Curriculum Vitae

Rintaro Saito

Institute for Advanced Biosciences,
Faculty of Environment and Information Studies,
Keio University

April 1, 2010

1 Profile

|	
Family Name	Saito
Given Name	Rintaro
Gender	Male
Date of Birth	November 5, 1972
Nationality	Japanese
Address	Fujisawa, Kanagawa, Japan
Marital Status	Married
Degree	Ph.D. (Keio University, 2000)
Affiliation	Institute for Advanced Biosciences
	Faculty of Environment and Information
	Studies, Keio University
	(Postal address: Endo 5322, Fujisawa
	Kanagawa 252-8520 Japan)
Title	Assistant Professor
Research Field	Bioinformatics
Research Interests	Computational genomic analyses (especially on non-coding regions)
	Interactome analyses
Tel	+81-466-49-3473
Fax	+81-466-49-3473
E-Mail	rsaito@sfc.keio.ac.jp
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)
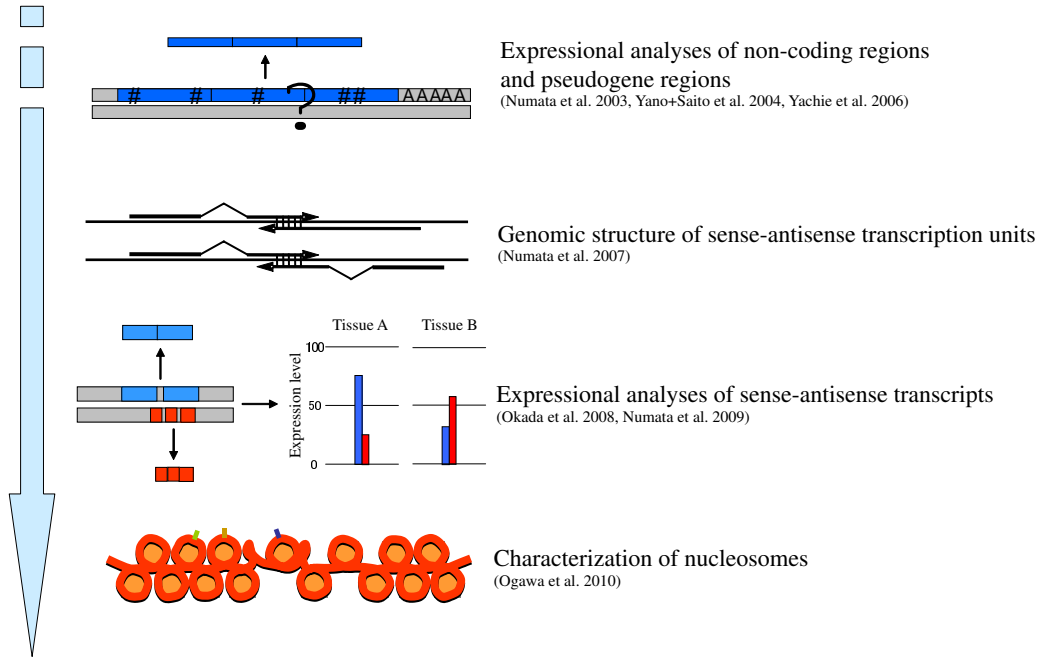
Figure 1: History of our research on genome sequence analyses

2 Researches

2.1 Characterization of genomic non-coding regions

Genomic sequences contain not only genes but also their regulatory information and how the organisms evolved at molecular level. We have especially been focusing on non-coding regions of the genomes (Figure1) and worked on computational characterization of these regions to investigate functionality of these regions and to find how these regions have evolved.

Although human genome contains >97% of non-protein-coding regions, their functions were unclear. After joining FANTOM project, we analyzed approximately 60,000 mouse cDNA sequences and surprisingly found that there are thousands of transcripts which do not encode proteins and thus there are many non-coding RNAs in a cell (Numata et al. 2003). Similarly we found that many pseudogenes are expressed in the cell (Yano and Saito et al. 2004). Subsequently, we developed a computational method (based on markov model which can assess RNA folding potential) to predict non-coding RNAs in *E. coli* (Yachie et al 2006). Expression of some of non-coding RNA candidates were experimentally validated. In this work, we found that some of non-coding RNAs were encoded in the antisense strand of protein-coding genes. Subsequently we characterized sense-antisense structures on the eukaryotic genomes using cDNA sequences (Numata et al. 2007). Furthermore, we used microarray to characterize expression patterns of these sense-antisense transcripts (Okada et al. 2008, Numata et al. 2009).

My recent interest is to integrate epigenetic information to characterize genomic sequences (Ogawa et al. 2010).
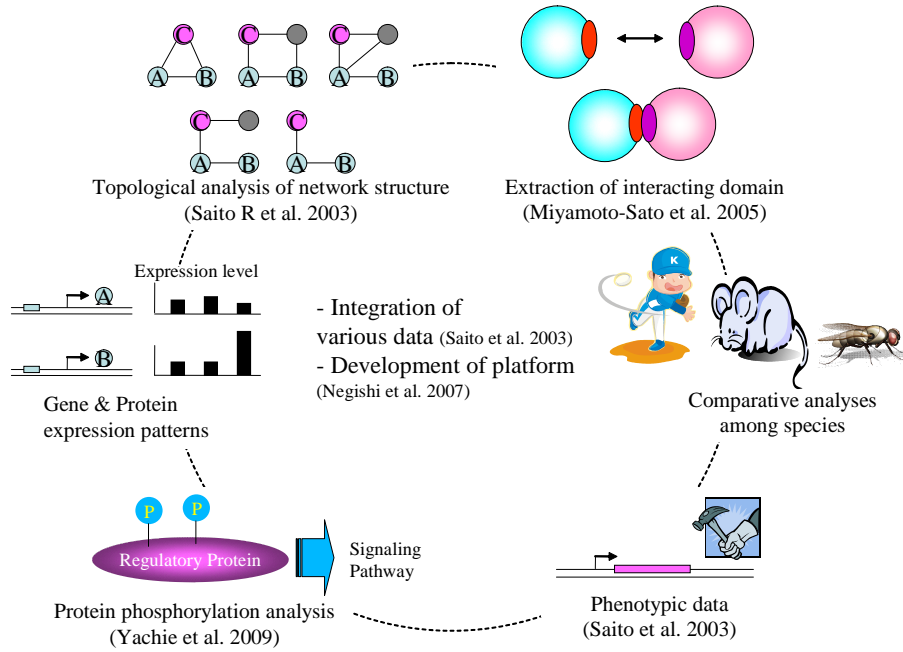
Figure 2: Interactome analysis through integration of various data and techniques

Besides these works, I am also interested in translation signals in mRNAs which were main topic of my research during my doctoral course in the late 90'th. I found that un-translated regions of mRNAs contain important information related to the mechanism of translation (Saito R et al. 1999).

We also have worked on assessment of translation initiation signals, codon biases, characterization of GC skews in the whole genome, upstream ORFs, etc.

2.2 Interactome analyses

Recent high-throughput techonologies such as yeast two-hybrid enabled us to obtain huge amount of molecular interactions in a cell. We have been working on computational analyses to extract valuable information from such huge networks. The research started in 2000 with development of method to eliminate false positives from protein-protein interaction (PPI) data - one of the essential problem to deal with high-throughput PPI data was large amount of false positives. Using topological characteristics specific to spurious interactions, we eliminated false positives (Saito et al. 2002, Saito et al. 2003). This allowed us to predict functions of uncharacterized proteins more efficiently.

The next step was to integrate PPI data with other genome-wide data. We first integrated PPI data with genome-wide expression data to find that reliable PPIs extracted by our method

are likely to have similar gene expression pattern. In addition, we integrated phenotypic data (lethality of gene deletion for each gene) to find that interacting proteins are likely to have similar phenotypes (Saito et al. 2003).

We are now trying to integrate various kinds of genome-wide data to extract biological knowledge from these huge amount of data (Figure 2). To this end, we have developed an platform eXpanda for such analyses (Negishi et al. 2007).

PPIs usually do not have directions. However protein phosphorylation data from mass spectrometry allowed us to analyze directions from kinase to substrates (Yachie et al. 2009).

3   Education History

1983	-	1986	Hawthorne Elementary School (Ottawa Ontario, CANADA)
1986	-	1988	Higashiyama Junior High School (Public school)
1988	-	1991	Keio High School (Private school)
1991	-	1995	Keio University Faculty of Environmental Information (Bachelor's degree, 1995)
1995	-	1997	Keio University Graduate School of Media and Governance Master Course (Master's degree, 1997)
1997	-	2000	Keio University Graduate School of Media and Governance Doctoral Course (Ph.D., 2000)

Thesis title: "Computer Analyses of Translation Initiation Sites of Genes"

4   Work Experiences

4.1   RIKEN Genomic Sciences Center (2000-2002)

- Participated in FANTOM[1] 1 and 2 to develop tools to analyze and visualize cDNA sequences (Kawai J et al. 2001, Okazaki Y et al. 2002).

- Developed a system to annotate coding region of given cDNA sequence (Furuno et al. 2003).

- Developed a method to extract reliable interactions from protein-protein interaction network based on topological information (Saito R et al. 2002, Saito R et al. 2003).

- Designed a database to store and view protein-protein interactions (Suzuki H et al. 2003).

- Developed a pipeline to extract non-coding RNA candidates from mouse cDNA sequences (Numata et al. 2003).

4.2   Institute for Advanced Biosciences, Keio University (2002-)

- Supervised ~25 students per year as director to conduct various researches and to publish papers.

---

[1]Functional Annotation of Mouse. http://fantom2.gsc.riken.jp/

- Developed a method to computationally extract non-coding RNA candidates from *M. musculus* (Numata et al. 2003) and *E. coli* (Yachie N et al.2006).

- Computationally analysed expression patterns of antisense RNAs and pseudogenes in *H. sapiens* and *M. musculus* (Yano Y et al. 2004, Okada et al. 2008, Numata et al. 2009).

- Characterized un-translated regions of mRNAs (Ozawa Y et al. 2002, Sato M et al. 2003, Ozawa Y et al. 2003, Osada Y et al. 2006, Mori K et al. 2007, Matsui M et al. 2007).

- Developed a platform **eXpanda**, which allows bioinformatics user to analyze and visualize various interactome (Negishi et al. 2007).

- Characterize protein-protein interaction and phosphorylation networks in cells (Arifuzzaman M et al. 2006, Yachie N et al. 2009).

- Characterized epigenetics status of genomic region (Ogawa R et al. 2010).

- Developed a method to assess codon biases of various prokaryotic genomes (Suzuki H et al. 2004, Suzuki et al. 2005, Suzuki et al. 2009).

- Organized spring (Spring Science Camp) and summer schools (Summer BioCollege, Keio Summer BioCamp) for high school students as a captain which was held at IAB. Also organized two-days school for investors (Bio Finance Guild for investors).

5 Current 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[2]

- Basic theory on algorithms including recursion and O-notation.
- Dynamic programming, sequence alignment.

---

[2]This course is for graduate students

- Probabilistic models (Hidden Markov Model).
- Machine-learning, maximum-likelihood method, Expectation-Maximization algorithm.
- Hierarchical clustering, k-means clustering, self-organization map.

6 Participated Projects and fundings

6.1 Funded projects

6.1.1 Understanding and Controls of Life's Function *via* Systems Biology

Funding organization Ministry of Education

Project name COE21 (Center of Excellence)

Duration 2002 to 2006

Funding 183,623,000 Yen[3] (Total)

Representative member Prof. Hiroshi Yanagawa

6.1.2 Extraction and Functional Analyses of Sense-Antisense RNAs in Humans

Funding organization New Energy and Industrial Technology Development Organization (NEDO)

Project name Functional RNA project

Duration H.17 to H.21

Funding 500,000 to 1000,000 Yen / Year.

Representative member Dr. Akio Kanai

6.1.3 Global analysis of transcription factor complexes using *in vitro* virus method

Funding organization Ministry of Education

Poject name Genome network project

Duration 2006 - 2008

Funding 50,000,000 Yen / Year

Representative member Prof. Hiroshi Yanagawa

---

[3]100 Yen = \$1.11 US (as of March 6, 2010)

6.2 Other fundings

6.2.1 Integrative analysis of protein-protein interactions and DNA microarray data to predict biological pathways in higher organisms (2001)

Funder Yazaki Kagaku Gijutsu Shinko Kinen Zaidan (Yazaki Memorial Science & Technology Promotion Foundation)

Funded All fees to travel and attend Intelligent Systems for Molecular Biology 2001 (Copenhagen, Denmark)

6.2.2 Construction and mining of reliable protein-protein interactions (2003)

Funder SFC project support

Funded 420,000 Yen

6.2.3 Computational prediction of expressed pseudogenes in *Homo sapiens* and *Mus musculus*

Funder Japan Society for the Promotion of Science

Funded All fees to travel and attend the conference "Genome and RNA: Expression and Functions" (Costa Rica)

6.2.4 Development of clustering algorithm for molecular interaction networks (2008)

Funder Keio University

Funded 500,000 Yen

7 Patent

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

8 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.)

9 Publications

9.1 Journals

1. Ogawa R, Kitagawa N, Ashida H, <u>Saito R</u>[+], Tomita M(2010) **Computational prediction of nucleosome positioning by calculating the relative fragment frequency index of nucleosomal sequences**. *FEBS Let* (In press)

2. Miyamoto-Sato E, Fujimori S, Ishizaka M, Hirai N, Masuoka K, <u>Saito R</u>, 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)

3. Numata K, Osada Y, Okada Y, <u>Saito R</u>, 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

4. Suzuki H, <u>Saito R</u>[+], Tomita M.(2009) **Measure of synonymous codon usage diversity among genes in bacteria**. *BMC Bioinformatics* 10:167.

5. FANTOM Consortium, Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, de Hoon MJ, Katayama S, Schroder K, Carninci P, Tomaru Y, Kanamori-Katayama M, Kubosaki A, Akalin A, Ando Y, Arner E, Asada M, Asahara H, Bailey T, Bajic VB, Bauer D, Beckhouse AG, Bertin N, Bjorkegren J, Brombacher F, Bulger E, Chalk AM, Chiba J, Cloonan N, Dawe A, Dostie J, Engstrom PG, Essack M, Faulkner GJ, Fink JL, Fredman D, Fujimori K, Furuno M, Gojobori T, Gough J, Grimmond SM, Gustafsson M, Hashimoto M, Hashimoto T, Hatakeyama M, Heinzel S, Hide W, Hofmann O, Hornquist M, Huminiecki L, Ikeo K, Imamoto N, Inoue S, Inoue Y, Ishihara R, Iwayanagi T, Jacobsen A, Kaur M, Kawaji H, Kerr MC, Kimura R, Kimura S, Kimura Y, Kitano H, Koga H, Kojima T, Kondo S, Konno T, Krogh A, Kruger A, Kumar A, Lenhard B, Lennartsson A, Lindow M, Lizio M, Macpherson C, Maeda N, Maher CA, Maqungo M, Mar J, Matigian NA, Matsuda H, Mattick JS, Meier S, Miyamoto S, Miyamoto-Sato E, Nakabayashi K, Nakachi Y, Nakano M, Nygaard S, Okayama T, Okazaki Y, Okuda-Yabukami H, Orlando V, Otomo J, Pachkov M, Petrovsky N, Plessy C, Quackenbush J, Radovanovic A, Rehli M, <u>Saito R</u>, Sandelin A, Schmeier S, Schonbach C, Schwartz AS, Semple CA, Sera M, Severin J, Shirahige K, Simons C, St Laurent G, Suzuki M, Suzuki T, Sweet MJ, Taft RJ, Takeda S, Takenaka Y, Tan K, Taylor MS, Teasdale RD, Tegner J, Teichmann S, Valen E, Wahlestedt C, Waki K, Waterhouse A, Wells CA, Winther O, Wu L, Yamaguchi K, Yanagawa H, Yasuda J, Zavolan M, Hume DA; Riken Omics Science Center, Arakawa T, Fukuda S, Imamura K, Kai C, Kaiho A, Kawashima T, Kawazu C, Kitazume Y, Kojima M, Miura H, Murakami K, Murata M, Ninomiya N, Nishiyori H, Noma

S, Ogawa C, Sano T, Simon C, Tagami M, Takahashi Y, Kawai J, 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.

6. Yachie N, <u>Saito R</u>[+], 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.

7. Okada Y, Tashiro C, Numata K, Watanabe K, Nakaoka H, Yamamoto N, Okubo K, Ikeda R, <u>Saito R</u>[+], 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

8. Suzuki H, <u>Saito R</u>[+], Tomita M (2007) **Variation in the Correlation of G + C Composition with Synonymous Codon Usage Bias among Bacteria**. *EURASIP J Bioinform Syst Biol* 61374

9. Matsui M, Yachie N, Okada Y, <u>Saito R</u>[+], Tomita M.(2007) **Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse**. *FEBS Let* 581(22):4184-8

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

11. Arakawa K, <u>Saito R</u>[+], Tomita M. (2007) **Noise-reduction filtering for accurate detection of replication termini in bacterial genomes**. *FEBS Let* 581(2):253-8

12. Numata K, Okada Y, <u>Saito R</u>[+], Kiyosawa H, Kanai A, Tomita M. (2007) **Comparative analysis of cis-encoded antisense RNAs in eukaryotes**. *Gene* 392(1-2):134-41

13. Mori K, <u>Saito R</u>[+], Kikuchi S+, Tomita M (2006) **Inferring rules of** *E. coli* **translational efficiency using an artificial neural network**. *Biosystems* 90(2):414-420

14. Osada Y, <u>Saito R</u>[+], Tomita M (2006) **Comparative analysis of base correlations in 5' untranslated regions of various species**. *Gene* 375:80-6

15. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, <u>Saito R</u>, 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

16. Yachie N, Numata K, <u>Saito R</u>[+], 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, <u>Saito R</u>, Kanai A, and Tomita M (2005) **Computational analysis of microRNA target recognition in** *Caenorhabditis elegans*. *Gene* 365:2-10

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

20. Yano Y*, <u>Saito R</u>[*], 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, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashidume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, <u>Saito R</u>, 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, <u>Saito R</u>, Osato N, Fukuda S, Sato K, Watahiki A, Hirozane-Kishikawa T, Nakamura M, Shibata Y, Yasunishi A, Kikuchi N, Yoshiki A, Kusakabe M, Gustincich S, Beisel K, Pavan W, Aidinis V, Nakagawara A, Held WA, Iwata H, Kono T, Nakauchi H, Lyons P, Wells C, Hume DA, Fagiolini M, Hensch TK, Brinkmeier M, Camper S, Hirota J, Mombaerts P, Muramatsu M, Okazaki Y, Kawai J, 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., <u>Saito R</u>., 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, <u>Saito R</u>, 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, <u>Saito R</u>, 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. Suzuki, H., <u>Saito R</u>., 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.

27. Ozawa, Y., <u>Saito R</u>., 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

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

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

30. <u>Saito R</u>, 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.

31. Ozawa Y, Hanaoka S, <u>Saito R</u>, 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.

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

33. Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., <u>Saito R</u>., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E.,Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N.,Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A.,Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E., and Hayashizaki, Y. (2002) **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs**. *Nature 420*: 563-573.

11

34. 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

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J, Fukuda S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito R., Kadota, K., Matsuda,H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa,H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R.,Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y. (2001) **Functional annotation of a full-length mouse cDNA collection**. *Nature* 409: 685-90.

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

42. Toda, Y., Saito R. and 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.

43. <u>Saito R</u>. and 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.

44. Osada Y, <u>Saito R</u>. and 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.

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

9.2 Book

9.2.1 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 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

10   Oral Presentations at International Conferences

1. <u>Saito R</u>, Ozawa Y, Fujimori S, Matsui M, Ushiama S, Kashima H, Yanagawa H, Miyamoto-Sato E, Tomita M.(2007) **Comprehensive Analysis of Domain-Domain Interactions Using In Vitro Virus**, *Genome Informatics Workshop* 2007 (Singapore) [Poster selected for oral presentation]

2. <u>Saito R</u>, Kochi H, Hirotsune S, Tomita M.(2005) **Computational Prediction of Expressed Pseudogenes in** *Homo sapiens* **and** *Mus musculus. Symposium on Structural and Evolutionary Genomics "Genome and RNA: Expression and Functions"*, 2005 (Puentarenas, Costa Rica)

3. <u>Saito R</u>. et al.(2004) **Construction of Reliable Protein-Protein Interaction Network in** *Escherichia coli. 2nd International E. coli Alliance Conference on Systems Biology* (Banff, Canada) [Invited talk]

4. <u>Saito R</u>. et al.(2003) **Refinement and Mining of Protein-Protein interaction network in** *Escherichia coli. Escherichia coli Conference. Towards New Biology in the 21st Century* (Awaji-shima, Japan)

5. <u>Saito R</u>, Tomita M.(1998) **On negative selection against ATG triplets near start codons in eukaryotic and prokaryotic genomes**. *International Conference on Bioinformatics of Genome Regulation and Structure* (BGRS'98) (Novosibirsk, Russia)

11 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