



## The Mammalian Protein–Protein Interaction Database and Its Viewing System That Is Linked to the Main FANTOM2 Viewer

Harukazu Suzuki, Rintaro Saito, Mutsumi Kanamori, et al.

*Genome Res.* 2003 13: 1534-1541

Access the most recent version at doi:[10.1101/gr.956303](https://doi.org/10.1101/gr.956303)

---

**References** This article cites 32 articles, 12 of which can be accessed free at:  
<http://genome.cshlp.org/content/13/6b/1534.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# The Mammalian Protein–Protein Interaction Database and Its Viewing System That Is Linked to the Main FANTOM2 Viewer

Harukazu Suzuki,<sup>1,3,6</sup> Rintaro Saito,<sup>1,5</sup> Mutsumi Kanamori,<sup>1</sup> Chikatoshi Kai,<sup>1</sup> Christian Schönbach,<sup>4</sup> Takeshi Nagashima,<sup>4</sup> Junko Hosaka,<sup>4</sup> and Yoshihide Hayashizaki<sup>1,2,3</sup>

<sup>1</sup>Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan; <sup>2</sup>Genome Science Laboratory, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan; <sup>3</sup>JST-CREST, Saitama 351-0198, Japan; <sup>4</sup>Biomedical Knowledge Discovery Team, Bioinformatics Group, RIKEN Genomic Sciences Center (GSC), Yokohama 230-0045, Japan

Here, we describe the development of a mammalian protein–protein interaction (PPI) database and of a PPI Viewer application to display protein interaction networks (<http://fantom2l.gsc.riken.go.jp/PPI/>). In the database, we stored the mammalian PPIs identified through our PPI assays (internal PPIs), as well as those we extracted and processed (external PPIs) from publicly available data sources, the DIP and BIND databases and MEDLINE abstracts by using FACTS, a new functional inference and curation system. We integrated the internal and external PPIs into the PPI database, which is linked to the main FANTOM2 viewer. In addition, we incorporated into the PPI Viewer information regarding the luciferase reporter activity of internal PPIs and the data confidence of external PPIs; these data enable visualization and evaluation of the reliability of each interaction. Using the described system, we successfully identified several interactions of biological significance. Therefore, the PPI Viewer is a useful tool for exploring FANTOM2 clone-related protein interactions and their potential effects on signaling and cellular communication.

[The protein–protein interaction data that have been derived from our experiments and are newly described in this paper have been submitted to the BIND database.]

Recent large-scale genome- and cDNA-sequencing projects have provided information on complete sets of genes in numerous organisms in which more than half of identified genes have been uncharacterized (Goffeau et al. 1996; The *C. elegans* Sequencing Consortium 1998; Adams et al. 2000; Kawai et al. 2001; Lander et al. 2001; Venter et al. 2001; Okazaki et al. 2002; Waterston et al. 2002). One of the most effective approaches to elucidating the function of the products of these unknown genes is large-scale analysis of protein–protein interactions (PPI), because most proteins act in complexes to regulate the biological processes of organisms from development to disease (Oliver 2000; Pawson and Nash 2000). To this end, high-throughput genome-wide screening of PPIs has been carried out for *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and higher organisms such as *Mus musculus* (Uetz et al. 2000; Walhout et al. 2000b; Ito et al. 2001; Suzuki et al. 2001). The experimentally produced PPIs are stored in several publicly available PPI databases. The major part of PPIs in many PPI databases is now yeast PPIs, because comprehensive PPI data sets are available. However, it is important to design

PPI databases that focus on other organisms, such as mammals, because large-scale analysis soon will produce a huge number of PPIs.

Many of the PPI data were derived through the yeast two-hybrid method, and the usefulness of the analysis appeared to be limited by the high number of false-positive interactions (Legrain et al. 2001). Because the affinity of each interaction is not appropriately captured in yeast two-hybrid systems, false-positive interactions complicate the identification of biologically relevant interactions. In addition, incorrect conclusions regarding biological importance may be derived from these erroneous interactions. von Mering et al. (2002) recently estimated that as many as 50% of interactions obtained from yeast two-hybrid screens are false-positives. Actually, yeast two-hybrid-derived PPI data from two independent research groups showed very little overlap (Uetz et al. 2000; Ito et al. 2001), although this finding may reflect the high false-negative rate of these data sets. In designing the PPI database, we therefore emphasized the evaluation of each interaction so that reliability can be considered for constructing the protein network. Current public PPI databases have not addressed this problem very well.

Recently, the FANTOM2 (the second Functional Annotation Meeting of Mouse cDNA) consortium reported the functional annotation of 60,770 full-length cDNAs from the RIKEN mouse encyclopedia project (Okazaki et al. 2002). Expert curators manually confirmed the computational annota-

<sup>5</sup>Present address: Institute for Advanced Biosciences, Keio University, Yamagata 997-0035, Japan

<sup>6</sup>Corresponding author.

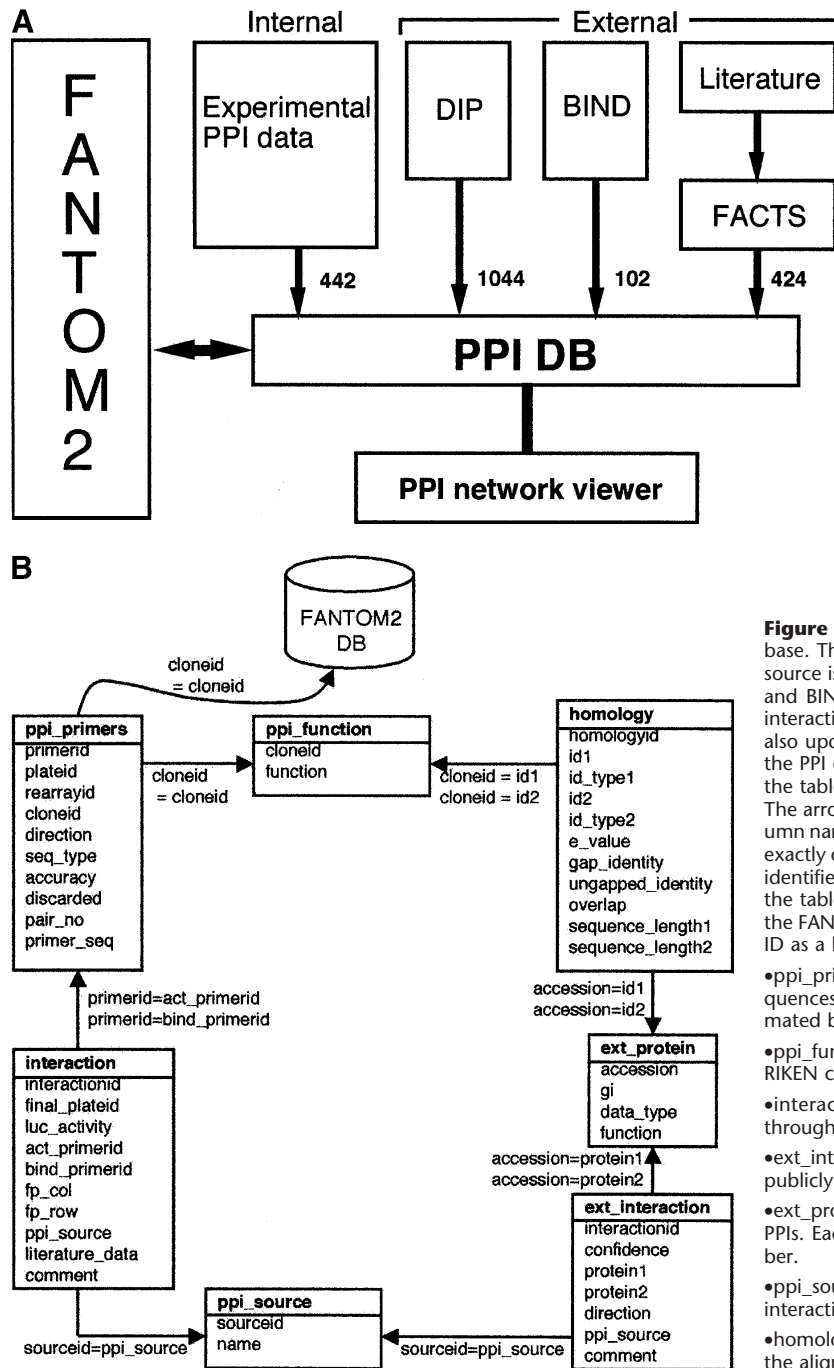
E-MAIL [rgscerg@gsc.riken.go.jp](mailto:rgscerg@gsc.riken.go.jp); FAX (81) 45-503-9216.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.956303>.

tion of the FANTOM2 clones to provide the appropriate gene name, gene symbol, coding sequence information, and relevant gene ontology (GO) terms. This information can be very useful in analysis of experimental PPI data, because we are systematically analyzing PPIs by using the FANTOM2 cDNAs (Suzuki et al. 2001). However, our experimental data are not comprehensive at present. In organisms with tens of thousands of genes, it is difficult to establish comprehensive PPI data sets, because the total number of experimental examinations is estimated to be far larger than that for budding yeast. Further, we may have lost information because of false-

negative PPIs generated by the high-throughput assay system. To compensate for the incompleteness of our PPI data sets, it was necessary to analyze the experimental PPI data together with publicly available information on mammalian PPIs.

Here, we describe the development of a mammalian PPI database and a PPI Viewer, a graphical tool that is linked to the main FANTOM2 viewer. We integrated publicly available PPI information with information we obtained from our PPI assays into the PPI database and also included data regarding the luciferase reporter activity or confidence of each interaction.



**Figure 1** RIKEN PPI database. (A) Overview of the database. The number of interactions obtained from each data source is shown near the arrows. The interactions from DIP and BIND public database are updated regularly, and the interactions from the internal experiments and FACTS are also updated in a timely fashion. (B) The table structure of the PPI database. Each box represents a table and contains the table name at the top, followed by the column names. The arrows and equation labels indicate corresponding column names between two tables. Using this correspondence, exactly one entry in the table at the end of the arrow can be identified by using a corresponding value in the column of the table at the beginning of the arrow. A specific entry in the FANTOM2 database can be retrieved by using the clone ID as a key. The various tables are:

- ppi\_primers: Contains primer IDs as well as primer sequences and clone IDs. A protein sequence can be estimated by using this table and the RIKEN cDNA database.
- ppi\_function: Includes the functional annotation for each RIKEN clone.
- interaction: Comprises a list of interactions identified through our assays (internal PPIs).
- ext\_interaction: Stores a list of interactions extracted from publicly available data sources (external PPIs).
- ext\_protein: Includes protein information of the external PPIs. Each protein is identified by its NCBI accession number.
- ppi\_source: Stores information regarding the origin of the interactions, i.e., RIKEN\_PPI assay, DIP, BIND, or literature.
- homology: Contains the results of homology searches and the alignment of two sequences.

A

List

SORT KEY	DATA TYPE
REARRAYID(ACT)	INTERNAL

INTERNAL Interaction List									
No	Phase2	RearrayID	CloneID	LUC.	ACT	BIND	ACT	BIND	Primer Seq.
1	NP00010E03	NP00010E03	5730542014	5730542014	2	(P1)ribosome 5-phosphate isomerase ...	(P1)ribosome 5-phosphate isomerase ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
2	NP00020P24	NP00020P24	4932422H19	4932422H19	3	(P1)tumor necrosis factor recept ...	(P1)tumor necrosis factor recept ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
3	NP00020P24	NP00020P24	4932422H19	4932422H19	3	(P1)tumor necrosis factor recept ...	(P1)tumor necrosis factor recept ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
4	NP00020P24	NP00020P24	4932422H19	4932422H19	3	(P1)tumor necrosis factor recept ...	(P1)tumor necrosis factor recept ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
5	NP00030M13	NP00030M13	6430404I09	6430404I09	3	(P1)TRAF family member associate ...	(P1)TRAF family member associate ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
6	NZ00043I21	PK00013018	2010302H03	4632428011	1	(P1)LIM domain only 4	(P1)acetylcholinesterase (Ache)	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
7	NZ00049H20	R00004C19	2700028L15	0610025O12	1	(P1)leukaryotic translation initi ...	(P1)inosine 5-phosphate dehydrog ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
8	NZ00072E23	PK00022K10	2700044K04	5430433B02	1	(P1)Mus musculus recombinant qua ...	(FA)related to CG10440 PROTEIN.	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
9	PK00003N03	NZ00063I24	5730435I22	1100049E17	3	(FA)transcription factor E2a	neuronal helix-loop-helix protel ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
10	PK00003N03	R00001564	5730435I22	1100049E17	3	(FA)transcription factor E2a	(FA)inhibitor of DNA binding 2	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
11	PK00003N03	ZK00010N01	5730435I22	1100049E17	3	(FA)transcription factor E2a	scleraxis=basic helix-loop-helix ...	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
12	PK00003N03	ZK00039G21	5730435I22	2010071A14	3	(FA)transcription factor E2a		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
13	PK00003N03	ZK00039P10	5730435I22	2610027O10	3	(FA)transcription factor E2a		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
14	PK00003N03	ZK00057A24	5730435I22	1500032O24	3	(FA)transcription factor E2a		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
15	PK00004D18	PK00004D18	5730466C23	5730466C23	2	(FA)related to hypothetical prot...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
16	PK00011A10	ZK00041E18	4432411E18	2410002G23	2	(P2)TTS4		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
17	PK00012Q01	R000002113	4631426H08	0610009O09	3	(FA)similar to Cg10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
18	PK00014H08	ZK00061E19	4921520P21	2610104C23	1	(FA)hypothetical protel...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
19	PK00016D23	R000020F04	4931499E17	1500016H19	1	(P2)myosin IXA		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
20	PK00016P03	R000020F04	4931428G15	1500016H19	2	(FA)TOLLIP PROTEIN.		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
21	PK00018C09	ZK00041E18	4932423I07	2410002G23	2	(P1)coatomer protein complex, su...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
22	PK00022K10	PK00013P05	5430433B02	4921507J22	1	(FA)related to CG10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
23	PK00022K10	PK00022K10	5430433B02	5430433B02	1	(FA)related to CG10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
24	PK00022K10	R000011D24	5430433B02	1300018I18	1	(FA)related to CG10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
25	PK00022K10	ZK00047B18	5430433B02	2410043M02	1	(FA)related to CG10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
26	PK00022K10	ZK00056B20	5430433B02	2700043D08	1	(FA)related to CG10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
27	PK00022K10	ZK00059G06	5430433B02	2310034C05	1	(FA)related to CG10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
28	PK00022K10	ZK00061F04	5430433B02	2610301B19	1	(FA)related to CG10440 PROTEIN		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
29	PK00025B17	ZK00044M24	9030402K04	2010309C18	1	(FA)homolog to human splicing fa...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
30	PK00026G24	ZK00011K15	9130019P20	1300007E16	1	(FA)homolog to KIAA1085 PROTE		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
31	PK00026M03	ZK00016C01	9130003H19	2310004M18	2	(FA)basic leucine zipper transc...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
32	PK00028F20	PK00012Q01	4833436C19	4833436A13	3	(FA)keratin complex 2 basic gene...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
33	PK00028F20	PK00028F20	4833436C19	4833436A13	3	(FA)keratin complex 2 basic gene...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
34	PK00028F20	R000008M21	4833436C19	1200007G13	2	(FA)keratin complex 2 basic gene...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
35	PK00028F20	ZK00013B01	4833436C19	2310021D12	3	(FA)keratin complex 2 basic gene...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
36	PK00028F20	ZK00029C01	4833436C19	2210407G07	3	(FA)keratin complex 2 basic gene...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
37	PK00028F20	R000002113	4833436A13	0610009O09	3	(P1)keratin complex 1 acidic gen...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
38	PK00032A02	R000023J09	4930476118	1810037B05	1	(P2)KIAA0808 gene product		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
39	PK00041F07	PK00041F07	6230405F17	6230405F17	2	(P2)(BC000313) Similar to nucle...		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
40	PK00048P19	R000001F08	6530403D07	0610008A10	3	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
41	PK00048P19	R000001P11	6530403D07	0610008A10	3	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
42	PK00048P19	R000004M17	6530403D07	0610008A10	3	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
43	PK00048P19	R000011D04	6530403D07	1300018I05	3	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
44	PK00048P19	R000022P18	6530403D07	1810021D23	2	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
45	PK00048P19	R000023J09	6530403D07	1810037B05	3	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
46	PK00048P19	R000023N17	6530403D07	1810038K19	3	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
47	PK00048P19	ZK00033B17	6530403D07	2610016H24	3	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
48	PK00048P19	ZK00033J11	6530403D07	2610016H15	1	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A
49	PK00048P19	ZK00034Q16	6530403D07	2810011E04	2	(FA)unclassifiable transcript		6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A	6AAG6AGCCGCGACCATGTGCAAGGCGGAG6A

Clone ID: 5730435I22, (FANTOM 2) Sequence ID: 6938, Rearray ID: PK00003N03, DOB Accession: AZ017617, (FANTOM 1) Sequence ID: 6938, MGI Clone Accession: MGI:1694362, MGI Marker Accession: MGI:98510 (DNA seq / AA seq / MGSAC / SeqMap / Menu/Order/RIKEN/HITS/FAST)

Library information

Strain: C57BL/6J, Tissue: whole body, Age: 8 days embryo

Cluster

RTS ID: T5528, RPS ID: PA526, # of clones: 4

Representative clone: S320058N22

Other Member(s): S320002E11, 5730435I22, 9430038K03

Summary image

version: 4, last modified: Oct 4 2002

Curated Gene Name: transcription factor E2a

Curated Gene Symbol: Tcf2a

Match Status 1: complete

Match Status 2: splice variant

Evidence: BLASTN, 99%, match=2493

DB reference: MGI:1694362, MGI:98510, DOB:018280

B

### Search

CLONEID:

ID:

KEYWORD:

LUC:

CONFIDENCE:

LEVEL:

### Assay Type List

No			Assay
1	<input type="checkbox"/>	<input checked="" type="checkbox"/>	RIKEN_PPI
2	<input type="checkbox"/>	<input checked="" type="checkbox"/>	DIP
3	<input type="checkbox"/>	<input checked="" type="checkbox"/>	BIND
4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	LITERATURE

### Interaction Link

APPLLET SIZE: 900-600

View Hide 150  fix

[\*] = Homology clone.

Figure 2 (Legend on facing page)

## RESULTS

### Data Source for Mammalian Protein–Protein Interactions (PPIs)

We reported the development of a novel assay system for systematic analysis of PPIs by the mammalian two-hybrid method (Suzuki et al. 2001). We found 442 interactions among approximately 6000 cDNAs, which were derived from mouse full-length cDNA-enriched libraries. To provide the interaction data set for the FANTOM2 meeting, we developed the PPI database and viewer, a graphical tool for displaying the PPI data (Fig. 1). In addition to the interactions identified through our assays (internal PPIs), we extracted 1044 and 102 mammalian PPIs (external PPIs), respectively, from the DIP and BIND databases (Bader et al. 2001; Xenarios et al. 2002), two publicly available data sources (Fig. 1A). In addition, we used the FACTS (Functional Association/Annotation of cDNA Clones from Text/Sequence Sources; <http://facts.gsc.riken.go.jp>) system (Nagashima et al. 2003) to extract PPIs related to the FANTOM2 clones from MEDLINE abstracts of the literature; although the extraction is still in progress, this effort yielded another 424 independent external PPIs. We deposited all these variously obtained interactions in the PPI database.

### Basic Structure of the PPI Database

The RIKEN PPI database is a relational database that runs on a Sybase platform, and is based on a table structure that efficiently accommodates both internal and external PPIs (Fig. 1B). Lists of internal and external PPIs are stored in the interaction and ext\_interaction tables, respectively. Detailed information regarding the proteins is stored in separate tables—ppi\_primers and ppi\_function for proteins in internal PPIs and ext\_protein for proteins in external PPIs. For obtaining deduced protein sequences from the FANTOM2 database, the ppi\_primers table also includes the clone ID and information about the primers that were used for construction of the assay samples. The homology table enables the integration of identical proteins from the internal and external protein sets, when pairs of proteins (within threshold values of homology and ungapped identity) are stored as identical proteins (see Methods). This table also contains a descriptive result of the alignment of two proteins—information that indicates how homologous the pairs are.

### Characteristics of the PPI Viewer

The PPI Viewer (<http://fantom21.gsc.riken.go.jp/PPI/>) consists of two parts, the Interaction List and the Interaction Net-

work, and displays information from the PPI database (Fig. 2). Users can scan all the deposited internal and external PPIs with the Interaction List, which contains brief summaries regarding the interactions (Fig. 2A). IDs in the internal Interaction List are linked to the FANTOM2 main viewer (Fig. 2A inset), enabling users to obtain further information on the clone (see also Fig. 2 legend). IDs and PPI source in the external Interaction List are also linked to the NCBI clone viewer and the MEDLINE abstracts, respectively (data not shown).

Information about PPIs beyond the target protein is shown in the Interaction Network to visually characterize the protein network (Fig. 2B). Proteins of interest can be searched by either IDs or keywords (see Table 1 and Fig. 2 legend). When users input the ID of the target protein, the protein interaction network is shown as nodes (proteins) and edges (interactions). The Interaction Network also can be displayed with the annotations for the proteins in the nodes (see Table 1). Each node is linked to the clone viewers as well as the Interaction List and each edge to the interaction information, and these linked tools can be accessed by clicking the appropriate node or edge. In addition, because the link between the FANTOM2 viewer and the PPI Viewer is live in both directions, users can access the PPI Viewer from the main FANTOM2 viewer.

In the Interaction Network, the thickness of the edge correlates with the luciferase reporter activity of each internal PPI (red edges in Fig. 2B). It has been shown that the average correlation coefficient of gene expression profiles that correspond to interacting pairs is significantly higher than those that correspond to random pairs (Grigoriev 2001). The average correlation coefficients for interactions whose luciferase reporter activity are 1, 2, and 3 are 0.002, 0.03, and 0.18 respectively, clearly showing that interactions with higher reporter activity are more likely to be coexpressed. We also displayed the confidence of each external PPI using the thickness of the edge (default value 2); redundant interactions from independent reports in each external data source were assigned confidence values of 3. Including information regarding luciferase activity and confidence facilitates visualization and evaluation of the reliability of each interaction in the PPI Viewer. In addition, the color (red, blue, or green) of the edges can be selected to indicate the respective data sources. When a redundant interaction is derived from at least two data sources, the edge is shown as a white line (see Fig. 2B for example). This approach is useful because PPIs revealed by at least two different methods have greater reliability (von Mering et al. 2002). In addition, for optimal visualization of the

**Figure 2** Screen display of the PPI Viewer. The viewer consists of the Interaction List (A) and the Interaction Network (B). The Interaction List displays all the deposited PPIs sorted according to the proteins' IDs. For internal PPIs, the list can be sorted according to bait (BIND) or prey (ACT) proteins. The listed data include the interaction pairs and their annotations, supplemented with information of the primer sequences for construction of the assay samples and the luciferase activity (LUC) of the interaction (internal PPIs) or the data source linked to the MEDLINE abstracts (external PPIs). Clone IDs in the list are linked to the main FANTOM2 viewer (*right lower inset*). Some clone IDs are linked to the RIKEN 3'-end-sequence database with homology information because we have used several RIKEN clones that correspond to mouse known genes and therefore have not been fully sequenced. Accession IDs (GI number) in the external interaction list are linked to the NCBI clone viewer. These links enable users to obtain detailed information on the clone. The Interaction Network (B) comprises three sections labeled Search, Assay Type List, and Interaction Link. In the Search section, users can search for a clone to display according to any part of letters or numbers in its clone ID, accession ID, or GI number. Users also can search for proteins by KEYWORD. For example, entering "death" as a keyword yields clone 2010323J02 (programmed cell death 6). LUC, CONFIDENCE, and LEVEL are parameters with which users can display interactions having the desired luciferase activity, confidence and within the desired number of interaction steps from the target protein, respectively. In the Assay Type List section, interactions from selected data sources can be color-coded (red, green, or blue). In the Interaction Link section, the protein interaction network for the target protein (shown as a red node) is displayed as edges and nodes. The luciferase activity (LUC) and confidence of an interaction are reflected by the thickness of its edge. For internal PPIs, an arrow indicates the direction of bait protein to prey protein in the mammalian two-hybrid assay. Self-interactions are shown as double boxes, and integrated genes are shown as a single node with an asterisk. The interaction network is labeled with IDs (default) or with both IDs and annotations (by clicking the View and Hide buttons). The APPLET SIZE and length of edges (pull-downs to the right of the Hide button) can be changed. Brief instructions are listed in Table 1.

**Table 1.** Brief Instructions for Using the Controls in the PPI Viewer

Section	Control	Usage
Interaction List (see Fig. 2A)	Submit	Select the appropriate SORT KEY and its corresponding DATA TYPE, then click the <b>Submit</b> button.
	Go Top	Return to the top of the menu.
Interaction Network (see Fig. 2B) Search	ID	1. Select a clone ID, accession ID, or GI number from the upper pulldown. 2. Enter any part of the letters or numbers of the ID in the entry box left to the <b>Search</b> button, and click the <b>Search</b> button.
	KEYWORD	1. Enter a word that may be included in the annotation of the proteins, and click the <b>Search</b> button. 2. Select the gene of interest from the lower pulldown.
	LUC	Choose a luciferase activity (LUC) to display the internal experimental PPIs (RIKEN_PPI) having that activity.
	CONFIDENCE	Choose a confidence to display the external public PPIs (DIP, BIND, and LITERATURE) having that confidence.
	LEVEL	Choose a level, to display proteins within the indicated number of interaction steps from the target protein.
	Reload	Click the <b>Reload</b> button to update the interactions display after changing any parameters.
	Go Top	Return to the top of the menu.
Assay Type List		1. If necessary, change the data source and display colors by clicking the appropriate radio buttons. 2. Click the <b>Reload</b> button after making any changes.
Interaction Link	APPLET SIZE View	Choose the desired applet size from the pulldown. Click to display each protein labeled with its annotation and ID number.
	Hide 100–300	Click to display each protein labeled only with its ID number. Choose the length of edge from the pulldown.
	Fix	Check the <b>Fix</b> checkbox to lock the display.

PPI network of a target gene, the network that is displayed can be adjusted with regard to the luciferase activity (LUC), confidence of each interaction, and the number of interaction steps from the target protein.

### Analysis of the Protein Interaction Network With the PPI Viewer

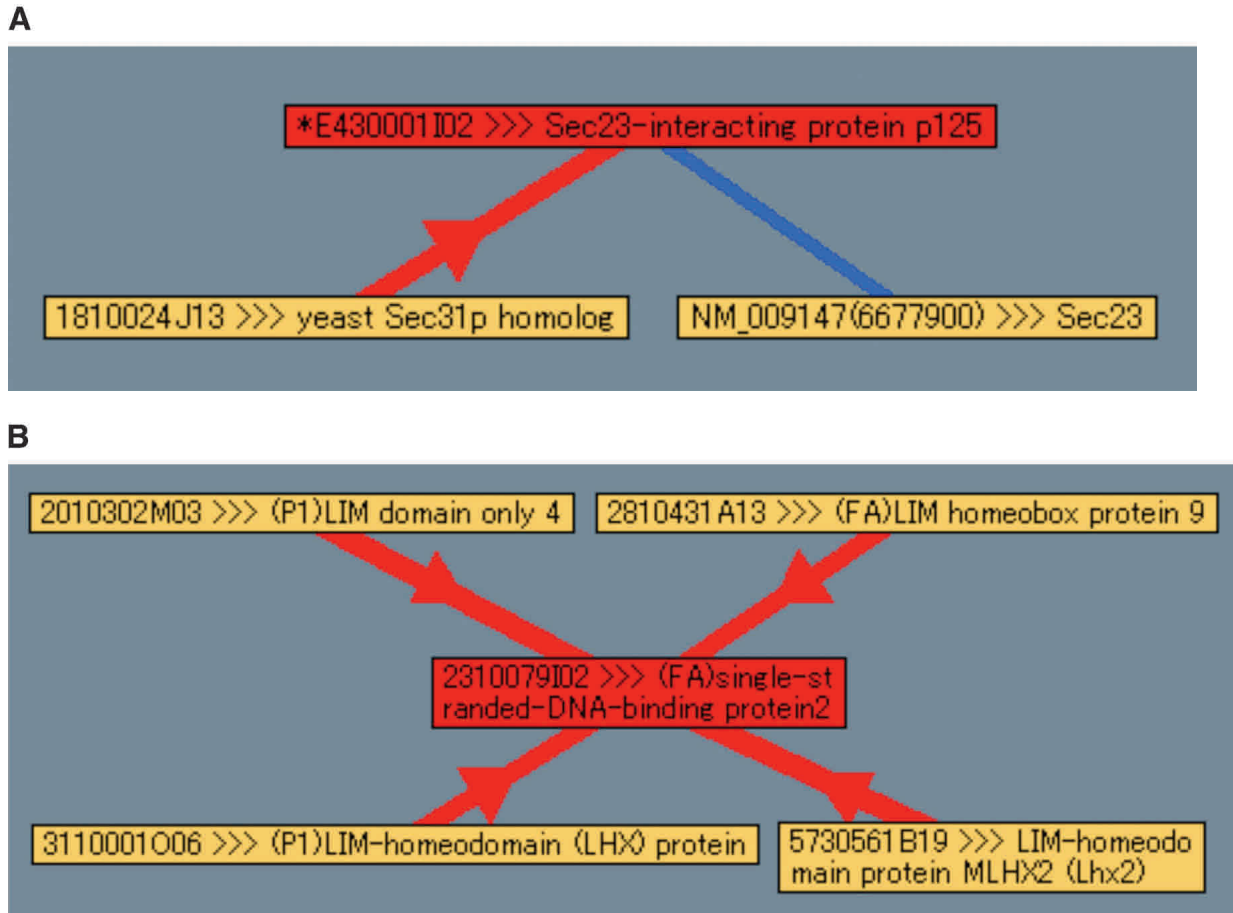
The PPI Viewer can be very useful for identifying interactions of significant biological interest in the network of interactions. Using the PPI Viewer, we have found a novel interaction partner (clone A930002B18) of TRAF2, a key adaptor molecule involved in tumor necrosis factor (TNF)-induced signaling pathway (Wajant and Scheurich 2001). We have verified the interaction experimentally and designated the partner as T2BP (TRAF2 binding protein; Kanamori et al. 2002). In addition, clone E430001I02 is a good example for significant integration of the internal and external interactions (Fig. 3A): The product of this gene interacts with the mouse homolog of the yeast Sec31 protein with a luciferase reporter activity of 3 and has been reported to be a Sec23-interacting protein (Tani et al. 1999). In yeast, Sec23 and Sec31 are both components of COPII vesicles, which are involved in the transport of proteins between the endoplasmic reticulum (ER) and the Golgi apparatus. Binding of a low-molecular-weight GTP-binding protein, Sar1p, to the ER membrane sequentially recruits Sec23/24p and Sec13/31p complexes, resulting in budding of the vesicles (Wieland and Harter 1999; Antonny and Schekman 2001). The product of E430001I02 is a mammalian protein that has no homolog in yeast. Our results suggest that E430001I02 is involved in facilitating the sequential entry of Sec23/24p and Sec13/31p into the ER membrane by physically bridging between Sec23p and Sec31p. Actually, it has

been reported that overexpression of the Sec23-interacting protein caused disorganization of the ER-Golgi intermediate compartment and Golgi apparatus (Tani et al. 1999).

Another example of the utility of our PPI database and viewer is clone 2310079I02—single-stranded DNA-binding protein 2 (SSBP2)—which is involved in 14 interactions, some of which seem to be false-positives. By displaying interactions with a luciferase reporter activity of 3 only, we found that this protein interacts with four proteins containing LIM domains, three of which are LIM homeodomain transcription factors; the remaining one is LIM domain only (LMO) protein LMO4 (Fig. 3B). The *SSBP2* gene recently was reported to be disrupted by an inversion-coupled translocation in acute myelogenous leukemia cell line ML3, and *SSBP2* is suggested to possess tumor-suppressor activity through gene dosage or other epigenetic mechanisms (Castro et al. 2002). In addition to the evidence that deregulated expression of the LMO family proteins LMO1, LMO2, and LMO4 are associated with oncogenesis, several reports suggest that other LIM domain proteins are also involved in leukemogenesis (Wu et al. 1996; Rabbitts 1998; Visvader et al. 2001; Kawamata et al. 2002). These results suggest that interaction of the LIM proteins with *SSBP2* plays an important role in the molecular mechanism of tumor-suppressor activity of *SSBP2*.

### DISCUSSION

In this report, we describe the development of a mammalian protein–protein interaction database and a PPI Viewer, which is linked to the main FANTOM2 viewer. Although the database and viewer were originally developed for internal, experimentally identified PPIs, we have incorporated PPIs extracted from the DIP and BIND public databases and the literature



**Figure 3** Analysis of interaction networks with the PPI Viewer. (A) Protein interactions involved in the formation of COPII vesicles. The components of COPII vesicles, Sec23 and Sec31, are connected by two interactions derived from internal (red edge) and external (blue edge) PPIs. (B) Interaction of LIM domain proteins with single-stranded-DNA-binding protein 2 (SSBP2, clone ID 2310079I02). Our PPI assay identified 14 interaction partners of this clone; but setting the luciferase reporter activity parameter to 3 reduced the number of displayed proteins to only 4.

(external PPIs). Connection between internal and external PPIs is not so frequently observed in the PPI Viewer at present because our experimental data are not comprehensive and because the annotation of the external PPIs derived from FACTS is still in progress. Nonetheless, the integrated information is useful for understanding mammalian protein networks in the face of limited data, as shown in Figure 3A for example, and will be more useful with increase in both the internal and external PPIs stored in the PPI database.

Because it can incorporate interactions from various resources into a single window, our PPI Viewer provides a powerful way of performing “interolog” analysis (Walhout et al. 2000a, 2000b). Interologs are homologous pairs of interacting proteins in different species and may be useful for biological analysis in two ways. First, interologs may decrease the number of false-negative interactions when the reference species is evolutionarily close to the target species, because in such cases, the interologs in the target species are likely to interact (Matthews et al. 2001). For example, a number of mammalian interologs among several transcription factors complemented interactions that have not been detected in our assay (blue and green edges in Fig. 2B). Second, information regarding interologs can be used to assess the confidence and biological significance of observed interactions (Walhout et al. 2000a,b).

The greater the evolutionary distance of the reference species, the more significant the interactions. Of the 442 internal PPIs, we have found 36 interologs when we used PPIs in the DIP database. Users having reference species information can conduct interolog analyses with the PPI Viewer.

There seem to be advantages and disadvantages in our system compared with other PPI databases, such as BIND and DIP. A unique feature of our database is that it incorporates information regarding the luciferase reporter activity of our experimental data and the confidence of the external PPIs. Furthermore, redundant interactions derived from at least two data sources are shown as white edges in the PPI Viewer. This feature enables users to elucidate the reliability of each interaction. The reliability of interactions has not been addressed satisfactorily in most public PPI databases, and it is generally accepted that available PPI information includes many false-positive interactions that may yield incorrect biological conclusions (Legrain et al. 2001). In this sense, it may be more valuable to apply the Interaction Generality measure, a means for computationally assessing the reliability of PPIs obtained in biological experiments (Saito et al. 2002), to the PPI Viewer. Another advantage of our system is that it permits the display of a large variety of information to find biologically relevant interactions. Users can efficiently visualize the

protein network around the protein of interest using several controls such as LUC, confidence, level, and color choice, while there are few publicly available PPI databases whose viewers suitably display such information. We showed a few example interactions whose biological significance was successfully suggested through the use of the PPI Viewer (Fig. 3). These interactions are essential in uncovering molecular mechanisms of the biological processes. However, it should be stressed that prior to further analysis interactions of interest should be confirmed by other methods, such as coimmunoprecipitation or in vitro binding assays, because the PPI data include many false-positive interactions and PPIs revealed by at least two different methods have greater reliability (von Mering et al. 2002). Finally, our database may be unique by focusing on mammalian PPI, which will be more useful both in future higher functional annotations and the proteomics studies in mammals. On the other hand, a disadvantage of our system compared with other databases is that operation of the viewer may be complicated because of the rich information displayed. Further, the search function to find proteins and/or interactions of interest may be less powerful than the display function, although part of this disadvantage could be compensated by a variety of search terms in the main FANTOM2 viewer linked to the PPI Viewer.

During the FANTOM2 meeting, the annotation of each gene was based on information regarding homology to known genes and/or motifs, and information obtained through use of the PPI database was not included in the annotation. Nonetheless, having the PPI database linked to the FANTOM2 database will be useful for detecting protein interactions of genes of interest, because PPI analysis is one of the most important approaches for uncovering the functions of uncharacterized genes. The goal of our PPI database is to construct a mammalian protein database by integrating all available genome-wide interaction data sets, where reliability of each interaction should be evaluated as described. We believe that such a database will be definitely essential both in future higher functional annotations and the proteomics studies.

## METHODS

### Data Source of the PPIs

The experimentally examined PPIs obtained from our high-throughput assay system were deposited in the PPI database. Of the 442 interactions we obtained, 145 were reported previously, and the remaining 297 interactions are newly published. In addition, mammalian PPIs were extracted from the publicly available PPI databases, DIP (<http://dip.doe-mbi.ucla.edu>) and BIND (<http://www.biond.org/>), and from MEDLINE abstracts by using the FACTS system (Nagashima et al. 2003). To reduce redundancy, we removed the 145 experimentally produced PPIs that we previously deposited in the BIND database from those we extracted from BIND. Further, we removed redundancy of the interactions with identical PubMed identifiers (PMID) from the publicly available PPIs. In the case of PPIs extracted by the FACTS system, 424 PPIs were selected from the originally extracted 696 PPIs. These extracted PPIs were integrated with our experimentally identified interactions; we then used the BLAST and FASTA programs to identify identical proteins. When the E-value was lower than  $1.0 \times 10^{-30}$  and the ungapped identity was more than 90% between two proteins, they were determined to be identical and were placed in the same node in the PPI Viewer. We similarly integrated the redundant proteins among the external PPIs and stored the homology information in the PPI

database. The interactions deposited in the database are available by request.

### PPI Viewer

Brief instructions for using each control in the PPI Viewer are shown in Table 1. When the luciferase reporter activity of an internal PPI is more than 3-, 5-, and 10-fold greater than the mean background in the bait- or prey-only samples, the activity is defined as 1, 2, and 3, respectively. The confidence of interactions from the publicly available databases and curated literature is defined as 2, and that of redundant interactions from independent reports in each external data set is assigned to be 3.

## ACKNOWLEDGMENTS

We thank E. Hiratsu, H. Shige and S. Katayama for their technical assistance, D. Silva, L. Socha, and I. Kurochkin for annotations, and K. Yoshida, H. Kusuda, and H. Murakami for database construction. We also thank the members of the Laboratory for Genome Exploration Research Group. This study has been supported by a Research Grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science, and Technology of the Japanese Government to Y.H., by CREST of JST (Japan Science and Technology Corporation) to Y.H. and by Presidential Research Grant for Intersystem Collaboration of RIKEN GSC to H.S.

## REFERENCES

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Antonny, B. and Schekman, R. 2001. ER export: Public transportation by the COPII coach. *Curr. Opin. Cell. Biol.* **13**: 438–443.
- Bader, G.D., Donaldson, I., Wolting, C., Ouellette, B.F., Pawson, T., and Hogue, C.W. 2001. BIND—The biomolecular interaction network database. *Nucleic Acids Res.* **29**: 242–245.
- Castro, P., Liang, H., Liang, J.C., and Nagarajan, L. 2002. A novel, evolutionarily conserved gene family with putative sequence-specific single-stranded DNA-binding activity. *Genomics* **80**: 78–85.
- The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282**: 2012–2018.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. Life with 6000 genes. *Science* **274**: 546, 563–567.
- Grigoriev, A. 2001. A relationship between gene expression and protein interactions on the proteome scale: Analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **29**: 3513–3519.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.* **98**: 4569–4574.
- Kanamori, M., Suzuki, H., Saito, R., Muramatsu, M., and Hayashizaki, Y. 2002. T2BP, a novel TRAF2 binding protein, can activate NF- $\kappa$ B and AP-1 without TNF stimulation. *Biochem. Biophys. Res. Commun.* **290**: 1108–1113.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* **409**: 685–690.
- Kawamata, N., Sakajiri, S., Sugimoto, K.J., Isobe, Y., Kobayashi, H., and Oshimi, K. 2002. A novel chromosomal translocation t(1;14)(q25;q32) in pre-B acute lymphoblastic leukemia involves the LIM homeodomain protein gene, *Lhx4*. *Oncogene* **21**: 4983–4991.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Legrain, P., Wojcik, J., and Gauthier, J.M. 2001. Protein–protein

- interaction maps: A lead towards cellular functions. *Trends Genet.* **17**: 346–352.
- Matthews, L.R., Vaglio, P., Reboul, J., Ge, H., Davis, B.P., Garrels, J., Vincent, S., and Vidal, M. 2001. Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or “interologs”. *Genome Res.* **11**: 2120–2126.
- Nagashima, T., Silva, D.G., Petrovsky, N., Socha, L.A., Suzuki, H., Saito, R., Kasukawa, T., Kurochkin, I.V., Konagaya, A., and Schönbach, C. 2003. Inferring higher functional information for RIKEN mouse full-length cDNA clones with FACTS. *Genome Res.* **13**: (this issue).
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Oliver, S. 2000. Guilt-by-association goes global. *Nature* **403**: 601–603.
- Pawson, T. and Nash, P. 2000. Protein–protein interactions define specificity in signal transduction. *Genes & Dev.* **14**: 1027–1047.
- Rabbitts, T.H. 1998. LMO T-cell translocation oncogenes typify genes activated by chromosomal translocations that alter transcription and developmental processes. *Genes & Dev.* **12**: 2651–2657.
- Saito, R., Suzuki, H., and Hayashizaki, Y. 2002. Interaction generality, a measurement to assess the reliability of a protein–protein interaction. *Nucleic Acids Res.* **30**: 1163–1168.
- Suzuki, H., Fukunishi, Y., Kagawa, I., Saito, R., Oda, H., Endo, T., Kondo, S., Bono, H., Okazaki, Y., and Hayashizaki, Y. 2001. Protein–Protein interaction panel using mouse full-length cDNAs. *Genome Res.* **11**: 1758–1765.
- Tani, K., Mizoguchi, T., Iwamatsu, A., Hatsuzawa, K., and Tagaya, M. 1999. p125 is a novel mammalian Sec23p-interacting protein with structural similarity to phospholipid-modifying proteins. *J. Biol. Chem.* **274**: 20505–20512.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. 2000. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**: 623–627.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Visvader, J.E., Venter, D., Hahm, K., Santamaria, M., Sum, E.Y., O’Reilly, L., White, D., Williams, R., Armes, J., and Lindeman, G.J. 2001. The LIM domain gene LMO4 inhibits differentiation of mammary epithelial cells in vitro and is overexpressed in breast cancer. *Proc. Natl. Acad. Sci.* **98**: 14452–14457.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. 2002. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**: 399–403.
- Wajant, H. and Scheurich, P. 2001. Tumor necrosis factor receptor-associated factor (TRAF) 2 and its role in TNF signaling. *Int. J. Biochem. Cell Biol.* **33**: 19–32.
- Walhout, A.J., Boulton, S.J., and Vidal, M. 2000a. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. *Yeast* **17**: 88–94.
- Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. 2000b. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**: 116–122.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Wieland, F. and Harter, C. 1999. Mechanisms of vesicle formation: Insights from the COP system. *Curr. Opin. Cell. Biol.* **11**: 440–446.
- Wu, H.K., Heng, H.H., Siderovski, D.P., Dong, W.F., Okuno, Y., Shi, X.M., Tsui, L.C., and Minden, M.D. 1996. Identification of a human LIM-Hox gene, hLH-2, aberrantly expressed in chronic myelogenous leukaemia and located on 9q33–34.1. *Oncogene* **12**: 1205–1212.
- Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M., and Eisenberg, D. 2002. DIP, the Database of Interacting Proteins: A research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**: 303–305.

## WEB SITE REFERENCES

- <http://fantom21.gsc.riken.go.jp/PPI/>; PPI Viewer.  
<http://facts.gsc.riken.go.jp/>; FACTS system.  
<http://dip.doe-mbi.ucla.edu/>; DIP database.  
<http://www.biddb.org/>; BIND database.

Received November 5, 2002; accepted in revised form March 3, 2003.