Construction of a modular yeast two-hybrid cDNA library from human EST clones for the human genome protein linkage map

Shao-bing Hua 1,*, Ying Luo 1,2, Mengsheng Qiu 1,3, Eva Chan 2, Helen Zhou 4, Li Zhu

Received 1 February 1998; received in revised form 28 April 1998; accepted 29 April 1998; Received by E.Y. Chen

Abstract

Identification of all human protein–protein interactions will lead to a global human protein linkage map that will provide important information for functional genomics studies. The yeast two-hybrid system is a powerful molecular genetic approach for studying protein–protein interactions. To apply this technology to generate a human protein linkage map, the first step is to construct two-hybrid cDNA libraries that cover the entire human genome. With a homologous recombination-mediated approach, we have constructed a modular human EST-derived yeast two-hybrid library in the Gal4 activation domain-based vector, pACT2. Quality analysis of this library indicated that the approach of constructing two-hybrid cDNA libraries from individually arrayed human EST clones is feasible, and such a two-hybrid library is suitable for detecting protein–protein interactions. This is also the first time that a comprehensive two-hybrid system cDNA library has been constructed from a collection of individually arrayed EST clones. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Functional genomics; Homologous recombination; IAP-1

1. Introduction

The Human Genome Project has generated a tremendous amount of DNA sequence data. To date, nearly 50 000 UniGenes have been identified (Boguski and Schulier, 1995; Miller et al., 1997). They represent approximately 50% of the estimated 100 000 genes in human cells (Rowen et al., 1997). However, only a minority of these UniGenes have known functions. The next major challenge is to study the functions of these genes as well as their regulatory pathways. Several technologies have been implemented in functional genomics studies. For example, oligonucleotide chips (Fodor et al., 1991; Pease et al., 1994), SAGE (Velculescu et al., 1995) and DNA microarrays (Schena et al., 1995) can be used to obtain large-scale or even genome-wide mRNA expression data. These technologies can provide a view of changes in gene expression patterns in response to different physiological states (DeRisi et al., 1997; Zhang et al., 1997). In addition, partial protein sequences from high-resolution, two-dimensional gels and electrospray mass spectrometry of protein complexes can be used to assign peptides to specific gene sequences (Yates et al., 1995; Wilm et al., 1996). These technologies, however, are mainly focused on individual genes and proteins. No information on regulatory pathways and protein–protein interactions, which play pivotal roles in physiological status, can be easily deduced from results obtained by employing these technologies. We rationalize that identification of all protein–protein interactions in human cells and further establishment of a complete human protein linkage map will provide important information to the functional genomics studies.
Yeast two-hybrid technology, developed by Fields and coworkers (Fields and Song, 1989; Chien et al., 1991; Bartel et al., 1993), has proven to be a successful molecular genetic approach for detecting interactions of proteins from different cellular compartments (Duree et al., 1993; Guan et al., 1994; Hsu et al., 1996; Mendelsohn and Brent, 1994; Schreiber-Agus et al., 1995), including extracellular ligand-receptor interactions (Ozenberger and Young, 1995). The advantages of two-hybrid systems over other biochemical methods, such as coimmunoprecipitation and comminoprecipitation, are apparent (Mendelsohn and Brent, 1994). It provides an easy and rapid in-vivo binding assay in eukaryotic cells. The genes encoding the interacting partners are readily available. Furthermore, the two-hybrid system can be modified for a rapid and high throughput screening procedure, which is becoming increasingly important in the functional genomics era. These advantages make the two-hybrid system a top choice for large-scale analysis of all human protein-protein interactions (Evangelista et al., 1996; Fields, 1997). In fact, a smaller scale bacteriophage T7 protein linkage map based on the yeast two-hybrid system has already been reported (Bartel et al., 1996).

In order to build a complete human protein linkage map by two-hybrid technology, it is critical to construct appropriate two-hybrid cDNA libraries in both two-hybrid activation domain (AD) vectors and DNA binding domain (DBD) vectors (Bartel et al., 1996). Ultimately, these libraries should include cDNA of all the human genes. However, currently available two-hybrid cDNA libraries were generated from limited number of individual tissues, organs, or cell lines. These libraries were not normalized; therefore, they have limited representations of low abundance cDNAs. These libraries are almost exclusively built in AD vectors. In addition, the identity of an interacting partner has to be determined by DNA sequencing. More desirable yeast two-hybrid system cDNA libraries for the human protein linkage map should include the following features: (1) The cDNA libraries should be normalized; therefore, the libraries have good representation of low abundance cDNAs (Soares et al., 1994; Bonaldo et al., 1996). (2) It is preferable to have long inserts, but not necessarily to have full-length cDNA, since frequently occurring stop codons in the S′ untranslated region of a cDNA would prevent the formation of the hybrid protein. (3) cDNA clone from every human gene should be included. (4) The identities of interacting partners are readily available. Here, we report the construction of an arrayed modular yeast two-hybrid cDNA library in a Gal4 AD vector pACT2 from human EST clones of the Merck-Washington University EST project (Gerhold and Caskey, 1996) by in-vivo homologous recombination (Hua et al., 1997) in a high throughput manner. Human EST clones generated by the Merck-Washington University EST project largely fulfill the above criteria. The project is targeted to cover the entire human genome. The quality of the constructed modular two-hybrid library was analyzed by sequencing random clones of the library as well as by test-library screening against a human apoptosis protein, IAP-1. The results indicated that the approach to construct two-hybrid cDNA libraries from individually arrayed human EST clones is feasible, and that the constructed two-hybrid library is suitable for detecting protein-protein interactions. A different modular two-hybrid library in AD vectors, as well as a two-hybrid library in DBD vectors, can be constructed in a similar manner. This is also the first two-hybrid cDNA library constructed from a large number of individually arrayed EST clones.

2. Materials and methods

2.1. Yeast strains and media

Saccharomyces cerevisiae strain Y 190 (MATa, ura3–52, his3–D0200, ade2–101, trpl–901, leu2–3,112, gal4A::gal80, URA3::GAL–lacZ, cyh2, LYS2::GAL–H153) was from Dr Stephen Elledge (Harper et al., 1993). This strain was used in conventional cDNA library screening. Yeast strain J693 (MATa, ura3–52, his3–D0200, ade2, trpl, leu2, gal4A::gal80, URA3::GAL–lacZ, cyh2, LYS2::GAL–H153) was from Dr. Rodney Rothstein. It was used in the construction of the human EST-derived two-hybrid cDNA library. Yeast media were from CLONTECH Laboratories, Inc. A LiA○PEG transformation kit (CLONTECH) was used for yeast transformation.

2.2. Human EST clones

Human EST clones were obtained from Genome Systems, one of the I.M.A.G.E. (Integrated Molecular Analysis of Genomes and their Expression) Consortium (Lennon et al., 1996) distributors. These human EST clones were arrayed in 384-well plates in Lawrence Livermore National Laboratories (LLNL) and sequenced by the Merck-Washington University EST project (Gerhold and Caskey, 1996) by in-vivo homologous recombination (Hua et al., 1997) in a high throughput manner. Human EST clones were normalized; therefore, the libraries have good representation of low abundance cDNAs. These were from CLONTECH Laboratories, Inc. A LiA○PEG transformation kit (CLONTECH) was used for yeast transformation.
2.3. Oligonucleotide primers

Human EST insert DNA was amplified by the polymerase chain reaction (PCR). The oligonucleotide primers used in PCR amplification are listed below. There were four distinct primers for each specific E. coli-based plasmid vector used in the original EST library construction. The first three primers of each set were forward primers. They were designed to amplify EST inserts in three open reading frames (see Fig. 1). The bases that were added to allow for coding all three open reading frames are underlined below. The fourth primer was the reverse primer that corresponded to the downstream sequence of the E. coli plasmid vector. Each primer consisted of two portions. The 5' sequence (64 bases) corresponded to the sequence flanking the multiple cloning sites (MCS) of the yeast two-hybrid AD vector pACT2, so that the PCR products of EST clones could be fused to the carboxy-terminus of the Gal4 activation domain in pACT2 by in-vivo homologous recombination (Hua et al., 1997). The 3' sequence (21 bases) of the primers, however, corresponded to sequences flanking the EST inserts in the E. coli plasmid vectors. Due to their lengths, these oligonucleotides were purified by polyacrylamide gel electrophoresis after synthesis. The sequences of the oligonucleotide primers are listed below:

For pT7T3D-Pac vector:

1-1: 5'-ACCCCACCAAACCCAAAAAAAGAGAT-CGTATGGCTTACCCATACGAT GTTCCAGAT-

1-2: 5'-ACCCCACCAAACCCAAAAAAAGAGAT-CGTATGGCTTACCCATACGAT GTTCCAGAT-

Fig. 1. Scheme of large-scale transfer of human EST clones to two-hybrid vector pACT2 by homologous recombination. The EST inserts were amplified in three open reading frames simultaneously by PCR. PCR-amplified EST fragments were mixed with linearized pACT2 DNA and cotransformed into yeast cells. The EST fragments were then integrated into pACT2 at the carboxy-terminus of the Gal4 activation domain (AD) via homologous recombination. Multiple cloning site (MCS) region is enlarged for viewing. '6' indicates the MCS region digested by BamHI and XhoI. Recombination sites are indicated by dotted cross-lines. Flanking regions of the MCS are represented by blank boxes and dotted boxes.
2.4. In-vivo cloning by homologous recombination

E. coli cultures containing EST clones were transferred from one 384-well plate to four 96-well plates by a 96-pin replicator (V and P Scientific, Inc.). Approximately 1 ml of E. coli culture was transferred into 100 ml of LB medium containing ampicillin (100 ng/ml). The 96-well E. coli plates were then incubated overnight on a microtiter-plate shaker at 37 °C. By using a 96-pin replicator, approximately 1 ml of confluent E. coli culture was transferred into each tube of a 96-tube PCR plate for direct amplification. The PCR mixture (a total of 50 nl final volume) contained 0.8 ml K lenTaq polymerase mix (CLONTECH), 1 ml of 10 mM 5′ end primer (0.33 nl of each ORF primer) and 1 ml of 10 mM 3′ end primer. The PCR amplification started with denaturing at 94 °C for 3 min, followed by 30 cycles of 15 s denaturation at 94 °C and a 2-min extension at 68 °C.

pACT2 plasmid DNA was digested with BamHI and XhoI. One microgram of digested plasmid DNA was mixed with a total of 25 ml of PCR products and transformed into yeast strain J693 for in-vivo recombination (Hua et al., 1997). The transformants were spread on an agar plate of SD medium lacking leucine (SD-L) and incubated at 30 °C for 4–5 days. A total of 10,000–100,000 transformants were routinely obtained on each agar plate. Yeast transformants on the plate were then harvested with 4 ml of liquid SD-L medium containing 20% glycerol, aliquoted into each well of a 96-deep-well box, and stored at −70 °C in multiple copies.

2.5. cDNA library screening

Human IAP-1, an apoptosis inhibitor (Liston et al., 1996), was used as a test bait in the library screening. Complete coding region of human IAP-1 cDNA was cloned in-frame at the carboxy-terminus of Gal4 binding domain (Hua et al., 1997). The transformants were spread on an agar plate of SD medium lacking leucine (SD-L) and incubated at 30 °C for 7 days. A total of 10,000–100,000 transformants were routinely obtained from each agar plate.

By using a 96-pin replicator, approximately 1 ml of confluent E. coli culture was transferred into each tube of a 96-tube PCR plate for direct amplification. The PCR mixture (50 nl final volume) contained 0.8 ml K lenTaq polymerase mix (CLONTECH), 1 ml of 10 mM 5′ and 3′ end primer (0.33 nl of each ORF primer) and 1 ml of 10 mM 5′ and 3′ end primer. The PCR amplification started with denaturing at 94 °C for 3 min, followed by 30 cycles of 15 s denaturation at 94 °C and a 2-min extension at 68 °C.

Human IAP-1, an apoptosis inhibitor (Liston et al., 1996), was used as a test bait in the library screening. Complete coding region of human IAP-1 cDNA was cloned in-frame at the carboxy-terminus of Gal4 binding domain in pGBT9 (Bartel et al., 1993), thus designated pGBT9/IAP-1. To screen the human EST derived two-hybrid cDNA library described above, the library was first amplified in 96-well plates overnight at 30 °C. Approximately 10,000–100,000 transformants were routinely obtained on each agar plate. Yeast transformants on the plate were then harvested with 4 ml of liquid SD-L medium containing 20% glycerol, aliquoted into each well of a 96-deep-well box, and stored at −70 °C in multiple copies.
DNA from these cDNA libraries was prepared and transformed into competent yeast cells (strain Y190) containing pGBT/IAP-1. Approximately 300 μg of plasmid DNA were used for each transformation. Transformants were then plated on SD-LWH in the presence of 45 mM 3-AT. Large colonies were picked for further HIS growth selection and β-galactosidase activity assay as described above.

2.6. Plasmid retrieval and sequencing

Plasmid DNA was retrieved from HIS/lacZ colonies by using a plasmid isolation kit (CLONTECH). cDNA inserts were amplified by PCR, and their sequences were determined. Plasmid constructs containing human IAP-1 and its interacting partners were subsequently re-transformed into yeast to verify their interactions.

3. Results

3.1. Large-scale transfer of EST inserts into yeast two-hybrid vectors

The Merck-Washington University EST project has sequenced more than 300,000 human EST clones, and their sequences have been deposited into the dbEST division of GenBank (see URL http://www-bio.lnl.gov/bbrip image/image.html for the updated information). These EST clones were picked from a number of individual cDNA libraries derived from various tissues/organisms at different developmental stages (Hillier et al., 1996). So far, approximately 50,000 UniGenes have been derived from the combination of these EST clones and other resources in GenBank (Boguski and Schuler, 1995; Miller et al., 1997). The majority of the cDNA libraries used to generate EST clones are normalized cDNA libraries (Bonaldo et al., 1996; Hillier et al., 1996). Two-hybrid cDNA library generated from such EST clones will, thus, have a high complexity and good representations of low abundance cDNAs. In addition, the two-hybrid library can be expanded in the future along with the progress of the Merck-Washington University EST project. The new EST clones can be subsequently transferred to two-hybrid vectors to generate additional non-overlapping two-hybrid cDNA libraries. Eventually, the two-hybrid system cDNA library generated from the EST pool will cover the entire human genome.

We obtained a total of 251,520 human EST clones (arrayed in 384-well plates) (LLNL plate nos 1–655). LLNL plate nos 7–16 were not used in the current study due to their lack of sequence information. Clone identification numbers were downloaded from ftp://humpty.llnl.gov/pub/image/outgoing/. The GenBank Accession Nos of these human EST sequences were downloaded from ftp://ncbi.nlm.nih.gov/repository/dbEST/. Downloaded data were stored in two tables in a Microsoft Access-based database (designated GeneNet database) and linked by IMAGE ID and LLNL ID. The GeneNet database can be used to search sequence, insert size, tissue origin, and plate location information of each EST clone.

The E. coli-based plasmid vectors used in the original EST library construction (Section 2.2) cannot be used with the yeast two-hybrid system since yeast expression elements are absent from these vectors. Transfer of EST inserts from these vectors to the yeast two-hybrid vectors by conventional molecular cloning (Sambrook et al., 1989) would require multiple steps of in-vitro manipulation, which is rather time-consuming and expensive. It is impractical to use such an approach for large-scale construction of a complete yeast two-hybrid cDNA library from human EST clones. We have previously reported a rapid cloning method to transfer EST inserts from their original E. coli-based vectors into the yeast two-hybrid expression vector in a single step by homologous recombination in yeast (Hua et al., 1997). We adopted this approach of in-vivo cloning by homologous recombination in the current study. The scheme of this procedure is illustrated in Fig. 1.

Due to the large number of clones in the human EST libraries, we decided to use a limited pooling method to transfer EST fragments from their original vectors to the two-hybrid system vector pACT2 in order to maximize the work efficiency of our resources. The pooling procedure is shown in Fig. 2. E. coli cells from each original 384-well EST plate were inoculated in four 96-well plates, using a 96-pin replicator. A mixture of three upstream primers specific to each open reading frame (ORF) of the clones was used since ORF information is not available for the EST clones. After PCR amplification, aliquots of PCR products from five tubes of each 96-tube PCR plate were run on an agarose gel to examine the quality of PCR products. In most cases, eight non-overlapping bands were visible (data not shown).

After the first pooling step, we generated 326 × 96-tube PCR plates. After the PCR amplification, each tube contained the EST inserts amplified from eight EST clones. We then pooled four rows of PCR products into one row (Fig. 2), resulting in a mixture of PCR products from 32 EST clones in a single well. These EST inserts were mixed with linearized pACT2 and transformed into yeast cells so that these inserts were integrated into
Fig. 2. Pooling procedures for generating the arrayed human EST-derived two-hybrid cDNA library. One original 384-well plate of LLNL EST clones is shown here (LL1 represents LLNL human EST plate no. 1). One block of the plate is enlarged to show the location index of four individual wells. LL1a01 means the a01 well of LL1 plate. E. coli cultures from one 384-well plate were transferred to four 96-well plates (designated E1–E4). The new locations of the above mentioned block are indicated by ‘Ω’. In addition, the pattern of location changes is described above each 96-well plate.

E. coli cultures from plates E1–E8 (plates E5–E8 derived from plate LL2 are not shown here) were pooled into one 96-tube PCR plate (P1) for amplification of EST inserts by PCR. Each well of plate P1 contained eight EST clones (their identities are shown here). After PCR amplification, four rows (indicated by shaded circles) were pooled into one row on a new 96-well plate. EST fragments in each well (derived from original 32 EST clones) were mixed with linearized pACT2 DNA and transformed into yeast cells. The transformants were spread on an SD-L agar plate. After incubation, yeast transformants from each SD-L plate were harvested and arrayed in an individual well of a 96-deep-well box in multiple copies.

The two-hybrid system vector by in-vivo homologous recombination (Hua et al., 1997) as described below. pACT2, a GAL4 activation domain vector, was chosen as the cDNA library backbone vector (Li et al., 1994). This vector can facilitate detection of weaker protein–protein interactions compared with other popular vectors such as pGAD424 (Legrain et al., 1994), partly due to its high expression promoter (Li et al., 1994). pACT2 plasmid DNA was first digested by restriction endonucleases BamHI and XhoI. PCR-amplified EST inserts and the digested vector DNA were cotransformed into yeast strain J693 in 96-deep-well
3.2. Sequencing verification

In order to verify that each well of the newly constructed two-hybrid cDNA library contained the EST clones that we originally pooled from, we retrieved plasmid DNA from four randomly selected wells of the human EST-derived two-hybrid cDNA library, and further transformed the DNA into E. coli. Plasmids from a total of 40 randomly picked E. coli colonies were prepared. Sequences of the EST inserts in these plasmids were determined from their 5' ends, i.e. upstream of the Gal4 binding domain junctions.

We found that 28 out of 40 sequenced clones (70%) had their matches in the EST database (data not shown). All the matched clones are derived from their anticipated wells. The results correlated very well with known sequence information generated by the Merck-Washington University EST project, in which only 72% of the original EST clones contained 5' end sequence information in dbEST. The rest of the EST clones from the Merck-Washington University EST project either had only 3' end sequence information or had no sequence information at all. The ratio of EST clones containing three open-reading frames, which was introduced through the PCR primer, was very close to 1:1:1. As observed previously (Hua et al., 1997), no mutations were present in the recombination regions of any of the sequenced clones. Only one clone (2.5%) of 40 sequenced clones did not contain an EST insert, consistent with our previous observations on in-vivo cloning by homologous recombination (Hua et al., 1997).

3.3. Test screening of two-hybrid system EST cDNA library

To demonstrate that this EST-derived two-hybrid cDNA library can be used for identifying protein-protein interactions, we performed a test-library screening using human IAP-1, an apoptosis inhibitor (Liston et al., 1996), as the bait protein. cDNA from the coding region of IAP-1 was fused in-frame to the carboxy-terminus of the Gal4 DNA-binding domain of pGBT9, resulting in pGBT9/IAP-1. All clones of our arrayed human EST-derived two-hybrid cDNA library were pooled after being individually amplified. The pooled library was then transformed with pGBT9/IAP-1 DNA, and further subjected to HIS growth selection and β-galactosidase color assay (see Section 2.5) to identify proteins interacting with human IAP-1. The identities of the positive clones were determined by sequencing after the plasmids were retrieved from yeast cells. To compare the library screening results of the human EST-derived two-hybrid library with the conventional two-hybrid libraries, we also screened three randomly selected conventional two-hybrid cDNA libraries (human brain, human liver, and human lymphocyte) against human IAP-1. The results are shown in Table 1.

From human EST-derived two-hybrid cDNA library screening, we have successfully identified 10 interacting partners of IAP-1, including TRAF2, a known interacting protein of IAP-1 (Shu et al., 1996; Uren et al., 1996), from EST clones that we originally pooled from, we retrieved plasmid DNA from four randomly selected wells of the human EST-derived two-hybrid cDNA library, and further transformed the DNA into E. coli. Plasmids from a total of 40 randomly picked E. coli colonies were prepared. Sequences of the EST inserts in these plasmids were determined from their 5' ends, i.e. upstream of the Gal4 binding domain junctions.

We found that 28 out of 40 sequenced clones (70%) had their matches in the EST database (data not shown). All the matched clones are derived from their anticipated wells. The results correlated very well with known sequence information generated by the Merck-Washington University EST project, in which only 72% of the original EST clones contained 5' end sequence information in dbEST. The rest of the EST clones from the Merck-Washington University EST project either had only 3' end sequence information or had no sequence information at all. The ratio of EST clones containing three open-reading frames, which was introduced through the PCR primer, was very close to 1:1:1. As observed previously (Hua et al., 1997), no mutations were present in the recombination regions of any of the sequenced clones. Only one clone (2.5%) of 40 sequenced clones did not contain an EST insert, consistent with our previous observations on in-vivo cloning by homologous recombination (Hua et al., 1997).

<table>
<thead>
<tr>
<th>EST clones</th>
<th>100</th>
<th>100</th>
<th>100</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total transformants (millions)</td>
<td>69</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distinct HIS+/lacZ+ clones</td>
<td>4</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Known partner</td>
<td>TRAF2</td>
<td>TRAF2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New interactions identified</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Comparison of library screening results of the human EST-derived yeast two-hybrid cDNA library with those of conventional two-hybrid libraries (human brain, human liver and human lymphocyte) using human IAP-1 as the bait protein.
library described in this report is suitable for generating the human protein-linkage map.

4. Discussion

We have successfully constructed a modular human EST-derived two-hybrid library from arrayed human EST clones generated by the Merck-Washington University EST project in a high-throughput manner. With the progress of the Merck-Washington University EST project, such a two-hybrid cDNA library could eventually cover the entire human genome. Analysis of the quality of our current EST-derived two-hybrid cDNA library by sequencing random clones and test library screening indicated that the modular library can be successfully used for identifying protein–protein interactions, and ultimately for the establishment of a complete human protein linkage map. Such a human protein linkage map will provide invaluable information in the functional genomics era. For example, it will provide functional hints for novel genes and reveal new functions for known genes. Many regulatory and metabolic pathways can be identified from the linkage map. Furthermore, the linkage map will also provide potential targets for therapeutic intervention.

As opposed to traditional library construction methods, we used a homologous recombination-mediated approach for constructing a yeast two-hybrid library from arrayed EST clones. Homologous recombination as a means of cloning in yeast has been introduced in the last decade (Orr-Weaver and Szostak, 1983; Ma et al., 1987). Here, we have successfully extended the technique to the large-scale construction of an arrayed two-hybrid cDNA library in a high-throughput manner. The most time-consuming step in the current process of library construction is harvesting yeast transformants from each plate and further transferring them into 96-deep-well boxes. One major advantage of an arrayed cDNA library is that the mating method (Bendixen et al., 1994; Finley and Brent, 1994) can be adopted for high-throughput screening of the library in the future.

In the current human EST-derived two-hybrid cDNA library, we used EST clones from 58 cDNA libraries derived from more than 28 human tissues/organs at different developmental stages (Hillier et al., 1996). Most of these EST libraries were prepared after cDNA normalization or subtraction (Bonaldo et al., 1996; Hillier et al., 1996), which improves library complexity and representation of low abundance mRNA. A two-hybrid cDNA library derived from such EST clones, thus, has some advantages over currently available conventional two-hybrid cDNA libraries, which are largely made from single tissues or organs, and not normalized. Comparison of the test library screening results between the human EST-derived cDNA library and a combination of three conventional two-hybrid libraries indicated that more positive interactions were identified from the human EST-derived two-hybrid cDNA library, even though significantly more transformants from the conventional libraries were screened.

Currently available human EST clones from the Merck-Washington University EST project do not cover the entire human genome. Therefore, the EST-derived two-hybrid library that we reported here is a modular library. Construction of additional modular libraries does not have to include EST clones present in the current library. During the process of library construction, we transferred all the available EST clones to the two-hybrid vector, rather than selecting UniGene EST clones before library construction. Limited duplication is necessary and even desirable for cross-referencing in the genome project. In addition, cDNA clones with different lengths that are encoded by a single gene are present in the human EST libraries. Information on interacting domains of a specific protein may become immediately available after analyzing the screening results when such a two-hybrid library is used for screening. However, we would recommend that UniGene clones should be used for constructing two-hybrid bait libraries in the DBD vectors in the future. It should also be noted that many EST clones do not contain the complete 5' sequences (Hillier et al., 1996); therefore, the amino termini of the encoded proteins may not be present. Protein interactions mediated by those missing amino termini will not be detected using the current library. Such a problem may be overcome in the future when additional modular two-hybrid libraries made from full-length cDNAs or cDNAs prepared by random priming become available.

The application of the human EST-derived two-hybrid cDNA library, as we propose, is largely to search for all possible interactions and thus to establish a global human protein linkage map. It should be kept in mind that protein–protein interactions detected in two-hybrid library screening might not exist in an individual tissue type at a specific developmental or disease stage. Detection of several interacting partners for a single protein does not provide any evidence for exclusive or simultaneous binding. Limitations to the two-hybrid approach include positive interactions that are detected by this method but do not occur in human cells. For example, proteins derived from large gene families or proteins that interact via short stretches of residues might fall into this category. A further potential problem is false-negative results. It might be impossible to detect certain interactions that are mediated by posttranslational modifications such as glycosylation and myristylation or that occur between proteins unable to fold appropriately in an intracellular environment. Therefore, integration of information provided by other technologies, such as genome-wide expression data
and/or intracellular localization data, into the human protein linkage map would greatly facilitate the discovery of disease genes and potential pharmaceutical targets, as well as the advancement of basic research in human biology.

Acknowledgement

We are grateful to Drs Stanley Fields, Rodney Rothstein, Stephen Elledge, Chiao-Chien Hwang, and Hong Jin for their invaluable supports and inputs on this project. We also thank Ann Holtz for her advice during this work, and John Ambrozik for critical reading of the manuscript.

References

during this work, and John Ambroziak for critical recombination in yeast. Plasmid 38, 91–96.

Acknowledgement

We are grateful to Drs Stanley Fields, Rodney Rothstein, Stephen Elledge, Chiao-Chien Hwang, and Hong Jin for their invaluable supports and inputs on this project. We also thank Ann Holtz for her advice during this work, and John Ambrozik for critical reading of the manuscript.

References

during this work, and John Ambroziak for critical recombination in yeast. Plasmid 38, 91–96.

Acknowledgement

We are grateful to Drs Stanley Fields, Rodney Rothstein, Stephen Elledge, Chiao-Chien Hwang, and Hong Jin for their invaluable supports and inputs on this project. We also thank Ann Holtz for her advice during this work, and John Ambrozik for critical reading of the manuscript.

References

during this work, and John Ambroziak for critical recombination in yeast. Plasmid 38, 91–96.

Acknowledgement

We are grateful to Drs Stanley Fields, Rodney Rothstein, Stephen Elledge, Chiao-Chien Hwang, and Hong Jin for their invaluable supports and inputs on this project. We also thank Ann Holtz for her advice during this work, and John Ambrozik for critical reading of the manuscript.

References

during this work, and John Ambroziak for critical recombination in yeast. Plasmid 38, 91–96.