Vol. 44, No. 2, February 1998 BIOCHEMISTRY *and* MOLECULAR BIOLOGY INTERNATIONAL
Pages 225-233

INCREMENT OF EFFICIENCY IN THE IDENTIFICATION OF NOBLE GENES BY COLONY HYBRIDIZATION ASSAY

(SCREENING OF LOW REDUNDANT CLONES FROM HUMAN FETAL CDNA LIBRARY)

J.W. Kim, I.A. Lee, Y.H. Lee, J.C. Song, Y.K. Choe, Y.S.Hahn[1], J.H.Chung[1], T.W. Chung and I.S. Choe*

Molecular and Cellular Biology Division, Korea Research Institute of Bioscience and Biotechnology, KIST, Taejon 305-600,Korea
[1]Department of Biological Sciences, KAIST, Taejon 305-701, Korea

Received November 3, 1997

SUMMARY

For the rapid identification of noble genes in a specific tissue by computer analysis from the cDNA sequences determined by single-pass cDNA sequencing, clone redundancy was one of the major obstacles. To facilitate the efficiency in identification of noble genes, it was necessary to reduce the number of clones to be sequenced by eliminating the redundant clones for a rapid analysis. In order to increase the probability of isolating noble sequences from the cDNA clones of human fetal liver tissue origin, colony hybridization assay was adopted and redundant clones were efficiently removed. Four cDNA clones highly redundant in the human fetal liver cDNA libraries including α-globin, γ-globin, serum albumin and H19 RNA sequences were selected as the probes. Two hundreds and sixty two cDNA clones were randomly selected and tested with the probes for hybridization properties. The identity of each cDNA clone giving positive or negative signals in the hybridization assay was determined by DNA homology search with the nucleic acid databases. Among the 76 clones giving positive signals, 57 clones (75%) were found to be identical to the probe sequences and could be eliminated by colony hybridization assay before neucleotide sequencing.

Key words : Human fetal liver cDNA, redundancy, prescreening, hybridization assay.

INTRODUCTION

A major effort has been the random nucleotide sequencing of the partial cDNA sequences isolated from the cDNA libraries of a specific tissue or a cell line in identifying novel genes and in understanding the structures and the functions of proteins. Partial nucleotide sequencing of the cDNA clones from the cDNA libraries of human brains (Adams *et al.*, 1992), liver (Okubo *et al.*, 1992), heart tissue (Liew *et al.*, 1994)and human fetal liver(choi, *et al.* 1995)has been performed.

1039-9712/98/020225-09$05.00/0

In an attempt to identify the novel cDNA sequences by single pass nucleotide sequencing, the redundancy of the expressed genes was a major obstacle to obtaining the high proportion of the cDNA clones represented with low frequency to the total number of clones isolated from the cDNA libraries. A random sequencing study with brain cDNA libraries revealed that only 46% of the cDNA clones contained novel sequences. The removal of the highly redundant cDNA clones prior to a sequence determination would lead to an increase in the probability of the identification of rare cDNA sequences. To accomplish this purpose, several approaches including differential hybridization with labeled cDNAs as the probes (Hoog, *et al.* 1991), the construction of subtractive cDNA libraries (Duguid, *et al.* 1990, Fargnoli, *et al.* 1990, Diatchenko, *et al.*1996, Deleersnijder, *et al.*1996) and the preparation of a normalized cDNA library (Sasaki, *et al.* 1994) have been devised to remove the redundant cDNA clones or to prepare the tissue specific cDNA clones of high quality.

To improve the probability of isolating novel cDNA sequences, Hoog *et al* synthesized the mixtures of cDNAs from poly(A)$^+$ RNAs prepared from liver, kidney, and heart tissue or prepubertal testis and used them as the probes. The cDNA hybridization experiment by Hoog *et al* revealed that approximately 30% of the cDNA clones in the prepubertal testicular cDNA library represented redundnant cDNA clones.

In a normalized cDNA library constructed by repeated dissociation and reassociation of double stranded short cDNA fragments in solution (Sasaki, *et al.* 1994), the frequency of the redundnant markers decreased. For example, the frequency of the most abundant marker, β-globin cDNAs decreased from 1.067% to 0.009% and the endogenous β-actin cDNAs belong to the abundant cDNA class decreased from 0.54% to 0.02%.

The removal of the frequently expressed cDNA clones in a cDNA library prior to their nucleotide sequencing would increase the ratio of the rare species of the cDNA clones represented in a library. In this study, it was found that the employment of the proper number of probes was essential for the efficient removal of the redundant cDNA sequences and for the efficient selection of noble genes from the human fetal liver cDNA library.

MATERIALS AND METHODS

Construction of cDNA library

Total RNAs were purified from the liver tissue of a 26 week old human fetus of Korean origin by the acid guanidine phenol chloroform (AGPC) method described by Chomczynski *et al.* with a minor modification (Chomczynski *et al.*, 1987). Poly(A)$^+$ RNAs were isolated from the total RNAs

using an oligo- dT cellulose affinity chromatography (Lin *et al.*, 1991 : Haqqi *et al.*, 1992). The cDNA libraries were constructed using a λ-ZAP cDNA cloning kit (Stratagene, La Jolla, CA) according to the manual provided by the manufacturer (Stratagene). 5μg of poly(A)$^+$ RNA was primed with an oligo-dT primer for the synthesis of cDNA using a commercial cDNA synthesis kit. After the ligation of EcoRI adaptors onto the cDNA, it was digested with XhoI, and finally ligated into EcoRI, XhoI-cut λ-ZAP vector from Stratagene. Ligated DNAs were packaged, *in vitro*, using a Gigapack II Gold packaging system (Short *et al.*, 1992).

DNA sequencing

Sequenase Kit Version 2.0 was purchased from USB Co. (Cleveland, Ohio, U.S.A.). Template DNAs for nucleotide sequencing were prepared from the phagemid by excision (Alting-Meese *et al.*, 1989). SK primer sequence (CGCTCTAGAACTAGTGGATC ; Korea Biotech. Co., Taejon, Korea) was used to determine the 5′-end nucleotide sequences of the cDNA clones. After the sequencing reactions, the samples were run onto 6% polyacrylamide/7M urea gel in 1X TBE buffer at a constant voltage of 1800V. Dideoxy chain termination reactions were performed with [^{35}S]-dATP obtained from Amersham.

Preparation of the probes for colony hybridization

For the preparation of probe DNAs, the plasmids containing α-globin, γ-globin, H19, and serum albumin sequences were used as the template DNAs for polymerase chain reactions (PCRs). PCRs were carried out under the following conditions. Each of 50μl reaction mixtures contained 100pg of the template DNA, 0.2mM each of dATP, dGTP, dCTP, 19:1 mixture of dTTP and DIG-dUTP (Boehringer Mannheim Co., Germany), 50ng each of synthesized primers for 5′-end (ATTAACCCTCACTAAAGAAC) and 3′-end (AATACGACTCACTATAGGGC) and 2.5 units of Taq DNA polymerase (PerkinElmer Cetus, U.S.A). The DNA was denatured at 94℃ for 7 minutes prior to the amplification reaction. Thirty cycles of the PCR was performed with the denaturation of DNA at 94℃ for 50 seconds, the annealing step at 55℃ for 50 seconds, and the polymerization reaction at 72℃ for one minute. The polymerization reaction in the last cycle of the PCR was extended for 10 minutes. The concentration of the amplified DNAs was accessed by the chemi-luminescent method using Lumigen PPD.

Colony hybridization.

The glycerol stocks were made after the mass excision of the phagemid from the λ-ZAP express vector (Stratagene, 1993), a single colony was picked and cultured for 16 hours in a 96 well microtiter plate containing 50μl Luria-Bertani broth (LB broth) medium and by adding an equal volume of 40% glycerol. Colonies from glycerol stocks were transferred onto a sheet of nylon membrane using a Multitransfer (Korea Biotech, Korea) and grown for 16 hours. After denaturation and neutralization reactions, the filters were hybridized with the DIG labeled probes according to the manufacture's protocol (BM Co., 1993). The hybridization solution contained 5X SSC, 1.0%(w/v) blocking reagent, 0.1% N-lauroylsarcosine and 25ng of the mixture of the heat denatured probe DNAs. 10μg/ml of sonicated cDNA purified from the clone containing only the vector sequence and bacterial DNAs purified from E. coli strain SOLR were used as competitor DNAs (Short., *et al* 1988).

RESULTS AND DISCUSSION

In our previous study, the cDNA libraries from human fetal liver tissues were constructed without modification (Kim, *et al*,. 1995) using a λ-Zap XR II system and partial cDNA sequences were determined. The partial cDNA sequences were examined for similarities to the GenBank nucleic acid database with a modified BLAST program. The cDNA sequences which did not show any database match in the first computer search were tested again using the BLAST e-mail server at the National Center for Biotechnology Information. Database entries which showed high homology ($P<0.001$) to the known sequences were extracted and alined with the sequences of the cDNA clones isolated in this study using the FASTA program. The results of the database match were confirmed by manual inspection of the FASTA results. From a set of 4,189 partial cDNA clones sequenced, 3,437 were informative complementary DNA sequences to nuclear encoded mRNAs. Another 732 clones were identified to be the vector DNA sequences or clones containing cDNA sequences shorter than 60 bp. The informative cDNA sequences of the 3,437 clones were classified into four groups as shown in Fig. 1 based on the criteria described below. 2,354 of the sequences (68.5% of the informative clones) consisted of matches to previously known human DNA sequences and designated as identified clones. The clones in this group satisfied the following criteria: Probability specified by Altschul *et al.*(1990) was below 10^{-5}; nucleic acid sequences matched with more than 60% identity and the number of matched nucleotides were larger than 60 bases. 187 clones (5.4%) matched the repetitive elements of the human genome and labelled repeating sequences. Alu sequences, L1 sequences, di- and tri-nucleotide repeats were included in this group. 216 clones (6.3%) matched the known expressed sequence tags (ESTs). 690 clones (20.0%) matched no known sequences in the databases as of December 1996.

Because of the clone redundancy, the 2,354 clones designated as the identified clones could be classified into only 286 kinds of independant genes. These redundancies were also found in other random cDNA sequencing studies. It is known that the most abundant clones in human brain cDNA libraries (Adams *et al.* 1992) are β-actin (0.6%) and myelin basic protein genes (0.5%), and cytochrome b (20.4%) and elongation factor 1α (18.5%) genes were the most frequently expressed genes in the mouse testis cDNA library (Hoog *et al.*, 1991). In the human fetal liver cDNA libraries, the frequency of globin genes including α-, β- and γ-globins is 807 (23.4%)

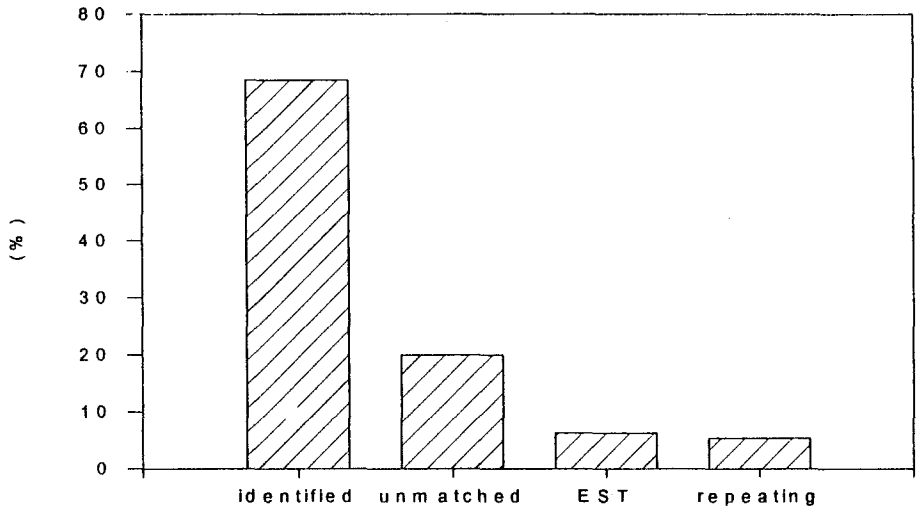

Fig. 1. Sequence analysis of 3,437 cDNA clones from human fetal liver
Informative 3,437 cDNA clones were classified into 4 groups as 2,354 identified clones((68.5%), 690 unmatched clones(20%), 216 EST(6.3%) and 187 clones with repeating element(5.4%)

among the total informative clones of 3,437 and the most abundant species(Fig.2). The number of genes among the 3,437 informative clones showing redundancies of more than 10 are 1,909 (55.6%).

To improve the efficiency of isolating noble cDNA clones, the highly redundant clones should be removed prior to the sequence determination. For the determination of efficiency in identifying the novel genes, two hundred and sixty two cDNA clones were randomly selected, sequenced and database matched. These 262 cDNA clones were also probed with four cDNA clones highly redundant in fetal liver cDNA libraries. The redundancies of the four clones were 500 for γ-globin (21.2%), 306 for serum albumin (12.9%), 282 for α-globin (12.1 %) and 55 for H19 RNA (2.3%) as shown in Fig.2. A mixture of these four cDNA sequences was hybridized to the 262 selected clones.

Fig.3. shows that 186 clones(71.0%) that had failed to hybridize with the four probe DNAs representing the highly redundant clones, were classified with the same criteria as described earlier. Table 1. and Fig.4. show that different patterns emerged when these results were compared with those of the random sequencing study. Marked differences were noted in the reduction of the redundant gene species. The composition of

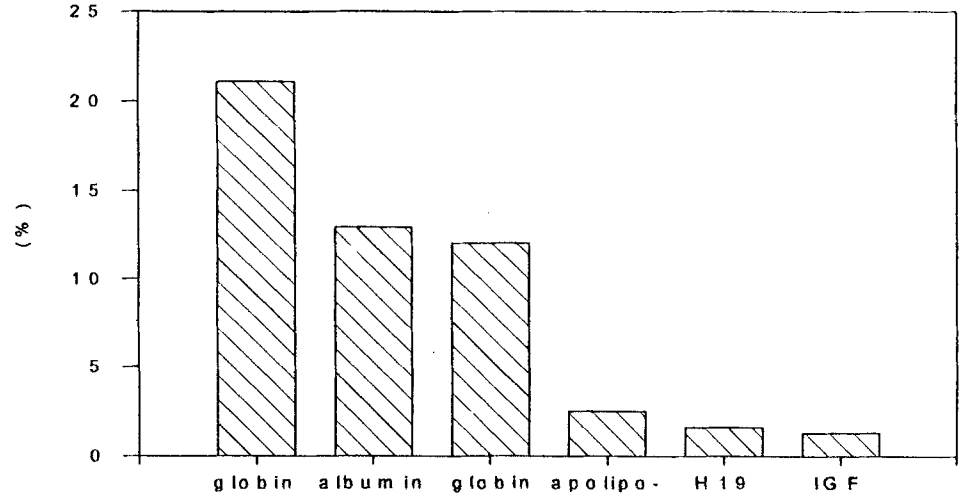

Fig. 2. Analysis of 2,354 identified clones that show high redundancy
In the human fetal cDNA library, 500 γ-globin clones was most redundant(21.2%)and followed by 306 albumin clones(12.9%), 282 α-globin clones(12.1%), 88 apolipoprotein clones(2.5%), 55 H19(1.6%), 44 insulin growth factor(1.3%).

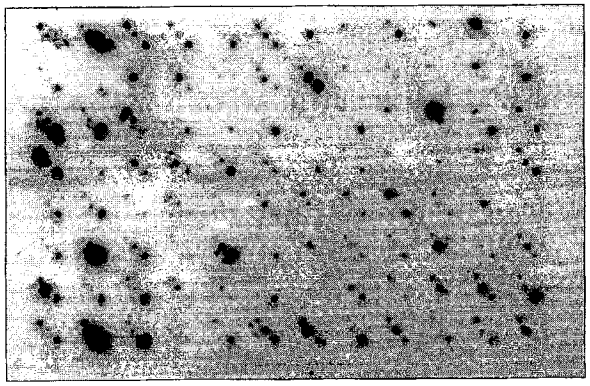

Fig. 3. Results of colony hybridization
262 clones were grown on NC paper on agar plates and followed by denaturation, and hybridization with ^{32}p-labelled prove, which they contain 4 redundant clones: γ-globin, albumin, α-globin and H19 and then exposure to X-ray film. after X-ray exposure, 76 spots had positive signal

Table 1. Classification of clones randomly sequenced and the clones prescreened prior to the sequence determination

classification	clones randomly sequenced		clones prescreened	
clones identified	68.5%	(2,354)	66.1%	(123)
clones with no match	20.0%	(690)	24.7%	(46)
EST	6.3%	(216)	5.9%	(11)
repeating sequences	5.4%	(187)	3.2%	(6)
Total	100.0%	(3,437)	100.0%	(186)
redundant clones	33.2%	(1,143)	12.9%	(34)

* Redundant clones are α & γ-globin, H19 RNA and serum albumin
* Number of clones classified is shown in parentheses

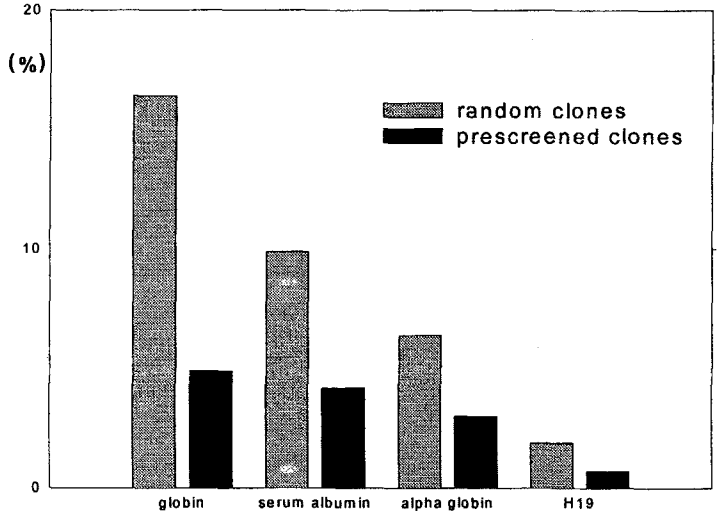

Fig. 4. The comparision of redundant clones between random clones and prescreened clones
To remove redundant clones prior to random sequencing determination, most redundant 4 clones including γ-globin, albumin, α-globin and H19 were commited to hybridized to randomly selected 262 cDNA clones. Out of 91 total redundant clones, 57 clones were removed after prescreening. In the case of gamma globin, 30 of 43 clones were removed

the identified clones out of the total number of informative clones was slightly lower than the composition determined by the random sequencing study. This reduction reflects the elimination of redundant clones. After prescreening, the ratio of redundant clones to the total number decreased by 10.3% when compared with the results of random sequencing as shown in Table 2.

Table 2. The ratio of 4 redundant clones in randomly sequenced and prescreened clones

redundant clones	number of clones (w/o prescreening)	number of clones (w/ prescreening)
gamma globin	14.5%(500)	4.9%(13)
serum albumin	8.9%(306)	4.2%(11)
alpha globin	8.2%(282)	3.0%(8)
H19	1.6%(55)	0.7%(2)
total	33.2%(1143/3437)	12.9%(34/262)

Out of 76 clones which gave positive signals to the four probe DNAs, 57 clones (75.0%) were confirmed to be the probe DNAs by nucleotide sequencing as shown in Table 3. These results suggested that the analysis of noble genes from a specific tissue could be carried out more rapidly and efficiently by eliminating the redundant clones using the colony hybridization assay .

The colony hybridization assay employing the cDNA mixture composed of 34 redundant clones as the probe was not effective for the increment of the efficiency in the screening of novel genes due to the high ratio of false positive signals (data not shown). This high level of false positive signals could be a result from the non-specific binding between the probes and the test cDNA sequences. For the efficient elimination of redundant clones before the sequence determination by the colony hybridization assay, the proper number of redundant cDNAs being used as the probe DNAs was one of the critical factors. In other studies performed by Duguid and Fargnoli, the mixtures of labeled total cDNAs prepared from various tissues were used as the probes in the differential cDNA screening (Duguid, 1990, Fargnoli, 1990). This strategy reduced the population of highly represented sequences in a cDNA library by selectively removing the common sequences shared by other libraries. With this strategy, the probability of isolating novel cDNA clones increased from 12% in the randomly sequenced group to 84% in the group modified by the differential cDNA screening method (Hoog, 1991).

In this study, a simple strategy to remove the highly redundant cDNA clones and to improve the probability of isolating novel genes was tested for the applification of a large scale human fetal liver cDNA sequencing. Using the proper number of four highly redundant cDNA clones instead of

Table 3. Analysis of 76 positive signalling clones in colony hybridization assay

clones		N. of clones	%
redundant clones		57	75%
	gamma globin	(30)	
	albumin	(15)	
	alpha globin	(9)	
	H19 RNA	(3)	
other clones		19	25%
		67	100%

total cDNA as the probe, we could effectively remove the most redundant clones in human fetal liver cDNA libraries without losing rare cDNA species due to the high level of false positive signals caused by employing an excessive number of cDNA species as the probe. This method is simple and convenient in increasing efficiency in isolating the novel genes and very useful for the screening of rare cDNA species.

REFERENCES

1. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulous, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McComb ie, W. R. and Ventor, J. C. (1991) Science. 252: 1651-1656.
2. Alting-Meese, M. A. and Short, J. M. (1989) Nucleic Acids Res. 17: 9494
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. 215: 403-410
4. Boehringer Mannheim (1993) in the DIG system user's guide for filter hybridization pp. 33-42, Boehronger Mannheim GmbH, Biochemica. Germany
5. Duguid, J. R. and Dinauer, M. C. (1990) Nucleic Acids Res. 18: 2789-2792
6. Hoog, C. (1991) Nucleic Acids Res. 19: 6123-6127.
7. Kim, J. W., Song, J. C., Lee, I. A., Lee, Y., Nam, M. S., Hahn, Y., Chung, J. H. and Choe, I. S. (1995) Kor. J. Biochem. Mol. Bio 28(5): 402-407
8. L. Diatchenchko, Y. F. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov and P. D. Siebert. (1996) Proc Natl. Acad. Sci. U. S. A. 93: 6025-6030
9. S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov and P. D. Siebert. (1996) Proc. Natl. Acad. Sci. U. S. A. 93: 6025-6030
10. Liew, C. C., Hwang, D. M., Fung, Y. W., Laurenssen, C., Cukerman, E., Tsui, S. and Lee, C. Y. (1994) Proc. Natl. Acad. Sci. U. S. A. 91: 10645-10649
11. Lin, X., Feng, X. H. and Watson, J. C. (1991) Proc. Natl. Acad. Sci. U. S. A. 88 : 6951-6955
12. Sasaki, Y. F., Ayusawa, D. and Oishi, M. (1994) Nucleic Acids Res. 22: 987-992
13. S. S. Choi, J. W. Yun, E. K. Choi, Y. G. Cho, Y. C. Sung and H. S. Shin. (1995) Mamm. Genome 6: 653-657
14. W. deleersnijder, G. Hong, R. Cortvrindt, C. Poirier, P. Tylzanowski, K. Pittois, E. Van Marck and J. Merregaert. (1996) J. Biol. Chem. 271: 19475-19482