CHAPTER 10

# CURRENT ATTEMPTS TO IMPROVE THE SPECIFICITY OF NUCLEIC ACIDS HYBRIDIZATION

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**: Being very useful and informative, many techniques based on nucleic acids hybridization suffer from the cross-annealing of repetitive DNA, presenting in reassociating samples. This "wrong" annealing causes "nonspecific" hybridization of nonorthologous DNA fragments, thus producing chimeric sequences and at the final stage significantly hampering the analysis of the resulting cDNA or genomic libraries. Such chimeras may constitute up to 40–60% of DNA libraries. Importantly, the number of chimerical clones positively correlates with the complexity of hybridizing genomic or cDNA mixtures. The hybridization specificity is a crucial factor determining both the fidelity the efficiency of all hybridization-based analytical techniques. In this chapter, I review the current attempts to increase the specificity of hybridization at both stages: during nucleic acids reassociation and at the stage of selection of proper hybrids. To this end, approaches based on chemical modifications, improving hybridization kinetics, and improving selection of perfectly matched duplexes, have been developed.

**Keywords**: Hybridization specificity, PCR selection effect, targeted genomic difference analysis (TGDA), thermodynamically stable duplexes, phenol emulsion reassociation technique (PERT), repetitive element, genomic repeat, chimerical clone, hybridization temperature, melting point, perfectly matched hybrids, mispaired DNA rejection (MDR), TILLING, mismatch-sensitive nuclease, mung bean nuclease, $C_o tA$ fraction.

**Abbreviations**: MDR, mispaired DNA rejection; PCR, polymerase chain reaction; PERT, phenol emulsion reassociation technique; RDA, representative differential analysis; RE, repetitive element; SH, subtractive hybridization; SSH, suppression subtractive hybridization; TGDA, targeted genomic difference analysis.



*A. Buzdin and S. Lukyanov (eds.), Nucleic Acids Hybridization,* 241–264.

**TABLE OF CONTENTS**



## 1. INTRODUCTION

Many popular (approximately 300 PubMed citations per year) experimental techniques for genome and transcriptome analysis, such as coincidence cloning (Chapter 8) and subtractive hybridization (SH) (Chapters 3, 6, 7), including representative differential analysis (RDA) (Lisitsyn and Wigler 1995) and suppression subtractive hybridization (SSH) (Diatchenko et al. 1996), are based on DNA hybridization in solution, followed by polymerase chain reaction (PCR) amplification of certain hybridized fractions (Sasaki et al. 1994; Nagayama et al. 2001).

The hybridization specificity is a crucial factor determining both the fidelity of the natural biological processes and the efficiency of hybridization-based analytical techniques. Hybridization specificity ($f$) is determined as a relative factor for match versus mismatch discrimination: $f = \exp\text{-}|\Delta G_{\text{m-mm}}/RT|$, where $G_{\text{m-mm}}$ is the free energy penalty for binding to sites that differ from the perfectly complementary sequences by a single base-pair substitution. If $G_{\text{m-mm}}$ is ~4 kcal/mol (Roberts and Crothers 1991), hybridization specificity will be ~1/100–1/1000, and theoretically it is possible to find a range of conditions (so-called stringency conditions) where perfect complexes will be substantially more stable than the complexes containing mismatches (Broude 2002).

However, being very useful and informative, the techniques based on nucleic acids hybridization are not free from some imperfections. The well-known disadvantage of complex DNA mixture hybridization is the cross-annealing of repetitive DNA, presenting in reassociating samples (Hames and Higgins 1985). This "wrong" annealing causes "nonspecific" hybridization of nonorthologous DNA fragments, thus producing chimeric sequences and at the final stage significantly hampering the analysis of the resulting cDNA or genomic libraries (see Figure 1). Such chimeras may constitute up to 40–60% of DNA libraries.

Importantly, the number of chimerical clones positively correlates with the complexity of hybridizing genomic or cDNA mixtures. This is probably the

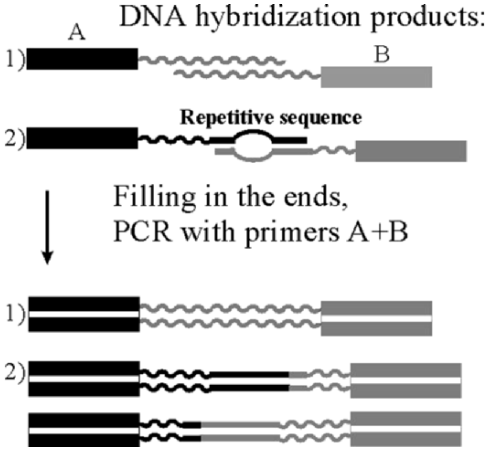


*Figure 1.* Along with the normal duplex formation, chimeric hybrids between the repetitive sequences may occur, especially when complex genomic mixtures are hybridized. In many techniques, these "wrong" hybrids are not properly recognized and are further amplified and included in the final clone libraries, thus forming up to 60% of the overall sequence information.

major reason that limits SH applications to the comparison of cDNA samples (representing only the modest part of genomic sequences) and of small genomes (such as prokaryotic, yeast or planarian DNA; see Chapters 1 and 3). SH-based approaches in their present form are hardly applicable to the recovery of differences between the complex genomic DNAs like mammalian ones. Two major related techniques dealing with complex genomic DNA subtractions, namely RDA (Lisitsyn and Wigler 1993; Lisitsyn and Wigler 1995) and targeted genomic difference analysis (TGDA) (Buzdin et al. 2002; Buzdin et al. 2003a), all employ dramatic genomic DNA simplification prior to hybridization step.

In the first case, this simplification is achieved by means of the so-called PCR selection effect, when the total pool of fragmented genomic DNA with ligated adapters is PCR preamplified for 50–100 cycles. This results in a great bias in different DNA fragment concentrations in the resulting amplicons: most of the sequences turn to be underrepresented or completely lost due to rather inefficient PCR amplification with *Taq* DNA polymerase, whereas the others, forming relatively small (~10% or less) fraction of the initial pool, are overrepresented because of an optimal length/GC-content ratio making them preferable targets for *Taq* polymerase. Therefore, RDA utilizes a random genomic DNA simplification based on the fragment size and GC-content. Consequently, the resulting pool of differential sequences, which appear after the subtraction stage, lacks most of the differential sequences presented in the genome originally. RDA is, thus, worth applicable for the recovery of some marker sequences, but cannot be used for comprehensive genome or cDNA analyses. The second approach TGDA is based on a specific PCR amplification of a group of genomic sequences of

interest (e.g. sequences flanking insertions of human retroelements amplified with primer(s) specific to the retroelement 5′- or 3′-terminus (Buzdin et al. 2002; Buzdin et al. 2003a; Mamedov et al. 2005)). These sequences are selectively amplified with the reasonable number of PCR cycles (25–40 depending on the requirement of nested PCR amplification), and are further subtracted resulting in a comprehensive library enriched in DNAs presented in one of the comparing samples (tracer) but absent from the others (driver). The strength of TGDA is the complete rather than random (as for RDA) recovery of differential sequences of the interest (different groups of repetitive elements (REs), multigene family members, pseudogenes, duplicated, or multiplicated genomic loci). At present, this technique has been successfully applied to the recovery of human-specific endogenous retroviruses (Buzdin et al. 2002), L1 retrotransposons (Buzdin et al. 2003a), and for the experimental identification of polymorphisms created in human populations by the insertions of *Alu* repeats (Mamedov et al. 2005).

However, both TGDA and RDA approaches will be inefficient for the complete comparisons of the whole genomes. Of course, such a comparison may be done by means of complete genome sequencing (which is extremely expensive and time consuming; note, even now multiple gaps in human genome assembly are not filled [http://genome.ucsc.edu/cgi-bin/hgGateway]), but it is low probable that in the nearest future it will be possible to compare, say, 100 individual human genomes through the complete genome sequencing. Theoretically, for many applications SH could become an alternative to shotgun genome sequencing (except for identification of "fine" differences like single nucleotide substitutions or microsatellite length polymorphisms). "For many applications" means screening for relatively large deletions, duplications, translocations, insertions of pseudogenes, and transposable elements, as well as of the exogenous sequences like retroviruses. To improve the existing in-solution nucleic acids hybridization techniques and, in particular, subtraction-based methods, two major approaches seem reasonable:

– To improve hybridization kinetics (in order to insure that only the most thermodynamically stable duplexes [i.e. those lacking mispaired nucleotides] are formed during the stage of nucleic acids hybridization)
– To improve the recognition of perfectly matched duplexes (e.g. by selectively PCR amplifying them).

Both approaches appear to be fruitful in many cases.

## 2. IMPROVING HYBRIDIZATION KINETICS

Coincidence cloning, SH, and other techniques described in this book, all follow the uniform rules of DNA hybridization in solution. Although hybridization kinetics was better studied for SH (Sverdlov and Ermolaeva 1994; Ermolaeva and Wagner 1995; Milner et al. 1995; Ermolaeva et al. 1996), all major conclusions and theoretical considerations defined for SH, will be true for other techniques based on nucleic acids hybridization in solution as well. SH has

become the practice since 1984, when Palmer and Lamar (1984) proposed a simple general idea of specific separation of enriched tracer–tracer homoduplexes from other components of the reannealed mixture: tracer DNA fragments (those containing differential sequences to be found) should have termini different from those of driver fragments (those serving as the background for differential tracer fragments; see Chapters 1, 3, 6, and 7). The authors prepared a mouse recombinant library enriched in Y chromosome fragments. The female mouse DNA (driver) was randomly cut into fragments, whereas the male DNA (tracer) was cleaved with *Mbo* I restriction endonuclease. Both DNAs were mixed at a ratio of 100:1, respectively, denatured, and reannealed. Only reassociated tracer homoduplexes contained sticky *Mbo* I ends at both termini and therefore could be selectively ligated to a *Bam* HI-digested pBR322 vector. This principle has been successfully applied to the isolation of DNA probes corresponding to a deletion spanning Duchenne muscular dystrophy locus (Kunkel et al. 1985).

The expected enrichment of the subtracted DNA with a sequence difference ($E^{d}$ ($t$) value) is expressed by a formula (Ermolaeva and Sverdlov 1996):

$$E^{d}(t) = (1 + RD_0t)/(1 + RT_0t)$$

where $R$ [$M^{-1}s^{-1}$] is the reassociation rate constant, and $D_0$ and $T_0$ are initial molar concentrations of driver and tracer, respectively. The maximum enrichment at $t \rightarrow \infty$ is $D_0/T_0$. For finite $t$ values, like 14 h (overnight incubation), the enrichment value increases as $RD_0$ increases. Thus, to reach better results, one should increase the values of $R$, $D_0$, or both. As an example, Lamar (Lamar and Palmer 1984) and Kunkel (Kunkel et al. 1985), with coauthors, clearly realizing that the rate of hybridization is crucial to achieve substantial enrichment, increased the $R$ value by addition of chemical accelerators, such as phenol, to reannealing mixtures.

However, mammalian genomes are too complex to reach sufficiently high $D_0$ values, and only major differences (like presence/absence of Y chromosome or extended deletions) can be isolated in such a way. As genome size increases beyond $5 \times 10^8$ bp (complexity comparable with that of arabidopsis or drosophila genomes), the kinetics of hybridization start to become an increasingly important factor limiting enrichment of the target (Milner et al. 1995). To enhance the kinetics of hybridization, increased hybridization times, higher driver concentrations, greater driver/tracer ratios, longer DNA fragments, and the use of techniques that enhance the rate of reassociation, e.g. phenol emulsion reassociation technique (PERT) (Kohne et al. 1977) or solvent exclusion (Barr and Emanuel 1990), may be effective. At a driver DNA concentration of 3.125 mg/ml the effective enhancement under PERT is only 2.2-fold (Kohne et al. 1977). However, compared to the rate in 1.0 M NaCl, the relative enhancement in the presence of 11% dextran sulphate and 1.5 M NaCl is 11.9-fold (Kohne et al. 1977; Barr and Emanuel 1990), although driver DNA concentration is limited to 1.0 mg/ml, reducing the achievable enhancement (Kohne et al. 1977; Barr and Emanuel 1990) to about fourfold.

### 2.1 Simplification of Hybridizing Mixtures

One of the widely used genomic subtraction schemes, called representational difference analysis (RDA) (Lisitsyn and Wigler 1995), involves selective PCR amplification of reannealed double-stranded tracer fragments, with all other types of hybrids not amplified. To ensure the selectivity, authors used special PCR adapters that functioned in the same way as *Mbo* I ends of the renatured tracer in previous works (Lamar and Palmer 1984; Kunkel et al. 1985). To increase the enrichment, RDA uses complexity reduction of driver and tracer DNA before subtraction. To this end, driver and tracer DNA fragments are repeatedly amplified so that the resulting driver and tracer represent a depleted pool of initial fragments as a result of the size-bias of PCR amplification. These simplified, fragmented genomes (or amplicons) are then used for subtraction. Thus, in the RDA technique (Lisitsyn and Wigler 1993), $D_0$ is increased by means of random genome simplification that allows molar concentrations of driver and tracer to be increased at the same mass concentrations. The simplification and repetitive cycles of subtraction make it possible to obtain very high enrichment of tracer with target fragments (Lisitsyn et al. 1994a). RDA was successfully used to clone DNA losses and amplifications in tumors (Carulli et al. 1998) and to generate specific genetic markers linked to a trait of interest (Lisitsyn et al. 1994b). An alternative to great PCR-cycle-number–based simplification of the hybridizing mixtures is the use of infrequent-cutter endonucleases providing very limited sets of fragments, which would not be too long for PCR amplification (see more detailed insight in Chapter 6, Section 3).

However, an evident RDA drawback is that due to random genome simplification only a minor fraction of the genome (2–10%) is actually compared, while 90–98% remains beyond the analysis. Therefore, isolation of genomic differences using this technique is, in a sense, a matter of luck. When TGDA is used (Buzdin et al. 2002), the genome simplification is targeted, the simplified fraction is not random but contains a fairly definite portion of the genome. The complexity $C$ of the simplified portion depends on the repetitive target content in the genome. With $N$ target repeats in the genome and an average size of the amplified fragments of ~256 bp (an average fragment produced by a frequent-cutter restriction endonuclease), $C = 256\ N$. In the case of human endogenous retroviral family HERV-K (which was the first target for TGDA application), which amount to about 2000 in number of representatives (Buzdin et al. 2003b), $C$ is as low as $\sim 5 \times 10^5$, which is only 0.017% of the whole human genome complexity. This results in a dramatic ($3.6 \times 10^7$) increase in the hybridization rate of the simplified versus the original genomic DNA, providing that mass concentrations during the subtraction are the same (Ermolaeva et al. 1996). The mass concentrations used by the authors (Buzdin et al. 2002; Buzdin et al. 2003a) (150 ng of driver DNA and 1.5 ng of tracer DNA per 1 μl) correspond to the driver and tracer molar concentrations of $5 \times 10^{-10}$ and $5 \times 10^{-12}$. At $R = 10^6$ (Hames and Higgins 1985), one could expect ~20-fold enrichment after 14 h of

hybridization. If the DNA is not simplified, the enrichment will be just negligible. The enrichment value 16, experimentally found by the authors, was in good agreement with the theory.

The successful application of TGDA partly depends on the divergence between the members of the repetitive element group under comparison (transposons, pseudogenes, etc.). If this divergence is high, oligonucleotide primers designed using the group consensus sequence may fail to prime PCR with the group members diverged too far from the consensus. However, the technique is aimed at the comparison of highly homologous REs forming evolutionarily young groups (sequence divergence less than 10%), with integrations polymorphic between closely related species or even within one species. The detailed TGDA protocol is given in the Section 4.1.

If sequences of interest are unique, then TGDA cannot be applied and a very serious problem appears when working with complex genomes: low reassociation rate of the nonrepetitive DNA. Indeed, DNA reassociation rate for each particular fragment is proportional to the square of its concentration; therefore, REs 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, so that ~99% reassociation may take months or years, an enormous background of repetitive sequences appears when reasonable (~days) hybridization time is used. In this case, a great majority of double-stranded molecules in solution are reassociated repeats, whereas unique sequences mostly still in a single-stranded form. Therefore, the true genomic sequence representations will be enormously biased in such clone libraries.

A rather efficient attempt to improve the situation is the addition of competitor DNA fractions containing 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 "$C_o$t A" and "$C_o$t 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) from the hybridization mixture 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 Section 4.2), or using

biotin–streptavidin systems. To our experience, the use of $C_o$t 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 REs, occupying approximately two thirds of human DNA (Lander et al. 2001; Venter et al. 2001). Nevertheless, the number of chimerical clones representing improperly matched duplexes remained high in such libraries (~56%). To achieve better results with this approach, one has to titer genomic and competitor DNA concentrations.

Another important parameter, hybridization temperature, is a well-known regulator of hybridization specificity: the lower the temperature, the more is the background. However, after the value of 65°C the temperature increase has little or no effect on the hybridization specificity (Chalaya et al. 2004). For instance, exactly the same proportion of chimerical clones (~56%) was produced when frequent-cutter enzyme-digested human genomic DNA was hybridized at 65°C and 85°C.

### 2.2 Chemical Modifications

As mentioned above, the hybridization temperature increase from 65°C to 85°C has absolutely no effect on the hybridization specificity (~100–300 bp long fragments of human genomic DNA hybridized). This means that both perfectly and imperfectly matched duplexes of such lengths are stable enough to be formed at 85°C. Therefore, using temperature as an instrument of hybridization specificity control requires shorter hybridizing fragments. However, it is frequently puzzling to unambiguously map fragments shorter than 100 bp in the genomic sequence. Titering hybridization temperature conditions from 85°C (when ~56% background duplexes are formed) to 94°C (when DNA is denatured) will be probably helpful for creating higher-hybridization-fidelity systems; however, this temperature interval is rather small, and the data obtained are frequently hardly reproducible.

The approach based on DNA chemical modifications, or synthesis of DNA analogs, makes it possible to increase the stability of complementary nucleotide interactions, thus to increase hybridization temperatures and, therefore, to significantly enlarge the operational temperature interval between the point where hybridization specificity is not sufficient, and the melting point (see, e.g. Mouritzen et al. 2003). The further "fine-tuning" of the temperature conditions will make it possible to find out the conditions ensuring the highest specificity of hybridization for genomic DNA fragments of a given length.

Another group of methods utilizes chemical modifications for discriminating mispaired versus perfectly matched DNA (or RNA) duplexes (Cotton et al. 1988). Under the special conditions, some chemicals do preferentially modify mismatched nucleotides due to their higher availability, and the following specific glycoside bond cleavage results in a degradation of "wrong" hybrids.

For mispaired C bases modification, hydroxylamine (NH2OH) is used, whereas mispaired T was originally modifying with osmium tetroxide (OsO4) (Cotton 1999), which is now replaced by potassium permanganate (KMSnO4) solution with a coadditive triethylammonium chloride due to high toxicity of the former (Roberts et al. 1997). Modified DNAs are further simultaneously cleaved by piperidine and purified. However, these products cannot be efficiently cloned or PCR-amplified, most probably, due to some background DNA modifications.

The main advantage of this approach includes nearly 100% efficiency in cleaving hybrids having mispaired C or T nucleotides, whereas its important shortcomings are multiple manipulations and the fact that toxic chemicals are required. Finally, the most serious drawback of all chemical modifications-based methods for improving hybridization selectivity is that the products cannot be further PCR-amplified and cloned.

## 3. IMPROVING SELECTION OF PERFECTLY MATCHED HYBRIDS

Improving selection of perfectly matched duplexes in hybridization mixtures is an alternative promising approach aimed to remove background chimerical sequences from the resulting clone libraries. This approach does not deal with the improvement of hybridization conditions, but, instead, it is focused on the exclusive amplification of the "proper" hybrids (Figure 1). To this end, hybridized DNA may be treated with some chemical reagents specifically modifying mispaired nucleotides or producing double- or single-strand DNA breaks there. However, the PCR amplification and cloning of such chemicals-treated DNA is problematic, and the second approach comprising hybridized DNA exposure to the nucleases specifically recognizing improperly matched DNA, seems to be advantageous. The method called mispaired DNA rejection (MDR), recently published by Chalaya and coauthors (Chalaya et al. 2004) makes it possible to almost completely exclude the chimerical sequences from analyzing DNA subsets (Figure 2). The technique is based on the observation that overwhelming majority of cross-hybridizingREs, although sharing considerable sequence similarity, are not entirely identical to each other. Their DNA heteroduplexes are therefore imperfectly matched, having quite a number of mispaired bases. These latter can form single nucleotide mismatches or even extended single-stranded DNA loop regions. All such structural deviations from normal properly paired DNA duplexes can be recognized and cut by certain enzymes, termed here mismatch-specific nucleases. Mispaired DNA sensitive nucleases, serving *in vivo* as reparation or viral life cycle machinery units, are now successfully employed by investigators for mutation detection. Such approaches are both simple and rather efficient, such as, for example, TILLING technique for large-scale mutation screening (Till et al. 2004), Surveyor mutation detection system (Qiu et al. 2004), and elegant high-fidelity technique for endonuclease/ligase-based mutation scanning by Huang and others (Huang et al. 2002). The most commonly used mismatch-specific nucleases are phage T7 endonuclease I (Babon

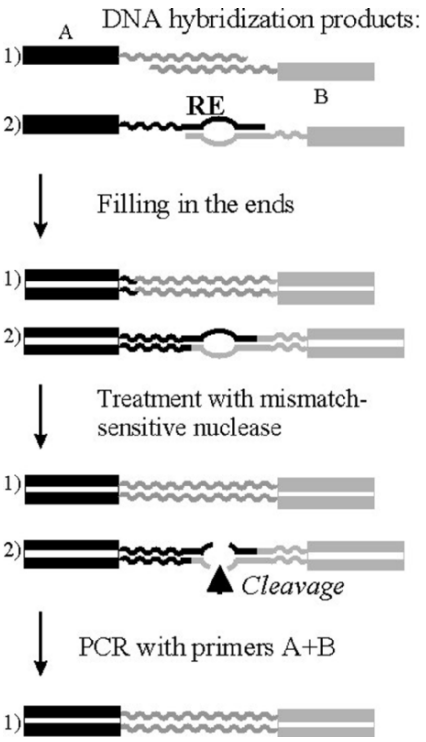


*Figure 2.* Outline of the mispaired DNA rejection (MDR) approach. Nonperfectly matched DNA hybrids, which constitute most of chimeric sequences, are recognized and cleaved by mismatch-sensitive nucleases, thus preventing their exponential amplification in the final PCR(s) and strongly increasing the occurrence of target nonchimeric sequences in the resulting clone libraries.

et al. 2003), T4 endonuclease VII (Mikhailov and Rohrmann 2002), modified bacterial endonuclease V (Huang et al. 2002), plant CEL I and Surveyor nucleases (Kulinski et al. 2000; Qiu et al. 2004). The authors demonstrated that these enzymes, cleaving DNA at mispaired base positions, can be used for eliminating chimerical hybrids from DNA hybridization mixtures, thus strongly reducing the number of background chimerical clones from 44–60% to 0–4%. MDR can be applied to both cDNA and genomic DNA subtractions of very complex DNA mixtures. This technique was also useful for the genome-wide recovery of highly conserved DNA sequences, as demonstrated by comparing human and pygmy marmoset genomes (Chalaya et al. 2004).

In order to investigate MDR efficiency, Chalaya et al. used the testing system (see Figure 3) comprising (1) digestion of mammalian genomic DNA with frequent-cutter enzyme, (2) ligation of different oligonucleotide suppression adapters (required for the PCR-suppression effect described in the Chapter 2) to digested DNA, (3) melt and annealing of two DNA portions harboring different

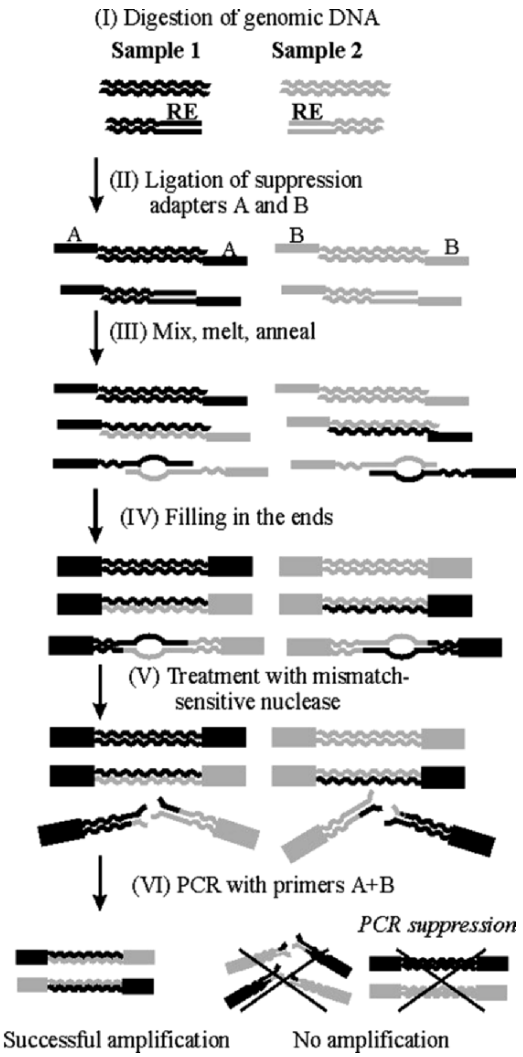


*Figure 3.* The testing system used to investigate MDR efficiency (see text). The use of MDR reduced the background chimeric clone proportion from 44–60% to 0–4%.

adapters, (4) filling-in the ends of DNA duplexes with DNA polymerase, (5) treatment with mismatch-sensitive nuclease, and (6) PCR amplification of heteroduplexes, that were not cleaved at the previous stage, with primers specific to both adapters using PCR-suppression effect, described in details earlier (Gurskaya et al. 1996).

Briefly, it includes the ligation of restriction fragments to a panhandle-like structure-forming adapter. The authors used standard adapters (Lavrentieva

et al. 1999) forming after ligation to restriction fragments ~40 bp long GC-rich inverted repeats at their termini. Therefore, such single-stranded DNA fragments contained self-complementary termini capable of forming strong intramolecular stem-loop structures. PCR of the DNA fragments with such termini is therefore suppressed in homoduplexes when primers targeted at the 5′-ends of the ligated adapters are used. In contrast, heteroduplex molecules have different termini unable to form stem-loop structures, and can be further efficiently PCR amplified in this system. Nested PCR with primers A2 and B2 increases the specificity of the amplification. This procedure thus ensures exclusive amplification of only the heteroduplex DNA. The control experiments had all of the stages mentioned above, except the (5) step, i.e. treatment of hybridized DNA with nucleases.

Two mismatched DNA sensitive nucleases were used: Surveyor nuclease that recognizes and cleaves mispaired DNA structures within DNA duplexes and mung bean nuclease, which degrades single-stranded DNA and, therefore, is able to attack loop structures in chimeric hybrids. Mammalian DNAs were chosen for model experiments because they stand among most complex eukaryotic genomes, thus producing very complex hybridization mixtures, far more complex than those of cDNAs. Thus, by solving the challenge of unwanted chimera formation for complex mammalian genome libraries, one may be assured that this obstacle will be surmounted for lower complexity libraries too (such as those of cDNAs or of less complex genomes).

The resulting DNA libraries were cloned into *Escherichia coli*, and random transformants from each library were sequenced. The authors applied the following criteria for the chimera detection: such sequences did not match genomic databases entirely, but their separate 5′- and 3′-terminal fragments did match the databases. Figure 4 depicts the results of the analysis of six DNA libraries. It is clear that the addition of $C_o$t A fraction and the hybridization temperature increase from 65°C to 85°C has essentially no effect on the number of chimerical clones, in contrast to the addition of mismatch sensitive nucleases. Both mung bean and Surveyor nucleases display the strong effect on the chimera formation, greatly reducing their number from 44–60% clones to 0–4%. Many sequenced inserts contained genomic REs, which is not surprising, as they constitute a major part of mammalian DNA (Lander et al. 2001). Such RE sequences even if they correspond to correct genomic loci may match different positions on the genomic DNA, thus making their exact mapping problematic. Therefore it is desirable to minimize the portion of such kind of sequences in the libraries. Interestingly, the proportion of REs containing inserts differed considerably among the libraries: $C_o$t A-libraries contained high number of REs independently on the addition of the nuclease (87–93% of the sequenced clones), $C_o$t A + /N-libraries – slightly smaller proportion of REs (76–78%), and finally $C_o$t A + /N + library (H6, mung bean nuclease added) had only 44% of RE-containing inserts. These data show that the best results in the library construction can be achieved with both (1) addition of RE-containing competitor

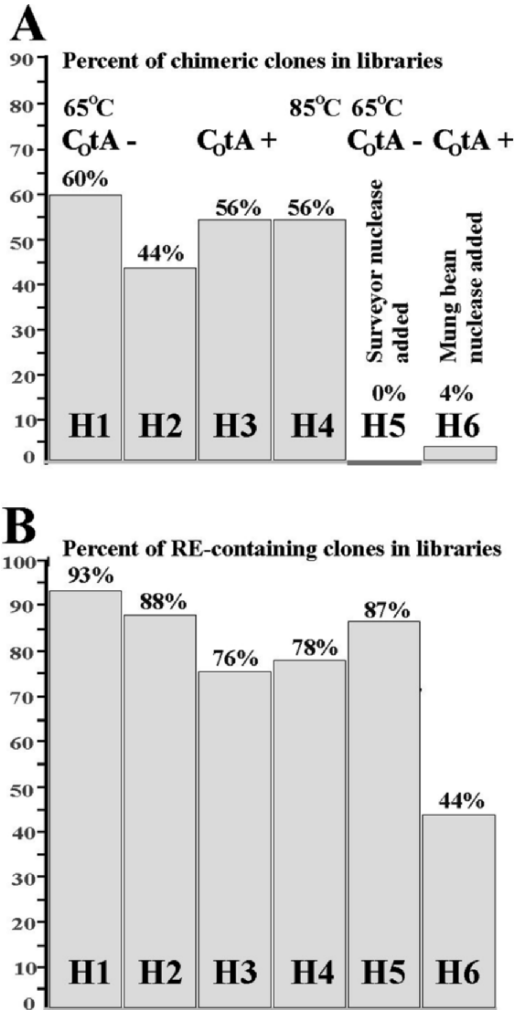


*Figure 4.* Comparison of six DNA libraries, created under different hybridization conditions with or without the use of MDR. H1, human–human DNA hybridization at 65°C (T65), without competitor $C_0$tA DNA ($C_0$tA-), no mismatch sensitive nucleases added (N-); H2, human–chimpanzee DNA, T65, $C_0$tA-, N-; H3, human–human DNA, T65, $C_0$tA added ($C_0$tA+), N-; H4, human–human DNA, T85, $C_0$tA + , N-; H5, human–chimpanzee DNA, T65, $C_0$t A-, Surveyor nuclease added; H6, human–human DNA, T65, $C_0$tA + , mung bean nuclease added. (A) Column height reflects the proportion of chimeric clones in analyzed libraries. The number of chimeric sequences is dramatically decreased in libraries, treated with mismatch sensitive nucleases. (B) Column height reflects the proportion of clone inserts, containing repetitive element (RE) sequences. It can be seen that the addition of $C_0$tA competitor DNA alone slightly decreases the number of RE-containing clones, but the combination of both $C_0$tA addition and nuclease digestion yields the best result in library construction.

DNA into hybridization mixture and (2) treatment of hybridized DNA with mismatch sensitive nucleases.

Also, MDR was applied to interspecies DNA hybridizations (analogous to coincidence cloning approach, Chapter 8). To this end in two hybridization experiments the authors hybridized human and chimpanzee DNA. Human and chimpanzee genomes are closely related, displaying ~98% sequence identity (Lander et al. 2001). The results suggest that MDR reduces the number of chimerical sequences from 44% even to the absence of detected chimeras. All sequenced inserts from Surveyor nuclease-treated library did contain sequences, highly conservative between the two genomes (average identity of 98.3%). Some inserts contained regions, evolutionary conserved among the sequenced mammalian genomes – these of human, chimpanzee, mouse, and rat. This observation suggests that MDR could also be applied for the recovery of evolutionary conserved sequences between different genomes. To investigate this, the authors performed another interspecies hybridization, between human and new world monkey *Callithrix pygmaea* genomes, followed by the subsequent digestion with Surveyor nuclease. The pygmy marmoset *C. pygmaea* genome is more divergent from human DNA than that of chimpanzee (human and new world monkey ancestor lineages separated roughly 45 million years ago (Sverdlov 1998; Sverdlov 2000), thus showing about 20% DNA sequence divergence (assuming the average nucleotide substitution rate in primate genomes to be $2.2 \times 10^{-9}$ bases per million years (Consortium 2002)). Seventy-one percent of cloned inserts represented moderately (~14%) divergent genomic repeats, which are believed to be present in both human and marmoset genomes, and the remaining 29% were unique sequences, most of which were conserved among human, chimpanzee, mouse, and rat genomes. To confirm the high conservation value of these sequences among human and marmoset, the corresponding loci from *C. pygmaea* genome were PCR-amplified and sequenced. Indeed, all sequenced marmoset loci displayed significant DNA conservation and similarity to the corresponding human loci with the average sequence identity of 95%, thus showing about fourfold slower mutation rate for these loci than neutral base substitution rate. Interesting enough, Surveyor nuclease was more efficient for DNA library refinement than mung bean nuclease, probably because of the ability to recognize and to cleave the DNA at the one-nucleotide mismatches, in contrast to mung bean nuclease, which is specific to more extended single-strand DNA loop regions (Figure 5).

The results presented above strongly suggest that MDR technique may provide a useful tool for the refinement of various DNA libraries obtained with the use of DNA reassociation, including subtractive and normalized genomic and cDNA libraries. The technique may also considerably improve genome wide recovery of evolutionary conserved sequences. The experimental techniques for identification of evolutionary conserved regions are required for the comparison of sequenced and/or unsequenced genomes, thus making MDR a universal method. Whenever the case the technique application will hopefully diminish

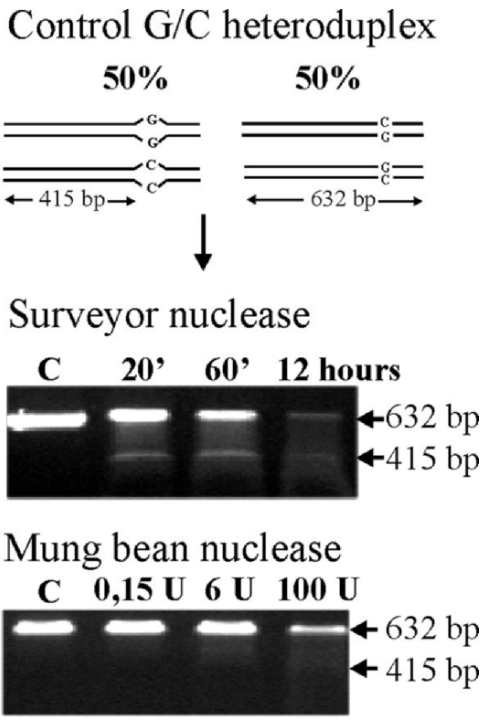


*Figure 5.* Comparison of Surveyor and mung bean nuclease digestion of control DNA. Control DNA, supplied with Surveyor mutation detection kit, contains 50% of perfectly matched 632 bp long DNA duplexes, and 50% of duplexes having single mismatched base pair (25% of G-G and 25% of C-C mispaired nucleotides). The DNA cleavage at the exact mismatch site gives 415 bp long products. Surveyor nuclease digestion: the 415 bp long fragment band is clearly seen in all experiments, thus showing specific mismatched DNA cleavage. Mung bean nuclease digestion: no detectable band at 415 bp could be detected, slight smearing appeared when large amounts of nuclease were used, thus suggesting the lack of efficient specific cleavage at single mispaired nucleotides by mung bean nuclease.

the confusions caused by cross-hybridization of closely related but however different paralogous sequences. MDR protocol is given in the Section 4.3.

## 4. PROTOCOLS

### 4.1 Targeted Genomic Difference Analysis

TGDA is schematically presented in Figure 6. The method includes two steps: whole-genome selective amplification of the flanks adjacent to interspersed REs (in our case human L1 retrotransposons and HERV-K endogenous retroviruses) in both genomic DNAs under comparison (Figure 6A), and then SH of the selected amplicons (Figure 6B). The first step is based on the PCR-suppression effect (Lukyanov et al. 1995; Lukyanov et al. 1997; see Chapter 2).

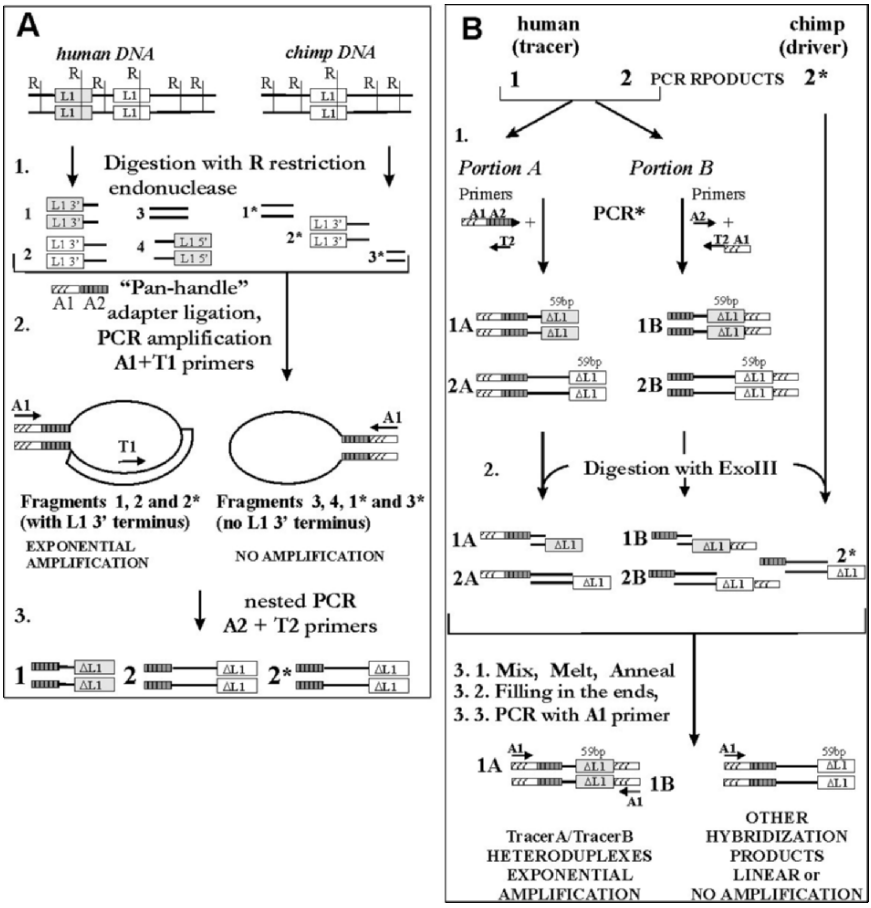


*Figure 6.* TGDA schematic representation. (A) Stages of selective amplification of L1-flanking genomic regions. Gray and open boxes denote human-specific and other L1s, respectively. R, positions of restriction sites. Hatched boxes designate suppression adapters used. Different types of restriction fragments are enumerated (with asterisks for chimp DNA). (B) Stages of subtractive hybridization (SH). Stage 1: PCR* carried out in accordance with the step-out PCR technique using A1 + A1A2 + T2, or A1 + A1T2 + A2 sets of primers.

Briefly, it includes digestion of the genomic DNAs with a frequent-cutter restriction enzyme, R (Figure 6A, stage 1, using *Alu* I), and ligation of the resulting restriction fragments to a stem-loop structure forming oligonucleotide adapter (Figure 6A, stage 2, see Chapter 2, Section 6). As a result, all DNA restriction fragments had inverted repeats at their termini. Therefore, the single-stranded fragments contained self-complementary termini capable of forming strong intramolecular stem-loop structures (panhandle-like structures, Figure 6A). PCR of the DNA fragments with such termini is suppressed when only one

primer targeted at the 5′-ends of the ligated adapter (Figure 6A, stage 2) is used. In contrast, a pair of A1- and T1-primers targeted at the single-stranded part of the stem-loop structure (Figure 6A, stage 2) can initiate DNA synthesis by DNA polymerase. The amplified DNA in this case will have different termini unable to form stem-loop structures, and can be further efficiently amplified with A1 + T1 primers. Nested PCR with A2 and T2 primers increases the specificity of the amplification. This procedure ensures efficient, nearly exclusive amplification of only the fragments that contain the target sequence, HERV-K LTR or L1 in this example. We used it to prepare amplicons containing the DNA flanks of the LTRs 5′-parts and L1 3′-termini both for human and chimpanzee DNAs. T1 and T2 primers directed the DNA synthesis to the outside of the target repeat.

The subtraction in shown schematically on Figure 6B. This allows direct isolation of sequences present solely in one of two related genomes (we will define them as targets), without any preliminary knowledge of the genome sequences. SH is based on reassociation of both genomic DNAs under comparison. After digestion and mixing at a large excess of one DNA (defined as driver DNA) over the other one (defined as tracer DNA), the resulting short fragments were denatured and cooled to reanneal. During the reassociation most of the tracer DNA hybridizes to the excess driver DNA, except for the targets, which form homoduplexes. The self-reassociated tracer is enriched in these reassociated target fragments, compared with the original tracer genome. In this example we tried to identify human specific targets, and therefore human and chimpanzee (our closest relative) DNAs were used as tracer and driver, respectively. We prepared two separate portions of the tracer DNA (Figure 6B, left, stage 1) by reamplification of the human amplicons obtained at the previous stage (Figure 6A, stage 3). We used the step-out PCR (Matz et al. 1999) with primers A1A2, A1, and T2, or A1T2, A1, and A2 for amplifications of portions A and B, respectively. The resulting portion A DNA fragments contained A1A2 sequence at one end and T2 sequence at the other, whereas the corresponding terminal sequences of the portion B fragments were A2 and A1T2.

The following 5′-protruding single-stranded termini formation (Figure 6B, stage 2) is a critical stage for the whole procedure: it prevents cross-hybridization of the repetitive parts common for all the amplicons, and ensures subsequent specific amplification of the double-stranded tracer A/B heteroduplexes formed during the subtraction process. To form 5′-protruding single-stranded termini, we digested A and B tracers with the nuclease *Exo* III until ~60 nt were removed from each 3′-end. The driver DNA was digested similarly to remove ~40 terminal nt.

Tracers A and B and a 100-fold excess of the driver (Figure 6B, stage 3) were mixed, melted, and allowed to reanneal. The resulting mixture contained single-stranded fragments of the both tracers and the driver, double-stranded hybrids formed between the tracers and the driver, homoduplexes formed as a result of self-reassociation of tracers A and B, and heteroduplexes formed by cross-reassociation of the tracers A and B complementary strands (tracer A/B fraction). Once the

protruding ends of the latter heteroduplexes have been filled-in with DNA polymerase, the heteroduplexes acquired targets for primer A1 at both termini and were the only fragments that could be exponentially amplified with this primer.

The PCR products were cloned in *E. coli*, ~500 random transformants were arrayed and further analyzed for each library.

#### *4.1.1 DNA Samples and oligonucleotides*

Genomic DNA was extracted from 20 samples of individual human placentas, human blood samples, or blood samples of chimpanzees using a genomic DNA purification kit (Promega). Suppression adapters are listed in Chapter 2, Section 6.1.

#### *4.1.2 Preparation of tracer and driver DNAs*

Digestion of human and chimpanzee DNAs, adapter ligation, and PCR amplification of LTR-flanking regions was all done as described (Lavrentieva et al. 1999). We amplified 1 ng aliquots of human amplicons according to step-out PCR procedure (Matz et al. 1999) with set A (0.01 μM A1A2, 0.2 μM A2, 0.2 μM T2) or set B (0.01 μM A2T2, 0.2 μM A2, 0.2 μM A1) of primers using 15 cycles at 95°C for 15 s, 57°C for 10 s,72°C for 90 s.

We digested 150 ng each of the resulting tracer A and B samples, and 3000 ng of the initial chimp amplicon (driver) with *Exo* III nuclease separately at 16°C using the following conditions: tracer A, 20 units of the *Exo* III, 11 min (40 terminal nt to be removed); tracer B, 20 units, 14 min (60 nt to be removed); driver, 400 units,11 min (40 residues to be removed). We mixed 15 ng each of the digested tracer A and B samples with 1500 ng of digested driver. DNA samples were purified by phenol/chloroform extraction, precipitated with ethanol, and dissolved in 5 μl of sterile water.

#### *4.1.3 Subtractive hybridization*

We mixed both tracer A/driver and tracer B/driver samples, transferred them into a hybridization buffer (0.5 M NaCl, 50 mM Hepes, pH 8.3,0.2 mM EDTA), denatured at 95°C for 10 min, and hybridized at 65°C for 14 hs. The final 15 μl mixture was diluted with 185 μl dilution buffer (50 mM NaCl, 5 mM Hepes, pH 8.3, 0.2 mM EDTA). We PCR-amplified 1 μl of this diluted mixture with 0.4 μM A1 primer. The PCR conditions were as follows: (1) 72°C for 6 min to fill in the ends of DNA duplexes; (2) 95°C for 15 s, 65°C for 10 s, 72°C for 90 s, 15 cycles. PCR products obtained were further cloned in *E.coli* using a TA-cloning system (Promega), and ~500 individual clones were sequenced for each library.

### 4.2 Using Competitor DNA to Decrease the Background of Genomic Repeats

#### *4.2.1 Starting material*

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

*Oligonucleotides.* We used the standard suppression adapters A1A2 (5′-*GTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT*-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′-*AGCGTGGTCGCGGCCGAGGT*-3′, and B2, 5′-*TCGAGCGGCCGCCCGGGCAGGT*-3′. Oligonucleotides were synthesized using an ASM-102U DNA synthesizer (Biosan, Novosibirsk, Russia).

### *4.2.2 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).

### *4.2.3 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). $C_0$tA fraction competitor DNA (Gibco BRL, USA) was added in 100x weight excess to the hybridization mixture. In control experiments, no $C_0$tA DNA was added.

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

### *4.2.4 PCR amplification of hybridization products and library construction*

*Nested PCR amplification.* DNA samples were dissolved in 100 μl of water and 1 μl 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.

## 4.3 Mispaired DNA Rejection

### *4.3.1 Starting material*

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

*Oligonucleotides.* We used the standard suppression adapters A1A2 (5′-*GTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT*-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′-*AGCGTGGTCGCGGCCGAGGT*-3′, and B2, 5′-*TCGAGCGGCCGCCCGGGCAGGT*-3′. Oligonu-cleotides were synthesized using an ASM-102U DNA synthesizer (Biosan, Novosibirsk, Russia).

### *4.3.2 DNA preparation for hybridization*

*Digestion of genomic DNA.* About 1 μg 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).

### *4.3.3 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, $C_0$tA 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/1 µg of hybridized DNA) to fill in the ends of DNA duplexes at 72°C for 20 min.

### *4.3.4 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 1× 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.

### *4.3.5 PCR amplification of hybridization products and library construction*

*Nested PCR amplification.* DNA samples were dissolved in 100 µl of water and 1 µl 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 NCBI (http://www.ncbi.nlm.nih.gov/BLAST) (Altschul et al. 1990). For multiple alignments the ClustalW program (Thompson et al. 1994) was used.

*4.3.6 PCR amplification of evolutionary conserved sequences*

A DNA sample of 40 ng of old world monkey *C. pigmaea* blood 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.

## REFERENCES

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

Babon JJ, McKenzie M, Cotton RG (2003) The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. Mol Biotechnol 23:73–81

Barr FG, Emanuel BS (1990) Application of a subtraction hybridization technique involving photoactivatable biotin and organic extraction to solution hybridization analysis of genomic DNA. Anal Biochem 186:369–373

Broude NE (2002) Stem-loop oligonucleotides: a robust tool for molecular biology and biotechnology. Trends Biotechnol 20:249–256

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, 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 (small star, filled). Genomics 81:149–156

Carulli JP, Artinger M, Swain PM, Root CD, Chee L, Tulig C, Guerin J, Osborne M, Stein G, Lian J, Lomedico PT (1998) High throughput analysis of differential gene expression. J Cell Biochem 30–31:286–296

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

Consortium MGS (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562

Cotton RG (1999) Mutation detection by chemical cleavage. Genet Anal 14:165–168

Cotton RG, Rodrigues NR, Campbell RD (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. Proc Natl Acad Sci USA 85:4397–4401

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

Ermolaeva OD, Sverdlov ED (1996) Subtractive hybridization, a technique for extraction of DNA sequences distinguishing two closely related genomes: critical analysis. Genet Anal 13:49–58

Ermolaeva OD, Wagner MC (1995) SUBTRACT: a computer program for modeling the process of subtractive hydridization. Comput Appl Biosci 11:457–462

Ermolaeva OD, Lukyanov SA, Sverdlov ED (1996) The mathematical model of subtractive hybridization and its practical application. Proc Int Conf Intell Syst Mol Biol 4:52–58

Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov KA, Vagner LL, Ermolaeva OD, Lukyanov SA, Sverdlov ED (1996) Equalizing cDNA subtraction based on

selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. Anal Biochem 240:90–97

Hames BD, Higgins SJ (eds) (1985) Nucleic acid hybridization: a practical approach. IRL Press, Oxford, Washington DC

Huang J, Kirk B, Favis R, Soussi T, Paty P, Cao W, Barany F (2002) An endonuclease/ligase based mutation scanning method especially suited for analysis of neoplastic tissue. Oncogene 21:1909–1921

Kohne DE, Levison SA, Byers MJ (1977) Room temperature method for increasing the rate of DNA reassociation by many thousandfold: the phenol emulsion reassociation technique. Biochemistry 16:5329–5341

Kulinski J, Besack D, Oleykowski CA, Godwin AK, Yeung AT (2000) CEL I enzymatic mutation detection assay. Biotechniques 29:44–46, 48

Kunkel LM, Monaco AP, Middlesworth W, Ochs HD, Latt SA (1985) Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. Proc Natl Acad Sci USA 82:4778–4782

Lamar EE, Palmer E (1984) Y-encoded, species-specific DNA in mice: evidence that the Y chromosome exists in two polymorphic forms in inbred strains. Cell 37:171–177

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.

Lisitsyn N, Wigler M (1993) Cloning the differences between two complex genomes. Science 259:946–951

Lisitsyn N, Wigler M (1995) Representational difference analysis in detection of genetic lesions in cancer. Methods Enzymol 254:291–304

Lisitsyn NA, Leach FS, Vogelstein B, Wigler MH (1994a) Detection of genetic loss in tumors by representational difference analysis. Cold Spring Harb Symp Quant Biol 59:585–587

Lisitsyn NA, Segre JA, Kusumi K, Lisitsyn NM, Nadeau JH, Frankel WN, Wigler MH, Lander ES (1994b) Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. Nat Genet 6:57–63

Lukyanov K, Diatchenko L, Chenchik A, Nanisetti A, Siebert P, Usman N, Matz M, Lukyanov S (1997) Construction of cDNA libraries from small amounts of total RNA using the suppression PCR effect. Biochem Biophys Res Commun 230:285–288

Lukyanov KA, Launer GA, Tarabykin VS, Zaraisky AG, Lukyanov SA (1995) Inverted terminal repeats permit the average length of amplified DNA fragments to be regulated during preparation of cDNA libraries by polymerase chain reaction. Anal Biochem 229:198–202

Mamedov IZ, Arzumanyan ES, Amosova AL, Lebedev YB, Sverdlov ED (2005) Whole-genome experimental identification of insertion/deletion polymorphisms of interspersed repeats by a new general approach. Nucleic Acids Res 33:e16

Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchik A (1999) Amplification of cDNA ends based on template-switching effect and step- out PCR. Nucleic Acids Res 27:1558–1560

Mikhailov VS, Rohrmann GF (2002) Binding of the baculovirus very late expression factor 1 (VLF-1) to different DNA structures. BMC Mol Biol 3:14

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

Mouritzen P, Nielsen AT, Pfundheller HM, Choleva Y, Kongsbak L, Moller S (2003) Single nucleotide polymorphism genotyping using locked nucleic acid (LNA). Expert Rev Mol Diagn 3:27–38

Nagayama K, Enomoto N, Miyasaka Y, Kurosaki M, Chen CH, Sakamoto N, Nakagawa M, Sato C, Tazawa J, Ikeda T, Izumi N, Watanabe M (2001) Overexpression of interferon gamma-inducible protein 10 in the liver of patients with type I autoimmune hepatitis identified by suppression subtractive hybridization. Am J Gastroenterol 96:2211–2217

Qiu P, Shandilya H, D'Alessio JM, O'Connor K, Durocher J, Gerard GF (2004) Mutation detection using Surveyor nuclease. Biotechniques 36:702–707

Roberts E, Deeble VJ, Woods CG, Taylor GR (1997) Potassium permanganate and tetraethylammonium chloride are a safe and effective substitute for osmium tetroxide in solid-phase fluorescent chemical cleavage of mismatch. Nucleic Acids Res 25:3377–3378

Roberts RW, Crothers DM (1991) Specificity and stringency in DNA triplex formation. Proc Natl Acad Sci USA 88:9397–9401

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

Sasaki H, Nomura S, Akiyama N, Takahashi A, Sugimura T, Oishi M, Terada M (1994) Highly efficient method for obtaining a subtracted genomic DNA library by the modified in-gel competitive reassociation method. Cancer Res 54:5821–5823

Sverdlov ED (1998) Perpetually mobile footprints of ancient infections in human genome. FEBS Lett 428:1–6

Sverdlov ED (2000) Retroviruses and primate evolution. Bioessays 22:161–171

Sverdlov ED, Ermolaeva OD (1994) [Kinetic analysis for subtractive hybridization of transcripts]. Bioorg Khim 20:506–514

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

Till BJ, Burtner C, Comai L, Henikoff S (2004) Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res 32:2632–2641

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