2954–2958 Nucleic Acids Research, 1995, Vol. 23, No. 15

Identification of differentially expressed genes by restriction endonuclease-based gene expression fingerprinting

Natalia B. Ivanova[1] and Alexander V. Belyavsky[1,2,*]

[1]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilov Str. 32, 117 984 Moscow, Russia and [2]New York Blood Center, 310 East 67 Street, New York, NY 10021, USA

Received April 7, 1995; Revised and Accepted June 30, 1995

EMBL accession nos X83884–83886 and X84016

ABSTRACT

A novel method for identification of differentially expressed genes has been developed. It is based on the consecutive restriction digestions of 3′ terminal cDNA fragments to produce a fingerprint of gene expression. cDNA molecules are synthesized using a biotinylated oligo(dT) primer, digested with a frequently cutting restriction endonuclease and the 3′-terminal restriction fragments are isolated using streptavidin microbeads. After amplification by PCR, cDNA fragments are immobilized again on streptavidin beads, radiolabeled and treated sequentially with a set of restriction endonucleases. The products of individual enzymatic reactions from two or more different RNA populations are resolved by polyacrylamide gel electrophoresis and compared to reveal differentially expressed genes. This strategy enabled us to identify and clone the fragments of five genes expressed differentially in murine thymus and spleen. One of the genes was found to encode terminal deoxynucleotidyl transferase; others are apparently previously unknown genes.

INTRODUCTION

Differential gene expression underlies many fundamental biological processes such as embryo- and organo-genesis, cell and tissue differentiation, long-term plasticity of the nervous system and cellular response to various stimuli. Therefore, the development of powerful methods for identification and cloning of differentially expressed genes is critical for studying these processes.

The classical approach to isolation of differentially expressed genes is the differential screening of a cDNA library with labeled probes derived from two or more different mRNA samples. Subtractive hybridization can be applied to enrich for sequences which are unique for one cell type (1–3).

Recently, a method for RNA characterization by an arbitrarily primed reverse transcription-coupled PCR (AP RT–PCR) was described (4,5). Patterns of gene expression are produced by low-stringency PCR with short, arbitrarily selected primers, followed by analysis of the products by high-resolution polyacrylamide gel electrophoresis (PAGE). Here we present an alternative approach to analysis of mRNA expression. It is based on the creation of non-overlapping sets of 3′ terminal cDNA restriction fragments which are resolved by high-resolution PAGE. The procedure gives rise to highly informative, fingerprint-like patterns, which can be used for comparative analysis of ensembles of expressed genes, as well as for cloning of differentially expressed genes. The utility of this approach, which we termed 'gene expression fingerprinting' (GEF), is illustrated in the present study by cloning five sequences expressed differentially in murine thymus and spleen.

MATERIALS AND METHODS

All commonly used DNA and RNA manipulations were performed according to (6). Enzymatic reactions were carried out as recommended by manufacturers, if not otherwise indicated.

RNA isolation

Total RNA was prepared from murine thymus and spleen by the method of guanidine thiocyanate-acid phenol extraction (7). Poly(A)$^+$ RNA was prepared by oligo(dT) cellulose chromatography (8).

cDNA synthesis

First strand cDNA synthesis was carried out using 200 U Superscript reverse transcriptase (Gibco-BRL, USA), 30 U RNasin ribonuclease inhibitor (Promega, USA), 10 pmol Bio-T_{13}-primer (5′ biotin-GGGAGGCCCTTTTTTTTTTTT) and 5 μg total RNA in a 20 μl reaction. Unreacted Bio-T_{13}-primer was removed by two cetyltrimethylammonium bromide precipitations (9). cDNA:RNA hybrids were tailed using 50 μM dGTP and 25 U terminal transferase (Boehringer Mannheim). After phenol extraction and isopropanol precipitation, the second cDNA strand was synthesized by 1.5 U Bio-Taq DNA polymerase (Biomaster, Russia) in 25 μl reaction containing 10 pmol C_{13}-primer (5′-AAGGAATTCCCCCCCCCCCC). The temperature profile of reaction was as follows: 1.5 min at 98°C, 1.5 min at 60°C and 20 min at 72°C.

* To whom correspondence should be addressed at: The Engelhardt Institute of Molecular Biology, Moscow, Russia

Generation of 3′-terminal cDNA fragments

Double-stranded cDNA was digested with 4 U *Sau*3A I or *Bam*HI (New England Biolabs, USA) for 60 min at 37°C. Biotinylated cDNA fragments were immobilized on MagneSphere streptavidin microbeads (Promega, USA) according to the manufacturer's recommendations, washed with 150 mM NaCl, 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA and ligated with 3 pmol of double-stranded adaptor:

primer #1 3′CAAAAAACTTCGAACCTCGGGTG-5′
primer #2 5′-GATCGTTTTTTGAAGCTTGGAGCCCAC-3′

using 0.2 U T4 DNA ligase (Boehringer Mannheim, Germany) in a 10 µl reaction.

After removal of unligated adaptor by washing beads with 150 mM NaCl, 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, cDNA fragments were amplified by PCR using Bio-Taq DNA polymerase and Bio-T13- and #1 primers. The reaction mixture was incubated for 1.2 min at 98°C, 1.5 min at 55°C and 3.5 min at 72°C for a total of 15 cycles.

The amplified cDNA was adsorbed onto streptavidin microbeads, treated with 100 mM NaOH to remove non-biotinylated strands and washed with 10 mM Tris-Cl pH 8.0, 0.2 mM EDTA. Three pmolH of #3 primer (5′-GTGGGTCCAAGCTTC) in a total volume of 10 µl were annealed to the immobilized strands, elongated for 5 min at 20°C with 10 µCi ^{32}P-dATP (3000 Ci/mmol) and then for 20 min at 37°C with cold dNTPs using 2 U Sequenase (US Biochemicals, USA). Streptavidin beads with immobilized fragments were treated sequentially with 2 U each of *Eco*RV, *Pst*I (Promega, USA), *Msp*I and *Hin*PI (New England Biolabs, USA) restriction endonucleases for 60 min at 37°C in a 10 µl reaction mix. Fragments released after each restriction digestion were collected, denatured and resolved by electrophoresis in 5% denaturing TBE–Urea polyacrylamide gel.

Cloning of the fragments of differentially expressed genes

Isolation of cDNA fragments was performed as described in (10). cDNA fragments were oligo (dG)-tailed by terminal transferase and amplified by PCR with C_{13}- and #1 primers as described above. Amplified fragments were purified by agarose electrophoresis, treated with *Sau*3A I and *Eco*RI enzymes and ligated into *Bam*HI, *Eco*RI-cleaved pUC18 vector.

Northern blot analysis

Samples of 2 µg of poly(A)$^+$ RNA from murine spleen and thymus were resolved by 1% formaldehyde agarose gel electrophoresis, transferred onto Hybond-N$^+$ nylon membrane (Amersham, UK) and probed with ^{32}P-labeled inserts from recombinant plasmids carrying cloned cDNA fragments according to manufacturer's protocol. After hybridization, blots were washed 3× at 65°C with 0.5× SSC, 0.4% SDS.

RESULTS AND DISCUSSION

The GEF protocol

The scheme of the proposed approach is shown in Figure 1. Double stranded cDNA is synthesized using 5′-biotinylated oligo(dT)-containing primer and digested with a frequently cutting restriction endonuclease. By selecting only 3′ terminal cDNA fragments via immobilization on streptavidin microbeads,

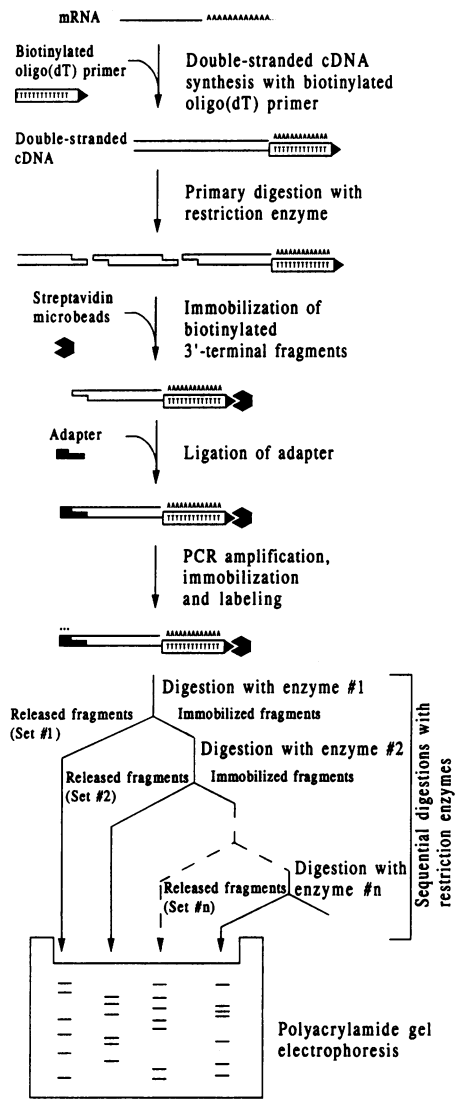

Figure 1. Scheme of the gene expression fingerprinting protocol used in the present study.

the complexity of the cDNA fragment pool is reduced and each mRNA species is represented by not more than one restriction fragment of the specific length and sequence. Immobilized fragments are ligated to adaptor, amplified by PCR with biotinylated T^{13} and adaptor primers, immobilized onto streptavidin microbeads at the biotinylated end and highly radiolabeled at the other end. Streptavidin beads with immobilized cDNA fragments are treated sequentially with different restriction enzymes and liberated labeled fragments are collected after each digestion. This consecutive enzymatic treatment accomplishes two goals simultaneously: (i) to generate labeled fragments of discrete lengths (immobilized fragments may have multiple 3′ ends due to variable positioning of oligo(dT)-primer on the mRNA poly(A) tails during the first strand synthesis); (ii) to split the complex set of cDNA fragments into subsets of less complexity. The products of individual enzymatic reactions are analyzed by PAGE and fingerprints from different mRNA samples are compared. For cloning, cDNA fragments correspon-

ding to differentially expressed genes are eluted from the gel, oligo (dG)-tailed and amplified by PCR with oligo(dC) and adaptor primers.

Testing the feasibility of the approach

To exemplify the validity of the proposed scheme, we analyzed RNA preparations from murine (line C57BL/6) thymus and spleen. The primary restriction digestion was performed with *Sau*3A I restriction enzyme. cDNA fragments for PAGE separation were generated by sequential cleavages by *Eco*RI, *Pst*I, *Msp*I and *Hin*PI enzymes. The corresponding fingerprints are highly specific for each secondary enzyme used (Fig. 2). Primary restriction digestion with *Bam*HI instead of *Sau*3A I resulted in quite different fingerprints (compare lines 2 and 5, Fig. 2). The number of differential bands (with an intensity difference of more than 3-fold between the two tissues) reached 11% of total bands for thymus and 8% for spleen.

Characterization of differentially expressed mRNAs

Four differential bands (arrows 1–4 in Fig. 2) were examined for differential gene expression. Cloned inserts were sequenced and used as probes for Northern blot hybridizations with poly(A)$^+$ RNA from thymus and spleen. Band 1 after cloning gave rise to two unrelated sequences, 1.1 and 1.2. On the Northern blots, both clones detected hybridizing mRNA species in the thymus, but not in the spleen sample (Fig. 3A and B respectively). Hybridization of the Northern blot with the cloned fragment 2 demonstrated that the corresponding mRNA was approximately 4- to 6-fold more abundant in the spleen than in the thymus (Fig. 3C). Fragment 3 was found to be identical to the 3′ terminal region of terminal deoxynucleotidyl transferase mRNA, and its expression pattern (Fig. 3D) was in a good agreement with the published data (11). Hybridization of the cloned fragment 4 with the RNA blot failed to reveal any discernible signal. To verify the expression pattern of the corresponding mRNA, the RT–PCR using (T)-primer and fragment 4-specific primer was performed with total RNA from spleen and thymus. Southern blots of amplified cDNAs were hybridized with another non-overlapping fragment 4-specific oligonucleotide. Hybridization signal corresponding to the fragment ~0.5 kb long was detected in the thymus, but not in the spleen sample (data not shown). Therefore, the presumed expression pattern was confirmed for all fragments which were taken for cloning.

With the exception of the fragment 3, no significant homology for the cloned sequences was found in the EMBL Data Library (Release 38). The sequences of fragments 1.1, 1.2, 2 and 4 were deposited in the EMBL Data Library under accession nos X83884, X83885, X83886 and X84016 respectively.

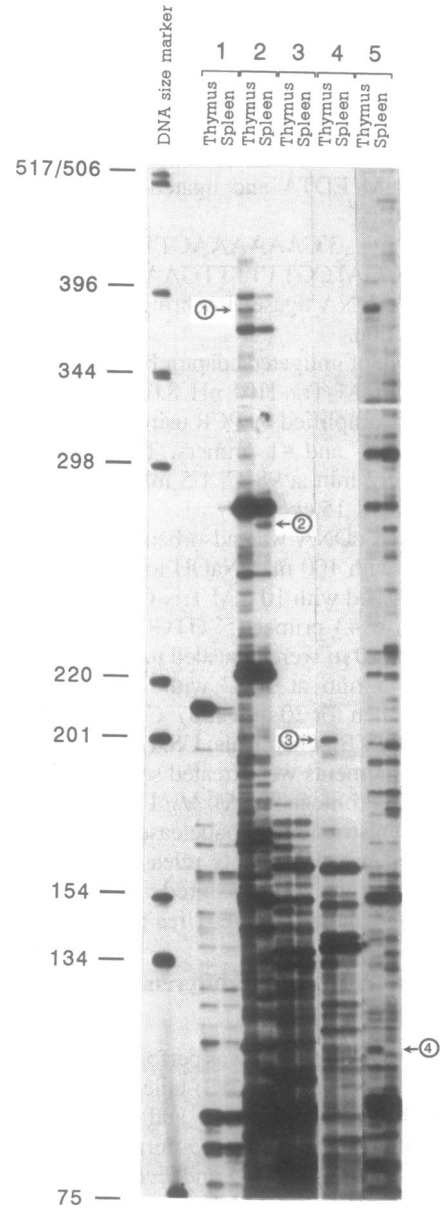

Figure 2. Fingerprint analysis of mRNA from murine thymus and spleen. Primary cleavages were performed with either *Sau*3A I (lane pairs 1–4) or *Bam*HI (5) enzymes. Secondary restriction digestions were performed sequentially with *Eco*RI (1), *Pst*I (2,5), *Msp*I (3) and *Hin*PI (4) enzymes. Fragments indicated as 1–4 were examined for differential gene expression.

The majority of mRNAs are involved in fingerprinting

To estimate the fraction of mRNA lacking appropriate restriction sites and therefore escaping the analysis, we performed a computer analysis of murine mRNAs with completely sequenced 3′ termini [a total of 659 mRNA sequences from the EMBL Data Library (Release 38)]. The length distributions of cDNA fragments after the primary digestion for *Sau*3A I and *Bam*HI enzymes (four and six base recognition sites respectively) are shown in Figure 4A. In the case of *Sau*3A I endonuclease, 93% of cDNAs contain cleavage sites within 1000–1500 bp from the site of polyadenylation, so that the corresponding fragments are within the range of efficient PCR amplification. In the secondary restriction digestion step, ~96% of the *Sau*3A I fragments can be cleaved with a small number of restriction endonucleases (8–10 enzymes) with four base recognition sites (Fig. 4B). This percentage might reach 100% if the oligo(dT)-primer used for the first strand synthesis is modified to include a restriction site for the last endonuclease in the set and one or two degenerate bases at the 3′ end to anchor the primer to the start of the poly(A) tail.

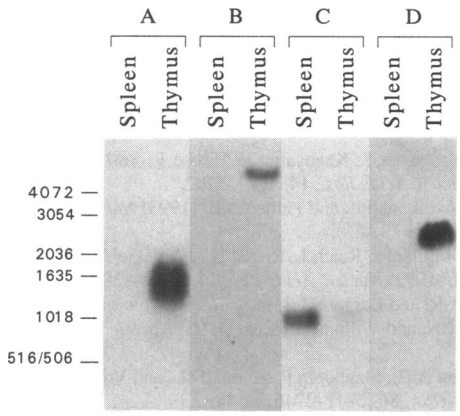

Figure 3. Northern blot analysis of poly(A)+ RNA isolated from murine thymus and spleen. Equal amounts of RNA (2 µg) were loaded onto each of the lanes. Blots were hybridized with ^{32}P-labeled cloned fragments 1.1 (**A**), 1.2 (**B**), 2 (**C**) and 3 (**D**). Positions of single-stranded DNA size marker fragments (1 kb ladder, Gibco-BRL, USA) are indicated on the left side.

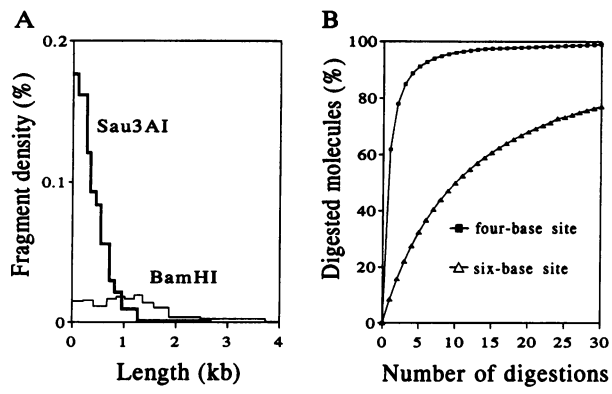

Figure 4. (**A**) Computer-simulated length distribution of cDNA fragments after primary restriction digestion by *Sau*3A I or *Bam*HI enzymes. The density of fragments is defined as a percentage of cDNA molecules cut within a 10 base pair interval at a given distance from the 3′ end, divided by 10. Integration over all fragment lengths gives the fraction of mRNA involved in fingerprinting. It is 93% of total mRNA for *Sau*3AI and 34% for *Bam*HI restriction enzymes. (**B**) Fraction of cleaved 3′ terminal cDNA fragments plotted as a function of number of sequential secondary restriction digestions (cases for enzymes with four- and six-base recognition sites are shown). Computations were made on the basis of length distribution for the *Sau*3A enzyme shown in Figure 4A. We assumed all frequencies of cleavage to be equal to 1/256 (four base pair recognition sites) or 1/4096 (six base pairs recognition sites) regardless of base content.

In contrast to four-base recognition enzymes, six-base recognition enzymes produce less complex subsets of fragments (Fig. 4B). The optimal set of enzymes for secondary digestions will therefore contain a few enzymes with six-base recognition sites followed by enzymes with five- and then with four-base sites. This allows production of more uniform fingerprints in terms of total number of bands per gel lane.

Comparison with other methods and further developments

Unlike hybridization methods, our approach does not impose a stage of hybridization with a very complex mixture of DNAs or RNAs. This feature may result in a considerable increase in the sensitivity limit for transcript detection. Another advantage of the proposed approach is that fingerprints from several mRNA populations can be easily compared side-by-side on one gel.

Compared to the AP RT–PCR method, the main advantage of the GEF approach is that a small number of consecutive enzymatic reactions is sufficient to have at least 80% and up to 93% of cDNA molecules involved in fingerprinting. Moreover, information from individual secondary restriction digestions is non-overlapping, provided that these reactions are complete. With AP RT–PCR, depending on the particular variant, hundreds or even thousands of amplification reactions are required for full-scale analysis of the population of mRNA molecules and the results of individual reactions, especially if exhaustive analysis is to be performed, will be overlapping to a significant extent.

Additionally, GEF may give more reliable and reproducible results. In the AP RT–PCR, up to 5% of the bands are not reproducible between duplicate samples (12), which is comparable with the percentage of genes expressed differentially between different tissues. Changes as small as 2°C in the annealing temperature may lead to a significant increase or decrease in the number of bands amplified by AP RT–PCR techniques (13). We believe that GEF approach results in a more robust procedure, since for the generation of cDNA fragments, highly specific and reproducible restriction enzyme reactions are used and PCR amplifications are performed under high stringency conditions. The fact that all sequences we cloned have confirmed their presumed specificity of expression argues strongly in favor of this assumption.

The scheme shown in Figure 1 is a single example of several possible variants of restriction nuclease-based gene expression fingerprinting. Synthesis of double-stranded cDNA can be performed in a number of ways, including a solid-phase synthesis approach (14,15). After the primary restriction digestion and adapter ligation, the population of cDNA fragments can be divided into subsets by PCR amplification with the modified oligo(dT)- and adapter primers containing one or two additional 3′ terminal bases; this strategy can also be combined with secondary restriction digestion as described above.

The major limitation of the present variant of the GEF approach is that the resolution of polyacrylamide gel electrophoresis rarely exceeds 300–400 bands, compared, for example, to 1000–2000 bands expected to be present in an average cDNA subset after 10 secondary restriction reactions. Therefore, an existing variant of the approach allows one to visualize the most abundant 15–30% of the mRNA population. Although the labeling efficiency of cDNA fragments is high enough to detect very rare transcripts, this potential can only be realized with the use of more efficient systems for DNA fragment separation. We therefore expect that a significant increase in the sensitivity and amount of information generated will result from the use of 2-D gel electrophoresis, which is capable of resolving up to 2000 fragments on a single gel (16,17).

ACKNOWLEDGEMENTS

We thank Drs Irina A. Fesenko and Igor B. Ivanov (Engelhardt Institute of Molecular Biology) for help in manuscript preparation and Drs Jan Visser and Trudy Pfaffenbach (The New York

Blood Center) for critical reading of the manuscript. This work was supported in part by the Russian Foundation for Fundamental Research (grant No. 94-04-11731a), The Netherlands Organization for the Advancement of Pure Research (NWO, grant No. 07-13-065), International Scientific Foundation (grant No. M2X000) and INTAS Foundation (grant No.93-3338).

REFERENCES

1 Sargent,T.D. (1987) *Methods Enzymol.*, **152**, 423–432.
2 Timberlake,W.E. (1980) *Dev. Biol.*, **78**, 497–510.
3 Hedrick,S.M., Cohen,E.A., Nielsen,E.A. and Davis,M.M (1984) *Nature*, **308**, 149–153.
4 Welsh,J., Chada,K., Dalal,S.S., Cheng,R., Ralph,D. and McClelland,M. (1992) *Nucleic Acids Res.*, **20**, 4965–4970.
5 Liang,P. and Pardee,A.B. (1992) *Science*, **257**, 967–971.
6 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
7 Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156–159.
8 Jacobson,A. (1987) *Methods Enzymol.*, **152**, 254–262.
9 Belyavsky,A.V., Vinogradova,T.V. and Rajewsky,K. (1989) *Nucleic Acids Res.*, **17**, 2919–2932.
10 Welsh,J., Petersen,C. and McClelland,M. (1991) *Nucleic Acids Res.*, **19**, 303–306.
11 Koiwai,O., Yokota,T., Kageyama,T., Hirose,T., Yoshida,S. and Arai,K. (1986) *Nucleic Acids Res.*, **14**, 5777–5792.
12 Liang,P., Averboukh,L. and Pardee,A.B. (1993) *Nucleic Acids Res.*, **21**, 3269–3275.
13 Bauer,D., Muller,H., Reich,J., Riedel,H., Ahrenkiel,V., Warthoe,P. and Strauss,M. (1993) *Nucleic Acids Res.*, **21**, 4272–4280.
14 Kimmel,A.R. and Berger,S.L.(1987) *Methods Enzymol.*, **152**, 307–316.
15 Lambert,K.N. and Williamson,V.M. (1993) *Nucleic Acids Res.*, **21**, 775–776.
16 Uitterlinden,A.G., Slagboom,P., Knook,D.L. and Vijg,J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2742–2746.
17 Hatada,I., Hayashizaki,Y., Hirotsune,S., Komatsubara,H. and Mukai,T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9523–9527.