912–914 Nucleic Acids Research, 1999, Vol. 27, No. 3

Systematic comparison of gene expression through analysis of cDNA fragments within or near to the protein-coding region

Youqiang Ke*, Chun Jing, Philip S. Rudland[1], Paul H. Smith and Christopher S. Foster

Department of Pathology, Faculty of Medicine and [1]School of Biological Sciences, Liverpool University, Liverpool L69 3BX, UK

Received November 5, 1998; Revised and Accepted December 10, 1998

ABSTRACT

Life is controlled by the timely and ordered expression of genes. Identification of important genes involved in specific physiological and pathological conditions requires efficient methods to analyse differential gene expression. We describe a novel strategy, namely complete comparison of gene expression (CCGE), for a systematic assessment of differentially expressed genes. Using the CCGE method, double-stranded cDNA is digested with two restriction enzymes that cut with different frequencies, the representative cDNA fragments are generated within or near to the protein-coding region. After being flanked by two different types of adapters, and amplified by a nested suppression PCR, the selected cDNA fragments, representing entire cDNA population, can be divided into 256 subsets; amplified and compared in a systematic manner.

Differential display (1) or mRNA fingerprinting (2) has been widely used to analyse the differentially expressed genes. However, the original technique suffered from the fact that only differences in relatively abundant mRNAs could be detected and many of those apparent differences were later found to be artifacts. Recently, several differential display-derivative methods have been developed to assess differential gene expression profiles in a more effective manner. However, these methods have been designed to analyse cDNA fragments from the 3′ ends of mRNAs (3–5), which are usually not in the protein-coding region. Although some of the recently developed strategies (6–10) may be employed to analyse other cDNA fragments, they may not have enough sensitivity for a more or less complete assessment of the entire cDNA species. Here we report a new approach, namely complete comparison of gene expression (CCGE), which compares the gene expression profiles between different types of cells by analysing the cDNA fragments within or near to the protein-coding region. Using the CCGE approach, a complete assessment of differential gene expression may be achieved through analysing all cDNA fragments, representing the entire cDNA population, in a systematic manner.

The details of the CCGE strategy are illustrated in Figure 1. After digestions with two restriction enzymes that cut with different frequencies, then ligated to two different types of adapters, and amplified by suppression PCR, the selected cDNA fragments can be divided into 256 subsets, amplified and separated in a systematic manner. To avoid the symmetrical recognition sequence of a chosen enzyme being included in the inner primer, RE1 and RE2 should be chosen preferably from the class IIS enzymes, which have cleavage sites at a precise distance from the recognition sites. To reduce the number of possible permutations of overhanging bases and hence the number of adapters needed to tag the entire cDNA fragments, the ideal choice for RE1 and RE2 should be those class IIS enzymes, such as *Ple*I (GAGTCN4/5) and *Mnl*I (CCTCN7/6), which generate a random one-base overhang.

Assuming the average size of cDNA in mammalian cells is ~1500 bp, a five-base recognition enzyme RE1 (step 1) cuts once every (4^5) 1024 bp on average; thus it is anticipated that most of the cDNA species will be cleaved by RE1 in step 1. A four-base recognition enzyme cuts once every (4^4) 256 bp on average; thus the RE2 will further cut each cDNA into an average of four shorter fragments. Therefore, it is expected that the majority of the cDNA fragments generated by RE1 should be cut at least once by RE2. However, it is possible that some cDNAs do not have a particular RE1 recognition site. Thus, one or two more rounds of analyses, using different RE1s, may be necessary for a complete assessment of all cDNA species.

In step 4, although all cDNA fragments generated by RE1 and RE2 may be tagged by A2 either at one or at both ends, only the outward fragment, within or near to the protein-coding region, may be amplified by PCR, because it has both A1 and A2 primer sequences attached. The A1 and A2 primer combinations cannot amplify the 3′ end fragments, which contain an A2 primer sequence at one end and a poly dT sequence at the other. For the cDNA fragments generated in the middle positions, which are 'tagged' by A2 at both ends, they cannot be amplified by the A2 outer primer (Fig. 1) because there is no complementary sequence in the opposite strand. Therefore, the nested PCR can amplify only one fragment from each cDNA molecule to form the template pool. The amplification of other cDNA fragments is suppressed by using the 'pseudo-double-stranded' adaptor (11).

To test the feasibility of CCGE, a benign rat epithelial cell line Rama 37 and three of its metastatic variants, which were generated by transfection of the parental Rama 37 cells with DNA extracted from malignant carcinoma cells (12,13), have been

*To whom correspondence should be addressed. Tel: +44 151 706 4515; Fax: +44 151 706 5859; Email: yqk@liv.ac.uk

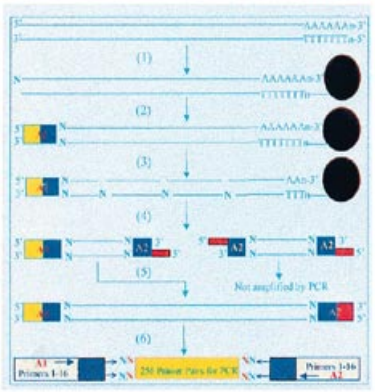

Figure 1. Schematic illustration of the detailed procedures of the CCGE strategy. N indicates A, T, G or C; n indicates 8–20. In step 1, the double-stranded cDNA is digested by a five-base recognition, class IIS restriction endonuclease (RE1) to generate cDNA fragments with a random one-base overhang. In step 2, the 3′ fragments are separated from other DNA residues by paramagnetic beads. A group of short DNA adapters (A1) is used to 'tag' the cDNA fragments at the 5′ end. A1 comprises four adaptors; each contains an outer (yellow) and an inner (blue) PCR primer sequence and a different 1 base (A, T, G or C) overhang. After the ligation reaction, the excess adaptors are removed. In step 3, the cDNA is cleaved by a four-base recognition, class IIS enzyme (RE2), to produce shorter fragments and to generate random one base overhangs at their ends. In step 4, another group of DNA adaptors (A2) is 'tagged' to the shorter cDNA fragments. A2 comprises four adaptors; each consists of a 'pseudo-double-stranded' DNA (11), with the longer strand containing an outer (red) and an inner (blue) primer sequence for PCR. A2 also has a random 1 base overhang (A, T, C or G) at its double-stranded end, which complements the overhangs at the end of the cDNA fragments. In step 5, the cDNA fragments are amplified by PCR, using both outer primers included in A1 and A2. In step 6, the products from step 5 are used as a template pool to perform further PCR, using the inner primers included in both A1 and A2. By extending one more base at the 3′ ends of the inner primers, 16 (4^2) primers can be derived from each end of the cDNA, using all possible permutations of the two bases at their 3′ ends. This yields a total of (16×16) 256 possible combinations for the inner primers, and hence the entire cDNA fragments can be selectively divided into 256 subsets for PCR. Each subset of cDNA fragments amplified by PCR can be separated by electrophoresis in a denaturing polyacrylamide gel. Measuring the peak areas or the intensities of the bands can identify the differentially expressed genes in a systematic manner.

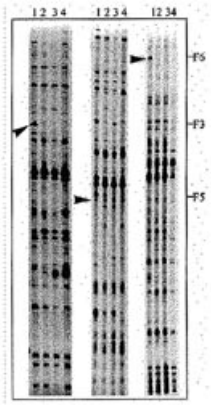

Figure 2. Analysis of differential gene expression between the benign and the metastatic cells. Three representative subsets of cDNA fragments obtained from CCGE PCR were subjected to PAGE. The dried gel was exposed to a Fuji RX X-Ray film for 48 h. Lane 1, the cDNA fragments originating from the benign Rama 37 cells; lanes 2–4; the cDNA fragments originating from the metastatic variants RMP2c-H, RMP2a-Lu and RMP2b-Lu, respectively. The arrowheads point to three bands which exhibit the most pronounced differences in their intensities between the benign and the malignant cells.

subjected to gene expression profile analysis. The RE1 used in this work was *Hph*I; RE2 was *Mnl*I. The nucleotide sequence of the upper strand of A1 was 5′-AACAAGCCACCGCCGCCA-CACACGCTGCTCACGCTGCTTCN-3′, where N was an overhanging base chosen from A, T, C or G, respectively. The lower strand of A1 was 5′-GAAGCAGCGTGAGCAGCGTGTGTG-GCGGCGGTGGCTTGTT-3′. The sequence of the longer strand of A2 was 5′-AAGCCGAATAACCCGCCTGTGATGCTCAT-GCTCACGGTCAN-3′, where N was an overhanging base selected from A, T, G or C, respectively. The short strand of A2 was a 20-base oligonucleotide complementary to the inner section of the longer strand: 5′-TGACCGTGAGCATGAGCATC-3′.

Total RNA was extracted from the benign parental Rama 37 cell line and three of its metastatic variants RMP2c-H, RMP2a-Lu and RMP2b-H (12) by the guanidinium–isothiocyanate method. The synthesis of double-stranded cDNA was performed according to the instructions supplied with the cDNA Synthesis Kit (Boehringer Mannheim Biochemica, Germany). We used 120 ng of total RNA from each cell line to start the CCGE. Ligations in steps 2 and 4 were conducted with 2 pmol of A1 and A2 in each reaction in a total volume of 20 µl at 16°C for 18 h, with 200 U of T4 ligase, 50 mM Tris–HCl (pH 7.4), 10 mM $MgCl_2$, 10 mM DDT, 1 mM ATP, and with 25 mg/ml BSA. The T4 ligase was inactivated at 65°C for 10 min. The nested PCR in step 5 was performed in a total volume of 20 µl, with 20 mM Tris–HCl (pH 7.4), 25 mM KCl, 0.05% Tween 20, 100 µg/ml BSA, 1 mM $MgCl_2$, 200 µM dNTPs, 200 nM of each of the primers, 1 U *Taq* DNA polymerase (Perkin Elmer, UK). The PCR consisted of 16 cycles at 94°C for 30 s, then at 65°C for 20 s and at 72°C for 30 s. At the end of the last cycle, the reaction was incubated at 72°C for 5 min. The PCR products were diluted into 1 ml, and 2 µl were taken from this diluted solution as template for the second round of PCR. The 'selective primer pairs' used in the PCR were defined by using the overhanging base and by extending a further base at the 3′ ends of both inner primers included in A1 and A2. The second PCR consisted of 40 cycles at 94°C for 30 s, then at 62°C for 30 s and at 72°C for 30 s. At the end of the last cycle, the reaction was incubated at 72°C for 5 min. The reagents and concentrations were the same as those used for the nested PCR, except that 100 nM of each primer, 2 µM dNTPs were used. In addition, 10 µCi of [^{35}S]dATP were incorporated in each reaction. PCR products (4 µl) and loading buffer (3 µl) were heated at 80°C for 2 min, and 5 µl of the mixture were subjected to electrophoresis in an 8% polyacrylamide gel, using a 60 cm gel tank (Bio-Rad, UK). The separated cDNA fragments were visualised by autoradiography. In this work, 10 of the 256 possible primer combinations were used to amplify the cDNA fragments. Nine cDNA fragments, which exhibited the most pronounced changes in their expression levels between the benign and the malignant cells, were removed from the denaturing gel and amplified by a further round of PCR (14). Eight of the nine fragments were successfully recovered and shown to have unique nucleotide sequences when determined in an automatic sequencer (Model 373, Applied BioSystems, US); two of them correspond to currently known genes and the remaining six are potentially novel. The representative expression profiles detected by three primer sets are shown in Figure 2. Arrows point to three bands, which exhibited the most pronounced differences between the benign and the malignant cells. The CCGE results were verified by using the slot blot

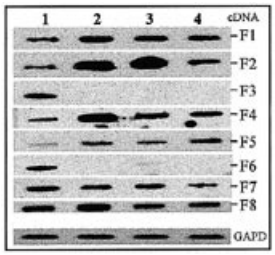

Figure 3. Measurement of gene expression by slot blot hybridisation. Nine sets of total RNA, each consisting of four samples (5 µg each), extracted from four different cell lines respectively (5), were loaded onto a nylon membrane (Hybaid-N, Amersham, UK) with a slot blot apparatus (Bio-Rad, UK). After the RNA was cross-linked with the membrane by exposure to UV light for 3 min, the membrane was cut into nine pieces, each containing a set of four RNA samples. Eight purified cDNA fragments, detected from eight separate panels of the CCGE electrophoretagrams (F3, F5 and F6 are shown in Figure 2; the other five fragments are not shown), were radioactively labelled and used as probes to hybridise to the eight sets of four RNA samples on the nylon membranes. A constitutively expressed GAPD cDNA was labelled in the same way and used to hybridise to the ninth set of RNA samples to standardise the hybridisation. Lane 1, the parental benign Rama 37 cells; lanes 2–4, the metastatic cell lines RMP2c-H, RMP2a-Lu and RMP2b-Lu, respectively.

hybridisation method to analyse the expression levels of the eight recovered cDNAs in the benign and the malignant cells. For seven of the eight cDNA fragments, the expression profiles detected by slot blot were very similar to those detected by CCGE (Fig. 3), whereas one cDNA fragment failed to show changes in hybridising mRNAs between the benign and the malignant cells, and therefore was probably a false positive (F8).

ACKNOWLEDGEMENTS

Y.K. would like to thank the North West Cancer Research Fund and the Research and Development Fund of Liverpool University for the support of a research grant.

REFERENCES

1 Liang,P. and Pardee,A.B. (1992) *Science*, **257**, 967–971.
2 Welsh,J., Chada,K., Dalal,S.S., Cheng,R., Ralph,D. and McCleland,M. (1992) *Nucleic Acids Res.*, **20**, 4965–4790.
3 Kato,K. (1995) *Nucleic Acids Res.*, **23**, 3685–3690.
4 Prashar,Y. and Weissman,S.M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 659–663.
5 Matz,M., Usman,N., Shagin,D., Bogdanova,E. and Lukyanov,S. (1997) *Nucleic Acids Res.*, **25**, 2541–2542.
6 Ivanova,N.B. and Belyavsky,A.V. (1995) *Nucleic Acids Res.*, **23**, 2954–2958.
7 Suzuki,H., Yaoi,T., Kawai,J., Hara,A., Kuwajima,G. and Watanabe,S. (1996) *Nucleic Acids Res.*, **24**, 289–294.
8 Kato,K. (1996) *Nucleic Acids Res.*, **24**, 394–395.
9 Habu,Y., Fukada-Tanaka,S., Hisatomi,Y. and Iida,S. (1997) *Biochem. Biophys. Res. Commun.*, **234**, 516–521.
10 Lopez-Nieto,C.G. and Nigam,S.K. (1996) *Nat. Biotechnol.*, **14**, 857–861.
11 Lukyanov,K.A., Matz,M.V., Bogdanova,E.A., Gurskaya,N.G. and Lukyanov,S.A. (1996) *Nucleic Acids Res.*, **24**, 2194–2195.
12 Ke,Y., Beesley,C., Smith,P., Barraclough,R., Rudland,P.S. and Foster,C.S. (1998) *Brit. J. Cancer*, **77**, 287–296.
13 Chen,H., Ke,Y., Oates,A.J., Barraclough,R. and Rudland,P.S. (1997) *Oncogene*, **14**, 1581–1588.
14 Wang,H., Xia,Z.B., Wang,Y.H., Li,C.H. and Sun,L.Y. (1998) *Mol. Biotechnol.*, **9**, 171–173.