Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 6025–6030, June 1996
Biochemistry

Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries

LUDA DIATCHENKO*, YUN-FAI CHRIS LAU†, AARON P. CAMPBELL†, ALEX CHENCHIK*, FAUZIA MOQADAM*, BETTY HUANG*, SERGEY LUKYANOV‡, KONSTANTIN LUKYANOV‡, NADYA GURSKAYA‡, EUGENE D. SVERDLOV‡, AND PAUL D. SIEBERT*

*CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303; †Division of Cell and Developmental Genetics, Department of Medicine, University of California, San Francisco, CA 94121; and ‡Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Science, Mikluho-Maklaya 16/10, V-437 Moscow 117871, Russia

Communicated by Mark M. Davis, Stanford University School of Medicine, Stanford, CA, February 1, 1996 (received for review October 26, 1995)

ABSTRACT **A new and highly effective method, termed suppression subtractive hybridization (SSH), has been developed for the generation of subtracted cDNA libraries. It is based primarily on a recently described technique called suppression PCR and combines normalization and subtraction in a single procedure. The normalization step equalizes the abundance of cDNAs within the target population and the subtraction step excludes the common sequences between the target and driver populations. In a model system, the SSH technique enriched for rare sequences over 1,000-fold in one round of subtractive hybridization. We demonstrate its usefulness by generating a testis-specific cDNA library and by using the subtracted cDNA mixture as a hybridization probe to identify homologous sequences in a human Y chromosome cosmid library. The human DNA inserts in the isolated cosmids were further confirmed to be expressed in a testis-specific manner. These results suggest that the SSH technique is applicable to many molecular genetic and positional cloning studies for the identification of disease, developmental, tissue-specific, or other differentially expressed genes.**

In higher eukaryotes, biological processes such as cellular growth and organogenesis are mediated by programs of differential gene expression. To understand the molecular regulation of these processes, the relevant subsets of differentially expressed genes of interest must be identified, cloned, and studied in detail. Subtractive cDNA hybridization has been a powerful approach to identify and isolate cDNAs of differentially expressed genes (1–3). Numerous cDNA subtraction methods have been reported. In general, they involve hybridization of cDNA from one population (tester) to excess of mRNA (cDNA) from other population (driver) and then separation of the unhybridized fraction (target) from hybridized common sequences. The latter step is usually accomplished by hydroxylapatite chromatography (3), avidin-biotin binding (1, 4, 5), or oligo(dT)$_{30}$-latex beads (2). Despite the successful identification of numerous important genes such as the T-cell receptors (3) by these methods, they are usually inefficient for obtaining low abundance transcripts. These subtraction techniques often require greater then 20 μg of poly(A)$^+$ RNA, involve multiple or repeated subtraction steps, and are labor intensive.

Recently a new PCR-based technique, called representational difference analysis, has been described that does not require physical separation of single-stranded (ss) and double-stranded (ds) cDNAs. Representational difference analysis has been applied to enrich for genomic fragments that differ in size or representation (6) and to clone differentially expressed cDNAs (7). However, representational difference analysis does not resolve the problem of the wide differences in abundance of individual mRNA species. Consequently, multiple rounds of subtraction are still needed (7). The mRNA differential display (8) and RNA fingerprinting by arbitrary primed PCR (9) are potentially faster methods for identifying differentially expressed genes. However, both of these methods have a high level of false positives (10, 11), biased for high copy number mRNA (12) and might be inappropriate in experiments in which only a few genes are expected to vary (11).

Here we present a new PCR-based cDNA subtraction method, termed suppression subtractive hybridization (SSH), and demonstrate its effectiveness. SSH is used to selectively amplify target cDNA fragments (differentially expressed) and simultaneously suppress nontarget DNA amplification. The method is based on the suppression PCR effect previously described by our laboratories: long inverted terminal repeats when attached to DNA fragments can selectively suppress amplification of undesirable sequences in PCR procedures (14, 15). We have recently applied the suppression PCR effect in chromosome walking (14) and rapid amplification of cDNA ends (15). The subtraction method described here overcomes the problem of differences in mRNA abundance by incorporating a hybridization step that normalizes (equalizes) sequence abundance during the couse of subtraction by standard hybridization kinetics. It eliminates any intermediate step(s) for physical separation of ss and ds cDNAs, requires only one subtractive hybridization round, and can achieve greater than 1,000-fold enrichment for differentially expressed cDNAs. We demonstrate the effectiveness of the SSH method by generating a testis-specific cDNA library and characterizing selected cDNA clones. Furthermore, we show that subtracted cDNA mixture can be used directly as a hybridization probe for screening recombinant DNA libraries, such as a human Y chromosome cosmid library, thereby identifying chromosome-specific and tissue-specific expressed sequences.

MATERIALS AND METHODS

Oligonucleotides. The following gel-purified oligonucleotides were used.

(*i*) cDNA synthesis primer: Pr16, 5′-TTTTGTACAAGCTT$_{30}$-3′.

(*ii*) Adapters:

adapter 1, 5′-GTAATACGACTCACTATAGGGCTCGAGCGG-
CCGCCCGGGCAGGT-3′
3′-CCCGTCCA-5′

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: SSH, suppression subtractive hybridization.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base (accession nos. H48477, H48478, H48931–H48939, H52858–H54046, H54559–H54560, H56769–H56778, and H64202–H64207).

adapter 2, 5′-TGTAGCGTGAAGACGACAGAAAGGGCG-
TGGTGCGGAGGGCGGT-3′
3′-GCCTCCCGCCA-5′.

(*iii*) PCR primers:

P1, 5′-GTAATACGACTCACTATAGGGC-3′

P2, 5′-TGTAGCGTGAAGACGACAGAA-3′

PN1, 5′-TCGAGCGGCCGCCCGGGCAGGT-3′

PN2, 5′-AGGGCGTGGTGCGGAGGGCGGT-3′.

Driver Preparation. Driver ds cDNA was synthesized from 2 μg each of different human poly(A)$^+$ RNA (CLONTECH), using the Great Lengths cDNA Synthesis Kit (CLONTECH) and 1 ng of oligonucleotide Pr16 as a primer. First- and second-strand cDNA synthesis and blunt-ending of DNA ends by T4 DNA polymerase were carried out according to the manufacturer's protocol. The resulting cDNA pellet was dissolved in 10 μl of deionized water and digested by *Rsa*I or *Hae*III in a 50 μl reaction mixture containing 15 units of enzyme (New England Biolabs) for 3 h. The cDNAs were then phenol-extracted, ethanol-precipitated, and resuspended in 7 μl of deionized water. The cDNA preparations from 10 human tissues were then mixed together in equal proportions. The final concentration of driver was ≈300 ng/μl.

Tester Preparation. *Rsa*I or *Hae*III digested ds tester cDNA was prepared as described above for the driver. Digested tester cDNA (1 μl) was diluted in 5 μl of H_2O. The diluted tester cDNA (2 μl) was then ligated to 2 μl of adapter 1 and adapter 2 (10 μM) in separate ligation reactions in a total volume of 10 μl at 16°C overnight, using 0.5 units of T4 DNA ligase (Life Technologies) in the buffer supplied from the manufacturer. After ligation, 1 μl of 0.2 M EDTA was added and the samples were heated at 70°C for 5 min to inactivate the ligase and stored at −20°C.

Subtractive Hybridization. Two microliters of driver ds cDNA (600 ng) was added to each of two tubes containing 2 μl of adapter 1- and adapter 2-ligated tester cDNA (20 ng). The samples were mixed, ethanol precipitated, and then resuspended in 1.5 μl of hybridization buffer [50 mM Hepes, pH 8.3/0.5 M NaCl/0.02 mM EDTA, pH 8.0/10% (wt/vol) PEG 8000]. The solution was overlaid with mineral oil, the DNAs were denatured (1.5 min, 98°C), and then allowed to anneal for 10 h at 68°C. After this first hybridization, the two samples were combined and a fresh portion of heat-denatured driver (≈150 ng) in 1.5 μl of hybridization buffer was added. The sample was allowed to hybridize for an additional 10 h at 68°C. The final hybridization was then diluted in 200 μl of dilution buffer (20 mM Hepes, pH 8.3/50 mM NaCl/0.2 mM EDTA), heated at 72°C for 7 min and stored at −20°C.

PCR Amplification. For each subtraction, we performed two PCR amplifications. The primary PCR was conducted in 25 μl. It contained 1 μl of diluted, subtracted cDNA, 1 μl of PCR primer P1 (5 μM), 1 μl of PCR primer P2 (5 μM), and 22 μl of PCR master mixture prepared using the Advantage cDNA PCR Core Kit (CLONTECH). PCR was performed with the following parameters: 75°C for 7 min; 30 cycles at (91°C for 30 sec; 68°C for 30 sec; 72°C for 2.5 min); and a final extension at 68°C for 7 min. The amplified products were diluted 10-fold in deionized water. Some of the product (1 μl) was then used as a template in secondary PCR for 10 cycles under the same conditions used for the primary PCR, except PCR primer P1 and P2 were replaced with nested PCR primer PN1 and PN2, respectively. In some cases, we also performed a secondary PCR amplification of only the 3′-ends of the subtracted cDNA fragments. Primer Pr16 was used in combination with one of the nested PCR primers. PCR was performed for 17 cycles (91°C, 30 sec; 60°C, 30 sec; 72°C, 2.5 min). The PCR products were analyzed by either 8 M urea/6% acrylamide or 2% agarose gel electrophoresis.

Cloning and Analysis of the Subtracted cDNA. Products from the secondary PCRs were inserted into pCRII using a T/A cloning kit (Invitrogen). Plasmid or cosmid DNAs were prepared using QIAwell 8 Plus Kit (Qiagen) according to the manufacturer's protocol. DNA sequencing was performed by the chain termination reaction manually or by automated means at the Biomolecular Resource Center (University of California, San Francisco). Nucleic acid homology searches were performed using the BLAST program through e-mail servers at the National Center for Biotechnology Information (National Institutes of Health, Bethesda).

The inserts in pCRII were amplified for 20 cycles under similar conditions as in the secondary PCR described above. Amplified inserts were then purified using Qiaquick Spin PCR Purification Kit (Qiagen), ^{32}P-labeled by random priming (15), and used as probes for Northern hybridization of Human Multiple Tissue Northern Blots (CLONTECH). Hybridization was performed in 5 ml of ExpressHyb solution (CLONTECH) using 5 × 10^6 cpm per 100 ng of cDNA probe or 5 × 10^7 cpm per 500 ng of cosmid probe for 1.5 to 16 h at 65°C in the

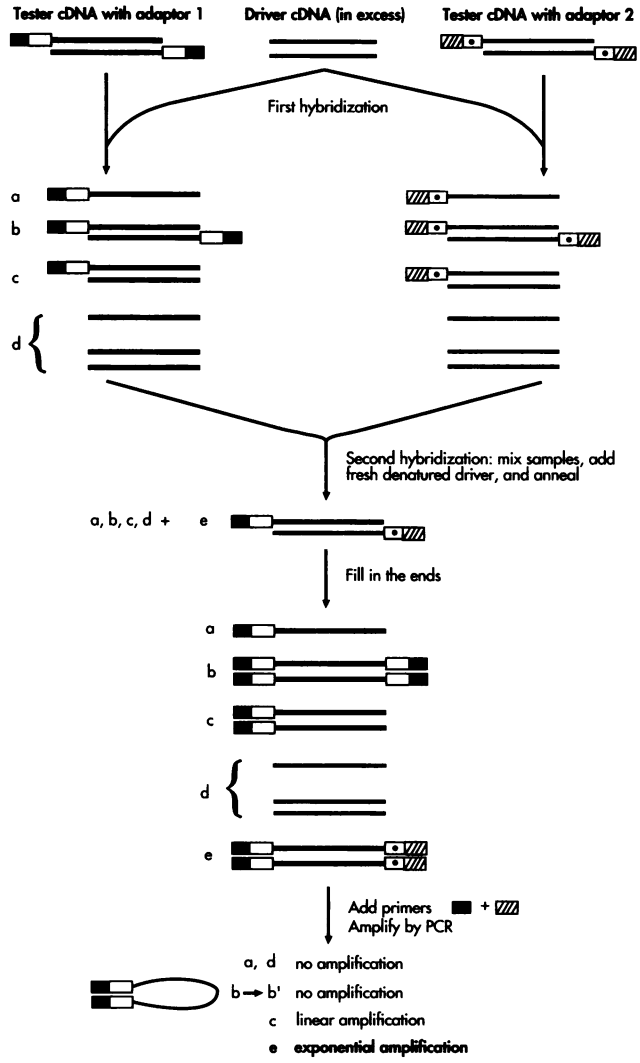

FIG. 1. Scheme of the SSH method. Solid lines represents the *Rsa*I digested tester or driver cDNA. Solid boxes represent the outer part of the adaptor 1 longer strand and corresponding PCR primer P1 sequence. Shaded boxes represent the outer part of the adaptor 2 longer strand and corresponding PCR primer P2 sequence. Clear boxes represent the inner part of the adaptors and corresponding nested PCR primers PN1 and PN2. Note that after filling in the recessed 3′ ends with DNA polymerase, types a, b, and c molecules having adapter 2 are also present but are not shown.

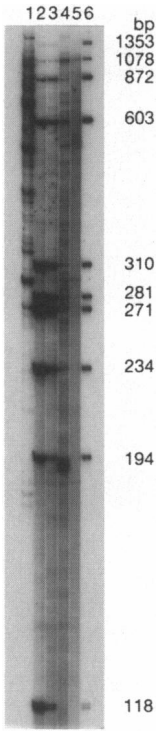

FIG. 2. Results of SSH in a model system using a reconstituted tester. ϕX174 DNA was added to skeletal muscle cDNA and subtracted against original skeletal muscle cDNA according to the SSH protocol described. One μCi of [α-^{32}P]dCTP (3000 Ci/mmol; 1 Ci = 37 GBq) was included in the secondary PCR step and the products were detected by autoradiography. Lane 1, PCR amplification of unsubstracted skeletal muscle cDNA with 0.1% ϕX 174 DNA added; lanes 2–5, final PCR results of the SSH procedure when ϕX 174 DNA was added to the tester in concentrations corresponding to 0.1%, 0.01% 0.001%, and 0.000% of the total tester DNA; lane 6, *Hae*III digest of the ϕX 174 ligated with adapters 1 and 2 and amplified using nested primers PN1 and PN2.

presence of 0.1 mg/ml sonicated heat-denatured salmon sperm DNA. Filters were washed at 65°C with 2× SSC and 0.5% SDS four times for 15 min each, and a high stringency wash with 0.2× SSC and 0.5% SDS at 65°C for 30 min. The cosmids (500 ng) and total subtracted cDNA mixture (100 ng) when used as probes were preannealed with human Cot-1 DNA (100 μg, CLONTECH) in 4× SSC and 0.1% SDS, at 65°C for 4 h before hybridization to the filters.

Hybridization Screening of Human Y Chromosome Cosmid Library. The human Y cosmid library (LL0YNC03 "M") was constructed from human Y chromosomes that were flow-sorted from a human-hamster somatic hybrid cell by the National Gene Library Project (Lawrence Livermore National Laboratory) using the Lawrist 16 cosmid vector as described (17, 18). It consists of approximately 12,000 independent colonies arrayed in 130 96-well microtiter dishes. The 3,072 colonies from plate no. 60–91 were replicated on nylon filters in a 2 × 2 format, grown overnight on Luria–Bertani broth plus kanamycin plates, and immobilized *in situ* according to established procedure (16). Hybridization was performed with [^{32}P]dCTP-labeled subtracted cDNA mixture as described above.

THE PRINCIPLE OF SUPPRESSION SUBTRACTIVE HYBRIDIZATION

A schematic representation of the SSH method is shown in Fig. 1. The differentially expressed cDNAs (target) are present in the "tester" cDNA but are absent (or present at lower levels) in "driver" cDNAs. The tester and driver ds cDNAs are first digested with a four-base cutting restriction enzyme that yields blunt ends. The tester cDNA fragments are then divided into two samples (1 and 2) and ligated with two different adapters (adapter 1 and adapter 2), resulting in two populations of tester, (1) and (2). The ends of the adapters are designed without phosphate groups, so that only the longer strand of each adapter can be covalently attached to the 5′-ends of the cDNA.

The SSH technique uses two hybridizations. First, an excess of driver is added to each sample of the tester. The samples are then heat-denatured and allowed to anneal. The ss cDNA tester fraction (a) is normalized, meaning concentrations of high and low abundance cDNAs become roughly equal. Normalization occurs because the reannealing process generating homo-hybrid cDNAs (b) is faster for the more abundant molecules, due to the second order kinetics of hybridization (19). Furthermore, the ss cDNAs in the tester fraction (a) are significantly enriched in cDNAs for differentially expressed genes, as "common" nontarget cDNAs form heterohybrids (c) with the driver.

In the second hybridization, the two samples from the first hybridization are mixed together. Only the remaining normalized and subtracted ss tester cDNAs are able to reassociate and form (b), (c), and new (e) hybrids. Addition of a second portion of denatured driver at this stage further enriches fraction (e) for differentially expressed genes. The newly formed (e) hybrids have an important feature that distinguishes them from hybrids (b) and (c) formed during first and second hybridizations. This feature is that they have different adapter sequences at their 5′-ends. One is from sample 1 and the other is from sample 2. The two sequences allow preferential amplification of the subtracted and normalized fraction (e) using PCR and

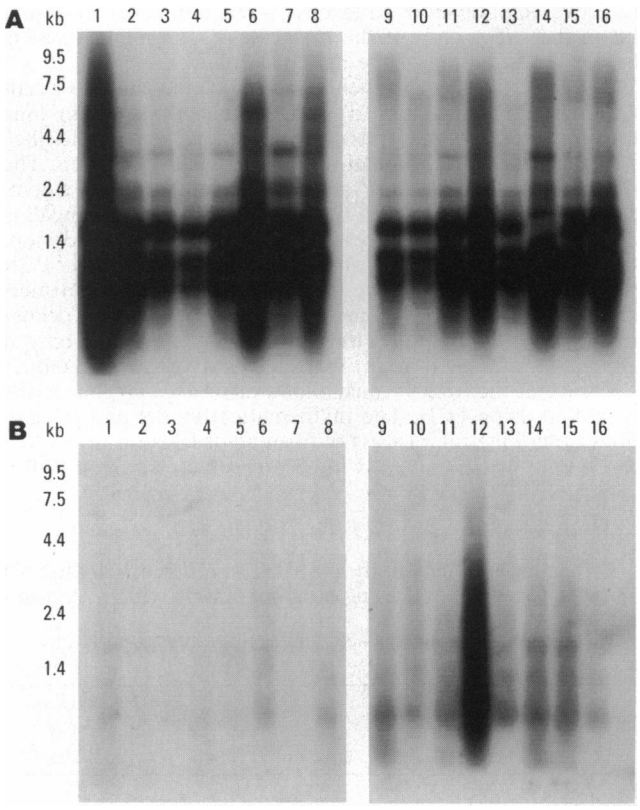

FIG. 3. Northern blot analysis of unsubtracted (*A*) and subtracted (*B*) testis cDNA. The multiple tissue Northern blots contained approximately 2 μg of poly(A)$^+$ RNA from human (1) heart, (2) brain, (3) placenta, (4) lung, (5) liver, (6) skeletal muscles, (7) kidney, (8) pancreas, (9) spleen, (10) thymus, (11) prostate, (12) testis, (13) ovary, (14) small intestine, (15) colon, and (16) peripheral blood leukocyte. The exposure times were 14 h (*A*) and 4 days (*B*).

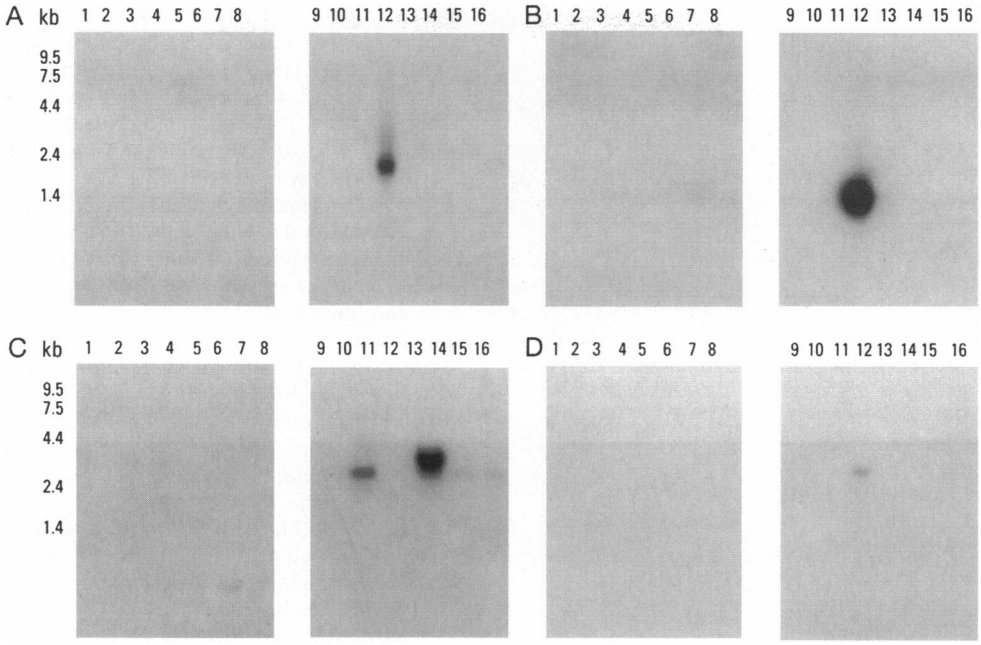

FIG. 4. Examples of Northern blot analysis with cloned testis-specific cDNAs. The subtracted cDNA was cloned into T/A cloning vector. The inserts of 10 randomly selected clones were hybridized with multiple tissues Nothern blots contained approximately 2 μg of poly(A)$^+$ RNA from human (1) heart, (2) brain, (3) placenta, (4) lung, (5) liver, (6) skeletal muscles, (7) kidney, (8) pancreas, (9) spleen, (10) thymus, (11) prostate, (12) testis, (13) ovary, (14) small intestine, (15) colon, and (16) peripheral blood leukocyte. Clones P-3, P-4, P-1, and P-7 are shown in *A–D*, respectively. The exposure times were 14 h (*A*), 5 h (*B*), 3 days (*C*), and 14 days (*D*).

a pair of primers, P1 and P2, which correspond to the outer part of the adapter 1 and 2, respectively. To accomplish this selective amplification, an extension reaction is performed to fill in the sticky ends of the molecules for primer annealing before the initiation of the PCR procedure.

In all PCR cycles, exponential amplification can only occur with type (e) molecules. Type (b) molecules contain long inverted repeats on the ends and form stable "panhandle-like" structures after each denaturation-annealing PCR step. The resulting "panhandle-like" structure cannot serve as a template for exponential PCR, because intramolecular annealing of longer adapter sequences is both highly favored and more stable than intermolecular annealing of the much shorter PCR primers (14, 15). This is the suppression PCR effect. Furthermore, type (a) and (d) molecules do not contain primer binding sites, and type (c) molecules can be amplified only at a linear rate. Only type (e) molecules have different adapter sequence at their ends which allows them to be exponentially amplified using PCR. The mathematical model and calculations describing the process of forming of fraction (e) as well as the rate of enrichment will be presented elsewhere (20).

RESULTS

Efficiency and Sensitivity of SSH. The efficiency of the SSH method was evaluated using a model system with a reconstituted artificial tester. To create the tester samples, we added various amounts of bacteriophage ϕX174 DNA as targets for subtraction in human skeletal muscle ds cDNAs. The same cDNAs without viral DNA were used as driver. The amount of ϕX174 DNA added was 0.1%, 0.01%, and 0.001% of the human cDNAs and corresponded to approximately 100, 10, and 1 copies of target restriction fragments per cell, respectively (21). Both the tester and the driver were digested by *Hae*III and used in SSH procedure. Fig. 2 shows the electrophoretic analysis of the final PCR products. Before subtraction, the bands corresponding to the ϕX174 DNA were indiscernible (Fig. 2, lane 1). After subtraction, the pattern of bands was almost identical (lanes 2–4) to that of control ϕX174 DNA (lane 6). Results of densitometric scanning of the autoradiogram indicated that there had been an enrichment of viral DNA fragments of 100-, 1,000-, and 5,000-fold for the samples containing 0.1%, 0.01%, and 0.001% of viral DNA, respectively.

Table 1. Summary of sequences and clones represented in the testis-specific cDNA library

Match category	No. clones represented (%)	No. known as testis-specific genes (%)
Exact human match	7 (13.5)	4 (7.7)
Known genes	5 (9.6)	4 (7.7)
ESTs	2 (3.8)	
Novel sequences	55 (86.5)	4 (7.7)
Nonexact human match	9 (17.3)	
Nonhuman match	4 (7.7)	4 (7.7)
No data base match	32 (61.5)	
Total	62 (100)	8 (15.4)

Generation and Cloning of Tissue Specific cDNAs. To demonstrate the usefulness of SSH, we have used this tech-

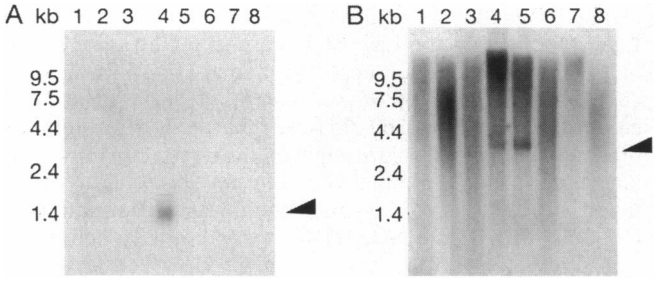

FIG. 5. Examples of Northern blot analysis of cosmid clones that specifically hybridized to the testis-specific subtracted probes. (*A*) TSPY cosmid. (*B*) Cosmid Y79-A8. The multiple tissue Northern blot contained $\approx$2 μg of poly(A)$^+$ RNA from human (1) spleen, (2) thymus, (3) prostate, (4) testis, (5) ovary, (6) small intestine, (7) colon, and (8) peripheral blood leukocyte. Arrowheads indicate the bands considered to be specific.

nique to generate a testis-specific cDNA library. cDNAs synthesized from human testis poly(A)$^+$ RNAs were subtracted against a mixture of cDNAs derived from poly(A)$^+$ RNAs of 10 different human tissues: heart, brain, placenta, lung, liver, skeletal muscle, kidney, spleen, thymus, and ovary. To examine the efficiency of subtraction, the mixtures of unsubtracted and subtracted testis-specific cDNAs were labeled and hybridized to Northern blots of multiple human tissues. The results are shown in Fig. 3. The unsubtracted testis cDNA probe (Fig. 3*A*) hybridized strongly to all RNA samples on the blot and possibly to the common or homologous species of mRNAs among the human tissues. The subtracted, testis-specific cDNA probe hybridized strongly to the testis RNA and very weakly to all other RNAs. Under similar conditions, previous studies suggested that for an individual cDNA to produce any signal in the Northern blot, its concentration needs to be at or greater than 0.1–0.3% of the cDNA mixture (19). Accordingly, our Northern blot analysis indicates that we achieved a high level of enrichment of testis-specific cDNA and, at the same time, a drastic reduction of highly abundant and/or common cDNAs.

The high efficiency of the testis-specific subtraction was subsequently confirmed after cloning and sequence analysis of selected testis-specific cDNAs. Ten randomly selected cDNA clones were used to probe human multiple tissue Northern blots. All 10 cDNA probes revealed unique mRNAs, which were expressed only in the testis. The autoradiograms of four of the resulting Northern blots are shown in Fig. 4. The exposure time ranged from 2 h to 2 weeks, suggesting testis-specific cDNAs of different abundance were represented.

The cDNA inserts from these 10 original clones and 52 additional clones (total of 62) were partially sequenced and analyzed for homology in the GenBank and EMBL data bases. A summary of the sequencing data are shown in Table 1. Four cDNA fragments were found to correspond exactly to known human testis-specific mRNAs. Four were found to have 75–85% homology with known mouse testis-specific mRNAs. These cDNAs probably represent the human homologues of the mouse genes. Two fragments had 96% homology with expressed sequence tags (ESTs) fragments for which information about tissue specificity was not available. We also found nine cDNA fragments that had at least 60% homology (over at least 60 nt) with other EST fragments. It is likely that they represent members of several novel gene families. One clone was identical to a mitochondria ribosomal RNA sequence that is expressed in all tissues. The remaining 32 clones demonstrated no significant matches with entries in the databases and potentially represent cDNAs from novel testis-specific genes. Of the 62 sequences, only 5 were redundant. These sequence analysis results suggest that the subtracted cDNAs are normalized and highly complex.

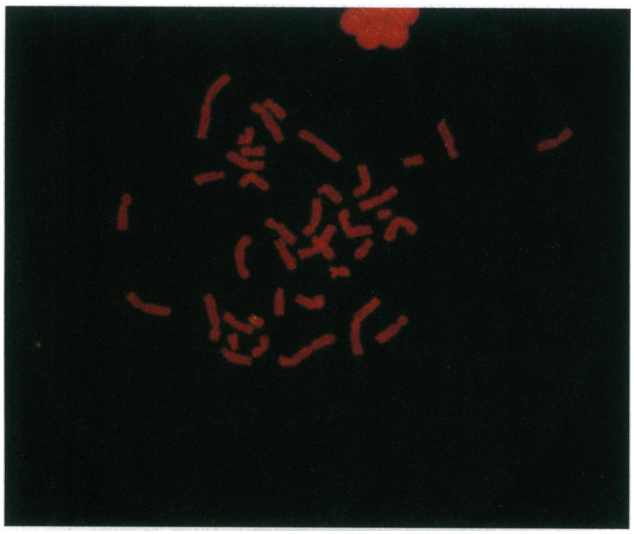

FIG. 6. Fluorescence *in situ* hybridization of human metaphase chromosomes using Y79-A8 cosmid DNA as probe. Positive signals were primarily located on band Yq211. Other human chromosomes also showed weak hybridizations, suggesting that homologous sequences might also be present in these portions of the human genome.

Identification of Tissue-Specific Genes from the Human Y Chromosome. Because the Northern blot analysis of the subtracted cDNA mixture showed very specific hybridization to the testis poly(A)$^+$ RNA, we suspected that this subtracted cDNA mixture could be used to screen a cosmid library constructed from flow-sorted human Y chromosomes to sucessfully identify functional sequences expressed in the testis. A total of 224 out of 3,072 cosmids showed specific hybridization. Of these, 25 were purified and tested for tissue-specific expression by Northern blot analysis. Eighteen cosmids showed positive hybridization with an abundant 1.3 kb mRNA (Fig. 5*A*, lane 4). The identical size of the mRNA suggested that these cosmids might contain the *TSPY* gene, which has at least 20–40 copies on the Y chromosome and is expressed only in testis (18). Subsequent Southern blot analysis of *Eco*RI-digested cosmids with a *TSPY* cDNA probe confirmed that all of these 18 cosmids did contain copies of the *TSPY* gene (data not shown). Seven other cosmids showed a smear of hybridization with all RNAs on the blots, suggesting the presence of repetitive sequences in these cosmid probes. Twelve additional non-*TSPY* cosmids were analyzed. Of these, four cosmids hybridized specifically with the testis RNA and were not related to *TSPY*. In particular, cosmid Y79-A8 hybridized with poly(A)$^+$ RNA from human testis and ovary (Fig. 5*B*, lanes 12 and 13, respectively). The hybridization of this cosmid to RNA from ovary suggests that a homologous gene (or sequences) might be present in other part of the human genome. Results of fluorescence *in situ* hybridization experiments showed that this cosmid is primarily located on band q211 of the Y chromosome (Fig. 6). There were additional hybridizations to other human chromosomes, confirming our postulation that homologous sequences are present elsewhere in the human genome. Further fluorescence *in situ* hybridization studies on the other three testis-specific cosmids also confirmed their locations on the human Y chromosome (data not shown).

DISCUSSION

The studies described here show that the SSH technique is simple and efficient for generating cDNAs highly enriched for differentially expressed genes of both high and low abundance. The high level of enrichment of rare transcripts has been achieved by the inclusion of a normalization step in the subtraction procedure. Based on the results of our reconstituted model, a 1,000- to 5,000-fold enrichment for rare (several molecules per cell) cDNAs can be achieved in a single round of SSH procedure. Successful normalization was clearly demonstrated by our Northern blot and sequence analyses of random cDNA clones from the subtracted testis cDNA library. All cDNA clones analyzed gave positive hybridization signals with the testis RNA in Northern blots and were present in different abundance. Fifty-seven of 62 clones had different sequences, and most of them represented unique and previously uncharacterized sequences. In addition, only one of 62 clones represent redundant sequences and it was derived from the mitochondrial rRNA.

Furthermore, our standard SSH procedure can be modified to increase the possibility of identifying quantitatively regulated transcripts between the tester and driver cDNA populations. Under our standard conditions, the driver cDNA would have eliminated most of the common sequences between the tester and driver cDNA samples during the first hybridization step. However, quantitatively different cDNA

species may still remain in the tester populations. To further eliminate common sequences, excess fresh driver cDNA needs to be added to the samples in the second hybridization step, thereby further subtracting quantitatively different but common sequences between the tester and driver populations. Hence, to retain the representation of the quantitatively different cDNAs in the final SSH products, the driver cDNA can be omitted in the second hybridization step, thereby allowing quantitatively regulated cDNAs in the tester samples to anneal and form hybrid e (Fig. 1) that are amplfiable in subsequent PCR.

SSH requires only one round of selective hybridization and PCR amplification of the target molecules, without any physical separation of ss and ds cDNAs. Inclusion of enhancer (e.g., PEG) in the hybridization solution further increases the efficiency of hybridization (19), and eliminates the usual steps of preamplification (7, 22) and/or cloning (23) of the cDNAs before the subtraction procedure. Using the protocol described in this study, we were able to perform cDNA subtraction starting with only 1–2 μg of poly(A)$^+$ RNA. The use of enhancer also allowed us to reduce the experimentation time to within 24 h.

One potential disadvantage of the SSH technique is the fact that under our standard procedure, a few micrograms of poly(A)$^+$ RNA from the two cell populations are needed. In some special cases, such quantity of RNAs may be difficult to obtain. To circumvent this problem, an amplification step for both the driver and tester cDNAs can be incorporated to generate sufficient quantities of both cDNA samples before initiating the SSH procedure. In such cases, separate adapter/primers will be ligated to the cDNA fragments and subsequently used for the PCR amplification (20). Nevertheless, we believe that avoiding the preamplification step is desirable because it may result in the loss of some sequences.

In our SSH protocol, we use a four-base recognition site restriction enzyme to digest cDNA into fragments. Of the many such restriction enzymes available, *Rsa*I was chosen because it generated the largest average size of fragments (≈600 bp). Although this may be a disadvantage when full-length cDNAs are desired, dividing each cDNA into multiple fragments has two important advantages. First, long DNA fragments may form complex networks that prevent the formation of appropriate hybrids, especially at high concentration required for efficient hybridization (our unpublished data). Second, cutting the cDNAs into small fragments provides better representation of individual genes. Derived from related but distinct members of gene families, cDNAs often have similar coding sequences that may cross-hybridize and be eliminated during the subtraction procedure (24). Further, different fragments from the same cDNA may vary considerably in terms of hybridization and amplification characteristics and may not hybridize or be amplified (7, 22). Thus, some fragments from differentially expressed cDNAs may be eliminated during the SSH procedure. However, other fragment from the same cDNA may be enriched and isolated. Once a small cDNA fragment is cloned and sequenced, numerous approaches, including several PCR based methods, can be used to quickly obtain corresponding full-length cDNAs (for review, see ref. 15).

The high level of enrichment, low background, and normalized abundance of cDNAs in the subtracted mixture make the method ideal for rapid cloning of cDNAs of differentially expressed genes. More importantly, the feasibility of using the uncloned subtracted cDNA mixture as a hybridization probe makes this technique versatile and powerful. As demonstrated in our studies, tissue-specific cDNA sequences from a particular chromosome can be identified using the subtracted testis-specific cDNA mixture as a probe in hybridization screening of a human Y chromosome cosmid library. Similarly, hybridization of subtracted tissue-specific cDNA probes to yeast artificial chromosome, bacterial artificial chromosome, or cosmid contigs could possibly be used to quickly identify potentially differentially regulated or tissue-specific genes. If a disease gene is mapped onto these DNA contigs, genes identified with subtracted tissue-specific cDNA probes can potentially serve as candidate genes involved in the corresponding disease(s), thereby enhancing many of the current positional cloning strategies.

We thank C.-C. Huang for the help in the analysis of sequencing data, M. Shen and S. Yu for excellent technical assistance, S. Colby for the critically reading the manuscript, and Theresa Provost for the helping in figure preparation. This work is partially supported by grants to Y.-F.C.L. from the National Institutes of Health and the Department of Veterans Affairs.

1. Duguin, J. L. & Dinauer, M. C. (1990) *Nucleic Acids Res.* **18,** 2789–2792.
2. Hara, E., Kato T., Nakada, S., Sekiya, L. & Oda, K. (1991) *Nucleic Acids Res.* **19,** 7097–7104.
3. Hendrick, S. M., Cohen, D. I., Nielsen, E. L. & Davis, M. M. (1984) *Nature (London)* **308,** 149–153.
4. Sargent, T. L. & Dawid, I. B. (1983) *Science* **222,** 135–139.
5. Davis, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2129–2194.
6. Lisitsyn, N., Lisitsyn, L. & Wigler, M. (1993) *Science* **259,** 946–951.
7. Hubank, M. & Schatz, D. G. (1994) *Nucleic Acids Res.* **22,** 5640–5648.
8. Liang, P. & Pardee, A. (1992) *Science* **257,** 967–970.
9. Welsh, J., Chada, K., Dalal, S. S., Ralph, D., Cheng, L. & McClelland, M. (1992) *Nucleic Acids Res.* **20,** 4965–4970.
10. Bauer, D., Warthoe, P., Rohde, L. & Struss, M. (1994) *PCR Methods and Applications* (Cold Spring Harbor Lab. Press, Plainview, NY), Supplement, pp. S97–S108.
11. Sompayac, L., Jane, S., Burn, T. C., Tenen, D. G. & Danna, K. J. (1995) *Nucleic Acids Res.* **23,** 4738–4739.
12. Bertioli, D. J., Schichter, U. H. A., Adams, M. J., Burrows, P. R., Steinbiss, H.-H. & Antoniw, J. F. (1995) *Nucleic Acids Res.* **23,** 4520–4523.
13. Welsh, J., Ralph, D. & McClelland, M. (1995) in *PCR Strategies,* eds. Innis, M. A., Gelfand, D. H. & Sninski, J. J., *et al.* (Academic, San Diego), pp. 249–276.
14. Siebert, P. D., Chenchik, A., Kellogg, D. E, Lukyanov, K. L. & Lukyanov, S. A. (1995) *Nucleic Acids Res.* **23,** 1087–1088.
15. Chenchik, A., Moqadam, L. & Siebert, P. D. (1996) in *A Laboratory Guide to RNA: Isolation, Analysis, and Synthesis,* ed. Krieg, P. (Wiley, New York), pp. 273–321.
16. Sambrook, J., Fritsh, E. L. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 8.46–8.49.
17. Van Dilla, M. L. & Deaven, L. L. (1990) *Cytometry* **11,** 208–218.
18. Zhang, J. S., Yang-Feng, T. L., Muller, U., Mohandas, T. K., de Jong, P. L. & Lau, Y.-F. C. (1992) *Hum. Mol. Genet.* **1,** 717–726.
19. Britten, R. L. & Davidson, E. H. (1985) in *Nucleic Acid Hybridization: A Practical Approach,* eds. Hames, B. D. & Higgins, S. J. (IRL, Oxford), pp. 3–15.
20. Gurskaya, N. G., Diatchenko, L., Chenchik, A., Siebert, P. D., Khaspekov, G. L., Lukyanpv, K. A., Vagner, L. L., Ermolaeva, O. D., Lukyanov, A. A. & Sverdlov, E. D. (1996) *Anal. Biochem.,* in press.
21. Lewin, B. (1985) *Genes* (Wiley, New York), pp. 230–236.
22. Wang, L. & Brown, D. D.(1991) *Proc. Natl. Acad. Sci. USA* **88,** 11505–11509.
23. Duguid, J. R., Rohwer, R. G. & Seed, B. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5738–5742.
24. Ko, M. S. H. (1990) *Nucleic Acids Res.* **18,** 5705–5711.