Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 9628–9632, November 1991
Genetics

Direct selection: A method for the isolation of cDNAs encoded by large genomic regions

(erythropoietin/guanine nucleotide-binding regulatory proteins/chromosome 7)

MICHAEL LOVETT*†, JUHA KERE‡, AND LINDA M. HINTON*

*Department of Molecular Genetics, Genelabs Inc., 505 Penobscot Drive, Redwood City, CA 94063; and ‡Department of Molecular Microbiology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110

Communicated by Sherman M. Weissman, July 24, 1991

ABSTRACT **We have developed a strategy for the rapid enrichment and identification of cDNAs encoded by large genomic regions. The basis of this "direct selection" scheme is the hybridization of an entire library of cDNAs to an immobilized genomic clone. Nonspecific hybrids are eliminated and selected cDNAs are eluted. These molecules are then amplified and are either cloned or subjected to further selection/amplification cycles. This scheme was tested using a 550-kilobase yeast artificial chromosome clone that contains the *EPO* gene. Using this clone and a fetal kidney cDNA library, we have achieved a 1000-fold enrichment of EPO cDNAs in one cycle of enrichment. More significantly, we have further investigated one of the "anonymous" cDNAs that was selectively enriched. We confirmed that this cDNA was encoded by the yeast artificial chromosome. Its frequency in the starting library was 1 in 1×10^5 cDNAs and after selection comprised 2% of the selected library. DNA sequence analysis of this cDNA and of the yeast artificial chromosome clone revealed that this gene encodes the β_2 subunit of the human guanine nucleotide-binding regulatory proteins. Restriction mapping and hybridization data position this gene (GNB2) to within 30–70 kilobases of the *EPO* gene. The selective isolation and mapping of *GNB2* confirms the feasibility of this direct selection strategy and suggests that it will be useful for the rapid isolation of cDNAs, including disease-related genes, across extensive portions of the human genome.**

The rapid identification of coding sequences within large genomic regions would considerably accelerate the isolation of clinically significant genes and aid in the construction of a human gene map. Unfortunately, by current methodologies, the task of identifying coding sequences is significantly difficult when the region of interest is 20–40 kilobases (kb) and becomes almost intractable when the region is several hundred kilobases long. Several techniques have been employed to address this problem. These include the use of CpG islands (1–3) as potential sign posts for the ends of some transcription units, the use of "zoo" blots to detect cross-species conservation of genomic sequences (4, 5), and the hybridization of radiolabeled cDNAs to arrayed genomic clones (6, 7). More recently, a number of schemes have been described that are based on the PCR (8) and are targeted at the selective amplification of cDNAs that contain human sequences from somatic cell hybrids (9). Exon trapping is another recently described and technically sophisticated scheme that is also targeted at the capture of expressed sequences from fairly large genomic regions (10, 11). Unfortunately, many of these schemes are time consuming and technically complex and in some cases also suffer from being low in sensitivity (for a review, see ref. 12).

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.

One straightforward approach to detecting coding regions within a large genomic region would be the use of a purified yeast artificial chromosome (YAC) DNA (13) to directly screen a cDNA library. However, this approach is time-consuming, labor-intensive, and only occasionally successful (14, 15). Prior to screening with a YAC, the repeated sequence elements within it must be suppressed by some form of blocking (15–17). This, coupled with the relatively high sequence complexity of the YAC, results in poor signal-to-noise ratios. Even when successful, this type of screen shows great variability, detecting only some of the true positives. In one study (15), which echoes our experience, a 180-kb YAC containing the human aldose reductase gene was used to directly screen a cDNA library in which the target cDNA was moderately abundant (1 in 10,000 clones). Although this screen succeeded in detecting aldose reductase cDNAs, it only detected ≈10% of those detected by an aldose reductase cDNA probe.

In light of this we have developed a selection scheme for the enrichment of cDNAs that are encoded by large genomic regions. We reasoned that a library of these enriched cDNAs could be either directly analyzed for true positives or screened by hybridization with the YAC DNA, with a much higher certainty of success. In this report we describe the application of this direct selection scheme to the isolation and identification of cDNAs encoded by a 550-kb YAC from human chromosome 7.

MATERIALS AND METHODS

Preparation of YAC DNAs. Yeast chromosomes were prepared and electrophoresed on a contour-clamped homogeneous electric field gel (18). The EPO YAC (550 kb) was excised and the DNA was digested *in situ* with *Mbo* I prior to extraction using Geneclean II (Bio 101, La Jolla, CA). Approximately 100 ng of YAC DNA was incubated in 0.5 M NaOH for 10 min at room temperature and neutralized with 1 M Tris·HCl (pH 7.5), and SSC was added to 5× (20× SSC is 3 M NaCl/0.3 M sodium citrate, pH 7.0). The DNA was applied to a GeneScreen filter (DuPont) in a dot-blot apparatus. The filter was washed with 10× SSC, UV-crosslinked, baked for 1 hr at 80°C, and prehybridized (19). For genomic clones from the regions surrounding the interleukin 3 (*IL3*) and granulocyte/macrophage colony-stimulating factor (GM-CSF) genes, ≈10 ng of the purified DNAs was bound to GeneScreen filter dots.

Preparation of cDNA. To prepare cDNA inserts, a 5-μl sample (5×10^8 independent clones from a library of 1×10^6 recombinants) of a commercially available oligo (dT)-primed fetal kidney cDNA library (Clontech) was subjected to 30

Abbreviations: YAC, yeast artificial chromosome; GM-CSF, granulocyte/macrophage colony-stimulating factor; PCR, polymerase chain reaction.

†To whom reprint requests should be addressed.

cycles of PCR with 10F and 10R vector primers. Approximately 1 μg of cDNA inserts was prepared from the entire library and the average length of this cDNA was ≈1 kb; 1 μg is thus roughly equivalent to 1×10^{12} molecules. PCR products were gel-purified and were electroeluted onto NA45 membranes (Schleicher & Schuell). cDNAs were then ethanol-precipitated, resuspended in 100 μl of 10 mM Tris·HCl, pH 8/1 mM EDTA, and desalted on a Sephadex G-50 spun column (20). cDNAs were preblocked using 100 μg of sheared total human DNA [average length, 400 base pairs (bp)]. In some cases, 1 μg of pBR322 DNA and 1 μg of yeast DNA were included. The cDNAs and genomic DNAs were mixed, boiled for 5 min, quenched in iced water for 5 min, and hybridized in a total volume of 1 ml for 90 min at 65°C. For IL3/GM-CSF phage and cosmid selections, we utilized an oligo (dT)-primed cDNA library derived from phorbol 12-myristate 13-acetate-activated T cells. Inserts were prepared in bulk (1 μg) by conventional methods and were reduced in size by digestion with a frequently cutting restriction enzyme. These small fragments were then ligated to oligonucleotide amplification linkers (21).

Hybridization, Washing, and Elution. Hybridization was conducted at 65°C for 24–48 hr in 1 ml of standard solution (19) containing cDNA at 1 μg/ml and ≈100 ng of genomic DNA on the filter (2×10^8 molecules for a 500-kb YAC). Post-hybridization washes were 1× SSC/0.1% SDS for 30 min at room temperature, 0.1× SSC/0.1% SDS for 30 min at room temperature, and 0.1× SSC/0.1% SDS, four changes of solution for a total of 6 hr at 65°C. The bound cDNAs were eluted from the filter in 50 μl of 50 mM NaOH for 15 min at room temperature. The solution was neutralized with an equal volume of 1 M Tris·HCl (pH 7.5) and was applied to a Sephadex G-50 spun column. Samples of this DNA were used directly in a PCR.

Amplification of the Eluted cDNAs. PCRs were conducted with the 10F and 10R primers used for initial amplification. In other cases we have used primer sets that were nested within the ends of the synthetic linkers. The eluted material (10 μl) was used in a 100-μl PCR mixture.

PCRs and Primers. All primers were synthesized in an ABI model 380B DNA synthesizer. The DNA sequences of the various primers were as follows: 10R vector primer for λgt10, GAGTATTTCTTCCAGGGTA; 10F, GCAAGTTCAGCCTGGTTAAG; T3 ZAPII primer, ATTAACCCTCACTAAAG; T7, AATACGACTCACTAAAG; 5′ EPO cDNA primer, CTCATCTGTGACAGCCGAGTC; 3′ EPO, CTCCATCCTCTTCCAGGCATAG; GNB2 5′ coding region primer (GB21.1), GGAAGCTCATCATCTCTGGGAC; GNB2 5′ reverse primer (GB21.2), GGAGCAGATGTTGTCCAACC; GNB2 3′ noncoding region primer (GB2.1), AAGATCTGGAACTAATGGCC; GNB2 3′ reverse primer (GB2.2), CCTCCGGCCCTCCAGAGGGG. PCRs were conducted using a Perkin–Elmer/Cetus PCR kit and a Perkin–Elmer thermocycler. The conditions for all PCRs involved a 1-min denaturation at 94°C, annealing for 1 min, and extension for 1 min at 72°C. The annealing conditions for primers were for 10F/10R, 55°C; for T3/T7, 37°C; for 5′ EPO/3′ EPO, 57°C; for GB21.1/GB21.2, 55°C; and for GB2.1/GB2.2, 55°C. Reactions were continued for 30 cycles.

Cloning the PCR Products. PCR products were digested with proteinase K for 1 hr at 50°C, phenol/chloroform-extracted, ethanol-precipitated, and digested with *Eco*RI. DNAs were gel-purified onto NA45 membranes, eluted, precipitated, resuspended, and cloned into the *Eco*RI site of λZAPII (Stratagene).

Screening with Radiolabeled YAC DNA. The purified YAC DNA was radiolabeled using a random-priming kit (Boehringer Mannheim) to a specific activity of 5×10^8 cpm/μg and was preblocked with 1 mg of sheared total human DNA (average length, 400 bp) plus 10 μg of pBR322 DNA, and in some hybridizations, 10 μg of sheared total yeast DNA (average length, 400 bp). Cot1 DNA (GIBCO/BRL) has also been substituted for total human DNA and used in this preblocking procedure. The DNAs were mixed, boiled for 5 min, quenched in ice water for 5 min, and then prehybridized in 1 ml of hybridization solution for 30 min at 65°C (13, 14). Duplicate plaque lifts from the starting and selected libraries were screened at 5–10 ng of radiolabeled probe per ml.

DNA Sequencing. PCR products from cDNA clones were gel-purified using Geneclean II and sequenced using T3 and T7 primers and a Pharmacia DNA sequencing kit. The PCR product from the YAC genomic DNA was sequenced using the 3′ noncoding primers (GB2.1 and GB2.2) that are specific to the *GNB2* gene.

RESULTS

The Selection Scheme. Fig. 1 shows a diagram of the direct selection scheme. Purified YAC DNA is denatured, immobilized on a solid or filter support, and then hybridized with an entire library of cDNA inserts. cDNA inserts can be either amplified from the vector using vector primers or, after fragmentation, ligated to a linker and subsequently amplified (21). This type of amplification cassette is indicated by the boxes on the ends of the cDNAs in Fig. 1. Likewise, a mixture of cDNA libraries or equalized abundance cDNA libraries can be used (22). Approximately 3–5% of cDNAs contain repetitive elements (23) and these must be blocked or eliminated from the hybridization prior to the selection step. The blocked cDNAs are hybridized to the YAC and specific cDNAs are eluted after post-hybridization washing. The eluted cDNAs are then amplified and can be either cloned using restriction sites in the end linker or used in additional selection cycles (see *Discussion*).

We have used shorter genomic clones ranging from 20 kb to 40 kb that include the *IL3* and GM-CSF genes on human chromosome 5, as substrates for direct selection. The data from these pilot studies are summarized in Table 1. In the

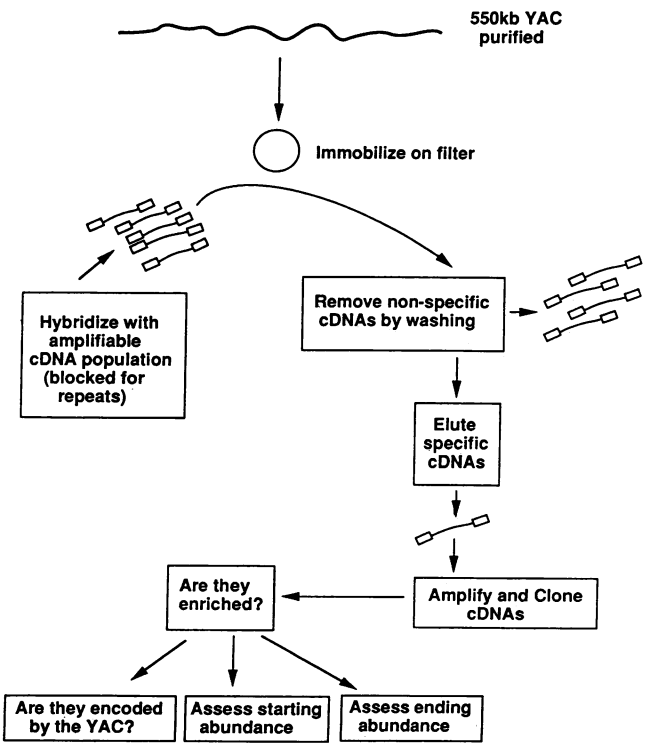

FIG. 1. Diagrammatic representation of the direct selection scheme (see text for details).

Table 1. Summarized enrichments of cDNAs

Genomic DNA	Gene	cDNA abundance		Fold enrichment
		Starting	Ending	
IL3/GM-CSF	*IL3*	1 in 250,000*	≈1 in 300†	≈800
	GM-CSF	1 in 250,000*	≈1 in 300†	≈800
EPO YAC	*EPO*	1 in 1,000,000‡	≈1 in 1000§	≈1000
	GNB2	1 in 100,000‡	≈1 in 100§	≈2000

The IL3/GM-CSF phage/cosmid was 20–40 kb and the EPO YAC was 550 kb.

*Starting abundance in T-cell cDNA library (1×10^6 clones screened). Amplification was conducted with an oligonucleotide linker/primer added to the digested cDNA.

†Ending abundance was assessed on 5×10^4 clones.

‡Starting abundance in fetal kidney cDNA library (1×10^6 clones screened). Amplification was conducted with vector primers.

§Ending abundance was assessed on 2.5×10^4 clones.

experiments described here the starting reagent was a 550-kb YAC clone that contains the human erythropoietin (*EPO*) gene. This gene has been localized to chromosome 7q21–22 (24, 25) and the YAC DNA was confirmed to be derived from this region by fluorescence *in situ* hybridization (data not shown). No other genes were known to reside within this large genomic clone prior to this study. A human fetal kidney cDNA library was used. This was shown to contain an EPO cDNA by the PCR and also by direct screening with an EPO cDNA probe.

cDNA inserts were derived by PCR using vector primers and a direct selection was conducted on the EPO YAC. Fig. 2*A* shows a gel electrophoretic comparison of the cDNA inserts from the starting and selected libraries. As expected, the sequence complexity in the selected library was clearly reduced when compared to the starting library. This was not due to preferential reamplification of a subset of the sequences present in the starting library: as shown in Fig. 2*B*, an additional 30 cycles of amplification slightly reduced the mean length of the starting library but did not result in a pattern similar to that of the selected library. A control hybridization of cDNAs to a "blank" filter resulted in no detectable fragments upon subsequent amplification (data not shown). Likewise, a control elution from a filter that contained the YAC but was not hybridized with the cDNA resulted in no detectable products (data not shown). The strong band at 800 bp in Fig. 2*A* was a common contaminant that could be eliminated from the selected material (see below).

The Degree of Enrichment. To gain a quantitative assessment of the degree of enrichment, plaque lifts from the starting and selected libraries were hybridized with an EPO cDNA. As is summarized in Table 1, a screen of one million cDNAs from the starting library resulted in the detection of only one EPO cDNA. In comparison, the selected library contained the EPO cDNA at a frequency of 1 in 1000 clones, an enrichment of 1000-fold. An example of this latter hybridization is shown on the right in Fig. 3*A* in which duplicate EPO positives are indicated on a screen of 5000 recombinant plaques. The low intensity of the detected signals in Fig. 3*A* is due to the high density of the screen and the use of a low-specific-activity cDNA probe. These positives were purified to homogeneity and subsequently confirmed by the PCR. From this result it appears that the selection system can enrich a positive control cDNA by ≈1000-fold even when that cDNA is extremely low in abundance in the starting cDNA.

Screening with the YAC. Prior to initiating a direct selection, we assessed whether direct YAC screening could detect the EPO cDNA in the starting fetal kidney cDNA library. Additional sets of plaque lifts made from the same starting library were hybridized with the gel-purified, radiolabeled, and blocked EPO YAC DNA. No duplicate positives were detected in this screen (data not shown) nor was the previously located EPO cDNA clone detected. After the selection and 1000-fold enrichment of the EPO cDNA, we again determined whether the YAC could detect EPO cDNAs within this lower-sequence-complexity library. Fig. 3*A* shows duplicate plaque lifts of 5000 recombinants from the

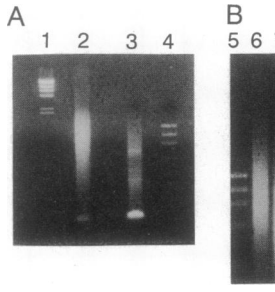

FIG. 2. Size distribution of the selected and eluted cDNAs. (*A*) Comparison of the eluted and amplified cDNAs (selected library with T3/T7 primers) with the starting library of cDNAs (10R/10F primers). (*B*) Comparison of these same two sets of cDNAs with a sample of the starting library that was reamplified with 10F and 10R primers for 30 cycles of the PCR. Marker tracks are a *Hin*dIII digest of λ DNA (lane 1) and a *Hae* III digest of φX174 replicative form DNA (lanes 4 and 5). Lanes: 2 and 6, starting library; 3 and 8, selected library; 7, starting library reamplified.

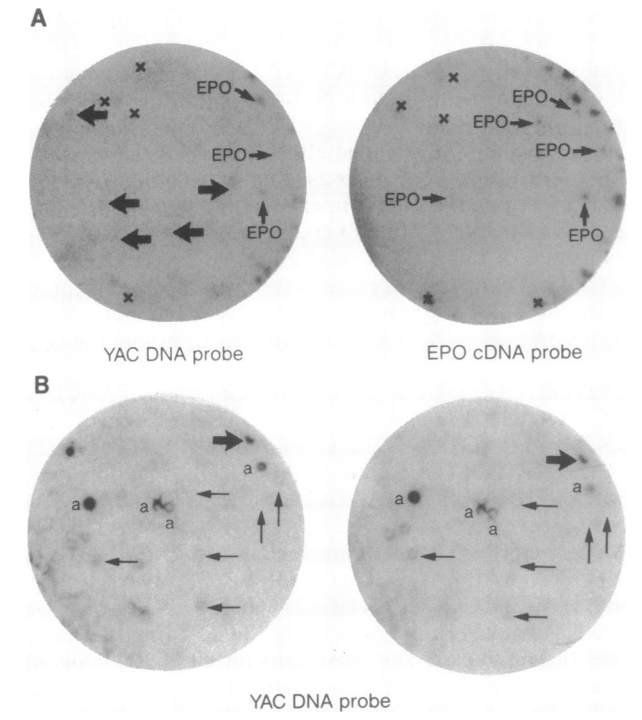

FIG. 3. Screen of the selected library with the radiolabeled EPO cDNA and with the radiolabeled YAC DNA. (*A*) Duplicate plaque lifts from a plating of 5000 selected cDNAs. The filter shown on the right was hybridized with the radiolabeled EPO cDNA. Only signals that were detected in duplicate are shown. The filter on the left was hybridized with the radiolabeled blocked EPO YAC DNA. Duplicate positives are shown and positives that overlap with EPO cDNAs are marked as EPO. Duplicate positives that were subsequently found to be GNB2 cDNAs are shown by arrows. (*B*) Duplicate lifts from a screen of 400 selected cDNAs, hybridized with radiolabeled YAC EPO. Both filter duplicates are shown. The duplicate positives marked "a" are selection artifacts. The large arrow shows the GNB2-positive clone that was picked. Smaller arrows indicate where other copies of this same cDNA were subsequently detected on this filter using a purified GNB2 cDNA as a probe.

selected library. The filter on the right was hybridized with the EPO cDNA probe as mentioned above. The filter on the left was hybridized with the EPO YAC that was gel-purified, radiolabeled, and blocked. In this experiment the radiolabeled YAC was blocked with both human DNA and yeast DNA prior to hybridization. In all cases the signals derived with the labeled YAC were low in intensity. Nevertheless, three of five EPO cDNAs that were present on this filter were detected by YAC hybridization on duplicate filters (cf., the left and right filters). This variability in signal detection when using YAC probes is not unusual, as discussed above.

In addition to the EPO cDNAs, the YAC also detected several other duplicate positive signals. These are indicated by the large arrows in Fig. 3*A*. As will be described below, the marked positives were subsequently found to be one cDNA species that is also encoded by the YAC and concomitantly enriched by the direct selection procedure.

Selection Artifacts. To unequivocally identify positive cDNAs, lower-density platings of the selected library were screened. Duplicate filters from one of these plates are shown in Fig. 3*B*. These filters were also screened with YAC DNA that was gel-purified, radiolabeled, and blocked. However, this screen differs from that shown above in that this YAC DNA was blocked with human DNA but not blocked with yeast DNA. This has important consequences for signal detection. The clones indicated by the a in Fig. 3*B* were purified and their DNAs were sequenced. These clones were derived from yeast rRNA or DNA and represent a common contaminant in commercially available cDNA libraries. This is probably a consequence of the use of yeast RNA as carrier during cDNA synthesis and its subsequent conversion to cDNAs. The 800-bp band present in the selected material shown in Fig. 2*A* was subsequently confirmed as a yeast ribosomal contaminant. We have found that this band can be largely eliminated by either blocking the cDNAs with yeast DNA prior to selection or by blocking the YAC DNA prior to hybridization screening.

The blocking of repetitive elements in the cDNA appears to have been successful since the spurious enrichment of cDNAs containing repetitive sequence elements did not appear to constitute a serious problem in this library or in selected cDNA libraries we have constructed from other genomic regions. Only 1% of the selected cDNAs detectably hybridized to radiolabeled total human genomic DNA (data not shown). In contrast, between 3% and 5% of the starting cDNA clones contain repetitive elements (23).

True Signals. The duplicate positive in Fig. 3*B* (marked with a large arrow) was not a yeast artifact and did not hybridize with total human DNA. This cDNA clone was purified and the cDNA insert was radiolabeled. This cDNA probe, when hybridized to a Southern blot containing various restriction digests of human genomic DNA, was found to be single copy (data not shown). When this probe was hybridized to the filters shown in Fig. 3, it was found to comprise 2% of the selected library. The arrows in Fig. 3*B* indicate where additional copies of this positive cDNA were identified. Three out of six of these are detectable in the YAC hybridization on both filters; one is detectable on one filter but not on the other, and two are not detectable on either filter. This again serves to underscore the weak signals and variable detection sensitivity inherent in screening with large blocked YAC DNAs. Nevertheless, a sufficiently high number of true positives were detectable within the enriched library, such that additional direct selection cycles were not immediately necessary.

The positive cDNA was next hybridized to the starting fetal kidney cDNA library and was found to be present at a frequency of one positive in 100,000 plaques (summarized in Table 1). This indicates that the cDNA was enriched by 2000-fold in the selected library relative to the starting

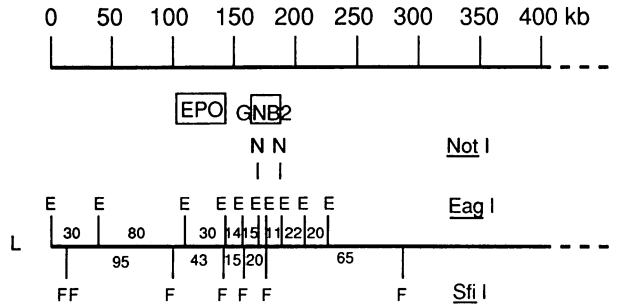

FIG. 4. Localization of the *GNB2* gene within the EPO YAC. (*Top*) A scale in kb is shown. (*Middle*) The deduced location of the *EPO* and *GNB2* genes. (*Bottom*) Positions of the *Not* I (N), *Eag* I (E), and *Sfi* I (F) sites within the YAC in this region are shown. The distances in kb between *Sfi* I and *Eag* I sites are shown. The left end of the YAC is denoted by the L.

library, an enrichment that is in good agreement with the other selections summarized in Table 1.

The Positive cDNA Is *GNB2* and Is Encoded by the YAC. The DNA sequence of the positive cDNA was determined and was found to share 100% identity with the gene encoding the β_2 subunit of guanine nucleotide binding protein (*GNB2*, ref. 26 and data not shown). The *GNB2* gene has been localized to the long arm of chromosome 7 and is highly homologous to the gene encoding the β_1 subunit of guanine nucleotide binding protein (*GNB1*) located on chromosome 1 (26, 27). However, our DNA sequence unequivocally identified the cDNA as being encoded by the *GNB2* locus. To confirm that the YAC did indeed contain the *GNB2* gene and not a closely related member of this gene family, we made use of the fact that the 3′ untranslated region of the *GNB2* gene is entirely specific to the *B2* gene (26). Primers specific to this region were used in a PCR conducted on the purified YAC to confirm the presence of these sequences within the genomic clone (data not shown). The DNA sequence of the resulting PCR products was also determined and found to be 100% identical to the *GNB2* gene, thus confirming the presence of the gene within the YAC.

Positioning of the *GNB2* Gene Relative to EPO. Prior to initiating our direct selections, a library of phage clones had been derived from the purified YAC clone and a long range map of this region had been embarked upon. The localization of *GNB2* within this map (Fig. 4) was based upon the following data. Hybridization of the GNB2 PCR products specific to the 5′ and 3′ ends of the gene identified one phage that contained a *Not* I site. However, a subsequent more-detailed PCR analysis revealed that this phage did not contain the 3′ end of the gene. The 5′ and 3′ GNB2 PCR products detected an ≈10-kb *Eag* I fragment on conventional Southern blots and an ≈20-kb *Not* I fragment on pulsed-field blots. Thus these data indicate that the *GNB2* gene is located 30–70 kb away from the previously localized *EPO* locus, within the small *Not* I fragment.

DISCUSSION

In this report we have described an approach to the isolation and identification of transcribed sequences from large genomic regions. We have applied this scheme to the isolation of cDNAs that are encoded by a large YAC clone and have identified one gene that is expressed at low levels and is closely linked to the *EPO* gene. Serendipitously, this selected cDNA was found to be encoded by a known gene. This demonstration of the sensitivity and selectivity of the technique naturally leads to the questions of how many other genes might be thought to reside on this YAC and whether any of them are related to *EPO* or *GNB2*. Interestingly, the

gene for the α_{i1} subunit of human guanine nucleotide-binding proteins has also been mapped to chromosome 7 band q21 (28, 29). However, it is unlikely that this gene is within the EPO YAC since a PCR on the YAC with primers specific to various parts of the *GNAI1* gene was negative (data not shown). Some estimate of the number of genes that might be encoded by a 500-kb genomic clone can be gained by considering estimates of total gene number and gene distribution (30). Based on these rather tenuous calculations, the EPO YAC might be expected to encode 10 or more genes. The selected library described here not only contains a high level of contaminating yeast clones (≈10%) but also contains at least four human cDNAs, that each comprise between 2% and 4% of the selected material. It is not yet known whether these are also encoded by the YAC.

It is clear that if a saturation transcription map of a large genomic region is required, then cDNAs from many tissue types will have to be sampled. One approach for sampling from many tissues, would be to employ multiplexing strategies, in which mixtures of cDNA libraries from various sources were "tagged" with different end linkers. In addition, the direct selection scheme is readily amenable to additional cycles of selection and amplification, to further increase the degree of enrichment. We have subjected the selected cDNAs to further cycles of amplification and observed a qualitatively higher level of EPO cDNA enrichment (data not shown), although the mean size of the PCR products decreased. With the refinement of these steps, it may be possible to enrich to a point at which random picking of positive cDNAs is successful, even for the very low abundance cDNAs described in this report.

Three other components of the direct selection scheme are also amenable to further development. The first is the choice of matrix on which to immobilize genomic clones. Our preliminary data indicate that streptavidin-coated beads combined with biotin-labeled YAC DNAs are efficient selection matrice. The second component that could be improved upon is the blocking and/or depletion of highly repetitive elements from genomic clones prior to immobilization or use as hybridization probes. The third component is the choice and construction of cDNA libraries. The depletion of repeats in cDNA libraries and the approximate normalization of cDNA abundance classes (22) are two of the areas that are open to improvement.

The future applications of the direct selection scheme are wide-ranging. We anticipate that when combined with arrayed chromosome-specific genomic DNA libraries, the selection technique could be used for the enrichment of chromosome-specific cDNAs. Combinations of large YAC contigs and multiplexed cDNA libraries should be particularly powerful, allowing for the selection of cDNAs from a number of tissues at once. The development of these types of strategies should lead to insights into gene distribution, the determination of tissue specific transcription maps across large genomic regions, and the rapid isolation of candidate cDNAs for many disease-related genes.

We thank Martha Guerra for her assistance with DNA sequencing and Rick Cuevas for help in preparation of the manuscript. We are grateful to Drs. Greg Reyes and Cynthia Edwards for their comments on the manuscript. This work was supported in part by Grants RO1 HG00368 and R44 HG00508 to M.L. from the National Center for Human Genome Research. J.K. was supported in part by the Academy of Finland, the Finnish Cultural Foundation, and Finska Läkaresällskapet.

1. Lindsay, S. & Bird, A. P. (1987) *Nature (London)* **327,** 336–338.
2. Lavia, P., Macleod, D. & Bird, A. (1987) *EMBO J.* **6,** 2773–2779.
3. Pontarotti, P., Chimini, G., Nguyen, C., Boretto, J. & Jordan, B. R. (1988) *Nucleic Acids Res.* **16,** 6767–6778.
4. Abe, K., Wei, J. F., Hsu, Y. C., Uehara, H., Artzt, K. & Bennett, D. (1988) *EMBO J.* **7,** 3441–3449.
5. Rommens, J. M., Iannuzzi, M. C., Kerem, B., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J. R., Tsui, L.-C. & Collins, F. S. (1989) *Science* **245,** 1059–1065.
6. Yokoi, T., Lovett, M., Cheng, Z. Y. & Epstein, C. J. (1986) *Hum. Genet.* **74,** 137–142.
7. Hochgeschwender, U., Sutcliffe, J. G. & Brennan, M. B. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8482–8486.
8. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharfe, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Ehrlich, H. A. (1988) *Science* **239,** 487–494.
9. Liu, P., Legerski, R. & Siciliano, J. (1989) *Science* **246,** 813–815.
10. Duyk, G. M., Kim, S. W., Myers, R. M. & Cox, D. R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8995–8999.
11. Buckler, A. J., Chang, D. D., Graw, S. L., Brook, D., Haber, D. A., Sharp, P. A. & Housman, D. E. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4005–4009.
12. Hochgeschwender, U. & Brennan, M. B. (1991) *BioEssays* **13,** 139–144.
13. Burke, D. T., Carle, G. F. & Olson, M. V. (1987) *Science* **236,** 806–812.
14. Wallace, M. R., Marchuk, D. A., Anderson, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Brereton, A., Nicholson, J., Mitchell, A. L., Brownstein, B. H. & Collins, F. S. (1990) *Science* **249,** 181–186.
15. Elvin, P., Slynn, G., Black, D., Graham, A., Butler, R., Riley, J., Anand, R. & Markham, A. F. (1990) *Nucleic Acids Res.* **18,** 3913–3917.
16. Sealey, P. G., Whittaker, P. A. & Southern, E. M. (1985) *Nucleic Acids Res.* **13,** 1905–1932.
17. Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J.-R., Burks, C., Sirotkin, K. M. & Goad, W. B. (1989) *Genomics* **4,** 273–289.
18. Burke, D. T. & Olson, M. V. (1991) *Methods Enzymol.* **194,** 251–270.
19. Church, G. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
21. Reyes, G. R. & Kim, J. P. (1991) *Mol. Cell. Probes*, in press.
22. Patanjali, S. R., Parimoo, S. & Weissman, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1943–1947.
23. Crampton, J. M., Davies, K. E. & Knapp, T. F. (1981) *Nucleic Acids Res.* **9,** 3821–3834.
24. Watkins, P. C., Eddy, R., Hoffman, N., Stanislovitis, P., Beck, A. K. & Galli, J. (1986) *Cytogenet. Cell. Genet.* **42,** 214–218.
25. Law, M. L., Cai, G. Y., Lin, F. K., Wei, Q., Huang, S. Z., Hartz, J. H., Morse, H., Lin, C. H., Jones, C. & Kao, F. T. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6920–6924.
26. Fong, H. K. W., Amatruda, T. T., III, Birren, B. W. & Simon, M. I. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3792–3796.
27. Blatt, C., Eversole-Cire, P., Cohn, V. H., Zollman, S., Fournier, R. E. K., Mohandas, L. T., Nesbitt, M., Lugo, T., Jones, D. T., Reed, R. R., Weiner, L. P., Sparkes, R. S. & Simon, M. I. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7642–7646.
28. Bray, P., Carter, A., Guo, V., Puckett, C., Kamholz, J., Spiegel, A. & Nirenberg, M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5115–5119.
29. Bloch, D. B., Bloch, K. D., Iannuzzi, M., Collins, F. S., Neer, E. J., Seidman, J. G. & Morton, C. C. (1988) *Am. J. Hum. Genet.* **42,** 884–888.
30. Bickmore, W. A. & Sumner, A. T. (1989) *Trends Genet.* **5,** 144–148.