Protein–Protein Interaction Panel Using Mouse Full-Length cDNAs
Harukazu Suzuki, Yoshifumi Fukunishi, Ikuko Kagawa, et al.

Genome Res. 2001 11: 1758-1765
Access the most recent version at doi:10.1101/gr.180101

Supplemental Material
http://genome.cshlp.org/content/suppl/2001/09/26/GR-1801RR.DC1.html

References
Article cited in:
http://genome.cshlp.org/content/11/10/1758.full.html#related-urls

Creative Commons License
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at http://creativecommons.org/licenses/by-nc/3.0/.

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:
http://genome.cshlp.org/subscriptions

© 2001, Published by Cold Spring Harbor Laboratory Press
Methods

Protein–Protein Interaction Panel Using Mouse Full-Length cDNAs

Harukazu Suzuki,1 Yoshifumi Fukunishi,1 Ikuko Kagawa,1 Rintaro Saito,1
Hiroshi Oda,1 Toshinori Endo,1 Shinji Kondo,1 Hidemasa Bono,1 Yasushi Okazaki,1
and Yoshihide Hayashizaki1,3

1Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan;
Genome Science Laboratory, RIKEN Tsukuba Institute, Tsukuba 305-0007, Japan; 3Department of Medicine, Tsukuba
University, Tsukuba 305-0006, Japan.

We have developed a novel assay system for systematic analysis of protein–protein interactions (PPIs) that is
characteristic of a PCR-mediated rapid sample preparation and a high-throughput assay system based on the
mamalian two-hybrid method. Using gene-specific primers, we successfully constructed the assay samples by
two rounds of PCR with up to 3.6 kb from the first-round PCR fragments. In the assay system, we designed all
the steps to be performed by adding only samples, reagents, and cells into 384-well assay plates using two types
of semiautomatic multiple dispensers. The system enabled us examine more than 20,000 assay wells per day.
We detected 145 interactions in our pilot study using 3500 samples derived from mouse full-length enriched
cDNAs. Analysis of the interaction data showed both several significant interaction clusters and predicted
functions of a few uncharacterized proteins. In combination with our comprehensive mouse full-length cDNA
clone bank covering a large part of the whole genes, our high-throughput assay system will discover many
interactions to facilitate understanding of the function of uncharacterized proteins and the molecular
mechanism of crucial biological processes, and also enable completion of a rough draft of the entire PPI panel in
certain cell types or tissues of mouse within a short time.

RESULTS

PCR-Mediated Rapid Sample Preparation

We have prepared the samples for PPI assay by PCR without
any cloning steps (Fig. 1). First, we have synthesized each
gene-specific forward primer possessing an 18-base common
sequence followed by the gene-specific sequence (Fig. 1A). We
constructed the samples by two rounds of PCR (Fig. 1B and
Methods). In the first PCR, we amplified each cDNA using the
gene-specific primer and the M13 universal primer to make
PCR-Mediated Rapid Sample Preparation and a high-throughput PPI as-
say system, that allowed us to obtain interaction data very
rapidly.

We detected 145 interactions in our pilot study using 3500 samples derived from mouse full-length enriched
cDNAs. Analysis of the interaction data showed both several significant interaction clusters and predicted
functions of a few uncharacterized proteins. In combination with our comprehensive mouse full-length cDNA
clone bank covering a large part of the whole genes, our high-throughput assay system will discover many
interactions to facilitate understanding of the function of uncharacterized proteins and the molecular
mechanism of crucial biological processes, and also enable completion of a rough draft of the entire PPI panel in
certain cell types or tissues of mouse within a short time.

As in the case of Saccharomyces cerevisiae (budding yeast), Caenorhabditis elegans, Drosophila, and Arabidopsis (Mewes et al. 1997; The C. elegans Sequencing Consortium 1998; Adams et al. 2000; The Arabidopsis Genome Initiative 2000), large-scale genome sequencing and cDNA libraries brought us a rough
draft of whole genes in higher organisms such as human and mouse, wherein many of the genes were novel ones of
unknown function (International Human Genome Sequencing Consortium 2001; The RIKEN Genome Exploration Research
Group Phase II Team and the FANTOM Consortium 2001; Venter et al. 2001). To uncover the function of each gene,
the matrix method has been performed (Uetz et al. 2000; Ito et al. 2001) and a
systematic analysis of PPIs in C. elegans has been started (Wal-
hout et al. 2000); both of these have been performed by the
matrix method using the interaction mating-mediated yeast
two-hybrid system (Colas and Brent 1998). In organisms with
several ten-thousands of genes, however, it seems less easy to
establish entire PPI panels (matrix) because the total number
of examinations in human or mouse is estimated to be far
larger than those for budding yeast or C. elegans. Here we
report two key developments to address this difficulty, PCR-
mediated sample preparation and a high-throughput PPI as-
say system, that allowed us to obtain interaction data very
rapidly.

Results

PCR-Mediated Rapid Sample Preparation

We have prepared the samples for PPI assay by PCR without
any cloning steps (Fig. 1). First, we have synthesized each
gene-specific forward primer possessing an 18-base common
sequence followed by the gene-specific sequence (Fig. 1A). We
constructed the samples by two rounds of PCR (Fig. 1B and
Methods). In the first PCR, we amplified each cDNA using the
gene-specific primer and the M13 universal primer to make
the protein coding sequence (CDS) fragment with the com-
mon 18-base sequence at the 5’ terminus. We also amplified
DNA fragments for human cytomegalovirus (CMV) immedi-
ate early promoter followed by the Gal4 DNA-binding do-
main or herpes virus VP16 transactivation activation domain,
in which both DNA fragments have a common sequence at the
3’ termini. We used the common sequence as a margin to
connect the first PCR products with the Gal4 or VP16 frag-
ments. In the second PCR (the overlapping PCR [Higuchi et al. 1988; Ho et al. 1989]) we amplified the first PCR products and the Gal4 or VP16 fragments to make the PCR products, in which protein derived from each cDNA was designed to be expressed as fusion proteins with the Gal4 DNA-binding domain or the VP16 transcriptional activation domain and was under the control of the CMV promoter (BIND and ACT samples, respectively). We successfully constructed assay
samples with up to 3.6 kb from the first-round PCR fragments with a success rate of more than 95% (Fig. 1C). Further, we transfected the BIND samples into CHO-K1 cells and detected the expressed fusion proteins with almost reasonable size by Western blotting analysis using a monoclonal antibody against the Gal4 DNA-binding domain (Fig. 1D). To confirm that the samples were applicable to the PPI assay, we applied the PCR-mediated positive control samples, BIND-inhibitor of differentiation (BIND-ID), and ACT-myogenic regulatory protein (ACT-MyoD), to the standard mammalian two-hybrid method according to the manufacturer’s method. We observed significant positive signals in the assay; the activity of the luciferase reporter gene (count) was 54,009 whereas those from the assay using either BIND-ID or ACT-MyoD were 3454 and 974, respectively.

High-Throughput Assay System

We established an assay system using 384-well assay plates that is based on the mammalian two-hybrid method (Dang et al. 1991; Fearon et al. 1992). There is a difficulty for efficient assay in the standard mammalian two-hybrid method because the cultured cells must be prepared in each well of tissue culture dishes before the assay. To facilitate high throughput, we designed all the assay steps to be performed by adding only samples, reagents, and cells into the assay plates using two types of semiautomatic multiple dispensers and computerized sample tracking (see Methods). The ACT samples were prepared in 96-well plates by mixing them with the culture medium supplemented with a plasmid for the luciferase gene. The BIND samples were prepared in the same way. All combinations consisting of 96 ACT samples and 4 BIND samples were prepared in each 384-well assay plate. The ACT and BIND samples were added into the plates by 96-channel dispensers and 8-channel workstations, respectively. After adding the ACT and BIND samples, the transfection reagent and suspended CHO-K1 cells were added into each well in this order by 384-channel dispensers and were suspended by pipetting several times. If the expressed proteins interact, transcription of the luciferase gene is activated (Fig. 1B). We measured luciferase activity by the detection reagent after incubation of the assay samples in a CO2 incubator overnight.

Figure 2 shows the sequence of procedures of the assay system. Because some BIND samples (Gal4 fusion proteins) increase luciferase gene expression without interaction with any ACT samples, we first performed a preassay to remove the BIND samples with high background. We found that −2% of BIND samples had high background values, and these samples were excluded from further analysis. In the first assay we used a mixture of BIND samples and ACT samples to increase efficiency. In a test experiment shown in Table 1, the interaction was detectable using 1/16 BIND-ID and 1/4 or 1/8 ACT-MyoD, although the luciferase activity decreased drastically depending on the dilution of positive control samples. To examine whether these dilution ratios were applicable to the actual sample mixture, the other test experiment was performed using 16-mixture (16-mix) of BIND samples and 6-mix of ACT mixture samples in which BIND-ID and ACT-MyoD were involved in one of the BIND and ACT samples, respectively. The interaction between the positive control samples was significantly detected; the luciferase activity of the positive combination was 2032, where mean and standard deviation (SD) for the BIND sample with BIND-ID were 800 and 419 (n = 24), respectively, and those for the ACT sample with ACT-MyoD were 1296 and 349 (n = 12), respectively. Similar results were also obtained using another positive control pair, BIND-SV40 large T-antigen and ACT-p53 (data not shown). Thus, we performed the first assay using a 16-mix of BIND samples and a 6-mix of ACT samples at one time in a well of the assay plates.

After measurement of the luciferase activity in the first assay, we analyzed these results statistically. Positive candidates were selected as those values revealing SD values both more than 3.0 for each 16-mix BIND sample and more than 2.0 for each 6-mix ACT sample (see Methods). The positive control wells containing BIND-ID and ACT-MyoD were detected as positives in 93% of the first assays in this condition. The mean SD values of the positive control wells were 7.14 and 4.11 for the BIND and ACT samples, respectively. Although the first assay is a rate-limiting step in our assay flow, our system enabled us to evaluate more than 20,000 assay wells per day (60 384-well assay plates per day). In the second assay, we examined the positive combinations in the first assay by the combination of a single BIND sample and a 6-mix ACT sample, and a 16-mix BIND sample and a single ACT sample, to identify the interaction pair candidates. Finally, we examined the positive assay with a single BIND sample and a single ACT sample to confirm the interaction pairs.

Results of the Pilot Study

In the pilot study, examinations were performed using 3500 BIND samples and 3400 ACT samples, which were derived from mouse full-length enriched cDNAs with sequence data available for the primer design (The RIKEN Genome Explora-
tion Research Group Phase II Team and the FANTOM Consortium 2001; http://genome.ssc.riken.go.jp/). As shown in Table 2, we detected 145 PPIs of which 27 were self-interactions (all interaction data are available at http://www.genome.org/ and http://genome.ssc.riken.go.jp/). Judging from the annotation for each interacting gene, the number of interactions between genes of known function, between genes of known and unknown function, and between genes of unknown function were 77, 39, and 29, respectively.

We analyzed the network of interactions by our software called PPI network viewer to confirm whether the detected interactions were biologically significant. We found several predictable protein network clusters. For example, we detected a quite large protein cluster with a basic helix-loop-helix (bHLH) motif (PPI numbers 26, 27, 53, 65, 66, 72, 73, 74, 119, and 141 in Table 2). It is well known that this motif appears often in some transcription factors and their regulatory proteins and is responsible for heterodimerization of the bHLH-containing proteins (Norton et al. 1998). We compared the interaction data of this cluster with those deposited in the ProNet PPI database (http://pronet.doubletwist.com/). We found four out of six interactions of the ProNet data (MTF2A-Id1B, MTF2A-Id2, MTF2A-Id3, and Pan-2-Id2) were detected in our study. The remaining two interactions were LYL1-MTF2A and LYL1-Pan-2, which have been established by the yeast two-hybrid method and/or immunoprecipitation (Miyamoto et al. 1996). We have also detected interactions among keratin family proteins and keratin-related protein (PPI numbers 7, 8, 9, 41, 85, and 91 in Table 2).

Recently, several groups tried to predict the function of uncharacterized proteins using yeast global PPI data (Schwikowski et al. 2000; Uetz et al. 2000; Walhout et al. 2000; Ito et al. 2001). The main concept behind such analyses is that known proteins interacting with uncharacterized proteins provide a valuable clue to the function of the uncharacterized proteins because many proteins play a role in the network of cellular biological processes by associating with other related proteins (guilt-by-association [Oliver 2000]). Considering the confidence of the functional prediction using the two-hybrid analysis data, we searched our pilot experiment data for uncharacterized proteins that interact with more than two known proteins of similar function. We found that Clone 2810048F05 is a good example of this situation; it interacts with the third largest subunit of RNA polymerase II (PPI number 133) and the β-subunit of transcription factor IIE (PPI number 62), indicating that 2810048F05 is involved in the transcription process. Another example is the CBFA2T1/MTG8 gene (3110001I23). CBFA2T1/MTG8 is a component of the nuclear receptor corepressor (Lutterbach et al. 1998; Wang et al. 1998). Our result showing CBFA2T1/MTG8 interacting with two isoforms of Id proteins (PPI numbers 66 and 72) indicates that Ids may play a role in the regulation of the nuclear receptor corepressor.

**DISCUSSION**

To construct assay samples efficiently from each cDNA, we have synthesized each gene-specific forward primer. The primer has a common sequence and we used it as the margin to connect two DNA fragments by overlapping PCR for the fusion proteins. The gene-specific primers are also useful for simple expression of the proteins both in vitro and in vivo, because the common sequence consists of the Shine-Dalgarno sequence for a prokaryotic ribosome-binding site, GAAGGA, and the Kozak consensus sequence for a eukaryotic translation initiation site, GCCGCACCAGT (Fig. 1A). We have expressed the proteins successfully in vitro using constructs in which the T7 promoter sequence was added upstream of the first-round PCR products by extension PCR (data not shown).

The main characteristic of our strategy for PPI search is quickness through all the steps from sample preparation through assay. The PCR-mediated sample could be prepared very quickly within 1 d because it is not necessary to recover the clones as plasmids from bacteria. The assay is based on the mammalian two-hybrid method in which the assay could be performed using transiently expressed proteins. Therefore, the PCR-mediated samples were applied directly to the assay. Further, incubation time necessary for the assay is only 20 h, which is faster than that for the yeast two-hybrid method, which takes at least several days. The results of the assay are quantified by measuring luciferase activity. These values are also suitable for quick judgment of positives (or positive candidates). In addition, the values for interaction pairs will be useful for evaluating the strength of each interaction because the luciferase activity roughly parallels the strength of interaction.

There are advantages and disadvantages in the PCR-mediated sample preparation. In addition to the rapid sample preparation described above, the direct use of the PCR products as samples could minimize the problem of mutations incorporated into the samples by thermostable DNA polymerase. Actually, we confirmed the expressed fusion proteins of reasonable size using the PCR-mediated samples (Fig. 1D). We showed successful construction of assay samples with up to 3.6 kb from the first-round PCR fragments (Fig. 1C). However, the concentration of assay samples has a tendency to decrease in first-round PCR fragments of >3 kb. Because the intensity of luciferase activity was affected strongly by the concentration of the assay samples, as shown in Table 1, it is plausible that the less-amplified assay samples may not be screened effectively.

Generally, the two-hybrid method is not completely reliable because the results usually contain many false positives. Actually, very little overlap of the yeast PPI data from two independent research groups using the yeast two-hybrid method was observed, even though part of the reason may be explained by the different conditions of the experiments such as the sample construction procedure, the selection strategy, and the depth of the examinations (Uetz et al. 2000; Ito et al. 2001). The false-positive interactions in the two-hybrid method could be classified into two types: less-reproducible interactions and physiologically less-significant interactions. Our assay strategy may have several advantages to minimize the false positives. First, because we have three examinations (the first, second, and final assays) before determining each interaction pair, such repeated assays should be expected to decrease less-reproducible interactions. This assertion was confirmed by the results of reexamination of some of the assays: almost 80% of the interaction pairs could be found again in the reexamination (data not shown). Second, we applied the results of the first assay to the statistical analysis because the luciferase activity was quantitative. The statistical selection of the positive candidates must be effective in reducing the false positives because it enabled exclusion of most of the pseudopositive values caused by high background in some BIND or ACT samples. Finally, because the assay system is performed in mammalian cells, expressed proteins are more...
<table>
<thead>
<tr>
<th>PPI No.</th>
<th>BIND clone ID</th>
<th>ACT clone ID</th>
<th>annotation</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0610006G13 (J04716) ferritin light chain [M]</td>
<td>1500001P18 (J03941) ferritin heavy chain [M]</td>
<td>(J03941) ferritin heavy chain [M]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0610007F07 (K01515) hypoxantheme phosphoribosyltransferase [M]</td>
<td>0610007F07 (K01515) hypoxantheme phosphoribosyltransferase [M]</td>
<td>(K01515) hypoxantheme phosphoribosyltransferase [M]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0610007L03 (U46837) SRB7 [H]</td>
<td>2010309C18 (U94662) TFG protein [M]</td>
<td>(U94662) TFG protein [M]</td>
<td>(U94662) TFG protein [M]</td>
</tr>
<tr>
<td>5</td>
<td>0610007L13 (X13752) delta-aminolevulinate dehydratase (AA 1-330) [M]</td>
<td>0610007L13 (X13752) delta-aminolevulinate dehydratase (AA 1-330) [M]</td>
<td>(X13752) delta-aminolevulinate dehydratase (AA 1-330) [M]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0610009D16 (D14811) antioxidant protein 1 [M]</td>
<td>0610009D16 (D14811) antioxidant protein 1 [M]</td>
<td>(D14811) antioxidant protein 1 [M]</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0610009O09 (M13955) type II mesothelial keratin K7 [H]</td>
<td>0610009O08 (U24223) alpha-complex protein 1 [H]</td>
<td>(U24223) alpha-complex protein 1 [H]</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0610009O09 (M13955) type II mesothelial keratin K7 [H]</td>
<td>2010309C18 (U94662) TFG protein [M]</td>
<td>(U94662) TFG protein [M]</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0610009O09 (M13955) type II mesothelial keratin K7 [H]</td>
<td>4610009O09 (M13955) type II mesothelial keratin K7 [H]</td>
<td>(M13955) type II mesothelial keratin K7 [H]</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0610009D16 (D14811) antioxidant protein 1 [M]</td>
<td>0610009D16 (D14811) antioxidant protein 1 [M]</td>
<td>(D14811) antioxidant protein 1 [M]</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0610009O09 (M13955) type II mesothelial keratin K7 [H]</td>
<td>0610009O08 (U24223) alpha-complex protein 1 [H]</td>
<td>(U24223) alpha-complex protein 1 [H]</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0610009D16 (D14811) antioxidant protein 1 [M]</td>
<td>0610009D16 (D14811) antioxidant protein 1 [M]</td>
<td>(D14811) antioxidant protein 1 [M]</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0610009O09 (M13955) type II mesothelial keratin K7 [H]</td>
<td>0610009O08 (U24223) alpha-complex protein 1 [H]</td>
<td>(U24223) alpha-complex protein 1 [H]</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0610009O09 (M13955) type II mesothelial keratin K7 [H]</td>
<td>2010309C18 (U94662) TFG protein [M]</td>
<td>(U94662) TFG protein [M]</td>
<td></td>
</tr>
</tbody>
</table>

Suzuki et al. 1762 Genome Research www.genome.org
likely to be in their native conformation with appropriate posttranslational modifications (Fagan et al. 1994). Thus, there may be fewer false interactions in the mammalian two-hybrid system than would occur in the yeast two-hybrid system. Nonetheless, this does not mean that all the interactions detected in our assay are physiologically significant. It is clear that we should examine the significance of each interaction by additional experiments such as immunoprecipitation.

We found 145 interactions using 3500 BIND samples and 3400 ACT samples. When we suppose that all possible combinations (1.2 × 10^7 combinations) were tested significantly, this means that interactions were detected approximately once every 80,000 combinations. Although it has not been well-known how often PPIs are found in mammals, the frequency seems to be relatively lower compared with the value of once per 30,000 combinations obtained by systematic analysis of PPIs in yeast (Uetz et al. 2000). The major reason is that we may overlook PPIs with moderate or weak affinity because we used the positive control samples with high affinity. The other reason for the observed low frequency of interactions without a chance to meet, the appropriate selection of genes involved in the process of interaction, and the complexity level handling at one time. Thus, it should be necessary to reduce the complexity level further in our assay system for the detailed analysis of PPI with various interaction affinities. The other reason for the observed low frequency of interactions with our PPI systems is that as a first attempt we did not examine the assay without selecting genes coexpressed together. Unlike yeast, higher organisms consist of many types of cells with various gene expression spectrums, in which each cell may express only some of the genes. Because proteins could not compose biologically significant interactions, the appropriate selection of genes for PPI search may enhance PPI discovery more efficiently. Actually, genes involving similar functions or the same pathways, some of which are likely to be associated with other, could be clustered together according to their expression profiles (Eisen et al. 1998; Miki et al. 2001).

Viewing PPIs globally through a network is valuable because it increases the confidence level for each interaction.

Table 2. (Continued)

<table>
<thead>
<tr>
<th>PPI No.</th>
<th>BIND clone ID</th>
<th>Annotation</th>
<th>ACT clone ID</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1500012F11</td>
<td>(AF074723) RNA polymerase transcriptional regulation mediator [H]</td>
<td>1200015A19</td>
<td>novel protein</td>
</tr>
<tr>
<td>51</td>
<td>1500016H19</td>
<td>(AB001740) p27 [H]</td>
<td>1500016H19</td>
<td>(AB001740) p27 [H]</td>
</tr>
<tr>
<td>52</td>
<td>1500016H19</td>
<td>(AB001740) p27 [H]</td>
<td>2700002C15</td>
<td>(Z31399) CC20 protein elu chain [M]</td>
</tr>
<tr>
<td>53</td>
<td>1500037O10</td>
<td>(U28068) neurogenic differentiation factor (neuroD) [M]</td>
<td>3300001C01</td>
<td>(U16321) MIF-2A protein [M]</td>
</tr>
<tr>
<td>54</td>
<td>1500040C19</td>
<td>(X93357) SYT [M]</td>
<td>0610043D20</td>
<td>(Z85979) histone H3.3A [M]</td>
</tr>
<tr>
<td>55</td>
<td>1600014M03</td>
<td>(U06753) acidic calponin [R]</td>
<td>2310007K12</td>
<td>(D37837) 65-kDa macrophage protein [M]</td>
</tr>
<tr>
<td>56</td>
<td>1700003P11</td>
<td>(AF151883) CGI-1-25 protein [H]</td>
<td>1110001O11</td>
<td>(AF119676) RAB25 [M]</td>
</tr>
<tr>
<td>57</td>
<td>1700022I15</td>
<td>(X75959) polyA binding protein [M]</td>
<td>1700021C22</td>
<td>(NM_015157) synaptonemal complex protein 3 [M]</td>
</tr>
<tr>
<td>58</td>
<td>1700025D04</td>
<td>(U81100) TRAF4 associated factor 1 [H] 2e-74</td>
<td>2310047M10</td>
<td>novel protein</td>
</tr>
<tr>
<td>59</td>
<td>1700026B03</td>
<td>(L33252) GTPase (Ran) [M]</td>
<td>2400006H24</td>
<td>(Z49574) ORF Y1074w [Sc] 1e-10</td>
</tr>
<tr>
<td>60</td>
<td>1700029P02</td>
<td>(AF019926) protein kinase [M]</td>
<td>2010309C18</td>
<td>(U94662) TFG protein [M]</td>
</tr>
<tr>
<td>61</td>
<td>1810041B23</td>
<td>(X63469) TFIIE-beta [H]</td>
<td>2810048P05</td>
<td>novel protein</td>
</tr>
<tr>
<td>62</td>
<td>1810043E06</td>
<td>(M60523) Id3 protein [M]</td>
<td>3300001C01</td>
<td>(U16321) MIF-2A protein [M]</td>
</tr>
<tr>
<td>63</td>
<td>1810043E06</td>
<td>(M60523) Id3 protein [M]</td>
<td>3100001C01</td>
<td>(U16321) MIF-2A protein [M]</td>
</tr>
<tr>
<td>64</td>
<td>1810043E06</td>
<td>(M60523) Id3 protein [M]</td>
<td>3100001C01</td>
<td>(U16321) MIF-2A protein [M]</td>
</tr>
<tr>
<td>65</td>
<td>1810043E06</td>
<td>(M60523) Id3 protein [M]</td>
<td>3100001C01</td>
<td>(U16321) MIF-2A protein [M]</td>
</tr>
<tr>
<td>72</td>
<td>2010016A14</td>
<td>(U43884) Id18 protein [M]</td>
<td>5730435I22</td>
<td>(X54594) Pan-2 [R]</td>
</tr>
<tr>
<td>73</td>
<td>2010016A14</td>
<td>(U43884) Id18 protein [M]</td>
<td>5730435I22</td>
<td>(X54594) Pan-2 [R]</td>
</tr>
<tr>
<td>74</td>
<td>2010016A14</td>
<td>(U43884) Id18 protein [M]</td>
<td>5730435I22</td>
<td>(X54594) Pan-2 [R]</td>
</tr>
<tr>
<td>75</td>
<td>2210407C07</td>
<td>(AB013608) cytokeratin 17 [M]</td>
<td>0610009009</td>
<td>(M13953) type II mesothelial keratin K7 [H]</td>
</tr>
<tr>
<td>76</td>
<td>2310015J09</td>
<td>(M27734) keratin type I [M]</td>
<td>3300001P10</td>
<td>(M19723) keratin K5 [H]</td>
</tr>
<tr>
<td>77</td>
<td>2610027O10</td>
<td>(A129753) basic helix-loop-helix factor Cor1 [M]</td>
<td>5730435I22</td>
<td>(X54594) Pan-2 [R]</td>
</tr>
<tr>
<td>78</td>
<td>2900002E16</td>
<td>(D38999) RNA polymerase II 3 (Rpo2-3) [M]</td>
<td>2810048P05</td>
<td>novel protein</td>
</tr>
<tr>
<td>79</td>
<td>3300001C01</td>
<td>(U16321) MIF-2A protein [M]</td>
<td>1110001H08</td>
<td>(M69293) Id2 protein [M]</td>
</tr>
</tbody>
</table>

Clone IDs and annotations of the participants in part of 145 interaction pairs are shown. The complete table is available as an on-line supplement at http://www.genome.org/ and also available at our Web site, http://genome.gsc.riken.go.jp/. Results of BLASTX (2.0.11) search were used for the annotation. Where similar genes with an E-value of ca. 5 are specified, their accession number in parenthesis, gene name, species (in brackets), and E-value are shown. In the case of identical or orthologous genes, their E-values are omitted. Other clones are described as novel proteins.

Abbreviations of species are as follows: A, Arabidopsis thaliana; C, Caenorhabditis elegans; G, Gallus gallus; H, Homo sapiens; M, Mus musculus; Mt, Mycobacterium tuberculosis; O, Oryctolagus cuniculus; R, Rattus norvegicus; Sc, Saccharomyces cerevisiae; Sp, Schizosaccharomyces pombe; X, Xenopus laevis.

All interaction data in this report have been submitted to the public database BIND (http://www.bind.ca/).
and because it suggests the function of uncharacterized proteins (Mayer and Hieter 2000). Thus, global analysis has been performed in yeast PPIs, which showed that most of the interactions were involved in a complex network; the results were used to reasonably predict the function of 29 previously uncharacterized proteins (Schwikowski et al. 2000). We have also predicted the function of a few previ-
ously uncharacterized proteins. Thus, computer analysis of the PPI data seems to be a very powerful tool to explore the functions of proteins and the molecular mechanisms of biolog-
ical processes. However, it should be stressed again that predictions made using such a bioinformatics approach must be confirmed by additional experiments.

The RIKEN mouse genome project has already collected a large number of mouse full-length cDNAs (The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium 2001; http://genome.gsc.riken.
go.jp/). Applying these cDNAs to the assay systems described here paved the way for us toward completion of the rough draft of a PPI panel covering most genes in certain cell types or tissues of a higher organism within a short time.

METHODS

Primers

Each gene-specific forward primer was designed manually as shown in Figure 1. In some cDNAs with partial coding se-
quence (CDS), we designed the primer at the 5’ end of the CDS. Other primers used in this work were as follows: FPCMV5, 5’-CCCATATGGCCCGCATGTGC-3’ and its nested primer FPCMV5, 5’-GCCATGCTGGCCATGTATTTGAC-3’; RKBBSD2, 5’-CATGGTGGGCCGTCCTTGGCCGGAATGCATCTAAGT-3’; RKBBACT2, 5’-CATGTTGGGCCGTCCTTGGCCGGAATGCATCTAAGT-3’; M13 universal forward primer P7, 5’-CGCCAGGGTTTTCCCAGTCACGACGATACAGTCAACTG-3’; M13 universal reverse primer P8, 5’-AACGCGGTATCCAGAATTTCACACAGGAAA-3’.

Construction of the Assay Samples

Plasmid vectors pACT, pBIND, and pGSLuc were from the CheckMate mammalian two-hybrid system (Promega). The protein CDS of each mouse cDNA was amplified with the corresponding gene-specific forward primer and P8 (or P7) primers (first PCR). Because some clones were difficult to am-
plify because of the high GC content around the initiation ATG, PCR was performed using modified dNTPs consisting of 250 µM each of dATP, dCTP, and dTTP, 167 µM of dGTP, and 83 µM of 7-deaza dGTP at the final concentration. The modi-
fied dNTPs improved amplification of the PCR products effi-
ciently. Fragments for the CMV promoter and the Gal4 DNA-
binding domain (BIND fragment) or the VP16 transcriptional activation domain (ACT fragment) were amplified from pBIND or pACT vectors by PCR using the primer sets FPCMV5 and RKBBSD2 or FPCMV6 and RKBBACT2, respectively. Overlapping PCR was performed to connect the CDS frag-
ments with the BIND or ACT fragments. One to two micro-
liters of the first PCR products was mixed with 0.5 µL of ACT or
BIND fragment and amplified in 100-µL reactions using the primer set FPCMV5 and P8 (or –21M13). All of the PCR prod-
ucts were checked by agarose gel electrophoresis before each subsequent step in the PCR protocol.

Western Blotting Analysis

Five microliters of the BIND samples were transfected into 2.4 × 10^5 CHO-K1 cells in 12-well culture plates. After 24 h of incubation, cells were washed once with ice-cold TBS and har-
vested using 100 µL of Lamuli sample buffer. The samples were boiled for 2 min and mixed well. Protein in Lamuli sample buffer (10 µL) was subjected to SDS-PAGE and trans-
fected electrically onto PVDF membrane. The membrane was blocked by TBS/0.05% Tween 20 (TBS/Tween) containing 5% skim milk for 1 h and incubated with anti-Gal4 DNA-binding domain monoclonal antibody (Santa Cruz; dilution 1:200) for 1 h. After washing with TBS/Tween, the membrane was incu-
bated with HRP-conjugated anti-mouse IgG2a (Santa Cruz; dilution 1:1000) for 1 h and washed again with TBS/Tween. Detection of the signal was performed using the ECL plus system (Amersham Pharmacia).

Assay

The PCR-mediated BIND and ACT samples were used directly in the assay without further purification. One-quarter micro-
liter each of the BIND and ACT samples, 30 ng of pGSLuc, and 9.5 µL of Opti-MEM (Lifetech) were transferred into a well of a 384-well plate. Ten microliters of the transfection reagent LF2000 (Lifetech) that was diluted 32-fold with Opti-MEM was added to the well and suspended by pipetting several times. After 20 min of incubation, 20 µL of CHO-K1 cells that were suspended in F12 medium at 1300 cells/µL were added into the transfection mixture and suspended well. The assay samples were incubated in a CO_2 incubator for 20 h. The luciferase activity after the incubation was measured with the Steady-Glo Luciferase Assay System (Promega). Two types of semiautomatic multiple dispensers, 96/384-channel dispensers (Biotec) and 8-channel workstations (Tecan), were used in the high-throughput assay.

Statistical Analysis

Positive candidates were selected statistically after the first assay. The assay data consisting of 24 16-mix BIND samples and 96 6-mix ACT samples were analyzed at one time. First, the SD of each assay point was calculated for each 16-mix BIND sample. These SD values were used for the calculation of the SD for each 6-mix ACT sample because of the large varia-
tion of the background values among the BIND samples. Posi-
tive candidates were selected as those values revealing SD both more than 3.0 for each 16-mix BIND sample and more than 2.0 for each 6-mix ACT sample. These values were de-
termined empirically.

ACKNOWLEDGMENTS

We especially thank C. Kai for his excellent technical assis-
tance. We also thank the members of the Laboratory for Ge-
nome Exploration Research Group. This study was supported by Special Coordination Funds and a Research Grant for the RIKEN Genome Exploration Research Project from the Science Technology Agency of the Japanese Government to Y.H. This work was also supported by a Grant-in-Aid for Scientific Research on Priority Areas and Human Genome Program, from the Ministry of Education, Science and Culture, and by a Grant-in-Aid for a Second Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health and Welfare to Y.H.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES


Colas, P. and Brent, R. 1998. The impact of two-hybrid and related

Received January 18, 2001; accepted in revised form July 5, 2001.