CHAPTER 1

NUCLEIC ACIDS HYBRIDIZATION: POTENTIALS AND LIMITATIONS

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: Several nucleic acids hybridization-based approaches, such as microarray, competitive genomic, and Southern or Northern blot hybridization, have become popular tools for specialists in biochemistry and in biomedicine, and are now in routine use. However, the potential of in-solution nucleic acids hybridization-based experimental techniques seems to be underestimated now. Examples are subtractive hybridization (SH), which allows one to efficiently find differences in genomic DNAs or in cDNA samples; coincidence cloning (CC), which, on the contrary, makes it possible to identify sequences that are present in all the samples under comparison; cDNA normalization, which is used for the smoothing of rare and frequent transcript content in cDNA libraries; and TILLING approach, which has demonstrated its great potential for the reverse genetics studies. Finally, several techniques are aimed at the large-scale recovery of DNA polymorphisms, including single nucleotide polymorphisms (SNPs). This book will focus on the above-mentioned and other recent developments in the area of nucleic acids hybridization, including attempts to improve its specificity. In this introductory chapter, I have tried to briefly characterize the current state of the art in in-solution nucleic acids hybridization techniques, and to define their major principles and applications. The advantages and shortcomings of these techniques will be discussed here.

Keywords: Nucleic acids hybridization, cDNA library construction, EST, differentially expressed genes, differential transcripts, differential sequence, microarray, competitive genomic hybridization, subtractive hybridization, coincidence cloning, rare transcript, frequent transcript, normalization, cDNA normalization, cDNA subtraction, genomic subtraction, polymorphism recovery, mutation, single nucleotide polymorphism, SNP, hybridization kinetics, hybridization rate, tracer, tester, driver, genome size, genome complexity, representational differences analysis (RDA), subcloning, restriction fragment length polymorphisms recovery, suppression subtractive hybridization, SSH, genomic polymorphism, Sanger sequencing, mispaired nucleotides, mutant strand, wild-type strand, mutant allele, wild-type allele, rapid amplification of cDNA ends, RACE, differentially methylated.

A. Buzdin and S. Lukyanov (eds.), Nucleic Acids Hybridization, 1–28.

Abbreviations: BAC, bacterial artificial chromosome; CC, coincidence cloning; cDNA, complementary DNA; CHS, covalently hybridized subtraction; dNTP, deoxyribonucleotidetriphosphate; EST, expressed sequence tag; GREM, genomic repeat expression monitor; mRNA, messenger RNA; MOS, mirror orientation selection; NGSCC, nonmethylated genomic sites coincidence cloning; PEER, primer extension enrichment reaction; PCR, polymerase chain reaction; PERT, phenol emulsion reassociation technique; RACE, rapid amplification of cDNA ends; RDA, representative differences analysis; RFLP, restriction fragment length polymorphism; RNAi, interfering RNA; RT, reverse transcription; SAGE, serial analysis of gene expression; SH, subtractive hybridization; SNP, single nucleotide polymorphism; SSH, suppression subtractive hybridization; YAC, yeast artificial chromosome.

TABLE OF CONTENTS

1. INTRODUCTION

Watson–Crick hybridization of complementary sequences in nucleic acids is one of the most important fundamental processes necessary for molecular recognition *in vivo* (Watson and Crick 1953), as well as nucleic acid identification and isolation *in vitro* (Southern 1992). The use of experimental techniques based on DNA hybridization in solution is advantageous for many applications, starting from representative complementary DNA (cDNA) library construction for expressed sequence tag (EST) sequencing to the identification of evolutionary conserved sequences, differentially expressed genes, or genomic deletions. Unique characteristics of many such techniques make them powerful competitors

for well-known approaches that are appreciated worldwide like microarray hybridization and competitive genomic hybridization. Examples are subtractive hybridization (SH), which allows one to efficiently find differences in genomic DNAs or in cDNA samples; coincidence cloning (CC), which, on the contrary, makes it possible to identify sequences which are common for all samples under comparison; and cDNA normalization, which is used for the smoothing of rare and frequent transcript content in cDNA libraries, thus being extremely useful for representative EST library construction. Moreover, several techniques deal with the large-scale DNA polymorphism recovery, including identification of single nucleotide polymorphisms (SNPs).

Nucleic acid hybridization in solution has few general advantages over hybridization with solid carrier-immobilized nucleic acids: faster hybridization kinetics, better discrimination of proper hybrids, and their availability for further polymerase chain reaction (PCR) amplification and cloning. Among such in-solution hybridization methods, SH is undoubtedly the most popular technique. Analysis of differential gene expression requires application of global approaches that represent a leading tool in postgenomic studies and include transcriptome and proteome analysis as well as methods allowing population-wide sequence and functional polymorphism analysis. Central to these new technologies are DNA chips designed for quantitative and qualitative uses (Brown and Botstein 1999) (see Chapter 11). Although they are now very useful and widely distributed, many popular DNA microarray techniques share a number of shortcomings:

1. The analysis is limited by a number of cDNAs or synthetic oligonucleotides applied on the chip. This number is usually significantly lower than the total gene quantity of the organisms under study. It creates, therefore, the problem that many genes escape such an analysis.
2. General transcriptome-wide chip techniques in their actual state hardly distinguish between different gene splice forms.
3. The expression of genes transcribed at low levels cannot be detected by using standard microarray approaches.
4. cDNA-based chips do not differentiate between many gene family members and/or between many transcripts containing repetitive DNA.
5. Microarray chips lack many natural RNAi cDNAs or synthetic oligonucleotides and therefore cannot be used for comprehensive studies of gene expression regulation at the level of RNA interference by small interfering RNAs.

However, most of these concerns can be effectively addressed by using specific variants of microchip technology, thus making microarrays a truly universal technique (reviewed in Chapter 11). Probably, the most important disadvantage of closed systems such as microarrays is that they require preliminary genomic sequence information in order to identify differentially expressed transcripts.

Open systems have the flexibility of identifying uncatalogued sequences: alternatively, differences in gene expression between two samples can be compared directly by methods such as differential display (Liang and Pardee 1992), differential cloning techniques (Sagerstrom et al. 1997), and combinations of these (Pardinas et al. 1998; Yang et al. 1999). These approaches have been successfully used to identify genes differentially expressed in two populations that exhibit large changes in expression levels, or genes that are expressed at high concentrations in terms of number of copies per cell.

However, these techniques have a low efficiency of identifying rare genes that are differentially expressed (Martin and Pardee 2000). This problem is exacerbated when the change in expression level of rare transcripts is small. Since genes expressed at low levels also play a role in establishing differentiated phenotypes, their identification is essential for a complete mechanistic understanding of cellular changes. The major advantage of SH lies in the ability to identify differentially expressed genes, irrespective of the level of expression, in the absence of sequence information. In addition to preparation of differential cDNA libraries, SH is also extremely useful for identification of genomic DNA differences (Diatchenko et al. 1996, 1999; Ermolaeva et al. 1996; Akopyants et al. 1998; Bogush et al. 1999).

2. CLONING THE DIFFERENCES: SUBTRACTIVE HYBRIDIZATION

2.1 Birth of a Method

SH was first used as early as 1966 by Bautz and Reilly (1966) to purify phage T4 mRNA. Recently, a number of groups have employed variations of the technique, both to clone cDNAs derived from mRNAs that undergo up- or down-regulation (cDNA subtraction), and to identify genomic deletions (genomic subtraction). This approach became well known since 1984 when Palmer and Lamar applied SH for the construction of mouse recombinant DNA libraries, enriched in Y chromosome sequences (Lamar and Palmer 1984). Since that date SH has significantly evolved, thus becoming one of the most important and effective tools in molecular biology. This truly universal approach is being used for a diverse set of applications like cloning and characterization of new genes, recovery of tissue-specific, malignancy-specific, or organism-specific transcripts, identification of genes differentially expressed at different stages of embryo development, cancer progression, regeneration, for the recovery of genes up- or down-regulated in response to external or internal stimuli, etc. (Cekan 2004; Ying 2004). The method is useful for the genome-wide comparison of bacterial DNAs (Bogush et al. 1999), for isolating species-specific loci (Buzdin et al. 2002; Buzdin et al. 2003), and polymorphic markers in both eukaryotic and prokaryotic genomes (Bogush et al. 1999; Nadezhdin et al. 2001). In addition, SH was shown effective for DNA subcloning from yeast artificial chromosomes (YACs) into smaller vectors (Zeschnigk et al. 1999), for mapping of genomic

rearrangements associated with cancer or chromosome abnormalities, even for filling in extended gaps in large-scale sequencing projects (Frohme et al. 2001).

However, being such a powerful instrument for molecular biology and biomedicine, SH usefulness is still underestimated. In this section, I have tried to elucidate all major techniques dealing with the subtraction of nucleic acids. As stated above, SH became sound in 1984 when Palmer and Lamar proposed a simple idea of a separation of hybrid molecules: double-stranded homohybrids of the "tracer" or "tester" DNA (a sample containing differential sequences to be identified), from heterohybrids tracer–driver and homohybrids driver–driver ("driver" is a sample containing reference nucleic acid sequences). SH is aimed at the isolation of a fraction of tracer-specific sequences absent from driver. The idea was that tracer and driver DNA would have different sequences on their termini (Lamar and Palmer 1984). The authors wanted to create a mouse recombinant DNA library enriched in Y chromosome sequences (Figure 1). Female mouse DNA (driver) was fragmented by sonication, whereas male DNA (tracer) was digested with *MboI* restriction endonuclease. Tracer DNA was mixed with 100-fold weight excess of a driver, denatured, and allowed to reanneal. Only reassociated tracer–tracer homoduplexes had sticky ends at both termini and could be ligated into the plasmid vector pBR322 digested by *BamHI* restriction endonuclease.

Soon afterwards, this principle was successfully employed for cloning of human DNA fragments absent in patients with Duchenne muscular dystrophy (Kunkel et al. 1985). Simultaneously, other research teams started using similar approaches for the recovery of messenger RNAs distinguishing analyzing samples (Chien et al. 1984; Kavathas et al. 1984), specific for certain cell type, tissue, or organism (Figure 2). Using a poly(A) + fraction of tracer RNA, cDNA first strands were synthesized, initial RNA was then degraded by the addition of NaOH, so that only cDNA first strands complementary to tracer mRNA remained in solution. The tracer was then mixed with taken in 100-fold or more weight excess of driver, which was a poly(A) + RNA fraction from another sample. In the resulting mixture, tracer fragments were either hybridized with the excess of driver complementary strands, or remained in a single-stranded form. The latter single-stranded fraction, which was enriched in tracer-specific sequences, was purified from driver and tracer–driver hybrids on a hydroxyapatite column (which provides column binding by double-stranded nucleic acids). Using purified single-stranded tracer, cDNA second strands are synthesized and the resulting double-stranded cDNA is ligated in either expression (if an additional round of subtraction is needed) or cloning vector (Chien et al. 1984; Kavathas et al. 1984). To increase hybridization rate, chemical accelerators such as phenol could be added to the hybridizing mixture (Travis and Sutcliffe 1988).

However, this approach did not become popular, probably due to three serious limitations: (1) great amounts of mRNA are needed; (2) the technique is extremely laborious; and (3) RNA degradation may cause severe problems at many stages. Recently, the first problem was solved by cloning tracer and driver

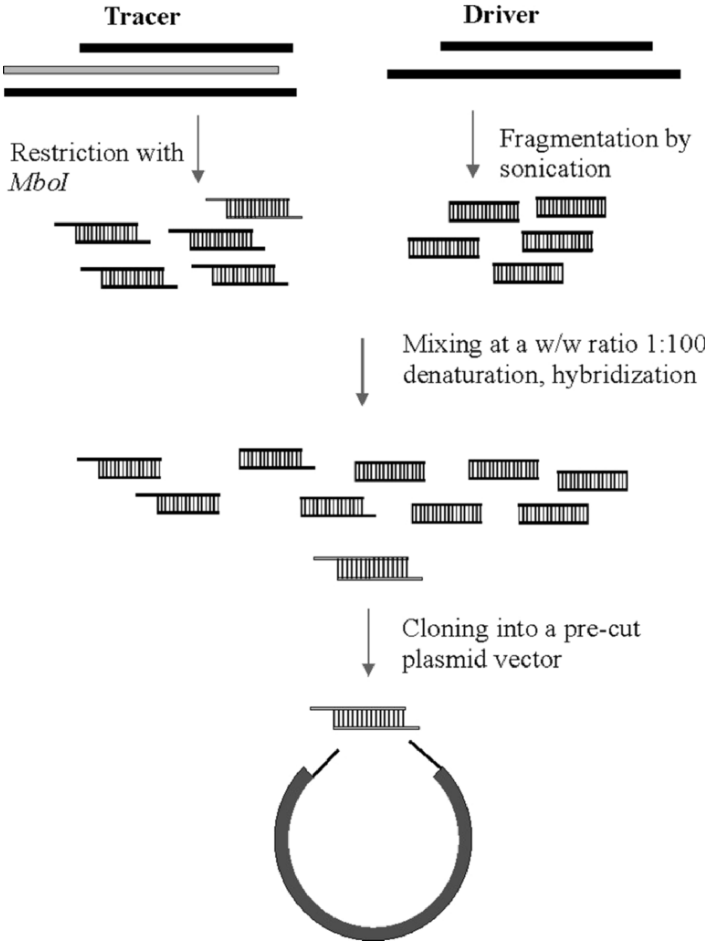

Figure 1. Genomic DNA subtraction scheme, proposed by Palmer and Lamar, based on tracer–tracer hybrids cloning using unique restriction sites.

cDNAs into special single- or double-stranded expression vectors. RNA was produced in *Escherichia coli*, thus making it possible to obtain large amounts for the hybridization (Palazzolo and Meyerowitz 1987; Kuze et al. 1989; Rubenstein et al. 1990). The third barrier was waived in part when driver mRNA was replaced by double-stranded cDNA (the use of single-stranded first strand cDNA was less cost-efficient). However, the latter improvement created a new problem: how to separate tracer–tracer duplexes from tracer–driver and driver–driver after hybridization?

In 1986, Welcher et al. (1986) created the first biotin–streptavidin subtraction system: biotinylated primers were used for driver cDNA synthesis, and the

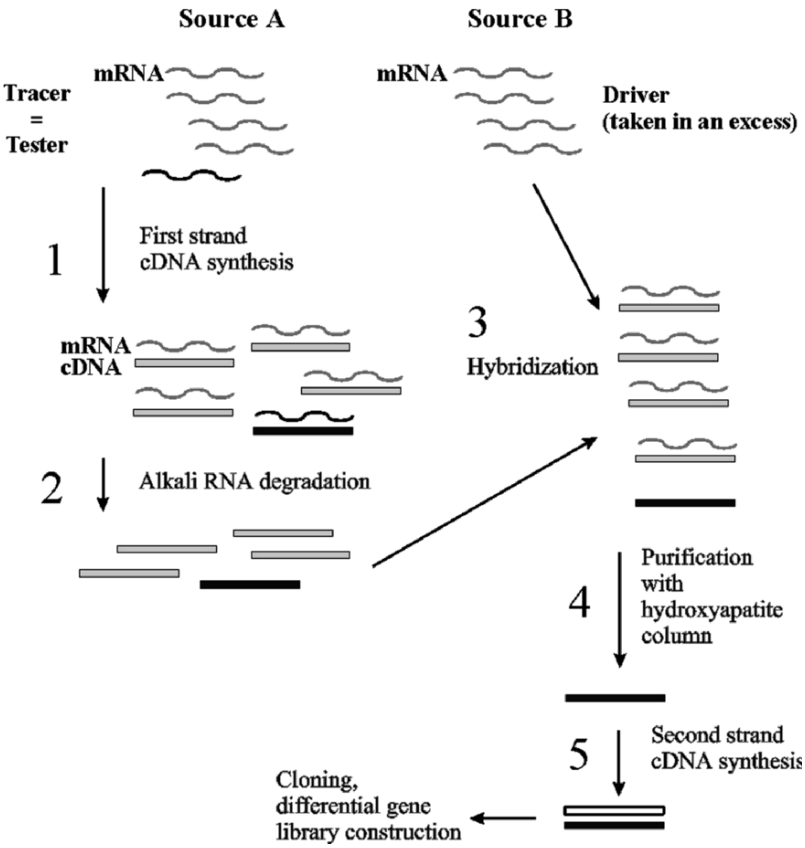

Figure 2. cDNA/mRNA subtraction scheme, utilizing alkali RNA degradation at stage 2.

subtraction products were incubated with streptavidin-coated copper granules. As a result, tracer–tracer duplexes remained in solution, whereas the granules bound all other hybridization products, except single-stranded tracer. This method of subtracted product separation became quite popular (some details being modified). In 1993, magnetic beads replaced copper granules as a solid-phase carrier for conjugated streptavidin in tracer–tracer hybrid purification (Lopez-Fernandez and del Mazo 1993; Sharma et al. 1993).

Overall, these early techniques for cDNA subtraction generally involved one or two rounds of hybridization and used (–) mRNA to drive hybridization to (+) cDNA tracer. However, the preparation of (–) mRNA in large amounts is not always a practical proposition; consequently, for less-abundant sequences, the concentration of driver is likely to be too low to drive hybridization to completion. The degree of enrichment is limited by the driver to tracer ratio, and a single round of hybridization will only enrich adequately those upregulated messages

that are rare in the (–) population, but highly abundant in the (+) population. Sequences that are only moderately abundant even after upregulation, or which are upregulated only to a limited extent, will still be obscured by a background of common sequences. Furthermore, the amount of cDNA remaining after hybridization can be tiny and the problem of cloning successfully such minute quantities of cDNA is not trivial.

2.2 PCR-assisted Subtractive Hybridization

Regardless of the method improvements mentioned above, SH is too laborious a process: in a period from 1984 to 1989, the use of SH was described in only 29 research papers (Figure 3). It was obviously PCR that made SH an easy, inexpensive, and widely used technique. Indeed, the researcher could now obtain substantial amounts of tracer and driver DNA in an easy inexpensive way with minute amounts of starting material. Moreover, the use of the PCR solved the problem of cloning of the subtraction products (Hla and Maciag 1990; Timblin

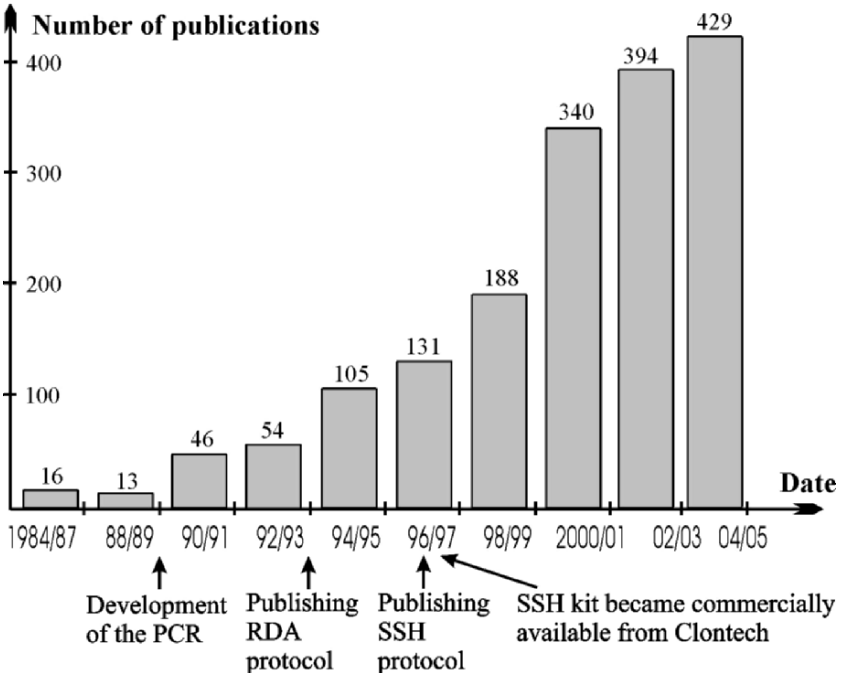

Figure 3. Citation dinamics of the subtractive hybridization approach in the literature. Publications in MedLine – indexed peer-reviewed journals were explored using PubMed server at National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/). The integral figure of 16 publications is shown here for the period 1984–1987.

et al. 1990; Hara et al. 1991). Consequently, since 1990 PCR is incorporated in essentially all subtraction protocols and the interest of the scientific community to SH has increased: the number of published SH applications per year is increasing by three- to fourfold (Figure 3).

PCR has also provided a solution to another important problem: before PCR one could not perform an effective subtraction for rare transcripts as their concentrations were small and the reassociation rate during SH was negligible; such transcripts, therefore, were escaping analysis. PCR made it possible to obtain unlimited amounts of DNA for hybridization, thus enabling detection of such infrequent RNAs (Hla and Maciag 1990; Timblin et al. 1990; Hara et al. 1991; Herfort and Garber 1991; Wang and Brown 1991). In addition to gene expression assays, since 1991 SH is being used for the recovery of differential sequences in bacterial genomes (Cook and Sequeira 1991), which is actually one of the most important applications of SH: identification of differences in genomic DNAs of virulent versus nonvirulent strains, between disease-causing bacterial species and their harmless relatives, is extremely important for both creating new diagnostic markers and targeting bacterial genes for new drugs development (Cook and Sequeira 1991; Cruz-Reyes and Ackers 1992).

However, genomic DNA subtraction efficiency depends greatly on the complexities of the DNAs under comparison, as learned from both experimental studies (Wieland et al. 1990; Clapp et al. 1993) and mathematical models simulating SH (Sverdlov and Ermolaeva 1994; Milner et al. 1995; Ermolaeva et al. 1996; Cho and Park 1998). As 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 enrichment of the target (Milner et al. 1995). Mammalian genomes are too complex to reach sufficiently high reassociation rate values, and only major differences (like presence or absence of Y chromosome or extended deletions) can be isolated in such a way. To enhance the kinetics of hybridization, increased hybridization times, higher driver concentrations, greater driver to 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; Laman et al. 2001) or solvent exclusion (Barr and Emanuel 1990), may be effective. Major considerations on the kinetical requirements for the effective subtraction of complex genomic mixtures and related formulas are given in more detail in Chapter 10.

In addition to poor reassociation rates, one more problem appears when complex eukaryotic DNAs are being compared: repetitive sequences (which form, for example, >40% of mammalian DNA: Lander et al. 2001; Venter et al. 2001) reassociate significantly faster than unique genomic sequences (Milner et al. 1995), and the resulting differential libraries are greatly enriched in repeats (Rubin et al. 1993). This creates a serious obstacle to attempts of direct higher eukaryotic genomic DNA comparison by means of SH. However, as reviewed by Sverdlov (1993) and Sverdlov and Ermolaeva (1994), an adequate reassociation rate theoretically could be obtained for complex genomic mixtures as well, if

single-stranded tracer and driver are used, as learned from the authors' mathematical model and preliminary experimental results (Sverdlov and Ermolaeva 1993; Ermolaeva and Sverdlov 1996). Unfortunately, to my knowledge, since 1996 when this idea was published, it was never used in practice for genome-wide comparisons of higher eukaryotic DNAs. So, until now the problem of poor SH applicability to complex nucleic acid mixtures remains unsolved. It should be mentioned here that all above considerations regarding SH reassociation kinetics are in fact more general, being true and actual for all the family of methods based on nucleic acids hybridization in solution as well.

2.3 First Worldwide Success: Representational Differences Analysis

Lisitsyn and Wigler 1993) reported a new SH-based approach termed "representative differences analysis" (RDA), which made the idea of SH quite popular (and which is actively in use until now; see Figure 3). RDA was applied first to the comparison of two mammalian genomes and cloning of the differential sequences. Kinetical limitations are waived here because of the random simplifications of the comparing genomic mixtures due to either the use of nonfrequent-cutter restriction enzymes or PCR selection effect.

In Figure 4, tracer and driver genomic DNAs are digested with the restriction endonuclease, and different oligonucleotide adapters are ligated to the fragmented DNAs. The resulting ligation mixtures are further subjects to 50–100 cycles of PCR amplification with adapter-specific primers. Due to well-known "PCR selection" effect, different fragments in complex genomic mixtures are PCR-amplified with very different efficiencies, primarily depending on the fragment size. Therefore, the major parts of initial DNA fragments are greatly underrepresented or lost in the final amplicons, which contain only 2–10% of the initial fragment diversity. The amplicons are further treated with mung bean nuclease to degrade 3′-terminal adapter sequences (which could cause a problem at the next step due to adapter–adapter cross-hybridization), tracer is then mixed with the excess of a driver, denatured, and allowed to hybridize. The hybridization mixture is then incubated with DNA polymerase to fill in the 3′-termini, followed by a PCR with the primer specific to the tracers' adapter. Only the tracer–tracer duplexes are amplified exponentially, whereas other hybridization products are amplified only linearly or not amplified at all. The exponentially amplified double-stranded duplexes can be easily cloned into a plasmid vector and sequenced. Alternatively, additional rounds of subtraction may be performed to increase library enrichment in differential sequences, presented solely in tracer. This technique is inexpensive, fast, and relatively easy; it permits working with small amounts of the starting material (genomic DNA or cDNA). It is not surprising, therefore, that RDA became quite popular for both transcription analyses and genomic marker recovery (Lisitsyn et al. 1994a, b, 1995; Ayyanathan et al. 1995; Drew and Brindley 1995; Lisitsyn and Wigler 1995; Schutte et al. 1995) (Figure 3).

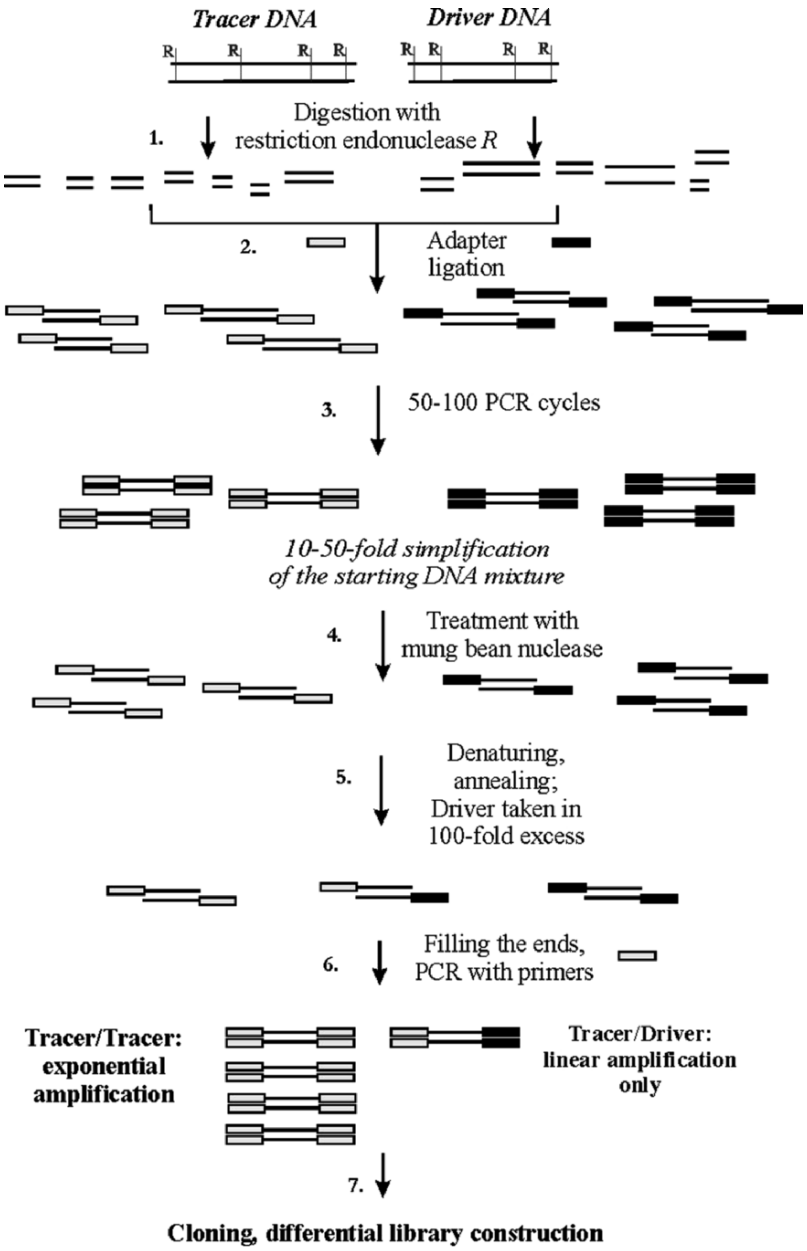

Figure 4. Representative differences analysis (RDA) application for identifying differential genomic DNA sequences. Initial genomic DNA mixture simplification is provided here by 50–100 cycles of PCR amplification at stage 3.

However, RDA has an obvious shortcoming: only small part of a transcriptome or genome is analyzed, whereas the majority (90–98%) escapes analysis. RDA still cannot provide genome-wide comparisons of the DNAs as its results are very fragmentary. Another limitation of the RDA use is the high background of the false-positive signals when cDNA libraries are compared: when the differences in transcription spectra are small, only a small number of tracer–tracer duplexes will be formed, and linear amplification of tracer–driver hybrids will create serious problems in the enriched subtracted library construction (Ayyanathan et al. 1995). This figure is even more pessimistic when such rare differential genes are poorly transcribed. Nevertheless, RDA is widely in use until now; some studies employing RDA are even entitled "genome/transcriptome-wide", which in fact is not the case, as we know now. To my opinion, RDA is advantageous for the recovery of differential marker sequences in DNAs under comparison.

With the increasing interest in SH and its applications, few reliable mathematical models of SH appeared and even two computational programs for the SH simulation *in silico* were published in 1995 (Ermolaeva and Wagner 1995; Milner et al. 1995). Below I briefly describe few successful and original experimental SH applications. The authors of the paper (Sallie 1995) proposed a new modification of the SH comprising the hybridization of subtracted products with filter-immobilized cosmid libraries of the candidate genomic loci. Such a use of SH, to the opinion of the authors, enabled unambiguous mapping of the genomic loci containing differentially transcribed genes. Authors of the next paper (Chen et al. 1995) applied RDA to the more efficient subcloning of the YAC fragments into smaller vectors, suitable for the insert sequencing. Genomes of YAC-containing yeasts (tracer) and those lacking YAC (driver) were fragmented and subtracted; after few cycles of subtraction, the products were cloned and sequenced. An overwhelming majority of the inserts contained sequences from the YAC. This application of the SH, although obviously interesting, did not become popular, probably due to the removal of YACs from large-scale sequencing strategies in favor of more stable bacterial artificial chromosomes (BACs).

A modification of the RDA was used to map cancer-specific deletions (Zeschnigk et al. 1999). Products of the subtraction of cancerous and normal genomes were hybridized with the ordered YAC library, thus making it possible to directly map differential sequences (no complete human genome sequence was available in databases at that time) and, therefore, to identify cancer-specific deletions. In more recent publication (Frohme et al. 2001), RDA was employed to fill the gaps during sequencing of the genome of *Xilella fastidiosa*. Fragments with already defined primary structure were subtracted from *Xilella* genome, differential sequences were hybridized with the complete *X. fastidiosa* genomic clone library and the positive clones (those containing differential sequences which were not sequenced before) were sequenced.

Another interesting approach was the SH-based technique for the restriction fragment length polymorphisms recovery, termed "RFLP subtraction" (Rosenberg et al. 1994). Genomic DNAs under comparison were digested by

restriction endonucleases and loaded separately on agarose gels. Following electrophoresis separation, specific zones (e.g. containing fragments from 100 to 500 base pairs (bp) in length) were excised for both tracer and driver, the DNAs were eluted from the agarose, and then used for the SH, resulting in amplicons enriched in differential fragments presenting in that zone in one sample DNA but absent from another one. The technique was proven to be very effective for the recovery of new RFLPs, genetic markers of a universal usefulness. This approach was significantly improved when the authors (Sasaki et al. 1994) performed subtraction directly in the gel. Both digested samples under comparison, one (driver) in a 100-fold weight excess over another one (tracer), were loaded on *the same* track of the gel and separated by electrophoresis. The gel was treated with NaOH to denature DNA, and then neutralized to hybridize denatured DNAs directly in the gel. Finally, the hybridized DNA was eluted from the agarose and was PCR-amplified to select for tracer–tracer duplexes. The authors obtained very high enrichment values, close to the theoretical maximum possible enrichment. This might be explained by very high local concentrations of the fragments of each type in gel which were significantly greater than those in solution. Unfortunately, this approach, which might be a perfect alternative to most popular SH techniques such as RDA and suppression subtractive hybridization (SSH), is very laborious.

2.4 Further Improvements: Suppression Subtractive Hybridization, Polymerase Chain Reaction Suppression Effect, and Normalization of cDNA Libraries

It may be seen from Figure 3 that the development of a new method called "suppression subtractive hybridization" (SSH) (Diatchenko et al. 1996; Jin et al. 1997) and further release of the corresponding kit from Clontech in 1996 caused a revolution in the use of the SH. The technique became more robust, more reproducible, and easier to perform. This clearly enhanced interest was accompanied by some "qualitative" changes in publications citing the SH: in 1984–1991 every third paper described any SH modification or improvement (at least from the point of view of the authors), in 1992–1995 every fifth, in 1998–1999 every tenth, and in 2000–2005 (when SSH was published and the kit became commercially available and appreciated) as rare as every 62nd paper. This might suggest that the popularization of SSH and RDA "killed" the creativity of the authors who were finally satisfied by the existing SH techniques. Probably, this means that these methods cannot be further improved, or that there is no need to improve them for most applications: SH became a routine reproducible procedure.

The principal advantage of the SSH over other SH techniques is the greatly reduced background of false-positive clones. In other methods, this background is caused by a linear amplification of the tracer–driver hybrids which is reduced to a minimum in SSH (described in detail in Chapter 3). Another advantage of the SSH is the custom normalization of cDNA libraries which can be performed in order to equalize the concentrations of different transcripts located in the libraries. Such equalization is needed to avoid discrimination of rare transcripts

during the subtraction. Normalization of cDNA libraries can be used for purposes other than SH as well, e.g. for the construction of representative EST libraries. This independent group of methods is described in Chapter 5.

Both advantages of the SSH are based on the "PCR suppression" effect (for detailed description, see Chapter 2). Approximately 40 nt long GC-rich linkers (termed "suppression adapters") are ligated to a fragmented double-stranded DNA (dsDNA). Further treatment with DNA polymerase builds the second strand of the adapters so that the initial DNA fragments are flanked by inverted GC-rich 40-nucleotide sequences. The methods' rationale is that the primers that are complementary to suppression adapters cannot efficiently anneal and initiate the PCR alone (Figure 5) due to significantly stronger intramolecular base pairing of the inverted repeats, which eliminates available

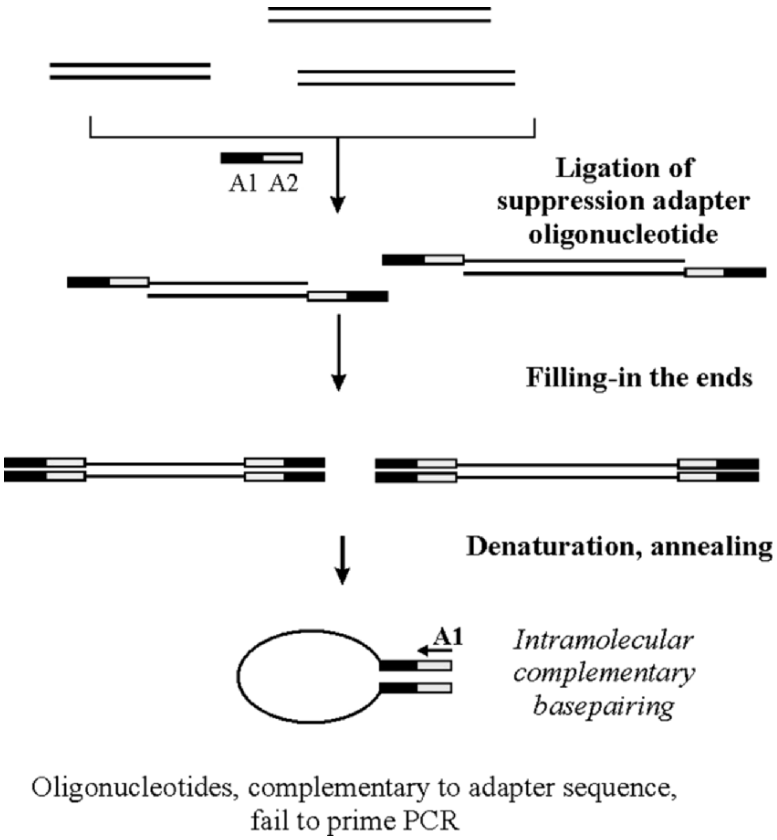

Figure 5. The principle of PCR suppression effect. GC-rich inverted repeats (suppression adapters) base pair intramoleculary, thus preventing annealing of shorter primer oligonucleotides, designed to the adapter sequence. The strength of such an approach is that it greatly reduces background amplification when adapter-specific PCR primers are used.

primer binding sites for the PCR (Chapter 2). The use of the PCR suppression effect prevents the background PCR amplification with adapter-specific primers.

Figure 6 depicts the schematic representation of the SSH and its application to identifying differences between two bacterial genomes. The DNAs (both tracer and driver) were digested with restriction endonucleases, and the tracer

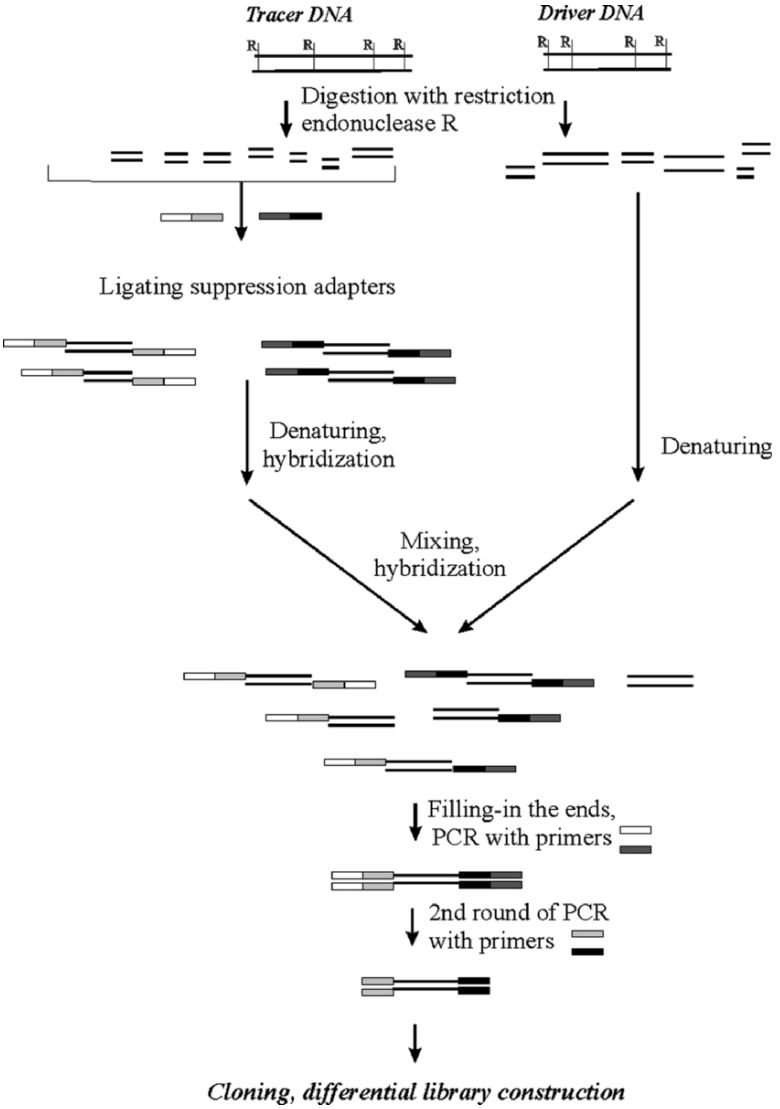

Figure 6. Suppression subtractive hybridization (SSH) scheme, adopted for the comparison of two bacterial genomes. When complex genomes like mammalian DNAs are compared, other approaches must be used.

DNA was then subdivided into two portions (tracer A and tracer B), and two different suppression adapters were ligated separately to these different portions. Nothing was ligated to driver DNA. Both tracer fractions were mixed with driver, denatured, and allowed to hybridize. In the subsequent PCR amplification with primers complementary to both suppression adapters used, only the tracer A or tracer B duplexes, enriched in tracer-specific sequences, could be exponentially amplified further (Akopyants et al. 1998).

Importantly, SSH is also used for comparing transcriptomes (Diatchenko et al. 1996), resulting in high-quality differential cDNA libraries (Chapter 3). In this application, PCR suppression effect made it possible to solve in part one of the most important general problems of cDNA analysis – loss of the rare transcripts from the resulting libraries. The method developed by Sergey Lukyanov's team for normalization of the cDNA libraries in solution (Figure 7; see also Chapter 5) permits an easy and effective smoothing of the concentrations of rare and abundant cDNAs. Two different suppression adapters are ligated to two portions of the fragmented double-stranded cDNA, separately denatured, and then allowed to reanneal for a short time. At this stage, mostly highly abundant sequences hybridize with each other. These fractions are then mixed and allowed to hybridize again without melting. Those cDNAs which did not form duplexes during the first hybridization may hybridize now to form double-stranded molecules. The hybridized cDNA ends are built-in and the mixture is further PCR-amplified with primers complementary to suppression adapters used (Figure 7). This results in exponential amplification of only those duplexes which were formed during the second, but not the first, hybridization. The final amplicon is, therefore, enriched in rare transcript replicas. Such a strategy is a good alternative to a more sophisticated laborious approach (Bonaldo et al. 1996), which utilizes a special subtraction of highly abundant transcripts from the total cDNA pools to enrich the libraries in rare cDNAs.

Although the proportion of background false-positive clones after the use of SSH (multiple rounds of subtraction may be used) is usually low, further improvement was reported recently (Rebrikov et al. 2000). In a technique termed "mirror orientation selection" (MOS) (Figure 8), the number of nondifferential clones is reduced based on the observation that the background, which is caused by reassociation of nontracer-specific molecules, appears just by chance, and each type of such background duplexes is presented by a small number of molecules, compared to the "proper", tracer-specific sequences. As the SSH products (Figure 8) harbor adapter sequences, flanking the tracer DNA in *both* tracer orientations (products A and A′ in the figure) rather than the removal of one adapter, denaturation and subsequent reannealing, followed by filling the ends, may result in tracer fragments flanked by the second adapter sequence from both sides. Using single-primer PCR, such tracer fragments can be exponentially amplified (Figure 8). However, this is not the case for background fragments, as they appear in the resulting SSH amplicons by chance, and each type is presented by very low concentrations (whereas the number of such

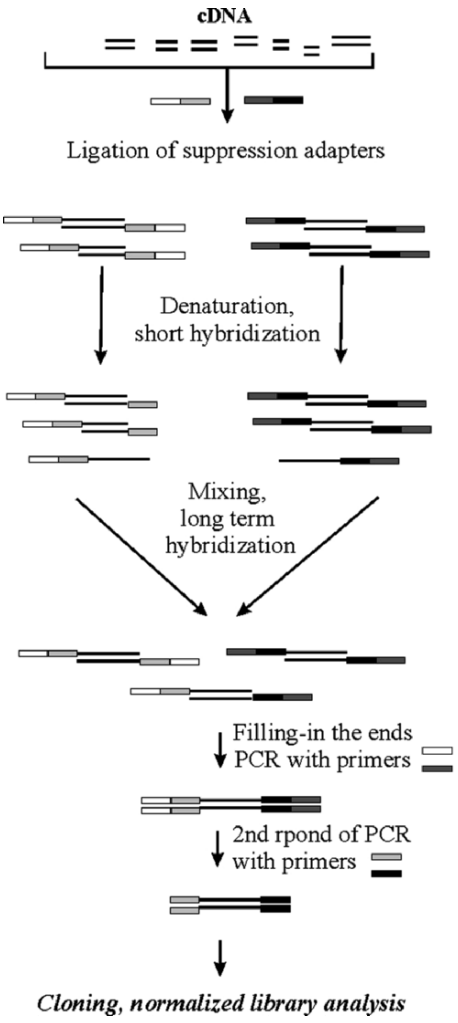

Figure 7. cDNA library normalization scheme using PCR suppression effect. Normalization results in smoothing concentrations of abundant and rare transcripts, which is important for higher transcript repertoire representation in the resulting differential cDNA libraries and for many other applications like EST sequencing.

types may be enormous), and the probability that they will form hybrids carrying adapter sequences at both termini is negligible in most cases (Figure 8, right panel).

Also, SSH may be used in combination with differential display (Pardinas et al. 1998) and microchip hybridization (Yang et al. 1999). The latter seems to be a very promising approach, as it gives an integral picture of a spaciotemporal differential gene expression. Importantly, the direct use of microarray hybridization usually discriminates rare transcripts, whereas preliminary SSH, especially with a stage of cDNA normalization, may greatly enhance both the sensitivity and reproducibility of the results.

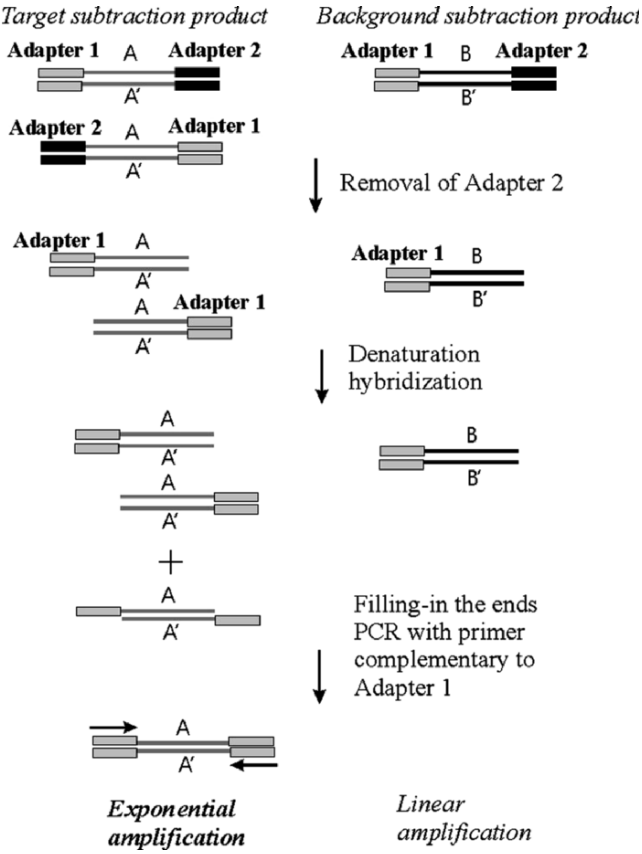

Figure 8. Mirror-oriented selection (MOS) scheme. MOS is aimed at reducing background produced by pseudoselective amplification of occasional unrelated sequences. MOS was found useful for both cDNA and low to medium complexity genomic DNA comparisons.

2.5 Covalently Hybridized Subtraction, Primer Extension Enrichment Reaction, and Other Promising Approaches in Subtractive Hybridization

In Sections 2.2–2.4, I described two most popular techniques based on SH which stand SH among the most effective, timely, and universal approaches in molecular biology like differential display, large-scale shotgun sequencing, EST sequencing, and serial analysis of gene expression (SAGE) (Carulli et al. 1998). As the number of published studies utilizing SH increases, regardless of the lack of serious popularized advances in its methodology during the last 5 years, the potential of SH remains high. In this section, some recent original and promising SH-based techniques will be reviewed. In Shao-Yao Ying's laboratory, a novel approach termed "covalently hybridized subtraction" (CHS) was proposed (Ying and Lin 1999) which utilizes covalent binding of tracer and driver after

hybridization. To this end, driver DNA (termed "subtracter" by the authors) was chemically modified so that DNA strands in all posthybridizational duplexes with driver DNA were covalently bound and, therefore, could not be further PCR-amplified, in contrast to tracer–tracer hybrids. This important modification of the SH is described in Chapter 7. Another interesting approach created to improve RDA technique is based on the protection of 3′-termini of tracer DNA by alpha-thiodeoxyribonucleotides (Kuvbachieva and Goffinet 2002). After hybridization with a nonmodified driver, the resulting mixture was treated with a mixture of *Exo*III nuclease that degrades unprotected 3′-ends and mung bean nuclease that digests single-stranded DNA (ssDNA). Only tracer–tracer hybrids remained in solution, which were further PCR-amplified, cloned, and sequenced.

In the new technique termed "primer extension enrichment reaction" (PEER), Ganova-Raeva et al. (2006) utilized a new original rationale of using short fragments of tracer for a primer extension reaction on the template of driver DNA. To this end, double-stranded tracer DNA was converted into small fragments by extensive endonuclease cleavage and then tagged by ligation to a specially designed adapter. The 3′-end of the adapter incorporates a recognition site for a class IIS restriction endonuclease. The fragments are cleaved with the IIS *Mme* I enzyme to create oligonucleotides with unique sequence at the 3′-end derived from the tracer and a 5′-end derived from the adapter. These adapter-tagged oligonucleotides are annealed to the driver DNA template and extended in the presence of biotinylated ddNTPs. This event blocks any further extension and allows the removal of the biotinylated molecules from the reaction by use of streptavidin-coated magnetic beads. Primers that share driver sequences are blocked and removed leaving only primers with unique sequences that can only be found in the tracer. In the presence of initial full-length tracer DNA, these oligonucleotides can prime an extension reaction from the fragments unique to the tracer (target capture). This step converts the tagged primers into DNA templates suitable for PCR amplification by oligonucleotides containing only the adapter sequences. The final step is expected to generate collection of fragments of different sizes that may be cloned and sequenced. The authors demonstrated that at least for some applications PEER was significantly more sensitive and effective than other variations of SH, including SSH. This very promising technique and its comparison with other methods of differential gene expression screening is described in more detail in Chapter 6.

In concluding this section I should mention a few shortcomings shared among the majority of the SH-based techniques. First, PCR amplification can bias representations of the amplified DNAs due to the *PCR selection* effect, when too many cycles are used (see Section 2.3). Second, SH sometimes does not differentiate between members of evolutionary young gene families sharing high nucleotide sequence identity. The same problem appears at both cDNA and genomic DNA levels when genomic repeats are studied. The last but not the least limitation is that the "small" differences in gene expression, which are less

than one order of magnitude, are usually hardly detectable in SH-based techniques. However, a number of SH modifications have been proposed to partly or completely address these shortcomings (see Chapters 2, 3, 5–7, and 10).

3. FINDING COMMON DNA: COINCIDENCE CLONING

In contrast to SH, which is aimed at the recovery of differential sequences residing in a tracer sample but absent from a driver sample, the approach termed "coincidence cloning" (CC; Chapter 8) was developed to find DNA fragments which are common to the samples under study. The approach is based on cloning 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 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. 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 were the first to exploit the technique of selective PCR suppression (PCR suppression effect was mentioned in Section 2.4 and will be described in detail in Chapter 2), which strongly increased the efficiency of CC (Azhikina et al. 2004, 2006; Azhikina and Sverdlov 2005).

Figure 9 represents a simple model of the use of CC for isolation of evolutionary conserved sequences shared by comparing genomes, reported by Chalaya et al. (2004). Genomic DNAs of human and of New World monkey marmoset *Callithrix pigmaea* were digested with frequent-cutter 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 of the ends with DNA polymerase (Figure 9) and treatment with mismatch-specific nucleases (Chalaya et al. 2004). These enzymes recognize improperly matched dsDNAs and cut such "wrong" hybrids, thus clearly enhancing hybridization specificity (see Chapter 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

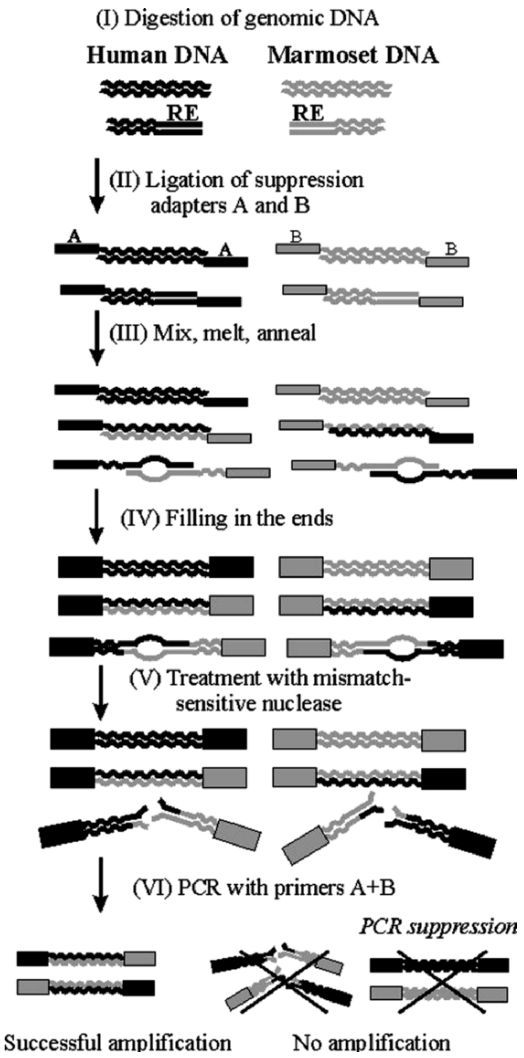

Figure 9. Schematic representation of mispaired DNA rejection (MDR) technique, whose rationale is the specific enzymatic degradation of mismatched hybrids (stage V), which results in significantly lower background mutual hybridization of the two comparing DNA samples. This approach has been demonstrated to be efficient for complex genomic mixtures like primate DNA.

result, the authors 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. The technique is based on the initial fragmentation with a methyl-sensitive restriction enzyme. 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 enzymes 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.

This amplicon is further hybridized to a new-suppression-adapter-ligated fragmented DNA from the genomic locus of interest (the authors analyzed methylation profiles of an ~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). Recently, the same group of authors combined NGSCC with SAGE, thus creating a new technique termed "RIDGES", which is significantly more informative than NGSCC, as its outcome is 10- to 20-fold more information about methylation sites per one sequenced clone (Azhikina et al. 2006).

However, the use of CC is not restricted to genomic DNA analysis. In particular, a 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, which results in the construction of a hybrid genomic DNA or cDNA library, enriched in promoter-active repeats, thus making it possible to create a comprehensive genome-wide map of such repetitive elements (Buzdin et al. 2006a, b). These and other applications of the CC are described in detail in Chapter 8.

4. HYBRIDIZATION IN SOLUTION FOR THE RECOVERY OF GENOMIC POLYMORPHISMS

Unlike SH, which generally deals with finding relatively long differential DNA fragments, this group of methods is aimed at the identification of very small, single nucleotide-scale differences between the comparing DNA samples. The study of such mutations reveals the normal functions of genes, proteins, noncoding RNAs, the causes of many malignancies, and the variability of responses among individuals. A plethora of SNPs are not deleterious by themselves, but are linked to phenotypes associated with diseases and drug responses, thus providing a great opportunity for their use in large-scale association and population studies. Moreover, SNPs are increasingly recognized as important diagnostic markers for the detection of drug-resistant strains of hazardous microorganisms.

An impressive number of research groups working in this field managed to identify 10 million SNPs in recent years. However, this figure seems negligible compared to the real number of SNPs and other mutations present in the

genomes. The ideal method for mutation and SNP recovery would detect mutations in large fragments of DNA and position them to single base-pair accuracy, and would be sensitive, precise, and robust. Currently, the need in mutation detection is reflected by the plethora of chemical, enzymatic, bioinformatical, and physically based techniques. Many mutation discovery methods quickly and effectively indicate the presence of a mutation in a sample region, but fail to resolve its characterization and localization; another family of methods permits precise mutation mapping, but in a more laborious and expensive way. The group of novel approaches for mutation detection based on DNA hybridization in solution, which combines high performance, cost-efficiency, reliability, and detailed mutation characterization, will be reviewed in Chapter 9.

At present, mutation discovery through Sanger sequencing is often thought of as the "gold standard" for mutation detection. This perception is distorted due to the fact that this is the *only* method of mutation identification, but this does not mean it is the best for mutation detection. The fact that many scanning methods detect 5–10% of mutant molecules in a wild-type environment immediately indicates that these methods are advantageous over sequencing, at least for some purposes. Using bioinformatical approaches, a large number of mutations (mostly SNPs) were recently discovered.

However, these methodologies require prior knowledge of target sequences, normally obtained through DNA sequencing, and mutation recovery in such case is usually performed by multiple sequence alignment of publicly available sequence data. Recent studies indicate that only a small percentage of mutations can be discovered using this approach and, in particular, that SNPs with low frequency are often missed. It is clear now that high-throughput methods for detecting these variations are needed for in-population screening for complex genetic diseases in which extended genomic loci, large genes, and/or several genes may be affected. To meet the need of these studies, several groups of approaches have been developed. All of them are based on the rationale that mutation-containing DNA molecule will form mismatches at the mutation site when hybridized to the reference wild-type DNA (Figure 10). Thus, when mutant and wild-type DNA are hybridized together, two complementary mismatches are formed. Therefore, the detection and correct position of such mismatches is the key for mutation recovery.

Such mispaired nucleotides can be identified directly or indirectly using very different chemical, enzymatic, or physical approaches, which offer excellent detection efficiencies coupled with high throughput and low unit cost. It should be noted that definition of the mutational change obviously requires a sequencing step, at least to confirm the results. But in this case sequencing is targeted, not a "fishing expedition" in which the region where mutation occurred is unknown and plenty of sequencing work is absolutely required. As a result, these methods are able to cut the costs of detecting a mutation by one order of magnitude or more. Briefly, chemical approaches utilize chemical cleavage or modification of the mispaired nucleotides, enzymatic techniques employ enzymatic recognition of the

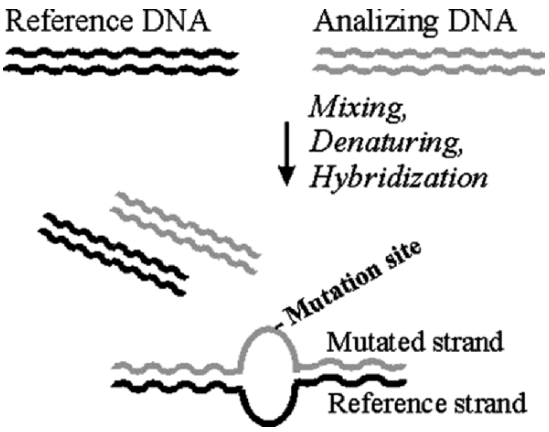

Figure 10. When hybridized, mutated and reference wild-type DNAs form heteroduplexes having mispaired regions, corresponding to the mutation sites.

mismatch (with further binding, cleavage, modification, or ligation of the DNA at the mispaired nucleotides), whereas physical methods look for a physical difference between the mutant strand and wild-type strands of DNA, being based either on physical isolation of imperfectly matched DNA hybrids (like electrophoretic separation) or on finding differences in mismatched versus perfect DNA hybrid physical peculiarities. All these approaches utilize nucleic acids hybridization in solution, and are described in more detail in Chapter 9, in comparison with each other and with direct sequencing-based approaches.

5. CONCLUSIONS

In this chapter I have briefly illustrated how the methods based on nucleic acids hybridization in solution could be helpful for a number of applications, and have provided a short overview underlining the methods' principles for each major technique, except for microarray hybridization, which is thoroughly reviewed in Chapter 11. A more detailed description will be provided in the following chapters of this book: Chapter 2 – PCR suppression effect; Chapter 4 – use of stem-loop oligonucleotides; Chapters 3, 6, and 7 – important modern variations of SH; Chapter 5 – normalization of the cDNA libraries; Chapter 8 – CC; Chapter 9 – hybridization-based mutation detection; and Chapter 10 – current attempts to improve the hybridization specificity.

A wide spectrum of experimental tasks covered by these approaches includes finding differential sequences in both genomic DNAs and cDNAs using microarrays, SH or ordered differential display, genome walking, multiplex PCR, cDNA library construction starting from small amount of total RNA, rapid amplification of cDNA ends (RACE), effective smoothing of the concentrations of rare

and abundant transcripts in cDNA libraries, recovery of promoter-active repeats and differentially methylated genomic DNA, identification of common DNA in genomic or cDNA sources, new gene mapping, finding evolutionary conserved sequences and both single-nucleotide and extended mutation discovery or large-scale monitoring. Of course, there is no panacea, an ideal method that would solve all technical problems that researchers face, but a combination of the above approaches will be likely fruitful for conducting successful research. The international team of the authors of this book has tried both to elucidate the current state of the art in hybridization techniques and to help the readers in choosing an appropriate method for performing an experiment in the most efficient way. We hope the book will be useful to all those interested in the modern life science methodologies.

REFERENCES

Akopyants NS, Fradkov A, Diatchenko L, Hill JE, Siebert PD, Lukyanov SA, Sverdlov ED, Berg DE (1998) PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proc Natl Acad Sci USA 95:13108–13113

Ayyanathan K, Francis VS, Datta S, Padmanaban G (1995) Development of specific DNA probes and their usage in the detection of Plasmodium vivax infection in blood. Mol Cell Probes 9:239–246

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

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

Bautz EK, Reilly E (1966) Gene-specific messenger RNA: isolation by the deletion method. Science 151:328–330

Bogush ML, Velikodvorskaya TV, Lebedev YB, Nikolaev LG, Lukyanov SA, Fradkov AF, Pliyev BK, Boichenko MN, Usatova GN, Vorobiev AA, Andersen GL, Sverdlov ED (1999) Identification and localization of differences between *Escherichia coli* and *Salmonella typhimurium* genomes by suppressive subtractive hybridization. Mol Gen Genet 262:721–729

Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 6:791–806

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37

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 (2003) Genome-wide targeted search for human specific and polymorphic L1 integrations. Hum Genet 112:527–533

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 non-repetitive 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

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 Suppl 30–31:286–296

Cekan SZ (2004) Methods to find out the expression of activated genes. Reprod Biol Endocrinol 2:68

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

Chen H, Pulido JC, Duyk GM (1995) MATS: a rapid and efficient method for the development of microsatellite markers from YACs. Genomics 25:1–8

Chien Y, Becker DM, Lindsten T, Okamura M, Cohen DI, Davis MM (1984) A third type of murine T-cell receptor gene. Nature 312:31–35

Cho TJ, Park SS (1998) A simulation of subtractive hybridization. Nucleic Acids Res 26:1440–1448

Clapp JP, McKee RA, Allen-Williams L, Hopley JG, Slater RJ (1993) Genomic subtractive hybridization to isolate species-specific DNA sequences in insects. Insect Mol Biol 1:133–138

Cook D, Sequeira L (1991) The use of subtractive hybridization to obtain a DNA probe specific for Pseudomonas solanacearum race 3. Mol Gen Genet 227:401–410

Cruz-Reyes JA, Ackers JP (1992) A DNA probe specific to pathogenic *Entamoeba histolytica*. Arch Med Res 23:271–275

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

Diatchenko L, Lukyanov S, Lau YF, Siebert PD (1999) Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. Methods Enzymol 303:349–380

Drew AC, Brindley PJ (1995) Female-specific sequences isolated from Schistosoma mansoni by representational difference analysis. Mol Biochem Parasitol 71:173–181

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

Frohme M, Camargo AA, Czink C, Matsukuma AY, Simpson AJ, Hoheisel JD, Verjovski-Almeida S (2001) Directed gap closure in large-scale sequencing projects. Genome Res 11:901–903

Ganova-Raeva L, Zhang X, Cao F, Fields H, Khudyakov Y (2006) Primer Extension Enrichment Reaction (PEER): a new subtraction method for identification of genetic differences between biological specimens. Nucleic Acids Res 34:e76

Hara E, Kato T, Nakada S, Sekiya S, Oda K (1991) Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. Nucleic Acids Res 19:7097–7104

Herfort MR, Garber AT (1991) Simple and efficient subtractive hybridization screening. Biotechniques 11:598, 600, 602–604

Hla T, Maciag T (1990) Isolation of immediate-early differentiation mRNAs by enzymatic amplification of subtracted cDNA from human endothelial cells. Biochem Biophys Res Commun 167:637–643

Jin H, Cheng X, Diatchenko L, Siebert PD, Huang CC (1997) Differential screening of a subtracted cDNA library: a method to search for genes preferentially expressed in multiple tissues. Biotechniques 23:1084–1086

Kavathas P, Sukhatme VP, Herzenberg LA, Parnes JR (1984) Isolation of the gene encoding the human T-lymphocyte differentiation antigen Leu-2 (T8) by gene transfer and cDNA subtraction. Proc Natl Acad Sci USA 81:7688–7692

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

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

Kuvbachieva AA, Goffinet AM (2002) A modification of representational difference analysis, with application to the cloning of a candidate in the Reelin signalling pathway. BMC Mol Biol 3:6

Kuze K, Shimizu A, Honjo T (1989) A new vector and RNase H method for the subtractive hybridization. Nucleic Acids Res 17:807

Laman AG, Kurjukov SG, Bulgakova EV, Anikeeva NN, Brovko FA (2001) Subtractive hybridization of biotinylated DNA in phenol emulsion. J Biochem Biophys Methods 50:43–52

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

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971

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

Lisitsyn NA, Lisitsina NM, Dalbagni G, Barker P, Sanchez CA, Gnarra J, Linehan WM, Reid BJ, Wigler MH (1995) Comparative genomic analysis of tumors: detection of DNA losses and amplification. Proc Natl Acad Sci USA 92:151–155

Lopez-Fernandez LA, del Mazo J (1993) Construction of subtractive cDNA libraries from limited amounts of mRNA and multiple cycles of subtraction. Biotechniques 15:654–6, 658–659

Martin KJ, Pardee AB (2000) Identifying expressed genes. Proc Natl Acad Sci USA 97:3789–3791

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

Nadezhdin EV, Vinogradova TV, Sverdlov ED (2001) Interspecies subtractive hybridization of cDNA from human and chimpanzee brains. Dokl Biochem Biophys 381:415–418

Palazzolo MJ, Meyerowitz EM (1987) A family of lambda phage cDNA cloning vectors, lambda SWAJ, allowing the amplification of RNA sequences. Gene 52:197–206

Pardinas JR, Combates NJ, Prouty SM, Stenn KS, Parimoo S (1998) Differential subtraction display: a unified approach for isolation of cDNAs from differentially expressed genes. Anal Biochem 257:161–168

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

Rosenberg M, Przybylska M, Straus D (1994) "RFLP subtraction": a method for making libraries of polymorphic markers. Proc Natl Acad Sci USA 91:6113–6117

Rubenstein JL, Brice AE, Ciaranello RD, Denney D, Porteus MH, Usdin TB (1990) Subtractive hybridization system using single-stranded phagemids with directional inserts. Nucleic Acids Res 18:4833–4842

Rubin CM, Leeflang EP, Rinehart FP, Schmid CW (1993) Paucity of novel short interspersed repetitive element (SINE) families in human DNA and isolation of a novel MER repeat. Genomics 18:322–328

Sagerstrom CG, Sun BI, Sive HL (1997) Subtractive cloning: past, present, and future. Annu Rev Biochem 66:751–783

Sallie R (1995) Isolation of candidate genes by subtractive and sequential (Boolean) hybridization: an hypothesis. Med Hypotheses 45:142–146

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

Schutte M, da Costa LT, Hahn SA, Moskaluk C, Hoque AT, Rozenblum E, Weinstein CL, Bittner M, Meltzer PS, Trent JM, et al. (1995) Identification by representational difference analysis of a homozygous deletion in pancreatic carcinoma that lies within the BRCA2 region. Proc Natl Acad Sci USA 92:5950–5954

Sharma P, Lonneborg A, Stougaard P (1993) PCR-based construction of subtractive cDNA library using magnetic beads. Biotechniques 15:610, 612

Southern EM (1992) Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. Biotechnology 24:122–139

Sverdlov ED (1993) Subtractive hybridization-a technique for extracting DNA s equences, discriminating between two closely-related genomes. Mol Gen Mikrobiol Virusol 6:3–12

Sverdlov ED, Ermolaeva OD (1993) Subtractive hybridization. Theoretical analysis, and a principle of the trap. Bioorg Khim 19:1081–1088

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

Timblin C, Battey J, Kuehl WM (1990) Application for PCR technology to subtractive cDNA cloning: identification of genes expressed specifically in murine plasmacytoma cells. Nucleic Acids Res 18:1587–1593

Travis GH, Sutcliffe JG (1988) Phenol emulsion-enhanced DNA-driven subtractive cDNA cloning: isolation of low-abundance monkey cortex-specific mRNAs. Proc Natl Acad Sci USA 85:1696–1700

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

Wang Z, Brown DD (1991) A gene expression screen. Proc Natl Acad Sci USA 88:11505–11509

Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737–738

Welcher AA, Torres AR, Ward DC (1986) Selective enrichment of specific DNA, cDNA and RNA sequences using biotinylated probes, avidin and copper-chelate agarose. Nucleic Acids Res 14:10027–10044

Wieland I, Bolger G, Asouline G, Wigler M (1990) A method for difference cloning: gene amplification following subtractive hybridization. Proc Natl Acad Sci USA 87:2720–2724

Yang GP, Ross DT, Kuang WW, Brown PO, Weigel RJ (1999) Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. Nucleic Acids Res 27:1517–1523

Ying SY (2004) Complementary DNA libraries: an overview. Mol Biotechnol 27:245–252

Ying SY, Lin S (1999) High-performance subtractive hybridization of cDNAs by covalent bonding between specific complementary nucleotides. Biotechniques 26:966–968, 970–972, 979 passim

Zeschnigk M, Horsthemke B, Lohmann D (1999) Detection of homozygous deletions in tumors by hybridization of representational difference analysis (RDA) products to chromosome-specific YAC clone arrays. Nucleic Acids Res 27:e30