

# Short Technical Reports

## Cloning of Deleted Sequences (CODE): A Genomic Subtraction Method for Enriching and Cloning Deleted Sequences

*BioTechniques* 31:788-793 (October 2001)

### ABSTRACT

*The deletion of specific genomic sequences is believed to influence the pathogenesis of certain diseases such as cancer. Identification of these sequences could provide novel therapeutic avenues for the treatment of disease. Here, we describe a simple and robust method called cloning of deleted sequences (CODE), which allows the selective cloning of deleted sequences from complex human genomes. Briefly, genomic DNA from two sources (human normal and tumor samples) was digested with restriction enzymes (e.g., BamHI, BglII, and BclI), then ligated to special linkers, and amplified by PCR. Tester (normal) DNA was amplified using a biotinylated primer and dNTPs. Driver (tumor) DNA was amplified using a non-biotinylated primer, but with dUTP instead of dTTP. After denaturation and hybridization, all the driver DNA was destroyed with uracil-DNA glycosylase (UDG), and all imperfect hybrids were digested with mung bean nuclease. Sequences deleted from the driver DNA but present in the tester DNA were purified with streptavidin magnetic beads, and the cycle was repeated three more times. This procedure resulted in the rapid isolation and efficient cloning of genomic sequences homozygously deleted from the driver DNA sample, but present in the tester DNA fraction.*

### INTRODUCTION

Different and successful approaches for subtraction at the cDNA level have been suggested and used (e.g., Reference 2). However, among all the genomic subtraction methods, only a modified variant called representational difference analysis (RDA) has produced reproducibly successful results (5,6). The

main idea behind RDA is the use of genomic subtraction for only a subset of genomic sequences (e.g., all *Bam*HI fragments less than 1 kb). Since the complexity of the genome is greatly reduced, the results of such modifications look promising. Another important point in RDA is that this method uses not only subtractive but also PCR kinetic enrichment, to purify restriction endonuclease fragments present in one DNA population, but not in the other. However, while having many advantages, RDA still has some limitations. The technique is complicated and susceptible to minor impurities. The differential product is usually between 250 and 350 bp in length, and this is not convenient for many applications (6). Another drawback is the relatively low productivity of RDA; only a few probes can be generated per experiment (5,6). The authors suggested that this limitation could be obviated by diminishing the number of rounds of hybridization/amplification or by increasing the complexity of representation. However, without simplification of driver and tester complexity, the method failed (5,7).

Recently, we developed a new procedure for the cloning of polymorphic (COP) sequences (4), and in this study we describe a new method for the rapid isolation of deleted genomic sequences that does not have all the limitations of RDA. The method is called cloning of deleted sequences (CODE) and is based on a modification of the COP procedure. Our major objectives were to develop a simple and reproducible protocol and to improve subtractive enrichment, thereby avoiding the excessive PCR kinetic enrichment steps that often generate small DNA products.

### MATERIALS AND METHODS

For the experimental procedures, we used DNA isolated from a small-cell lung carcinoma cell line, ACC-LC5. This cell line contains a homozygous 0.7 Mb deletion at 3p21.3–22 (3) and was used as a source for DNA A, the driver. DNA isolated from normal human lymphocytes was used as the control DNA (DNA B, the tester).

Two oligonucleotides, B1subtr1 5'-GATCCGCGGCCGCGGTCCTCCAAA-

AGGGTCAGTGCTGG-3' and B1subtr2 5'-CCCAGCACTGACCCTTTTGGACCGCGGCCGCG-3', were used to create the B1subtr1/2 linker. Annealing was carried out in a final volume of 100  $\mu$ L containing 20  $\mu$ L B1subtr1 (100  $\mu$ M), 20  $\mu$ L B1subtr2 (100  $\mu$ M), 10  $\mu$ L 10 $\times$  M buffer (Roche Molecular Biochemicals, Mannheim, Germany) and 50  $\mu$ L water. The reaction mixture was boiled for 8 min and allowed to cool slowly at room temperature. These oligonucleotides were selected because they were used successfully in previous work (4,10). Other oligonucleotides can be designed, for instance to clone *Hind*III fragments.

Two micrograms each of DNA A and DNA B, at concentrations of 50  $\mu$ g/mL, were digested with 20 U *Bam*HI, *Bgl*II, and *Bcl*I (Roche Molecular Biochemicals) at 37°C for 5 h. Upon completion of digestion, the enzymes were heat-inactivated for 20 min at 65°C.

Approximately 0.5  $\mu$ g each digested DNA was ligated overnight in the presence of a 50-M excess of B1subtr1/2 linker at room temperature. PCR of the tester product (DNA B with B1subtr1/2) was performed in 100  $\mu$ L of a solution containing 67 mM Tris-HCl (pH 9.1), 16.6 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.5 mM MgCl<sub>2</sub>, 0.1% Tween<sup>®</sup> 20, 200  $\mu$ M dNTPs, 100 ng tester product DNA, 400 nM biotinylated Pbsub primer (5'-biotin-CAGCACTGACCCTTTTGGGACC-3'), and 5 U *Taq* DNA polymerase. PCR of the driver product (DNA A with B1subtr1/2) was performed in 20 tubes using the Antiuniv primer (the same sequence as above but without biotin) and the following modifications: dUTP (300  $\mu$ M) was used instead of dTTP, and 2.5 mM MgCl<sub>2</sub> was used rather than 1.5 mM MgCl<sub>2</sub>. The PCR cycling conditions were 72°C for 5 min, followed by 25 cycles of 95°C for 1 min, 60°C for 1 min, and 72°C for 2.5 min, and a final extension at 72°C for 5 min.

All PCR-amplified DNA A samples were pooled (2 mL) and mixed with 20  $\mu$ L PCR-amplified DNA B (for subtraction, we used a DNA B:DNA A ratio of 1:100). The pooled sample was concentrated with ethanol to 100  $\mu$ L, purified using a JETquick<sup>™</sup> PCR Purification Spin Kit (Genomed GmbH, Bad Oeynhausen, Germany), and dissolved in 100  $\mu$ L water. This DNA mixture was fur-

# Short Technical Reports

ther concentrated with ethanol to 6  $\mu$ L and boiled for 10 min under mineral oil.

Subtractive hybridization was performed for 40 h in 9  $\mu$ L buffer containing 0.4 M NaCl, 100 mM Tris-HCl (pH 8.5), and 1 mM EDTA. After hybridization, the mixture was diluted to 200  $\mu$ L and extracted with an equal volume of chloroform:isoamyl alcohol (24:1) to

remove the mineral oil. Treatment with uracil-DNA glycosylase (UDG) (Roche Molecular Biochemicals) was performed in a buffer containing 70 mM HEPES-KOH (pH 7.4), 1 mM EDTA, and 1 mM dithiothreitol, with 30 U UDG at 37°C for 4 h. The DNA was then precipitated with ethanol and dissolved in 25  $\mu$ L TE buffer (10 mM Tris-

HCl, pH 8.0, 0.5 mM EDTA). To this, 3  $\mu$ L 10 $\times$  MBN buffer (30 mM sodium acetate, pH 4.6, 50 mM NaCl, 1 mM zinc acetate, and 0.001% Triton<sup>®</sup> X-100) and 20 U mung bean nuclease (Roche Molecular Biochemicals) were added, and the reaction was incubated at 37°C for 30 min. The reaction was stopped by the addition of EDTA to a final concentration of 1 mM.

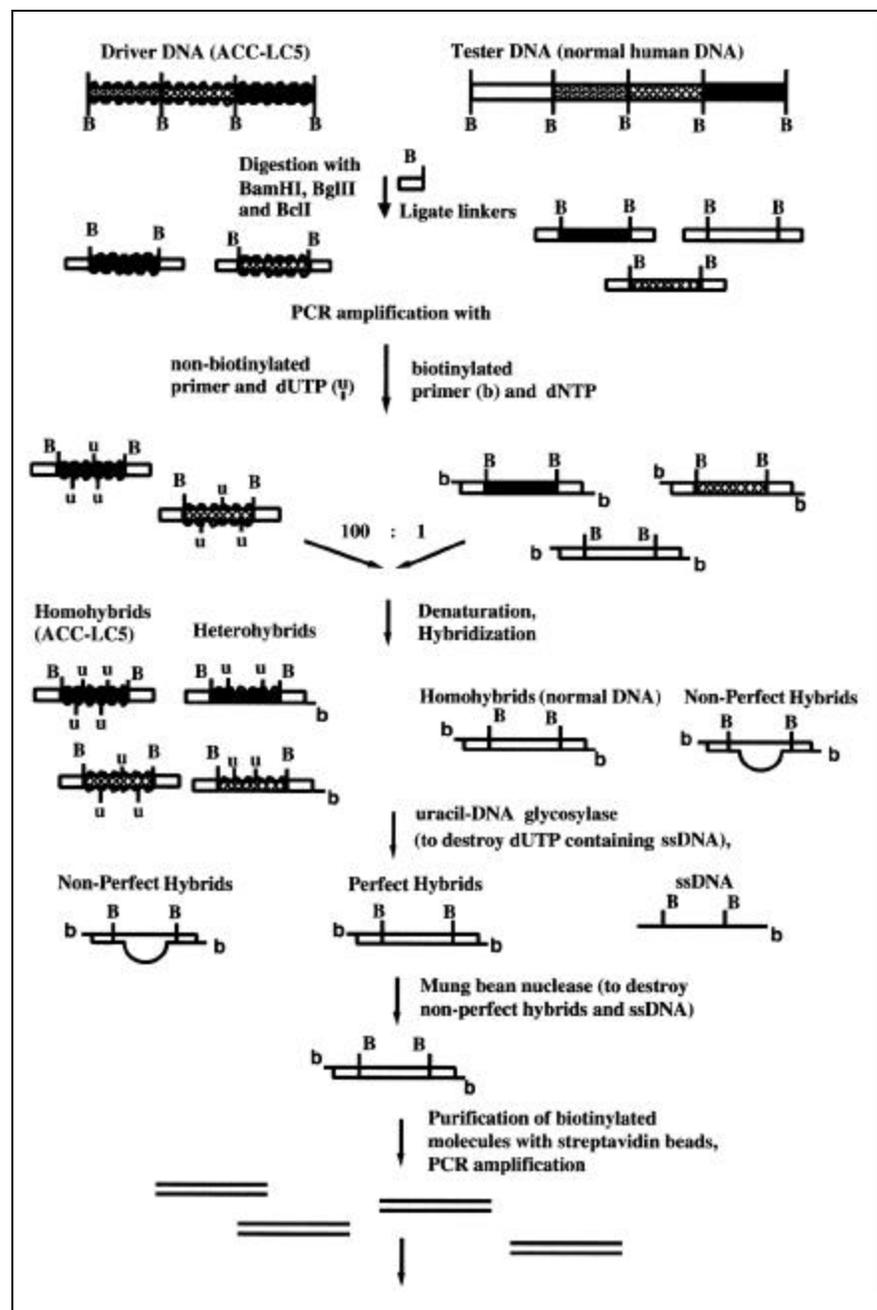
The subtracted DNA was purified with streptavidin-coupled Dynabeads<sup>®</sup> M-280 (DynaL A.S, Oslo, Norway) according to the manufacturer's instructions, and dissolved in 20  $\mu$ L TE buffer. Approximately 0.5  $\mu$ L of this DNA preparation was PCR-amplified, as described above for DNA B but using only 15 cycles, before subjecting the amplified DNA to a second round of hybridization. The whole procedure was repeated three more times (four cycles altogether).

The final subtraction product was PCR-amplified, purified with the JETquick PCR Purification Spin Kit, and digested with *NotI*. This DNA preparation was ligated into the pBC KS(+) vector (Stratagene, La Jolla, CA, USA), which had been pre-digested with *NotI* and dephosphorylated with alkaline phosphatase (Roche Molecular Biochemicals).

## RESULTS

An overview of the CODE procedure is shown in Figure 1 and described in Materials and Methods. The products of DNA amplification (1–2 kb) were denatured, hybridized, and treated with UDG (which destroyed all the driver DNA) and mung bean nuclease (which digested ssDNA and all the imperfect hybrids). The resulting tester homohybrids were concentrated with streptavidin beads and subjected to three more rounds of subtraction. The final PCR product was amplified and cloned into a suitable vector.

In this study, we have compared a lung tumor cell line, ACC-LC5, which contains a 0.7-Mb homozygously deleted region at 3p21.3–22, with normal control DNA. It is not known whether this cell line contains homozygous deletions on other chromosomes. This normal DNA is not a completely



**Figure 1. Flowchart diagram of the CODE procedure.** B, *Bam*HI or *Bgl*III or *Bcl*I; b, biotin; u, dUTP. Wavy lines designate DNA strands originating from the driver DNA. Homologous DNA fragments are marked in the same way. Mismatches in dsDNA hybrids are pinpointed by loops. They can be present in both homohybrids and heterohybrids.

# Short Technical Reports

appropriate control because it was isolated from another individual. Therefore, we should expect polymorphic sequences to be cloned, as well as those that are deleted.

Twenty-four random clones were tested by Southern blot analysis. Five of the clones were deleted in the tumor sample, five clones were not deleted, three clones failed to show a specific hybridization signal, and 11 were polymorphic. In Figure 2, several examples of Southern hybridization are shown. Analysis of the sequences from the deleted clones shows that at least two of them are from the 3p21.3–22 region deleted in ACC-LC5 (2). Other clones may represent deletion polymorphisms (6).

## DISCUSSION

These experiments demonstrate that the CODE method is a simple, effective, and robust procedure that can successfully isolate deleted genomic sequences. In contrast to RDA, we used three restriction enzymes in the CODE method to generate PCR products. This increased the complexity of the representations at least 3-fold. This method differs principally from RDA in that it

does not exploit the enrichment that arises from the difference between exponential and linear amplifications. This ensures that the biased influences generated by the PCR are kept to a minimum, and the subtractive enrichment becomes the most critical step.

Another important difference between RDA and RFLP subtraction methods (5,8) and the CODE method, which arises because of the simultaneous digestion with three enzymes, is that the CODE method allows the cloning of polymorphic probes that can detect larger or smaller DNA fragments in the driver DNA. The RDA and RFLP subtraction methods yield probes that detect only larger fragments in the driver DNA. Therefore, all polymorphic probes from RDA/RFLP subtraction experiments detected two alleles. The small allele (e.g., 0.6 kb) was always present in the tester but absent in the driver DNA, where the large DNA fragment (e.g., 7 kb) occurred. With the CODE method, probes that detect polymorphic fragments of similar lengths can be differentially cloned, as shown in Figure 2 (clones 6 and 7); see also Reference 4.

The general scheme of the CODE procedure is similar to the COP procedure (4). However, several important

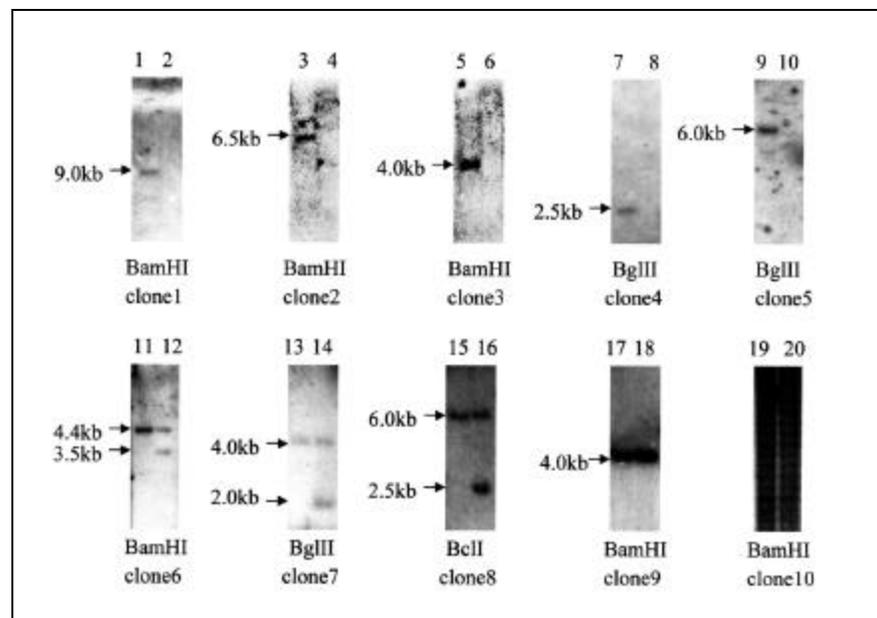
differences exist. To clone homozygously deleted sequences is a considerably more difficult task than to clone hemizygotously deleted polymorphic DNA fragments. We needed to increase the selectivity of the method at least 50-fold (6,8), and to minimize the reduction of complexity to maximize the number of different DNA fragments available for the subtraction. That is why we used fewer cycles for the PCR amplification of the subtracted products in the CODE procedure compared with the COP procedure. In the CODE procedure, the main force for differential cloning is the subtractive hybridization. This is why we used only two rounds of enrichment in the COP procedure, whereas in the CODE procedure we used four rounds.

A cDNA subtraction method using PCR amplification with dUTP and UDG treatment has been published recently (9). Our CODE procedure is significantly different from this cDNA subtraction method. Importantly, the subtraction efficiency of the Sugai et al. (9) method was very low (more than 90% of the clones were present in both mRNA preparations), despite the lower complexity of mRNA compared with genomic DNA.

The disadvantage of the CODE procedure is that it results in cloning DNA fragments that may or may not contain active components of genes (exons, introns, promoter regions, etc.). Deletions of nonfunctional regions occur frequently in the human genome (6). It will be important to demonstrate that the method can be used with enzymes containing CG pairs in their recognition sites, which are thus associated with CpG islands, and therefore genes (1). Our preliminary results show that the CODE method successfully works with *NotI* (Li et al., manuscript in preparation).

## REFERENCES

1. Bird, A.P. 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* 3:342-347.
2. Diatchenko, L., Y.F. Lau, A.P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov et al. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* 93:6025-6030.



**Figure 2. Southern analysis of genomic DNAs digested with different enzymes (shown in the figure).** The DNA was isolated from normal human DNA (odd lanes) and tumor DNA (even lanes). The genomic DNA was transferred to nylon filters and hybridized with inserts of the indicated clones. Clones 1–5 contain DNA fragments deleted in ACC-LC5. Clones 6–8 represent polymorphisms. Clone 9 is present both in the tester and driver DNA, and clone 10 contains repeated sequences.

3. **Ishikawa, S., M. Kai, M. Tamari, Y. Takei, K. Takeuchi, H. Bandou, Y. Yamane, M. Ogawa, and Y. Nakamura.** 1997. Sequence analysis of a 685-kb genomic region on chromosome 3p22-p21.3 that is homozygously deleted in a lung carcinoma cell line. *DNA Res.* 4:35-43.
4. **Li, J., F. Wang, V. Zabarovska, C. Wahlestedt, and E.R. Zabarovsky.** 2000. Cloning of polymorphisms (COP): enrichment of polymorphic sequences from complex genomes. *Nucleic Acids Res.* 28:e1.
5. **Lisitsyn, N., N. Lisitsyn, and M. Wigler.** 1993. Cloning the differences between two complex genomes. *Science* 259:946-951.
6. **Lisitsyn, N.A., N.M. Lisitsina, G. Dalbagni, P. Barker, C.A. Sanchez, J. Gnarr, W.M. Linehan, B.J. Reid, and M.H. Wigler.** 1995. Comparative genomic analysis of tumors: detection of DNA losses and amplification. *Proc. Natl. Acad. Sci. USA* 92:151-155.
7. **Lucito, R., J. West, A. Reiner, J. Alexander, D. Esposito, B. Mishra, S. Powers, L. Norton, and M. Wigler.** 2000. Detecting gene copy number fluctuations in tumor cells by microarrays analysis of genomic representations. *Genome Res.* 10:1726-1736.
8. **Rosenberg, M., M. Przybylska, and D. Straus.** 1994. "RFLP subtraction": a method for making libraries of polymorphic markers. *Proc. Natl. Acad. Sci. USA* 91:6113-6117.
9. **Sugai, M., S. Kondo, A. Shimizu, and T. Honjo.** 1998. Isolation of differentially expressed genes upon immunoglobulin class switching by a subtractive hybridization method using uracil DNA glycosylase. *Nucleic Acids Res.* 26:911-918.
10. **Zabarovska, V., J. Li, O. Muravenko, L. Ledorova, E. Bagra, I. Ernberg, C. Wahlestedt, G. Klein, and E.R. Zabarovsky.** 2000. CIS—cloning of identical sequences between two complex genomes. *Chromosome Res.* 8:1-7.

*This work was supported by research grants from Swedish Cancer Society, Pharmacia & Upjohn, Karolinska Institute, the Royal Swedish Academy of Sciences, Ingabritt och Arne Lundbergs Forskningss-tiftelse, and Åke Wiberg Foundation. Address correspondence to Drs. E.R. Zabarovsky and J. Li, MTC and CGR, Karolinska Institute, Box 280, S-171 77, Stockholm, Sweden. e-mail: eugzab@ki.se, jinli@ki.se*

Received 13 October 2000; accepted 12 April 2001.

**J. Li<sup>1</sup>, F. Wang<sup>1</sup>, V. Kashuba<sup>1</sup>,  
C. Wahlestedt<sup>1</sup>, and  
E.R. Zabarovsky<sup>1,2</sup>**

<sup>1</sup>*Karolinska Institute  
Stockholm, Sweden*

<sup>2</sup>*Russian Academy of Sciences  
Moscow, Russia*

## **Cost-Effective Method to Synthesize a Fluorescent Internal DNA Standard for Automated Fragment Sizing**

*BioTechniques* 31:793-800 (October 2001)

### **ABSTRACT**

*We describe a simple and cost-effective method for the synthesis of an internal fluorescently labeled DNA standard for fragment sizing using an automated DNA sequencer. A set of primer pairs labeled with ROX was developed to amplify 12 DNA fragments, 58–417 bp, derived from a conserved region of plant chloroplast DNA. These amplified fragments were mixed together, constituting a fluorescent internal DNA size marker. The precision of the size standard was evaluated by estimating the size of 20 alleles that were amplified at four dinucleotide microsatellite loci with the synthesized size standard and the commercial internal sizing standard, GeneScan® Rox-500. A number of intra-gel and inter-gel comparisons were run, and an analysis of variance was carried out. No significant difference was observed between the size estimates obtained with the synthesized DNA standard and the commercial standard. This facile and general PCR-based method for the synthesis of internal standards allows for significant savings in the implementation of large genotyping experiments using microsatellite or AFLP markers.*

### **INTRODUCTION**

The use of fluorescence detection and internal lane standards to size PCR products automatically (10) is an established procedure that is widely used in several areas of research in human, animal, plant, and microbial genetics, including individual identification, genetic mapping, and mutation detection.

High-throughput fluorescent genotyping requires a considerable amount of automation for the accurate and efficient generation of genetic marker data. Automated DNA sequencers and accompanying software products are