center for Genomic Science, University of Alabama at Birmingham, Birmingham, AL 35294, USA, [18]Center for Integrative Medical Science (IMS), RIKEN, Yokohama 230-0045, Japan, [19]Institute of Genetics, Canadian Institutes of Health Research, and McGill University, Montreal, Quebec, H3A 0G4 Canada, [20]Estonian Genome Center, University of Tartu, Tartu 51010 Estonia, [21]Department of

Medical Science, Ministry of Public Health, Nonthaburi 11000 Thailand, [22]National Institute of Biomedical Genomics and Indian Statistical Institute, Kalyani 741251 India, [23]Center for Human Genetics, University of Leuven (KU Leuven), B-3000 Leuven, Belgium, [24]DeBartolo Family Personalized Medicine Institute, Moffitt Cancer Center, Tampa, FL 33612 USA, [25]Genome Canada, Ottawa, Ontario K2P 1P1, Canada, [26]Institute of Medical Science, University of Tokyo, 108-8639 Tokyo, Japan, [27]Research and Academic Affairs, Hadassah Hebrew University Hospital, Jerusalem 91120, Israel, [28]Office of Human Research Affairs, Partners HealthCare, Boston, MA 02199, USA, [29]Department of Pharmacy, School of Health Sciences, University of Patras, Patras, 26504 Greece, [30]Laboratory for Molecular Medicine, Partners Healthcare Systems, Boston, MA 02139, USA, [31]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA, [32]Carmel Medical Center Department of Community Medicine and Epidemiology, Clalit National Personalized Medicine Program, Haifa 34362, Israel, [33]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-9305, USA, [34]Vanderbilt University School of Medicine, Nashville, TN 37232, USA, [35]Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA, [36]Department of Biotechnology, Ministry of Science and Technology, Govt., New Delhi 110 003 India, [37]Duke-National University of Singapore Graduate Medical School, Singapore 169857, Singapore, [38]Swedish Research Council, SE-101 38 Stockholm, Sweden, [39]University of Queensland, St. Lucia, QLD 4067 Australia, [40] Genomic Medicine Institute, Geisinger Health System, Danville, PA 18510, USA, [41]National University of Singapore, Singapore 119228, Singapore, [42]Center for Applied Genomics and Precision Medicine, Duke University, Durham, NC 27710, USA. manolio@nih.gov geoffrey.ginsburg@dm.duke.edu

We have been involved with the Genomic Medicine Meetings organized by National Institute of Health, USA (http://www.genome.gov/27549225), since 2013. The first message was published as a summary of our meeting discussions. G2MC's home page: http://iom.nationalacademies.org/activities/research/genomicbasedresearch/innovation-collaboratives/global_genomic_medicine_collaborative.aspx

Around the world, innovative genomic-medicine programs capitalize on singular capabilities arising from local health care systems, cultural or political milieus, and unusual selected risk alleles or disease burdens. Such individual efforts might benefit from the sharing of approaches and lessons learned in other locales. The U.S. National Human Genome Research Institute and the National Academy of Medicine recently brought together 25 of these groups to compare projects, to examine the current state of implementation and desired near-term capabilities, and to identify opportunities for collaboration that promote the responsible practice of genomic medicine. Efforts to coalesce these groups around concrete but compelling signature projects should accelerate the responsible implementation of genomic medicine in efforts to improve clinical care worldwide.

The international landscape of genomic medicine was explored from the viewpoints: (1) Haves and have nots. Selected current and desired genomic medicine capabilities across participating countries and regions (number surveyed = 25). (2) Genomic medicine: Barriers to implementation. (3) Opportunities for international collaborations in regard to evidence generation, health information technology, education / workforce development, pharmacogenomics, policy for (i) data sharing and regulatory issues and (ii) costs and benefits. Then multinational teamworks were organized for the above items.

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. Attenuated familial adenomatous polyposis with desmoids caused by an APC mutation

Ikenoue T[43], Yamaguchi K[43], Koumura M, Imoto S, Yamaguchi R, Shimizu E, Kasuya S, Shibuya T, Hatakeyama S[43], Miyano S, Furukawa Y[43]: [43]Division of Clinical Genome Research, Advanced Clinical Research Center

As a result of our Cancer Clinical Sequence Project, we present a case of attenuated familial adenomatous polyposis (AFAP) with a family history of desmoids and thyroid tumors. This patient had no colonic polyps but did have multiple desmoids. Genetic analysis identified a 4-bp deletion in codon 2644 (c.7932_7935delTTAT: p.Tyr2645LysfsX14) of the adenomatous polyposis coli (APC) gene. In cases with limited numbers of colonic polyps and desmoids, AFAP may be caused by a mutation in the 3′ region of APC.

b. Detection of APC mosaicism by next-generation sequencing in an FAP patient

Yamaguchi K[43], Komura M, Yamaguchi R, Imoto S, Shimizu E, Kasuya S, Shibuya T, Hatakeyama S[43], Takahashi N[43], Ikenoue T[43], Hata K[44], Tsurita G[44], Shinozaki M[44], Suzuki Y[44], Sugano S[46], Miyano S, Furukawa Y[43]: [44]Department of Surgery, Research Hospital, Institute of Medical Science, The University of Tokyo, [45]Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, [46]Laboratory of Functional Genomics, Graduate School of Frontier Sciences, The University of Tokyo

As a result of our Cancer Clinical Sequence Project based on whole genome sequencing, the following result is obtained. Familial adenomatous polyposis (FAP) of the colon is characterized by multiple polyps in the intestine and extra-colonic manifestations. Most FAP cases are caused by a germline mutation in the tumor-suppressor gene APC, but some cases of adenomatous polyposis result from germline mutations in MUTYH, POLD1 or POLE. Although sequence analysis of APC by the Sanger method is routinely performed for genetic testing, there remain cases whose mutations are not detected by the analysis. Next-generation sequencing has enabled us to analyze the comprehensive human genome, improving the chance of identifying disease causative variants. In this study, we conducted whole-genome sequencing of a sporadic FAP patient in which we did not find any pathogenic APC mutations by the conventional Sanger sequencing. Whole-genome sequencing and subsequent deep sequencing identified a mosaic mutation of c.3175G>T, p.E1059X in ~12% of his peripheral leukocytes. Additional deep sequencing of his buccal mucosa, hair follicles, non-cancerous mucosa of the stomach and colon disclosed that these tissues harbored the APC mutation at different frequencies. Our data implied that genetic analysis by next-generation sequencing is an effective strategy to identify genetic mosaicism in hereditary diseases.

Publications

1. Ikenoue T, Yamaguchi K, Koumura M, Imoto S, Yamaguchi R, Shimizu E, Kasuya S, Shibuya T, Hatakeyama S, Miyano S, Furukawa Y. Attenuated familial adenomatous polyposis with desmoids caused by an APC mutation. *Human Genome Variation*. 2: 15011, 2015.
2. Manolio TA, Abramowicz M, Al-Mulla F, Anderson W, Balling R, Berger AC, Bleyl S, Chakravarti A, Chantratita W, Chisholm RL, Dissanayake VH, Dunn M, Dzau VJ, Han BG, Hubbard T, Kolbe A, Korf B, Kubo M, Lasko P, Leego E, Mahasirimongkol S, Majumdar PP, Matthijs G, McLeod HL, Metspalu A, Meulien P, Miyano S, Naparstek Y, O'Rourke PP, Patrinos GP, Rehm HL, Relling MV, Rennert G, Rodriguez LL, Roden DM, Shuldiner AR, Sinha S, Tan P, Ulfendahl M, Ward R, Williams MS, Wong JE, Green ED, Ginsburg GS. Global implementation of genomic medicine: We are not alone. *Sci Transl Med*. 7(290): 290ps13, 2015.
3. Yamaguchi K, Komura M, Yamaguchi R, Imoto S, Shimizu E, Kasuya S, Shibuya T, Hatakeyama S, Takahashi N, Ikenoue T, Hata K, Tsurita G, Shinozaki M, Suzuki Y, Sugano S, Miyano S, Furukawa Y. Detection of APC mosaicism by next-generation sequencing in an FAP patient. *J Hum Genet*. 60(5): 227-231, 2015.