第2回「計算科学による新たな知の発展・統合・創出」シンポ
ー計算科学の戦略と次世代スーパーコンピューター

「ゲノム研究における超高速計算機システムの活用」

つくば国際会議場（エポカルつくば）
２００６年4月5日（水）

国立遺伝学研究所
生命情報・DDBJ研究センター
五條堀　孝

膨張し続ける生命情報はどこに保管されているのか

NCBI USTPO

GenBank

1982年～

米国

NIH

毎年1回ずつの持ち回り：
国際諮問委員会・国際実務者会議
（三島・ケンブリッジ・ワシントンDC）

DDBJ DNA Data Bank of Japan JPO

DDBJ

1986年～

日本

遺伝研

EBI EPO

EMBL

1980年～

欧州(17ヶ国)

英国・EBI

遺伝とは、親のもつ性質を子に伝えることをいい、その担い手は遺伝子です。

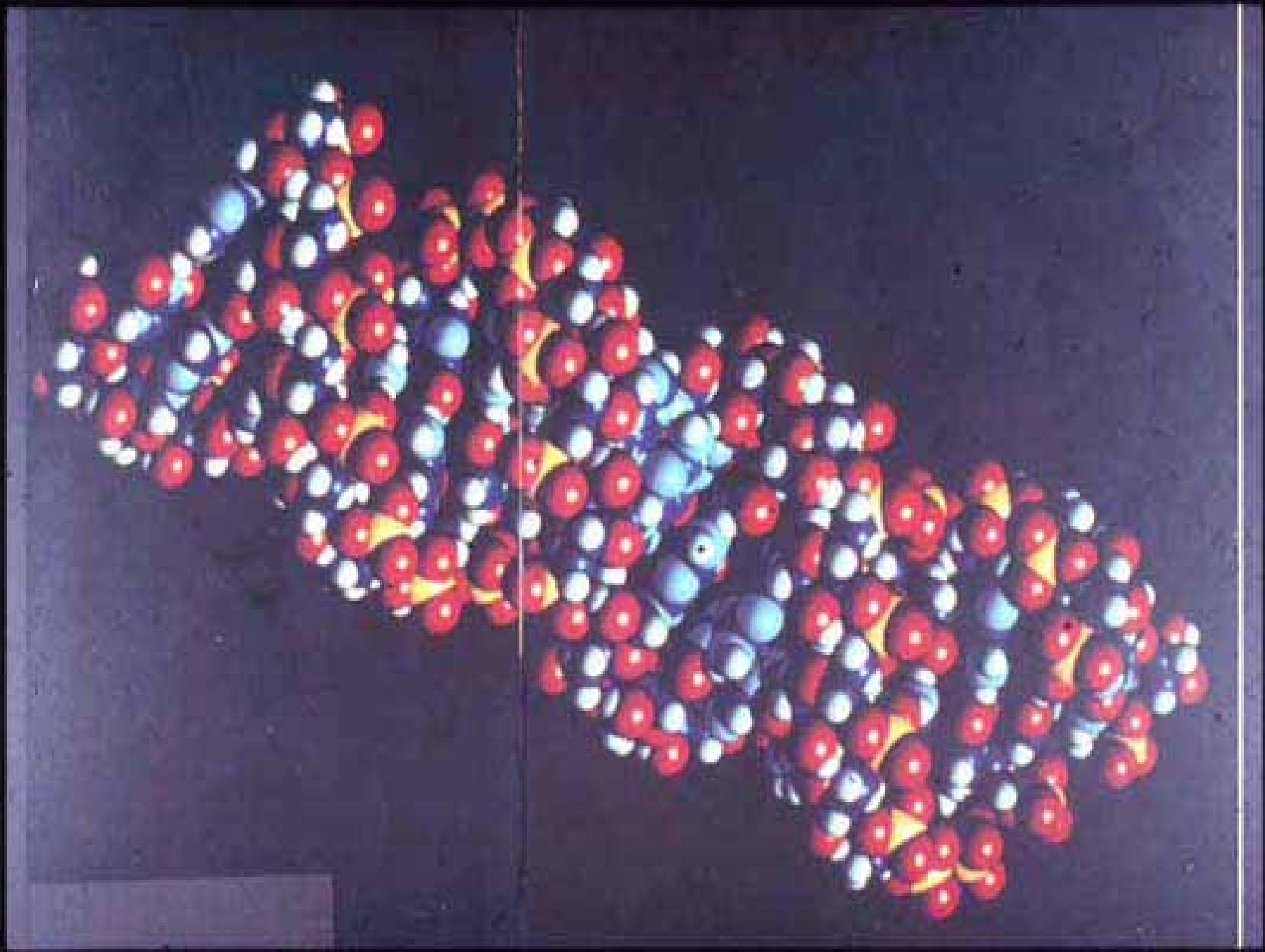
遺伝子は、生物を構成する細胞の核にある染色体という棒状の構造体に、DNAという化学物質としてきざまれています。

DNAには、アデニン、グアニン、シトシン、チミンの4種類の塩基があり、その配列が生体機能を担うタンパク質を決定したり遺伝子の活動を制御しています。

すなわち、DNAは生命の設計図と考えられます。

ヒトの場合、DNAに蓄えられている情報量は、24種類ある染色体1セットあたり塩基対にして約30億あるといわれています。

これらの配列を解析することにより、遺伝子の変異が原因と考えられている4,000以上の疾患の発症機構の解明をはじめ、生命現象全般を解き明かすための有益な知見の獲得が期待されます。



Components of DNA and RNA

- DNA: Adenine (A), Thymine (T), Guanine (G), Cytosine (C)

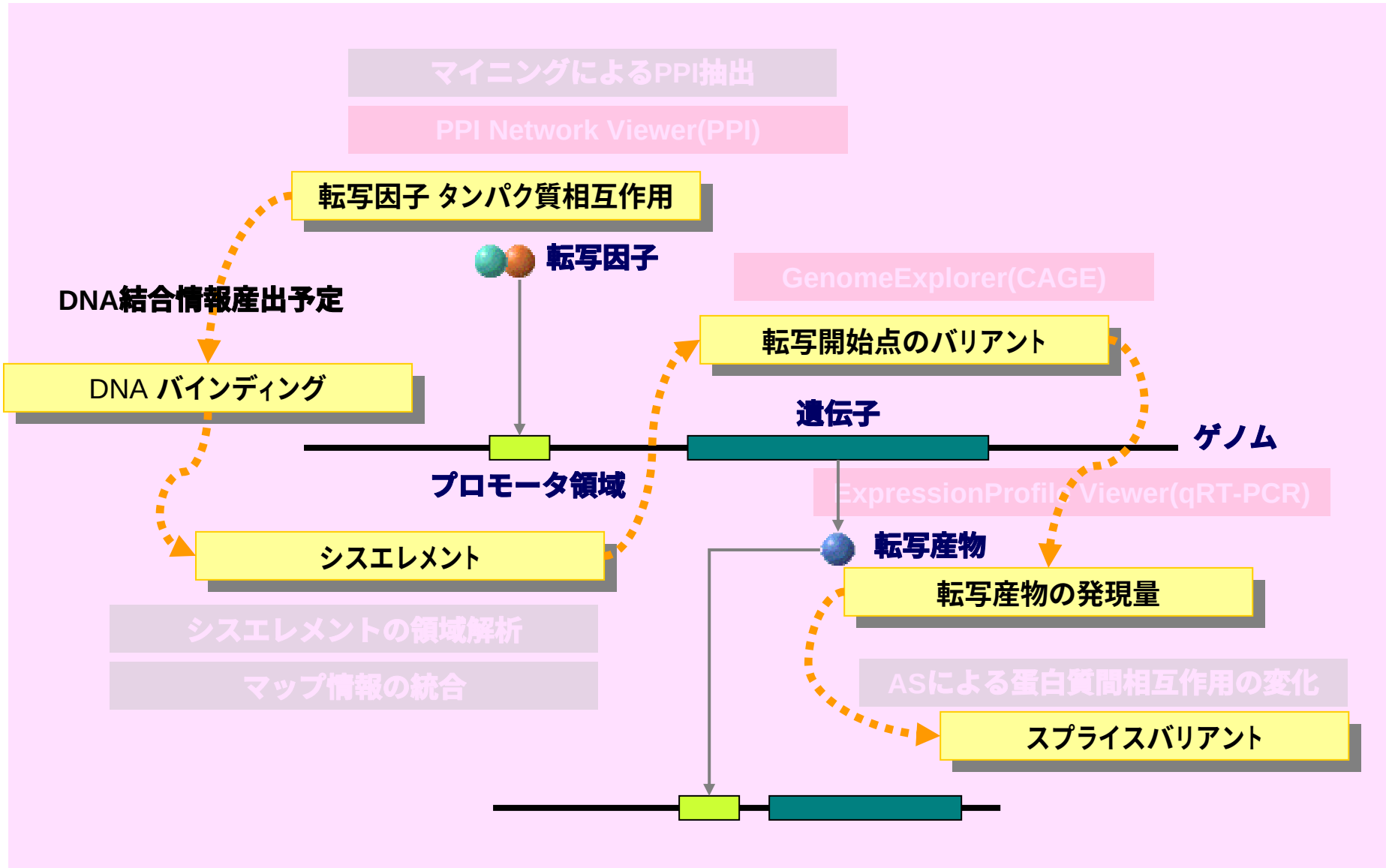
RNA: Adenine (A), Uracil (U), Guanine (G), Cytosine (C)


ORIGIN
1 ttccagcgtt gcgagacggt cggttccaag tgggcctgaa cgcgggggag agccggatct
61 gtcctcggga actgcaaggc cctgtgagcg gcagcactga gatcccgccc gcagctgctg
121 gaagcgtcga agctcagcgg gcccgcgcca tgacctgtgc ttagaactca tcctggcccg
181 cagagcctgc cgcgagtccc tgccgtcccc tgtgccgggc tcttggagcc actttcccca
241 gcggaactca gcccgcggct cggactccgg cgggacctgc tcggacgaat gacgccgccg
301 gattcaagca ctgtcttcct gttggccctg acaatcatag ccagcacctg ggctctgacg
361 cccactcact acctcaccaa gcatgacgtg gagagactaa aagcctcgct ggatcgccct
421 ttcacaaatt tggaatctgc cttctactcc atcgtgggac tcagcagcct tggtgctcag
481 gtgccagatg caaaggaagc atgtacctac atcagatcta accttgatcc cagcaatgtg
541 gattccctct tctacgctgc ccaggccagc caggccctct caggatgtga gatctctatt
601 tcaaatgaga ccaaagatct gcttctgcca gctgtcagtg aggactcatc tgttacccag
661 atctaccatg cagttgcagc tctaagtggc tttggccttc ccttggcatc ccaagaagca
721 ctcagtgccc ttactgctcg tctcagcaag gaggaggctg tgctggcaac agtccaggct
781 ctgcaggcag catcccacct gtcccagcag gctgacctga ggagcatcgt ggaggagatt
841 gaggaccttg ttgctcgcct ggatgaactc gggggcgtgt atctccagtt tgaagaagga
901 ctggaaacaa cagcgttatt tgtggctgcc acctacaagc tcatggatca tgtgggact
961 gagccatcca ttaaggagga tcaggtcatc cagctgatga acgcgatctt cagcaagaag
1021 aactttgagt ccctctccga agccttcagc gtggcctctg cagctgctgt gctctgcat
1081 aatcgctacc acgtgccagt tgtggttgtg cctgaggact ctgcttccga cactcatgaa
1141 caggctatct tgcggttgca agtcaccaat gttctgtctc agcctctgac tcaggccact
1201 gttaaactag aacatgctaa atctgttgct tccagagcca ctgtcctcca gaagacatcc

Genome. DNA, Genes

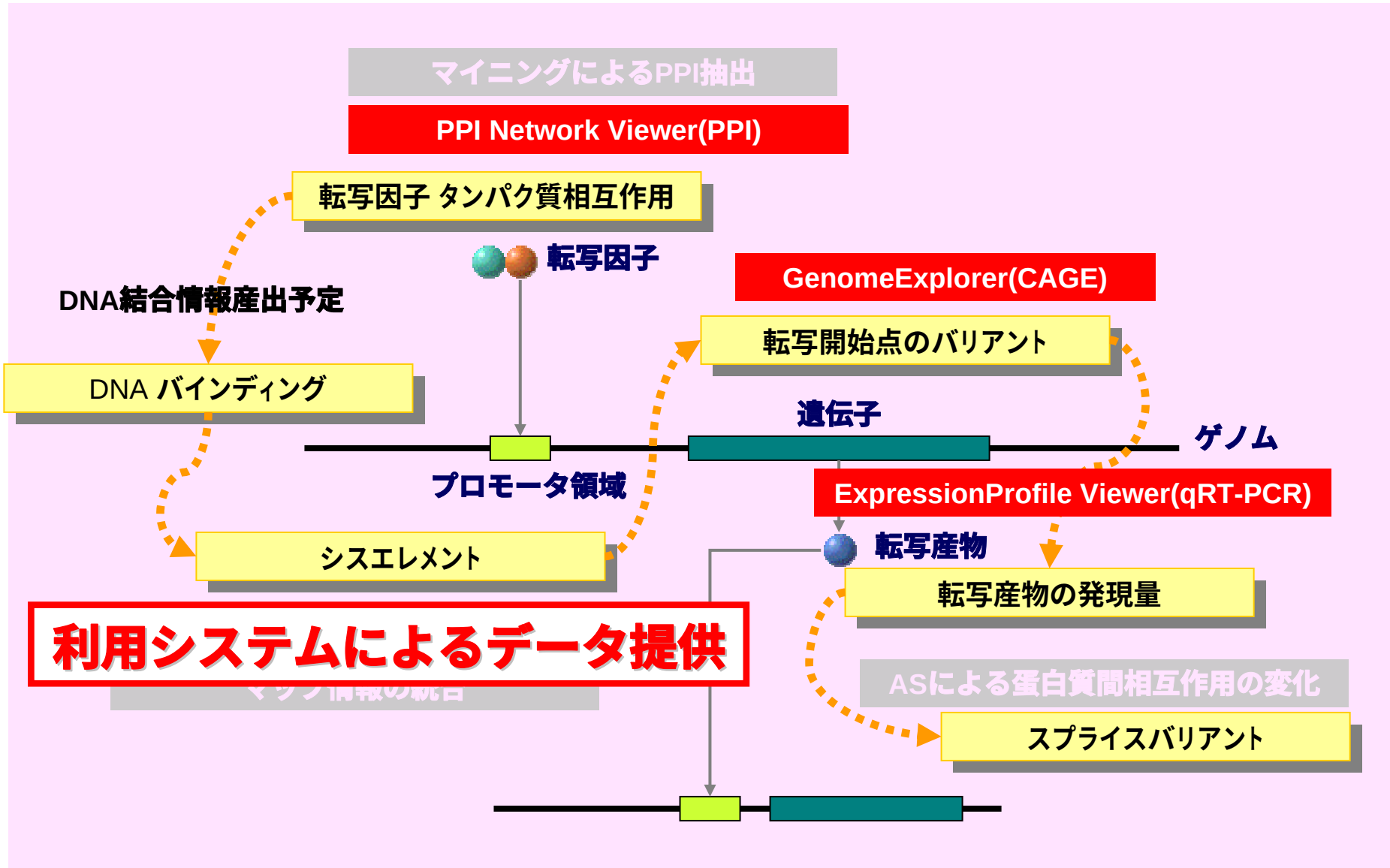
トランスクリプトームから見た遺伝子構造

転写制御ネットワークの解明に向けてゲノムネットワークプラットフォームでは様々なアプローチから研究を進めています。

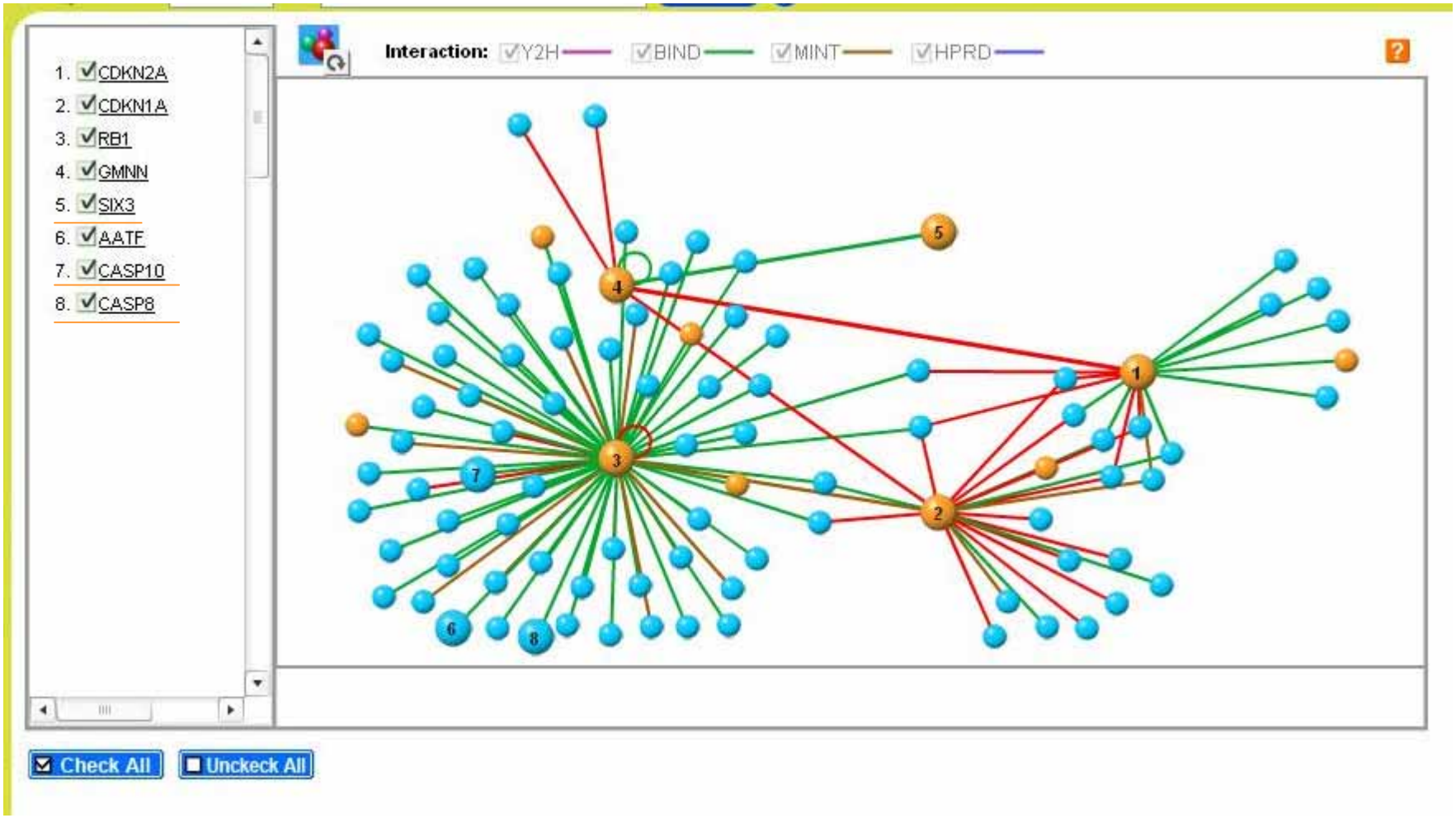


トランスクリプトームから見た遺伝子構造

転写制御ネットワークの解明に必要なコンポーネントを表現する「利用システム」の開発を進めています



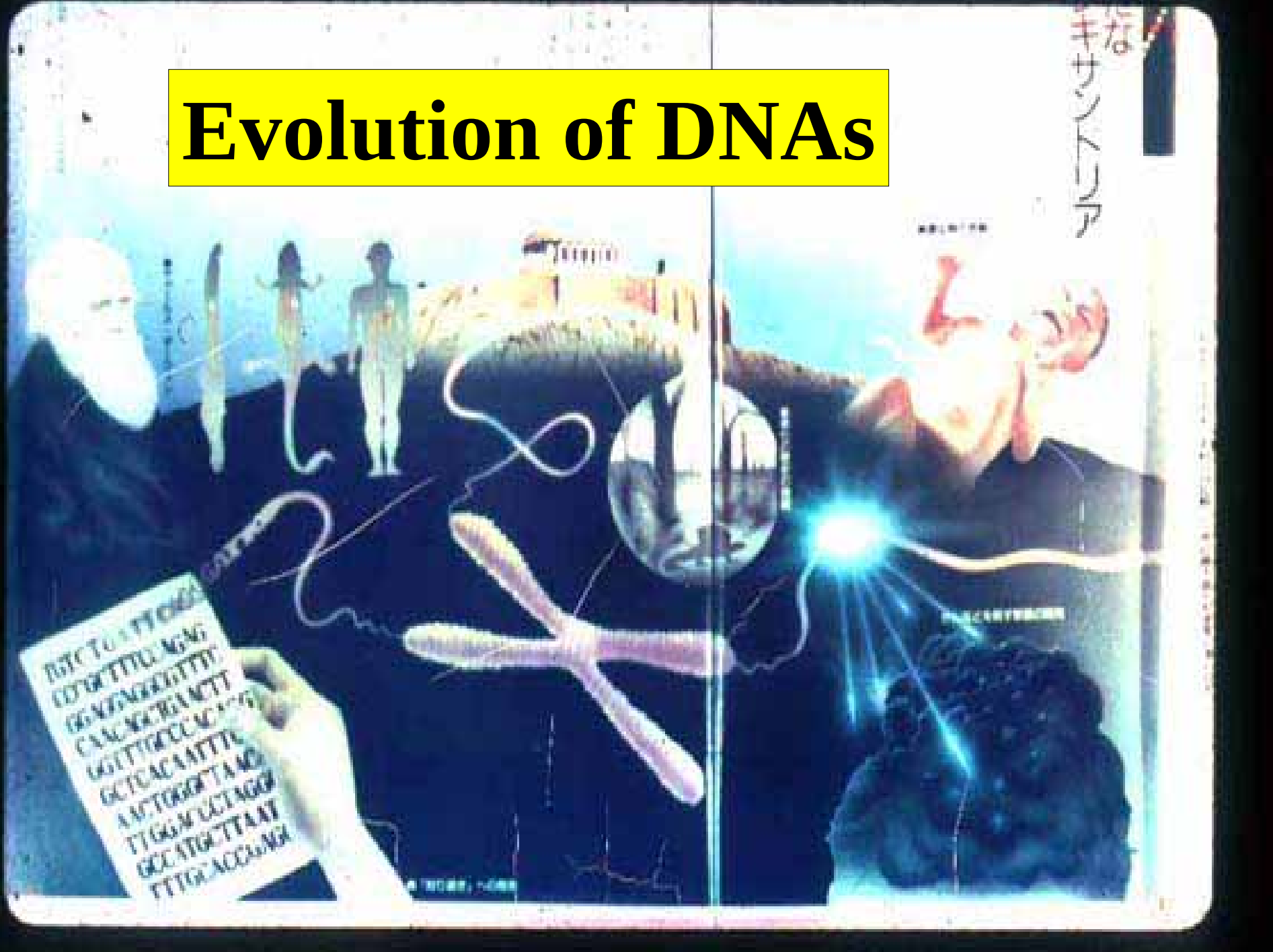
PPIデータを統合的に用いたネットワーク抽出例



Six3,Geminiを介したアポトーシスおよび細胞増殖の制御に関わるネットワーク候補例（caspase他）

A、T、C、Gのたった4つの塩基でヒトのすべてがきまる。

Evolution of DNAs



ゲノム情報・関連情報の爆発的増加

2003年4月14日
ヒトゲノム全塩基配列の解読完了の発表

1) Nature (2001) 409:860-921

2) Nature (2001) 409:685-690

articles

Functional annotation of a full-length mouse cDNA collection

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium*

* A full list of authors appears at the end of the paper

The RIKEN Mouse Gene Encyclopaedia Project, a systematic approach to determining the full coding potential of the mouse genome, involves collection and sequencing of full-length complementary DNAs and physical mapping of the corresponding genes to the mouse genome. We organized an international functional annotation meeting (FANTOM) to annotate the first 21,076 cDNAs to be analysed in this project. Here we describe the first RIKEN clone collection, which is one of the largest described for any organism. Analysis of these cDNAs extends known gene families and identifies new ones.

RIKEN Genome Exploration Research Group Phase II team: J. Kawai[1,2], A. Shinagawa[1], K. Shibata[1,2], M. Yoshino[1], M. Itoh[1,2], Y. Ishii[1], T. Arakawa[1], A. Hara[1], Y. Fukunishi[1,2], H. Konno[1,2], J. Adachi[1], S. Fukuda[1,2], K. Aizawa[1,2], M. Izawa[1], K. Nishi[1], H. Kiyosawa[1], S. Kondo[1], I. Yamanaka[1] & T. Saito[1]

FANTOM Consortium: Y. Okazaki[1], T. Gojobori[1], H. Bono[1], T. Kawaguchi[1], R. Saito[1], K. Kadota[1], H. Matsuda[1], M. Ashburner[2], S. Batalov[3], T. Casavant[4], W. Fleischmann[5], T. Gaasterland[6], C. Gissi[7], B. King[8], H. Kochiwa[11], P. Kublbeck[12], S. Lewis[13], Y. Matsuo[14], I. Nikaido[15], E. Pesole[16], J. Quackenbush[17], L. M. Schriml[18], F. Staubli[19], R. Suzuki[11], M. Tomita[11], L. Wagner[20], T. Washio[15], K. Sakai[21], T. Okido[1], M. Furuno[1], H. Aono[1], R. Baldarelli[22], G. Barsh[23], J. Blake[24], D. Boffelli[25], N. Bojunga[26], P. Carninci[1], M. F. de Bonaldo[27], M. J. Brownstein[27], C. Bult[27], C. Fletcher[27], M. Fujita[28], M. Gariboldi[29], S. Gustincich[30], D. Hill[31], M. Hofmann[32], D. A. Hume[33], M. Kamiya[1,2], N. H. Lee[17], P. Lyons[33], L. Marchionni[27], J. Mashima[1], J. Mazzarelli[27], P. Mombaerts[34], P. Nordone[27], B. Ring[27], M. Ringwald[35], I. Rodriguez[27], H. Sakamoto[36], H. Sasaki[37], K. Sato[1,38], C. Schönbach[39], T. Seya[21], Y. Shibata[1], K.-F. Storch[20], H. Suzuki[1], K. Toyo-oka[20], K. H. Wang[40], C. Weitz[20], C. Whittaker[41], L. Wilming[42], A. Wynshaw-Boris[43], K. Yoshida[1], Y. Hasegawa[1], H. Kawaji[1,44] & S. Kohtsuki[1]

General organizer: Y. Hayashizaki[1,2,44]

1. Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan; 2. GENE SIL 3-1-1 Koyadai, Tsukuba, Ibaraki 305-8074 Japan; 3. Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan; 4. NTT Software Corporation, 223-1 Yamashita-cho, Naka-ku, Yokohama, Kanagawa, 231-0034, Japan; 5. Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan; 6. European Molecular Biology Laboratory European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK; 7. Genomics Institute of the Novartis Research Foundation, 3115 Merryfield Row, San Diego, California 92121, USA; 8. The Cumberland Laboratory for Computational Genomics, University of Iowa Iowa City, Iowa 52242, USA;

2

Nature 409:860-921 (2001)

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

DNA sequence databases

GenBank, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bldg. 38A, 8600 Rockville Pike, Bethesda, Maryland 20894, USA

EMBL, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

DNA Data Bank of Japan, Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima-shi, Shizuoka-ken 411-8540, Japan

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

The RIKEN Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team*

*A full list of authors appears at the end of this paper

Only a small proportion of the mouse genome is transcribed into mature messenger RNA transcripts. There is an international collaborative effort to identify all full-length mRNA transcripts from the mouse, and to ensure that each is represented in a physical collection of clones. Here we report the manual annotation of 60,770 full-length mouse complementary DNA sequences. These are clustered into 33,409 'transcriptional units', contributing 90.1% of a newly established mouse transcriptome database. Of these transcriptional units, 4,258 are new protein-coding and 11,665 are new non-coding messages, indicating that non-coding RNA is a major component of the transcriptome. 41% of all transcriptional units showed evidence of alternative splicing. In protein-coding transcripts, 79% of splice variations altered the protein product. Whole-transcriptome analyses resulted in the identification of 2,431 sense–antisense pairs. The present work, completely supported by physical clones, provides the most comprehensive survey of a mammalian transcriptome so far, and is a valuable resource for functional genomics.

With the availability of draft sequences of the human genome[1,2], increasing attention has focused on identifying the complete set of mammalian genes, both protein-coding and non-protein-coding. As various analyses[3,4] have noted, different annotation criteria lead to different sets of predicted genes, and even the true number of protein-coding genes remains uncertain. Biases in the training sets used for optimizing gene-finding algorithms lead to systematic biases in the types of genes that are predicted computationally—and so important genes and gene classes can be missed. A systematic analysis of transcripts from human chromosome 21 using dense oligonucleotide arrays revealed that the apparent transcriptional output of processed cytoplasmic messenger RNAs exceeds the predicted exons by an order of magnitude[5].

30,694 new cDNA clones (FANTOM2 new set) and the 21,076 FANTOM1 clone set. We also present a global analysis of the mouse transcriptome based upon an integration of the FANTOM2 sequences with all available mouse mRNA data from the public sequence databases[10]. This data set provides the most comprehensive view so far of the transcriptional potential and diversity within the mouse genome, and serves as an important model for analysing the transcriptomes of other higher eukaryotes.

To enable a global description of the transcriptome, we use the term transcriptional unit (TU) to describe a segment of the genome from which transcripts are generated. A TU is defined by the identification of a cluster of transcripts that contains a common core of genetic information (in some cases, a protein-coding

sequence novel full-length mouse cDNAs. The initial results and validation of our approach have been reported previously and resulted in the functional annotation of 21,076 full-length cDNAs (FANTOM1)[11]. Here we describe the characterization and annotation of the FANTOM2 clone set (60,770), consisting of

and subtracted, with most from C57BL/6J mice, as described elsewhere[11,14,15]. Information about tissue source and other information regarding these libraries is available in Supplementary Information section 1.

1,442,236 sequences were grouped into 171,144 5′-end clusters

3）マウス全長ｃＤＮＡアノテーション

Nature (2002) 420: 563-573

理研・遺伝研ＤＤＢＪ　他　国際

Science. 2005 Sep 2;309(5740):1559-63.

The transcriptional landscape of the mammalian genome.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schonbach C, Sekiguchi K, Semple CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusic V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group).

The genome sequence and structure of rice chromosome 1

Takuji Sasaki*, Takashi Matsumoto*, Kimiko Yamamoto*, Katsumi Sakata*, Tomoya Baba*, Yuichi Katayose*, Jianzhong Wu*, Yoshihito Niimura†, Zhukuan Cheng‡, Yoshiaki Nagamura*, Baltazar A. Antonio*, Hiroyuki Kanamori*, Satomi Hosokawa*, Masatoshi Masukawa*, Koji Arikawa*, Yoshino Chiden*, Mika Hayashi*, Masako Okamoto*, Tsuyu Ando*, Hiroyoshi Aoki*, Kohei Arita*, Masao Hamada*, Chizuko Harada*, Saori Hijishita*, Mikiko Honda*, Yoko Ichikawa*, Atsuko Idonuma*, Masumi Iijima*, Michiko Ikeda*, Maiko Ikeno*, Sachie Ito*, Tomoko Ito*, Yuichi Ito*, Yukiyo Ito*, Aki Iwabuchi*, Kozue Kamiya*, Wataru Karasawa*, Satoshi Katagiri*, Ari Kikuta*, Noriko Kobayashi*, Izumi Kono*, Kayo Machita*, Tomoko Maehara*, Hiroshi Mizuno*, Tatsumi Mizubayashi*, Yoshiyuki Mukai*, Hideki Nagasaki*, Marina Nakashima*, Yuko Nakama*, Yumi Nakamichi*, Mari Nakamura*, Nobukazu Namiki*, Manami Negishi*, Isamu Ohta*, Nozomi Ono*, Shoko Saji*, Kumiko Sakai*, Michie Shibata*, Takanori Shimokawa*, Ayahiko Shomura*, Jianyu Song*, Yuka Takazaki*, Kimihiro Terasawa*, Kumiko Tsuji*, Kazunori Waki*, Harumi Yamagata*, Hiroko Yamane*, Shoji Yoshiki*, Rie Yoshihara*, Kazuko Yukawa*, Huisun Zhong*, Hisakazu Iwama†, Toshinori Endo§, Hidetaka Ito§, Jang Ho Hahn||, Ho-Il Kim||, Moo-Young Eun||, Masahiro Yano*, Jiming Jiang‡ & Takashi Gojobori†

4）イネゲノム１染色体配列決

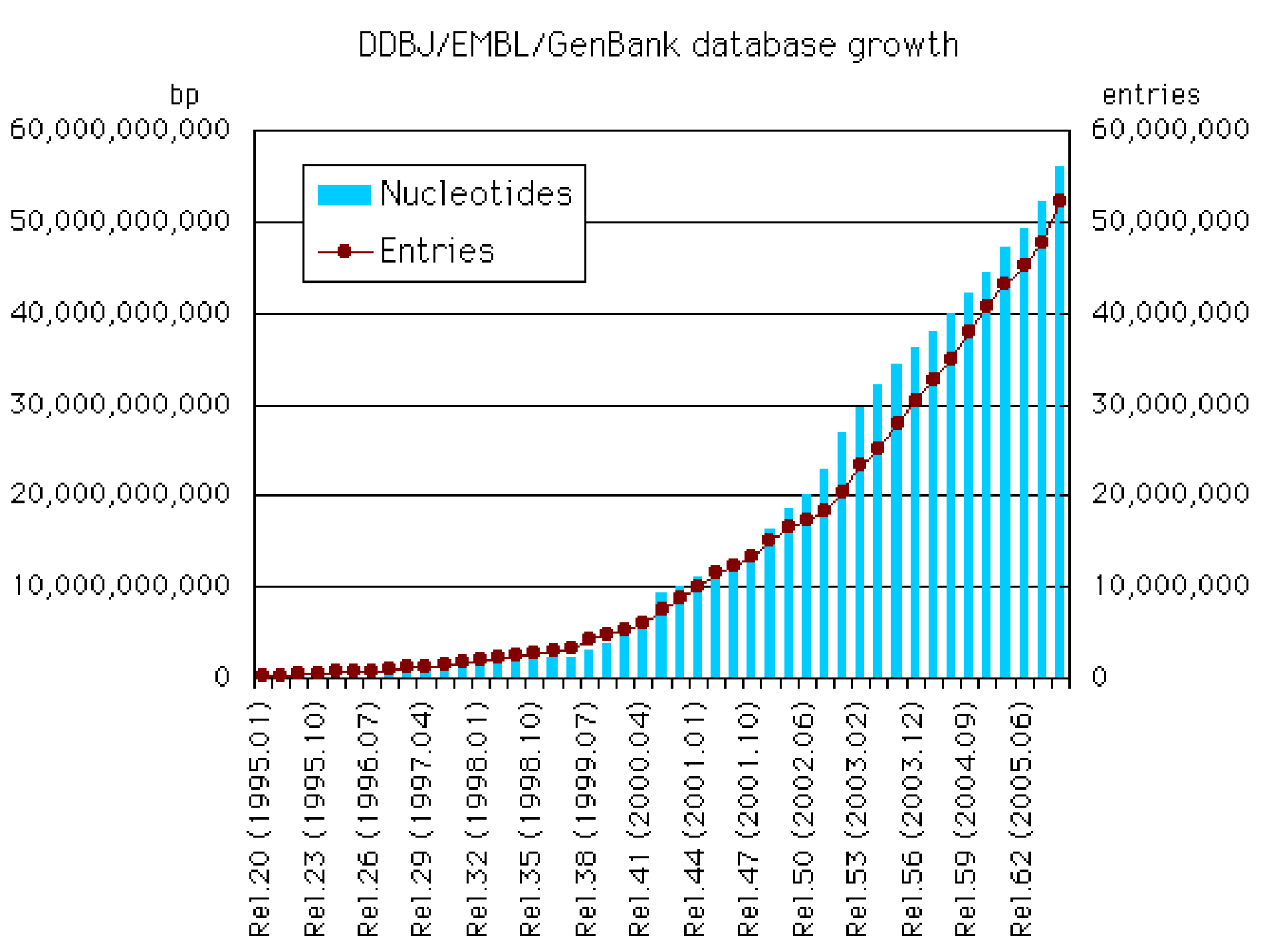
Nature (2002) 420:312-316

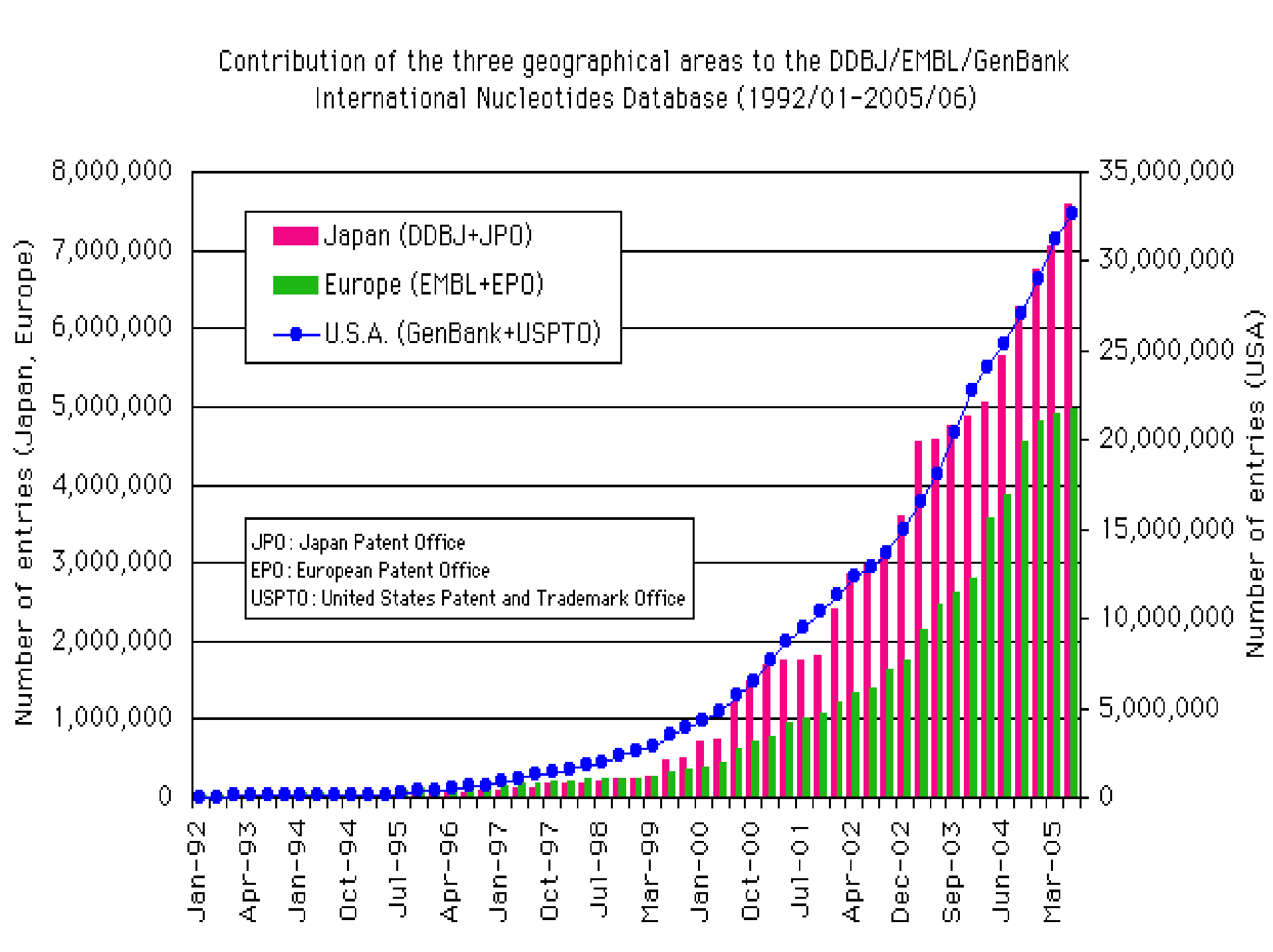
生資研・遺伝研ＤＤＢＪ

The map-based sequence of the rice genomeInternational Rice Genome Sequencing Project*

Nature (2005) 436: August 11

International Rice Genome Sequencing Project (Participants are arranged by area of contribution and then by institution.)
Physical Maps and Sequencing: Rice Genome Research Program (RGP) Takashi Matsumoto1, Jianzhong Wu1, Hiroyuki Kanamori1, Yuichi Katayose1, Masaki Fujisawa1, Nobukazu Namiki1, HiroshiMizuno1, Kimiko Yamamoto1, Baltazar A. Antonio1, Tomoya Baba1, Katsumi Sakata1, Yoshiaki Nagamura1, Hiroyoshi Aoki1, Koji Arikawa1, Kohei Arita1, Takahito Bito1, Yoshino Chiden1, Nahoko Fujitsuka1, Rie Fukunaka1, Masao Hamada1, Chizuko Harada1, Akiko Hayashi1, Saori Hijishita1, Mikiko Honda1, Satomi Hosokawa1, Yoko Ichikawa1, Atsuko Idonuma1, Masumi Iijima1, Michiko Ikeda1, Maiko Ikeno1, Kazue Ito1, Sachie Ito1, Tomoko Ito1, Yuichi Ito1, Yukiyo Ito1, Aki Iwabuchi1, Kozue Kamiya1, Wataru Karasawa1, Kanako Kurita1, Satoshi Katagiri1, Ari Kikuta1, Harumi Kobayashi1, Noriko Kobayashi1, Kayo Machita1, Tomoko Maehara1, Masatoshi Masukawa1, Tatsumi Mizubayashi1, Yoshiyuki Mukai1, Hideki Nagasaki1, Yuko Nagata1, Shinji Naito1, Marina Nakashima1, Yuko Nakama1, Yumi Nakamichi1, Mari Nakamura1, Ayano Meguro1, Manami Negishi1, Isamu Ohta1, Tomoya Ohta1, Masako Okamoto1, Nozomi Ono1, Shoko Saji1, Miyuki Sakaguchi1, Kumiko Sakai1, Michie Shibata1, Takanori Shimokawa1, Jianyu Song1, Yuka Takazaki1, Kimihiro Terasawa1, Mika Tsugane1, Kumiko Tsuji1, Shigenori Ueda1, Kazunori Waki1, Harumi Yamagata1, Mayu Yamamoto1, Shinichi Yamamoto1, Hiroko Yamane1, Shoji Yoshiki1, Rie Yoshihara1, Kazuko Yukawa1, Huisun Zhong1, Masahiro Yano1, Takuji Sasaki (Principal Investigator)1 ; The Institute for Genomic Research (TIGR) Qiaoping Yuan2, Shu Ouyang2, Jia Liu2, Kristine M. Jones2, Kristen Gansberger2, Kelly Moffat2, Jessica Hill2, Jayati Bera2, Douglas Fadrosh2, Shaohua Jin2, Shivani Johri2, Mary Kim2, Larry Overton2, Matthew Reardon2, Tamara Tsitrin2, Hue Vuong2, Bruce Weaver2, Anne Ciecko2, Luke Tallon2, Jacqueline Jackson2, Grace Pai2, Susan Van Aken2, Terry Utterback2, Steve Reidmuller2, Tamara Feldblyum2, Joseph Hsiao2, Victoria Zismann2, Stacey Iobst2, Aymeric R. de Vazeille2, C. Robin Buell (Principal Investigator)2; National Center for Gene Research Chinese Academy of Sciences (NCGR) Kai Ying3, Ying Li3, Tingting Lu3, Yuchen Huang3, Qiang Zhao3, Qi Feng3, Lei Zhang3, Jingjie Zhu3, Qijun Weng3, Jie Mu3, Yiqi Lu3, Danlin Fan3, Yilei Liu3, Jianping Guan3, Yujun Zhang3, Shuliang Yu3, Xiaohui Liu3, Yu Zhang3, Guofan Hong3, Bin Han (Principal Investigator)3; Genoscope Nathalie Choisne4, Nadia Demange4, Gisela Orjeda4, Sylvie Samain4, Laurence Cattolico4, Eric Pelletier4, Arnaud Couloux4, Beatrice Segurens4, Patrick Wincker4, Angelique D'Hont5, Claude Scarpelli4, Jean Weissenbach4, Marcel Salanoubat4, Francis Quetier (Principal Investigator)4; Arizona Genomics Institute (AGI) and Arizona Genomics Computational Laboratory (AGCol) Yeisoo Yu6, Hye Ran Kim6, Teri Rambo6, Jennifer Currie6, Kristi Collura6, Meizhong Luo6, Tae-Jin Yang6, Jetty S. S. Ammiraju6, Friedrich Engler6, Carol Soderlund6, Rod A. Wing (Principal Investigator)6; Cold Spring Harbor Laboratory (CSHL) Lance E. Palmer7, Melissa de la Bastide7, Lori Spiegel7, Lidia Nascimento7, Theresa Zutavern7, Andrew O'Shaughnessy7, Sujit Dike7, Neilay Dedhia7, Raymond Preston7, Vivekanand Balija7,W. Richard McCombie (Principal Investigator)7; Academia Sinica Plant Genome Center (ASPGC) Teh-Yuan Chow8, Hong-Hwa Chen9, Mei-Chu Chung8, Ching-San Chen8, Jei-Fu Shaw8, Hong-Pang Wu8, Kwang-Jen Hsiao10, Ya-Ting Chao8, Mu-kuei Chu8, Chia-Hsiung Cheng8, Ai-Ling Hour8, Pei-Fang Lee8, Shu-Jen Lin8, Yao-Cheng Lin8, John-Yu Liou8, Shu-Mei Liu8, Yue-Ie Hsing (Principal Investigator)8; Indian Initiative for Rice Genome Sequencing (IIRGS), University of Delhi South Campus (UDSC) S. Raghuvanshi11, A. Mohanty11, A. K. Bharti11,13, A. Gaur11, V. Gupta11, D. Kumar11, V. Ravi11, S. Vij11, A. Kapur11, Parul Khurana11, Paramjit Khurana11, J. P. Khurana11, A. K. Tyagi (Principal Investigator)11; Indian Initiative for Rice Genome Sequencing (IIRGS), Indian Agricultural Research Institute (IARI) K. Gaikwad12, A. Singh12, V. Dalal12, S. Srivastava12, A. Dixit12, A. K. Pal12, I. A. Ghazi12, M. Yadav12, A. Pandit12, A. Bhargava12, K. Sureshbabu12, K. Batra12, T. R. Sharma12, T. Mohapatra12, N. K. Singh (Principal Investigator)12; Plant Genome Initiative at Rutgers (PGIR) Joachim Messing (Principal Investigator)13, Amy Bronzino Nelson13, Galina Fuks13, Steve Kavchok13, Gladys Keizer13, Eric Linton Victor Llaca13, Rentao Song13, Bahattin Tanyolac13, Steve Young13; Korea Rice Genome Research Program (KRGRP) Kim Ho-Il14, Jang Ho Hahn (Principal Investigator)14; National Center for Genetic Engineering and Biotechnology (BIOTEC) G. Sangsakoo15, A. Vanavichit (Principal Investigator)15; Brazilian Rice Genome Initiative (BRIGI) Luiz Anderson Teixeira de Mattos16, Paulo Dejalma Zimmer16, Gaspar Malone16, Odir Dellagostin16, Antonio Costa de Oliveira (Principal Investigator)16; John Innes Centre (JIC) Michael Bevan17, Ian Bancroft17; Washington University School of Medicine Genome Sequencing Center Pat Minx18, Holly Cordum18, Richard Wilson18; University of Wisconsin–Madison Zhukuan Cheng19, Weiwei Jin19, Jiming Jiang19, Sally Ann Leong20
Annotation and Analysis: Hisakazu Iwama21, Takashi Gojobori21,22, Takeshi Itoh22,23, Yoshihito Niimura24, Yasuyuki Fujii25, Takuya Habara25, Hiroaki Sakai23,25, Yoshiharu Sato22, Greg Wilson26, Kiran Kumar27, Susan McCouch26, Nikoleta Juretic28, Douglas Hoen28, Stephen Wright29, Richard Bruskiewich30, Thomas Bureau28, Akio Miyao23, Hirohiko Hirochika23, Tomotaro Nishikawa23, Koh-ichi Kadowaki23 & Masahiro Sugiura31


DDBJ/EMBL/GenBank database growth
bp
entries
60,000,000,000
50,000,000,000
40,000,000,000
30,000,000,000
20,000,000,000
10,000,000,000
0
60,000,000
50,000,000
40,000,000
30,000,000
20,000,000
10,000,000
0
Nucleotides
Entries
Rel.20 (1995.01)
Rel.23 (1995.10)
Rel.26 (1996.07)
Rel.29 (1997.04)
Rel.32 (1998.01)
Rel.35 (1998.10)
Rel.38 (1999.07)
Rel.41 (2000.04)
Rel.44 (2001.01)
Rel.47 (2001.10)
Rel.50 (2002.06)
Rel.53 (2003.02)
Rel.56 (2003.12)
Rel.59 (2004.09)
Rel.62 (2005.06)


Contribution of the three geographical areas to the DDBJ/EMBL/GenBank International Nucleotides Database (1992/01-2005/06)
Number of entries (Japan, Europe)
Number of entries (USA)
8,000,000
7,000,000
6,000,000
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
0
35,000,000
30,000,000
25,000,000
20,000,000
15,000,000
10,000,000
5,000,000
0
Japan (DDBJ+JPO)
Europe (EMBL+EPO)
U.S.A. (GenBank+USPTO)
JPO : Japan Patent Office
EPO : European Patent Office
USPTO : United States Patent and Trademark Office
Jan-92
Apr-93
Jan-94
Oct-94
Jul-95
Apr-96
Jan-97
Oct-97
Jul-98
Mar-99
Jan-00
Oct-00
Jul-01
Apr-02
Dec-02
Sep-03
Jun-04
Mar-05

Biological hierarchy

Evolution

Ecosystem (Environments)

Population (Human population)

Individual (Human)

Organ (Lung, Stomach)

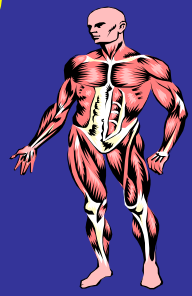
Tissue (Epidermal tissue)

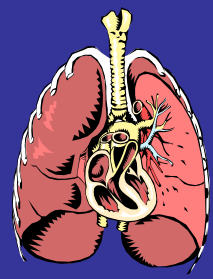
Cell (Red blood cell)

Organella (Mitochondria)

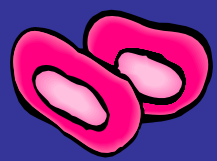
Bio-molecules (DNA, RNA, Proteins)

Molecules (H_2O, O_2)

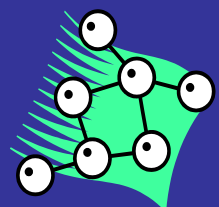




Integration



Red Blood Cell



Post-genome epoc: Future perspectives

23,149 genes

~40,000 genes

Organism → Health

Organ

Tissue

Cell

Molecule

- Phisiome
- Phenome

Interactome

Metabolome

Localizome

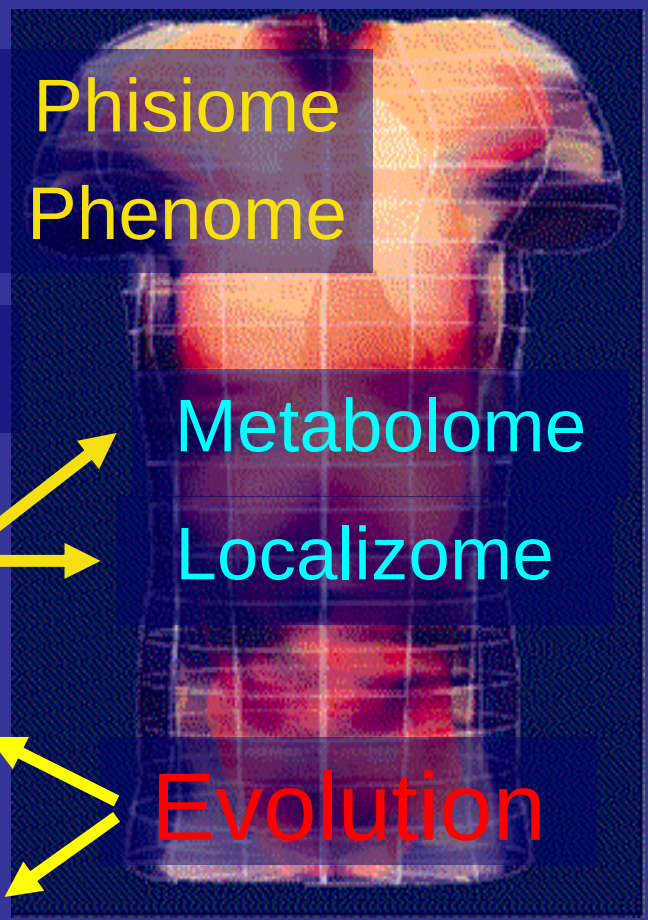
Ribonome

Proteome

Genes → Transcriptome

Evolution

Genome → Regulome



アノテーション（情報付加）
と
データベースの構築

“Human Full-Length cDNA Annotation Invitational” (H-Invitational)

August 25 - September 5, 2002

- Systematic Identification of Human Genes and its Biological Significance -

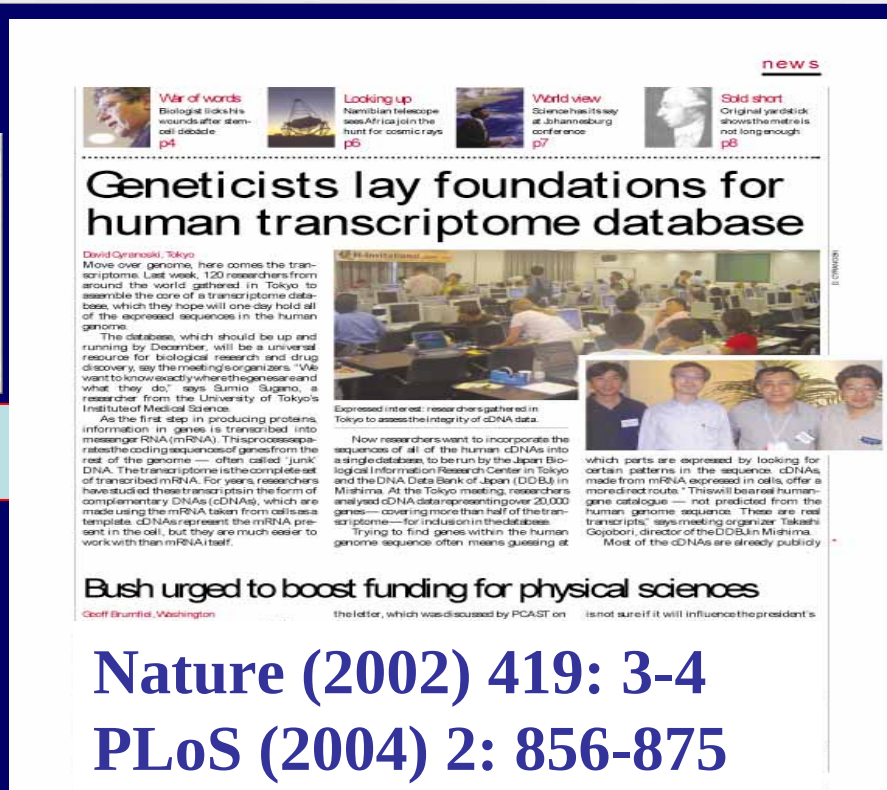
Co-organized by BIRC/AIST and DDBJ/NIG

Attended by more than 118 people from 40 organizations such as

BIRC, DDBJ, NCBI, EBI, Sanger Centre,NCI-MGC, DOE, NIH, DKFZ, CNHGC(Shanghai), RIKEN, Tokyo U, MIPS, CNRS, MCW, TIGR, CBRC, Murdoch U, U Iowa, Karolinska Int., WashU, U Cincinnati, Tokyo MD U, KRIBB, South African Bioinfor Inst, U College London, Reverse Proteomics Res. Inst., Kazusa DNA Inst, Weizmann Inst, Royal Inst. Tech. Sweden, Penn State U, Osaka U, Keio U, Kyushu U, TIT, Ludwig Inst. Brazil, Kyoto U, German Can.Inst., and NIG

Supported by

JBIC, METI, MEXT, NIH, and DOE


news

War of words
Biologist licks his wounds after stem-cell debacle
p4

Looking up
Namibian telescope sees Africa join the hunt for cosmic rays
p6

World view
Science has its say at Johannesburg conference
p7

Said short
Original yardstick shows the metre is not long enough
p8

Geneticists lay foundations for human transcriptome database

David Cyranoski, Tokyo

Move over genome, here comes the transcriptome. Last week, 120 researchers from around the world gathered in Tokyo to assemble the core of a transcriptome database, which they hope will one day hold all of the expressed sequences in the human genome.

The database, which should be up and running by December, will be a universal resource for biological research and drug discovery, say the meeting's organizers. "We want to know exactly where the genes are and what they do," says Sumio Sugano, a researcher from the University of Tokyo's Institute of Medical Science.

As the first step in producing proteins, information in genes is transcribed into messenger RNA (mRNA). This process separates the coding sequences of genes from the rest of the genome — often called 'junk' DNA. The transcriptome is the complete set of transcribed mRNA. For years, researchers have studied these transcripts in the form of complementary DNAs (cDNAs), which are made using the mRNA taken from cells as a template. cDNAs represent the mRNA present in the cell, but they are much easier to work with than mRNA itself.

Expressed interest: researchers gathered in Tokyo to assess the integrity of cDNA data.

Now researchers want to incorporate the sequences of all of the human cDNAs into a single database, to be run by the Japan Biological Information Research Center in Tokyo and the DNA Data Bank of Japan (DDBJ) in Mishima. At the Tokyo meeting, researchers analysed cDNA data representing over 20,000 genes — covering more than half of the transcriptome — for inclusion in the database.

Trying to find genes within the human genome sequence often means guessing at which parts are expressed by looking for certain patterns in the sequence. cDNAs, made from mRNA expressed in cells, offer a more direct route. "This will be a real human-gene catalogue — not predicted from the human genome sequence. These are real transcripts," says meeting organizer Takashi Gojobori, director of the DDBJ in Mishima.

Most of the cDNAs are already publicly

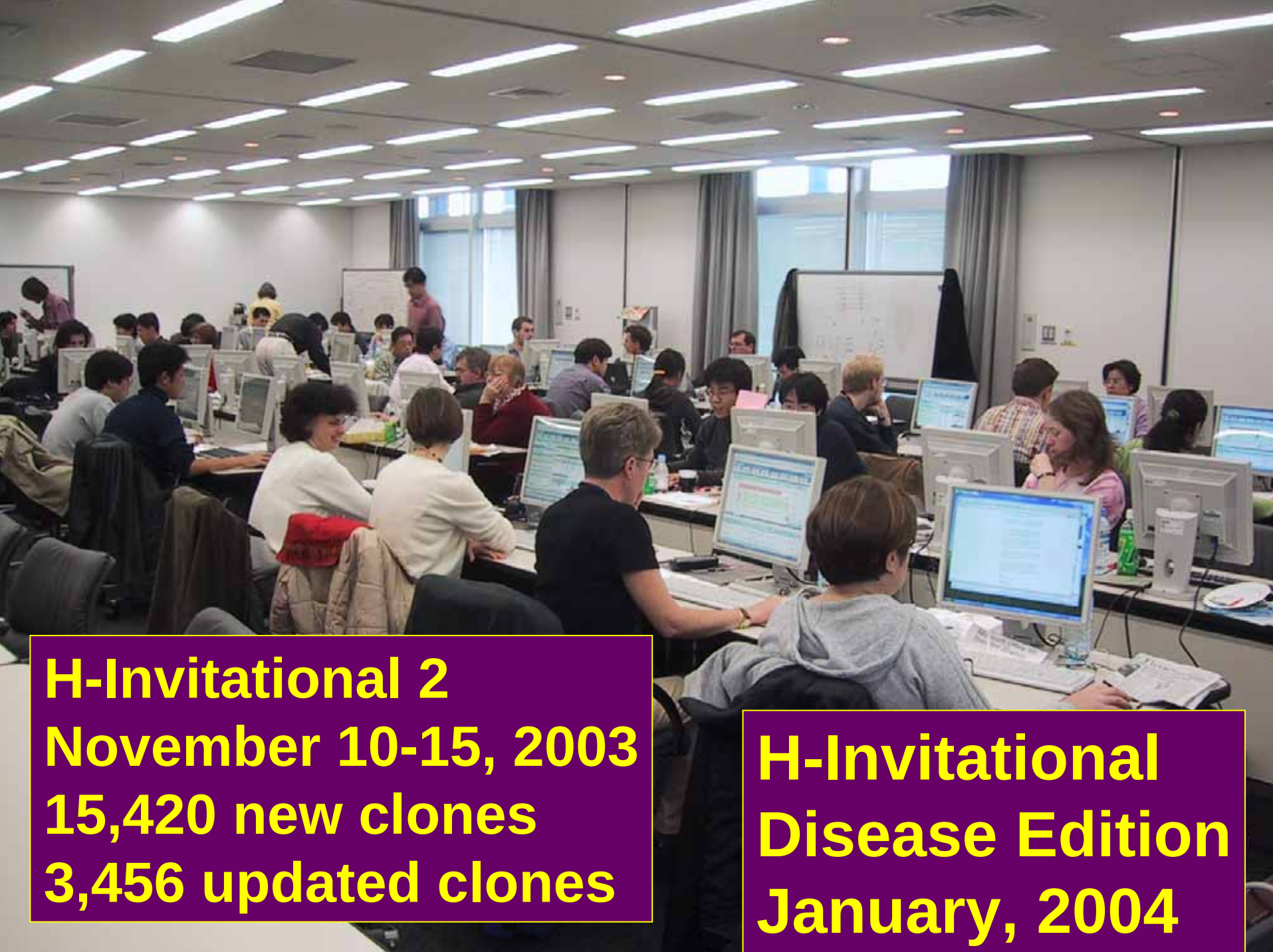
Bush urged to boost funding for physical sciences

Geoff Brumfiel, Washington

the letter, which was discussed by PCAST on

is not sure if it will influence the president's

Nature (2002) 419: 3-4
PLoS (2004) 2: 856-875

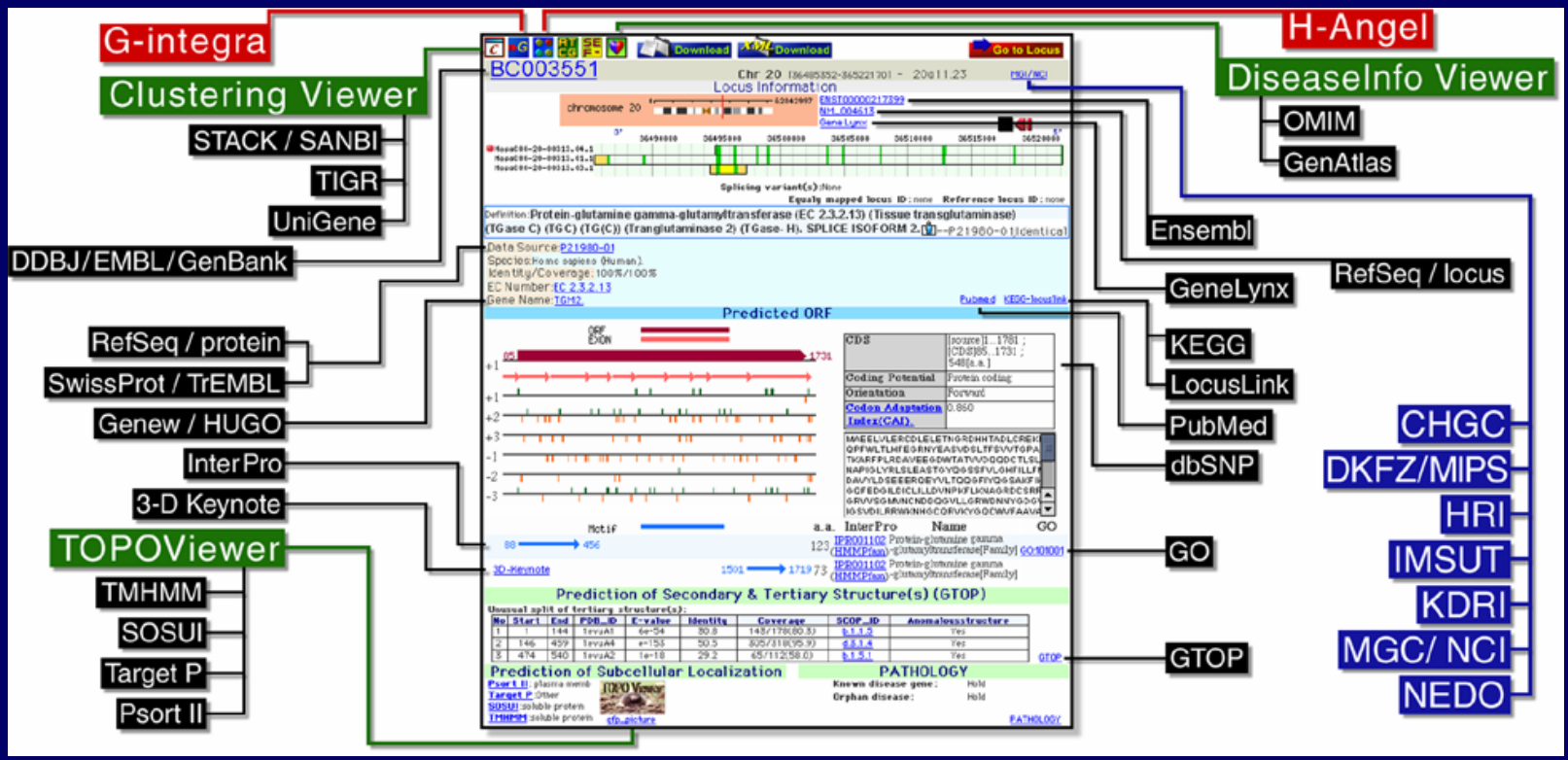


H-Invitational 2
November 10-15, 2003
15,420 new clones
3,456 updated clones

H-Invitational
Disease Edition
January, 2004

Most Interesting Findings in H-Invitational

(10) **このプロジェクトで産生されたすべての情報を格納した**H-InvDB **統合データベースを構築した。これは、**human full-length cDNA annotations **の**a comprehensive database**である。また、**known disease-related genes**と**loci co-localized with 694 orphan pathologies (mapped but not cloned)**のデータベースを構築した。公開用データベースもほぼできて、まもなく公開する。**



War of words
Biologist licks his wounds after stem-cell debacle
p4

Looking up
Namibian telescope sees Africa join the hunt for cosmic rays
p6

World view
Science has its say at Johannesburg conference
p7

Sold short
Original yardstick shows the metre is not long enough
p8

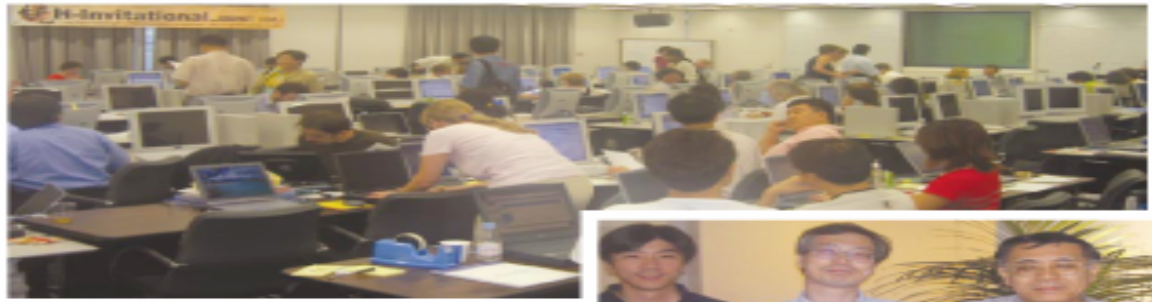
Geneticists lay foundations for human transcriptome database

David Cyranoski, Tokyo

Move over genome, here comes the transcriptome. Last week, 120 researchers from around the world gathered in Tokyo to assemble the core of a transcriptome database, which they hope will one day hold all of the expressed sequences in the human genome.

The database, which should be up and running by December, will be a universal resource for biological research and drug discovery, say the meeting's organizers. "We want to know exactly where the genes are and what they do," says Sumio Sugano, a researcher from the University of Tokyo's Institute of Medical Science.

As the first step in producing proteins, information in genes is transcribed into messenger RNA (mRNA). This process separates the coding sequences of genes from the rest of the genome — often called 'junk' DNA. The transcriptome is the complete set of transcribed mRNA. For years, researchers have studied these transcripts in the form of complementary DNAs (cDNAs), which are made using the mRNA taken from cells as a template. cDNAs represent the mRNA present in the cell, but they are much easier to work with than mRNA itself.



Expressed interest: researchers gathered in Tokyo to assess the integrity of cDNA data.

Now researchers want to incorporate the sequences of all of the human cDNAs into a single database, to be run by the Japan Biological Information Research Center in Tokyo and the DNA Data Bank of Japan (DDBJ) in Mishima. At the Tokyo meeting, researchers analysed cDNA data representing over 20,000 genes — covering more than half of the transcriptome — for inclusion in the database.

Trying to find genes within the human genome sequence often means guessing at which parts are expressed by looking for certain patterns in the sequence. cDNAs, made from mRNA expressed in cells, offer a more direct route. "This will be a real human-gene catalogue — not predicted from the human genome sequence. These are real transcripts," says meeting organizer Takashi Gojobori, director of the DDBJ in Mishima.

Most of the cDNAs are already publicly

Nature (2002) 419: 3-4
PLoS Biol. (2004) 2: 1-21

sical sciences

(PCAST) will call for a funding increase in fields such as physics and engineering, to match the five-year doubling of biomedical research at the National Institutes of Health, which will be completed next year.

"All evidence points to a need to improve funding levels for physical sciences and certain areas of engineering," says a draft of

fresh air," says Michael Lubell, director of public affairs at the American Physical Society in Washington DC, which has advocated an increase for the past few years.

Lubell believes that the proposal will influence Bush and the Office of Management and Budget, which sets the president's annual budget — although he

is not sure if it will influence the president's budget request for 2004, which is being drawn up now for release in February. "The 2004 presidential budget will be extraordinarily tight," he says.

Some scientific societies were less happy that the letter failed to mention other disciplines, such as space science and mathematics, whose funding has languished in recent years. "I'm a little concerned about the way they worded this letter," says Samuel Rankin, director of the American Mathematical Society in Washington DC.

But Lubell says he doesn't think that these fields are being left out. "I think they are implicitly included," he says.

NEWS

This Week

PAGE 373
How songbirds set their compass

375
China mu
gravitatic
wave pro

GENOMICS

New Global Database Lends a Hand to Gene Hunters

TOKYO—Genetics researchers received a new tool today. A database of annotated, full-length human complementary DNAs (cDNAs), compiled by an international team led by Japanese researchers, has been opened for public access. The database is expected to be a boon for research related to drug development, gene hunting, molecular evolution, and comparative genomics.

"This unique database should prove very valuable for understanding the human transcriptome [all the messenger RNA transcribed from genes]," says Sumio Sugano, a molecular biologist at the University of Tokyo's Institute of Medical Science and one of the organizers of the effort. cDNA embodies the protein-coding sequences of genes, as captured from messenger RNA expressed in different tissue. It represents actual protein-encoding genes and doesn't have to be inferred, as is the case when working with the genomic sequence data.

The new database pools and builds on information from six major cDNA projects, including the Mammalian Gene Collection of the U.S. National Institutes of Health, the German Human cDNA project, work at the Chinese National Human Genome Center in Shanghai, and three projects in Japan. Takashi Gojobori, deputy director of the Japan Biological Information Research Center in Tokyo, says that scientists involved in the various projects were looking for ways to make their work more accessible to all researchers and to augment the value of individual collections, for example, which focus on cancer-related cDNAs or cDNAs of genes with unknown functions. "One project can't cover everything, but put them all together and they are very complementary," he says.

Crunch time. Human genome scientists met in Japan in 2002 for a marathon annotation session.

Japanese researchers took the lead in creating the database because of the large number of projects here and because Gojobori, who is also a professor of bioinformatics at the National Institute of Genetics in Mishima, got a 5-year, $22 million grant in 1999 as one of a number of Millennium Projects focusing on genetics. Additional support came from cDNA projects in China, Germany, and the United States.

To create the initial data set, the six centers pooled information on more than 41,000 cDNAs. To validate, standardize, and annotate the cDNAs, Gojobori hand-picked 100-plus genome scientists from 40 institutions in 12 countries for a 10-day session in summer 2002. The event was called the "Human Full-Length cDNA Annotation Invitational," or "H-Invitational," Gojobori says, "because we wanted only the best genome scientists to participate to ensure a high standard of quality."

The resulting database (www.h-invitational.jp) includes data on more than 20,000 unique cDNA sequences, including everything known about function, structure, tissue expression patterns, disease relationships, and orthologs in common experimental animals. Researchers who want a copy of a particular cDNA clone can contact the appropriate institute. And that is just a down payment. A second annotating marathon held last November produced data on 15,000 additional clones, which will be uploaded as soon as they are processed. The database is also prepared to accept submissions.

Analysis of the annotated data has already [illegible] several thousand [illegible]didate genes. And [illegible]ar biologist at the [illegible]Center in Heidel-[illegible]n cDNA consor-[illegible]and the clones to-[illegible]"investigate the [illegible]cale."

[illegible] database as the [illegible]information on [illegible] of Japan's major [illegible]. Japan was late [illegible]bution to human [illegible]rts, contributing only about 6% of the total. In contrast, the three Japanese projects contributed about 60% of the cDNA data going into the H-Invitational Database. "I think Japanese researchers wanted to make a unique contribution to genomic efforts," he says.

–DENNIS NORMILE

[illegible] versity jobs this year in what he terms "an exceptional and immediate effort." He has promised more jobs in 2005. The government had already agreed to "unfreeze" €294 million from the 2002 and 2003 budgets and reinstate 120 full-time civil service jobs.

"This is a big day for French research," [illegible] ballooning public deficit. And some tough negotiations are expected over a law to reform research, due to be unveiled at the end of the year; a number of controversial proposals will be on the table, including financing R&D projects rather than labs. –BARBARA CASASSUS

Barbara Casassus is a writer based in Paris.

Science (2004) 304: 368

News Front Page
World
UK
England
Northern Ireland
Scotland
Wales
Business
Politics
Health
Education
Science/Nature
Technology
Entertainment

Have Your Say
Magazine
In Pictures
Week at a Glance
Country Profiles
In Depth
Programmes
BBC SPORT
BBC WEATHER
CBBC news
BBC ON THIS DAY

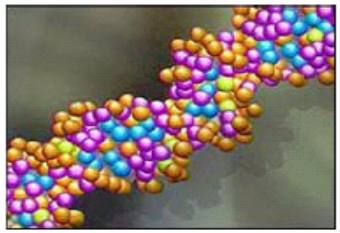
Last Updated: Tuesday, 20 April, 2004, 13:58 GMT 14:58 UK

E-mail this to a friend Printable version

Scientists decipher 21,000 genes

By Paul Rincon
BBC News Online science staff

An international team of 152 scientists has published a detailed map of more than 21,000 human genes.



The work is a major advance in our understanding of the genome

The work is seen as a major advance in the efforts to make sense of the genome, the DNA code that guides the building and maintenance of our bodies.

Sequencing of the human genome was officially finished in 2003, but scientists still need to interpret this vast resource of raw information.

The H-Invitational Consortium's work should aid the investigation of disease

The consortium, led by Takashi Gojobori of the Institute of Advanced Industrial Science and Technology in Japan [illegible]

"The genome wasn't designed by a computer programmer, from top to bottom. It keeps evolving all the time."

[illegible] sets

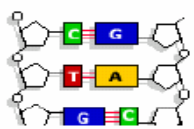
[illegible] the

[illegible] nes

and identify 5,155 new gene candidates.

"The gene is a very nebulous concept," co-investigator Anthony J Brookes, of the Karolinska Institute in Stockholm, Sweden, told BBC News Online.

THE DNA MOLECULE



BBC News (2004) April 20

ゲノム大量情報解析の時代

ー計算科学への渇望的期待ー

Horizontally transferred gene candidates detected in over 150 prokaryotic complete genomes

○Yoji Nakamura[1], Takeshi Itoh[2], Hideo Matsuda[3] and Takashi Gojobori[1,2]

Nature Genetics (2004) 36:760-766

バクテリア（細菌）のゲノムには、他の細菌から遺伝子が移入されるという現象の解明に向けて。
＝＞グリッド・コンピューティング

日本経済新聞

遺伝子を効率解析

遺伝研など グリッド技術利用

国立遺伝学研究所、理化学研究所などは、多数のパソコンを結んで一つの作業を分散処理する「グリッドコンピューティング」を利用して遺伝子を解析することに成功した。バイオ情報処理分野で、大規模なグリッドコンピューティングシステムを実動させたのは世界でも珍しい。研究チームは、膨大な遺伝子情報の高速処理にめどがついたと見ている。

参加したのは遺伝研生命情報・DDBJ研究センター、理研ゲノム科学総合研究センター、北陸先端科学技術大学院大学知識科学研究科、東京医科歯科大学生命情報学研究室と科学技術振興事業団。五団体で合計二百二十九台のパソコンをインターネットで結び、遺伝研DDBJセンターの微生物の遺伝子情報を分担して解析した。

微生物百十九種の約三十五万の遺伝子を分類させたところ、パソコン一台では従来五十日かかっていたものが二日で終わった。また、百十二種の微生物について約二十五万の遺伝子が微生物間を移動したかどうか調べ、進化の系統樹を描く計算をさせたところ、パソコン一台で四年半かかった時間が十八日でできたという。

科学

グリッドコンピューティングはバイオ分野でも欠かせない技術。研究者なら誰でも参加できる「オープンバイオインフォマティクス（OBI）」と呼ぶシステム作りも進んでいる。今回の実験もシステムの一環で、実験を統括した理研の小長谷明彦プロジェクトディレクターは「技術の有効性を実証できた」と説明している。

日本経済新聞
2003年8月18日(月)
朝刊

(理研・小長谷研との協力
富士通による協力)

遺伝研ＤＤＢとＮＴＴ西日本の共同実験

- 家庭用パソコン３０００台。グリッドコンピューティングを実施中。
- そのうち５００台はデータグリッドとして公開データベースを家庭用パソコンディスク上に公開。日本初
- 遺伝研ＤＤＢＪユーザーがＷＥＢから自由にこのシステムを相同性検索で使用できるよう公開中。
- 秘密保持の実証や実行時間の効率性を実験中。
- ＤＤＢＪ活動を広くアピール。一般社会のへ啓蒙活動を展開。

H16.3.11 静

保存版◇天 524169
00(二)NHKニュース10
▽盧武鉉大統領の弾劾問題で与野党対立激化
▽限りなく人に近づくロボット▽アテネ中継
五輪野球場 8508
00(S)(字)女神の恋(再)(終)
松本明子ほか 964256
15(S)(字)夢・音楽館 アスカ
▽迫真のソロライブ！桃井薫激のチャゲアス友情の絆 1807782
50 あすを読む「チェチェンの展望」 139071
0.00 N天◇15(S)(字)歴史が動いた(再) 7483706
1.05ガッテン(再) 4084469
1.50(S)天才タイルダーの肖像 加藤大治郎 1360541
2.41(S)映像散歩 9923164
4.00生き物選(終)◇45視点(再)
5.00 おはよう 40331096
8.15(字)てるてる 783725
8.30 N 112541

◇小ペット 91904
00(S)趣味・スポーツボウリング スペアを確実にクリアする 142530
25 わくわく授業 ポストの謎さがし 4716701
50 視点・論点 中村靖彦
00(S)100語英語◇英ビジネスワールド 188
30(S)スペイン語 97188
0.00(S)(字)トップランナー
日常を描く名手行定勲 田中麗奈が語る監督像
◇(S)ハッチ 5367812
1.00くらしの旅 3800725
1.30 日本史 6012270
2.00多人間講座(再) 斎藤環
若者の心(終) 5436980
2.50フォーユー 8106831
3.50高校物理(再) 5170305
4.20家庭総合(再) 7582812
4.50 春の島根 3482386
5.00(二)漢詩◇10自然◇名曲
5.30 ビジネス 78657
6.00(S)伊語会話 9928

00(S)(字)ドールハウス「危機一髪!/湯けむり旅情」
松下由樹 安達祐実
小池栄子 賀集利樹
中尾ミエ 80343
54 筑紫哲也NEWS23
米・保守派の論客が語る〝ならず者国家〟▽長嶋監督ほか 96119091
11.50 N天
55 ロンロバ！ アイドル最後の大物 4763121
0.35(S)Pooh！ アクション映画だ 3316541
0.50move！move！
0.55(S)からぁず ホワイトデー情報ほか 1495909
1.25名門アサ(終) 社民党に特捜最前線 7668218
1.55 テレショップ
(2.50終) 5771560
4.30(S)(N)バード 5953980
5.00 あさがけ 68218
6.00 ウォッチ 68906034

00 ダウンタウンDX
藤谷美和子カモ抜きのカモそば注文&小堺涙オシリ宅配&松本明子衝撃(終)赤面 793817
54(字)キスイヤ 彼女は今も元彼が好き…彼とどっちを取る？ 83124879
11.24(N)出来事 アフガン少年が歓喜…義足職人の贈り物▽58(S)MAX
新庄を支えるルーキーに独占密着 39462966
0.23MOBI サイドカーを作ろう▽出川のアメリカ珍道中 6937725
0.53 厳選こだわりのランチ20店安い！うまい！特集(再) 65506725
2.48 ホームショッピング
(3.18終) 6311522
5.00(S)買物 54299
5.30 ズームイン!!SUPER 94893980

54(S)ニュースステーション
大阪のカラスも…鳥インフルエンザ感染拡大で▽逮捕された元外務省課長補佐の告発第2弾〝機密費〟▽柔道野村忠宏選手 5415701
11.10(S)エコアップしずおか
15 銭形金太郎 激貧続々26歳かわいい貧娘の雪利用した生活術▽貧乏洗濯(終)テク 8902411
0.10 天
0.16パーティーへようこそ キャイーンがアダルト(終)宴会潜入 2280218
1.16(S)カーグラ 1852980
1.46(S)CDN 7550560
2.16 健康の扉 1851251
2.46(S)僕らはゲレンデファイターズin野沢
(3.44終) 3094541
4.40 買い物 9161218
5.10(S)演歌 5556218
5.40 食堂◇(N) 624724

00(S)(字)白い巨塔「最期の審判」山崎豊子作
唐沢寿明 江口洋介
黒木瞳 伊藤英明
西田敏行 91256
54(S)ビビビ・テレしず
00(S)VVV6 鎌倉の和風スイーツ！ 48508
30 ニュースJAPAN
その後…〝鶏感染〟は▽55(S)すぼると！ベッカム詳報！ 9216121
0.35くさデカ(再) (終)スープシチュー戦 2752522
1.05(S)テツand@トモ！
1.10 ロバートホール
5987725
1.35(S)CLダイジェスト
8強決定!! 2240473
2.35(S)直行便TV
(3.05終) 2781034
5.00 あかんたれ(再)
404164
5.25 めざまし 98148034

桜咲くまで
★SBSテレビ、後1・30

すべてを知った良子（大塚良重(念)）は身を引く決意をする

（大塚良重）は、互いに三浦（国広富之）のことを話していたと気付く。すべてを知った良子は、三浦と道子のために身を引くことを決意。しかしそれ以来、道子は三浦を避けるようになっていく。

（再春館製薬所、キユーピー㈱提供）

市毛良枝）と良子

ット曲を生み出ャゲ&アスカのロで登場。過去と最新のソロ曲〝トークでは、の曲をカバーしる司会の桃井か音楽論を交わつわる話題をオ式で送る。アメサス州の空港事が電光掲示板に停車してい後ろから、脇見衝突する。このついて、驚くべわれる。また、感リバボー」は、〝を抱えるあるソードを紹介す

朝日（朝刊）2月4日(水) P10 経済

巨大頭脳へ大実験

PC3000台ネットで結集

NTT西と国立遺伝学研

NTT西日本の高速通信サービス「フレッツ」加入者3千人を募り、常時接続している個人のコンピューターをつなぎ、「一つの仮想コンピューター」を立ち上げ、大規模に情報検索する、という実験をNTT西日本と国立遺伝学研究所が始める。「グリッド」と呼ばれる技術で、複数のコンピューターの中央演算処理装置（CPU）やハードディスクの余剰能力を寄せ集め、1台の高性能コンピューターを作り上げる仕組みだ。

3千台のパソコンの余剰能力を合わせれば、最大で通常のパソコンの約1500倍の能力が発揮できるという。

協力に応じた人はまず、NTTから送られる計算ソフトをパソコンに導入。NTTが用意するサーバーに接続する必要がある。

実験では、研究者が手元のコンピューターで、DNAの配列検索などの指示を打ち込むと、NTT西の管理サーバーを通じて検索依頼が各協力者のパソコンに出される。それぞれのパソコンが計算処理を分担し、検索結果は、NTT西の解析サーバーに集められ、まとめて研究者に回答する。

発表日 3月2日
関連記事 西日本他

毎日（朝刊）2月4日(水) P2 総合

家庭のパソコン3000台結んで〝スパコン〟に

今月から実験

NTT西など

NTT西日本と国立遺伝学研究所は3日、一般家庭のパソコンの余剰能力を集約して1台の高性能コンピューターに見立て、たんぱく質や遺伝子解析を行う共同実験を今月中旬～4月末に行うと発表した。同研究所が保有する世界最大の遺伝子データバンクの塩基配列データを各パソコンのハードディスクに分散保存、研究者が通信ネットワークを介し、必要に応じてデータ検索できる。将来的に、病気の遺伝子の発見や新薬開発への応用が期待されている。

複数のコンピューターを通信回線で結び、データの高速処理に役立てる「グリッド技術」を活用した。

共同実験では、NTTのブロードバンドサービスである「Bフレッツ」、「フレッツ・ADSL」の利用者から3000人程度の参加者を募る。家庭用パソコンのハードディスクやCPU（中央演算処理装置）などの余剰能力を集めることで、これまで同研究所内の高性能コンピューターが行っていた役割を肩代わりさせる。意味を成さない単位にデータを分散管理することで、安全を確保できる。

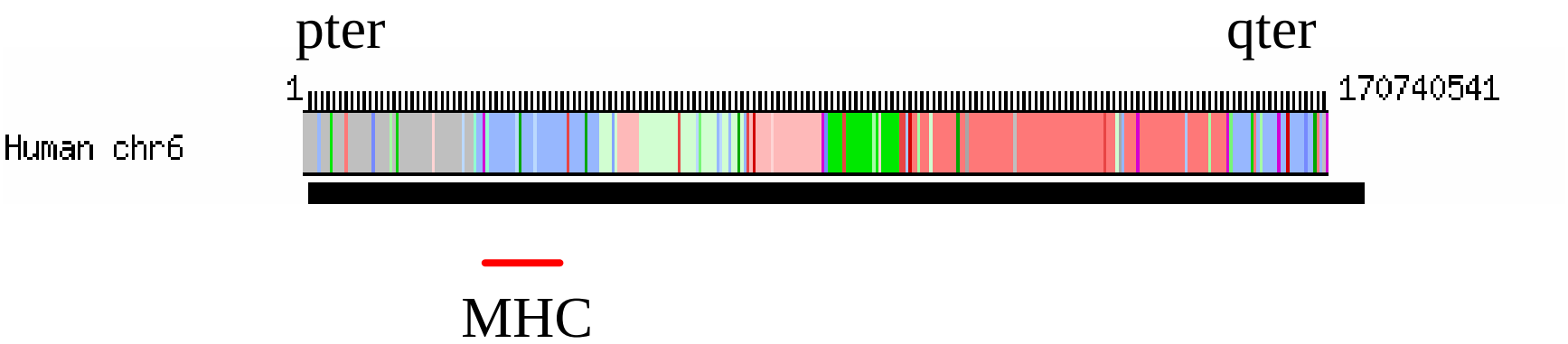
同研究所は14万種の生物の遺伝子データを保存。年々、データ量は増える一方だが、実験が成功すれば、高性能コンピューターを拡充する必要がなくなり、データの一極集中も避けられる。

実験参加者は3月中旬まで、ホームページ（http://www.bioathome.jp/）上で受け付ける。NTT西日本エリアの「Bフレッツ」「フレッツ・ADSL」の利用者が対象で、参加費用は無料。NTTのグループ企業が提供するウェブ上のショッピングモールで、最大3750円相当の買い物の割引ができるポイントももらえる。【岡田功】

発表日 3月2日
関連記事 西日本他

Whole Genome Alignment of Human and Mouse

Human Chromosome 6



Corresponding Mouse Chromosomes:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, Y

Highly Conserved Upstream Sequences for Transcription Factor Genes and its Evolutionary Implication to Regulatory Network

H. Iwama and T. Gojobori

Proc. Natl. Acad. Sci. (2004)

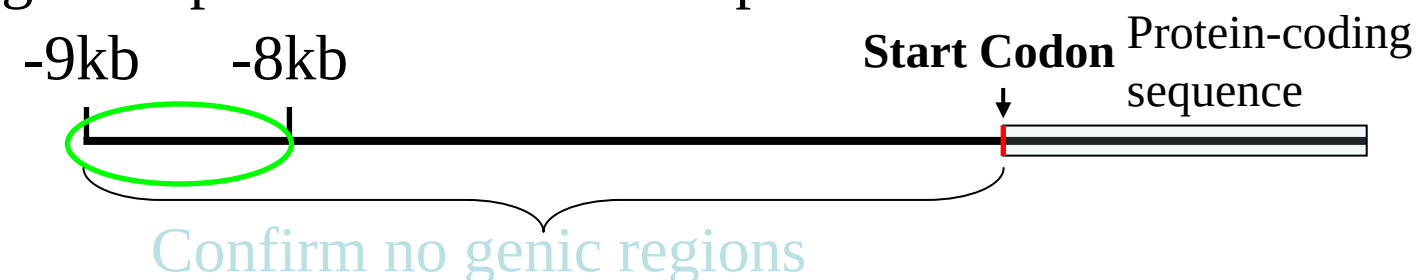
2004 Dec 7;101(49):17156-61.

Check genomic sequences and annotations

(1) Check if a genomic sequence 5'-upstream of translation start site is fully available along the 9-kb stretch in the particular genome contig sequence.

(2) See genomic contig annotations to check if the 9-kb upstream sequences do NOT include any genic region (even if it is only predicted).

Cut out 8-kb genomic sequences upstream of translation start site only when **no** genic region is present within 9-kb upstream stretch.



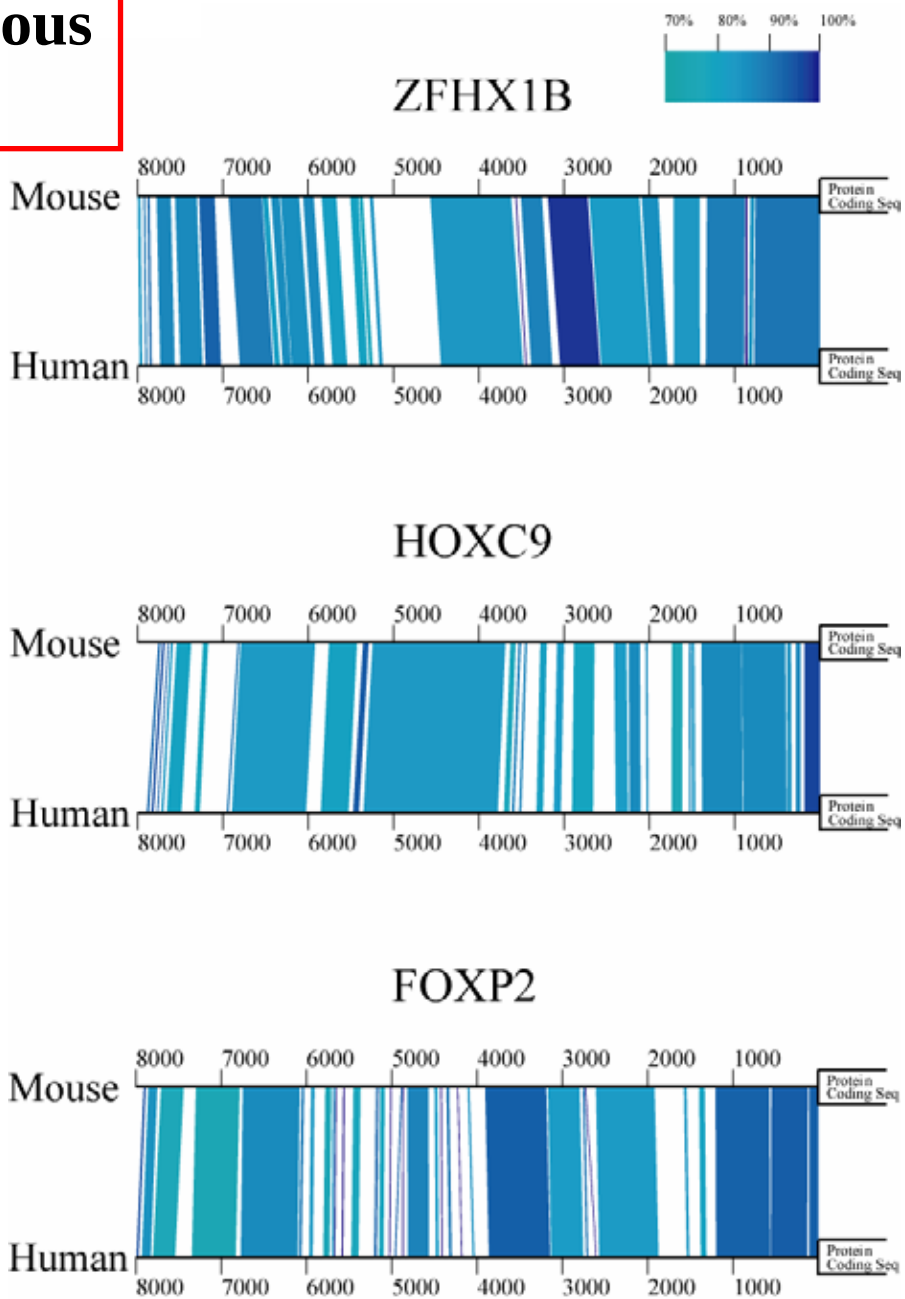
8kb-upstream orthologous genomic sequence set

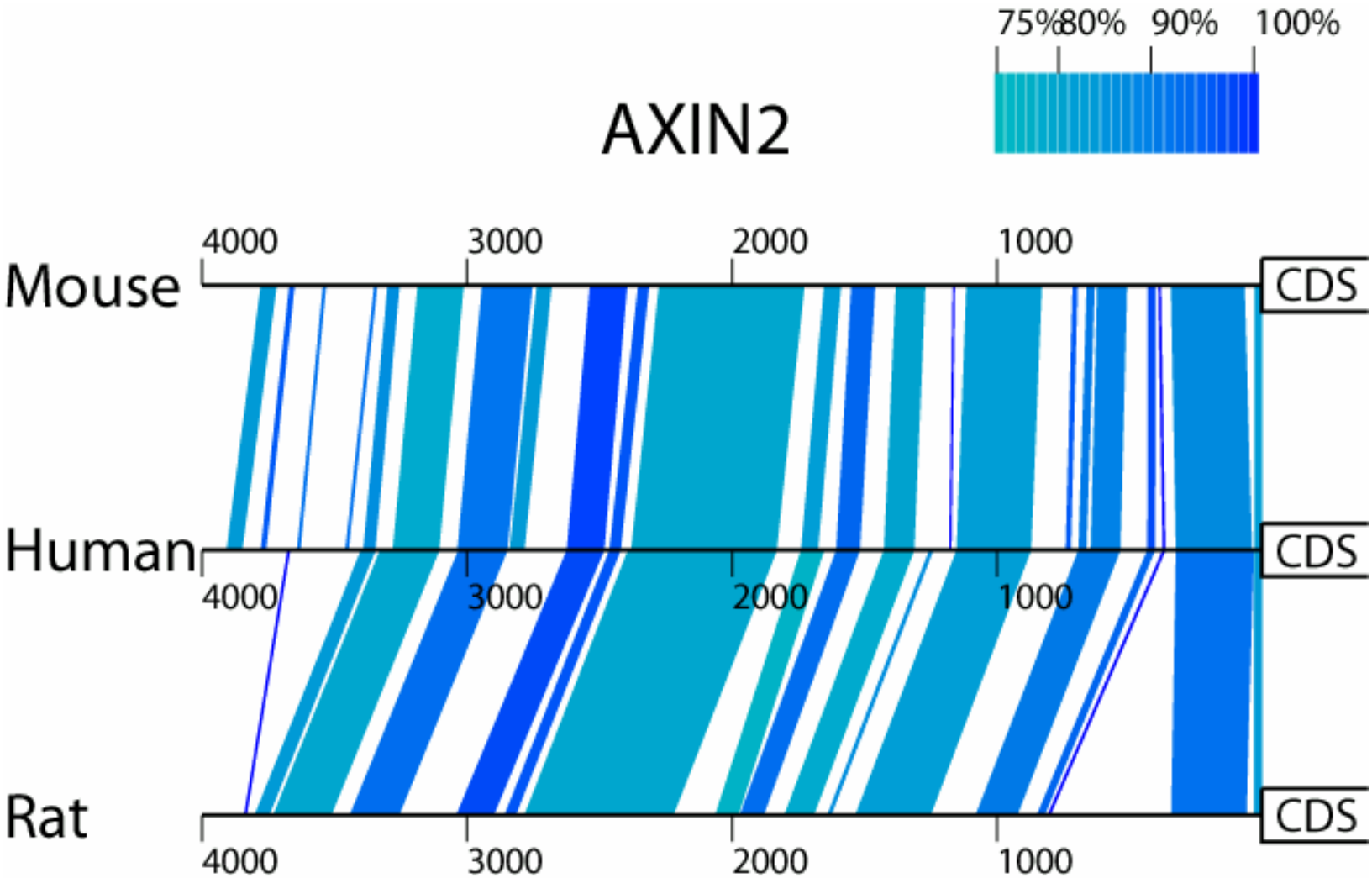
In total, 3,750 human-mouse orthologous upstream sequence pairs were obtained.

Top-3 Upstream-Conserved Human-Mouse Orthologous Genes

Within the top-10 upstream-conserved genes, 9 genes were ***transcription factor*** genes. ($p < 2*10^{-8}$)

62 genes of the top-200 upstream-conserved genes were also ***transcription factor*** genes. ($p < 5*10^{-15}$)





Reliability of our alignment method was shown by comparing human-mouse and human-rat orthologous upstream conserved regions.

Top-30 Upstream-Conserved Orthologues

Rank	Official Gene Symbol	Number of Identical Sites	Gene Name
1	ZFHX1B	6000	zinc finger homeobox 1b
2	HOXC9	5455	homeo box C9
3	FOXP2	5402	forkhead box P2
4	LHX2	4912	LIM homeobox 2
5	NR4A3	4873	nuclear receptor subfamily 4, group A, member 3
6	OTX2	4601	orthodenticle homolog 2 (Drosophila)
7	PITX2	4536	paired-like homeodomain transcription factor 2
8	NR4A2	4413	nuclear receptor subfamily 4, group A, member 2
9	INHBA	4400	inhibin, beta A (activin A, activin AB alpha polypeptide)
10	SIX1	4398	sine oculis homeobox homolog 1 (Drosophila)
11	NTNG2	4393	netrin G2
12	PAX6	4362	paired box gene 6 (aniridia, keratitis)
13	SP8	4235	Sp8 transcription factor
14	BAI3	4178	brain-specific angiogenesis inhibitor 3
15	MLLT10	4110	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 10
16	EYA1	4069	eyes absent homolog 1 (Drosophila)
17	OTP	4055	orthopedia homolog (Drosophila)
18	DNAJB5	3995	DnaJ (Hsp40) homolog, subfamily B, member 5
19	PROX1	3932	prospero-related homeobox 1
20	MEF2C	3931	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)
21	ELAVL2	3860	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)
22	HOXD4	3857	homeo box D4
23	NR2F1	3841	nuclear receptor subfamily 2, group F, member 1
24	PAX2	3838	paired box gene 2
25	DLL1	3809	delta-like 1 (Drosophila)
26	HOXD3	3802	homeo box D3
27	PCDH7	3793	BH-protocadherin (brain-heart)
28	NRXN3	3767	neurexin 3
29	CDK6	3763	cyclin-dependent kinase 6
30	LDB1	3753	LIM domain binding 1

Yellow-shaded are TF genes.

19 of top 30 were TF genes. ($p < 5*10^{-12}$)

高度な生命現象理解への応用

ー医薬学・臨床と産業への応用ー

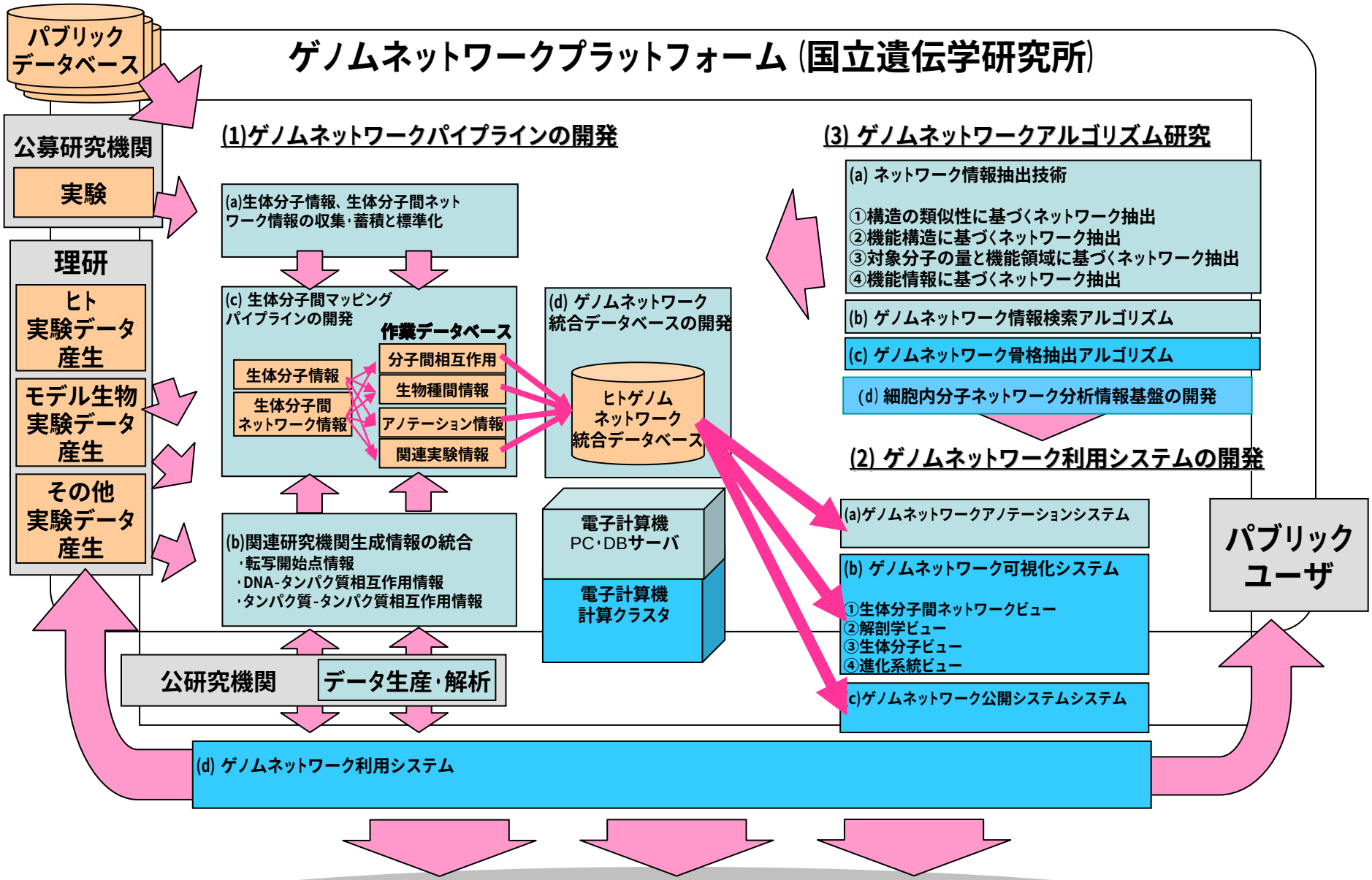
新プロジェクトの推進

ゲノムネットワーク・プロジェクト

- 転写制御ネットワークの解明
- タンパク質間相互作用の解明
- 情報プラットフォーム・統合データベースの構築
- 疾病・創薬などへの応用

日立製作所、日立ソフト、富士通、三井情報開発

ゲノムネットワークプラットフォームの構造



超高速検索システムの開発

- 富士通との共同開発
(シュンサク技術の応用)

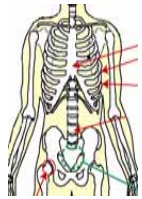
・ベンチャーの立ち上げ
国際バイオインフィマティクス研究所
国立大学教授１４名、日立、富士通他

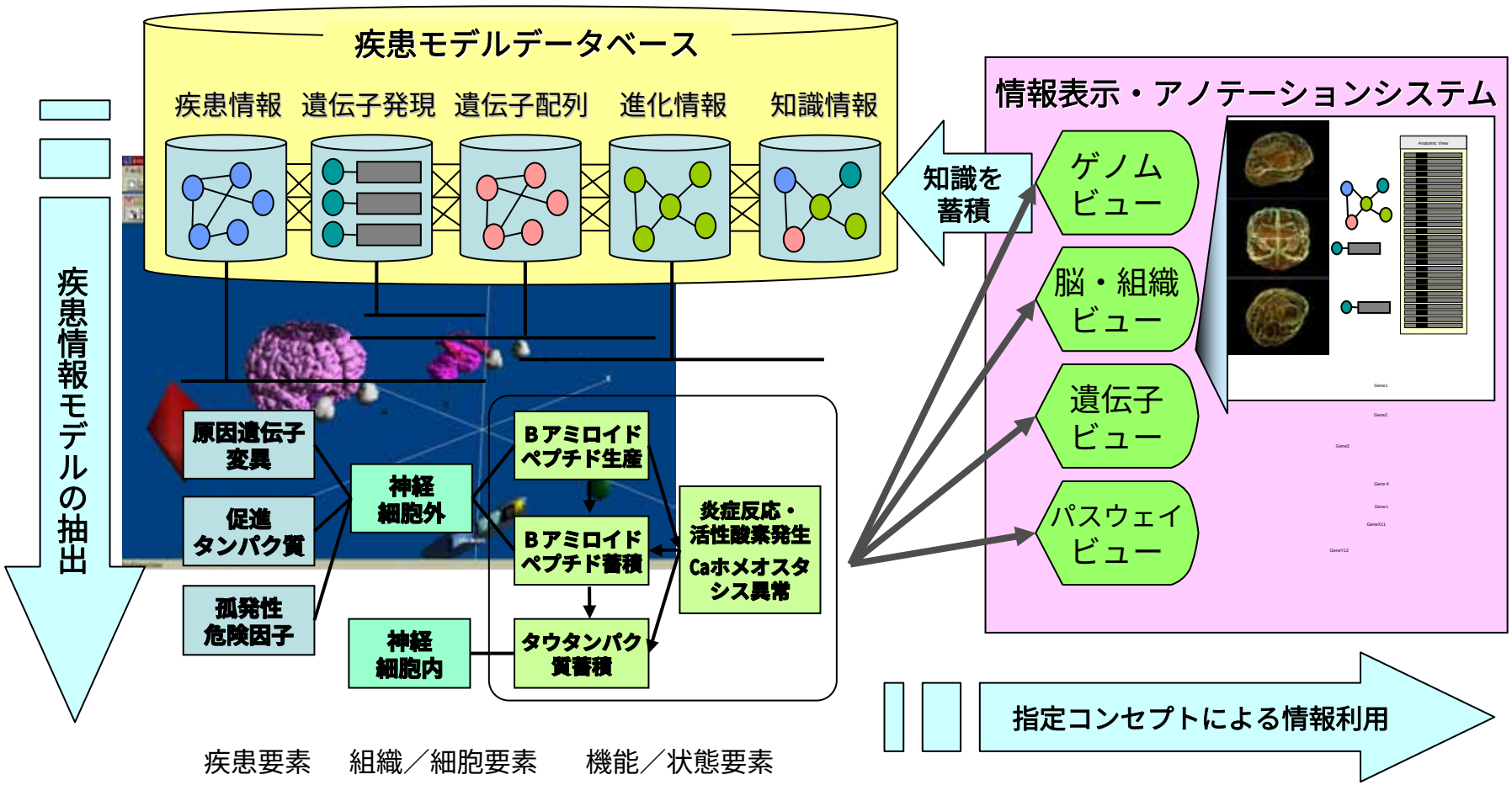
病態のシステム的理解と疾患情報モデルの構築

特定領域研究「応用ゲノム」
(代表　辻省次)

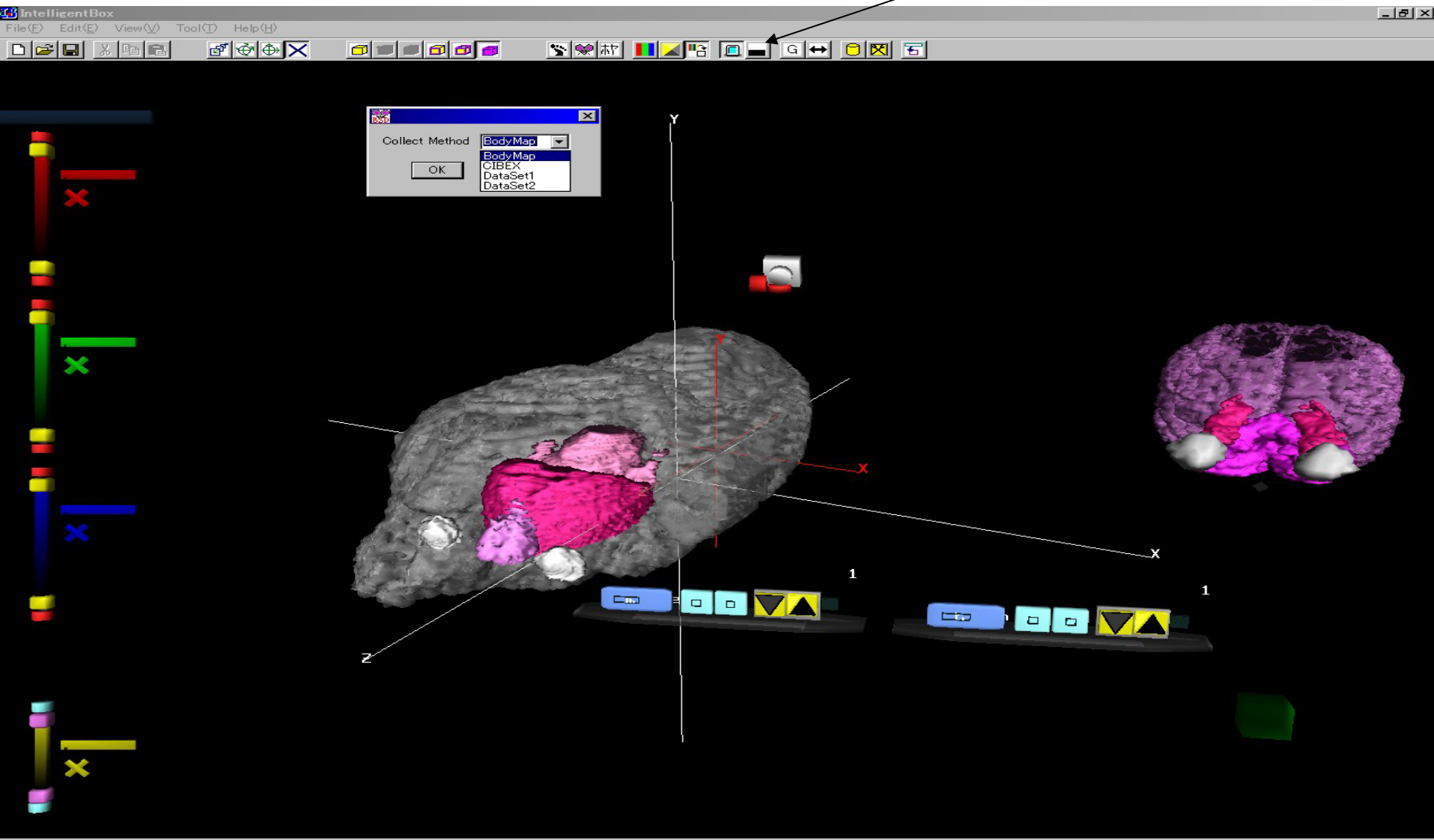
＜方法＞

疾患情報モデルDBシステム概略図





マウス脳との比較

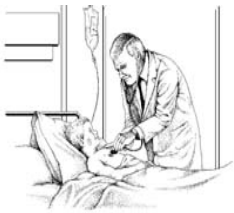


Color Control

「期待されるリサーチ・モデルと事業モデルの融合」

インターネット

病院



サンプル

医師

①血液採取

④解析結果

セキュリティ

ゲノム臨床データベース解析センター

DB管理

病気遺伝子DB　健康管理DB　副作用DB　……　その他DB

遺伝子探索研究

連鎖不均衡解析

ハプロタイプ頻度解析

薬剤関連遺伝子多型解析

ディプロタイプ解析

PPI・パスウェイ解析

遺伝子発現制御・プロテオーム

病変組織・器官の画像・イメージデータ

公共DB

統合DB

商用DB

その他DB

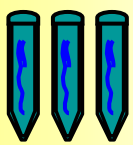
③データ

②検査

⑤リスク評価・診断・治療

タイピング研究開発センター

DNA・プロテオミックス・細胞



SNP
マイクロサテライトマーカー
遺伝子発現
抗原抗体反応
タンパク質量分析
シグナルトランズダクション
細胞同定

生命倫理委員会

戦略提案

ライフサイエンスにおける生命情報

(要点) 様々な生命情報が世界中で日々産生されている。

各生物種完全ゲノム情報(233種)

文献情報

生命情報の飛躍的な膨張

基礎研究情報 (実験)

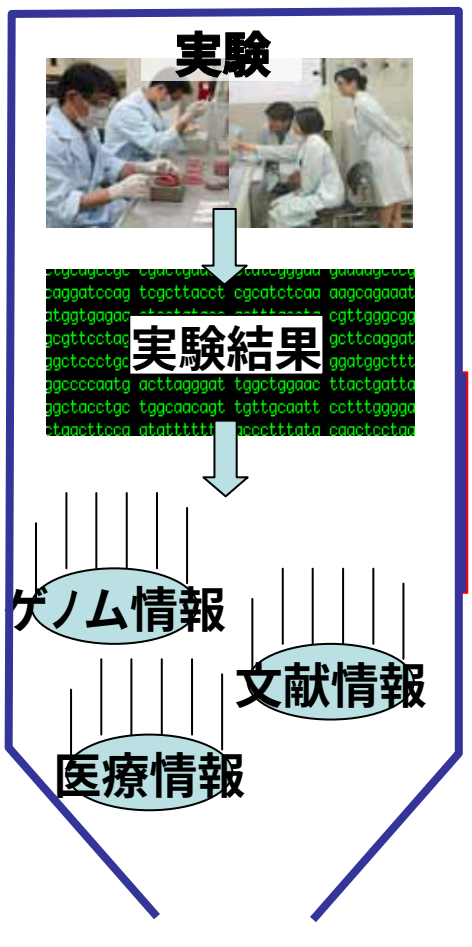
健康･医療情報 (臨床)

ポストゲノム時代のライフサイエンス

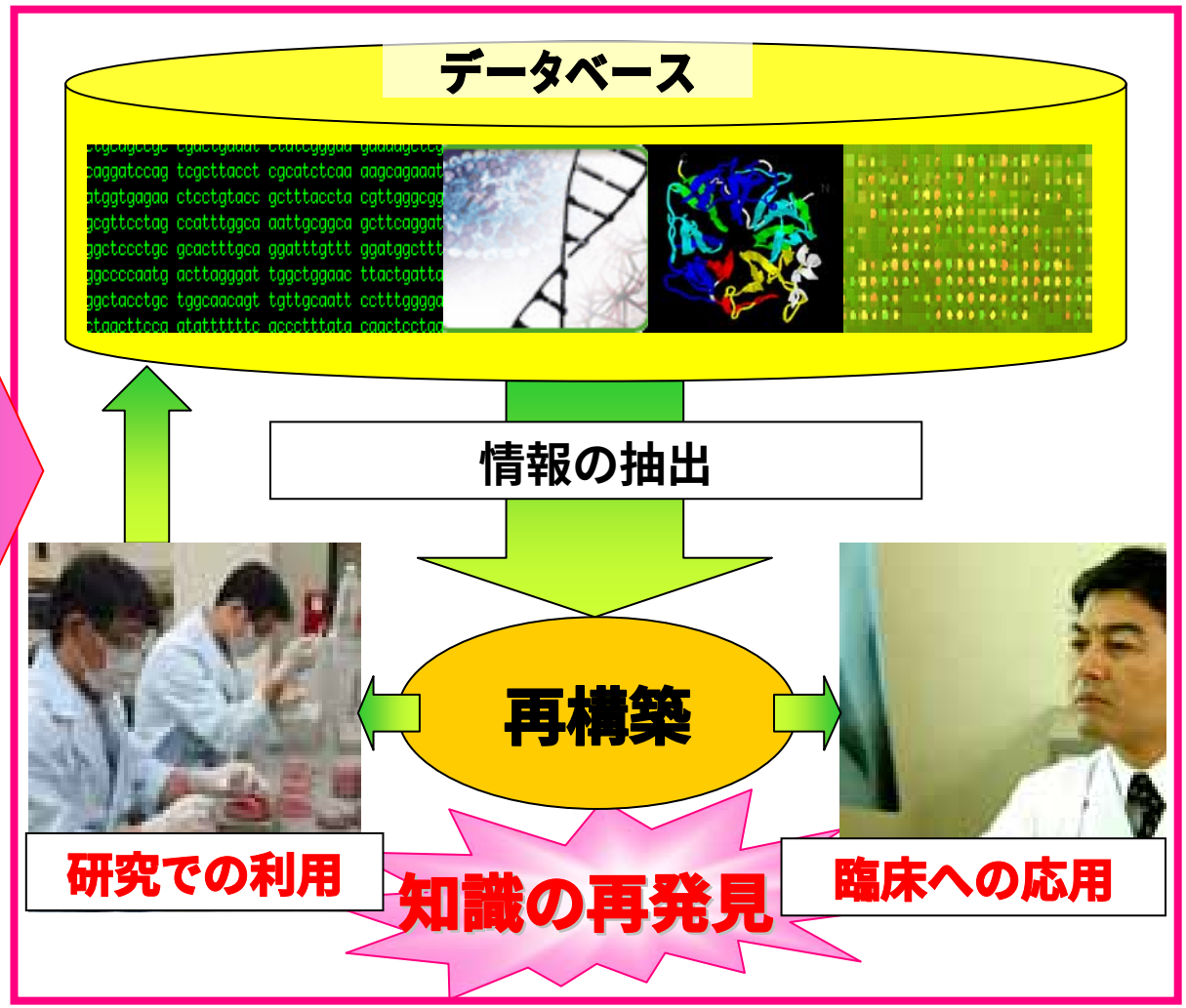
(要点) 生命情報を扱うことからさまざまな研究が始まる。

生命情報の氾濫

データベースから始める研究

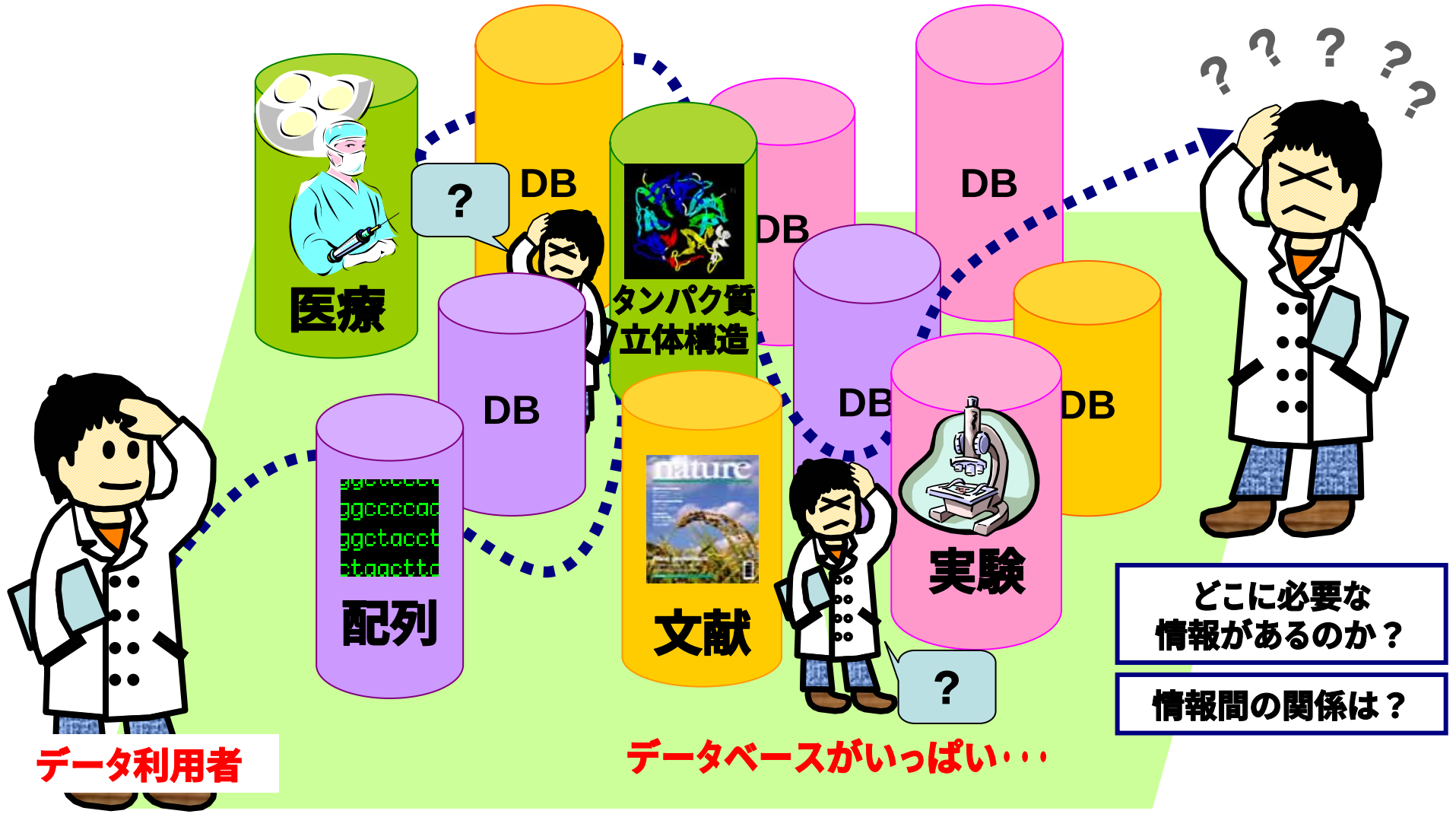


パラダイムシフト

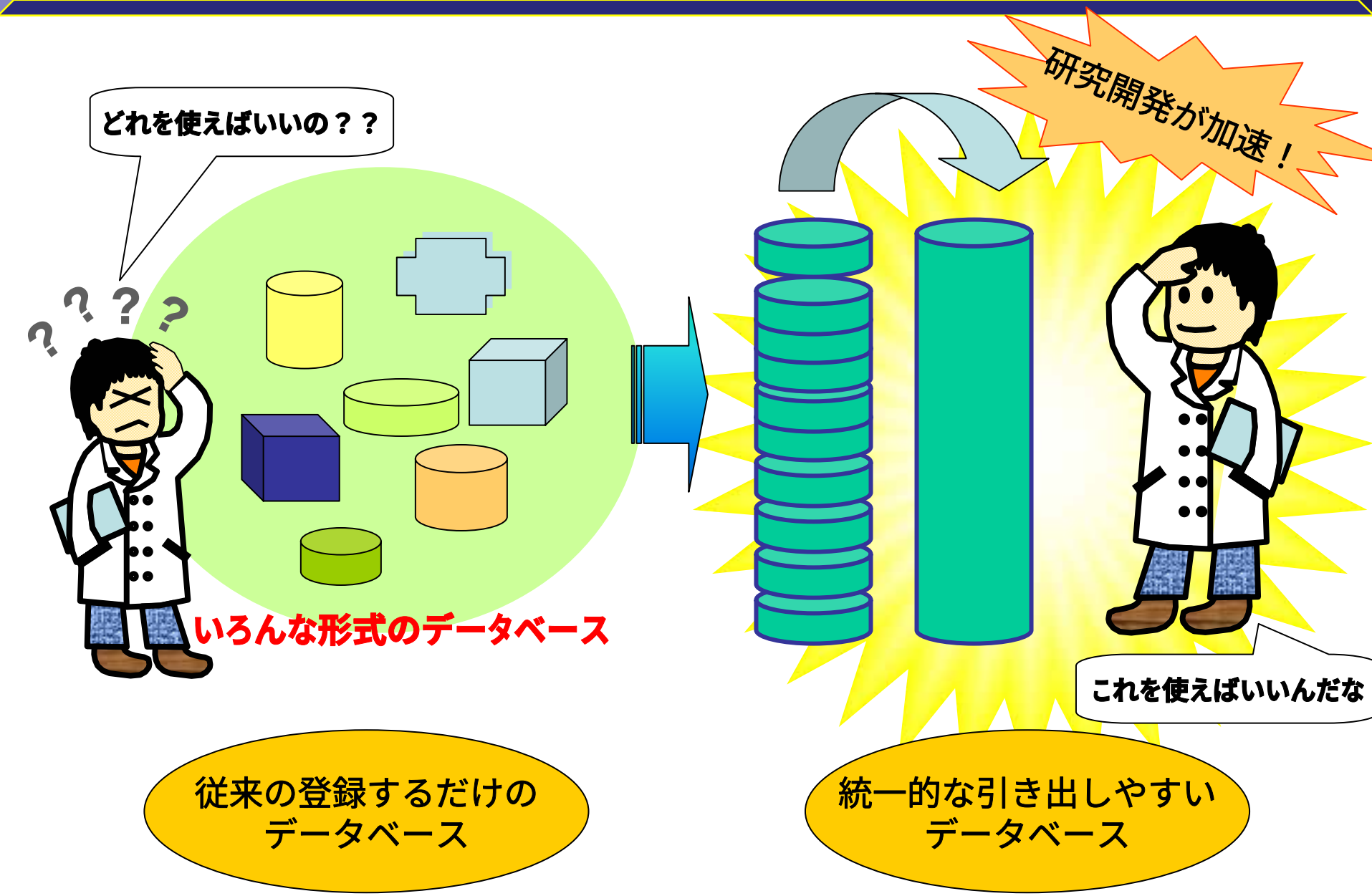


必要な情報を取得するにあたっての問題

(要点) 独自の形式で保存されていて統一的に集められていない情報は、研究開発の深刻な障害になっている。



散在した情報を統一的で且つ引き出しやすい形式へ変革を！



四位一体のバイオインフォマティクス

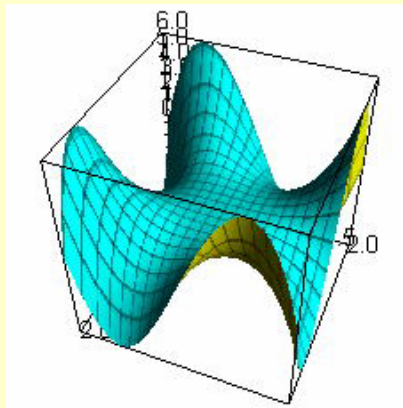
(要点) 4つの標榜テーマを機軸にバイオインフォマティクスを発展させる

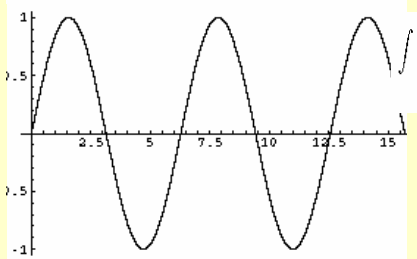
ナショナルデータセンター設立には何が必要か？

(要点) 計算機資源の増強と生命情報分野のソフトウエア開発及びR＆Dが必要である。

ライフサイエンス統合データベース

総合データベース　大型高性能計算機　ソフトウエア開発　人材育成





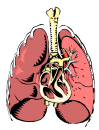
$$\int \frac{dx}{\cos x} = \int \frac{dx}{\sin\left(x+\frac{\pi}{2}\right)} = \int \frac{dt}{\sin t} = \log\left|\tan\frac{t}{2}\right| = \log\left|\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right|$$

ライフサイエンス統合データベースの将来像

(要点) 進行中プロジェクトの継続・統合により実用性を追求したデータの体系的な集合体による社会貢献を目指す。

生態系
社会
個体
臓器
細胞
分子

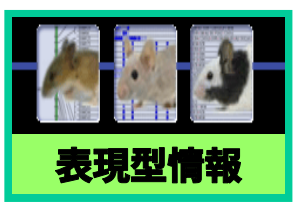

健康・医療情報

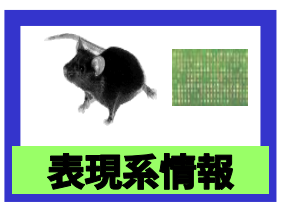



Red Blood Cell

情報

基礎研究情報


表現型情報


表現系情報

ライフサイエンス統合データベース

●実験情報
・マウス表現型データベース
・機能性RNAデータベース
・
・

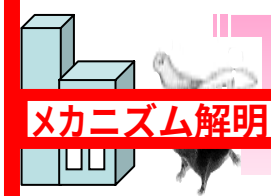
●医療情報
・SNPハプロタイプ
・集団遺伝学データベース
・家系データベース
・
・

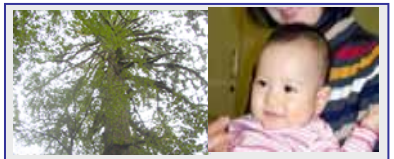
●文献情報
・PubMedCentralJapanの確立
・バイオテキスト・マイニング
・
・

●DDBJ関連の基本整備


新薬探索
製薬会社


治療
病院


メカニズム解明
実験施設



国民・社会
健康社会の実現

産業
研究機関・医療機関等による情報利用

「ペタコン・プロジェクトとの緊密連携」

|

「連携センターの設立」

Biological hierarchy

Evolution

Ecosystem (Environments)

Population (Human population)

Individual (Human)

Organ (Lung, Stomach)

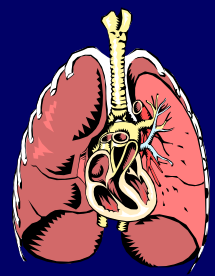
Tissue (Epidermal tissue)

Cell (Red blood cell)

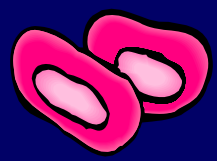
Organella (Mitochondria)

Bio-molecules (DNA, RNA, Proteins)

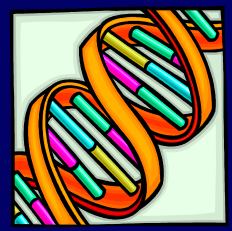
Molecules (H_2O, O_2)

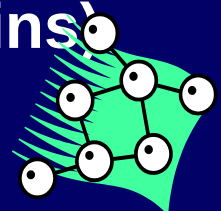


Integration

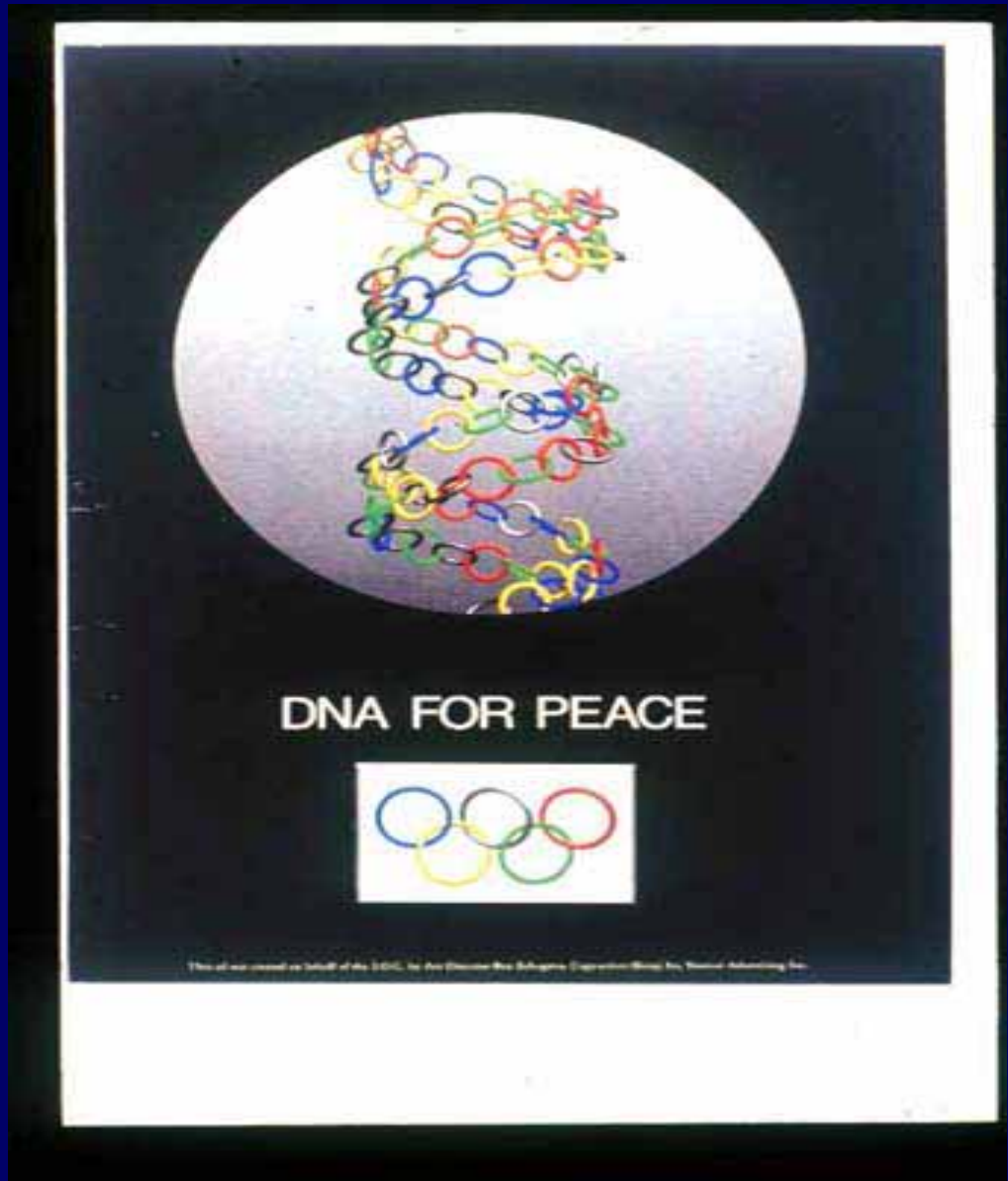


Red Blood Cell





社会に貢献するＤＮＡ情報！



The financial support was given from:

- **MEXT** (Ministry of Education, Science, Sports, and Culture, Japan)
- **METI** (Ministry of Economy, Trade, and Industry, Japan)
- **JBIC** (Japan Biological Informatics Consortium, Japan)
- **NIH** (National Institutes of Health, US)
- **DOE** (Department of Energy, US)
- **CNRS** (Centre National de la Recherche Scientifique, France)