Direct selection of cDNAs using whole chromosomes

Sylvie Rouquier¹, Barbara J. Trask², Sylvie Taviaux¹, Ger van den Engh², Sylvie Diriong¹, Gregory G. Lennon³ and Dominique Giorgi¹,*

¹Centre de Recherche de Biochimie Macromoléculaire, CNRS UPR 9008, BP 5051, 1919 Route de mende, 34033 Montpellier cédex, France, ²Department of Molecular Biotechnology FJ-20, University of Washington, Seattle, WA 98195, USA and ³Lawrence Livermore National Laboratory, PO Box 808, L-452, Livermore, CA 94551, USA

Received July 10, 1995; Revised and Accepted September 25, 1995

ABSTRACT

We have developed a method for direct selection of cDNAs using whole chromosomes as target DNA. Double-strand cDNAs were synthesized from human fetal brain polyadenylated mRNAs. Flow-sorted chromosomes 17 and 19 were amplified by degenerate oligonucleotide primed polymerase chain reaction (DOP-PCR) and used to capture ds cDNAs by an improved magnetic bead capture protocol. To demonstrate the capabilities of this method, the selected cDNAs were used as probes in FISH experiments. The selected cDNA populations specifically painted chromosomes 17 or 19 on metaphase spreads. These results demonstrate that it is possible to do chromosome painting using cDNA probes and that this method is a means to rapidly select expressed sequences encoded by any portion of the genome.

INTRODUCTION

Rapid identification and isolation of transcribed sequences from defined chromosomal regions is the rate-limiting step in positionally cloning genes responsible for inherited disorders. Conventional strategies for finding transcripts include hybridization of genomic fragments directly to cDNA libraries or Northern blots (1,2), isolation of sequences conserved between divergent species (3) and identification of CpG islands as potential markers of transcriptional units (4). These methods are labor intensive when applied to isolate single experiment, complex populations of cDNAs specific for each chromosome. Our results indicate that this technique is a useful tool to rapidly select genes present on each chromosome and to facilitate the construction of an expression map of the human genome.

MATERIALS AND METHODS

cDNA preparation

Double-stranded (ds) cDNA was synthesized from 200 ng of polyadenylated human fetal brain (HFB) RNA (Clontech, Palo Alto, CA) using the Uni-Amp Plus kit (Clontech) with a combination of oligo dT and random priming and then blunt-ended with T4 DNA polymerase, following the procedure recommended by the manufacturer. A Uni-Amp adapter designed for subsequent PCR was then ligated to both ends of 2 ng of the cDNA molecules in a 10 µl reaction. The Uni-Amp primer was used to amplify 1/4 of the ligation reaction by PCR. One-tenth of this first amplification was submitted to a second PCR reaction with the same primer. Ten secondary PCR reactions were pooled, extracted once with phenol/chloroform, once with chloroform, ethanol precipitated and resuspended in 200 µl of TE. The DNA concentration was determined by OD260 measurement.

Flow-sorted chromosome preparation

An immortalized human lymphoblast cell line (KH18 obtained from Gary Silverman, Harvard) was the source of flow-sorted

* To whom correspondence should be addressed
chromosome 17 and 19 as the genomic targets for the cDNA selection. Chromosomes were isolated using the polyclone procedure (18), stained with chromomycin and Hoechst 33528 and analyzed in a flow cytometer (MoFlo, Gier van den Engh, unpublished). 500 chromosomes of each type were flow-sorted in 0.2 ml PCR tubes containing 30 μl sterile water.

**DOP1.** Flow-sorted chromosomes were frozen, thawed and then subjected to DOP-PCR amplification with the primer 6MW as described elsewhere (19,20) using a Cetus DNA thermal cycler 9600 (Perkin Elmer). The primary cycling conditions consisted of an initial denaturation for 5 min at 95°C followed by five low annealing temperature cycles of 94°C for 30 s, 30°C for 60 s, a ramp to 72°C over 120 s and 72°C for 120 s. These steps were followed immediately by 30 high annealing temperature cycles of 94°C for 30 s, 62°C for 45 s and 72°C for 120 s. In the final cycle, this last step was extended to 10 min.

**DOP2.** One micro liter of the primary (DOP1) PCR product was submitted to a secondary DOP-PCR reaction using only the 30 high annealing temperature cycles. Five microliters of each 50 μl PCR reaction was run on a 0.5x TBE–1% agarose gel to check for PCR product.

One microgram of the DOP2 amplified chromosomal DNA was biotinylated using a nick-translation kit (BioNick, GIBCO/BRL) following the procedure recommended by the manufacturer. Biotinylated DNA was purified from unincorporated biotin using a Sephadex. G-50 (Pharmacia) spin column followed by ethanol precipitation.

cDNA selection

One microgram of HFB cDNA was prehybridized with 5 μg of human Cot-1 DNA (GIBCO/BRL) in 100 μl of hybridization buffer (6x SSC, 0.1% SDS) at 65°C for 16 h, to block human repetitive sequences within the cDNA. 100 ng of the DOP2 amplified and biotinylated chromosomal DNA was added to the solution and the hybridization was continued at 65°C for 24 h.

After hybridization, hybrids were immobilized on streptavidin-coated magnetic beads (Dynabeads M-280, Dynal Inc). Non-hybridized cDNA was removed by washing the beads 3-fold in 0.1x SSC, 0.1% SDS at room temperature, followed by three washes in 0.1x SSC, 0.1% SDS at 65°C. The bound cDNA was eluted by incubating the beads for 20 min in 50 μl of 50 mM NaOH at room temperature and then neutralized with 50 μl of 1 M Tris at pH 7.5. The eluate was then desalted by chromatography over a Sephadex G-50 spin column and amplified by PCR. Approximately 1 μg of amplified, eluted cDNA was submitted to a secondary cycle of enrichment by recycling through the above process, including the prehybridization step. After the second round of enrichment, the eluted cDNAs were amplified again by PCR, eluted and desalted as above. All the PCR reactions were performed with Uni-Amp primer for 35 cycles, following the conditions specified by the manufacturer (Clontech).

Fluorescence in situ hybridization (FISH)

Probes for FISH (chromosome painting) were prepared by using 5 μl of DOP1 amplified DNA in a DOP2 reaction as described above except that the dTTP concentration was lowered to 80 and 200 μM of biotin-16-dUTP (Boehringer Mannheim) was added in a total reaction volume of 50 μl. After purification on Sephadex G-50 columns, 100 ng of each probe were precipitated with 10 μg of Cot1 DNA to suppress repetitive sequences and used for hybridization on metaphase chromosomes spreads as described (21).

The double selected cDNAs were labeled with biotin by subjecting 5 μl of the products to a PCR amplification with the Uni-Amp adapter, using biotin-16 dUTP and reduced dTTP as described above. 500 ng of probe was precipitated with 10 μg of Cot1 DNA as described and used for hybridization on metaphase chromosomes spreads.

The sites of hybridized biotinylated probes were subsequently labeled with fluorescein (FITC) (22). The chromosomes were counterstained with propidium iodide (PI).

Images were collected using a Leica confocal laser scanning microscope equipped with a krypton-argon ion laser with a major emission line at 488 nm for FITC excitation and 568 nm for PI excitation. Planapochromat lenses (40x) were used and the untreated images were directly transferred from the VME bus of a Leica Motorola 68040 to a Silicon Graphics (Mountain View, CA) IRIS Indigo workstation (R3000). Images were deconvoluted, gamma mapped and converted to SGI raster format by using Convert software (23; N.J.C. Lamb, unpublished procedure). For gamma mapping, significant differences in pixel intensity between images obtained by the CCD camera and those obtained by the CLSM were automatically corrected by assigning the maximal and minimal image intensity range and fitting images data within this range. Under such circumstances, grey image backgrounds from the CCD camera were reassigned to black. Figures were merged with ADD2 (N.J.C. Lamb, unpublished procedure), assembled with SGI Showcase 2.01 and printed directly as postscript files by using a Kodak Colorse thermal sublimation printer.

**RESULTS**

**Direct selection of cDNAs**

Figure 1 summarizes the principle of the method: 500 chromosomes 17 or 19 were flow-sorted directly into 0.2 ml PCR tubes. The chromosomes were amplified by DOP-PCR as previously described (19,20). Aliquots of this first reaction were in turn used in a secondary PCR to generate more material. Typical profiles of PCR products after one and two rounds of amplification were obtained (19,24). The products appeared on a 1% agarose gel as a smear with a large size distribution centered around 600 bp (not shown). One microgram of this DNA template was biotinylated by nick-translation for subsequent coupling to streptavidin-coated magnetic beads.

We generated ds HFB cDNAs from HFB polyadenylated mRNAs. A universal adapter was ligated on both sides of the ds cDNAs. This ligation step allowed us to PCR amplify the ds cDNAs at each step using a single primer. We used this property to generate the amount of cDNA needed for the capture experiments. The yield of ds cDNAs synthesized from polyadenylated mRNAs was very high. We obtained >50 μg of ds cDNA using the conditions described in Materials and Methods. This primer did not amplify any sequence from human genomic DNA in the same PCR conditions (data not shown). This property allowed us to rule out possible genomic DNA contamination, which could occur during the selection process. Because ~10% of cDNA molecules contain Ala repeats or other repetitive sequences (15), we blocked repetitive sequences using Cot1 DNA before hybridizing DOP amplified chromosomes and ds
Double strand cDNA synthesis from Human Fetal Brain polyA+ mRNA

Flow-sorted chromosomes

Prehybridization in eppendorf tube
-blocking of repeated sequences

Flow-sorting

First round of selection

DOP-PCR amplification

Chromosome painting

65°C

Uotinylation

Chromosome painting

Hybridization

65°C

cDNA capture using
magnetic beads coated with streptavidin

high stringency washes

cDNA elution

PCR amplification

Second round of selection

cDNA chromosome painting

Figure 1. Outline of the cDNA direct-selection and cDNA-painting procedures. First round of selection: double-strand cDNAs were synthesized from human fetal brain mRNAs and an amplification cassette was ligated on both sides (hatched squares). Repetitive sequences were blocked by prehybridization with Cot1 DNA. Flow-sorted chromosomes were DOP-PCR amplified and biotinylated. cDNAs and biotinylated chromosomes were hybridized in solution. The resulting hybrids were captured on streptavidin-coated magnetic beads. During the following steps a magnet was used. High stringency washes were performed to remove non-specific cDNAs. The retained cDNAs were then eluted by alkaline treatment, PCR amplified using the amplification cassettes and then recycled through a second cycle of selection using the same protocol. The selected cDNAs issued from the second cycle were then biotinylated by PCR and used for FISH experiments on metaphase spreads.

cDNAs to each other. The size distribution of the selected cDNAs after one and two rounds of direct selection ranged from 200 bp to 1.0 kb with an average size of 400-500 bp (Fig. 2). We suspect that the short sizes are due to a combination of cDNA synthesis protocol and PCR amplifications. After the second round of selection (Fig. 2, lanes 2 and 4), the background is reduced and more particular bands are visible within the smear, indicating a further enrichment as described by others (16,17).

Characterization of DOP-amplified flow-sorted chromosomal material

To ascertain the success of the chromosome amplifications by DOP-PCR, we hybridized biotinylated DOP-amplified chromosomes 17 and 19 to metaphase chromosomes. The results are shown in Figure 3A and B. Each probe hybridizes specifically and strongly the starting chromosome. There is no detectable background or cross-hybridization to other chromosomes, demonstrating that the chromosome-17 and -19 sorts were not contaminated by other chromosomes and that repetitive sequences were suppressed by the blocking step. The chromosomes are densely and completely painted, indicating that amplification by DOP-PCR provides material representing a high coverage of the starting DNA.

Chromosome painting with selected cDNA

We used a one-step strategy to demonstrate the specificity and complexity of cDNAs selected by the capture process. The cDNAs eluted after two rounds of selection were biotin-labeled by PCR and hybridized to metaphase spreads. As shown in Figure 3C and D, the populations of selected cDNA molecules hybridize exclusively to the chromosomes used in each hybrid-selection experiment. cDNAs selected with DOP-amplified chromosome 17 hybridize to chromosome 17 and cDNAs selected with DOP-amplified chromosome 19 to chromosome 19. cDNAs label
The human genome contains between 60,000 and 100,000 genes, complement the genetic linkage map of the human chromosomes. This map would enable us to study their functions and possible implication in pathology. In this study, we report a strategy that combines two previously described methods (direct selection of cDNAs and whole amplification of chromosomes by DOP-PCR) to isolate a complex mixture of cDNAs from two human chromosomes. During the course of this work, related but different approaches have been described, which make use of either DOP-amplified microdissected homogeneously staining regions immobilized on filters or target DNA to select genes present in high copy in an aberrant portion of a chromosome (29) or the so-called 'PrepISH' technique using a combination of in situ hybridization and chromosome microdissection (30).

Direct selection has proved to be an effective method for isolating genes from several large genomic regions in other studies (10,11,13,14). Usually, region-specific expressed sequences are isolated by hybridizing a cDNA population to a cloned genomic target immobilized on filters or streptavidin-coated magnetic beads (15). Our procedure employs direct selection using whole chromosomes, but avoids the cloning step. This procedure is made possible by advances in universal whole-genome amplification by PCR. These universal amplification procedures have made it possible to generate whole-chromosome or region-specific paints from a few copies of chromosomes isolated by flow sorting or microdissection. The paints have been used in FISH analysis of normal and aberrant chromosomes (29,31).

We chose chromosomes 17 and 19 as our test subjects for several reasons: (i) they are small (85 and 62 megabases, respectively); (ii) they are nonacrocentric and thus lack ribosomal genes; (iii) they have a high GC content thought to reflect a high gene content (32,33). We chose human fetal brain tissue as the starting cDNA population, because it has been estimated to express ~30,000 different genes (34).

A double-stranded population of cDNA molecules were prepared from polyA+ human fetal brain mRNA using conventional procedures. We ligated ds cDNA to Uni-Amp adapter and amplified these DNA products using the adapter as primer sequence. Using this cDNA population for the selection overcomes the problem of the limited complexity of amplified cDNA libraries (14). The possibility of biased PCR amplification of complex ds cDNA populations using Uni-Amp primers was tested by Hozier et al. (30). The representation of cDNAs in the population was not significantly altered by the PCR, with the possible exception of a bias toward smaller molecules. In our study, the cDNA molecules captured after two rounds of selection/amplification with flow-sorted amplified chromosomes are not full-length and rather small (400-500 bp), probably due to this PCR bias for smaller molecules.

We used a simple method to assay the specificity and complexity of the selected cDNAs by hybridizing them back to metaphase chromosomes. When the selected cDNAs are hybridized by FISH to metaphase chromosomes, they label specifically the chromosome that was used for the capture. Chromosomes 17 and 19 are known to be highly rich in R bands. Previous studies have shown that genes are predominantly located in R bands (32,33). In these studies using FISH and chromosome in situ suppression hybridization (CIS), the authors demonstrated that chromosomes densely from end to end illustrating that selected material is a complex mixture of cDNAs mapping throughout the chromosomes. In Figure 3C the two chromosomes 17 are not evenly painted. Indeed, the chromosome 17 at top of the picture is less densely painted with one of its short arm chromatids appearing weakly labeled which is due to the optical section used in the confocal analysis. The most striking feature on this chromosome is that both chromatids of the long arm are labeled at the telomeres and at their proximal moieties, which is visualized as an absence of painting in the middle of the long arm. No other chromosome is consistently labeled. However some nonspecific signals are visible that may result from the relatively large size of the hybridized fragments (25). Discrete signals that could result from a genuine hybridization are also seen; for example on both chromosomes 2 (Fig. 3C) and both chromosomes 1 (Fig. 3D). Nevertheless these same signals are not observed significantly on other metaphase spreads (data not shown). As a control experiment, if the Uni-Amp primer is omitted in the PCR reaction to generate FISH probes of selected cDNAs, neither product nor FISH signals were obtained. These controls indicate that the observed FISH signal stems from cDNAs and not genomic contamination.

**DISCUSSION**

The human genome contains between 60,000 and 100,000 genes, which as mRNA represent 2–3% of the total sequence information (26). Approximately 3000–3500 genes have been identified and cloned to date, of which ~2500 have been mapped to human chromosomes (15,27). One of the ultimate goals of the human genome project is to identify, sequence and map all the genes and to study their functions and possible implication in pathology. In order to achieve this goal, we want to be able to make a transcription map of each chromosome. This map would complement the genetic linkage map of the human chromosomes (28) and thereby facilitate the rapid isolation of candidates genes for linked genetic diseases.

Different new approaches to select for genes present in a given portion of the genome have been developed in recent years. However, these procedures have in common, to use as genomic DNA source preisolated DNA clones, usually either YACs or cosmids or somatic cell hybrid DNA. There is thus a need for methods that can rapidly select cDNAs from entire chromosomes.
3. FISH on metaphase chromosomes of DOP-PCR amplified chromosome 17 (A) or chromosome 19 (B). FISH of a population of cDNAs obtained by two of direct selection using chromosome 17 (C) or chromosome 19 (D) as target DNA. The probes were labelled with biotin and detected with Avidin-FTTC (green). Chromosomes were counterstained with propidium iodide (red). Images were collected by confocal laser scanning microscopy (Leica) using an ion laser with a emission line at 488 nm for FTTC and 568 nm for PI excitation.

...slands are concentrated in these bands. They also reported most of the CpG islands are located in DNA fractions where island distances are within a range of 15–500 kb: the ty being <100 kb. Using biotinylated CpG probes in FISH ments, they showed that chromosomes 19 and 17 were etely painted, except the long arm of chromosome 17 ting a lack of hybridization in its middle, which reflects a vel of CpG islands in this region (33). Interestingly, due to her labeling, we observed clearly the same picture with one chromosomes 17 (on top of Fig. 3C) hybridized with our selection probe. This result is in good agreement with the y of CpG islands in this region (32,33) and could confirm gene number in this portion of the chromosome. Due to mplexity of our cDNA probe and the limit of resolution of metaphase chromosomes (~1 megabase to discriminate different probes (35)), it is not abnormal to observe complete ng and no banding pattern despite the preferential gene ution in R bands. Discrete signals are also seen on both homologues of other chromosomes (Fig. 3C and D). These same signals are not observed accurately on other metaphase spreads suggesting that they result from background hybridizations. Nevertheless, due to the weak intensity of hybridization signals of individual cDNAs, some of these signals could correspond to genuine hybridizations which could reflect the presence of gene families related to sequences contained in the source chromosome. In such cases the method could be interestingly applied to trap related genes encoded by different chromosomes, but this raises the question of the stringency conditions used in the selection process. Indeed, our method contains no specific means for controlling perfectly the degree of sequence match necessary for common elements to be recovered. In this direction more stringent conditions have been used in the end-ligation coincident sequence cloning cDNA selection method (12). This approach is more efficient at excluding less perfectly matched material. After the cDNA–DNA hybrids are formed, oligonucleotides complementary to the genomic DNA linkers are ligated to the cDNA...
strand. Thus, this step increases the specificity of the capture by permitting to recover only hybrids that are, in addition, perfectly base-paired at their ends. This is however not compatible with DOP-PCR since it needs defined sequence linkers to amplify the genomic material. Although the annealing and stringent wash conditions used in our method restrict the degree of mismatch in a tolerable way, it would be very useful for specific purposes to design conditions monitoring the level of pairing between cDNAs and target DNA.

Our method, while not pretending to select all the genes on a given chromosome, does rapidly yield a large number of genes on the chromosome. It is impossible to define precisely the complexity of the mixture of selected cDNAs. However, the cDNAs do label chromosomes specifically and from end to end, indicating a relatively high degree of complexity. Based on the FISH results, we estimate that in each experiment we isolated a mixture of several hundred cDNAs. Due to the mode of cDNA synthesis (oligo DT and random priming), different molecules representing different or overlapping parts of individual mRNAs may have been selected. The use of cDNA from other tissues should add to the collection of genes for a chosen chromosome and at the same time provide information on the time or place of expression. Thus, this cDNA-selection approach should be very useful in the development of a transcription map for each human chromosome.

More focused investigations on the individual selected cDNAs could be done by different approaches: (i) by sub-cloning and sequencing the selected cDNAs; (ii) by analysis of the corresponding genomic sequence to determine if the cDNA was selected by a true gene or a pseudogene present on the target chromosome; (iii) by using this complex mixture of cDNAs to probe a normalized, cDNA library spotted on high density filters (36). This last method should allow direct access to full-length cDNA clones and avoid redundant sequencing of cDNAs.

Finally, this cDNA selection could be used to isolate candidate disease-genes from large chromosomal regions (several megabases) such as those defined by linkage analysis. By generating target DNA sequences from DOP-amplified microdissected chromosomes using this quick and simple method, one could select a complex mixture of region-specific genes without the need for YAC or cosmid contigs spanning the region. The identified cDNA spanning the region could always be placed on the available YAC or cosmid maps by hybridization or PCR.

In summary, we describe here a new procedure for quickly isolating chromosome-specific cDNAs. This technique should facilitate the establishment of transcription maps for each human chromosome and the search for candidate disease-genes.

**ACKNOWLEDGEMENTS**

We thank Drs Anne Fernandez, Ned Lamb and Nicole Lautredou for confocal microscopy analysis. We thank Prof. Jacques Demaille for his support and his constant interest to the work. S.R. is supported by a fellowship from the Association pour la Recherche contre le Cancer (ARC). B.T. and G.v.d.E. are supported by US DOE grants DE-FG06-93ER61553 and DE-FG06-93ER61662. G.L. is supported under the auspices of the DOE by LLNL under contract no. W-7405-Eng-48.

**REFERENCES**