

Bioinformatics Center -Bioknowledge Systems-



Prof
KANEHISA, Minoru
(D Sc)



Assoc Prof
GOTO, Susumu
(D Eng)



Instr
NAKAYA, Akihiro
(D Sc)



PD
OKUNO, Yasushi
(D Pharm Sci)



PD
VERT, Jean-Phillipe
(D Sc)



PD
BICHET, Adeline
(D Sc)

Assoc Instr (pt)

HATTORI, Masahiro
IGARASHI, Yoshinobu
KATAYAMA, Toshiaki

Res

HIRAKAWA, Mika
PARK, Keun-joon

Research Student

LEVCHENKO, Maria

Students

NAKAO, Mitsuteru (D3)
YOSHIZAWA, Akiyasu (D3)
OKUJI, Yoshinori (D2)
KOTERA, Masaaki (D1)
LIMVIPHUVADH, V (D1)
MINOWA, Yosuke (D1)
ITOH, Masumi (M2)
FUJITA, Masashi (M1)
MORIYA, Yuki (M1)
OKUDA, Shujiro (M1)
YAMADA, Rikuhiro (M1)
YAMADA, Takuji (M1)

Visitors

ROLOFF, Tim Louis Pasteur University, January 2001 - August 2001
PALMER, Andrew University of Wisconsin-Madison, July 2001 - August 2001
SELINGER, Douglas Harvard Medical School, July 2001 - August 2001

Scope of Research

Owing to continuous developments of high throughput experimental technologies, ever increasing amounts of data are being generated in genomics and proteomics. We have been developing bioinformatics technologies for analyzing a large number of genes or proteins at a time, toward the understanding and utilization of higher order functional information of the cell or the organism. The suite of databases and associated software that we develop is called KEGG and is made publicly available as part of the GenomeNet service (<http://www.genome.jp>).

Research Activities (Year 2001)

Grants

Kanehisa M, Deciphering genetic and molecular networks by comparative genomics and systematic interaction experiments. Genome Frontier Project, MEXT.

Kanehisa M, Biological systems database and genome information science. Research for the Future Program, JSPS.

Kanehisa M, BRITE: deductive database of the genome and the biological system based on binary relations. Bioinformatics Research and Development, JST.

Goto S, Construction and retrieval of highly integrated biological databases. Grant-in-Aid for Scientific

Research on Priority Areas (C) "Genome Information Science", MEXT.

Nakaya A, Extraction of correlated gene clusters by parallel data mining, Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Information Science", MEXT.

Award

Kanehisa M, Okawa Publications Prize, *Invitation to Post-genome Informatics*, Okawa Foundation for Information and Telecommunications, 29 November 2001

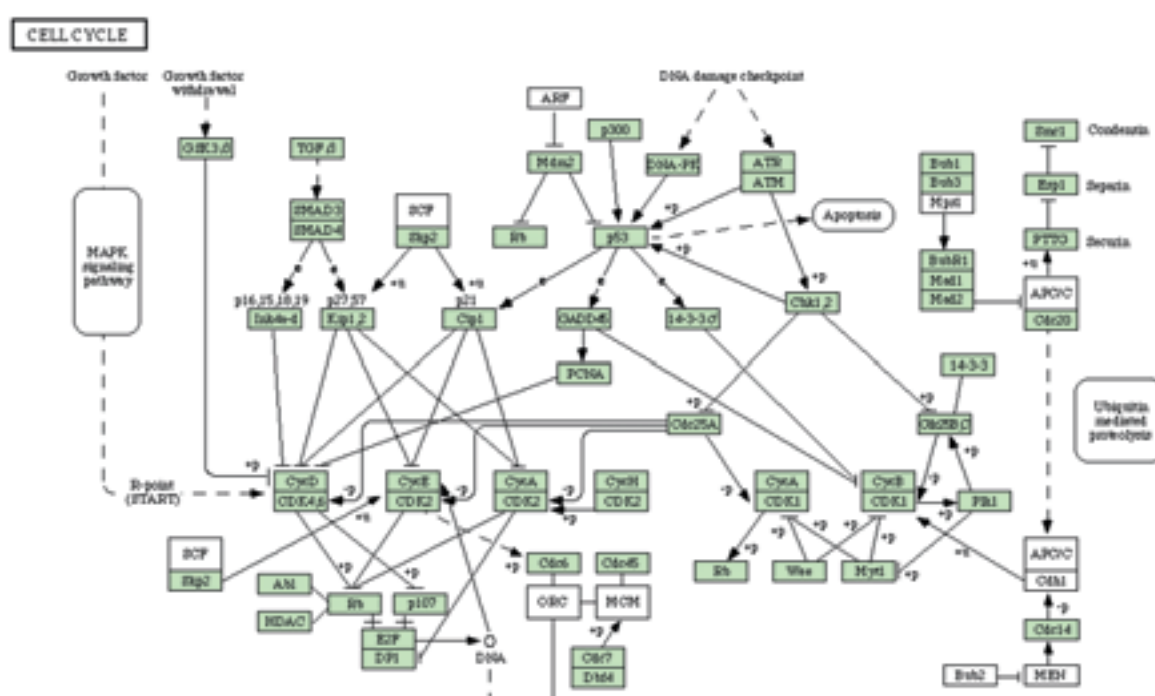
Topics

KEGG/SSDB is a new database for exploring the protein universe

SSDB (Sequence Similarity Database) is a new addition to the KEGG suite of databases. It contains the information about amino acid sequence similarities among all protein coding genes in all known genomes. The parallel supercomputer has been used to compute sequence similarities of 5×10^{10} pairs of genes, as well as continuous updates resulting from newly determined genomes. The data in SSDB can be considered as a huge graph consisting of protein-coding genes as nodes and similarity relations as edges. The graph algorithms that we have developed are used to perform extensive analyses of this graph, such as the gene



clustering of orthologs and paralogs, genome comparisons and functional predictions.



Human cell cycle pathway and its comparison to viral genomes

Molecular mechanisms of the eukaryotic cell cycle regulation have been studied extensively in the past decade. We have assembled current knowledge from published literature and constructed yeast and human cell cycle regulatory pathway diagrams under the KEGG project. Compared to other works, our presentation provides an overall picture on the control flows involving various molecular interactions in the eukaryotic cell division. The pathway diagrams can be used for gene function assignment in other organisms, comparative network analysis of complex biological pathways, visualization and correlation analysis of microarray gene expression data, among others.

Here we show the result of mapping homologous viral genes onto the pathway diagrams by sequence similarity searches. We have constructed a database of viral genes from a set of complete viral genomes, called vGENES. As a result of homology searches of vGENES entries against pathway components, it is shown that many viruses have counterparts of the cell cycle regulatory genes. For example, viruses have G1 Cyclin/CDK and its regulators or G1 transcription initiators, but do not have any subunit of a large protein complex. Such a tendency suggests that viruses carry only those genes that can critically affect initiation of the host cell's proliferative activities.