Isolation of cDNA clones using yeast artificial chromosome probes

P.Elvin*, G.Slynn, D.Black, A.Graham, R.Butler, J.Riley, R.Anand and A.F.Markham
Department of Biotechnology, ICI Pharmaceuticals, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

Received March 9, 1990; Revised and Accepted May 25, 1990

ABSTRACT

The cloning of large DNA fragments of hundreds of kilobases in Yeast artificial chromosomes, has simplified the analysis of regions of the genome previously cloned by cosmid walking. The mapping of expressed sequences within cosmid contigs has relied on the association of genes with sequence motifs defined by rare-cutting endonucleases, and the identification of sequence conservation between species. We reasoned that if the contribution of repetitive sequences to filter hybridisations could be minimised, then the use of large cloned DNAs as hybridisation probes to screen cDNA libraries would greatly simplify the characterisation of hitherto unidentified genes. In this paper we demonstrate the use of this approach by using a YAC, containing 180kb of human genomic DNA including the aldose reductase gene, as a probe to isolate an aldose reductase cDNA from a λgt11 human foetal liver cDNA library.

INTRODUCTION

Technological advances in molecular biology have allowed the separation of large DNA fragments, up to several megabases in size, by pulse field gel electrophoresis (PFGE) (1). More recently, the cloning of several hundred kilobase fragments as Yeast artificial chromosomes (YACs) has significantly advanced the feasibility of analysing large stretches of DNA (2). While linkage analysis of defined DNA markers allows the assignment of specific gene loci to areas comprising hundreds of kilobases of the genome, detailed mapping of large regions of DNA requires painstaking analysis, often involving the sub-cloning of manageable sized pieces of DNA in cosmid vectors. An example of the effort involved in such an undertaking was demonstrated by the analysis of the human major histocompatibility complex class III region. Chromosome walking in this region required 61 cosmid genomic inserts to cover 563kb DNA (3). Based upon the observation that some, but not all, genes are frequently associated with HTF-islands (4,5), eleven novel transcripts were identified by analysis of the cosmids from this 563kb region. A more recent example of the application of this powerful technology has been the cloning of the cystic fibrosis gene (6). This was hampered by the presence of regions of DNA that were apparently difficult to clone. Therefore, cloning of the cystic fibrosis gene required the construction and screening of many sub-libraries in cosmids and phage, mapping of the cloned fragments, and analysis of DNA regions conserved between species.

The use of large DNA fragments, such as those cloned into YAC vectors (2), as hybridisation probes would provide an alternative route to identify uncharacterised genes from a particular region of the genome. The experiments described herein, using a YAC comprising ~180kb of human genomic DNA and known to contain the aldose reductase gene, demonstrate the successful use of large cloned DNAs as hybridisation probes to screen cDNA libraries. In addition to identifying aldose reductase cDNAs we were able to rapidly isolate other transcripts encoded within this genomic fragment.

METHODS

Aldose reductase cDNA

A human aldose reductase cDNA probe was generated by the PCR using the amplimers A3: d(ATGGCAAGCCGTCTCCTGCTCAACAACGGC) and A4: d(GGCAAAGAGAAGTCTTGCTGAAAGGATTCC), which are found at the 5′ end of the coding sequence and at the end of the 3′ non-coding region of the cDNA respectively (7) (Fig 1). The PCR was carried out using 1μl of human foetal liver cDNA library in λgt11 (Clontech, 7.4×10^9 pfu ml^{-1}) as template, in a buffer comprising 10mM Tris HCl pH 8.3, 50mM KCl, 1.5mM $MgCl_2$, 0.01% gelatin, 200μM each dNTP, 100pmol of each amplimer, in a 100μl reaction volume. Samples were heated to 100°C for 5 minutes to denature the DNA and the PCR performed with 2 units Thermus aquaticus DNA polymerase (AmpliTaq, Cetus Corporation). The thermal cycles consisted of annealing at 65°C for 2 minutes, DNA synthesis at 72°C for 2 minutes and denaturation at 92°C for 2 minutes, for a total of 35 cycles. The 1300bp PCR product, generated by amplimers A3 and A4, was purified from a 1% low melting-point agarose gel.

Aldose reductase YAC

We screened a three genome equivalent human genomic DNA library of 40,000 clones in the vector pYAC4 (8) using a combination of hybridisation and PCR screening methods (9,10).

* To whom correspondence should be addressed

For hybridisation we used a probe generated from the aldose reductase cDNA (above). For PCR screening we used oligonucleotide primers from an intron of 325bp in the aldose reductase gene, to generate a 320bp product (Graham et al, unpublished observations). The PCR was carried out essentially as described in the text above. Screening identified a total of four recombinant clones, one of which (4FA1) was used to isolate the YAC by PFGE using a 'Waltzer' apparatus (11).

The clone 4FA1 was also used to construct a restriction map using partial digests of BssHII, NaeI, SfiI, and XhoI restriction enzymes. Restriction fragments were separated by PFGE on a 1.5% agarose gel at 150V/300mA with a pulse time of 25 seconds over a period of 24 hours. Fragment sizes and enzyme sites were determined from Southern blots (12) hybridised with pYAC4 end specific probes (2). The aldose reductase cDNA was localised in the restriction map by hybridisation of the partial digest blot to the cDNA probe.

Purification of YAC 4FA1

Yeast cells from a 10ml overnight culture were harvested as described previously (13) and used to prepare agarose plugs at a concentration of 2×10^8 cells per plug ($\sim2\mu$g DNA per plug). A total of 15 plugs, equivalent to approximately 600ng 4FA1, were loaded into a single long slot of a preparative PFGE gel as described (13). The gel comprised a 1.5% agarose support gel with a 1.0% Sea Plaque agarose fractionating gel. Following electrophoresis the YAC was excised from the gel, the agarose melted, and stored at 4°C in aliquots equivalent to $\sim$100ng each.

cDNA library screening

Approximately 2×10^4 recombinants from a human foetal liver cDNA library were plated onto E. coli Y1090 host cells on 9cm diameter LB agar plates. Two replica plaque lifts on Hybond-N (Amersham) were prepared from each plate according to the manufacturer's instructions, and hybridised with either YAC or cDNA probes. Plaques identified in a first round of hybridisation were removed from the plate using a sterile pasteur pipette and transferred to 200μl SM medium (14) containing 1μl chloroform, for storage at 4°C. The titre of these phage stocks was determined and appropriate dilutions prepared for replating at approximately 100 plaques per 9cm diameter plate. Duplicate plaque lifts were prepared for a second round of hybridisation with YAC and cDNA probes to select single aldose reductase positive plaques.

Probe preparation and hybridisation conditions

The cDNA or YAC DNA sequences were labelled with α ^{32}P dCTP (3000Ci mmol^{-1}, NEN) by the random priming method of Feinberg and Vogelstein (15). For the cDNA the labelling reaction (25ng DNA, 50μCi dCTP) was carried out for 30 minutes at 37°C; the YAC was labelled ($\sim$100ng DNA in agarose, 150μCi dCTP) for at least 16 hours at 37°C. Labelling reactions routinely resulted in specific activities of $\sim10^9$ cpm μg^{-1}.

To reduce the effect of vector and human repeat sequences in the hybridisation reaction the labelled YAC was denatured in a boiling water bath for 10 minutes in the presence of sheared human placental DNA (Type XIII, Sigma) and sheared pBR322 DNA. The vector and human sequences were then allowed to reassociate to between Cot 125–250 in 5× SSC at 65°C (16). Typically, reassociation was carried out in a final volume of 400μl with a probe concentration of 0.25μg ml^{-1}, 2.8mg ml^{-1} sheared human placental DNA and 50μg ml^{-1} pBR322 DNA.

Filters were prehybridised in a buffer containing 5× SSC, 5× Denhardts, 200μg ml^{-1} sheared salmon sperm DNA (Type III, Sigma), 0.1% SDS and 6% PEG 6000 for at least 6 hours at 65°C. For filters subsequently hybridised with the YAC probe, 100μg ml^{-1} denatured sheared human placental DNA was included in the prehybridisation reaction.

Hybridisations were carried out in 5× SSC, 2× Denhardts, 200μg ml^{-1} sheared salmon sperm DNA, 0.1% SDS, and 6% PEG 6000 with a probe concentration of 10^6 cpm ml^{-1} (cDNA) or $\sim2\times10^6$ cpm ml^{-1} (YAC) for approximately 16 hours at 65°C. Following hybridisation, filters were washed in 2× SSC, 0.1% SDS for 20 minutes, followed by 0.5× SSC, 0.1% SDS for 20 minutes at 65°C. The filters were then wrapped in Saran Wrap and exposed to Kodak X-AR film at −70°C for the times indicated.

Amplification and direct sequencing of positive λgt11 plaques

cDNA clones identified by hybridisation to 4FA1 were plaque purified and transferred to 200μl SM containing 1μl chloroform. From these phage stocks, 1μl was used in a PCR reaction using two λgt11 primers which flank the λgt11 EcoRI cloning site : d(GGTGGCGACGACTCCTGGAGCCCG) and d(TTGACACCAGACCAACTGGTAATG). Reaction conditions were as described previously except that annealing was carried out at 60°C. Amplified DNAs were desalted and excess dNTPs and primers removed by centrifuge-driven dialysis on a Centricon 100 microconcentrator (Amicon) (17). Direct sequencing of PCR products was performed as described previously (18).

RESULTS

The YAC, 4FA1, was shown to contain the entire sequence of the aldose reductase gene comprising approximately 10kb of genomic DNA. PFGE and Southern blot analysis allowed the construction of the map shown in Fig 1b. The aldose reductase coding sequences are contained in a 24kb BssHII fragment.

Initial cDNA library screening was carried out at a high plaque density of approximately 1000 plaques per 9cm diameter plate. The human genomic sequences of 4FA1 identified a total of 43 phage plaques, that were discernible from non-specific background signal. The aldose reductase cDNA, hybridised to duplicate filters, identified 33 positive plaques. A typical pair of duplicate filters are shown in Figure 2a and b. The number of positive plaques identified by each probe is shown in Table 1, approximately 10% of those plaques identified with the YAC probe were also shown to be positive with the aldose reductase cDNA. Subsequently, plaque purification of clones shown to hybridise to both the aldose reductase cDNA and 4FA1 was done by replating phage at a lower density of $\sim$100 plaques per plate. Duplicate filters hybridised to cDNA or YAC probes showed complete agreement in signals. Results typical of this second round of hybridisation screening are shown in Figure 2 c–e.

Single plaques identified by the probe 4FA1 in the second round of screening of the low density plates were picked as described previously. In order to test for the presence of aldose reductase sequences in these λgt11 recombinants, the PCR was carried out as described using the λgt11 primers. The sequence of the PCR product was determined by direct sequencing using an aldose reductase specific sequencing primer (Figure 3). We have also used the λgt11 PCR amplimers as sequencing primers and determined the sequence of the ends of the cDNA clone. Sequence data obtained from one such plaque is shown in Figure

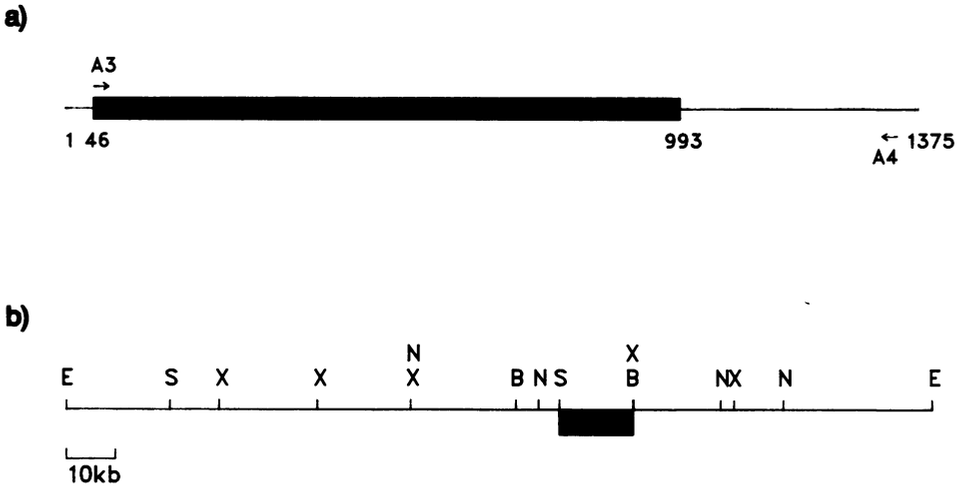

Figure 1. Schematic representation of the human aldose reductase cDNA and gene locus. a) Human aldose reductase cDNA (8). The coding region is shown in a solid box together with 45bp of 5′, and 382bp of 3′ untranslated sequence. The oligonucleotides A3 (nucleotides 46−75) and A4 (nucleotides 1316−1345) were used to generate a 1300bp cDNA. b) Human aldose reductase gene locus. The YAC 4FA1 contained the 180kb EcoRI genomic fragment which was mapped by partial restriction enzyme digest and PFGE. Sites indicated are BssHII (B), NaeI (N), SfiI (S), and XhoI (X). The solid box indicates the region containing the aldose reductase gene.

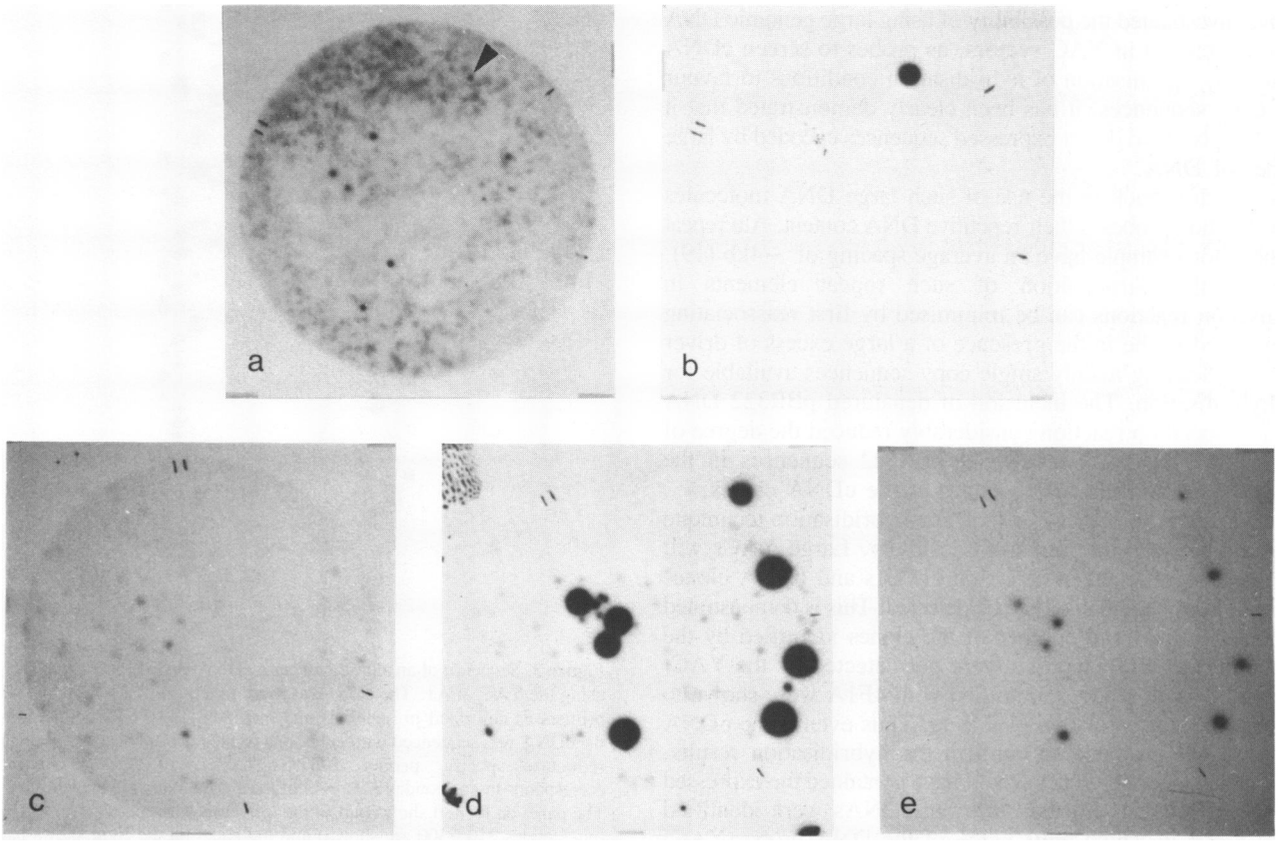

Figure 2. Duplicate filters from first (a and b) and second (c, d and e) rounds of hybridisation screening with YAC 4FA1, and the aldose reductase cDNA. In (a) eight cDNAs are identified with the YAC probe, the plaque shown by the arrow was aldose reductase positive when the duplicate filter was hybridised with the aldose reductase cDNA probe (b). When this area was replated and screened with the YAC (c) and the cDNA (d and e) probes, all of the plaques identified with the YAC were seen to be aldose reductase positive. All filters were washed to 0.5× SSC, 0.1% SDS at 65°C; exposure times at −70°C were (a) 3 days, (b) 5 hrs, (c) 5 days, (e) 5 hrs and (d) 20 minutes.

Table 1. Numbers of positively hybridising plaques per 9cm plate, detected in the first (high plaque density) round of screening. Duplicate filters were hybridised with either the aldose reductase cDNA or 4FA1 YAC probes. Where each probe identified the same recombinant, a coincident signal was recorded.

Plate No.	Probe YAC	Probe cDNA	Coincident signal	Plate No.	Probe YAC	Probe cDNA	Coincident signal
1	8	1	1	11	2	–	–
2	1	–	–	12	1	–	–
3	4	3	1	13	–	1	–
4	1	3	–	14	1	2	–
5	–	4	–	15	–	1	–
6	–	1	–	16	5	1	–
7	4	3	–	17	4	–	–
8	2	2	–	18	2	2	–
9	2	1	1	19	–	3	–
10	3	3	–	20	3	2	1

3, confirming the identity of the 4FA1 positive plaque as an aldose reductase cDNA.

Finally, twenty-two of the recombinants that were identified by the YAC probe, but did not hybridise to the aldose reductase cDNA, were picked from the high density plates as described. For subsequent analysis, the cDNAs were isolated by PCR using the λgt11 amplimers; PCR products were obtained from 14 of the 22 recombinants. By Southern blot analysis 4 of the 14 cDNAs were shown to hybridise to the YAC 4FA1 probe.

DISCUSSION

We have investigated the possibility of using large genomic DNA fragments, cloned in YAC vectors, as probes to screen cDNA libraries. By optimisation of hybridisation conditions to favour single copy sequences, it has been clearly demonstrated that it is indeed possible to detect expressed sequences encoded by large stretches of DNA.

A major drawback in the use of such large DNA molecules as hybridisation probes is their repetitive DNA content. Alu repeat sequences for example have an average spacing of ~4kb (19). However, the participation of such repeat elements in hybridisation reactions can be minimised by first reassociating the denatured probe in the presence of a large excess of driver DNA (16), leaving mainly single copy sequences available for filter hybridisation. The inclusion of denatured pBR322 DNA in the reassociation reaction considerably reduced the degree of background hybridisation between pBR322 sequences in the pYAC4 vector and phage sequences of the cDNA clones.

It is important to remember that this hybridisation technique is working close to the limits of sensitivity. Large YACs will in fact be less efficient in detecting cDNAs and cDNA clones with small inserts are unlikely to be detected. This is demonstrated by the fact that a large number of the clones identified by the aldose reductase cDNA probe were not detected by the YAC. However, 4 of 43 plaques identified with 4F1A were shown to be aldose reductase positive (Table 1). Thus even if the cDNA probe was not available to confirm the hybridisation results, ~10% of the selected clones would have contained the expressed sequence of interest. Furthermore the cDNAs were identified using a probe in which only ~1% of the DNA probe (YAC) represented the target sequence.

Some discrepancy in the numbers of clones detected by cDNA and YAC probes is not unexpected. A consideration of gene number, and the clustering of genes within the genome (20), might predict that a genomic probe covering 180kb would hybridise to more than one cDNA. The detection of cDNAs by the YAC which were not detected by the aldose reductase cDNA probe is thus not surprising. We have shown that Southern

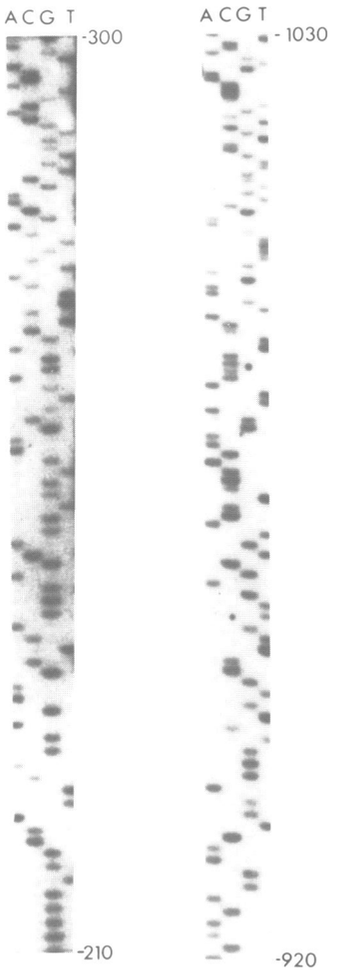

Figure 3. Sequence of an aldose reductase cDNA λgt11 clone which was isolated using the YAC 4FA1. The clone was used directly for PCR with the two λgt11 primers as described in materials and methods. The PCR product comprising the cDNA was sequenced with either one of the λgt11 primers, or with an aldose reductase specific primer d(GTGATCCCCAAGTCTGTGACA) which corresponds to nucleotides 823–843 of the aldose reductase cDNA sequence (8). The panel on the left shows part of the sequence obtained with the λgt11 primer (nucleotides 210–300 are shown) and the panel on the right shows part of the sequence obtained with the aldose reductase specific primer (nucleotides 920–1030 are shown).

blotting identified four cDNA clones that represented transcribed sequences in the YAC clone that were unrelated to aldose reductase, suggesting the existence of at least one other active gene in close proximity to the aldose reductase gene.

The aldose reductase gene described here comprises approximately 10kb of genomic sequence, and was shown to be contained in a 24kb BssHII fragment. However, for many large genes in which the coding sequences are spread over great distances, the general applicability of the technique described here may be limited by the spacing of coding sequences. It is conceivable that YAC clones may comprise intron sequences alone, or contain only part of a gene sequence poorly represented in a cDNA library. The successful application of the method may also depend on the relative abundance of target cDNAs in the library screened, and the fact the tissue specific gene expression may entail the screening of more than one cDNA library.

In summary, PFGE and YAC cloning technology is enabling large regions of the human genome to be characterised with respect to disease linked markers. The use of large cloned DNAs as probes to screen cDNA libraries will complement existing strategies for the identification of coding sequences, and allow researchers to move rapidly from uncharted areas of the genome to defective genes, without the need to construct and screen sub-genomic libraries.

REFERENCES

1. Schwartz, D.C. and Cantor, C.R. (1984) Cell **37** 67–75.
2. Burke, D.T., Carle, F.G., Olson, M.V. (1987) Science **236** 806–812.
3. Sargent, C.A., Dunham, I., Campbell, R.D. (1989) EMBO.J. **8** 2305–2312.
4. Lindsay, S. and Bird, A.P. (1987) Nature **327** 336–338.
5. Bird, A.P. (1986) Nature **321** 209–213.
6. Riordan, J.R., Rommens, J.M., Keren, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J., Drumm, M.L.,Iannuzzi, M.C., Collins, F.S., Tsui, L. (1989) Science **245** 1066–1073.
7. Graham, A., Hedge, P.J., Powell, S.J., Riley, J., Brown, L., Gammack, A., Carey, F., Markham, A.F. (1989) Nucleic Acids Res. **17** 8368.
8. Anand, R., Riley, J.H., Butler, R., Smith, J.C.S., Markham, A.F. (1990) Nucleic Acids Res. **18** 1951–1956.
9. Green, E.D., Olson, M.V. (1990) Proc. Natl. Acad. Sci. USA **89** 1213–1217.
10. Heard, E., Davies, B., Feo, S., Fried, M. (1989) Nucleic Acids Res. **17** 5861.
11. Southern, E.M., Anand, R., Brown, W.R.A., Fletcher, D.S. (1978) Nucleic Acids Res. **15** 5925–5943.
12. Southern, E.M. (1975) J. Mol. Biol. **98** 503–517.
13. Anand, R. and Southern, E.M. (1990) In Rickwood, D. and Hames, B.D. (eds) Gel electrophoresis of nucleic acids. IRL Press, Oxford pp 101–123.
14. Maniatis, T., Fritsch, E.F., Sambrook, J. (1982) Molecular cloning : A Laboratory Manual. Cold Spring Harbour Laboratory, New York.
15. Feinberg, A.P., Vogelstein, B. (1984) Anal. Biochem. **137** 266–276.
16. Sealey, P.G., Whittaker, P.A., Southern, E.M. (1985) Nucleic Acids Res. **13** 1905–1932.
17. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharfe, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Ehrlich, H.A. (1988) Science **239** 487–494.
18. Newton, C.R., Kalsheker, N., Graham, A., Powell, S.J., Gammack, A., Riley, J., Markham, A.F. (1988) Nucleic Acids Res. **16** 8233 –8243.
19. Hwu, H.R., Roberts, J.W., Davidson, E.H., Britten, R.J. (1986) Proc. Natl. Acad. Sci. USA **83** 3875–3879.
20. Bickmore,W.A. and Sumner, A.T.(1989) Trends in Genetics **5** 144–148.