CHAPTER 8

COINCIDENCE CLONING: ROBUST TECHNIQUE FOR ISOLATION OF COMMON SEQUENCES

ANTON A. BUZDIN

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 16/10 Miklukho-Maklaya, 117997 Moscow, Russia
Phone: +(7495) 3306329; Fax: +(7495) 3306538; E-mail: anton@humgen.siobc.ras.ru

Abstract: Coincidence cloning (CC) is aimed at finding DNA fragments, which are common to the samples under study. The nature of these input DNAs may be genomic or cDNA, cloned or uncloned. The approach is based on cloning identical (or almost identical) nucleotide sequences belonging to different fragmented genomic DNA or cDNA pools, while discarding sequences that are not common to both. Early versions of the CC technique were not very efficient. Their most serious disadvantage was rather low selectivity, so that the resulting libraries of the fragments contained large amounts of sequences unique to one of the two sets of DNA fragments under comparison. To avoid this, Azhikina and colleagues were the first to exploit the technique of selective polymerase chain reaction (PCR) suppression, which strongly increased the efficiency of CC. Another important problem is the "nonspecific" imperfect hybridization between nonorthologous repetitive elements or short sequence-similar sequences, which produces chimeric clones representing an impressive fraction of the libraries (up to 60%, when complex genomic mixtures). An important improvement in this technique comprises treatment of the hybrids with the nucleases, specifically recognizing single-nucleotide mismatches or more extended loop regions. This results in digestion of improperly matched hybrids (primarily chimeras), whereas perfect, nonchimeric heteroduplexes are greatly enriched in the final mixture (up to 96% or more).

coincidence cloning; PCR, polymerase chain reaction; PS, PCR suppression effect; SAGE, serial analysis of gene expression.

TABLE OF CONTENTS

1. INTRODUCTION

Unlike subtractive hybridization, which is aimed at the recovery of differential sequences, the approach termed "coincidence cloning" (CC) was developed to find DNA fragments, which are common to the samples under study. According to Rebeca Devon and Anthony Brookes, who were among the inventors of this method, "the term *coincidence cloning* encompasses a wide range of methodologies, the aim of which is to isolate DNA sequences which occur in both of two input DNA sources. The nature of these input DNAs may be genomic or cDNA, cloned or uncloned. If the input DNAs are genomic then the product will be enriched for useful markers co-occurring between the two. If the input DNAs comprise one genomic resource and one cDNA resource the product will contain genes mapping to that particular genomic region" (Devon and Brookes 1996).

The approach is based on cloning identical (or almost identical) nucleotide sequences belonging to different fragmented genomic DNA or cDNA pools, while discarding sequences that are not common to both (Devon and Brookes 1996). By comparing genomic DNA fragments with fragments of any similarly fragmented locus (cloned in the form of bacterial artificial chromosomes – BACs, cosmids etc.), one can select and identify the genomic fragments belonging to this locus.

To this end, both fragmented DNAs under comparison are specifically tagged (e.g. by ligating different terminal adapter oligonucleotides), mixed, denatured, and hybridized, followed by the isolation of duplexes having both specific tags (i.e. "heterohybrid" products derived from both samples, which are common to both tagged DNA mixtures). The former step is the key stage of the whole procedure,

as an efficient isolation of proper hybrids provides construction of CC libraries, truly enriched in common sequences (Figure 1).

   Three major approaches for distinguishing heterohybrids from homohybrids can be mentioned. First (*cloning selection*), hybridizing input DNAs (source A and B) may be flanked by different restriction enzyme recognition sites (e.g. by introducing such sites in ligated terminal adapters). Restriction endonuclease-treated hybridization products are ligated into the vector predigested to produce sticky ends complementary to those appeared in heterohybrids, and random transformants are screened. Second (*physical separation*), one of the source DNAs can be immobilized on a solid support (e.g. source A DNA bound to magnetic beads) and, following hybridization, the support-attached fraction (which contains unhybridized source A molecules, source A homohybrids, and heterohybrids source A/B) is separated from the liquid-phase fraction (all other types of DNAs). The solid-phase fraction is further polymerase chain reaction
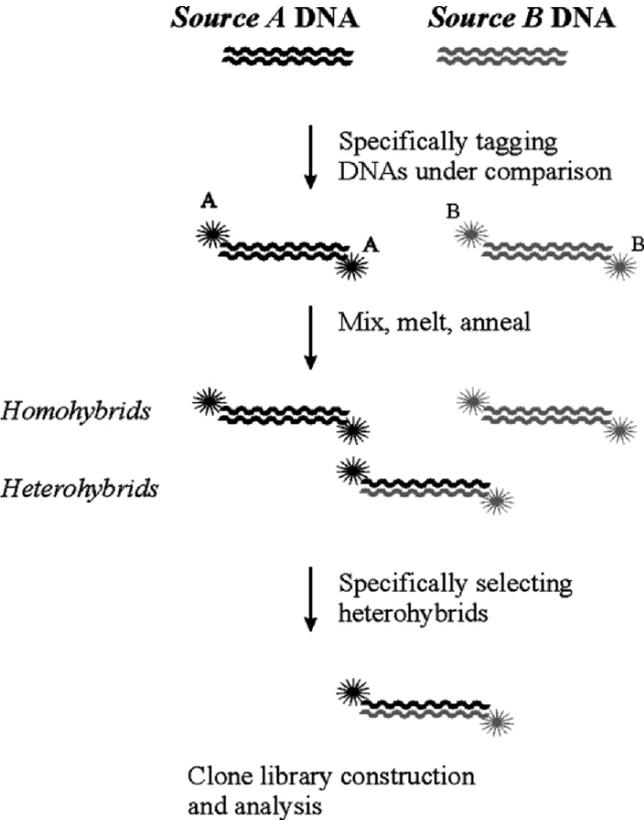

Figure 1. Generalized scheme for coincidence cloning (CC) approaches. Proper isolation of source 1–2 heteroduplexes is the key procedure in all CC-based techniques.

(PCR) amplified with primers specific to source B linkers, resulting in the more or less specific amplification of source A/B heterohybrids. Third (*PCR-only-based approaches*), source A and B DNAs are ligated to different oligonucleotide linkers, and further hybridization products are amplified with pairs of source A- and B-specific primers (described below in this chapter).

Early versions of the CC technique were not very efficient and, therefore, have not been widely used. Their most serious disadvantage was rather low selectivity, so that the resulting libraries of the fragments contained large amounts of sequences unique to one of the two sets of DNA fragments under comparison. To avoid this, Azhikina and colleagues exploited the technique of selective PCR suppression (PCR suppression effect is described in detail in Chapter 2; Diatchenko et al. 1996), which strongly increased the efficiency of CC (Azhikina et al. 2004; Azhikina and Sverdlov 2005; Azhikina et al. 2006).

Another important problem is the "nonspecific" imperfect hybridization between nonorthologous repetitive elements or short sequence-similar sequences, which produces chimeric clones representing an impressive fraction of the libraries (up to 60%, when complex genomic mixtures like mammalian DNAs are hybridized). Such chimeric clones make the library analysis problematic, as (1) it is frequently difficult to distinguish chimeras from "true" hybrids, especially for unsequenced genomes, and (2) more sequencing work is needed to get substancial data sets. The recent paper by Chalaya et al. (2004) describes an important improvement in this technique, which comprises treatment of the hybrids with the nucleases, specifically recognizing single-nucleotide mismatches, or more extended loop regions. This results in digestion of improperly matched hybrids (primarily chimeras), whereas perfect, nonchimeric heteroduplexes are greatly enriched in the final mixture (up to 96% or more).

Also, when considering CC, one has to clearly realize that kinetical requirements are extremely important for the success of the whole procedure (see Chapters 1 and 10). When very complex DNA mixtures are hybridized, it is very difficult to obtain satisfactory reassociation values for a reasonable time. In particular, when genome size increases beyond 5×10^8 bp (complexity comparable with that of arabidopsis or drosophila genomes), the kinetics of hybridization start to become an increasingly important factor limiting reassociation (Milner et al. 1995). To enhance the kinetics of hybridization, increased hybridization times, higher DNA concentrations, longer DNA fragments, and the use of techniques that enhance the rate of reassociation (reviewed in Chapter 10, Section 2), are recommended.

The last (but not the least) problem is a very low reassociation rate of non-repetitive DNAs. Indeed, DNA reassociation rate for each particular fragment is proportional to the square of its concentration; therefore, repetitive elements presented in a genome by ~10 (some pseudogenes), ~1000 (several mammalian endogenous retroviral families), or ~1,000,000 (human *Alu* retrotransposons) copies will hybridize, respectively, 10^2-, 10^6-, and 10^{12}-fold faster than fragments representing unique genomic sequences. As the latter's reassociation rate is

incomparably lower, ~99% reassociation may take months or years, an enormous background of repetitive sequences appears when reasonable (approximately days) hybridization time is used. In this case, a great majority of double-stranded molecules in solution are hybridized repeats, whereas unique sequences mostly remain in a single-stranded form. Interestingly, the famous primate-specific *Alu* retrotransposons were discovered first using CC of sequences co-occurring in a set of comparing genomes (Nelson et al. 1989; Aslanidis and de Jong 1991). Therefore, the true genomic sequence representations will be enormously biased in the resulting libraries.

A rather efficient attempt to improve the situation is the addition of competitor DNA fractions enriched in genomic repeats (Sambrook and Russell 2001) into hybridization mixture. Such competitors are mostly fractions of quickly reassociating double-stranded DNA, purified from single-stranded DNA using hydroxyapatite column chromatography. Such fractions, for example, commercially available "C_ot A" and "C_ot B" DNAs from Gibco BRL (USA) are greatly enriched in genomic repeats and may be used to decrease the background of repetitive sequences in cloned libraries. To this end, the initial genomic DNAs to be hybridized must be fragmented (either by sonication or digestion with restriction endonucleases), tagged (through the ligation of adapter sequences, by incorporation of biotin or other signal molecules), denatured and allowed to hybridize in the presence of competitor DNA, taken in a 100–1000-fold weight excess.

The major part of genomic repeats presenting in the sample DNA will hybridize to competitor DNA. At the next stage, it is crucial to isolate the "proper" hybrids (those formed by original genomic DNA fragments) while discarding genomic–competitor DNA duplexes and single-stranded DNAs. This can be done, for example, by using selective PCR amplification of the proper hybrids (see protocol in Chapter 10, Section 4.2), or using biotin–streptavidin systems. To our experience, the use of C_ot A DNA taken in 100-fold weight excess results in a decrease of genomic repeat-containing clones from ~93% to 76–78% of the libraries, when nonsimplified frequent-cutter endonuclease-digested human genomic DNA is hybridized (Chalaya et al. 2004), which is significantly closer to the natural genomic occurrence of repetitive elements, occupying approximately two thirds of human DNA (Lander et al. 2001; Venter et al. 2001).

To measure CC efficacy, the characteristic value termed *enrichment factor* was proposed. The enrichment factor is given by the ratio of the relative mass representation of a target sequence in the product DNA mixture, over its relative mass representation in the most complex of the two DNA sources (Devon and Brookes 1996).

2. CLONING SELECTION OF HETERODUPLEXES

This approach utilizes preliminary labeling of hybridizing DNA sources with ligated oligonucleotide adapters, harboring different restriction sites. For example (Figure 2), source A DNA is tagged by linkers having *Not*I restriction site,
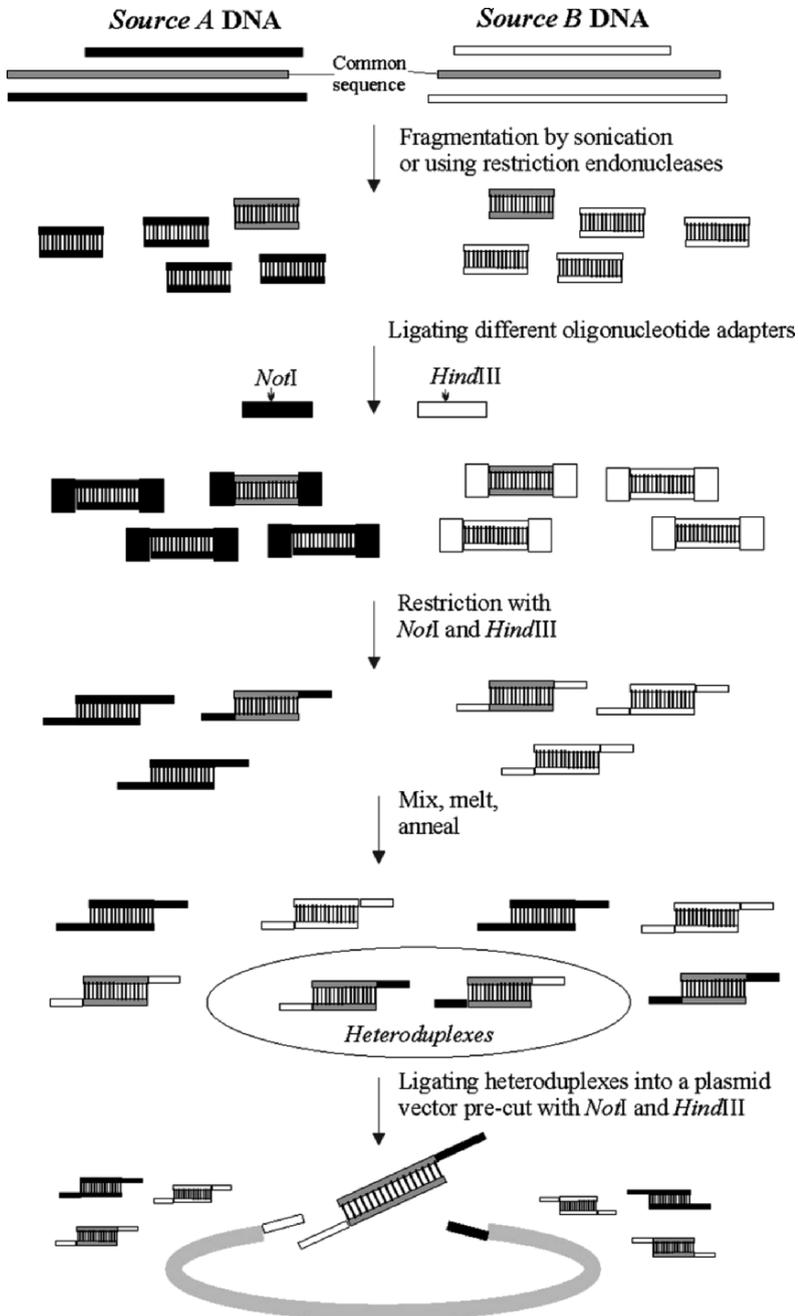
Figure 2. Variant of coincidence cloning (CC) utilizing selective cloning of the heteroduplexes using their unique combination of terminal restriction sites, differing them from all types of homoduplexes.

whereas linkers for source B have recognition sequence for *Hind*III enzyme. Both DNA samples are mixed, denatured, and allowed to hybridize. Following filling-in the DNA termini, the mixture is treated with both restriction enzymes and ligated into the plasmid vector having sticky ends compatible to those produced by *Not*I and *Hind*III restriction endonucleases, respectively. Theoretically, only heterohybrids derived from both sources will be ligated and further cloned into *Escherichia coli*. Using this approach, an enrichment factor of 10–20-fold could be achieved (Nelson et al. 1989). Cloning selection is limited by the recovery of significant levels of background products, mostly due to the cloning of source A- or B-only hybrids into an incompletely digested vector and the cloning of chimeric duplexes discussed above (Devon and Brookes 1996). For these reasons, other CC modifications have been developed.

3. PHYSICAL SEPARATION OF HYBRIDS

To improve the separation of hybrids, one of the source DNAs can be immobilized on a solid carrier (e.g. source A DNA bound to nylon filter or to strepta-vidin-coated magnetic beads via biotinylated primer). Prior hybridization, one or both DNA sources can be preblocked for repeats by adding an excess of $C_o t$ fractions. Following hybridization (Figure 3), the support-attached fraction (which contains unhybridized source A molecules, source A homohybrids, and heterohybrids source A/B) is separated from the soluble fraction (all other types of DNAs). The optimization of hybridization and washing conditions are critical in this type of experiment. The solid-phase fraction is further PCR amplified with primers specific to source B linkers, resulting in the more or less specific amplification of source A/B heterohybrids. The enrichment factor value (~1000-fold) increased dramatically comparing to the previous group of methods (Lovett et al. 1991; Parimoo et al. 1991), but, however, was still insufficient to reliably recover unique sequences or rare transcripts (Devon and Brookes 1996). Regardless of further development of some improved related versions of CC (Brookes et al. 1994), heterohybrid isolation remained puzzling in many applications, and a problem of unwanted chimeric hybrid formation has not been solved in these techniques. The CC in that form, therefore, did not become popular among the research community. The use of the PS effect, which will be described in the next section, revolutionerized CC and made it a method of choice: heterohybrid selection became an effective, easy, and quick procedure.

4. PCR-ONLY-BASED APPROACHES

These approaches, relying on the PS effect, have been developed recently by Tatyana Azhikina and colleagues (Azhikina et al. 2004; Azhikina and Sverdlov 2005; Azhikina et al. 2006). In general, in all its applications, PS allows the amplification of wanted sequences and simultaneously suppresses the amplification of unwanted ones. Pan handle-like stem-loop DNA constructs for PS are
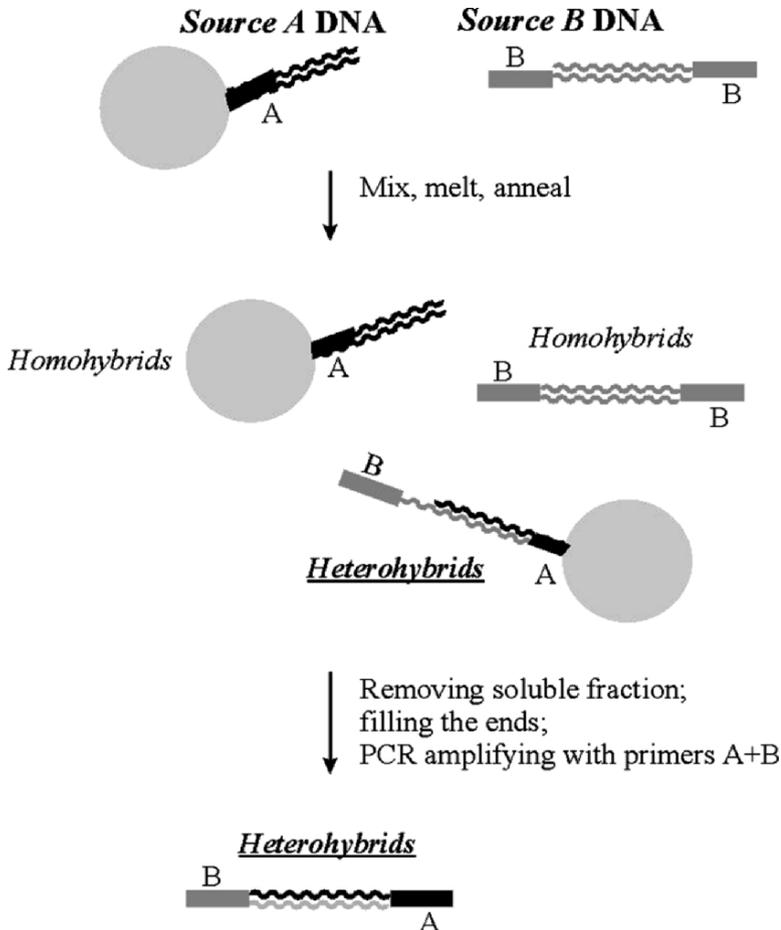
Figure 3. Selection of heteroduplexes based on both (1) attaching source A DNA to a solid carrier (magnetic beads) and (2) PCR amplification using combination of source A- and B-specific primers.

created by ligation of long guanine–cytosine (GC)-rich adapters to DNA or cDNA restriction fragments (Figure 4) (Luk'ianov et al. 1994). As a result, each single-stranded DNA fragment is flanked by terminal inverted repeats (i.e. by self-complementary ends). During PCR, on denaturing and annealing, the self-complementary ends of each single strand form duplex stems, converting each fragment into a large pan handle-like stem-loop structure. The formation of stable duplex structures at the fragment ends makes the PCR with the adapter-primer (A-primer) alone relatively inefficient, because the intramolecular annealing of the complementary termini is kinetically favored and more stable than the intermolecular annealing of shorter A-primers. This effect is therefore

A1 A2

**Ligation of
suppression adapter
oligonucleotide**

Filling-in the ends

Denaturation, annealing

A1 *Intramolecular
complementary
basepairing*

Oligonucleotides, complementary to adapter sequence,
fail to prime PCR
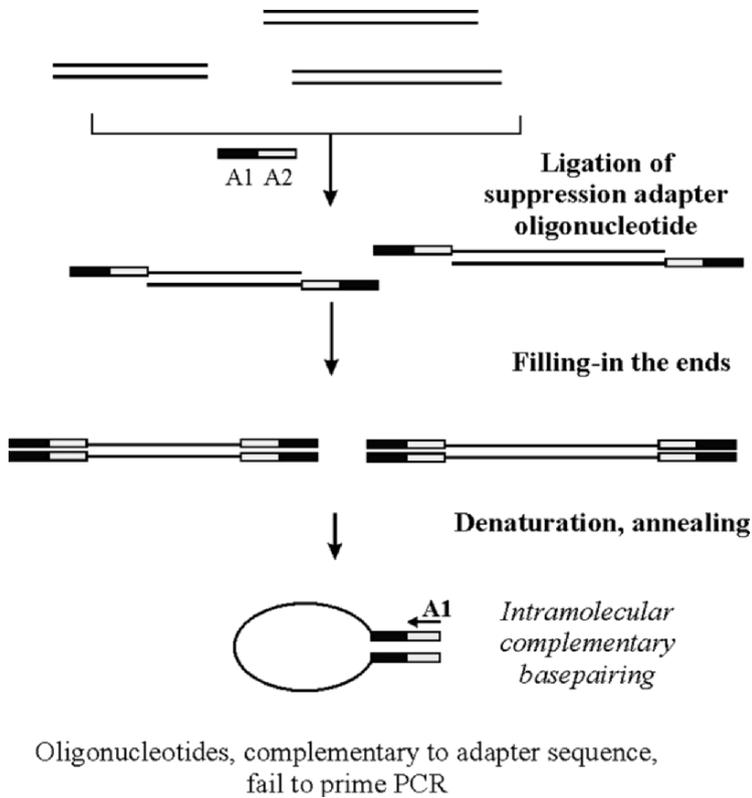
Figure 4. Principle of the PCR suppression (PS) effect. DNA molecules flanked by GC-rich ligated adapter oligonucleotides termed suppression adapters (~40 bp long artificial inverted terminal repeats) form intramolecular terminal duplexes, thus preventing terminal adapter-specific primer annealing and, consequently, inhibiting the PCR.

called PCR suppression (Luk'ianov et al. 1994). However, PCR is efficient in the presence of both A-primers and target-primers (T-primers, targeted at the specific sequences in the single-stranded loops). The T-primer anneals to its target and is used by DNA polymerase to initiate DNA synthesis. The newly synthesized product has two termini, which are not complementary and, thus, cannot fold into a stem-loop structure. This fragment is, therefore, not subject to the PS effect, and is efficiently amplified. Consequently, only the fragments containing the target are exponentially amplified by PCR, while the background fragments without the target remain inert.

Figure 5 represents a simple model of the use of CC for isolation of evolutionary conserved sequences shared by comparing genomes, recently reported by Chalaya et al. (Chalaya et al. 2004). Genomic DNAs of human and of New World monkey marmoset *Callithrix pigmaea* were digested with frequent-cutter

(I) Digestion of genomic DNA

Human DNA   **Marmoset DNA**

RE   RE

(II) Ligation of suppression
adapters A and B

(III) Mix, melt, anneal

(IV) Filling in the ends

(V) Treatment with mismatch-
sensitive nuclease

(VI) PCR with primers A+B

PCR suppression

Successful amplification   No amplification
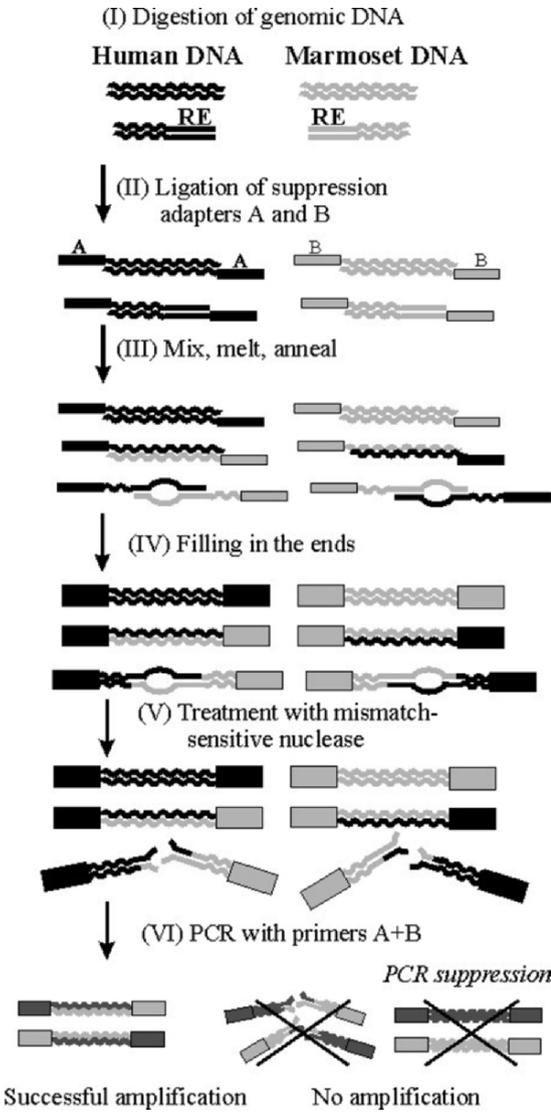
Figure 5. Schematic representation of the mispaired DNA rejection technique rationale. Not only exactly matched identical sequences, but also a number of background chimeric duplexes, which are usually products of hybridization between repetitive elements (REs), are generally PCR amplified and appear in DNA libraries. The addition of mismatch-sensitive nucleases makes it possible to selectively cleave background duplexes containing imperfectly matched regions and, therefore, to enrich the resulting library in target sequences.

restriction endonuclease, and two different sets of suppression adapters were ligated to them. Samples were then mixed, denatured, and allowed to reanneal, followed by filling the ends with DNA polymerase (Figure 5) and treatment with mismatch-specific nucleases to decrease production of chimeric hybrids. These enzymes recognize improperly matched double-stranded DNAs and cut such "wrong" hybrids, thus clearly enhancing hybridization efficacy (see also Chapters 9 and 10). At the next stage, hybridization products are subjected to PCR with primers specific to the suppression adapters used, so that only human–*C. pigmaea* hybrid molecules are amplified. As a result, we managed to create a genomic library highly enriched in evolutionary conserved sequences shared by human and *C. pigmaea* genomes.

Another successful application of the CC is the new technique called "non-methylated genomic sites coincidence cloning (NGSCC)", which results in a set of sequences that are derived from the genomic locus of interest and contain an unmethylated CpG site. In this case, annealing mixture complexities were not as high as for whole-genome digest hybridizations, and treatment with mismatch-sensitive nucleases was not needed. The technique is based on the initial fragmentation with a methyl-sensitive restriction enzyme (Figure 6). To simplify the DNA sets to be compared, they can be additionally digested with a frequent-cutter that is not sensitive to methylation of its target site, e.g. *Alu*I. As a result, the lengths of the fragments can be restricted to a size that is optimal for subsequent PCR amplifications, usually up to 1.5 kb. Different suppression adapters are then ligated to sticky ends produced by methyl-sensitive restriction enzyme and to blunt ends created by *Alu*I. Further, PCR amplification with primers specific to both adapters used results in the amplicon of genomic fragments having unmethylated CpG site at one terminus and *Alu*I restriction site in another one.

This amplicon is further hybridized to a new-suppression-adapter-ligated fragmented DNA from the genomic locus of the interest (the authors analyzed methylation profiles of a ~1 Mb long human genomic locus *D19S208-COX7A1* from chromosome 19). In the following nested PCR, only those unmethylated CpG-containing fragments that match to D19S208-COX7A1 genomic locus, were amplified. Sequencing of the resulting libraries derived from initial genomic DNAs from healthy and cancerous tissues enabled authors to create the first large-scale comprehensive tissue- and cancer-specific methylation map for that locus (Azhikina and Sverdlov 2005). More recently, the same group of authors combined NGSCC with serial analysis of gene expression (SAGE) thus creating new technique termed "RIDGES", which is significantly more informative than NGSCC, as its outcome in 10–20-fold more information about methylation sites per one sequenced clone (Azhikina et al. 2006).

As mentioned above, the use of CC is not restricted to genomic DNA analysis. In particular, recently published technique termed genomic repeat expression monitor (GREM) utilizes CC of preamplified 3′-terminal genomic flanking regions of the repetitive elements with the set of cDNA 5′-terminal parts. This results in construction of a hybrid genomic DNA/cDNA library,
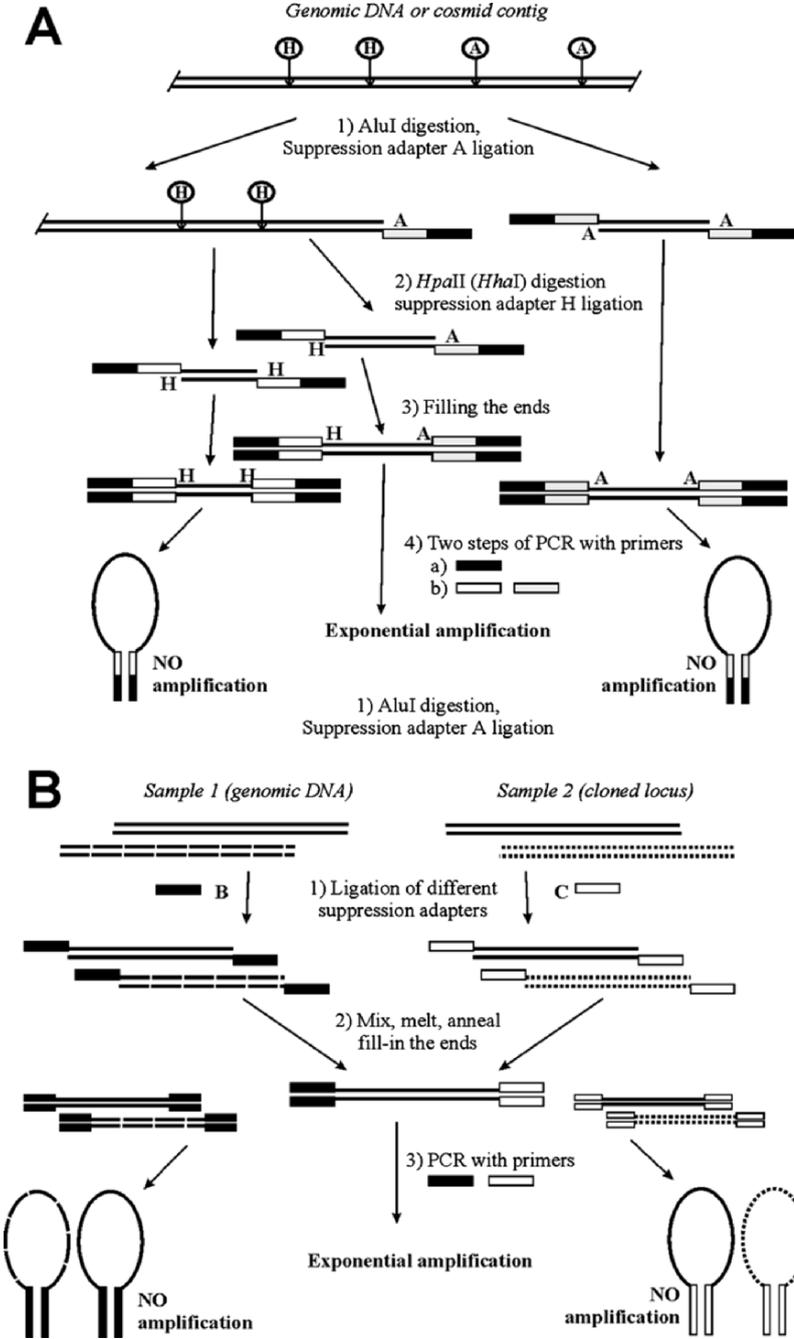
Figure 6.

enriched in promoter-active repeats, thus making it possible to create a comprehensive genome-wide map of such repetitive elements (Buzdin et al. 2006b; Buzdin et al. 2006a). We applied GREM for the analysis of long terminal repeats (LTRs) of mostly human-specific family of endogenous retroviruses called *HS LTRs* (Buzdin et al. 2002). The GREM technique outlined in Figure 7 consists of three major stages: (1) synthesis of full-length cDNA libraries whose clones include specific oligonucleotide adapters exactly tagging the cDNA 5′-ends, (2) selective PCR amplification of genomic repeat-flanking regions, and (3) hybridization of the genomic repeat-flanking regions to the cDNA with a subsequent PCR amplification of the genome–cDNA heteroduplexes.

The first stage of GREM is aimed at the amplification of full-length cDNAs tagged at the 5′-ends with a specific adapter oligonucleotide (CS in our case).The tagging is achieved due to the "cap-switch" effect in the process of cDNA synthesis. Having reached the 5′-end of the mRNA template, oligo(dT)-primed reverse transcriptase adds a few additional deoxycytidine nucleotides to the 3′-end of the cDNA. An oligonucleotide with an oligo-ribo(G) sequence at its 3′-end hybridizes to the deoxycytidine stretch to form a primer, which allows reverse transcriptase to switch templates and to continue replicating to the end of the oligonucleotide. This technique allows one to precisely tag the cDNA 5′-ends that correspond to transcription start sites (Figure 7). Prior to the hybridization at stage (3), the cDNA was digested with *Alu*I restriction endonuclease to get shorter fragments and to avoid further background amplification of hybrids with read-through transcripts driven in the sense orientation with respect to the LTR direction (Figure 7, stage 1, step "*Alu*I digestion"). *Alu*I was chosen because the HS LTR consensus sequence lacks restriction sites of this frequent-cutter endonuclease. The treatment of cDNA with *Alu*I (Figure 7) suppresses the yield of sense read-through LTR containing products at the following stage (see below).

At the second stage, we selectively PCR amplified genomic regions flanking the 3′-termini of HS LTRs. The cDNA hybridization with the amplicon obtained was used to select the cDNA molecules that contain HS LTRs at their 5′-termini. The amplification of genomic flanking regions is a critical step ensuring the specificity of the whole procedure. Nested PCRs result in selective amplification of all target repeat-flanking sequences, whereas cDNA

Figure 6. cont'd. Summary of the NGSCC procedure. (A) Selective amplification of *Alu* I-*Hpa* II/*Hha* I fragments. Cosmid contig/genomic DNA is depicted as horizontal lines. *Alu* I and *Hpa* II/*Hha* I recognition sites are marked by "A" and "H", respectively. A and H suppression adapters are shown as boxes in identical (external) parts are unshaded, and different (internal) parts are shaded black (for A) or gray (for H). Complementary sequences are indicated by hatching. (B) General scheme for the Selective Suppression of PCR (SSP)-assisted coincidence cloning (CC) procedure. Fragments unique for each of the samples under comparison are shown as dotted and dashed lines. Identical fragments are depicted as continuous lines. B and C suppression adapters are represented by blank and filled boxes, respectively.
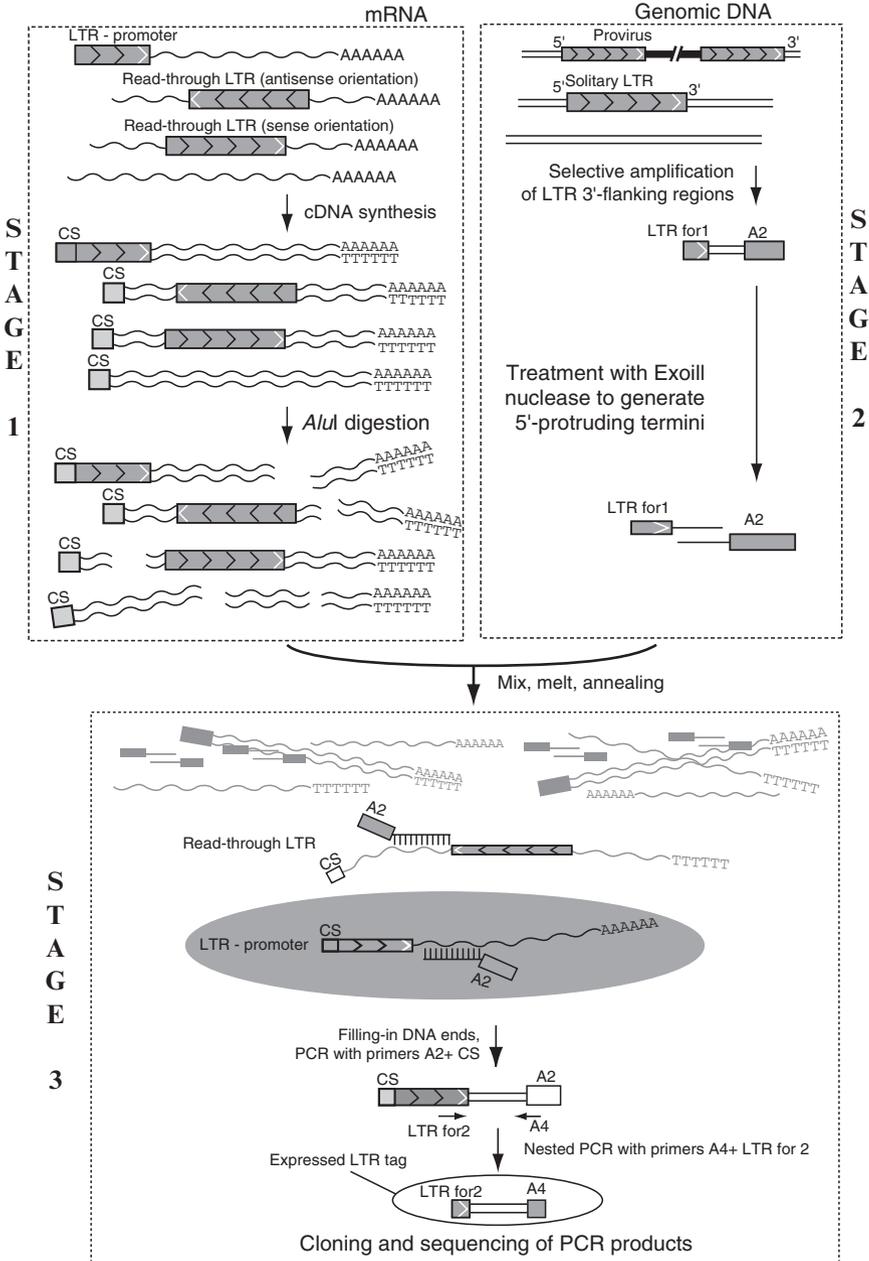
Figure 7. Schematic representation of the GREM technique (for details, see Section 6.2). The procedure includes three major stages: (1) genome-wide amplification of the genomic DNA flanking the 3′-ends of target repetitive elements (here, HS LTRs). Treatment of the resulting amplicon with *Exo*III generates 5′-protruding ends to be used at the third stage. (2) A double-stranded oligo d(T)-primed cDNA library is synthesized for tissues where expression of repetitive elements is to be studied.

(*Continued*)

amplification would not provide similar selectivity, as the exact locations of transcription start sites within repeat sequences may vary for different individual repetitive elements and, therefore, the design of suitable primers for PCR would be problematic.

To amplify genomic LTR flanking regions, we digested human genomic DNA with *Alu*I, ligated the fragments obtained to a 45 nt long GC-rich synthetic linker oligonucleotide (A1A2), and performed a series of nested PCR amplifications using HS LTR specific and adapter specific primers. As mentioned above, the HS LTR consensus sequence lacks *Alu*I restriction sites, whereas this endonuclease normally produces DNA fragments too short to be subject to PCR fragment size selection (Rebrikov et al. 2000). As mentioned above, the use of GC-rich suppression adapters minimizes background PCR amplification and results in almost 100% selective amplification of the expected fraction of the genome. The amplified LTR flanking sequences were treated with *Exo*III exonuclease to generate 5′-protruding termini required at stage (3) of GREM and to avoid any background cross-hybridization between LTR-containing sequences. We have recently demonstrated (Buzdin et al. 2002; Buzdin et al. 2003a) that *Exo*III may be used to remove adapter sequences from hybridizing mixtures. Under the conditions used, *Exo*III removes nucleotides slowly enough (~5 nucleotides per minute) to more or less precisely excise ~30 HS LTR 3′-terminal nucleotides from the amplicons. At the last step, the digested cDNA was hybridized to the LTR 3′-flanking genomic fragments. To selectively amplify the heteroduplexes containing genomic LTR flanking regions and cDNA 5′-terminal fragments generated due to LTR promoter activity, we used PCR with the CS primer against 5′-cDNA tags and A2 primer specific to the adapters ligated to the genomic DNA. This PCR step was followed by an additional nested PCR with primers A4 and LTRfor3 to increase the specificity of amplification (Figure 7).

As a result, only heteroduplexes, but not duplexes of cDNA not relevant to LTR expression or containing read-through LTRs, were amplified. A potential background of transcripts containing LTRs read-through in the sense direction was supposed to be negligible. A careful inspection of human transcribed sequence databases revealed in total 38 transcripts containing read-through HS LTRs, among them only four LTRs in the sense orientation. An *in silico* simulation of *Alu*I digestion suggested a complete removal of all such transcripts from GREM libraries.

Figure 7. cont'd. At this stage cDNAs are tagged by a linker oligonucleotide (CS) at the RNA transcription start sites using the "cap-switch" effect. cDNAs are then digested with *Alu* I restriction endonuclease that has no recognition sites within HS LTRs. This step precludes amplification of LTR sequences read-through in the sense orientation. (3) Finally, the genomic DNA amplicon (stage 1) is hybridized to the 5′-tagged cDNAs (stage 2). The protruding DNA ends are filled in with DNA polymerase, and the hybrids obtained (expressed LTR tags – ELTs) are nested PCR amplified with primers specific to the flanking genomic DNA adapter and cDNA 5′-terminal tag sequence, respectively.

The finally obtained amplified heteroduplexes were further cloned and sequenced. Every particular heteroduplex contained a 3′-HS LTR terminal portion, a fragment of the 3′-flanking genomic DNA, and an adapter sequence. Importantly, GREM makes it possible to characterize promoter activity of the repetitive elements in both qualitative and quantitative ways, as the number of particular heteroduplexes linearly correlates with the transcriptional activity of the corresponding promoters.

5. FUTURE PROSPECTS

The use of PCR suppression effect makes CC a method of choice for a number of applications involving finding a common fraction of nucleic acids in the samples under study. Incorporation of a stage of treatment with mismatch-sensitive nucleases significantly increases the fidelity of CC, which is especially important when complex sources like mammalian genomic DNAs are analyzed. The comparison of only two sources per one experiment may seem important limitation, but pooling several DNA samples in one source may, at least partly, solve this problem. CC, therefore, has bright perspectives to be widely used for (1) finding common transcripts in the analyzing tissue probes (e.g. for the recovery of genes coexpressed in cancer samples), (2) for experimental mapping of known or unknown transcripts on both characterized or uncharacterized genomic contigs, (3) for genome wide recovery of individual transcriptionally active repetitive elements, their mapping and quantification of their promoter activity, (4) for finding methylated or unmethylated CpGs in megabase-scale genomic contigs and, finally, (5) for the facile identification of evolutionary conserved sequences, shared by the comparing genomes.

6. PROTOCOLS

6.1 Cloning Similarities in Genomic DNAs

6.1.1 Starting material

DNA samples. In our experiments, we extracted DNA from four mixed human blood samples, or from blood samples of chimpanzee *Pan paniscus* and marmoset *C. pigmaea* using a genomic DNA purification kit (Promega lot #A7710) according to the manufacturers' recommendations.

Oligonucleotides. We used the standard suppression adapters A1A2 (5′-*GTAA-TACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT*-3′) and B1B2 (5′-*CGACGTGGACTATCCATGAACGCATCGAGCGGCCGCCCGGGCAGGT*-3′). For nested PCR amplifications, the following primers specific for the suppression adapter set were used: A1, 5′-*GTAATACGACTCACTATAGGGC*-3′, and B1, 5′-*CGACGTGGACTATCCATGAACGCA*-3′. A2, 5′-*AGCGTGGTCGCG-CCGAGGT*-3′, and B2, 5′-*TCGAGCGGCCGCCCGGGCAGGT*-3′. Oligonucleotides were synthesized using an ASM-102U DNA synthesizer (Biosan, Novosibirsk, Russia).

6.1.2 DNA preparation for hybridization

Digestion of genomic DNA. About 1μg of genomic DNA was digested with 10 units of frequent-cutter blunt end-producing restriction endonuclease *Alu*I (Fermentas) at 37°C, for 2 h. DNA was phenol–chloroform extracted, ethanol precipitated, and dissolved in 25 μl of sterile water.

Ligation of the suppression adapters. The suppression adapter ligation was done as described previously in this book (Lavrentieva et al. 1999). We used T4 DNA ligase (Promega) and suppression adapters A1A2 and B1B2 (see above), annealed to 10 nt long oligonucleotide complementary to the adapter 3′-terminal part, A3 and B3, respectively). Ligated DNA was purified using Quiagen PCR product purification kit, ethanol precipitated and dissolved in 5 μl of hybridization buffer (0.5 M NaCl, 50 mM Hepes, pH 8.3, 0.2 mM EDTA).

6.1.3 DNA hybridization

We mixed 800 ng of each of both DNA samples assigned for hybridization in a volume of 8 μl of 1x hybridization buffer, denatured at 95°C for 10 min, and hybridized at 65°C or 85°C for 50 h. The final 8 μl mixture was diluted with 72 μl of dilution buffer (50 mM NaCl, 5 mM Hepes, pH 8.3, 0.2 mM EDTA). In some experiments, C_0tA fraction competitor DNA (Gibco BRL, USA) was added in 100x weight excess to the hybridization mixture.

Filling in the termini of hybridized DNA. We used AmpliTaq DNA polymerase (1 unit per 1 μg of hybridized DNA) to fill in the ends of DNA duplexes at 72°C for 20 min.

6.1.4 Hybridized DNA treatment with mismatch sensitive nucleases

About 100 ng aliquots of hybridized DNA were digested with 1 μl Surveyor nuclease (Transgenomic, USA) in 20 μl of 1x buffer supplied by the manufacturer, overnight incubation at 42°C, or treated with 0.1 unit of mung bean nuclease (Promega) at 37°C for 15 min. DNA samples were phenol–chloroform extracted and ethanol precipitated.

6.1.5 PCR amplification of hybridization products and library construction

Nested PCR amplification. DNA samples were dissolved in 100 μl of water and 1μl was PCR amplified with 0.2 μM primers specific for the used suppression adapter set: A1 and B1. The PCR conditions were as follows: 95°C for 15″, 65°C for 10″,72°C for 90″, 15 cycles. To increase the amplification specificity, we used an additional round of nested PCR for 500-fold dissolved products of the latter PCR with 0.2 μM primers A2 and B2, under the same cycling conditions. The number of nested PCR cycles varied substantially depending on the particular hybridization.

Clone library construction. The PCR products obtained were cloned in *E. coli* strain DH5α using a TA-cloning system (Promega). We sequenced positive clones by the dye termination method using an applied biosystems 373 automatic DNA sequencer.

DNA sequence analysis. We used BLAT search (http://genome.ucsc.edu/cgi-bin/hgBLAT) to map clone inserts within human and chimpanzee genomes. Homology searches against GenBank were done using the BLAST web server at National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/BLAST) (Altschul et al. 1990). For multiple alignments the ClustalW program (Thompson et al. 1994) was used.

6.1.6 PCR amplification of evolutionary conserved sequences

As much as 40 ng of Old World monkey *C. pigmaea* blood DNA sample were PCR amplified using multiple sets of 0.2 μM unique genomic primers flanking the presumable conserved genomic loci. The resulting PCR products were analyzed on 1.2% agarose gels and sequenced.

6.2 Cloning and Presice Mapping of Transcribed Repetitive Elements

We applied this approach termed GREM to identify at a genome, wide scale of promoter active human-specific endogenous retroviruses and their LTRs (HS LTRs, see Section 4 (Figure 7). Below both adapter and LTR-specific primer structures are listed, to be adopted by the user depending on the specific task (which type of genomic repeats will be studied and which suppression adapter set will be used).

6.2.1 Starting material

Oligonucleotides. For linker ligation we used the standard suppression adapter A1A2 (5′-*GTAATACGACTCACTATAGGGCAGTCGACGCGTGCCCGGT-CCGAC*-3′) annealed to short oligonucleotide complementary to its 3′-end A3, 5′-*GTCGGACCGGGC*-3′. For nested PCR amplifications, the following primers specific for the suppression adapter set were used: A1, 5′-*GTAATACGAC-TCACTATAGGGC*-3′, A2, 5′-*AGCGTGGTCGCGGCCGAGGT*-3′, and A4, 5′-*TCGACGCGTGCCCGGTCCGACCT*-3′. LTR-specific primers were as follows: LTRfor1 (5′-*GTCTTGTGACCCTGACACATCC*-3′), LTRfor2, 5′-*CCTCCATATGCTGAACGCTG*-3′, and LTRfor3, 5′-*GGGGCAACCCACCC-CTAC*-3′. Oligonucleotides were synthesized using an ASM-102U DNA synthesizer (Biosan, Novosibirsk, Russia). For cap-swith based cDNA synthesis, the following oligonucleotides purchased from Clontech (USA) were used: CDS (5′-*AAGCAGTGGTATCAACGCAGAGTAC(T)$_{30}$*-3′), riboCS, 5′-*TAACAACGCA-GAGTACGC$_R$G$_R$G$_R$G*-3′ with three 3′-terminal ribonucleotides, and CS, 5′-*TAACAACGCAGAGTACGCGG*-3′.

DNA samples. We extracted DNA from three mixed human placentas using a genomic DNA purification kit (Promega lot #A7710) according to the manufacturers' recommendations.

Tissue sampling. Testicular parenchyma and seminoma were sampled from a surgical specimen under non-neoplastic conditions. Representative samples were divided into two parts, one of which was immediately frozen in liquid nitrogen, and the other was formalin-fixed and paraffin-embedded for histological analysis.

6.2.2 RNA isolation and cDNA synthesis

Total RNA was isolated from frozen samples pulverized in liquid nitrogen using an RNeasy mini RNA purification kit (Qiagen). All RNA samples were further treated with DNase I to remove residual DNA. Full-length cDNA samples were obtained according to a cap switch effect-based SMART cDNA synthesis protocol (Clontech, BD Biosciences) using an oligo(dT)-containing primer (CDS), PowerScript reverse transcriptase (Clontech, BD Biosciences), and a riboCS oligonucleotide. When PowerScript reverse transcriptase reaches the 5′-end of the mRNA, the enzyme's terminal transferase activity adds a few additional deoxycytidine nucleotides to the 3′-end of the cDNA. The riboCS oligonucleotide, which contains three guanineribonucleotide residues at its 3′-end, basepairs with the deoxycytidine stretch, creating an extended template. Reverse transcriptase then switches templates and continues the replication to the end of the oligonucleotide. The resulting full-length single-stranded cDNA contains 5′-terminal sequences complementary to the riboCS oligonucleotide. An Advantage 2 Polymerase mix (Clontech), CS, and CDS oligonucleotides were used to synthesize the second cDNA strands and to PCR-amplify double-stranded cDNA. Prior to further hybridization in the GREM procedure, 1 µg of cDNA was digested with 10 units of *Alu*I restriction endonuclease (Fermentas) for 3 h at 37˚C. This enzyme was used because the HS LTR consensus sequence lacks *Alu*I recognition sites.

6.2.3 Selective amplification of genomic regions flanking HS LTRs

Selective amplification of LTR 3′-flanking regions was based on the PCR suppression effect described in detail elsewhere (Siebert et al. 1995; Lavrentieva et al. 1999; Buzdin et al. 2002). Human genomic DNA (1 µg) was digested with 10 units of *Alu*I (Fermentas) restriction endonuclease, ethanol precipitated and dissolved in 20 µl of sterile water. Then, 100 pmol of annealed suppression adapters A1A2/A3 were ligated overnight to 300 ng of the digested DNA using three units of T4 DNA ligase (Promega) at 16°C. The ligated DNA was purified using Quiaquick purification columns (Quiagen) and eluted with 50 µl of water. About 1 µl of the eluted DNA was PCR amplified with the HS LTR-specific primer LTRfor1 and adapter-specific primer A1 using the following cycling program: (1) 72°C, 1′, (2) 95°C, 1′, and (3) 95°C, 15″; 65°C, 15″; 72°C, 1′ for 20 cycles. The PCR products were 500-fold diluted and used as templates for nested

PCR with the downstream HS LTR-specific primer LTRfor2 and adapter-specific primer A2 under the same cycling conditions, for 22 cycles. The amplified LTR flanking sequences were treated with *Exo*III exonuclease (Promega) to generate 5′-protruding termini exactly as described in (Buzdin et al. 2002; Buzdin et al. 2003a).

6.2.4 GREM procedure

The technique includes hybridization of PCR amplified genomic sequences flanking repetitive elements (HS LTRs in our case) with cDNA, followed by selective amplification and cloning of hybrid DNA duplexes (see Figure 7). About 100 ng of *Exo*III-treated LTR flanking sequences, obtained as described above, were mixed with 300 ng of cDNA in 4 μl of hybridization buffer (0.5 M NaCl, 50 mM HEPES, pH 8.3, 0.2 mM EDTA), overlaid with mineral oil, denatured at 95°C for 5 min and hybridized at 68°C for 14 h. The final mixture was diluted with 36 μl of dilution buffer (50 mM NaCl, 5 mM HEPES, pH 8.3, 0.2 mM EDTA), and 1 μl of the diluted hybridization mixture [BG1] was PCR-amplified with 0.2 μM adapter-specific primer A2 and 0.2 μM cDNA 5′-end-specific primer CS under the following conditions: (1) 72°C for 5 min to fill in the ends of DNA duplexes, (2) 95°C for 15″, 65°C for 15″, 72°C for 1′30″, 8 cycles. The PCR products were 500-fold diluted and reamplified by nested PCR for 20 cycles (95°C, 15″, 65°C, 15″, 72°C, 1′30″) with 0.2 μM nested adapter-specific primer A4, and 0.2 μM HS LTR 3′-end-specific primer LTRfor3. The final PCR products were cloned in *E. coli* using a pGEM-T vector system (Promega) and sequenced by the dye termination method using an Applied Biosystems 373 automatic DNA sequencer.

6.2.5 DNA sequence analysis and repetitive element transcriptional status control

DNA sequence analysis. The human specific HERV-K LTR group (HS) consensus sequence was taken from our previous work (Buzdin et al. 2003b). LTR flanking regions were investigated with the RepeatMasker program (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker; Smit AFA and Green P, unpublished data). Homology searches against GenBank were done using the BLAST web server at NCBI (http://www.ncbi.nlm.nih.gov/BLAST) (Altschul et al. 1990). To determine genomic locations of LTR flanking regions, the UCSC genome browser and BLAT searches (http://genome.ucsc.edu/cgi-bin/hgBLAT) were used.

Transcriptional status control. For RT–PCR control of LTR transcriptional status, we used pairs of primers, one of which was specific to the 3′-terminal part of a particular HS LTR (primer sequences not shown), and the other specific to a unique sequence within the corresponding genomic LTR 3′-flanking region. Prior to the RT–PCR analysis, the priming efficiency of the primers was preexamined by genomic PCRs at temperatures varying depending on the primer combination used. These PCRs were done for 19, 22, 25, and 28 cycles, with 40 ng of the human genomic DNA template

isolated from testicular parenchyma. The RT–PCR was done with cDNA samples of the same tissue, an equivalent of 20 ng total RNA being used as template in each PCR reaction performed in a final volume of 40 µl. About 5 µl aliquots of the reaction mixture after 21, 24, 27, 30, 33, 36, and 39 cycles of the amplification were analyzed by electrophoresis in 1.5% agarose gels. In all cases, the transcriptional status was determined from the number of PCR cycles needed to detect a PCR product of the expected length and the PCR product concentration measured using a Photomat system and the Gel Pro Analyzer software.

6.3 Finding Methylated or Unmethylated CpGs in Large Genomic Contigs

6.3.1 *Materials and trivial protocols*

Growth and transformation of *E. coli* cells, preparation of plasmid DNA, polyacrylamide gel electrophoresis (PAGE) and agarose gel electrophoresis, as well as other standard manipulations are performed as described by Sambrook et al. (1989). A model cosmid library representing the *D19S208-COX7A1* locus on human chromosome 19 was provided by Dr. Lisa Stubbs (Lawrence Livermore National Laboratory). The cells were grown overnight at 37°C in 5 ml of LB medium supplemented with kanamycin (20 mg/ml). Cosmid DNA was isolated using a Wizard Plus Miniprep DNA Purification System (Promega) according to the manufacturers' recommendations. DNA and samples from normal testis and seminoma tissues were kindly provided by Dr. Lyudmila Leppik. Clone inserts were sequenced with an Amersham Biosciences/Molecular Dynamics MegaBACE4000 Capillary Sequencer.

6.3.2 *Selective amplification of AluI-HpaII fragments*

Amplicon preparation from AluI-HpaII fragments obtained from cosmids. Cosmid DNA samples were digested with *Alu*I (MBI Fermentas, Vilnius, Lithuania), and suppression adapter A1A2 from the Section 6.2.1 was ligated to the resulting fragments as described in Chapter 2. The ligated fragments were then further digested with *Hpa*II (MBI Fermentas) and then ligated to suppression adapter Hpa (obtained by annealing an equimolar mixture of oligonucleotides 1 and 2: *5′-GTAATACGACTCACTATAGGGCAGGGCGTG GTGCGGAGGGCGGC-3′*(1) and 5′-*CGGCCGCCCTCC-3′* (2)). Suppression fragments were used as templates in a two-step selective amplication. The first PCR was performed using 10 ng of ligated DNA as template in a 25 µl volume containing 10 pmol of external primer I (Advantage 2 PCR kit,Clontech; 5′-*GTAATACGACTCACTATAGGGC*-3′). After a 5 min incubation at 72°C the mixture was PCR amplified for 25 cycles (94°C for 30″, 66°C for 30″, and 72°C for 90″). The second-step PCR was performed for 10 cycles (94°C for 30″, 68°C for 30″, 72°C for 90″) using 1 µl of the one tenths

diluted amplicon as template and internal primers I-Alu (5′-*AGCGTG-GTCGCGGCCGAGAG*-3′) and I-Hpa (5′-*AGGGCGTGGTGCGGAGGGCGGC*-3′). In order to remove flanking sequences corresponding to the primers used in the second PCR step, the amplicons were digested with *Alu*I and *Hpa*II. To fill in the ends after the *Hpa*II reaction, the samples were treated with the Klenow fragment of *E. coli* DNA polymerase I (Promega) and purified by phenol–chloroform extraction followed by ethanol precipitation. The suppression adapter B1B2 (obtained by annealing an equimolar mixture of oligonucleotides 3 and 4: 5′-*TGTAGCGTGAAGACGACAGAATCGAGCG-GCCGCCCGGGCAGGT*-3′ (3) and 5′-*ACCTGCCC*-3′ (4)) was ligated to the set of fragments so obtained.

Amplicon preparation from genomic AluI-HpaII fragments. The amplicons were prepared as described for cosmid DNA with two exceptions: (1) the second PCR step was performed with the 5′-phosphorylated primer I-Hpa, and (2) the PCR product was digested with *Alu*I only. After digestion, the set of fragments was purified and ligated to suppression adapter C (obtained by annealing an equimolar mixture of oligonucleotides 5 and 6: 5′-*TGTAGCGTGAAGAC-GACAGAAAGTCGACGCGTGCCCGGGCTGGT*-3′ (5) and 5′-*ACCAGC-CC*-3′ (6)).

6.3.3 The NGSCC procedure

Two mixes, relating *Hpa*II/normal and *Hpa*II/tumor genomic DNAs, respectively, were prepared, each containing 200 ng of the cosmid amplicon and 6 µl of the appropriate genomic DNA amplicon in 3 µl of HB hybridization buffer (50 mM HEPES pH 8.3; 0.5 M NaCl, 0.02 mM EDTA pH 8.0, 10% w/v PEG 8000). The mixtures were overlaid with ~30 µl of mineral oil and incubated for 3 min at 99°C (denaturing) and then for 18 h at 68°C (reannealing). After this, 200 µl of prewarmed HB (68°C) was added to each tube and 1 µl thereof was used in the following PCR as template. Reactions were performed in a 25 µl volume containing 10 pmol of external primer II (5′-*TGTAGCGTGAAGAC-GACAGAA*-3′). After preincubation for 5 min at 72°C, reaction mixtures were subjected to PCR for 25 cycles of 94°C C for 30″ s, 66°C C for 30′ and 72°C for 90″. The second PCR was performed for 10 cycles of 94°C for 30″, 68°C for 30″ and 72°C for 90″, using 1 µl of the one tenths diluted first PCR product as the template in a 25 µl volume containing 10 pmol of the internal primers II-B and I-Hpa (5′-*TCGAGCGGCCGCCCGGGCAGGT*-3′ and 5′-*GTAATAC-GACTCACTATAGGGC*-3′, respectively). The two PCR product mixtures were ligated into a pGEM-T vector (Promega) and cloned in *E. coli*. The clones were arrayed in microtiter plates, and the inserts were sequenced. The sequences obtained were mapped by comparison with those deposited in GenBank using the BLAST (Altschul et al. 1990) web server at NCBI (http://www.ncbi.nlm. nih.-gov/BLAST). The data were further analyzed using the Draft Human Genome Browser (http://genome.ucsc.edu/ goldenPath/hgTracks.html).

Bisulfite sequencing. To approve the true DNA methylation status, bisulfite sequencing was performed for some differential CpGs found using NGSCC. Primers specific for bisulfite modified DNA were designed using MethPrimer software (http://itsa.ucsf.edu/_urolab/methprimer/). Bisulfite modification was performed according to (Olek et al. 1996). Agarose beads were used directly in two successive PCRs. The first and second reactions were performed for 40 and 25 cycles, respectively (95°C for 20″, Annealing temperature varied upon individual primer combinations – for 20″, 72°C for 40″), in a 25 μl volume containing 10 pmol of each primer and 1 U of Taq DNA Polymerase (GibcoBRL). The fragments obtained were sequenced using the standard PCR products sequencing protocol.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Aslanidis C, de Jong PJ (1991) Coincidence cloning of Alu PCR products. Proc Natl Acad Sci USA 88:6765–6769

Azhikina T, Gainetdinov I, Skvortsova Y, Batrak A, Dmitrieva N, Sverdlov E (2004) Non-methylated genomic sites coincidence cloning (NGSCC): an approach to large scale analysis of hypomethylated CpG patterns at predetermined genomic loci. Mol Genet Genomics 271:22–32

Azhikina T, Gainetdinov I, Skvortsova Y, Sverdlov E (2006) Methylation-free site patterns along a 1-Mb locus on Chr19 in cancerous and normal cells are similar. A new fast approach for analyzing unmethylated CCGG sites distribution. Mol Genet Genomics 275:615–622

Azhikina TL, Sverdlov ED (2005) Study of tissue-specific CpG methylation of DNA in extended genomic loci. Biochemistry (Mosc) 70:596–603

Brookes AJ, Slorach EM, Morrison KE, Qureshi SJ, Blake D, Davies K, Porteous DJ (1994) Cloning the shared components of complex DNA resources. Hum Mol Genet 3:2011–2017

Buzdin A, Khodosevich K, Mamedov I, Vinogradova T, Lebedev Y, Hunsmann G, Sverdlov E (2002) A technique for genome-wide identification of differences in the interspersed repeats integrations between closely related genomes and its application to detection of human-specific integrations of HERV-K LTRs. Genomics 79:413–422

Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E (2006a) At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. J Virol 80:10752–10762

Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E (2006b) GREM, a technique for genome-wide isolation and quantitative analysis of promoter active repeats. Nucleic Acids Res 34:e67

Buzdin A, Ustyugova S, Gogvadze E, Lebedev Y, Hunsmann G, Sverdlov E (2003a) Genome-wide targeted search for human specific and polymorphic L1 integrations. Hum Genet 112:527–533

Buzdin A, Ustyugova S, Khodosevich K, Mamedov I, Lebedev Y, Hunsmann G, Sverdlov E (2003b) Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. Genomics 81:149–156

Chalaya T, Gogvadze E, Buzdin A, Kovalskaya E, Sverdlov ED (2004) Improving specificity of DNA hybridization-based methods. Nucleic Acids Res 32:e130

Devon RS, Brookes AJ (1996) Coincidence cloning. Taking the coincidences out of genome analysis. Mol Biotechnol 5:243–252

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (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

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Lavrentieva I, Broude NE, Lebedev Y, Gottesman II, Lukyanov SA, Smith CL, Sverdlov ED (1999) High polymorphism level of genomic sequences flanking insertion sites of human endogenous retroviral long terminal repeats. FEBS Lett 443:341–347

Lovett M, Kere J, Hinton LM (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. Proc Natl Acad Sci USA 88:9628–9632

Luk'ianov SA, Gurskaia NG, Luk'ianov KA, Tarabykin VS, Sverdlov ED (1994) [Highly-effective subtractive hybridization of cDNA]. Bioorg Khim 20:701–704

Milner JJ, Cecchini E, Dominy PJ (1995) A kinetic model for subtractive hybridization. Nucleic Acids Res 23:176–187

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 86:6686–6690

Olek A, Oswald J, Walter J (1996) A modified and improved method for bisulphite based cytosine methylation analysis. Nucleic Acids Res 24:5064–5066

Parimoo S, Patanjali SR, Shukla H, Chaplin DD, Weissman SM (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

Rebrikov DV, Britanova OV, Gurskaya NG, Lukyanov KA, Tarabykin VS, Lukyanov SA (2000) Mirror orientation selection (MOS): a method for eliminating false positive clones from libraries generated by suppression subtractive hybridization. Nucleic Acids Res 28:e90

Sambrook J, Russell DW (2001) Molecular cloning: a practical manual. CSHL Press, Cold Spring Harbour, New York

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 23:1087–1088

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al. (2001) The sequence of the human genome. Science 291:1304–1351