[CANCER RESEARCH 58, 5326–5328, December 1, 1998]

Advances in Brief

An Improved Method for Construction of Directionally Cloned cDNA Libraries from Microdissected Cells

Lisa A. Peterson, Monica R. Brown, Alex J. Carlisle, Elise C. Kohn, Lance A. Liotta, Michael R. Emmert-Buck, and David B. Krizman[1]

Laboratory of Pathology, National Cancer Institute, Bethesda, Maryland 20892

Abstract

Here, we developed an improved method for constructing microdissected cDNA libraries, based on strand-switching properties of reverse transcriptase, followed by PCR amplification with primers to mediate unidirectional insert cloning. Using RNA from microdissected ovarian carcinoma cells, we constructed a cDNA library consisting of 1.3×10^6 unidirectional recombinants with an average insert size of 500 bp. Single-pass sequencing of 100 clones with the T7 primer revealed 89 inserts derived from known genes, anonymous expressed sequence tags (ESTs), or novel sequences. Among these clones were known genes and ESTs previously found in cDNA libraries from bulk ovarian tissue RNA, sequences seen for the first time in an ovarian-derived library, and novel sequences not previously seen in any cDNA library. These results demonstrate a methodology for constructing quality cDNA libraries that are cloned in a unidirectional fashion, are complex and diverse, and reflect the tissue of origin.

Introduction

The multistep nature of tumorigenesis is reflected in progressive, pathologically defined cellular stages caused by aberrant patterns of gene and protein expression. Determination of these altered patterns of gene/protein expression may help define particular stages of tumorigenesis that may lead to improved diagnosis, identification of therapeutic targets, and better prognostic capabilities. The specific genetic alterations that drive the progression of a microscopic lesion to become a malignant tumor are difficult to determine due to the extreme amount of cellular heterogeneity that occurs in abnormally growing tissue. LCM[2] can be used to specifically procure microscopic amounts of homogeneous populations of cells (1). Thus, molecular genetic analysis of cells obtained by LCM can lead to powerful insights into genetic alterations that give rise to disease. We previously described a cDNA library from epithelial cells derived by microdissection of prostatic intraepithelial neoplasia (2). Results from this work indicate cDNA libraries from small numbers of microdissected cells can be constructed and are of value in establishing gene expression profiles within a histologically defined, homogeneous population of cells. Large-scale sequencing of thousands of individual clones from similar cDNA libraries can provide unique gene expression fingerprints. This forms the basis of the CGAP (3). To date, more than 15 CGAP cDNA libraries from microdissected tissue have been established and sequenced to various depths (http://www.ncbi.nlm.nih.gov/ncicgap/). However, inserts from these libraries were cloned bidirectionally, and because large-scale sequencing routinely relies on single-pass sequencing using the same primer for all clones, optimal information retrieval from current microdissected cDNA libraries is diminished.

To overcome this limitation, we applied an improved method in the construction of a cDNA library from laser capture-microdissected cells of a papillary serous ovarian carcinoma lesion. First strand cDNA was synthesized using a modified oligo(dT) primer, while addition of a second oligonucleotide served to direct the strand-switching event (4, 5). This strand switch methodology for construction of cDNA libraries is based on the properties of reverse transcriptase to add nucleotides on the 3′ end of a newly synthesized strand of cDNA. It has been shown that the human HIV reverse transcriptase preferentially adds purine nucleotides ($A > G \gg T \geq C$) when it reaches the end of the synthesizing strand of cDNA (4). Recently, it has been observed that the RnaseH-minus Superscript II reverse transcriptase enzyme exclusively adds three to four C residues on the 3′ end of first strand cDNA.[3] We used this property of reverse transcriptase to anchor a known sequence at the 3′ end of newly synthesized first strand cDNA and incorporated a PCR-based unidirectional UDG cloning approach for final library construction (6). This method will enhance high-throughput sequencing of microdissected cDNA libraries, thus increasing the efficiency of gene expression profiling of cells representing a distinct histological origin and tumorigenic stage.

Materials and Methods

Clinical Information, Microdissection, and Library Construction. The sample used for this analysis was obtained from the Cooperative Human Tissue Network and was received with all identifying information removed. According to Cooperative Human Tissue Network reports, this specimen was taken from a 74-year-old women with FIGO stage IV poorly differentiated papillary serous ovarian carcinoma. Ten-μm frozen sections were cut, mounted onto plain glass slides, and stained with H&E. Histological diagnosis was confirmed prior to microdissection. LCM was used to selectively procure a pure population of ovarian tumor cells, as described previously (1). Approximately 15,000 cells were dissected for RNA isolation and cDNA library synthesis.

This protocol was performed as outlined in Fig. 1. Total cellular RNA was obtained from ~15,000 microdissected cells and treated with DNase I, as described previously (1). The total amount of RNA obtained was not determined. A standard reverse transcription reaction was carried out using Superscript II reverse transcriptase (Life Technologies, Inc.) on total RNA in a volume of 10 μl with 500 ng of the modified oligo(dT)-anchored primer 3′AP [5′-GCGCGATCGCTCGACATCGATACGAC$(T)_{30}$VN-3′] for priming first strand cDNA and a 1 μM concentration of the strand switch primer 5′SWITCH (5′-TTCGGCTGCGAGAAGACGACAGAAGGG-3′). This reaction was carried out at 42°C for 30 min and stopped by placing it on ice.

A total of 5 μl of the cDNA were placed directly in a 100-μl volume PCR under standard conditions using 2.5 units of Taq polymerase (Boehringer Mannhiem) and a 0.5 μM concentration of each of the following amplification

Received 8/28/98; accepted 10/12/98.

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

[1] To whom requests for reprints should be addressed, at Laboratory of Pathology, National Cancer Institute, Advanced Technology Center, 134F, 8717 Grovemont Circle, Bethesda, MD 20892. Phone: (301) 435-5155; Fax: (301) 402-0043; dkrizman@helix.nih.gov.

[2] The abbreviations used are: LCM, laser capture microdissection; CGAP, Cancer Genome Anatomy Project; EST, expressed sequence tag.

[3] Clontech and Life Technologies, Inc., personal communication.

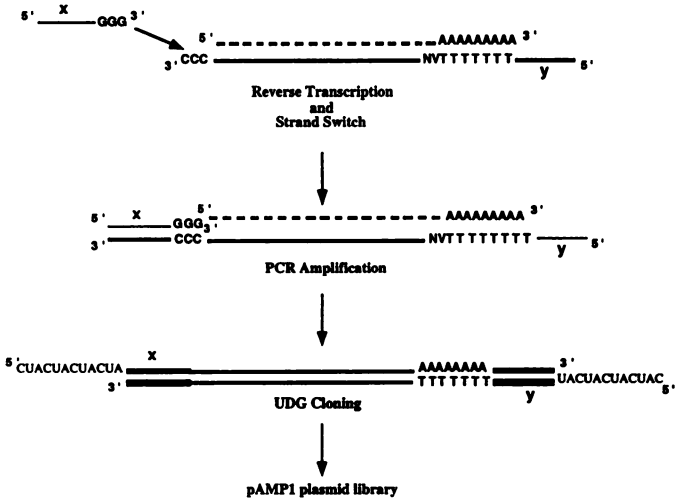

Fig. 1. Improved strategy for construction of unidirectionally cloned cDNA libraries from microdissected cells. mRNA is primed by a modified oligo(dT) primer, and first strand cDNA is synthesized until the 5′ end of each mRNA transcript is reached. The reverse transcriptase enzyme adds three to four C residues that are then used to extend the 3′ end of each cDNA strand beyond the 5′ end of each mRNA transcript in a sequence-dependent manner, mediated by the strand switch primer. PCR amplification of first strand cDNA is performed with two different primers specific for the *x* and *y* sequences that reside on either end of each cDNA molecule. The amplified library is subsequently cloned in the unidirectional UDG cloning vector pAMP1.

primers: 5′PCR (5′-CUACUACUACUATTCGGCTGCGAGAAGACGAC-AGAA-3′) and 3′PCR (5′-CAUCAUCAUCAUGATCGCTCGACATCGAT-ACGAC-3′). The triplet-repeat ends were designed to clone PCR product into the unidirectional UDG cloning vector pAMP1 (Life Technologies, Inc.). The cycling conditions consisted of the following steps: an initial 3-min denaturing step at 94°C; 1 min at 80°C, at which time the Taq was added (hot start procedure); 30 cycles at 94°C for 15 s, 68°C for 15 s, and 72°C for 3 min; and a final extension of 72°C for 5 min. The PCR product was size selected by passing through a CHROMA SPIN-400 column (cutoff of <400 bp) to remove small PCR product, unincorporated primers, primer/primer artifacts, nucleotides, Taq, and buffers according to manufacturer's recommendations (Clontech). The 100-μl volume flow-through was precipitated by adding 50 μl of a polyethylene glycol solution (30% PEG-8000–30 mM $MgCl_2$), centrifuged at 14,000 rpm for 20 min at room temperature, washed with 70% ethanol, and resuspended in 10 μl of H_2O. A total of 1 μl of PCR product was cloned into the vector pAMP1 according to manufacturer's recommendations to determine plating efficiency and total number of recombinants. The remaining 9 μl of PCR product were stored at −20°C for future scale-up of the library. The cloning reaction was transformed into DH10B chemical-competent bacteria, a fraction of the cells were plated on L-amp plates, and 100 clones were randomly picked for sequencing.

Library Analysis. For determination of cDNA insert size, 20 clones were picked at random and used in 50-μl PCRs with M13f and M13r primers to amplify inserts. A total of 5 μl from each reaction was analyzed on a 1.2% agarose gel, and product size was determined by comparison to a 100-bp ladder marker (Life Technologies, Inc.). Seqwright, Inc., (Houston, TX), performed sequencing under standard sequencing conditions using the T7 primer. Sequencing reactions were analyzed on an ABI 377 automatic sequencing apparatus. Individual clone sequences were compared to various sequence data bases, including GenBank and dbEST (7), by the BLAST program, accessed through the National Center for Biotechnology Information at the following http server: http://www.ncbi.nlm.nih.gov/ (8).

Results

Preliminary quality assessment of RNA from the lesion of papillary serous ovarian carcinoma tissue section was performed by RT-PCR using actin-specific primers (data not shown). Total RNA was obtained from ~15,000 cells by LCM, and a cDNA library was subsequently established. A tenth of the library PCR was analyzed on a 1.2% agarose gel and a homogeneous smear of PCR product ranging from 100 to 1500 bp in size was observed (data not shown). This PCR product was size selected by column purification and used in a cloning reaction to establish a library of 1.3×10^6 primary recombinants. PCR analysis determined an average insert size of 500 bp, with a range of 300 bp to 1.2 kb (data not shown). To further assess the quality of the library, single-pass sequencing of 100 clones was performed and the sequences analyzed by BLASTN with the combined characterization of all inserts shown in Table 1 (8).

Sequence analysis demonstrates that 90% (90 of 100) of the inserts derived from known genes (55), anonymous ESTs (30), novel transcripts (4), or expressed Alu repeats (1). All inserts were cloned in a unidirectional fashion and sequencing with the T7 primer revealed presence of a poly(A) stretch at the beginning of each sequence. In addition, there were no clones of mitochondrial, bacterial, or rRNA origin, and there were no clones without an insert. Also, no chimeric clones were found. However, 10% (10 of 100) of the clones contained inserts that were too short to contain analyzable sequence and were derived from very short (2–10 nucleotides) stretches of cDNA sequence. This library shows a sequencing success rate of 89%, as determined by dividing the number of clones derived from known genes, anonymous ESTs, or novel sequences by the total number of attempted sequences (89 of 100).

The 89 clones representing known genes, anonymous ESTs, and novel sequences were clustered according to the frequency with which each distinct sequence was found. A total of 44 distinct clusters were identified, giving a diversity factor of 49.4% (44 of 89). The most frequently represented transcripts were *ribosomal protein L14* (22.5%, 20 of 89), EST nz83d03.s1 (5.6%, 5 of 89), *ribosomal protein S15a* (4.5%, 4 of 89), and EST qi74b01.x1 (4.5%, 4 of 89). In addition, six clusters were identified three times each (3.4%), three clusters identified two times each (2.2%), and the remaining 32 clusters were singletons (1.1%). The transcripts identified by this sequencing effort are listed in Table 2.

Many known genes found in this cDNA library have been seen in other cDNA libraries made from bulk ovarian tissue, including *torsinA* (*DYT1*), *vacuolar H+ATPase proton channel1* (*ATP6C*), and *elongation factor 1-delta* (*EF-1 delta*). Moreover, other known genes have appeared for the first time in an ovarian cDNA library, including the *Treacher Collins syndrome* gene (*TCOF1*) and the *EBI3-associated protein p60* gene. Many anonymous ESTs seen here have also been seen in other ovarian libraries, whereas 13 anonymous ESTs found in this library have not been found in any other ovarian cDNA library, and 4 novel ESTs were found that have never been seen in any cDNA library.

Table 1 *Cumulative sequence analysis of 100 papillary serous ovarian cancer cDNA clones*[a]

Characterization	No. of clones (%)
Known genes	55 (55%)
Anonymous ESTs	30 (30%)
Novel ESTs	4 (4%)
Expressed Alu	1 (1%)
Mitochondrial	0
Ribosomal RNA	0
Bacterial sequence	0
Vector only	0
Short insert (<100 bp)	10 (10%)
Total	100

[a] Individual clone sequences were compared to sequence data bases by the BLAST program. Identification of each clone was made, and clones were categorized accordingly. Novel sequences showed no significant similarities to existing data base sequences; however, all four demonstrated weak similarities to ESTs from dbEST.

Table 2 *Known genes and anonymous ESTs identified by random sequencing of 100 clones from this cDNA library*[a]

Known genes	Anonymous ESTs	
EF-1 delta	AA639821	AI053970
KNP-1q	AA548598	AA262694
torsin A (DYT1)	AA501824	AI053431
ribosomal protein L38	AI053982	AI054362
ribosomal protein L14	AA732021	AI053635
ribosomal protein S19	AA738360	AA810861
ribosomal protein S10	AI053896	AI053633
ribosomal protein S9	AI053666	AI005356
ribosomal protein S15a	AA743577	AA470810
ribosomal protein L37a	AI053741	AI054420
IgG Fc receptor (*hFcRn*)	AA480341	AI053844
Integrin β-5 subunit	AA602572	AI053515
Treacher Collins syndrome (*TCOF1*)	AA805286	
EB13-associated protein p60		
Vacuolar ATPase proton channel (*ATP6C*)		

[a] The known genes are listed by name, whereas the anonymous ESTs are listed by GenBank accession number.

Discussion

High-throughput, large-scale EST sequencing projects aim to determine adequate minimal sequence from hundreds to thousands of library subclones to make gene identification possible. This is best accomplished by maintaining a rigid working environment with little to no variations in the methodology. Thus, all cDNA libraries used in this approach should have uniform clone propagation, DNA preparation capabilities, and sequencing potential. The most advantageous vector is one that grows to high copy number, is ampicillin resistant, and contains one or more of the following primer sites to mediate sequencing reactions: M13f, M13r, T3, or T7. In addition, inserts should be cloned in a unidirectional fashion so that sequencing all clones will yield 3′ ends of transcripts [poly(A) end] with a high degree of accuracy and consistency. The cDNA library reported here fits these criteria. The high copy vector pAMP1 is ampicillin resistant and, because inserts have been cloned unidirectionally, 3′ ends of all transcripts can be readily sequenced with the T7 primer.

Using the described methodology, we constructed a cDNA library from microdissected papillary serous ovarian carcinoma cells. To perform a preliminary quality assessment, we sequenced a small number of clones from this library. All analyzable clones derived from known genes, anonymous ESTs, novel ESTs, or expressed Alu repeats with no clones containing bacterial, rRNA, mitochondrial, or vector-only sequence. The absolute number of chimeric clones could not be determined because each clone was sequenced only once from a single direction; however, no chimeric clones were found with the sequence that we were able to analyze. Moreover, because the diversity of this library is high (49.4%), it can be favorably compared to the diversity of both microdissected cDNA libraries and bulk (nonmicrodissected) tissue cDNA libraries that have been constructed and previously sequenced within CGAP.

This methodology has been shown to result in an ovarian-derived cDNA library that is diverse, complex, and devoid of contaminating clones and that has been found to contain genes previously shown to be expressed in ovarian tissue. Moreover, this library has demonstrated the presence of anonymous ESTs never seen in any previous ovarian library, as well as novel sequences, indicating the potential for gene discovery. To our knowledge, this cDNA library is the first to directly reflect the gene expression profile of a homogeneous, tissue-derived population of papillary serous ovarian carcinoma cells, as well as being the first microdissected cDNA library to be cloned in a unidirectional manner. This advancement in methodology has increased the use of microdissected cDNA libraries and improved the ease with which they can be sequenced. Future efforts include construction of unidirectional cDNA libraries representative of all stages of ovarian cancer progression (normal, borderline invasive, invasive, and metastatic cancer) and large-scale sequencing of these libraries to establish a gene expression profile of ovarian cancer progression.

References

1. Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., Goldstein, S., Weiss, R. A., and Liotta, L. A. Laser capture microdissection. Science (Washington DC), *274:* 998–1001, 1996.
2. Krizman, D. B., Chuaqui, R. F., Meltzer, P. S., Trent, J. M., Duray, P. H., Linehan, W. M., Liotta, L. A., and Emmert-Buck, M. R. Construction of a representative cDNA library from prostatic intraepithelial neoplasia (PIN). Cancer Res., *56:* 5380–5383, 1996.
3. Strausberg, R. L., Dahl, C. A., and Klausner, R. D. New opportunities for uncovering the molecular basis of cancer. Nat. Genet., *16:* 515–516, 1997.
4. Peliska, J. A., and Benkovics, S. J. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science (Washington DC), *258:* 1112–1118, 1992.
5. Patel, P. H., and Preston, B. D. Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. Proc. Natl. Acad. Sci. USA, *91:* 549–553, 1994.
6. Rashtchian, A., Buchman, G. W., Schuster, D. M., and Berninger, M. S. Uracil DNA glycosylase-mediated cloning of polymerase chain reaction-amplified DNA: application to genomic and cDNA cloning. Anal. Biochem., *206:* 91–97, 1992.
7. Boguski, M. S., Lowe, T. M., and Tolstoshev, C. M. dbEST-database for "expressed sequence tags." Nat. Genet., *4:* 332–333, 1993.
8. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. J. Mol. Biol, *215:* 403–410, 1990.

An Improved Method for Construction of Directionally Cloned cDNA Libraries from Microdissected Cells

Lisa A. Peterson, Monica R. Brown, Alex J. Carlisle, et al.

Cancer Res 1998;58:5326-5328.