A novel method to isolate the common fraction of two DNA samples: hybrid specific amplification (HSA)

Frédéric Lecerf, L. Foggia, Philippe Mulsant, Agnès Bonnet, François Hatey

A novel method to isolate the common fraction of two DNA samples: hybrid specific amplification (HSA)

F. Lecerf*, L. Foggia, P. Mulsant, A. Bonnet and F. Hatey

Laboratoire de Génétique Cellulaire, INRA Toulouse, Chemin de Borde Rouge, BP 27, F-31326 Castanet Tolosan, France

Received April 30, 2001; Revised and Accepted July 11, 2001

ABSTRACT

Hybrid specific amplification (HSA) is a novel simple method elaborated in order to isolate the common fraction of two DNA samples while avoiding the background due to repeated sequences. The method is based on the suppressive PCR principle, associated with a Cot1 pre-hybridization step. In recent work we demonstrated that hyperprolificity observed in Booroola ewes is associated with a mutation in the bone morphogenetic protein receptor IB gene (*BMPR-IB*). We applied HSA between ovarian cDNA and DNA from four BAC clones containing *BMPR-IB* in order to test for the presence of other genes expressed in ovary and to isolate additional *BMPR-IB* exon sequences. Of the 460 clones obtained, none contained repeated sequences. We successfully obtained 37 clones representing the major part of *BMPR-IB* coding sequence, together with 5′- and 3′-UTR sequences. Here we have successfully applied HSA to a particular tissue, but it should be possible to trap the common fraction of two DNA samples, whatever their nature.

INTRODUCTION

The identification and localization of genes of interest are the main goals in positional cloning. Typically, candidate regions are contained in a set of recombinant BACs or YACs, and candidate genes have to be searched for in a large stretch of DNA sequence. This step has often been an obstacle to gene discovery. Different methods have been elaborated to select region-specific cDNAs based on filter or oligo(dT)–latex hybridization and amplification (1–3), large scale genomic fragment screening (4–8), sequence analysis (9,10), exon trapping (11) or PCR methods (12). Each method has its own advantages, but no single method is rapid and simple.

Here we describe a novel and simple method termed hybrid specific amplification (HSA) based on the suppressive PCR principle (13), which selects and amplifies the common sequences of two complex DNA samples while eliminating repeated sequences. The use of oligonucleotide adapters that form strong clamps ensures the specificity of the method, such that only fragments with two different adapters will be amplified whereas fragments with one type of adapter will not be selected due to the suppressive effect on PCR of the adapters.

Our previous work showed that a mutation in the bone morphogenetic protein receptor IB gene (*BMPR-IB*) is associated with hyperprolificity in Booroola sheep (14). The Booroola locus is contained in a set of four ovine BACs. In human the corresponding chromosomal region has been shown to contain two identified genes: *BMPR-IB* and *UNC5C* (only *BMPR-IB* is expressed in the ovary).

In this context, we have used HSA to analyze ovarian genes contained in the four BACs. Preliminary results using HSA directly between BAC DNA and ovarian cDNA yielded a high percentage of repeated sequences (data not shown). In order to eliminate the repeated sequences, we introduced a Cot1 DNA prehybridization step in the method. The major part of the *BMPR-IB* sequence was obtained with HSA.

MATERIALS AND METHODS

Biological materials

Four ovine BACs (15) were used: 118b4, 300f2, 320h12 and 382c4. The sequences of the human BAC 115m7 (accession no. AC009920), orthologous to the *FecB* locus, and the ovine *BMPR-IB* cDNA sequence (accession no. AF298885) are available. Sequences reported in this paper have been deposited in the GenBank database (accession nos AF373711 and AF373712). Ewe ovaries were collected from a local slaughterhouse and placed in RNAlater (Ambion), prior to RNA extraction.

Oligonucleotides

The oligonucleotides and primer sequences (16) were those from suppression subtractive hybridization (SSH). The length of adapter A1S was modified from 8 to 12 nt.

Adapters. A1L (large), 5′-GTAATACGACTCACTATAGGGC-TCGAGCGGCCGCCCGGGCAGGT-3′; A1S (small), 5′-ACC-TGCCCGGGC-3′; A2L, 5′-TGTAGCGTGAAGACGACAG-AAAGGGCGTGGTGCGGAGGGCGGT-3′; A2S, 5′-ACCG-CCCTCCG-3′. Adapters were prepared as previously described (17).

PCR primers. P1, 5′-GTAATACGACTCACTATAGGGC-3′; P2, 5′-TGTAGCGTGAAGACGACAGAA-3′; PN1, 5′-TCG-AGCGGCCGCCCGGGCAGGT-3′ (nested primer 1); PN2, 5′-AGGGCGTGGTGCGGAGGGCGGT-3′ (nested primer 2).

*To whom correspondence should be addressed. Tel: +33 5 61 28 50 27; Fax: +33 5 61 28 53 08; Email: flecerf@toulouse.inra.fr

Sample preparation

Ovarian mRNA was extracted as described (18). Double-stranded cDNA was generated with the SMART PCR cDNA Synthesis Kit (Clontech). Aliquots of 2 µg of Qiagen-prepared BAC DNA and 1 µg of SMART cDNA PCR product were digested by *Rsa*I (15 U) for 90 min at 37°C. Digested samples were precipitated and resuspended in 10 µl of H_2O and then diluted (1/6). Diluted samples (2 µl) were then ligated to 2 µl of one adapter (10 µM) with 0.5 U T4 DNA ligase (Appligene) in a total volume of 10 µl for 3 h at 26°C. After ligation, 1 µl of EDTA–glycogen (0.2 M EDTA, 1 mg/ml glycogen) was added and samples were heated at 70°C for 5 min to inactivate the ligase.

Cot1 DNA

Ovine DNA extracted from fresh blood (19) was first sonicated for 90 min. The mean length of the sonicated DNA was 400 bp. Cot1 DNA was prepared as previously described (20). The resulting DNA was phenol extracted, ethanol precipitated and resuspended in H_2O. Aliquots of 5 µg of Cot1 DNA were digested with *Rsa*I (15 U), precipitated and resuspended in 10 µl.

Hybridization

Aliquots of 1.5 µl of digested BAC DNA with adapters and 1.5 µl of digested Cot1 DNA were denatured with 1 µl of 4× hybridization buffer supplied by Clontech (50 mM HEPES pH 8.3, 0.5 M NaCl, 0.02 mM EDTA pH 8.0, 10% w/v PEG 8000) at 98°C for 90 s and incubated at 68°C for 4 h. Then, 1.5 µl of digested-ligated cDNA mixed with 0.5 µl of hybridization buffer were denatured and added to the tube for a 10 h hybridization step at 68°C. As a control, 1.5 µl of digested-ligated BAC DNA and 1.5 µl of digested-ligated cDNA were denatured and incubated in 1 µl of hybridization buffer at 68°C for 10 h. The hybridization reaction was diluted in 20 µl of dilution buffer (20 mM HEPES–HCl pH 8.3, 50 mM NaCl, 0.2 mM EDTA pH 8.0) and heated at 72°C for 7 min.

PCR amplification

Two successive PCR amplifications were performed. The 25 µl PCR amplification mixture contained 1 U Platinum *Taq* DNA polymerase (Gibco BRL), 2.5 µl of 10× PCR buffer (Gibco BRL), 100 µM dNTP, 2.5 mM $MgCl_2$ and 1 µl of each primer (10 µM). The first PCR was carried out using 1 µl of diluted hybridization mixture and primers P1 and P2 with the following parameters in a Perkin Elmer 9600 thermocycler: 5 min initial elongation at 75°C, 30 cycles of 30 s at 66°C, 150 s extension at 68°C and 10 s at 94°C, followed by 7 min final extension at 68°C. A second PCR was done using 2 µl of the first PCR reaction and nested primers PN1 and PN2 with the same parameters as those mentioned above, except for the annealing temperature (68°C), extension temperature (72°C) and cycle number (15).

Cloning and analysis of HSA products

PCR products were inserted into pCR 2.1 TOPO using the TA Cloning Kit from Invitrogen. Inserts were amplified for 25 cycles using the PN1 and PN2 nested primers, monitored by 1% agarose gel electrophoresis and single strand conformation analysis (SSCA) for uniqueness. PCR products were dotted on a Hybond-N nylon filter and hybridized with a ^{32}P-labeled Cot1 DNA probe obtained by random priming (21). Hybridization was done in duplicate.

DNA sequencing was performed with the PN1 and PN2 nested primers by a chain termination reaction using an ABI dRhodamine Terminator Cycle Sequencing Ready Reaction Kit from Applied Biosystem on a Perkin Elmer 3700 sequencing apparatus. Primers were designed for all additional clones (*BMPR-IB* and unidentified clones). RT–PCR was carried out on 1 ng of ovine ovarian cDNA and PCR amplifications were conducted on BACs, genomic DNA and SMART cDNA.

Sequence analysis and alignments were carried out with GCG software (Genetics Computer Group, 1991) and Multalin (22).

RESULTS

Principle of hybrid specific amplification

A schematic representation of the method is shown in Figure 1. This method is based on the suppressive PCR effect. The different hybrid products are described in Figure 1. Fractions 1 and 2 belong either to the non-common fraction of the two DNA samples or to an abundant fraction of the samples (normalization effect). Only one kind of adapter is present at both ends of these DNA fragments. During PCR the adapters will form a 'panhandle-like' structure with a strong clamping effect (13). This conformation prevents annealing of the primers and thus amplification of the fragment during each PCR step. This PCR suppressive effect was confirmed on *Rsa*I-digested genomic DNA ligated with a single type of adapter (data not shown). Fraction 3 represents heteroduplexes between genomic DNA (or cDNA) and Cot1 DNA and can only be amplified at a linear rate. The use of Cot1 DNA with no adapter allows attenuation of the repeated sequences in genomic DNA. Fraction 4 will not amplify due to the absence of adapters. Only fraction 5, which represents the common fraction of the two samples and has two different adapters, will be exponentially amplified.

Characterization of clones

HSA was applied to ovarian cDNA and the four BACs spanning the *FecB* locus. Insert sizes in the selected clones varied from 100 to 1000 bp. Screening for clones containing repeated DNA was performed by filter hybridization with ^{32}P-labeled Cot1 DNA. In the absence of Cot1 prehybridization in HSA virtually all clones were positive. When the prehybridization step was included only 20 of 480 clones were positive (Fig. 2), confirming that elimination of repeated sequences was efficient. The Cot1-positive clones were excluded from further analysis.

Cot1-negative clones were analyzed on agarose gels and 192 clones with an insert size of >200 bp were selected. A redundancy screening test was performed by SSCA; 52 clones were selected to be sequenced and 44 informative sequences were obtained.

Each clone was first compared to the ovine *BMPR-IB* cDNA sequence and to the human BAC 115m7. Sequence alignments allowed a distinction between three different classes of fragments (Fig. 3A). (i) Four fragments showing a complete sequence match between cDNA and genomic DNA were

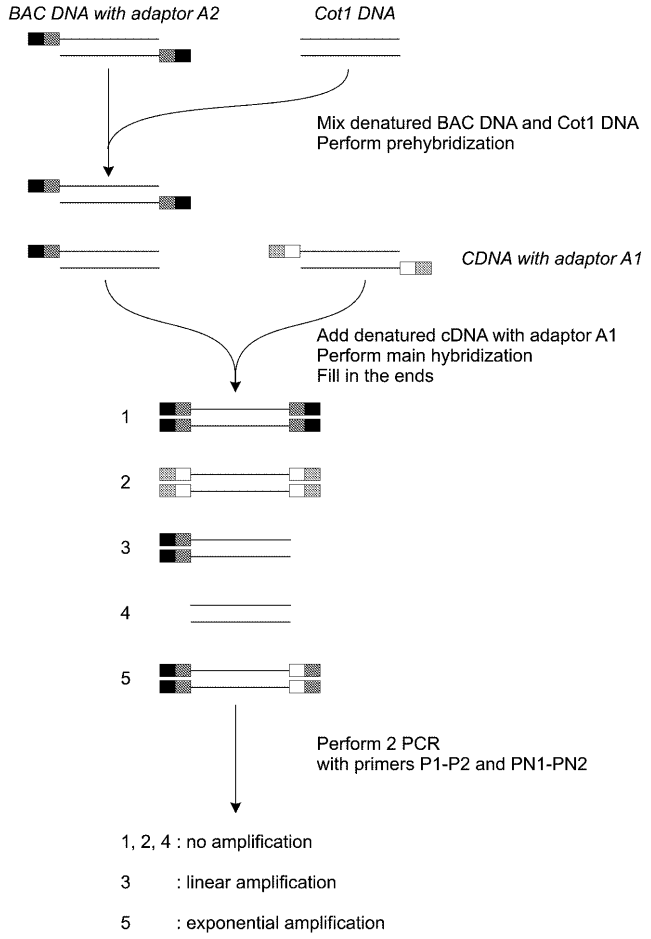

Figure 1. Schematic description of the HSA principle (only the major class of hybridization products is represented). The BAC and cDNA samples are first digested with *Rsa*I, to generate blunt end fragments, and each is ligated to a different adapter. To neutralize repeated sequences, a prehybridization step between the genomic sample and *Rsa*I-digested Cot1 DNA is performed. The denatured cDNA sample is then immediately added to the reaction tube for the main hybridization. Finally, samples are heated at 72°C for 5 min ('end-filling' step) and a two-step PCR amplification is performed to select the common fraction of the samples: only fraction 5 will be exponentially amplified.

isolated, one corresponding to 5′-UTR and three to 3′-UTR sequences not previously identified. 5′-RACE PCR and RT–PCR with primers for known *BMPR-IB* exons and clone sequences confirmed that these sequences were effectively *BMPR-IB*

Table 1. Summary of the clones obtained with HSA

Category	No.
Clones picked	480
Clones selected after Cot1 hybridization	460
Clones selected after agarose gel screening	192
Clones selected after SSCA	52
Informative sequences	44
BMPR-IB clones	37
Different *BMPR-IB* fragments	9
Other clones	7

sequences. (ii) We also observed cDNA fragments spanning consecutive exons between *Rsa*I restriction sites, spanning exons 3–7 and 9–10. No fragments spanning exons 7–9 and 11–12 were observed. (iii) The last class of clones is genomic fragments composed of exon and flanking intron sequence. In addition, one clone (marked by an asterisk in Fig. 3A) flanked by *BMPR-IB* intron sequences matches the exon 11 sequence. The principle of HSA could not explain how this clone was isolated.

On sequence alignment seven clones did not match with *BMPR-IB* sequences or with any identified gene contained in the genetic interval (e.g. *UNC5C*). Further RT–PCR analysis confirmed that these clones could have been isolated due to the presence of contaminant DNA. Therefore, HSA did not trap any other ovarian genes.

Table 1 summarizes the results obtained. With the use of Cot1 DNA only 4–5% of clones were assigned as repeated sequences. Amongst all sequenced clones, 85% could be effectively identified as *BMPR-IB* sequences. The major part of the coding sequence of *BMPR-IB* was obtained by HSA and the method isolated fragments from the 5′- and 3′-UTR regions.

DISCUSSION

Amongst the *BMPR-IB* sequences we obtained two categories: 5′- and 3′-UTR sequences (fragments identical between cDNA and genomic DNA) and coding sequences which spanned consecutive exons or were composed of exon and flanking intron sequences. The hybridization process which enabled the

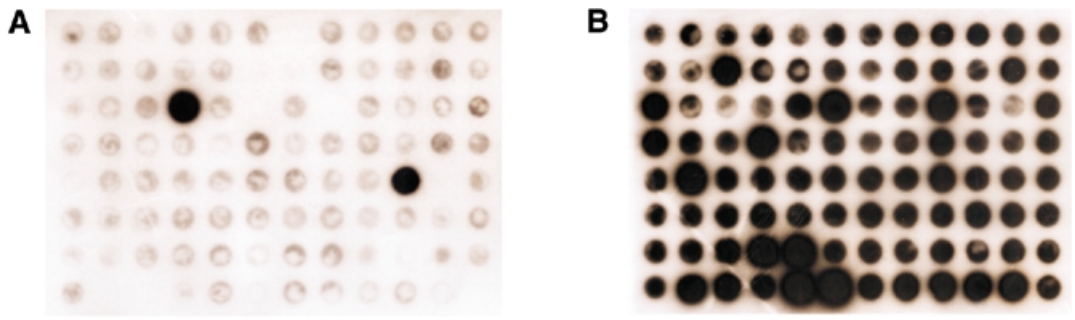

Figure 2. Dot-blot analysis of clones obtained after HSA with (**A**) or without (**B**) Cot1 prehybridization. The probe used was Cot1 DNA.

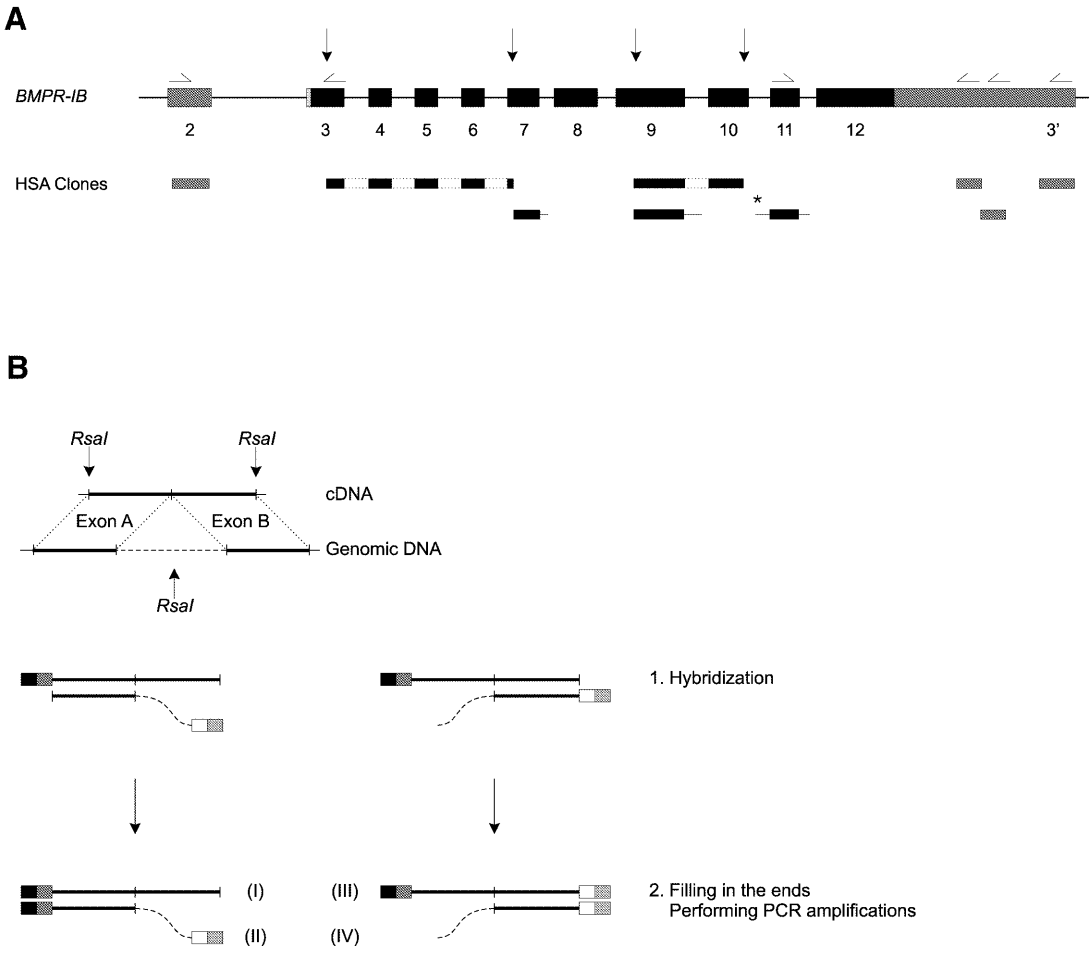

Figure 3. (**A**) Schematic representation of the HSA clones over the *BMPR-IB* locus. Boxes indicate known exons of the gene; black boxes represent the coding sequences (the start codon is situated in exon 3). The cDNA sequence is numbered according to previously known exons. Exon 1 is not shown in the figure, as it is located in a BAC excluded from the present HSA experiment. Arrows show the *Rsa*I recognition sites. Three kinds of clones were obtained: some contained in only one exon, others spanning several exons and, finally, hybrid clones containing exons and introns. Primers used for RT–PCR experiments are indicated with horizontal arrows. (**B**) Isolation of clones composed of consecutive exons or an exon with a flanking intron sequence. After the main hybridization, only the common exonic sequence matches between the cDNA and the genomic DNA. During the 'end-filling' step the second adapter is synthesized, allowing exponential amplification of molecules (II) and (III).

generation of such hybrid molecules is depicted in Figure 3B. Hybridization actually occurs at only one end of the fragment. During the 'end-filling' step the second adapter is synthesized on the other strand. This kind of fragment can thus be exponentially amplified. Characterization of such sequences requires sequence comparison against databases that are largely available, especially since the recent human genome sequencing efforts.

The application of HSA to our model was highly beneficial. One fragment corresponds to the 5′-region, whose existence was confirmed by 5′-RACE (data not shown). This sequence could be a non-coding exon of *BMPR-IB*, resulting from a different promoter or alternative splicing system (23). Three fragments of the 3′-UTR region were obtained. These fragments were sequenced and added to the *BMPR-IB* sequence. Apart from the inner fragments (exons 7–9 and 11–12), HSA isolated the major part of the *BMPR-IB* gene. Another objective of the application of HSA was to look for other ovarian genes. None of the other seven clones could be identified as ovarian genes. Computational analysis and RT–PCR experiments also rejected this hypothesis.

One major improvement of HSA is the attenuation of repeated sequences. Thus, in most other methods, and in HSA without a Cot1 DNA prehybridization step, final analysis of the clones obtained reveals at least 25% of clones containing repeated sequences (12). Introducing a Cot1 prehybridization step prevents this problem. Only 5% of the selected clones were assigned as repeated sequences.

A lot of methods have been elaborated to look for genes contained in BACs or YACs. Each strategy has its own advantages and is complementary to the others. Exon trapping works well but is labor intensive and may fail due to an interaction between host cell genes and the genes of interest. Moreover, the exon trapping method is not tissue specific. Results from computational prediction methods are only partially reliable and give inconsistent results. Direct hybridization selection has also found some use, but it is problematical with rarely expressed genes and repeated sequences.

We have described an additional approach to this problem. In contrast to other methods, HSA is a relatively simple and easy method to select chromosome-specific cDNA avoiding the repeated sequence background. Furthermore, HSA does not require specific equipment or large amounts of biological material. Moreover, this method has a wide range of applications as it can isolate every common fraction between two samples, whatever their nature.

ACKNOWLEDGEMENTS

We are very grateful to F. Benne for collecting ewe ovaries. We also thank J. Rallières for technical assistance. We are also very grateful to L. Schibler for providing the ovine BACs. We thank J. Gellin and A. Goureau for advice on carrying out the Cot1 DNA production. We are very grateful to C. Greenland and A. Vignal for revision of the manuscript.

REFERENCES

1. Lovett,M., Kere,J. and Hinton,L.M. (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl Acad. Sci. USA*, **88**, 9628–9632.
2. Parimoo,S., Patanjali,S.R., Shukla,H., Chaplin,D.D. and Weissman,S.M. (1991) cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl Acad. Sci. USA*, **88**, 9623–9627.
3. Tagle,D.A., Swaroop,M., Lovett,M. and Collins,F.S. (1993) Magnetic bead capture of expressed sequences encoded within large genomic segments. *Nature*, **361**, 751–753.
4. Troen,B.R. (1987) Colony screening of genomic cosmid libraries. *Methods Enzymol.*, **151**, 416–426.
5. Elvin,P., Slynn,G., Black,D., Graham,A., Butler,R., Riley,J., Anand,R. and Markham,A.F. (1990) Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res.*, **18**, 3913–3917.
6. Baxendale,S., Bates,G.P., MacDonald,M.E., Gusella,J.F. and Lehrach,H. (1991) The direct screening of cosmid libraries with YAC clones. *Nucleic Acids Res.*, **19**, 6651.
7. McAlinden,T.P. and Krawetz,S.A. (1994) A practical method to screen libraries of cloned DNA. *Anal. Biochem.*, **218**, 237–238.
8. Simmons,A.D., Overhauser,J. and Lovett,M. (1997) Isolation of cDNAs from the Cri-du-chat critical region by direct screening of a chromosome 5-specific cDNA library. *Genome Res.*, **7**, 118–127.
9. Uberbacher,E.C. and Mural,R.J. (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl Acad. Sci. USA*, **88**, 11261–11265.
10. Zhang,M.Q. (1997) Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl Acad. Sci. USA*, **94**, 565–568. [Erratum, 1997, *Proc. Natl Acad. Sci. USA*, **94**, 5495]
11. Duyk,G.M., Kim,S.W., Myers,R.M. and Cox,D.R. (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl Acad. Sci. USA*, **87**, 8995–8999.
12. Hamaguchi,M., O'Connor,E.A., Chen,T., Parnell,L., McCombie,R.W. and Wigler,M.H. (1998) Rapid isolation of cDNA by hybridization. *Proc. Natl Acad. Sci. USA*, **95**, 3764–3769.
13. Siebert,P.D., Chenchik,A., Kellogg,D.E., Lukyanov,K.A. and Lukyanov,S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.*, **23**, 1087–1088.
14. Mulsant,P., Lecerf,F., Fabre,S., Schibler,L., Monget,P., Lanneluc,I., Pisselet,C., Riquet,J., Monniaux,D., Callebaut,I. *et al*. (2001) Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Mérino ewes. *Proc. Natl Acad. Sci. USA*, **98**, 5104–5109.
15. Vaiman,D., Billault,A., Tabet-Aoul,K., Schibler,L., Vilette,D., Oustry-Vaiman,A., Soravito,C. and Cribiu,E.P. (1999) Construction and characterization of a sheep BAC library of three genome equivalents. *Mamm. Genome*, **10**, 585–587.
16. Diatchenko,L., Lau,Y.F., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D. *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.
17. Lisitsyn,N. and Wigler,M. (1993) Cloning the differences between two complex genomes. *Science*, **259**, 946–951.
18. Chomczynski,P. and Sacchi,N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.
19. Montgomery,G.W. and Size,J.A. (1990) Extraction of DNA from sheep white blood cells. *N. Z. J. Agric. Res.*, **33**, 437–441.
20. Britten,R.J., Graham,D.E. and Neufeld,B.R. (1974) Analysis of repeating DNA sequences by reassociation. *Methods Enzymol.*, **29**, 363–418.
21. Feinberg,A.P. and Vogelstein,B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Addendum Anal. Biochem.*, **137**, 266–267.
22. Corpet,F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*, **16**, 10881–10890.
23. Baur,S.T., Mai,J.J. and Dymecki,S.M. (2000) Combinatorial signaling through BMP receptor IB and GDF5: shaping of the distal mouse limb and the genetics of distal limb diversity. *Development*, **127**, 605–619.