Curriculum Vitae

Rintaro Saito

University of California, San Diego
Departments of Medicine and Bioengineering

February 28, 2011

1 Profile

Family Name	Saito
Given Name	Rintaro
Gender	Male
Date of Birth	November 5, 1972
Nationality	Japanese
Address	San Diego, California, U.S.A.
Marital Status	Married
Degree	Ph.D. (Keio University, 2000)
Affiliation	University of California, San Diego Departments of Medicine and Bioengineering (Postal address: 9500 Gilman Drive La Jolla, CA 92093-0688)
Title	Visiting Faculty
Research Field	Bioinformatics
Research Interests	Computational genomic analyses (especially on non-coding regions) Interactome analyses
Tel	+1-858-822-4667
Fax	+1-858-822-4246
E-Mail	risaito@ucsd.edu
Web Page	http://www.bioinfo.sfc.keio.ac.jp/class/bioinfo-a/WEB_RS
Skype ID	golgo8028
Hobby	Skiing, Driving, GO (Traditional Asian board game)

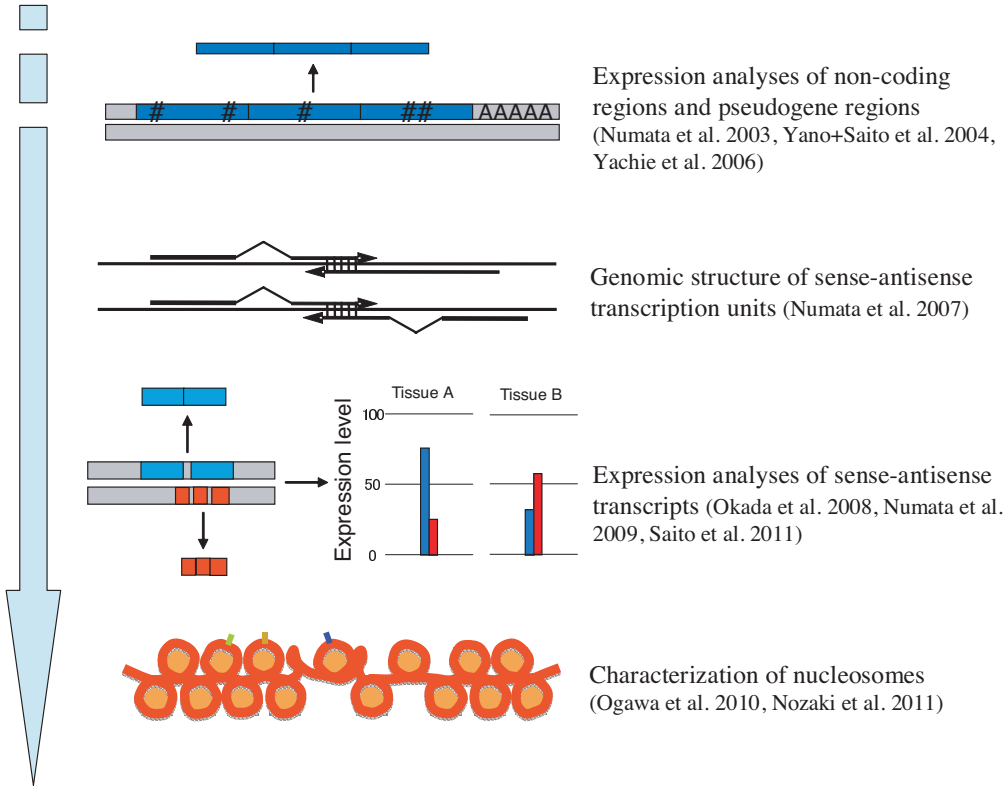

Figure 1: History of our research on genome sequence analyses

2 Researches

2.1 Characterization of non-coding genomic regions

Genomic sequences contain not only genes but also their regulatory information as well as information about how the organisms have evolved at the molecular level. We have especially been focusing on non-coding regions of the genomes and worked on computational characterization of these regions to investigate their functionality and to find how they have emerged (Figure 1) .

Although the human genome contains $> 97\%$ of non-protein-coding regions, their functions remain unclear. Our first step was to demonstrate that much of these regions are transcribed. After joining the FANTOM project at RIKEN Genomic Sciences Center, we computationally analyzed 60,770 of mouse full-length cDNA sequences and surprisingly found that there are thousands of non-coding transcripts which do not encode proteins in the cell [23,30]. Our analyses also showed that at least a small percentage of pseudogenes, which had been thought as inactive gene fossils, are expressed in human and mouse cell [20]. Subsequently, we developed a computational method

(based on markov model which can assess RNA folding potential) to predict non-coding RNAs in *E. coli* [16]. The expression of some of non-coding RNA candidates predicted by our method was experimentally validated by RT-PCR and Northern blotting.

Our next focus was characterization of antisense RNAs in various species, as we have found that some of non-coding predicted RNAs were encoded in the antisense strand of protein-coding genes [16]. We mapped cDNA sequences onto the genomes to find that there are thousands of such sense-antisense structures in various species [13]. We then analyzed expression patterns of these sense-antisense transcripts using our custom microarray [1,3,6,9]. We designed microarray probes specific to the antisense strand of known genes and discovered hundreds of novel antisense transcripts in human and mouse. The expression levels of some of them were significantly altered in several tissues including cancer tissues, giving insights into the roles of these antisense transcripts.

We are now integrating epigenetic information to characterize genomic sequences; nowadays, a huge amount of epigenetic information about DNA methylations and histone modifications is available. We started with computational prediction of nucleosome positions on the genomes based on genomic sequence patterns [5]. We are analyzing correlations between genomic modification patterns, genomic features and features of transcripts encoded in the corresponding genomic region [2].

In addition, we are also interested in translation signals in mRNA sequences which was my main research focus during my doctoral course in the late 1990's. We found characteristic sequence patterns around start codons which are related to the mechanism and evolution of translation initiation [39]. We have extended these analyses to characterization of translation initiation signals using free energy [38] and information theory [15]. We also have worked on computational analyses of translation termination signals [26,28], codon biases [7, 10, 18, 19] and upstream Open Reading Frames (ORFs) [11].

Related publications

[*] Marked authors contributed equally to the work, [+] Corresponding author. All reviewed by referees.

1. Saito R[*], Kohno K[*], Okada Y, Osada Y, Numata K, Kohama C, Watanabe K, Nakaoka H, Yamamoto N, Kanai A, Yasue H, Murata S, Abe K, Tomita M, Ohkohchi N, Kiyosawa H (2011) **Comprehensive Expressional Analyses of Antisense Transcripts in Colon Cancer Tissues Using Artificial Antisense Probes**. *BMC Medical Genomics* (accepted)

2. Nozaki T, Yachie N, Ogawa R, Saito R[+], Tomita M (2011) **Computational analysis suggests a highly bendable, fragile structure for nucleosomal DNA**. *Gene* (in press)

3. Watanabe Y, Numata K, Murata S, Osada Y, Saito R, Nakaoka H, Yamamoto N, Watanabe K, Kato H, Abe K, Kiyosawa H (2010) **Genome-wide analysis of expression modes and DNA methylation status at sense-antisense transcript loci in mouse**. *Genomics* 96(6):333-41

4. Kratz A, Arner E, Saito R, Kubosaki A, Kawai J, Suzuki H, Carninci P, Arakawa T, Tomita M, Hayashizaki Y, Daub CO. **Core promoter structure and genomic context reflect histone 3 lysine 9 acetylation patterns**. *BMC Genomics* 11:257

5. Ogawa R, Kitagawa N, Ashida H, Saito R^{+}, Tomita M(2010) **Computational prediction of nucleosome positioning by calculating the relative fragment frequency index of nucleosomal sequences**. *FEBS Let* 584(8):1498-502

6. Numata K, Osada Y, Okada Y, Saito R, Hiraiwa N, Nakaoka H, Yamamoto N, Watanabe K, Okubo K, Kohama C, Kanai A, Abe K, Kiyosawa H.(2009) **Identification of novel endogenous antisense transcripts by DNA microarray analysis targeting complementary strand of annotated genes**. *BMC Genomics* 10:392

7. Suzuki H, Saito R^{+}, Tomita M.(2009) **Measure of synonymous codon usage diversity among genes in bacteria**. *BMC Bioinformatics* 10:167.

8. FANTOM Consortium, Suzuki H, $\cdots$103 authors$\cdots$, Saito R, $\cdots$32 authors$\cdots$, Hume DA; Riken Omics Science Center, Arakawa T, $\cdots$20 authors$\cdots$, Hayashizaki Y.(2009) **The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line**. *Nat Genet* 41(5):553-62.

9. Okada Y, Tashiro C, Numata K, Watanabe K, Nakaoka H, Yamamoto N, Okubo K, Ikeda R, Saito R^{+}, Kanai A, Abe K, Tomita M, Kiyosawa H^{+} (2008) **Comparative expression analysis uncovers novel features of endogenous antisense transcription**. *Hum Mol Genet* 17(11):1631-40

10. Suzuki H, Saito R^{+}, Tomita M (2007) **Variation in the Correlation of G + C Composition with Synonymous Codon Usage Bias among Bacteria**. *EURASIP J Bioinform Syst Biol* 61374

11. Matsui M, Yachie N, Okada Y, Saito R^{+}, Tomita M.(2007) **Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse**. *FEBS Let* 581(22):4184-8

12. Arakawa K, Saito R^{+}, Tomita M. (2007) **Noise-reduction filtering for accurate detection of replication termini in bacterial genomes**. *FEBS Let* 581(2):253-8

13. Numata K, Okada Y, Saito R^{+}, Kiyosawa H, Kanai A, Tomita M. (2007) **Comparative analysis of cis-encoded antisense RNAs in eukaryotes**. *Gene* 392(1-2):134-41

14. Mori K, Saito R^{+}, Kikuchi S^{+}, Tomita M (2006) **Inferring rules of *E. coli* translational efficiency using an artificial neural network**. *Biosystems* 90(2):414-420

15. Osada Y, Saito R^{+}, Tomita M (2006) **Comparative analysis of base correlations in 5' untranslated regions of various species**. *Gene* 375:80-6

16. Yachie N, Numata K, Saito R^{+}, Kanai A, Tomita M (2006) **Prediction of non-coding and antisense RNA genes in *Escherichia coli* with Gapped Markov Model**. *Gene* 372:171-81

17. Watanabe Y, Yachie N, Numata K, Saito R, Kanai A, and Tomita M (2005) **Computational analysis of microRNA target recognition in** *Caenorhabditis elegans*. *Gene* 365:2-10

18. Suzuki H, Saito R+, Tomita M.(2005) **A problem in multivariate analysis of codon usage data and a possible solution**. *FEBS let* 579(28):6499-504

19. Suzuki H, Saito R+, Tomita M.(2004) **The 'weighted sum of relative entropy': a new index for synonymous codon usage bias**. *Gene*. 335:19-23.

20. Yano Y*, Saito R*, Yoshida N, Yoshiki A, Wynshaw-Boris A, Tomita M, Hirotsune S(2004) **A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene**. *J Mol Med* 82(7):414-22

21. Kikuchi S, ···64 authors···, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y; Rice Full-Length cDNA Consortium; National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team; Foundation of Advancement of International Science Genome Sequencing & Analysis Group; RIKEN. (2003) **Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice**. *Science* 301(5631):376-9.

22. Carninci P, Waki K, Shiraki T, Konno H, Shibata K, Itoh M, Aizawa K, Arakawa T, Ishii Y, Sasaki D, Bono H, Kondo S, Sugahara Y, Saito R, ···32 authors···, Hayashizaki Y. (2003) **Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia**. *Genome Res* 13(6B):1273-89

23. Numata, K., Kanai, A., Saito R., Kondo, S., Adachi, J., Wilming L. G., Hume, D. A., RIKEN GER Group Members, Hayashizaki, Y. and Tomita, M.(2003) **Identification of putative non-coding RNAs amongst the RIKEN mouse full-length cDNA collection**. *Genome Res* 13(6B):1301-1306.

24. Furuno M, Kasukawa T, Saito R, Adachi J, Suzuki H, Baldarelli R, Hayashizaki Y, Okazaki Y.(2003) **CDS annotation in full-length cDNA sequence**. *Genome Res* 13(6B):1478-87.

25. Nagashima T, Silva DG, Petrovsky N, Socha LA, Suzuki H, Saito R, Kasukawa T, Kurochkin IV, Konagaya A, Schonbach C.(2003) **Inferring higher functional information for RIKEN mouse full-length cDNA clones with FACTS**. *Genome Res* 13(6B):1520-33.

26. Ozawa, Y., Saito R., Washio, T, and Tomita, M.(2003) **Comparative study of translation termination sites and release factors (RF1 and RF2) in prokaryotes**. *J Mol Evol* 56(6):665-72

27. Sato, M., Umeki, H., Saito R+, Kanai, A. and Tomita, M. (2003) **Computational analysis of stop codon readthrough in** *D. melanogaster*. *Bioinformatics* 19(11):1371-80

28. Ozawa Y, Hanaoka S, Saito R, Washio T, Nakano S, Shinagawa A, Itoh M, Shibata K, Carninci P, Konno H, Kawai J, Hayashizaki Y, Tomita M.(2002) **Comprehensive sequence analysis of translation termination sites in various eukaryotes**. *Gene* 300(1-2):79-87.

29. Sakurai A, Fujimori S, Kochiwa H, Kitamura-Abe S, Washio T, Saito R, Carninci P, Hayashizaki Y, Tomita M. (2002) **On biased distribution of introns in various eukaryotes**. *Gene* 300(1-2):89-95.

30. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, ···128 authors···, Hayashizaki Y (2002) **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs**. *Nature 420*: 563-573.

31. Kochiwa H, Suzuki R, Washio T, Saito R, Bono H, Carninci P, Okazaki Y, Miki R, Hayashizaki Y, Tomita, M. (2002) **Inferring alternative splicing patterns in mouse from a full-length cDNA library and microarray data**. **Genome Res** 12(8): 1286-93

32. Sakai H, Washio T, Saito R, Shinagawa A, Itoh M, Shibata K, Carninci P, Konno H, Kawai J, Hayashizaki Y, Tomita M.(2001) **Correlation between sequence conservation of the 5' untranslated region and codon usage bias in** *Mus musculus* **genes**. *Gene* 276(1-2): 101-5.

33. Sakai H, Imamura C, Osada Y, Saito R, Washio T, Tomita M.(2001) **Correlation between Shine-Dalgarno sequence conservation and codon usage of bacterial genes**. *J Mol Evol* 52(2): 164-70.

34. Kawai J, ···22 authors···, Saito R, ···71 authors···, Hayashizaki Y (2001) **Functional annotation of a full-length mouse cDNA collection**. *Nature* 409: 685-90.

35. Saito R, Ozawa Y, Kuzuno N, Tomita M (2000) **Computer analysis of potential stem structures of rRNA operons in various procaryote genomes**. *Gene* 259(1-2): 217-22.

36. Toda Y, Saito R, Tomita M (2000) **Characteristic sequence pattern in the 5- to 20-bp upstream region of primate** *Alu* **elements**. *J Mol Evol* 50(3):232-7.

37. Saito R, Tomita M (1999) **Computer analyses of complete genomes suggest that some archaebacteria employ both eukaryotic and eubacterial mechanisms in translation initiation**. *Gene* 238(1): 79-83.

38. Osada Y, Saito R, Tomita M (1999) **Analysis of base-pairing potentials between 16S rRNA and 5' UTR for translation initiation in various prokaryotes**. *Bioinformatics* 15(7-8): 578-81.

39. Saito R, Tomita M (1999) **On negative selection against ATG triplets near start codons in eukaryotic and prokaryotic genomes**. *J Mol Evol* 48(2): 213-7.

2.2 Interactome analyses

Recent high-throughput technologies such as the yeast two-hybrid system allowed to obtain information about a very large number of molecular interactions (mainly protein-protein interactions) in a cell. We have been working on computational analyses of such large interaction datasets to extract valuable functional information from them. The research started in 2000 with development of methods to eliminate false positives from protein-protein interaction (PPI) data - one of the main problems in dealing with high-throughput PPI data was the large amount of false positives within these datasets. Using topological features which are frequently observed in spurious interactions, we developed methods to eliminate these false positives [48,50].

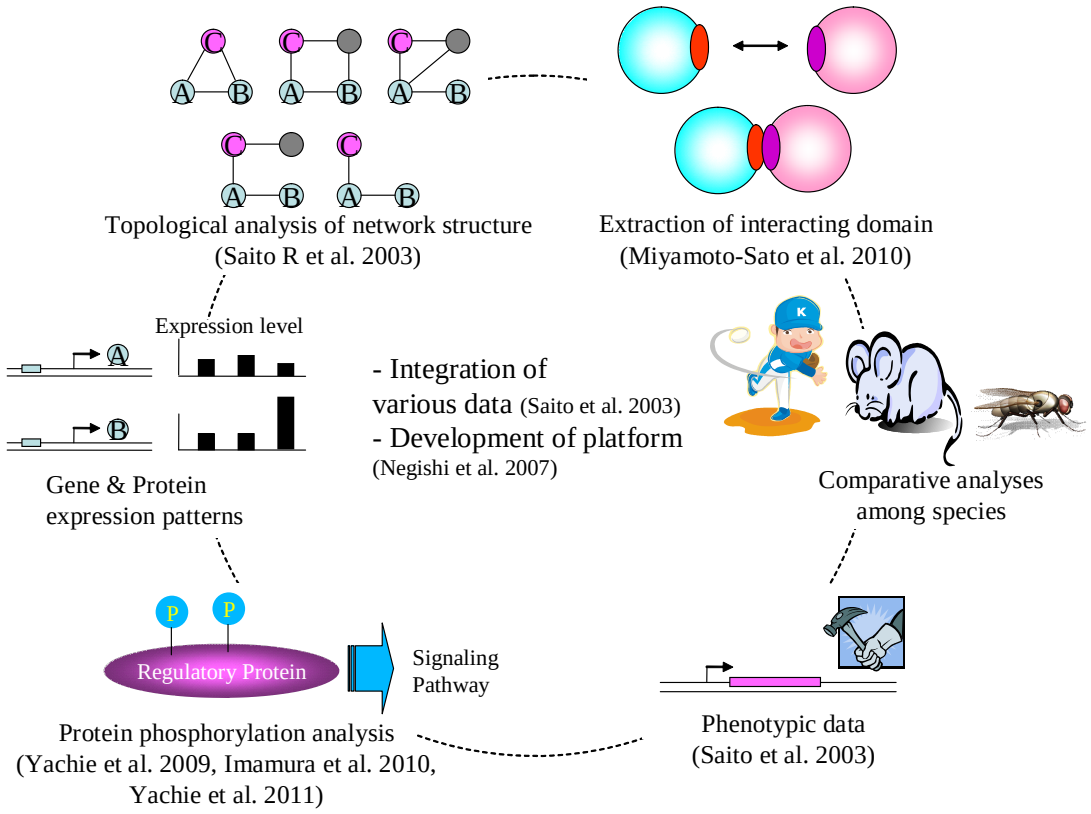

Figure 2: Interactome analysis through integration of various data and techniques

The next step was to integrate PPI data with other types of genome-wide data. We first integrated PPI data with genome-wide expression data to find that similarity in expression pattern supports true PPI. Then we integrated phenotypic data (lethality of gene deletion) to find that proteins participating in specific interactions are likely to have similar phenotypes [49].

As various genome-wide data became available, we further attempted to efficiently integrate these data to extract novel biological knowledge (Figure 2), keeping computational prediction of novel signal transduction pathways as our final goal. To do so, we first looked for genome-wide data which we should integrate.

Domain-domain interaction data were one of the candidates. We collaborated with experimental group using *in vitro* virus (IVV) technology to identify novel protein binding domains [43]. By combining the IVV data with our computational methods, we predicted not only PPIs but also their domain-domain interactions, some of which were validated by pull-down assay ([41], Manuscript in preparation).

As PPI networks are usually represented as un-directed graph, it is difficult to infer signaling pathway from these data because signaling pathways are usually "directed". Recent development in mass spectrometry allowed to obtain genome-wide protein phosphorylation status data which can be used to assign signaling directions from kinase to substrates within signaling pathways. We began to computationally characterize phosphorylation sites in phospho-protein sequences and found the "rich-gets-richer" process (phospho-proteins having many phospho-sites are likely to gain more phospho-sites) during molecular evolution of phospho-proteins [44]. We also developed preliminary methods to roughly draw signaling pathways based on time-course data of phosphorylation patterns [42].

To integrate, analyze and visualize various kind of genome-wide data, we have also developed a software platform, **eXpanda**, which provides useful APIs for bioinformaticians to study interaction networks [45].

Related publications

* Marked authors contributed equally to the work, + Corresponding author. All reviewed by referees.

40. Yachie N, Saito R+, Sugiyama N, Tomita M, Ishihama Y (2011) **Integrative features of the yeast phosphoproteome and protein-protein interaction map**. *PLoS comput biol* 7(1):e1001064

41. Ozawa Y, Saito R+, Fujimori S, Kashima H, Ishizaka M, Yanagawa H, Miyamoto-Sato E, Tomita M (2010) Protein complex prediction via verifying and reconstructing the topology of domain-domain interactions. BMC Bioinformatics (In press)

42. Imamura H, Yachie N, Saito R+, Ishihama Y, Tomita M (2010) **Towards the systematic discovery of signal transduction networks using phosphorylation dynamics data**. *BMC Bioinformatics* 11(1):232

43. Miyamoto-Sato E, Fujimori S, Ishizaka M, Hirai N, Masuoka K, Saito R, Ozawa Y, Hino K, Washio T, Tomita M, Yamashita T, Oshikubo T, Akasaka H, Sugiyama J, Matsumoto Y, Yanagawa H.(2010) **A comprehensive resource of interacting protein regions for refining human transcription factor networks**. *PloS ONE* (In press)

44. Yachie N, Saito R+, Sugahara J, Tomita M, Ishihama Y(2009) *In silico* **analysis of phosphoproteome data suggests a rich-get-richer process of phosphosite accumulation over evolution**. *Mol Cell Proteomics* 8(5):1061-71.

45. Negishi Y, Nakamura H, Yachie N, Saito R+, Tomita M (2007) **eXpanda: an Integrated Platform for Network Analysis and Visualization**. *In silico biology* 7, 0013

46. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, Ara T, Nakahigashi K, Huang HC, Hirai A, Tsuzuki T, Nakamura S, Altaf-Ul-Amin M, Oshima T, Baba T, Yamamoto N, Kawamura T, oka-Nakamichi T, Kitagawa M, Tomita M, Kanaya S, Wada C, Mori H (2006) **Large-scale identification of protein-protein interaction of** *Escherichia coli* **K-12** *Genome Res* 16(5):686-91

47. Suzuki, H., Saito R., Kanamori, M., Kai, C., Schonbach, C., Nagashima, T., Hosaka, J., Hayashizaki, Y.(2003) **The mammalian protein-protein interaction database and its viewing system that is linked to the main FANTOM2 viewer**. *Genome Res* 13(6B):1534-1541.

48. Saito R., Suzuki, H., Hayashizaki, Y.(2003) **Construction of reliable protein-protein interaction networks with a new interaction generality measure**. *Bioinformatics* 19(6):756-63.

49. Saito R, Suzuki H, Hayashizaki Y.(2003) **Global insights into protein complexes through integrated analysis of the reliable interactome and knockout lethality**. *Biochem Biophys Res Commun* 301(3):633-40.

50. Saito R., Suzuki, H. and Hayashizaki, Y. (2002) **Interaction generality, a measurement to assess the reliability of a protein-protein interaction**. *Nucleic Acids Res* 30(5): 1163-8.

51. Kanamori M, Suzuki H, Saito R, Muramatsu M, Hayashizaki Y.(2002) **T2BP, a Novel TRAF2 Binding Protein, Can Activate NF-kappaB and AP-1 without TNF Stimulation**. *Biochem Biophys Res Commun* 290(3):1108-13.

52. Suzuki H, Fukunishi Y, Kagawa I, Saito R, Oda H, Endo T, Kondo S, Bono H, Okazaki Y, Hayashizaki Y.(2001) **Protein-protein interaction panel using mouse full-length cDNAs**. *Genome Res* 11(10):1758-65.

3 Career

1983.6	-	1986.5	Hawthorne Elementary School (Ottawa Ontario, CANADA)
1986.6	-	1988.3	Higashiyama Junior High School (Public school)
1988.4	-	1991.3	Keio High School (Private school)
1991.4	-	1995.3	Keio University Faculty of Environmental Information (Bachelor's degree, 1995)
1995.4	-	1997.3	Keio University Graduate School of Media and Governance Master Course (Master's degree, 1997)
1997.4	-	2000.3	Keio University Graduate School of Media and Governance Doctoral Course (Ph.D., 2000) Thesis title: "Computer Analyses of Translation Initiation Sites of Genes"
2000.4	-	2002.3	RIKEN Genomic Sciences Center (Researcher)
2002.4	-	2010.1	Institute for Advanced Biosciences, Keio University (Assistant professor)
2010.2	-		University of California, San Diego Departments of Medicine and Bioengineering (Visiting Faculty)

4 Experienced teaching classes

Programing For Genome Analysis

- Basic Perl/Python programming.
- Basic analysis of genomic sequences.

Genome Informatics

- Basic mathematical statistics (Gaussian distribution, χ^2 distribution, etc.).
- DNA/RNA sequence analysis using information theory (Derivation of entropy, mutual information, Kullback-Leibler divergence, etc. and their applications to DNA/RNA sequence analysis).
- Prediction of RNA secondary structure by free energy optimization.

Molecular And Cellular Biology 3

- Chemical structure of membrane proteins.
- Chemical network of energy metabolism.

Bioinformatics Algorithms[1]

- Basic theory on algorithms including recursion and O-notation.
- Dynamic programming, sequence alignment.
- Probabilistic models (Hidden Markov Model).
- Machine-learning, maximum-likelihood method, Expectation-Maximization algorithm.
- Hierarchical clustering, k-means clustering, self-organization map.

[1] This course is for graduate students

5 Patent

Assessment of Protein-Protein Interactions (JP-A-2003-194813 (P2003-194813A))

6 Skills

English Fluent (Lived in Ottawa, Canada from 1983 to 1986.)

Bioinformatics Familiar with computational genomic sequence analysis and protein-protein interaction analysis. Experienced analyzing many gene expression data. Familiar with bioinformatics tools such as BLAST, FASTA, clustalw, HMMer, RNAfold. Experienced using cytoscape.

Programming Python (Main), Perl (Familiar), C (Familiar), R, C++, Java, Prolog, Microsoft VBA, Bourne shell, Awk, etc.

Biological experiments Experienced some basic techniques from 2003 to 2004 (RT-PCR, electrophoresis, Western Blotting, vector construction, transformation, etc.)

7 Book Publication

7.1 Fundamental Bioinformatics

Book title Fundamental Bioinformatics[a]

Book subtitle Approach from genome analysis programming[b]

Author Rintaro Saito

Supervisor Masaru Tomita

Published Date 2005-07-25

Publisher Saiensu-sha Co., Ltd.

Pages 168

Language Japanese

ISSN 4910054700756

[a] Translation for "Bioinformatics no kiso".
[b] Translation for "Genome kaiseki programming wo chushin ni".

Contents

- Genome sequence analysis using information theory (Entropy, relative entropy, mutual information)
- Statistics (Z-test, χ^2-test)
- Dynamic programming and application to sequence alignment and prediction of RNA secondary structure (Derivation of Zuker's formula)
- Hidden Markov Model, Viterbi's algorithm, forward and backward algorithms, paramater estimation using Baum-Welch algorithm.
- Hierarchical clustering and k-means clustering of gene expression data.
- Analysis of codon biases using principal component analysis (PCA), correspondence analysis (CA) and self-organization map (SOM). Mathematical derivation of PCA and CA.
- Appendices
  - Recursion, O-notation
  - Derivation of Gaussian distribution and χ^2 distribution.
  - Lagrange multipliers
  - Maximum likelihood estimation, expectation maximization algorithm

8 References

Masaru Tomita, Ph.D.

Director and Professor
Institute for Advanced Biosicences
Faculty of Environmental and Information Studies
Keio University
Tel: +81-466-47-5099
Fax: +81-466-47-5099
E-Mail: mt@sfc.keio.ac.jp
Web Page: http://www.iab.keio.ac.jp/en/content/view/34/124/

Yasushi Ishihama, Ph.D.

Associate Professor
Institute for Advanced Biosicences
Faculty of Environmental and Information Studies
Keio University
Tel: +81-235-29-0571
Fax: +81-235-29-0536
E-Mail: y-ishi@ttck.keio.ac.jp
Web Page: http://www.iab.keio.ac.jp/en/content/view/43/124/

Haruo Suzuki, Ph.D.

Research Associate
Department of Population Medicine and Diagnostic Sciences
College of Veterinary Medicine
Cornell University
Tel: +1-607-253-4228
Fax: +1-607-256-5608
E-Mail: hs568@cornell.edu