Curriculum Vitae

Personal details

Name: Natsuhiko KUMASAKA, PhD

Address: Wellcome Sanger Institute
Wellcome Genome Campus
Hinxton, Cambridge, CB10 1SA
UK

Contact: Phone: +44 0 122 349 6866
Email: nk5@sanger.ac.uk
http://kumasakanatsuhiko.jp

Position: Staff Scientist

Academic qualifications

BSc (Engineering), Department of Mathematics, Keio University, March 2003.
MSc (Engineering), School of Fundamental Science and Technology, Keio University, March 2005.
PhD (Science), School of Fundamental Science and Technology, Keio University, September 2007.

Research and professional specialities

Applied statistics; Population genetics; Cellular genomics; Bioinformatics; Data visualisation.

Employment

Research assistant at Information Technology Center (ITC), Faculty of Science and Technology, Keio University, Japan. (Apr. 2005 – Sep. 2007)

Postdoctoral Research Fellow of the 21st Century Centre of Excellence (COE) Programme "Integrative Mathematical Sciences" at Keio University, Japan. (Oct. 2007 – Mar. 2008)

Postdoctoral researcher at Center for Genomic Medicine, RIKEN, Japan. (Apr. 2008 – Mar. 2010).

Research scientist at Center for Genomic Medicine, RIKEN, Japan. (Mar. 2010 – Mar. 2012)

Postdoctoral fellow at Wellcome Sanger Institute, UK. (Apr. 2012 – Mar. 2016)

Staff scientist at Wellcome Sanger Institute, UK. (Apr. 2016 – present)

Honours / distinctions

Visiting Research Fellow at Department of Computer-Oriented Statistics and Data Analysis, Augsburg University, Germany (Professor Antony Unwin) (Oct. 2007 – Dec. 2007)

Best Presentation Award, the 2009 Japanese Joint Statistical Meeting (Sep. 2009).

Sanger Early Career Innovation Awards "High-throughput analysis of chromatin accessibility QTLs with smaller numbers of samples and cells using assay for transposase-accessible chromatin followed by sequencing" (Nov. 2013)

Research experience

At Keio University (Apr. 2003 – Mar. 2008): I had worked on data visualisation and proposed a high-dimensional (multivariate) data visualisation, called *Textile Plot* ([1, 2, 3, 4]). I was also involved in the DandD (Data and Description) project to develop a new environment for working with data.

At RIKEN (Apr. 2008 – Mar. 2012): I had extended the textile plot to Single Nucleotide Polymorphism (SNP) data [5]. I had also proposed a complementary approach to principal component analysis so that one can perform population structure analyses using the limited numbers of subjects and SNPs [6]. I had developed a genotyping method of Copy Number Polymorphisms (CNPs) based on a Bayesian Gaussian mixture model [7] followed by the first CNP based genome-wide association study of smoking quantity in a Japanese population [8]. I had also been involved in various genome-wide association studies of common diseases and other quantitative traits [9, 10, 11, 12, 13].

At Wellcome Sanger Institute (Apr. 2012 – present): In 2012, I have joined Wellcome Sanger Institute as a Postdoctoral fellow. Until 2014, I had been involved in a project revealing transcriptional variation among human induced pluripotent stem cells (hIPSCs) derived from different tissues along with human ES cells. Our study demonstrated that most cellular variability between hIPSCs from different individuals reflects genetic differences rather than cellular reprogramming or residual memory of the tissue of origin [14]. From 2014, I had been developing a new statistical approach, called RASQUAL, for association mapping that models genetic effects and accounts for biases in sequencing data in a single probabilistic framework. We had applied this method to generate a map of chromatin accessibility QTLs and showed that it improves fine-map regulatory variants and links distal regulatory elements with gene promoters [15].

I am currently working on mapping distal regulatory interactions using the chromatin accessibility QTLs. I have developed a novel statistical approach, a pairwise hierarchical model, which uses a principle of Mendelian Randomisation embedded within a Bayesian hierarchical model. We demonstrated that our model can be used to identify hierarchies of regulatory elements within a genetic locus and prioritise putative causal variants in genome-wide association studies, validating the BLK locus using CRISPR/Cas9 editing [16].

Educational experience

At Keio University: Teaching assistant at Department of Mathematics (under Professor Ritei Shibata):

Statistical Science and Its Exercise, data analysis using S-PLUS/R/Textile Plot, 3yr undergrad Spring (in Japanese) (2003 – 2007)

Data Analysis and Its Exercise, data analysis using S-PLUS/R/Textile Plot, 3yr undergrad Autumn (in Japanese) (2003 – 2007)

At RIKEN: Teaching assistant at The Institute of Medical Science, The University of Tokyo (under Dr. Naoyuki Kamatani):

Primer of Statistical Genetics -linear and logistic regression methods- (in English) (8th Jul. 2010)

Professional membership / activities

Member of the Japan Statistical Society (2003 – 2012).

Member of the Japan Society of Human Genetics (2008 – 2012).

Member of the International Genetic Epidemiology Society (2009 – 2012).

Publications

1. Kumasaka, N., Shibata, R. (2006) Implementation of textile plot. *Proceedings in COMPSTAT 2006* , 581-9.
2. Kumasaka, N., Shibata, R. (2007) The textile plot environment (in Japanese). *Proceedings of the Institute of Statistical Mathematics, Special Issue: Statistical Data Visualization* **55**, 47-68.
3. Kumasaka, N., Shibata, R. (2008) High-dimensional data visualisation: the Textile Plot. *Computational Statistics & Data Analysis* **52**, 3616-3644.
4. Kumasaka, N. (2008) ANOVA on the textile plot. *Proceedings in COMPSTAT 2008* , 79-87.
5. Kumasaka, N., Nakamura, Y., Kamatani, N. (2010) The Textile Plot: a new linkage disequilibrium display of multiple-single nucleotide polymorphism genotype data. *PLoS ONE* **5**, e10207.
6. Kumasaka, N., Yamaguchi-Kabata, Y., Takahashi, A., Kubo, M., Nakamura, Y., Kamatani, N. (2010) Establishment of a standardized system to perform population structure analyses with limited sample size or with different sets of snp genotypes. *Journal of Human Genetics* **55**, 525-33.
7. Kumasaka, N., Fujisawa, N., Hosono, N., Okada, Y., Takahashi, A., Nakamura, Y., Kubo, M., Kamatani, N. (2011) PlatinumCNV: a Bayesian Gaussian mixture model for genotyping copy number polymorphisms using SNP array signal intensity data. *Genet. Epidemiol.* **35**, 831-844.
8. Kumasaka, N., Aoki, M., Okada, Y., Takahashi, A., Ozaki, K., Mushiroda, T., Hirota, T., Tamari, M., Tanaka, T., Nakamura, Y., et al. (2012) Haplotypes with copy number and single nucleotide polymorphisms in CYP2A6 locus are associated with smoking quantity in a japanese population. *PLoS ONE* **7**, e44507.
9. Nuinoon, M., Makarasara, W., Mushiroda, T., Setianingsih, I., Wahidiyat, P. A., Sripichai, O., Kumasaka, N., Takahashi, A., Svasti, S., Munkongdee, T., et al. (2010) A genome-wide association identified the common genetic variants influence disease severity in β^0-thalassemia/hemoglobin e. *Human Genetics* **127**, 303-14.

10. Okada, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Kamatani, Y., Hosono, N., Tsunoda, T., Matsuda, K., Tanaka, T., Kubo, M., et al. (2011) Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Hum Mol Genet* **20**, 1224-31.

11. Okada, Y., Hirota, T., Kamatani, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Higasa, K., Yamaguchi-Kabata, Y., Hosono, N., Nalls, M. A., et al. (2011) Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.* **7**, e1002067.

12. Okada, Y., Yamazaki, K., Umeno, J., Takahashi, A., Kumasaka, N., Ashikawa, K., Aoi, T., Takazoe, M., Matsui, T., Hirano, A., et al. (2011) HLA-Cw*1202-B*5201-DRB1*1502 haplotype increases risk for ulcerative colitis but reduces risk for Crohn's disease. *Gastroenterology* **141**, 864-71.

13. Yamazaki, K., Umeno, J., Takahashi, A., Hirano, A., Johnson, T. A., Kumasaka, N., Morizono, T., Hosono, N., Kawaguchi, T., Takazoe, M., et al. (2012) A genome-wide association study identifies 2 susceptibility loci for Crohn's disease in a Japanese population. *Gastroenterology* .

14. Rouhani, F., Kumasaka, N., de Brito, M. C., Bradley, A., Vallier, L., Gaffney, D. (2014) Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLoS Genet.* **10**, e1004432.

15. Kumasaka, N., Knights, A., Gaffney, D. (2016) Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nature Genet.* **48**, 206-213.

16. Kumasaka, N., Knights, A. J., Gaffney, D. J. (2019) High-resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nat. Genet.* **51**, 128–137.

Conference Papers (most recent)

1. N. Kumasaka, F. Rouhani, L. Vallier, A. Bradley, D. Gaffney (2012) Transcriptional and epigenetic variation in human induced pluripotent stem cells (Abstract/Program 3507T); Presented at the 62nd Annual Meeting of The American Society of Human Genetics, November 8, 2012 in San Francisco, California (Poster presentation).

2. N. Kumasaka, F. Rouhani, A. Bradley, L. Vallier, D. Gaffney (2013) Genetic variation is a major source of transcriptional variation in human induced pluripotent stem cells. (Abstract/Program 185); Presented at the 63rd Annual Meeting of The American Society of Human Genetics, October 24, 2013 in Boston, MA (Platform presentation).

3. N. Kumasaka, A. Knights, D. Gaffney (2014) Combining allele-specific and population signal boosts power for association mapping of multiple DNA sequence-based cellular traits. (Abstract/Program 1881S); Presented at the 64th Annual Meeting of The American Society of Human Genetics, October 19, 2014 in San Diego, California (Poster presentation).

4. N. Kumasaka, A. Knights, D. Gaffney (2015) Fine-mapping cellular trait QTLs with RASQUAL and ATAC-seq. (Abstract/Program 167); Presented at the 65th Annual Meeting of The American Society of Human Genetics, October 8, 2015 in Baltimore, MD (Platform presentation).