Human Genome Center
Laboratory of Genome Database

In analyzing human genome data, the importance of maintaining databases of various facts and knowledge is unquestionable. Thus, the main mission of our laboratory is to provide worldwide genome-research communities with useful resources, including supercomputer facilities and Internet services. Not only maintaining established databases but also the development of newer databases and technologies for better mining biological and medical content from accumulated data is our important project.

1. Development of SIGNAL-ONTOLOGY, an ontology for the cell signaling system

Takako TAKAI and Toshihisa TAKAGI

In the post-genome sequencing era, the most significant issue is the reconstruction of living organisms in computer, based on their genome information. Thus, reconstruction and analysis of molecular interactions among gene products, pathways, and networks could be addressed as its first step. SIGNAL-ONTOLOGY is a challenge to model cell signaling system on this prospect. Comparing with the modeling of metabolic pathways, modeling the cell signaling pathways has difficulties in the extensive diversification of molecular interactions and the lack of proper identifiers of functions in pathways. An approach to solve these problems is a development of ontology for the cell signaling system. With this, we will supply common references of biochemical and cellular functional annotations to the cell signaling system, which will play a similar role with the molecular index provided by NC-IUBMB for the metabolic system. Fundamentally, ontology is an abstraction of intrinsic nature of the domain knowledge. Therefore, our SIGNAL-ONTOLOGY will serve as references not only for signaling pathways but also for many biological applications such as comparative genome analysis, genome annotation, knowledge-based database integration, knowledge extraction from texts, gene expression profile analysis, and cell signaling simulation. SIGNAL-ONTOLOGY is open to the public from http://ontology.ims.u-tokyo.ac.jp/signalontology/.

2. Knowledge representation of signal transduction pathways

Ken-ichiro FUKUDA and Toshihisa TAKAGI

Signal transduction is a common term used to define diverse topics that encompass a large body of knowledge about the biochemical mechanisms. Since most of the knowledge of signal transduction resides in scientific articles and is represented by natural languages or diagrams, there is a need of a knowledge representation model for signal transduction pathways that can be as readily processed by computers as it is easily understood by humans. The knowledge representation model is based on a compound graph structure and is designed to handle the diversity and hierarchical structure of pathways. The model is implemented as a HiLog program and a number of biological queries are demonstrated on it.

3. Automatic dictionary building by extraction of technical terms and their attributes from literature on signal transduction

Yoshiyuki KOBAYASHI and Toshihisa TAKAGI

Biomedical literature is a rich source of information that is not accessible from annotated databases. We developed an automatic method to build a dictionary on signal transduction from literatures. This

method collects not only facts of signal transduction, but also relations among facts. It will help the process of ontology building, and the generated dictionary can be regarded as a summary of literatures. In this method, we did not apply full natural language processing (NLP) techniques, because building a full NLP system requires enormous linguistic knowledge in machine-readable form, which is not available. Our method is to collect and categorize technical terms of biology and their attributes by some heuristics on surface characteristics of language notations on biological literatures. As an information source, we do not use sentences in literature but nominal compounds extracted from the sentences, since the syntactic complexity of nominal compounds is lower than that of sentences. We identify nominal compounds with a shallow parser and some heuristics on sequence patterns of parts of speech. Our technical term extraction method has 91% precision and our term relation identification has 84% precision. Finally, our method can generate a fairly good dictionary from literatures automatically.

4. Comparison of affected-sib-pair statistics for multilocus genetic disease models in multimarker analysis

Osamu OGASAWARA and Toshihisa TAKAGI

For complex diseases, recent interests have been focused on methods that take into account the joint effect at interacting loci. It is known that these methods can be much more powerful than the single locus analysis. We are developing a computer software that estimates the number of the alleles identical by descent (IBD) between an affected sib pair at arbitrary positions in the genome using the Lander-Green algorithm. Our software also calculates some affected sib pair statistics that can analyze the joint genetic effect at arbitrary number of loci. The reliability of our software was investigated for several epistatic and heterogeneous genetic models. The result showed that an extension of restricted likelihood ratio test is much more powerful than the other statistics, especially for the "Dom or Rec" heterogeneity and "Dom and Rec" epistasis models. For the "Dom or Dom" heterogeneity and "Dom and Dom" epistasis models, the power of restricted likelihood ratio test and that of mean test were nearly the same.

5. An algorithm clustering genes and time points simultaneously

Yasuhiro KOUCHI and Toshihisa TAKAGI

One of the most generic methods to analyze gene expression profiles is the clustering of genes, based on their expression similarity. Especially, the analysis of their time courses is useful to exploit the regulatory network between genes. Usually, the similarity is calculated from the profile over the whole genes or time points, not from the specific genes or points. The main idea of this study is to extract clusters comprising gene subsets and time point subsets simultaneously. Thus, this study enables us to extract gene subsets that show similar profiles at some specific time points with comparatively small computational complexity, without any prior knowledge. First, initial cluster-seeds, which contain small number of genes and points, are generated. Then, each cluster-seed is grown to local maximal with step-by-step additions of genes and time points, keeping the cluster score increasing. Finally, pairs of clusters are merged if the score of the merged cluster is higher than the sum of these two scores. We are now applying this method to some real biology data.

6. Inferring genetic networks from DNA microarray data by multiple regression analysis

Mamoru KATO, Tatsuhiko TSUNODA[1], and Toshihisa TAKAGI:[1]SNP research center, RIKEN

Inferring gene regulatory networks by differential equations from the time series data of a DNA microarray is one of the most challenging tasks in the post-genomic era. However, there have been no studies actually inferring gene regulatory networks by differential equations from genome-level data. The reason for this is that the number of parameters in the equations exceeds the number of measured time points. We here succeeded in executing the inference, not by directly determining parameters but by applying multiple regression analysis to our equations. We derived our differential equations and steady state equations from the rate equations of transcriptional reactions in an organism. Verification with a number of genes related to respiration indicated the validity and effectiveness of our method. Moreover, the steady state equations were more appropriate than the differential equations for the microarray data used.

7. The evolution and classification of the MAP kinase signaling pathways

Asako KOIKE and Toshihisa TAKAGI

The MAP kinase pathways composed of consecutive protein phosphorylations play vital roles in eukaryotic intracellular signal transduction. In this work, an evolutionary model of MAP kinase pathways is presented based on their sequence homology, the gene structure, and the phylogenetic trees constructed from the multiple sequence align-

ment of the kinase domain. As a result, it is found that kinases of the AGC group and the CaMK group in *S. cerevisiae* are orthologous with prokaryotic kinases and eukaryotic kinases expanded from these groups. Further, MAPKKK and MAPKK are likely to be caused by gene duplications of MAPK. The correlation between the similarity of gene products and the gene order on the genome has been also evidenced. The correlation coefficients of phylogenetic trees of MAPK vs. MAPKK, MAPKK vs. MAPKKK, and MAPK vs. MAPKKK are quite high, for example, 0.96-1.00 in *S. pombe*. This indicates that the MAPK pathways have co-evolved. In addition, the relationship between MAPK pathways of *S. cerevisiae* and those of vertebrates is clarified by the phylogenetic trees and the exon-intron structure. The vertebrate MAP kinase pathways concerning apoptosis, cell differentiation, cytokine production, and cell growth have evolved from the osmosensor pathway of *S. cerevisiae* and those concerning growth factors have evolved from the mating and pseudohyphal pathways of *S. cerevisiae*.

8. Prediction of transcription factors using local regions of biased amino acids

Goro TERAI, Kenta NAKAI, and Toshihisa TAKAGI

Elucidation of transcriptional regulatory network is important to understand many biological phenomena, including cell-cycle, differentiation, and development, at the molecular level. Discovery of transcription factors, which play a primary role in transcription regulation, helps us to elucidate a novel regulation between genes. Thus, we are developing a method predicting if a protein is a transcription factor or not when its amino acid sequence is given. It is known that local regions of biased amino acid composition (low complexity regions) are frequently observed in eukaryotic transcription factors. As for transcription factors of *Saccharomyces cerevisiae*, we found that low complexity regions, which were rich in D, E, Q, N, S, T, P, K, and R, were significantly frequently observed. We used the existence of these regions as main criteria for prediction of transcription factors. Prediction result concerning subcellular localization was also used to increase overall prediction accuracy. These criteria were used to train decision rules for discriminating known transcription factors of *S. cerevisiae*. The derived rules could detect 44 transcription factors in the training data with 18 false positives. Applying these rules to proteins of unknown function, we detected 88 candidates of novel transcription factors. Among them, we found that a gene product, YCR087C-A, is referred to be a transcription factor by its annotation: This protein has a glutamine-rich region, and no significant motifs are found in its amino acid sequence in our analysis, but is annotated to have a putative zinc finger domain. Thus, our method may be able to find novel transcription factors without significant homology.

9. DBTBS: A database of *Bacillus subtilis* promoters and transcription factors

Takahiro ISHII[2], Ken-ichi YOSHIDA[3], Goro TERAI, Yasutaro FUJITA[3] and Kenta NAKAI: [2]PharmaDesign, Inc and [3]Fukuyama University

With the completion of determining its entire genome sequence, one of the next major targets of *Bacillus subtilis* genomics is to clarify the whole gene regulatory network. To this end, the results of systematic experiments should be compared with the rich source of individual experimental results accumulated so far. Thus, we constructed a database of the upstream regulatory information of *B. subtilis* (DBTBS). Its current version is constructed by surveying 291 references and contains the information of 90 binding factors and 403 promoters. For each promoter, all of its known *cis*-elements are listed according to their positions, while these *cis*-elements are aligned to illustrate its consensus sequence for each transcription factor. All probable transcription factors coded in the genome were classified with the Pfam motifs. Furthermore, to facilitate the discovery of uncharacterized *cis*-elements, we added the alignment of upstream 300 nt of *B. subtilis* genes with orthologous genes of *Bacillus halodurans* and *Bacillus stearothermophilus.* Our database is accessible at http://elmo.ims.u-tokyo.ac.jp/dbtbs/.

10. Amino acid sequence analysis of polytopic membrane proteins

Mitsuhiro ARAKI, Toshihisa TAKAGI, and Kenta NAKAI

Although most transmembrane segments (TMSs) of α-type polytopic membrane proteins are composed of hydrophobic amino acids, about 8.4% of TMSs are relatively less hydrophobic and are hard to be distinguished with relatively hydrophobic loop segments. To understand why such less hydrophobic segments span the membrane, we systematically analyzed their sequence features using a non-biased database of experimentally-verified polytopic membrane proteins. Inspired by a recent experimental result (Ota *et al., Mol. Cell* 2, 495, 1998), we assumed that each TMS has its own topogenic tendency determined by the distribution of neighboring positive residues. Namely, a normal TMS tends to have the N_{in}/C_{out} orientation if it has more positive charges on its nearby N-terminal side; otherwise a TMS tends to be in the N_{out}/C_{in} orientation. Surprisingly, the

presence of less hydrophobic segments highly correlated with such topogenic tendencies of their flanking TMS on the C-terminal side. This explains why less hydrophobic segments span the membrane for 93 % of cases (excluding five segments on C-termini) and why relatively hydrophobic loop segments do not for 72 % of cases. Our result not only implies the mechanism of membrane protein assembly but also will be useful for improving TMS prediction methods.

11. A system for analyzing transcription regulatory regions

Katsuhiko MURAKAMI[4], Yoshihiro OHTA[5], Koji TANIKAWA[4], Hiroki NAKAE[4], Shigeo IHARA[4], and Toshihisa TAKAGI:[4]Life Science Group, Hitachi, Ltd. and [5]Central Research Laboratory, Hitachi, Ltd.

We have developed a transcription regulatory region analysis system. The system comprises 1) automatic "additional motifs" finding modules, 2) automatic data distribution modules, which can divide GenBank and EPD data into user-defined appropriate groups of biological species, 3) Pol II promoter prediction modules similar to PromFD, which can deal with both databases and computationally extracted motifs. Unlike PromFD, the system includes a YEBIS-like module, which produces TF-type motifs as hidden Markov models (HMM) from promoter sequences. The HMM can represent spacer-included motifs as well as widely used position weight matrix type motifs. A preliminary test of the performance of the Pol II promoter prediction (a function of the system) produced encouraging results. 4) Input/output GUI as part of a WWW-based client/server system. 5) Interfaces for direct accesses to transcription factor databases, such as TFD, TRANSFAC, IMD, and other "additional motifs". Users can access those databases from analysis result windows. 6) A distributed processing system which allows the analysis, especially the search of transcription factor databases and motifs, to be performed in a very short time. This system will accelerate the study of regulatory region *in silico* in various situations.

12. HGREP: Human Genome REconstruction Project

Tetsushi YADA, Yasushi TOTOKI, Yoshiyuki SAKAKI, and Toshihisa TAKAGI

On June 2000, working draft sequences which are estimated to cover 85% of the human genome were released, and the time was ripe for tracing the outline of the genome and for finding genes. However, since these sequences are too fragmentary, we cannot use them in raw format for the above purposes. Therefore, we launched a new project named HGREP (Human Genome REconstruction Project). In HGREP, (1) we sort the draft sequences by chromosome and assemble them, and (2) we make annotations such as genes and repeats on the sequences. These results are distributed via internet (http://hgrep.ims.u-tokyo.ac.jp/) and are continuously updated.

Although other research groups are doing similar attempts, such as Ensembl (http://www.ensembl.org/), HGREP is unique in the following points. (1) Ensembl assembles the draft sequences based on DNA fingerprint data, while HGREP assembles them based on sequence similarity. Therefore, HGREP realizes the reliable reconstruction of human chromosomes. (2) HGREP annotates genes by using a novel gene finding program named DIGIT (in progress). DIGIT enables us to predict exact gene structures with 10% higher accuracy compared to existing gene finding programs.

HGREP is a joint project between the Laboratory of Genome Database (University of Tokyo, Institute of Medical Science, Human Genome Center) and the Human Genome Research Group (Genomic Sciences Center, RIKEN).

13. Database and network services for sequence interpretation and information retrieval

a. Wide Area Network

Wide-area computer network is an essential component of the infrastructure for genome research. Thus, we are collaborating with the "GenomeNet" activity at Kyoto University, in cooperation with the IMNet and WIDE computer network groups. Currently, a 6Mbps line from Tokyo to Kyoto and a 6Mbps line to the US are maintained.

b. Computer system

For database and computational services, a supercomputer system is maintained. The system includes:

* SGI-CRAY T94/4128 (vector computer)
  Hitachi SR2201 (massively parallel computer with distributed memory architecture)
* SGI-CRAY Origin2000 (parallel computer with distributed shared memory architecture)
* Sun Ultra Enterprise 10000 (parallel computer with shared memory architecture)
* Sony Petasite (mass storage tape device)
* Sun Ultra1 and SGI Octane (workstations)

c. Database services

We support various database services through the Internet (http://www.hgc.ims.u-tokyo.ac.jp/

database.html). Not only standard databases of biological sequences, structures, and literature, but also locally-developed smaller databases are made publicly available by either e-mail or WWW.

Publications

Yoshida, M., Fukuda, K., and Takagi, T. PNADCSS: a workbench for constructing a protein name abbreviation dictionary, Bioinformatics 16:169-175, 2000.

Ono, T., Hishigaki, H., Tanigami, A., and Takagi, T. Automated extraction of information on protein-protein interactions from biological literature. Bioinformatics, 17: 155-161, 2001.

Hishigaki, H., Nakai, K., Ono, T., Tanigami, A., and Takagi, T. Assessment of prediction accuracy of protein function from protein-protein interaction data. Yeast, 18: 523-531, 2001.

Kato, M., Tsunoda, T., and Takagi, T. Inferring genetic networks from DNA microarray data by multiple regression analysis. Genome Informatics 11:118-128, 2000.

Nakai, K. Protein sorting signals and prediction of subcellular localization. Adv. Protein Chem. 54:277-344, 2000.

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O. and Zhang, M.Q. An alternative exon database (AEDB) and its statistical analysis, DNA and Cell Biol., 19: 739-756, 2000.

Ishii, T., Yoshida, K., Terai, G., Fujita, Y., and Nakai, K. DBTBS: A database of *Bacillus subtilis* promoters and transcription factors, Nucleic Acids Res. 29:278-280, 2001.

Kaminuma, T., Takai-Igarashi, T., Nakano, T., and Nakata, K. Modeling of signaling pathways for endocrine disruptors. BioSystems 55:23-31, 2000.

Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D. K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A. The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. Nature. 405:311-319, 2000.

Yamasaki ,Y., Watanabe, T. K., Okuno, S., Ono, T., Oga, K., Mizoguchi-Miyakita, A., Goto, Y., Shinomiya, H., Momota, H., Miyao, H., Hayashi, I., Asai. T., Suzuki, M., Harada, Y., Hishigaki, H., Wakitani, S., Takagi, T., Nakamura, Y., and Tanigami, A. Quantitative trait loci for lipid metabolism in the study of OLETF x (OLETF x Fischer 344) backcross rats. Clin. Exp. Pharmacol, Physiol. 27:881-886, 2000.

Ono, K., Tanaka, T., Tsunoda, T., Kitahara, O., Kihara, C., Okamoto, A., Ochiai, K., Takagi, T., and Nakamura, Y. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res. 60:5007-5011, 2000.

McCarthy, L. C., Bihoreau, M. T., Kiguwa, S. L., Browne, J., Watanabe, T. K., Hishigaki, H., Tsuji, A., Kiel, S., Webber, C., Davis, M. E., Knights, C., Smith. A., Critcher, R., Huxtall. P., Hudson, J. R. Jr., Ono, T., Hayashi, H., Takagi, T., Nakamura, Y., Tanigami, A., Goodfellow, P. N., Lathrop, G. M., James, M. R. A whole-genome radiation hybrid panel and framework map of the rat genome. Mamm. Genome 11:791-795, 2000.

Watanabe, T. K., Ono, T., Okuno, S., Mizoguchi-Miyakita, A., Yamasaki, Y., Kanemoto, N., Hishigaki, H., Oga, K., Takahashi, E., Irie, Y., Bihoreau, M. T., James, M. R., Lathrop, G. M., Takagi, T., Nakamura, Y., Tanigami, A. Characterization of newly developed SSLP markers for the rat. Mamm. Genome 11:300-305, 2000.

高木利久，冨田勝編集：ゲノム情報生物学（ポストシークエンスのゲノム科学6. 松原謙一・榊佳之監修）．中山書店 2000.

菱垣晴次，高木利久：バイオインフォマティクスによるタンパク質－タンパク質相互作用の解析．遺伝子医学 4(4) 78-82, 2000.

高木利久，矢田哲士，寺井悟朗：ゲノムデータベース . 実験医学（増刊「ゲノム医学とこれからのゲノム医療」）18: 1629-1634, 2000.

中井謙太（分担執筆）：実験医学別冊 ポストゲノム時代の実験講座1)．辻本豪三･田中利男編．羊土社 121-126, 163-167, 2000.

中井謙太：タンパク質の構造予測と機能予測，遺伝子医学 4(3):377-382, 2000.

高井貴子：パスウェイデータベース．BIOINDUSTRY 17(7): 5-14, 2000.

高井貴子：化学物質の安全性情報とゲノムサイエンスの融合．食品衛生化学雑誌 41(6):J392-J395, 2000.

Human Genome Center

Laboratory of Genome Structure Analysis

The main project in our laboratory is to identify and collect human genes en masse in the form of full-length cDNA clones. The sequence informations of full-length cDNA are indispensable for elucidating gene structures, such as exons and introns as well as promoters. Furthermore, full-length cDNA clones are indispensable resource for functional analysis of genes. Thus, the direction of our Laboratory is a mass determination of gene structures and functions. Following are topics in the year 2000.

1. Identification and isolation of human full-length cDNA clones by 1 pass sequencing

Yutaka Suzuki, Hiroko Kozuka-Hata, Junko Mizushima-Sugano, Tadashi Sato, Kiyomi Yoshitomo-Nakagawa, Yoshihiro Omori, Takusi Togashi and Sumio Sugano

We have sequenced 5' end of randomly picked cDNA clones from full-length enriched cDNA libraries made by "oligo-capping" method. We have sequenced about 130,000 clones this year. Of these clones, about 1/2-2/3 of them contained already known genes. About 50% of the known clones seemed to be full. With Helix Institute, we also sequenced about 400,000 clones. Now, we have about 20,000 putative full-length cDNA clones with unknown function. Using 5' end 1 pass sequence data, we identified mRNA start sites of many genes and now making human promoter data using these data.

With FLJ cDNA sequencing consortium, the entire sequence was determined for 8060 clones out of 20,000. The average length of cDNA is about 2200bp which distribute from 1kb to 5kb. 5032 clones had ORF longer than 120 amino acid residues (AA). The average ORF length is about 390 AA. About 16% of these clones had membrane-spanning sequence and 3.6% signal sequences. Further more, about 25 % of the clones with ORF longer than 120 AA had some type of motifs or showed some homology to known proteins. We are also mapping these fully sequenced clones to the draft sequence of the human genomes. The sequence data were deposited on the Genbank database and the clones will be available from several suppliers.

2. Identification of differentially expressed genes in metastatic site

Junichi Imai, Manabu Watanabe and Sumio Sugano

We have analyzing differetially expressed genes in lung metastatic model using differential display method. Metastasis of a primary tumor is a multistage process, and the interactions of tumor cells with host stromal cells must influence this process. These interactions may regulate the changes of the multiple gene expression in both tumor cells and host stromal cells at the metastasized site. In the course of characterizing these changes, we have identified overexpression of the c-met proto-oncogene at the metastasized lung by using the mRNA differential display technique. Immunohistochemical staining analysis showed that Met protein was overexpressed in tumor cells at the metastasized site. The c-met encodes a transmembrane tyrosine kinase identified as the receptor for hepatocyte growth factor/scatter factor (HGF/SF). HGF/SF was expressed at lung tissue. The Met was phosphorylated at the metastasized lung. Moreover, the overexpression of c-met was a induction process of transcriptional level, not a selection process. Finally, the c-met was also overexpressed at the metastasized lung by injection of both MC-1 fibrosarcoma cells and B16 melanoma cells. These findings suggest that the HGF/SF-Met signaling may be involved in metastasis.

3. Functional analysis of proteins identified by full-length cDNA clones

Yoshihiro Omori, Takami Komatsu, Takushi Togashi, Masaaki Oyama, Munetomo Hida, Yutaka Suzuki, Sumio Sugano

Function of new genes identified by full-length cDNA clones were first analyzed by sequence homology. Many cDNA clones showed some degree of homology with previously known genes. One interesting example of such cDNA is a homolog of angiotensin converting enzyme (ACE) on X chromosome. ACE is a major target for anti-hypertension drags and thought to be the single gene in human genome. Now, the presence of new ACE homolog, ACE2, becomes clear. It is interesting how this newly found ACE2 correlate with the pathology of hypertension.

Homolog search revealed that there was significant number of cDNAs which showed similarity to transcription factors. The expression analysis showed that some of them were expressed in the tissue specific fashion. These tissue specific transcription factors will be very interesting targets for the understanding of development and the function of tissues.

In order to facilitate the functional analysis of the proteins, we are now developing a mass expression capacity of the proteins from cDNA. We are also developing the "proteomics" capacity for the high through-put protein identification and interaction analysis.

4. Monky cDNA project

Munetomo Hida, Yutaka Suzuki, Sumio Sugano

In collaboration with Prof. Momoki Hirai in Faculty of Science and Dr. Katsuyuki Hashimoto in National Institute of Infectious Diseases, we started monky cDNA identification similar to that of human described above. The target organ for the isolation of full-length cDNAs is brain. We made "Oligo-capping" cDNA libraries from various parts of Macaca brain and more than 40,000 cDNA clones were sequenced at their 5' end and the comparison between human data is in progress.

Publications

Suzuki, Y., Ishihara, D., Sasaki, M., Nakagawa, H., Hata, H., Tsunoda, T., Watanabe, M., Komatsu, T., Ota, T., Isogai, T., Suyama, A., Sugano, S., Statistical analysis of 5' untranslated region of human mRNA using "Oligo-capping" cDNA libraries. *Genomis* 64: 286-297, 2000.

Ohmori, Y., Tanigami, A., Sugano, S Comparative PCR: A Simple and Sensitive Method to Quantify Low-Abundant mRNA Species. Genomics 67: 140-145, 2000.

Iizawa, M., Han, H.-J., Furukawa, Y., Nakajima, Y., Sugano, S., Ogawa, M., Nakamura, Y. Isolation and chromosomal assignment of a novel human gene, CORO1C, homologous to coronin-like actin-binding proteins. Cytogenet. Cell Genet. 88: 221-224, 2000.

Akashi, H., Han, H.-J., Iizawa, M., Nakajima, Y., Furukawa, Y., Sugano, S., Imai, K, Nakamura, Y. Isolation and characterization of a novel gene encoding a putative seven-span transmembrane protein, TM7SF3. Cytogenet. Cell Genet. 88: 305-309, 2000.

Akashi, H., Han, H.-J., Iizaka, M., Nakajima, Y., Furukawa, Y., Sugano, S., Imai, K., Nakamura, Y. Isolation and characterization of a human cDNA encoding a protein hohologous to the 7.2-kDa protein (subunit X) of bovine ubiquinol-cyctochrome C reductase. J. Hum. Genet. 45: 43-46, 2000.

Yang, Z-Q., Imoto, I., Fukuda, Y., Pimkhaokham, A., Shimada, Y., Imamura, M., Sugano, S., Nakamura, Y., Inazawa, J. Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. Cancer Research. 60: 4735-4739, 2000.

Yoshikawa, T., Nagasugi, Y., Azuma, T., Kato, M., Sugano, S., Hashimoto, K., Masuho, Y., Muramatsu, M., Seki, N. Isolation of novel mouse genes differentially expressed in brain using cDNA microarray. Biochemical and Biophysical Research Communications. 275: 532-537, 2000.

Osada, N., Kusuda, J., Suzuki, Y., Sugano, S., Hashimoto, K. Sequnce analysis, gene expression, and chromosomal assignment of mouse Borg4 gene and its human orthologue. J. Hum. Genet. 45: 374-377, 2000

Togashi, T., Choi, D.-K., Taylor, T. D., Suzuki, Y., Sugano, S., Hattori, M., Sakaki, Y. A novel gene, DSR1, from the distal Down syndrome critical region on chromosome 21q22.2. *DNA Res*. 7: 207-212, 2000.

Kawamoto T., Shishikura T., Ohira M., Takayasu H., Morohashi A., Takada N., Takahashi M., Suzuki Y., Sugano S., Hori T., Nakagawara A. Association between favorable neuroblastoma and high expression of the novel metalloproteinase gene, nbla3145/XCE, cloned by differential screening of the full-length-enriched oligo-capping neuroblastoma cDNA libraries. Med Pediatr Oncol. 35: 628-631, 2000.

Ohira M., Shishikura T., Kawamoto T., Inuzuka H.,

Morohashi A., Takayasu H., Kageyama H., Takada N., Takahashi M., Sakiyama S., Suzuki Y., Sugano S., Kuma H., Nozawa I., Nakagawara A. Hunting the subset-specific genes of neuroblastoma: Expression profiling and differential screening of the full-length-enriched oligo-capping cDNA libraries. Med Pediatr Oncol. 35: 547-549, 2000.

Davuluri R. V., Suzuki Y., Sugano S., Zhang M. Q. CART classification of human 5' UTR sequences. Genome Res. 10: 1807-1816, 2000.

Campbell H. D., Kamei M., Claudianos C., Woollatt E., Sutherland G. R., Suzuki Y., Hida M., Sugano S., Young I. G. Human and mouse homologues of the drosophila melanogaster tweety (tty) gene: A novel gene family encoding predicted transmembrane proteins. Genomics. 68: 89-92, 2000.

Takasaki J., Kamohara M., Matsumoto M., Saito T., Sugimoto T., Ohishi T., Ishii H., Ota T., Nishikawa T., Kawai Y., Masuho Y., Isogai T., Suzuki Y., Sugano S., Furuichi K. Related The molecular characterization and tissue distribution of the human cysteinyl leukotriene CysLT(2) receptor. Biochem Biophys Res Commun. 274: 316-322, 2000.

Watanabe, M., Sugano, S., Togashi, T., Imai, J., Uchida, K., Yamaguchi, R., Tateyama, S. Molecular cloning and phylogenetic analysis of canine β-casein. *DNA sequence* in press.

16. Watanabe, M., Tateyama, S., Togashi, T., Uchida, K., Yamaguchi, R., Shimizu, T., Sugano, S. Identification of canine α-lactalbumin. J Veterin. Med. Sci. in press

Watanabe, M., Yoshida, K., Hida, M., Kato, H., Uchida, K., Yamaguchi, R., Tateyama, S., Sugano, S. Cloning, expression analysis and chromosomal mapping of GTPBP2, a novel member of the G protein family. Gene in press

Watanabe, J., Sasaki, M., Suzuki Y., Sugano S. FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum. Nucl. Acid Res. in press

Watanabe, M., Sugano, S, Imai, J., Yoshida, K., Onodera, R., Amin, M. R., Uchida, K., Yamaguchi, R., Tateyama, S. Inhibition of cell proliferation, suppression of tumourigenecity, and induction of differentiation of the canine mammary tumour cell line by sodium phenylacetate. Res. Yet. Sci., in press

Human Genome Center

Laboratory of DNA Information Analysis

The aim of the research at this laboratory is to establish computational methodologies for discovering and interpreting information of nucleic acid sequences, proteins and some other experimental data arising from researches in Genome Science. Our current concern is to realize a system which can deal with the relationship between sequence information and biological functions by extracting biological knowledge encoded on sequences and by using knowledge bases developed so far. Apart from the research activity, the laboratory has been providing bioinformatics software tools and has been taking a leading part in organizing an international forum for Genome Informatics.

1. Computational Strategies for Analyzing Gene Expression Profiles

a. Algorithms for inferring qualitative models of biological networks

Tatsuya Akutsu, Satoru Miyano and Satoru Kuhara[1]:[1]Graduate School of Agriculture, Kyushu University

Modeling genetic networks and metabolic networks is an important topic in bioinformatics. We propose a qualitative network model which is a combination of the Boolean network and qualitative reasoning, where qualitative reasoning is a kind of reasoning method well-studied in Artificial Intelligence. We also develop algorithms for inferring qualitative networks from time series data and an algorithm for inferring S-systems (synergistic and saturable systems) from time series data, where S-systems are based on a particular kind of nonlinear differential equation and have been applied to the analysis of various biological systems.

b. Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function

Tatsuya Akutsu, Satoru Miyano, Satoru Kuhara[1]

Due to the recent progress of the DNA microarray technology, a large number of gene expression profile data are being produced. How to analyze gene expression data is an important topic in computational molecular biology. Several studies have been done using the Boolean network as a model of a genetic network. We proposes efficient algorithms for identifying Boolean networks of bounded indegree and related biological networks, where identification of a Boolean network can be formalized as a problem of identifying many Boolean functions simultaneously. The algorithm is obtained by combining fast matrix multiplication with the randomized fingerprint function for string matching. Although the algorithm and its analysis are simple, the result is non-trivial and the technique can be applied to several related problems.

2. Knowledge Discovery Systems

a. HypothesisCreator: a framework for knowledge discovery application development

Hideo Bannai, Yoshinori Tamada[2], Osamu Maruyama and Satoru Miyano:[2]Department of Mathematics, Tokai University

HypothesisCreator is a software library designed

to assist the knowledge discovery process, by facilitating the development of knowledge discovery systems and computational knowledge discovery experiments. We demonstrate the use of HypothesisCreator in developing computational experiments to analyze several kinds of genomic data, showing the diversity of HypothesisCreator's applicability and its usefulness. The implementation of HypothesisCreator is based on the idea of view-scope, a generalized form of attribute. The consideration of view-scope allows the seamless integration of various knowledge in databases, newly created attributes, and even experts' intuitions. For a typical user of HypothesisCreator, the task consists of designing how to look at the data (the designing of view-scope), and then applying hypothesis generation algorithms to these view-scopes. The current implementation of the library provides numerous view-scopes and several hypothesis generation algorithms, which can be combined to create new view-scopes to suit the user's needs. If the ready-made view-scopes do not suffice, the implementation of new view-scope classes by the user is not so difficult since HypothesisCreator is an object oriented class library. We are currently applying HypothesisCreator to several different kinds of data, e.g., (a) finding gene regulatory sites, (b) characterization of subcellular localization signals. The details of the experiments will be reported elsewhere.

HypothesisCreator is free software distributed under the GNU General Public License.It will be available fromhttp://www.HypothesisCreator. net/.

b. Polynomial-time learning of elementary formal systems

Satoru Miyano, Ayumi Shinohara[3] and Takeshi Shinohara[4]:[3]Department of Informatics, Kyushu University and [4]Department of Artificial Intelligence, Kyushu Institute of Technology

An elementary formal system (EFS) is a lgic program consisting of definite clauses whose arguments have patterns instead of first-order terms. We investigate EFSs for polynomial-time PAC-learnability. A definite clause of an EFS is hereditary if every pattern in the body is a subword of a pattern in the head. With this new notion, we show that H-EFS(m,k,t,r) is polynomial-time learnable, which is the class of languages definable by EFSs consisting of at most m hereditary definite clauses with predicate symbols of arity at most r, where k and t bound the number of variable occurrences in the head and the number of atoms in the body, respecftively. The class defined by all finite unions of EFSs in H-EFS(m,k,t,r) is also polynomial-time learnable. We also show an interesting series of NC-learnable classes of EFSs. As hardness results, the cflass of regular pattern languages is shown not polynomial-time learnable unless RP=NP. Furthermore, thhre related problem of deciding whether there is a common subsequence which is consistent with given positive and negative examples is shown NP-complete.

c. A simple greedy algorithm for finding functional relations:efficient implementation and average case analysis

Tatsuya Akutsu, Satoru Miyano and Satoru Kuhara[1]

Inferring funcgtional relations from relational databases is important for discovery of scientific knowledge because many experimental data in sceince are represented in the form of tables and many rules are represented in the form of functions. A simple greedy algorithm has been known as an approximation algorithm for this problem. In this algorithm, the original problem is reduced to the set cover problem and a well-known greedy algorithm for the set cover is applied. We show an efficient implementation of this algorithm that is specialized for inference of functional relations. If one functional relation for one output variable is required, each iteration step of the greedy algorithm can be executed in linar time. If functional relations for multiple output variables are required, it uses fast matrix multiplication in order to obtain non-trivial time complexity bound. In the former case, the algorithm is very simple and thus practical. We also show that the algorithm can find an exact solution for simple functions if input data for each funcion are generated uniformly at random and the size of the domain is bounded by a constant.

d. Intelligent system for topic survery in MEDLINE by keyword recommendation and learning text characteristics

Miyako Tanaka[5], Sanae Nakazono[6], Hiroshi Matsuno[6], Hideki Tsujimoto[7], Yasuhiko Kitamura[7] and Satoru Miyano:[5]Ube National College of Technology, [6]Faculty of Science, Yamaguchi University and [7]Osaka City University

We have implemented a system for assisting experts in selecting MEDLINE records for database construction purposes. This system has two specific features: The first is a learning mechanism which extracts characteristics in the abstracts of MEDLINE records of interest as patterns. These patterns reflect selection decisions by experts and are used for screening the records. The second is a keyword recommendation system which assists and supplements experts' knowledge in unexpected cases. Combined with a conventional keyword-based information retrieval system, this system may provide an efficient and comfortable environment for MEDLINE record selection by experts. Some compu-

tational experiments are provided to prove that this idea is useful.

3. Simulation Systems

a. Hybrid Petri net representation of gene regulatory network

Hiroshi Matsuno[6], Atsushi Doi[6], Masao Nagasaki and Satoru Miyano

It is important to provide a representation method of gene regulatory networks which realizes the intuitions of biologists while keeping the universality in its computational ability. We propose a method to exploit hybrid Petri net (HPN) for representing gene regulatory networks. HPN is an extention of Petri nets which have been used to represent many kinds of systems including stochastic ones in the field of computer science and engineering. Since HPN has continuous and discrete elements, it can easily handle biological factors such as protein and mRNA concentrations. We demonstrate that, by using HPNs, it is possible to translate biological facts into HPNs in a natural manner. t should be also emphasized that a hierarchical approach is taken for our construction of the genetic switch mechanism of lambda phage which is realized by using HPNs. This hierarchical approach with HPNs makes easier the arrangement of the compoinents in the gene regulatory network based on the biological facts and provides us a prospective view of the network. We also show some computational results of the protein dynamics of the lambda phage mechanism that is simulated and observed by implementing HPN on a currently available tool.

b. Genomic Object Net: object oritented reprentation of biological systems

Hiroshi Matsuno[6], Atsushi Doi[6], Rainer Drath[8] and Satoru Miyano:[8]University of Ilmenau

One of the most important and interesting topic in the field of bioinformatics is to develop the tool simulating biological phenomenon such as gene expressions and biochemical reactions. The required conditions for realizing the effective simulation tool are:(1)Acceptable technical expression of the tool to biologists, (2)Easy to describe biological facts and biological phenomenon on computers, (3)Easy to get the tool through Internet, and (4)Easy to simulate the biological phenomenon on the tool. Our solution to the problems is "exploiting hybrid Petri net (HPN) technique for describing biological systems''. We showed that, by using HPN, the genetic switch mechanism of lambda phage can be realized on computer in a natural manner, and protein and mRNA concentrations of the mechanism can be successfully simulated by using Visual Object Net++ which is a gene ral purpose system description tool based on HPN technique. We also showed the following examples of biological systems describing and simulating on Visual Object Net++; (a) Circadian rhythms in Drosophila, (b) Delta-Notch lateral inhibitory, and (c) Apoptosis induced by protein Fas. We then introduce our next strategy "Genomic Object Net Project" which may lead us to the development of new efficient bio-simulation tools.

4. RNA Secondary Structure Prediction

a. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots

Tatsuya Akutsu

We developed dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. For a basic version of the problem (i.e., maximizing the number of base pairs), we developed an $O(n^4)$ time exact algorithm and an $O(n^{4-\delta})$ time approximation algorithm. The latter one outputs, for most RNA sequences, a secondary structure in which the number of base pairs is at least 1-ε of the optimal, where ε, δ are any constants satisfying $O < \varepsilon, \delta < 1$.

5. Multiple Alignment Algorithm

a. Approximation algorithms for local multiple alignment

Tatsuya Akutsu, Hiroki Arimura[3] and Shimozono Shimozono[4]

We studied the local multiple alignment problem, which is also known as the general consensus patterns problem. Local multiple alignment is, given proten or DNA sequences, to locate a region (i.e., a substring) of fixed length from each sequence so that the score determined from the set of regions is optimized. We consider the following scoreing schemes: the score indicating the averate information content, the score defined by Li et al., and the sum-of-pairs score. We prove that multiple local alignment is NP-hard under each of these scoring schemes. In addition, we prove that multiple local alignment is APX-hard under the averate information content scoring. It implies that unless P=NP there is no polynomial time algorithm whose worst case approximation error can be arbitrarily specified. Several related theoretical results are provided. We also made computational experiments on approximation algorithms for local multiple alignment under the averate information content scoring. The results suggest that the Gibbs sampling algorithm proposed by Lawrence et al. is the best.

6. Algorithms

a. On the approximation of largest common subtrees and largest common point sets

Tatsuya Akutsu and Magnus M. Halldorson[9]:[9] University of Iceland

We considered the approximability of the largest common subtree and the largest common point-set problems, which have applications in molecular biology. It is shown that the problems can not be approximated within a factor of $n^{1-\varepsilon}$ in polynomial time for any $\varepsilon > 0$ unless NP $\subseteq$ ZPP, while a general search algorithm which approximates both problems within a factor of $O(n/\log n)$ is presented. For trees of bounded degree, an improved algorithm which approximates the largest common subtree within a factor of $O(n/\log^2 n)$ is presented. Moreover, several variants of the largest common subtree problem are studied.

Publications

Akutsu, T. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. Discrete Applied Mathematics 104:45-62, 2000.

Akutsu, T. A local search algorithm for local multiple alignment: special case analysis and application to cancer classification. Genome Informatics 11:337-338, 2000.

Akutsu, T., Arimura, H., Shimozono, S. On approxi mation algorithms for local multiple alignment. RECOMB 2000, 4:1-7, 2000.

Akutsu, T., Halldorson, M.M. On the approximation of largest common subtrees and largest common point sets. Theoretical Computer Science 233:33-50, 2000.

Akutsu, T., Miyano, S., Kuhara, S. Algorithms for inferring qualitative models of biological networks. Bioinformatics 16:727-734, 2000.

Akutsu, T., Miyano, S., Kuhara, S. Algorithms for identifying Boolean networks and related biologi cal networks based on matrix multiplication and fingerprint function. J. Comput. Biol. 7:331-343, 2000.

Akutsu, T., Miyano, S., Kuhara, S. A simple greedy algorithm for finding functional relations: efficient implementation and average case analysis. Lecture Notes in Artificial Intelligence 1967:86-98, 2000.

Akutsu, T., Miyano, S., Kuhara, S. Inference of nonlinear biological systems by using linear programming. In Currents in Computational Molecular Biology (Universal Academy Press, Inc.). pp.8-9, 2000.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., Miyano, S. Characterization of subcellular localization signals using Hypotheis Creator. Genome Informatics 11:362-363, 2000.

Doi, A., Matsuno, H., Miyano, S. In Currents in Computational Molecular Biology (Universal Academy Press, Inc.). pp.26-27, 2000.

Dunker, K.A., Konagaya, A., Miyano, S., Takagi, T. (eds.). Genome Informatics 2000. Universal Academy Press, Inc. 2000.

Matsuno, H., Doi, A., Drath, R., Miyano, S. Genomic Object Net: object oritented reprentation of biological systems. Genome Informatics 11: 229-230, 2000.

Matsuno, H., Doi, A., Nagasaki, M., Miyano, S. Hybrid Petri net representation of gene regulatory network. PSB 2000, 5:341-352, 2000.

Maruyama, O. and Miyano, S. Design Aspects of Discovery Systems. IEICE Trans. Inf. and Sys. E83-D(1): 61-70, 2000.

Maruayama, O., Tamada, Y., Matsumoto, S., Miyano, S. ViewDesigner: a tool for designing views on data in discovery system HypothesisCreator. In Currents in Computational Molecular Biology (Universal Academy Press, Inc.). pp.22-23, 2000.

Miyano, S., Shinohara, A., Shinohara, T. Polynomial time learning of elementary formal systems. New Generation Computing 18:217-242, 2000.

Miyano, S. (ed.). Surveys on Discovery Science. IEICE Trans. Information and Systems E83-D (1), 2000.

Miyano, S., Shamir, R., Takagi, T. (eds.). Currents in Computational Molecular Biology (Universal Academy Press, Inc.), 2000.

Nagasaki, M., Onami, S., Miyano, S., Kitano, H. In Currents in Computational Molecular Biology (Universal Academy Press, Inc.). pp.28-29, 2000.

Ohta, N., Hachisu, Y., Akutsu, T., Fujiyama, A. Automatic spot measurement for genetic spot array images. Genome Informatics 11:262-263, 2000.

Shamir, R., Miyano, S., Istrail, S., Pevzner, P., Wateman, M. (eds.). RECOMB 2000 (ACM Press), 2000.

Shinohara, A., Iida, K., Takeda, M., Maruyama, O., Miyano, S., Kuhara, S. Finding sparse gene networks. Genome Informatics 11:249-250, 2000.

Shinohara, A., Takeda, M., Moriyama, T., Maruyama, O., Goto, T., Miyano, S., Muta, S., Tashiro, K., Kuhara, S. In Currents in Computational Molecular Biology (Universal Academy Press, Inc.). pp.24-25, 2000.

Somogyi, R., Kitano, H., Miyano, S., Zheng, Q. (eds.) Molecular Network Modeling and Data Analysis,

PSB 2000 (World Scientific), 5:291-292, 2000.
Tamada, Y., Bannai, H., Maruyama, O., Miyano, S. HypothesisCreator: a framework for knowledge discovery application development. Genome Informatics 11:360-361, 2000.
Tanaka, M., Nakazono, S., Matsuno, H., Tsujimoto, H., Kitamura, Y., Miyano, S. Intelligent system for topic survery in MEDLINE by keyword recommendation and learning text characteristics. Genome Informatics 11:73-82, 2000.
Yasuda, T., Bannai, H., Onami, S., Miyano, S., Kitano, H. In Currents in Computational Molecular Biology (Universal Academy Press, Inc.). pp.34-35, 2000.
阿久津達也，麻生川稔，小長谷 明彦：分子生物情報学の現状と動向 , 人工知能学会誌 15:3-10, 2000.
松野浩嗣，宮野悟：バーチャルセル構築のためのプラットフォーム．bit 32(8):22-31, 2000.
丸山　修，宮野　悟：ゲノム知識発見．実験医学 18(12):1635-1642, 2000.
丸山　修，宮野　悟（分担執筆）：発見システムとヒューマンエキスパートのインテグレーション（第18章）．発見科学とデータマイニング．bit 別冊．共立出版．pp.179-189, 2000.
宮野　悟：バイオインフォマティクスの新しい流れ．コンピュータソフトウェア 17(6):57-62, 2000.
宮野　悟，丸山　修：配列データからの知識発見技術．BIO Industry 7:37-44, 2000.
宮野悟，丸山　修（分担翻訳）：コンピュータ支援による科学的知識の発見（第13章）．発見科学とデータマイニング．bit 別冊．共立出版．pp.179-189, 2000.
森下真一，宮野　悟（編）：発見科学とデータマイニング．bit 別冊．共立出版．2000.

Human Genome Center
Laboratory of Molecular Medicine

The major goal of the Human Genome Project is to identify genes predisposing to diseases, and develop new diagnostic and therapeutic tools. We have been attempting to isolate genes involving in carcinogenesis and also those causing or predisposing to other diseases such as cardiovascular disease, bone disease, and some allergic diseases. By means of technologies developed through the genome project including high-resolution genetic maps, a large scale DNA sequencing, and the differential display method, we have isolated a number of biologically and/or medically interesting genes.

1. Genes playing significant roles in human cancer

a. Genes that are inducible by p53

Hirofumi Arakawa, Hiroshi Tanaka, Tatsuya Yamaguchi, Kenji Shiraishi, Seisuke Fukuda, Kuniko Matsui, Shu Okamura, Ching C. Ng, Yoshiki Takei, and Yusuke Nakamura

The p53 gene is inactivated more frequently than any other gene in a wide range of human cancers. In its normal state it exerts a tumor-suppressing function by regulating cell-cycle arrest or by inducing apoptosis. Its encoded product, a transcription factor, binds to DNA in a sequence-specific manner to activate transcription of target genes such as p21/WAF1, MDM2, BAX, and GADD45. Among the known targets, p21/WAF1 is a well-known mediator in the p53 signaling pathway that induces cell-cycle arrest at the G1 phase through inhibition of cyclin-dependent kinase. Without functional p21/WAF1, cells having wild-type p53 fail to arrest the cell cycle, therefore a normal complement of normal target genes appears to be essential for p53 to function effectively.

In addition to its important functions in cell cycle arrest and apoptosis, accumulating evidences implied a possible role of p53 in DNA repair to various genotoxic stresses; for example, fibroblasts homozygous for mutant p53 are deficient in global DNA repair. Wild-type p53 protein has been shown to be required for global genomic nucleotide excision repair in UV-induced DNA damage and to protect cell death after irradiation. Inactivation of p53 protein enhanced sensitivity to multiple chemotherapeutic drugs. Toward understanding these importnat physiological functions, identification of additional genes regulated by p53 is an inevitable step.

To establish another efficient approach for isolating p53-target genes, we established a colon-cancer cell line SW480-LOWp53-1 carrying a wild-type p53 transgene that is inducible under control of the lactose operon. Induction of this transgene by isopropyl-β-D-thiogalactoside (IPTG) arrests growth of the transfected cells. To investigate cellular responses related to the p53-signaling pathway to induce growth arrest, we applied a differential-display method to screen mRNAs isolated from this cell line, and looked for genes whose expression was activated or suppressed after induction of wild-type p53.

We isolated a novel p53-inducible gene, termed p53R2, using the differential-display method to examine mRNAs in a cancer-derived human cell line that carries a highly regulated wild-type p53 expression system. The p53R2 gene contained a p53-binding sequence in intron 1 and encoded a 351-amino-acid peptide that showed striking similarity to ribonucleotide reductase small subunit (R2). R2 plays an important role in DNA synthesis during cell division[1]. Expression of the novel homologue was

induced by gamma- and UV- irradiation and also by adriamycin treatment in a wild-type p53-dependent manner, while R2 was not. Induction of p53R2 expression in p53-deficient cells caused G2/M arrest and prevented cells from death in response to adriamycin. Inhibition of endogenous p53R2 expression in cells, which have an intact p53-dependent DNA damage checkpoint, reduced ribonucleotide reductase activity, DNA repair and cell survival after exposure to various genotoxins. Our results suggest that p53R2 is a novel ribonucleotide reductase gene that is directly involved in the p53 checkpoint for repair of damaged DNA. The discovery of p53R2 provides solid evidence to clarify a relationship between the ribonucleotide reductase activity involved in repair of damaged DNA and p53 tumor suppression.

We also reported an important p53-target gene, fractalkine, that is a CX3C-type chemokine that induces chemotaxis of monocytes and cytotoxic T-cells. Using the differential display method for examining gene expression in p53-defective cells transfected by adenovirus containing wild-type p53, we observed that *fractalkine* was induced by ectopic expression of p53. An electromobility-shift assay showed that a potential p53-binding site present in the promoter of the fractalkine gene could bind to p53 protein. Moreover, a heterogeneous reporter assay indicated that this promoter sequence possessed p53-dependent transcriptional activity. The strong induction of *fractalkine* when p53 protein was expressed by a wild-type transgene in p53-defective cells brought to light a novel role for p53; that is, potential elimination of damaged cells by the host immune-response system through transcriptional regulation of fractalkine. Our results imply a pivotal role of p53 in immuno-surveillance to prevent cells from undergoing malignant transformation.

Through direct cloning of p53-binding sequences from human genomic DNA, we have isolated a novel gene, designated p53AIP1 (p53-regulated Apoptosis Inducing Protein 1), whose expression is inducible by wild-type p53. Ectopically expressed p53AIP1, which is localized within mitochondria, leads to apoptotic cell death through dissipation of mitochondrial $\Delta\psi m$. We have found that upon severe DNA damage, Ser46 on p53 is phosphorylated and apoptosis is induced. In addition, substitution of Ser46 inhibits the ability of p53 to induce apoptosis, and selectively blocks expression of p53AIP1. Our results suggest that p53AIP1 is likely to play an important role in mediating p53-dependent apoptosis, and phosphorylation of Ser46 regulates the transcriptional activation of this apoptosis-inducing gene.

b. APC, β-catenin, and Axin in human cancer

Yoichi Furukawa, Seiji Satoh, Yataro Daigo, Tatsushi Katoh, Nobutomo Miwa, Tadashi Nishiwaki, Teru Kawasoe, Hideyuki Ishiguro, Manabu Fujita, Takashi Tokino, and Yusuke Nakamura

Wnt signaling is crucial for development and organogenesis. β–catenin is stabilized by the Wnt signaling stimulus, accumulates in the cytoplasm, binds to T-cell factor/lymphocyte enhancer binding factor (TCF/Lef), and then up-regulates downstream genes. Mutations of *CTNNB1* (β–catenin) or *APC* (Adenomatous Polyposis Coli) have been reported in a variety of human neoplasms including colon cancers and hepatocellular carcinomas (HCCs). Since HCCs tend to show accumulation of β–catenin more often than mutations in *CTNNB1* itself, we looked for mutations of *AXIN1*, another key factor for Wnt signaling, in six HCC cell lines and 100 primary HCCs. Among the four cell lines and 87 HCCs in which no *CTNNB1* mutations were detected, we identified *AXIN1* mutations in three cell lines and six mutations in five of the primary HCCs. In cell lines containing mutations of either gene, we observed increased DNA-binding of TCF associated with β–catenin in nuclei. Adenovirus mediated gene transfer of wild-type *AXIN1* induced apoptosis in hepatocellular and colorectal cancer cells that had accumulated β–catenin as a consequence of either *APC*, *CTNNB1*, or *AXIN1* mutation, suggesting that Axin could be an effective therapeutic molecule for suppressing growth of a wide range of hepatocellular and colorectal cancers.

We also isolated a novel murine gene, *Drctnnb1a* (down-regulated by *Ctnnb1*, a), whose expression was experimentally down-regulated in response to the activated form of β-catenin. To investigate a possible role of DRCTNNB1A in cancers, we also isolated the human homologue, *DRCTNNB1A*, whose deduced product was 91% identical to the murine protein. The transcript was expressed in all human tissues examined, and we assigned the genomic location of *DRCTNNB1A* to chromosomal band 7p15.3 by *in situ* hybridization. Expression of *DRCTNNB1A* in SW480 colon-cancer cells was significantly increased in response to reduction of intracellular β-catenin by adenovirus-mediated transfer of the β-catenin-binding domain of APC into the cells. Furthermore, we documented reduced expression of *DRCTNNB1A* in 12 of 15 primary colorectal cancers examined, compared with corresponding adjacent non-cancerous mucosae. Our results implied that *DRCTNNB1A* is one of the genes involved in the β-catenin-Tcf/Lef signaling pathway, and that reduced expression of *DRCTNNB1A* may have some role in colorectal carcinogenesis.

Furthermore, we found that expression of murine monocyte chemotactic protein-3 (mMCP-3) was suppressed by activated β-catenin. Inversely, expression of MCP-3 in human colon-cancer cells was induced by depletion of β-catenin following adenovirus-me-

diated transfer of wild-type *APC* genes into the cells. A reporter-gene assay indicated that accumulation of β-catenin in the nucleus suppressed activity of the MCP-3 promoter through a putative Tcf/LEF-binding site, ATCAAAG, but when the promoter sequence contained a two-base substitution in the binding site it failed to suppress reporter-gene (luciferase) activity. An electrophoretic mobility-shift assay using the putative Tcf/LEF-binding sequence revealed interaction of the candidate sequence with the β-catenin complex. Furthermore, induction of MCP-3 cDNA into HT-29 colon-cancer cells increased expression of two markers of differentiation, alkaline phosphatase and carcinoembryonic antigen. Our results implied that activation of β-catenin through the Tcf/LEF signaling pathway may participate in colonic carcinogenesis by inhibiting MCP-3-induced differentiation of colorectal epithelial cells.

In addition, we confirmed that expression of a murine gene encoding NBL4 (novel band 4.1-like protein 4) was up-regulated by activation of β-catenin. To examine a possible role of NBL4 in cancer, we isolated the human homologue of the murine *NBL4* gene by matching m*NBL4* against the human EST database followed by 5' rapid amplification of cDNA ends (5'RACE). The cDNA of *hNBL4* encoded a protein of 598 amino acids that shared 87% identity in amino acid sequence with murine NBL4 and 71% with zebrafish NBL4. A 2.2-kb hNBL4 transcript was expressed in all human tissues examined except heart, brain, and lung. We determined its chromosomal localization at 5q22 by fluorescence *in situ* hybridization. Expression of *hNBL4* was significantly reduced when β-catenin was depleted in SW480 cells, a human cancer cell line that constitutionally accumulates β-catenin. The results support the view that NBL4 is an important component of the β-catenin/Tcf pathway and is probably related to determination of cell polarity or proliferation.

To investigate the mechanisms underlying hepatocellular carcinogenesis, we attempted to identify genes regulated by β-catenin/Tcf complex in a human hepatoma cell line, HepG2, in which an activated form of β-catenin is expressed. By means of cDNA microarray, we isolated a novel human gene, termed *MARKL1* (*MAP/microtubule affinity-regulating kinase like 1*), whose expression was down-regulated in response to decreased Tcf/LEF1 activity. The transcript expressed in liver consisted of 3,529 nucleotides that contained an open reading frame of 2,256 nucleotides, encoding 752 amino acids homologous to human *MARK3*. Expression levels of *MARKL1* were markedly elevated in 8 of 9 HCCs in which nuclear accumulation of β-catenin were observed, which may suggest that MARKL1 plays some role in hepatocelular carcinogenesis.

c. Growth suppression of human ovarian cancer cells by adenovirus-mediated transfer of the *PTEN* Gene

Motoko Unoki, Mieko Matsushima, and Yusuke Nakamura

Among gynecological malignancies, ovarian cancer is the leading cause of death. Despite introduction of new chemotherapeutic agents into treatment protocols, to date no overall improvement has been achieved in long-term survival. Hence, developing alternative strategies is a matter of urgency. The *PIK3CA* gene on chromosome 3q26, which encodes the catalytic subunit of PI3K, is frequently increased in copy number in ovarian cancers; PI3K mediates a major growth-control pathway, acting both to stimulate cell growth and to block apoptosis. This pathway is antagonized by a tumor suppressor gene (*PTEN*) on chromosome 10q23, which encodes a phosphatidylinositol phosphatase. The PTEN product opposes activation of PIP2 and PIP3, second messengers downstream of PI3K. Although mutations of the *PTEN* gene were reported to be rare in ovarian cancers, gene transfer may be an effective therapy for this type of tumors, if a high dose of the PTEN product is able to block the activated PI3K-mediated cell growth pathway. Transfection of a PTEN expression plasmid into glioma cells indeed can suppress growth by arresting cells in the G1 phase. Moreover, adenovirus-mediated *PTEN* gene transfer into glioma cells is able to suppress tumorigenicity and induce apoptosis initiated by disruption of cells' interactions with the extracellular matrix.

We used cDNA microarrays containing 4009 cDNAs to examine changes in gene-expression profiles when exogenous PTEN was induced in PTEN-defective cells. The microarrays and subsequent semiquantitative RT-PCR analysis revealed transcriptional stimulation of 99 genes and repression of 72 genes. Some of the differentially expressed genes already had been implicated in cell proliferation, differentiation, apoptosis, or cell cycle control. For example, over-expression of PTEN induced transactivation of cyclin-dependent inhibitor 1B (*p27Kip1*) and 2B (*p15INK4B*), members of the TNF receptor family; TNF-associated genes; and members of the Notch-signaling and Mad families. To our knowledge this is the first report of transactivation of those genes by PTEN. The genes differentially expressed in our experiments also included many whose correlation with cancer development had not been recognized before. Our data should contribute to a greater understanding of the broad spectrum of ways in which PTEN affects intracellular signaling pathways. Analysis of expression profiles with microarrays appears to be a powerful approach for identifying anti-cancer genes and/or disease-specific targets for cancer therapy.

d. cDNA microarrat analysis of cancer

Osamu Kitahara, Yoichi Furukawa, Toshihiro Tanaka, Chikashi Kihara, Kenji Ono, Renpei

Yanagawa, Eiji M. Nita, Hideaki Ogasawara, Jyunich Okutsu, Hitoshi Zenbutsu, Norihiko Shiraishi, Toshihisa Takagi, Yusuke Nakamura, and Tatsuhiko Tsunoda

To identify a set of genes evincing altered expression during colorectal carcinogenesis, we used our original microarray (containing cDNAs corresponding to 9184 human genes) to compare expression profiles of colorectal-cancer cells from eight tumors with corresponding non-cancerous colonic epithelia. These cell populations had been rendered homogeneous by laser-capture microdissection. Altered expression was observed for 907 genes; of those, 56 showed increased levels (more than two-fold) and 851 showed decreased expression (less than half of normal) in cancer cells. Subsequent examination of ten selected genes (five from each category) by semi-quantitative RT-PCR substantiated the reliability of the microarray analysis. This extensive list of genes identified in these experiments provides a large body of potentially valuable information with respect to colorectal carcinogenesis, and also represents a source of novel targets for cancer therapy.

To identify genes involved in development or progression of ovarian cancer, we analyzed gene-expression profiles of nine ovarian tumors using a DNA microarray consisting of 9121 genes. Comparison of expression patterns between the carcinomas and corresponding normal ovarian tissues enabled us to identify 55 genes that were commonly up-regulated and 48 that were down-regulated in the cancer specimens. When the five serous adenocarcinomas were analyzed separately from the four mucinous adenocarcinomas, we identified 115 genes that were expressed differently between the two types of tumor. Investigation of these genes should help to disclose the molecular mechanism(s) of ovarian carcinogenesis, and define molecular separation of the two most common histological types of ovarian cancer.

2. Genes responsible for other diseases

a. Bronchial asthma

Sachiyo Takeoka, Motoko Unoki , Yoshihiro Onouchi , Satoru Doi[2], Hiroshi Fujiwara[2], Akihiko Miyatake[3], Kimie Fujita[4], Ituro Inoue[1], Yusuke Nakamura, and Mayumi Tamari:[1]Division of Genetic Diagnosis, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, Japan, [2]Osaka Prefectural Habikino Hospital, Osaka, Japan, [3]Miyatake Asthma Clinic, Osaka, Japan and [4]College of Nursing, University of Shiga, Japan

The complex etiology of bronchial asthma (BA), one of the most common inflammatory diseases throughout the world, involves a combination of various genetic and environmental factors. A number of linkage and association studies have been performed to shed light on the genetic background of BA, but the genetic aspects are still poorly understood. In the course of a project to screen the entire human genome for single nucleotide polymorphisms (SNPs) that might represent useful markers for large-scale association analyses of common diseases and pharmacogenetic traits, we identified six SNPs within the gene encoding IkB-associated protein (IKAP), a regulator of the NF-kB signal pathway. Most of the SNPs were in linkage disequilibrium each other. A strong allelic association between BA in childhood and two SNPs, T3214A (Cys1072Ser) and C3473T (Pro1157Leu), was observed (P=0.000004 for T3214A and P=0.0009 for C3473T). T3214A was also associated with BA in adult (P=0.000002), while C3473T was not (P=0.056). To confirm the above results, haplotype frequencies with six SNPs were estimated and compared between BA patients and controls. A strong association with the BA in childhood and a specific haplotype, 819T, 2295G, 2446A, 2490A, 3214A, and 3473T (haplotype TGAAAT), (P=0.00004, Odds ratio 2.94, 95%CI=2.48-3.4), where two amino-acid substitutions are present. Interestingly, the other haplotype TACGTC, in which the last five nucleotides were different from the haplotype TGAAAT, was inversely correlated with the BA phenotype (P=0.002, Odds ratio 9.83, 95%CI=8.35-11.31). These results indicated that specific variants of the *IKAP* or a variant in linkage disequilibrium with the specific haplotype might be associated with mechanisms responsible for early-onset BA.

Although intensive studies have attempted to elucidate the genetic background of bronchial asthma (BA), one of the most common of the chronic inflammatory diseases in human populations, genetic factors associated with its pathogenesis are still not well understood. We surveyed 29 possible candidate genes for this disease for single nucleotide polymorphisms (SNPs), the most frequent type of genetic variation, in genomic DNAs from Japanese BA patients. This effort identified 33 SNPs, only four of which had been reported previously, among 14 of those genes. We performed association studies using 585 BA patients and 343 normal controls for these SNPs. Of the 33 SNPs tested, 31 revealed no positive association with BA, but a T924C polymorphism in the thromboxane A2 receptor gene showed significant association ($\chi 2=4.71$, p=0.030), especially with respect to adult patients ($\chi 2=5.56$, p=0.018). Our results suggest that variants of the TBXA2R gene or some nearby gene(s) may play an important role in the pathogenesis of adult BA.

b. Rheumatoid arthritis

Ryo Yamada[1], Toshihiro Tanaka, Motoko Unoki, Tatsuo Nagai[2], Tetsuji Sawada[2], Yozo Ohnishi,

Tatsuhiko Tsunoda[1], Masao Yukioka[3], Akira Maeda[4], Kenji Suzuki[2], Hiroomi Tateishi[4], Takahiro Ochi[5], Yusuke Nakamura, Kazuhiko Yamamoto[2]: [1]SNP Research Center, Institute of Physical and Chemical Research (RIKEN), Tokyo, Japan, [2]Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, [3]Department of Orthopedic Surgery, Yukioka Hospital, [4]Sasayama- Hospital, Hyogo College of Medicine and [5]Department of Orthopedics, Osaka University Medical School

Genetic variants of interleukin-3 (IL-3), a well-studied cytokine, may have a role in the pathophysiology of rheumatoid arthritis (RA) but reports on this association sometimes conflict. We designed a case-control study to investigate association between RA and a single nucleotide polymorphism (SNP) in the IL-3 promoter region. Comparison of RA cases with controls yielded a chi square value of 14.28 (p=0.0002) with a genotype relative risk of 2.24 [95% CI, 1.44, 3.49]. When we compared younger-onset female cases with female controls, the SNP revealed an even more significant correlation, with a chi square value of 21.75 (p=0.000004) and a genotype relative risk of 7.27 [2.80, 18.89]. The stronger association we observed in this clinically distinct subgroup (females, early onset) within a region where linkage disequilibrium was not significantly extended suggested that the genuine RA locus should locate within or close to the IL-3 gene. We combined genotypic data for SNPs on eight other candidate genes with our IL-3 results, to estimate relationships between pairs of loci and RA by maximum-likelihood analysis. We discuss here the utility of combining genotypic data in this way to identify possible contributions of various genes to this disease.

c. Ulcerative colitis and Crohn's disease.

Kennoki Kyo, Tetsuichiro Muto[1], Hirokazu Nagawa[1], G. Mark Lathrop[2], Yusuke Nakamura: [1]Department of Surgical Oncology, The University of Tokyo and [2]Centre National de Genotypage, Evry, France

Ulcerative colitis (UC) and Crohn's disease (CD), the major forms of inflammatory bowel disease (IBD), are multifactorial disorders of unknown etiology. Earlier we reported a possible association of susceptibility to UC with rare alleles of a variable number of tandem repeat (VNTR) locus within the "MUC3" gene. However, the structure of "MUC3" has never been determined in its entirety because the long stretches of tandem repeats within it make cloning extraordinarily difficult. In the study reported here, we obtained evidence that the "MUC3" locus actually consists of two genes, MUC3A and MUC3B, both of which encode membrane-bound mucins with two epidermal growth factor-like motifs. We now can describe the complete 3' terminal structures of both genes. We also analyzed multiple single-nucleotide polymorphisms (SNPs) present in 3' exonic sequences, to investigate whether those sequence variations could account for person-to-person differences in susceptibility to IBDs. The results revealed that a non-synonymous SNP of MUC3A involving a tyrosine residue with a proposed role in cell signaling may confer genetic predisposition to CD (p=0.0132).

Publications

Y. Utada, S. haga, T. Kajiwara, F. Kasumi, F. Akiyama, G. Sakamoto, Y. Nakamura, and M. Emi: Allelic loss at the 8p22 region as a prognostic factor in large and estrogen receptor negative breast carcinomas. Cancer 88:1410-1416, 2000

Y. Utada, S. haga, T. Kajiwara, F. Kasumi, F. Akiyama, G. Sakamoto, Y. Nakamura, and M. Emi: Mapping of target regions of allelic loss in primary breast cancers to 1-cM intervals on genomic contigs at 6q21 and 6q25.3. Jpn. J Cancer Research 91:293-300, 2000

S. Isaka, Y. Takei, T. Tokino, Y. Miyoshi, K. Koyama, M. Suzuki, E. Takahashi, C. Azuma, Y. Murata, and Y. Nakamura: p53-inducible gene, TP53TG5, which suppresses growth and shows cell cycle-dependent transition of expression. Genes Chromosmes & Cancer 4:345-352, 2000

T. Tokino and Y. Nakamura: The role of p53-target genes in human cancer. Crinical reviews in oncology/hematology 33:1-6, 2000

H. Nakagawa, K. Koyama, Y. Murata, M. Morito, T. Akiyama,and Y. Nakamura: EB3, a novel member of the EB1 family preferentially expressed in the central nervous system, binds to a CNS-specific APC homologue. Oncogene in press

T. Nishiwaki, Y. Daigo, T. Kawasoe, and Y. Nakamura: Isolation and mutational analysis of a novel human cDNA, DEC1 (deleted in esophageal cancer 1), derived from the tumor-suppressor locus on chromosome 9q32. Genes Chromosomes & Cancer 27:169-176, 2000

H. Akashi, H.-J. Han, M. Iizaka, Y. Nakajima, S. Sugano, K. Imai and Y. Nakamura: Isolation and characterization of a human cDNA encoding a protein homologous to the 7.2-kDa protein (subunit X) of bovine ubiquinol-cytochrome C reduc-

tase. J Human Genetics 45:43-46, 2000
H. Nakagawa, K. Koyama, Y. Murata, M. Morito, T. Akiyama, Y. Nakamura: APCL, a CNS-specific homologue of APC, Binds to p53-Binding protein 2 and translocates it to the perinucleus. Cancer Research 60:101-105, 2000
T. Kinjo, M. Isomura, T. Iwamasa, and Y. Nakamura: Molecular cloning and characterization of two novel genes on chromosome 8p21.3. J. Human Genetics 45:12-17, 2000
M. Iizaka, H.-J. Han, H. Akashi, Y. Nakajima, S. Sugano, M. Ogawa, and Y. Nakamura: Isolation and chromosomal assignment of a novel human gene, hCRNN4, homologous to coronin-like actin-binding proteins. Cytogenet. Cell Genet. In press
C. Kihara, T. Seki, Y. Furukawa, H. Yamana, Y. Kimura, P. van Schaardenburgh, K. Hirata, and Y. Nakamura: Mutations in zinc-binding domains of p53 as a prognostic marker of esophageal-cancer patients. Jpn. J. Cancer Research 91:190-198, 2000
H. Akashi, H.-J. Han, M. Iizaka, Y. Nakajima, Y. Furukawa, S. Sugano, K. Imai and Y. Nakamura: Isolation and characterization of a novel gene encoding a putative seven-span transmembrane protein, TM7SF3 Cytogenet. Cell Genet. 88:305-309, 2000
T. Oyama, Y. Miyoshi, K. Koyama, H. Nakagawa, T. Yamori, T. Ito, H. Matsuda and Y. Nakamura: Isolation of a novel gene on human chromosome 8p21.3-p22 whose expression is reduced significantly in human colorectal cancers with liver metastasis. Genes Chromosomes & Cancer 29:9-15, 2000
S. Satoh, Y. Daigo, Y. Furukawa, T. Katoh, N. Miwa, T. Nishiwaki, T. Kawasoe, H. Ishiguro, M. Fujita, T. Tokino, Y. Sasaki, S. Imaoka, M. Murata, T. Shimano, Y. Yamaoka, and Y. Nakamura: Axin mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of wild-type Axin. Nature Genetics 24:245-250, 2000
R. Yamada, T. Tanaka, Y. Ohnishi, K. Suematsu, M. Minami, T. Seki, S. Tohma, K. Yamamoto, Y. Nakamura: Identification of 142 single nucleotide polymorphisms in 41 candidate genes for rheumatoid arthritis in the Japanese population. Human Genetics 106:293-297, 2000
Y. Ohnishi , T. Tanaka , R. Yamada , K. Suematsu, M. Minami, K. Fujii , N. Hoki , K. Kodama , S. Nagata , T. Hayashi , N. Kinoshita , H Sato, H. Sato, T. Kuzuya, H. Takeda, M. Hori, and Y. Nakamura: Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population. Human Genetics 106:288-292, 2000
M. Futamura, H. Arakawa, K. Matsuda, T. Katagiri, S. Saji, Y. Miki and Y. Nakamura: Potential role of BRCA2 in a mitotic checkpoint after phosphorylation by hBUBR1. Cancer Research, 60:1531-1535, 2000
18. H. Iwasa, T. Itoh, R. Nagai, Y. Nakamura, and T. Tanaka: Twenty single nucleotide polymorphisms (SNPs) and their alleleic frequencies in four genes responsible for familial long QT syndrome in the Japanese population. J. Human Genetics 45:182-183, 2000
M. Unoki, S. Furuta, Y. Onouchi, O. Watanabe, S. Doi, H. Fujiwara, A. Miyatake, K. Fujita, M. Tamari and Y. Nakamura: Association studies of 33 single nucleotide polymorphisms (SNPs) in 29 candidate genes for bronchial asthma: Positive association of a T924C polymorphism in the thromboxane A2 receptor gene Human Genetics 106:44-446, 2000
S. Okuno, T.K. Watanabe, T. Ono, Y. Yamasaki, Y. Goto, H. Miyao, T. Asai, N. Kanemoto, K. Ogawa, A. Mizoguchi-Miyatake, T. Takagi, E. Takahashi, Y. Nakamura, and A. Tanigami: Genetic determinants of plasma triglyceride levels in (OLETF X BN) X OLETF backcross rats. Genomics 62:356-368, 2000
Z.-Q. Yang, M.A. Yoshida, Y. Fukuda, N. Kurihara, Y. Nakamura, and J. Inazawa: Molecular cytogenetic analysis of 17 renal cancer cell lines: Increased copy number at 5q31-33 in cell lines from nonpapillary carcinomas. Jpn. J. Cancer Research 91:156-163, 2000
H. Tanaka, H. Arakawa, K. Shiraishi, T. Yamaguchi, Y. Takei, and Y. Nakamura: A ribonucleotide reductase gene involved in a p53 dependent DNA damage checkpoint. Nature, 404:42-49, 2000
H. Ishiguro, Y. Furukawa, Y. Daigo, Y. Miyoshi, Y. Nagasawa, T. Nishiwaki, T. Kawasoe, M. Fujita, S. Satoh,ÇP N. Miwa, Y. Fujii, and Y. Nakamura: Isolation and characterization of human *NBL4*, a gene involved in the β-catenin/Tcf signaling pathway. Jpn. J. Cancer Res. 91:597-603, 2000
Y.-M. Lin, T. kato, S. Satoh, Y. Nakamura and Y. Furukawa: Identification of novel polymorphisms in the AXIN1 and CDX-2 genes. J. Human Genetics, 45:254-256, 2000
K. Shiraishi, S. Fukuda, T. Mori, K. Matsuda, T. Yamaguchi, C. Tanikawa, M. Ogawa, Y. Nakamura and H. Arakawa: Identification of fractalkine, a CX3C-type chemokine, as a direct target of p53. Cancer Research 60:3722-3726, 2000
L. C. McCarthy, M.-T. Bihoreau, S. L. Kiguwa, J. Browne, T. K. Watanabe, H. Hishigaki, A. Tsuji, S. Kiel, C. Webber, M. E. Davis, C. Knights, A. Smith, R. Critcher, P. Huxtall, J. R. Hudson, Jr., T. Ono, H. Hayashi, T. Takagi, Y. Nakamura, A. Tanigami, P. N. Goodfellow, G. M. Lathrop and M. R. James: A whole-genome radiation hybrid panel andframework map of the rat genome. Mammalian Genome, In press
T. Kawasoe, Y. Furukawa, Y. Daigo, T. Nishiwaki, H. Ishiguro, M. Fujita, S. Satoh, N. Miwa, Y. Nagasawa, Y. Miyoshi, M. Ogawa and Y.

Nakamura: Isolation and characterization of a novel human gene, DRCTNNB1A, whose expression is down-regulated by β-catenin Cancer Research 60:3354-3358, 2000

28. M. Nakanishi, C. Sakakura, Y. Fujita, R. Yasuoka, H. Aragane, K. Koide, A. Hagiwara, T. Yamaguchi, Y. Nakamura, T. Abe, J. Inazawa, and H. Yamagishi: Genomic alterations in primary gastric cancers analyzed by comparative genomic hybridization and clinicopathological factors. Hepato-Gastroenterology 47:658-662, 2000

T.K. Watanabe, T. Ono, S. Okuno, A. Mizoguchi-Miyashita, Y. Yamasaki, N. Kanemoto, H. Hishigaki, K. Oga, E. Takahashi, Y. Irie, M. Bihoreau, M. R. James, G.M. lathrop, T. Takagi, Y. Nakamura, and A. Tanigami: Characterization of newly developed SSLP markers for the rat. Mammalian Genome 11:300-305, 2000

M. Higashiyama, Y. Miyoshi, K. Kodama, H. Yokouchi, K. Takami, M. Nishijima, T. Nakayama, H. Kobayashi, K. Minamigawa, and Y. Nakamura: p53-regulated GML gene expression in non-small cell lung cancer: a promising relationship to cisplatin chemosensitivity. Europian Journal of Cancer 36:489-495, 2000

M. Horie, Y. Mitsumoto, H. Kyushiki, N. Kanemoto, A. Watanabe, Y. Taniguchi, N. Nishino, T. Okamoto, M. Kondo, T. Mori, K. Noguchi, Y. Nakamura, E. Takahashi, and A. Tanigami: Identification and characterization of TMEFF2, a novel survival factor for hippocampal and mesencephalic neurons. Genomics 67:146-152, 2000

32. M. Murata, K. Iwao, Y. Miyoshi, Y. Nagasawa, M. Yabu, S. Himeno, K. Imanishi, M. Ohsawa, H. Wada, S. Tominaga, T. Shimano, T. Kobayashi, and Y. Nakamura: Activation of β-catenin gene by interstitial deletions involving exon 3 as an early event in colorectal tumorigenesis. Cancer Letters 159:73-78, 2000

H. Akashi, H.-J. Han, M. Iizaka, and Y. Nakamura: Growth-suppressive effect of NSAIDs on 11 colon-cancer cell lines and fluorescence differential display of genes whose expression is influenced by sulindac. I. J. Cancer, in press

T. Yamauchi, M. Tada, K. Houkin. T. Tanaka, Y. Nakamura, S. Kuroda, H. Abe. T. Inoue, K. Ikezaki, T. Matsushima, M. Fukui: Linkage of familial Moyamoya disease (spotaneous occlusion of the circle of Willis) to chromosome 17q25. Stroke 31:930-935, 2000

S. Matsumoto, K. Minobe, Y. Utada, K. Furukawa, M. Onda, G. Sakamoto, F. Kasumi, Y. Nakamura, and M. Emi: Loss of heterozygosity at 3p24-p25 as a prognostic factor in breast cancer. Cancer Letter 152:63-69, 2000

W.-M. Qu, T. Miyazaki, M. Terada, L.-M. Lu, M. Nishihara, A. Yamada, S. Mori, Y. Nakamura, H. Ogasawara, C. Yazawa, S. Nakatsuru, and M. Nose: Genetic dissection of vasculitis in MRL/lpr lupus mice: a novel susceptibility locus involving the CD72^c Allele. Eur. J. Immunolo. 30:2027-2037, 2000

M. Murata, K. Iwao, Y. Miyoshi, Y. Nagasawa, T. Ohta, K. Shibata, K. Oda, H. Wada, S. Tominaga, Y. Matsuda, M. Ohsawa, Y. Nakamura and T. Shimano: Molecular and biological analysis of carcinoma of the samll intestine: β-catenin gene mutation by interstitial deletion involving exon 3 and replication error phenotype. Am. J. Gastroenterology 95:1576-1580, 2000

K. Ono, T. Tanaka, T. Tsunoda, O. Kitahara, C. Kihara, A. Okamoto, K. Ochiai, T. Takagi, and Y. Nakamura: Identification by cDNA microarray of genes Involved in ovarian carcinogenesis. Cancer Research, in press

T. Seki, T. Tanaka, and Y. Nakamura: Genomic structure and multiple single-nucleotide polymorphisms (SNPs) of the thiopurine S-methyltransferase(TPMT) gene. J. Human Genetics, in press

K. Oda, H. Arakawa, T. Tanaka, K. Matsuda, C. Tanikawa, T. Mori, H. Nishimori, K. Tamai, T. Tokino, Y. Nakamura, and Y. Taya: Isolation of a novel gene encoding a mitochondrial protein, *p53AIP1*, a potential mediator of p53-dependent apoptosis, and its regulation by Ser46-phosphorylated p53. Cell 102:849-862, 2000

K. Yamada, H. Tomita, K. Yoshiura, S. Kondo, K. Wakui, Y. Fukushima, S. Ikegawa, Y. Nakamura, T. Amemiya, and N. Niikawa: An autosomal dominant posterior polar cataract locus maps to human chromosome 20p12-q12. Eur. J. Human Genet. 8:535-539, 2000

Y. Utada, M. Emi, S. Haga, M. Yoshomoto, F. Kasumi, F. Akiyama, G. Sakamoto, S. Haga, T. Kajiwara, and Y. Nakamura: Allelic loss at 1p34-36 predict poor prognosis in node-negative breast cancer. Clinical Cancer Research 6:3193-3198, 2000

Z.-Q. Yang, I. Imoto, Y. Fukuda, A. Pimkhaokham, Y. Shimada, M. Imamura, S. Sugano, Y. Nakamura, and J. Inazawa: Identification of a novel gene, GASC1, with an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. Cancer Research 60:4735-4739, 2000

M.-S. Wu, K. Tani, H. Sugiyama, H. Hibino, K. Izawa, T. Tanabe, Y. Nakazaki, H. Ishii, J. Ohashi, H. Hohjoh, T. Iseki, A. Tojo, Y. Nakamura, Y. Tanioka, K. Tokunaga, and S. Asano: MHC (Major Histocompatibility Complex)-DRB genes and polymorphisms in common marmoset. Journal of Molecular Evolution 51:214-222, 2000

A. Pimkhaokham, Y. Shimada, Y. Fukuda, N. Kurihara, I. Imoto, Z.-Q. Yang, M. Imamura, Y. Nakamura, and J. Inazawa: Nonrandoom chromosomal imbalances in esophageal squamous cell carcinoma cell lines: possible involvement of the ATF3 and CENPF genes in the 1q32 amplicon. Jpn. J. Cancer Research 91:1126-1133, 2000

J. Sasaki, K. Ishikawa, K. Kobayashi, E. Kondo-Iida, M. Fukuyama, H. Mizusawa, S. Takashima, Y. Sasakihara, Y. Nakamura, and T. Toda: Neuronal expression of the fukutin gene. Human Molecular Genetics in press

M.E. Nita, S.K. Ono-Nita, N. Tsuno, O. Tominaga, T. Takenoue, E. Sunami, J. Kitayama, Y. Nakamura, and H. Nagawa: Bcl-XL antisense sensitizes human colon cancer cell line to 5-fluorouracil. Jpn. J. Cancer Res. 91:825-832, 2000

T. Nakajima, M. Kutabayashi, Y. Ohyama, Y. Kaneko, T. Furukawa, T. Itoh, Y. Taniguchi, T. Tanaka, Y. Nakamura, M. Hiraoka, and R. Nagai: Characterization of S818L mutation in HERG c-terminus in LQT2; Modification of activation-deactivation gating properties. FEBS Letters 481:197-203, 2000

M. Fujita, Y. Furukawa, Y. Nagasawa, M. Ogawa and Y. Nakamura: Down-regulation of monocyte chemotactic protein-3 (MCP-3) expression by activated β-catenin. Cancer Research, in press

K. Kyo, T. Muto, H. Nagawa, G. M. Lathrop, and Y. Nakamura: Distinct variants of the intestinal mucin gene MUC3A associated with ulcerative colitis and Crohn's disease. J. Human Genetics in press

51. O. Kitahara, Y. Furukawa, T. Tanaka, C. Kihara, K. Ono, R. Yanagawa, E. M. Nita, H. Ogasawara, J. Okutsu, H. Zenbutsu, N. Shiraishi, T. Takagi, Y. Nakamura, and T. Tsunoda: Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Research, in press

T. Itoh, K. Kikuchi, Y. Odagawa, S. Takata, K. Yano, S. Okada, N. Haneda, S. Ogawa, O. Nakano, Y. Kawahara, H. Kasai, T. Nakayama, T. Fukutomi, H. Sakurada, A. Shimizu, Y. Yazaki, R. Nagai, Y. Nakamura, and T. Tanaka: Correlation of genetic etiology with response to β-adrenergic blockade among symptomatic patients with familial long-QT syndrome. J. Human Genetics in press

S. Takeoka, M. Unoki, Y. Onouchi, S. Doi, H. Fujiwara, A. Miyatake, K. Fujita, Ituro Inoue, Y. Nakamura, and M. Tamari: Amino-acid substitutions in the *IKAP* significantly increase a risk of childhood bronchial asthma. J. Human Genetics, in press

R. Yamada, T. Tanaka, M. Unoki T. Nagai, T. Sawada, Y. Ohnishi, T, Tsunoda, M. Yukioka, A. Maeda, K. Suzuki, H. Tateishi, T. Ochi, Y. Nakamura, and K. Yamamoto: Association between an SNP in the promoter of the human interleukin-3 gene and rheumatoid arthritis (RA) in Japanese patients, and maximum-likelihood estimation of combinatorial effect of two genetic loci on susceptibility to the disease. Am. J. Human Genetics in press

M. Matsushima-Nishiu, M. Unoki, K. Ono, T. Tsunoda, T. Minaguchi, H. Kuramoto, M. Nishida, T. Satoh, T. Tanaka, and Y. Nakamura: Microarray analysis of gene-expression profiles in endometrial cancer cells expressing exogenous PTEN. Cancer Research, in press

T. Kato, S. Satoh, H. Okabe, O. Kitahara, K. Ono, C. Kihara, T. Tanaka, T. Tsunoda, Y. Yamaoka, Y. Nakamura, and Y. Furukawa: Isolation of a novel human gene, *MARKL1*, homologous to *MARK3* and its involvement in hepatocellular carcinogenesis. Neoplasia in press

C. Suzuki, M. Unoki, and Y. Nakamura: Identification and allelic frequencies of novel single-nucleotide polymorphisms in the DUSP1 and BTG1 genes. J Human Genetics in press

Human Genome Center

Laboratory of Functional Genomics

We are interested in sequence-based analysis of human genome, functional analysis of the yeast genome, molecular mechanism regulating mammalian circadian rhythms, and the hunting of genes with unique expression patterns.

1. Sequence analysis of human chromosome 21

Masahira Hattori[1], Todd Taylor[1], Kouhei Maekawa, Kunihiko Takamatsu, Tadayuki Takeda[1] and Yoshiyuki Sakaki:[1]RIKEN, Genomic Sciences Center, Yokohama

Human chromosome 21 is approximately 50 Mb (1 Mb=1000 kb) in size and represents a model for physical mapping of the human genome. Several neurological genetic diseases are associated with this chromosome, including Down's syndrome (DS), familial Alzheimer disease (FAD), familial amyotrophic lateral sclerosis, and progressive myoclonic epilepsy. For the analysis of complex genomes, we established BAC (insert size: -170 kb) and fosmid (insert size: -35 kb) screening systems using total human and sorted chromosome 21 DNA to facilitate the contig construction. Using this system, we have constructed contigs covering about 50% of the chromosome 21 including APP, Down syndrome critical region, D21A8 to APP region, q11.1 centromeric region, q22.3 telomeric region and D21S226 to AML region. These contigs were subjected to sequencing.

In the framework of an international collaboration, we have sequenced 33,546,361 base pairs (bp) of DNA with very high accuracy, the largest contig being 25,491,867 bp. Only three small clone gaps and seven sequencing gaps remain, comprising about 100 kilobases. Thus, we achieved 99.7% coverage of 21q. We also sequenced 281,116 bp from the short arm. The structural features identified include duplications that are probably involved in chromosomal abnormalities and repeat structures in the telomeric and pericentromeric regions. Analysis of the chromosome revealed 127 known genes and 59 pseudogenes.

Comparative sequence analysis is an effective method to identify functional elements and to help elucidate their role in the pathogenesis of DS, which exhibits various phenotypes. Toward this end, a 1.35 Mb segment of the syntenic region in mouse was sequenced. This genomic region has a GC content of average level (42.4% for human and 45.4% for mouse). The frequency of repetitive elements was 41.6% and 27.7% for human and mouse, respectively. The gene distribution and direction of transcription are highly conserved between the two species. We found 38 conserved non-coding regions (>75% nucleotide identity over >100bp) through comparative analysis. This study provides valuable data that should help us to understand the molecular mechanism of and facilitate the identification of medically important genes and regulatory elements.

2. Comprehensive analysis of protein-protein interactions in the budding yeast

Takashi Ito[2], Kazuhisa Ota[2], Hiroyuki Kubota, Ritsuko Ozawa, Tomoko Chiba[2] and Yoshiyuki Sakaki:[2]Cancer Research Institute, Kanazawa Univ.

Protein-protein interactions play crucial roles in the execution of various biological functions. Accordingly, their comprehensive description would considerably contribute to the functional interpretation of fully-sequenced genomes, that are flooded with novel genes of unpredictable functions. We pre-

viously developed a system to examine two-hybrid interactions in all possible combinations between the ~6,000 proteins of the budding yeast *Saccharomyces cerevisiae*. We have recently completed the comprehensive analysis using this system to identify 4,549 two-hybrid interactions among 3,278 proteins, thereby providing the largest dataset ever generated (http://genome.c.kanazawa-u.ac.jp/Y2H/). Unexpectedly, these data do not largely overlap with those obtained by the other large-scale two-hybrid project performed at CuraGen and hence have substantially expanded our knowledge on the protein interaction space or interactome of the yeast.

Cumulative connection of these binary interactions generates a single huge network linking the vast majority of the proteins. Bioinformatics-aided selection of biologically relevant interactions highlights various intriguing subnetworks. They include, for instance, the one that had successfully foreseen the involvement of a novel protein in spindle pole body function as well as the one that may uncover a hitherto unidentified multiprotein complex potentially participating in the process of vesicular transport. Our data would thus significantly expand and improve the protein interaction map for the exploration of genome functions, that eventually leads to thorough understanding of the cell as a molecular system.

We are trying to extract protein interaction modules from these binary interaction data using a novel "two-hybrid footprinting" method and a data mining approach. These modules will be used for the perturbation of cognate interactions to examine their biological roles and lay the basis for general interaction rules to predict other interactions. For instance, we revealed that a novel GI domain occurring at the N-termius of the eIF2 kinase Gcn2p mediates its association with Gcn1p and that the interaction is essential for the activation of the kinase and hence stress response at the translational level. The GI domain is found in various proteins of apparrently unrelated origins, and hence is assumed to be a novel modular domain for specific protein-protein interactions.

3. Transcriptomic approach to gene expression networks in the budding yeast

Fumihito Miura, Miyuki Onda, Kazuhisa Ota, Takashi Ito and Yoshiyuki Sakaki

For the systematic analysis of gene-gene interactions or gene regulatory networks, we developed a PCR-based knock-in strategy to make a transcription factor constitutively active by replacing its activation domain with that of VP16. This strategy would facilitate the search of downstream target genes even in the absence of activating signals from upstream, which cannot be known a priori for novel transcription factors. To achieve rapid and accurate expression profilinig, we are developing a composite system using DNA microarray and our unique adapter-tagged competitive PCR for rapid for the initial screening and confirmative quatitation of differential gene expression, respectively.

4. A novel G protein-coupled receptor gene expressed in striatum

Kazuyuki Mizushima, Kazuhisa Ota, Takashi Ito and Yoshiyuki Sakaki

Differential display screening for region-specific transcripts in rat brain revealed a novel striatum-specific transcript encoding an orphan G protein-coupled receptor (GPCR) designated as Strg/Gpr88 for striatum-specific GPCR. We isolated its homologues from human and mouse, and mapped them to chromosomes 1p21.3 and 3G1, respectively. These loci are syntenic to each other, thereby suggesting their orthology. The predicted primary sequences of Strg/Gpr88 proteins are highly conserved between human and rodents, and show highest homology to receptors for biogenic amines. However, *Strg/Gpr88* lacks some residues conserved in all known biogenic amine receptors and hence may represent a novel subtype of GPCRs. Northern blot and *in situ* hybridization analyses revealed that *Strg/Gpr88* transcripts are expressed almost exclusively in striatum in both human and rodents. Remarkable conservation in primary structure and unique expression pattern may imply a role for *Strg/Gpr88* in fundamental functions of striatum such as the control of motor behavior. To directly reveal the function of *Strg/Gpr88* at the level of animal, gene targeting experiments are currently underway. We are also interested in the molecular mechanisms for the striatum-specific expression. To identify relevant cis-elements, we are analyzing transgenic mice bearing variously truncated forms of *Strg/Gpr88* promoter fused to *lacZ* or *GFP* gene. For the search of candidates for trans-acting factors involved in the expression, we perfromed a yeast one-hybrid screening to revealed that the bHLH transcription factor SEF2 specifically binds to an evolutionarily conserved pair of E-boxes in this gene.

5. Analyses of genes with unique expression patterns

Yuriko Hagiwara-Takeuchi, Yoichi Yamada, Kohji Okamura, Takashi Ito and Yoshiyuki Sakaki

We developed a unique Allelic Message Display (AMD) screening for imprinted genes and identified a paternally expressed gene *Impact*, encoding a remarkably conserved protein of unknown function. We isolated its human homolog *IMPACT* and showed that it is loosely imprinted in some people

but not in the others (i.e. polymorphically). We thus elucidated the genome structures for both mouse and human genes to reveal a unique CpG island which bears characteristic tandem repeats, is subjected to parent-of-origin dependent methylation, and is lost from the non-imprinted human gene. The role for this island in genomic imprinting is being analyzed. We also identified and are analyzing a novel gene subjected to tissue-specific imprinting in addition to some mono-allelically expressed genes using nuclear-transplanted mice.

6. Molecular mechanisms regulating mammalian circadian clock

Hajime Tei, Akiko Hida, Rika Numano, Nobuya Koike, Shihoko Kojima, Yoko Sato, Satomi Shiozuka, Matsumi Hirose, Moe Kimura, Miyuki Shimada and Yoshiyuki Sakaki

Many biochemical, physiological and behavioral processes in many organisms exhibit circadian rhythms. Circadian rhythms are driven by autonomous oscillators and entrained by daily light-dark cycles. The transcription of *Per1*, a mammalian clock gene, oscillates in a circadian manner in the mouse suprachiasmatic nucleus (SCN; the central pacemaker of the mammalian circadian clock) with a peak in the daytime and a trough at night. In addition, the expression of m*Per1* in the SCN is induced immediately by a light pulse even at night. Function of the circadian expression of the mammalian *Per1* gene is a key question for the regulation of circadian rhythms. For elucidation of the molecular mechanisms controlling the mammalian circadian clock, the genomic sequences of the human and mouse *Per1* genes in addition to their transcriptional start sites have been determined. Both of the genomic sequences consist of 23 exons spanning approximately 16 kb. Comparisons of both genes revealed five and one conserved segments in the 5' flanking regions and the first introns, respectively. These conserved segments contained several potential regulatory elements such as five E-boxes (the binding site for the Clock-Bmal1 complex). Transfection analyses using a series of deletion and point mutants of the m*Per1*::*luc* reporter showed that each of the five E-boxes was functional for the *Per1* induction mediated by Clock and Bmal1. Second, We generated a *Per1*::*luc* transgenic rat line in which luciferase is rhythmically expressed under the control of the mouse *Per1* promoter, and have used it to study mammalian circadian organization. Light emission from cultured suprachiasmatic nuclei (SCN) of these rats was invariably and robustly rhythmic. Liver, lung, and skeletal muscle expressed damped circadian rhythms *in vitro*. In response to advances and delays of the environmental light cycle, the circadian rhythm of light emission from the SCN shifted more rapidly than did the rhythm of locomotor behavior. We hypothesize that self-sustained circadian oscillators in the SCN entrain damped circadian oscillators in the periphery to maintain adaptive phase control. Third, we constructed transgenic rat lines with constitutive expression of the mouse *Per1* gene using *Elongation 1 alpha* or *Neural specific enolase* promoters. Both the circadian period of locomoter activity and entrainment to light-dark cycles were severely affected in several transgenic lines. In addition, we measured the expression of the native (rat) *Per1* and *Per2* genes in the SCN and retina of the transgenic lines under DD conditions. The circadian expression of endogenous *Per1* and *Per2* genes was diminished in the transgenic lines. These results clearly indicate that the circadian expression and light induction of the mammalian *Per1* gene is involved in rhythm generation and/or entrainment of the circadian clock.

Publications

Suzuki, T., Nishiyama, K., Yamamoto, A., Inazawa, J., Iwaki, T., Yamada, T., Kanazawa, I. and Sakaki, Y. Molecular cloning of a novel apoptosis-related gene, human Nap1 (NCKAP1), and its possible relation to Alzheimer disease. Genomics 63: 246-254 (2000).

Slavov, D., Hattori, M., Sakaki, Y., Rosenthal, A., Shimizu, N., Minoshima, S., Kudoh, J., Yaspo, M-L., Ramser, J., Reinhardt, R., Reimer, C., Clancy, K., Rynditch, A. and Gardiner, K. Criteria for gene identification and features of genome organization: analysis of 6.5 Mb of DNA sequence from human chromosome 21. GENE 247: 215-232 (2000).

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.-S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.-K., Soeda, E., Ohki, M., Takagi, T., Sakaki Y.; Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A. Zimmermann, W., Rosenthal A.; kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nakagome, K., Mitsuyama, S., Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Brandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloecker, H.; Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Haaf, T., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and Yaspo, M-L. The DNA sequence

of human chromosome 21. Nature 405: 311-319 (2000).

Togashi, T., Choi, D.-K., Taylor, T. D., Suzuki, Y., Sugano, S., Hattori, M. and Sakaki, Y. A novel gene, DSCR5, from the distal Down syndrome critical region on chromosome 21q22.2. DNA Research 7: 207-212 (2000).

Yatsuki, H., Watanabe, H., Hattori, M., Joh, K., Soejima, H., Komoda, H., Xin, Z.-H., Zhu, X.-K., Higashimoto, K., Nishimura, M., Kuratomi, S., Sasaki, H., Sakaki Y. and Mukai, T. Sequece-based structural features between Kvlqt1 and Tapa1 on mouse chromosome 7F4/F5 corresponding to the Beckwith-Wiedemann syndrome region on human 11p15.5: long-stretches of unusually well conserved intronic sequences of Kvlqt1 between mouse and human. DNA research 7: 195-206 (2000).

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. and Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature 407: 81-86 (2000).

Chen, J., Wang, Y.-P., Wang, Y., Nakajima, T., Iwasawa, K., Hikiji, H., Sunamoto, M., Choi, D.-K., Yoshida, Y., Sakaki, Y. and Toyo-oka, T. Autocrine action and its underlying mechanism on nitric oxide on intracellular Ca2+ homeostasis in vascular endothelial cells. J. Biol. Chem. 275: 28739-28749 (2000).

Ohira, M., Kageyama, H., Mihara, M., Furuta, S., Machida, T., Shishikura, T., Takayasu, H., Islam, A., Nakamura, Y., Takahashi, M., Tomioka, N., Sakiyama, S., Kaneko, Y., Toyoda, A., Hattori, M., Sakaki, Y., Ohki, M., Horii, A., Soeda, E., Inazawa, J., Seki, N., Kuma, H., Nozawa, I. and Aakagawara, A. Identification and characterization of a 500-kb homozygously deleted region at 1p36.2-p36.3 in a neuroblastoma cell line. Oncogene 19: 4302-4307 (2000).

Nakagawara, A., Ohira, M., Kageyama, H., Mihara, M., Furuta, S., Machida, T., Takayasu, H., Islam, A., Nakamura, Y., Takahashi, M., Shishikura, T., Kaneko, Y., Toyoda, A., Hattori, M., Sakaki, Y., Ohki, M., Horii, A., Soeda, E., Inazawa, J., Seki, N., Kuma, H., Nozawa, I. and Sakiyama, S. Identification of the homozygously deleted region at chromosome 1p36.2 in human neuroblastoma. Med. Pediatr. Oncol. 35: 516-521 (2000).

Matsuo, R., Murayama, A., Saitoh, Y., Sakaki, Y. and Inokuchi, K. Identification and cataloging of genes induced by long-lasting long-term potentiation in awake rats. J. Neurochem. 74: 2239-2249 (2000).

Levanon, D., Glusman, G., Bangsow, T., Ben-Asher, E., Male, D.A., Avidan, N., Bangsow, C., Hattori, M., Taylor, T.D., Taudien, S., Blechschmidt, K., Shimizu, N., Rosenthal, A., Sakaki, Y., Lancet, D. and Groner, Y. Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. GENE 262: 23-33 (2001).

Wang, Y., Chen, J., Wang, Y., Taylor, C.W., Hirata, Y., Hagiwara, H., Mikoshiba, K., Toyo-Oka, T., Omata, M. and Sakaki, Y. Crucial Role of Type 1, but Not Type 3, Inositol 1,4,5-Trisphosphate (IP(3)) Receptors in IP(3)-Induced Ca(2+) Release, Capacitative Ca(2+) Entry, and Proliferation of A7r5 Vascular Smooth Muscle Cells. Circ. Res. 88: 202-209 (2001).

Shiose, A., Kuroda, J., Tsuruya, K., Hirai, M., Hirakata, H., Naito, S., Hattori, M., Sakaki, Y, and Sumimoto, H. A Novel Superoxide-producing NAD(P)H Oxidase in Kidney. J. Biol. Chem. 276: 1417-1423 (2001).

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409: 860-921 (2001).

Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S. and Sakaki, Y. Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc. Natl. Acad. Sci. USA 97: 1143-1147 (2000).

Iijima, Y., Ito, T., Oikawa, T., Eguchi, M., Eguchi-Ishimae, M., Kamada, N., Kishi, K., Asano, S., Sakaki, Y. and Sato, Y. A new ETV6/TEL partner gene, ARG (ABL related gene or ABL2), identified in an AML-M3 cell line with a t(1;12)(q25;p13) translocation. Blood 95: 2126-2131 (2000).

Kubota, H., Sakaki, Y. and Ito, T. GI domain-mediated association of the eukaryotic initiation factor 2_ kinase GCN2 with its activator GCN1 is required for general amino acid control in budding yeast. J. Biol. Chem. 275: 20243-20246 (2000).

Mizushima, K., Miyamoto, Y., Tsukahara, F., Hirai, M., Sakaki, Y. and Ito, T. A novel G-protein coupled receptor gene expressed in striatum. Genomics 69: 314-321 (2000).

Okamura, K., Hagiwara-Takeuchi, Y., Li, T., Vu, T. H., Hirai, M., Hattori, M., Sakaki, Y., Hoffman, A. R. and Ito, T. Comparative genome analysis of the mouse imprinted gene *Impact* and its nonimprinted human homolog *IMPACT*: Toward the structural basis for species-specific imprinting. Genome Res. 10: 1878-1889 (2000).

Hida, A., Koike, N., Hirose, M., Hattori, M., Sakaki, Y. and Tei, H. The human and mouse *Period1* genes: five well-conserved E-boxes additively contribute to the enhancement of m*Per1* transcription. Genomics 65: 224-233 (2000).

Yamazaki, S., Numano, R., Abe, M., Hida, A., Takahashi, R., Ueda, M., Block, G.D., Sakaki, Y., Menaker, M. and Tei, H. Resetting central and peripheral circadian oscillators in transgenic rats. Science 288: 682-685 (2000).

Stokkan, K-A., Yamazaki, S., Tei, H., Sakaki, Y. and Menaker, M. Entrainment of the circadian clock in the liver by feeding. Science 291: 490-493 (2001).