Isolation of cDNAs from the Cri-du-chat critical region by direct screening of a chromosome 5-specific cDNA library.

A D Simmons, J Overhauser and M Lovett

Genome Res. 1997 7: 118-127
Access the most recent version at doi:10.1101/gr.7.2.118

RESEARCH

Isolation of cDNAs from the Cri-du-chat Critical Region by Direct Screening of a Chromosome 5-Specific cDNA Library

Andrew D. Simmons,[1] Joan Overhauser,[2] and Michael Lovett[1,3]

[1]Departments of Otorhinolaryngology, Molecular Biology and Oncology, and The McDermott Center, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75235; [2]Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Chromosome-specific cDNA libraries are new tools for the isolation of genes from specific genomic regions. We have used two YACs than span the ~2-Mb cri-du-chat critical region (CDCCR) of chromosome 5p to directly screen a chromosome 5-specific (CH5SP) fetal brain cDNA library. To compare this library with other sources for new gene discovery, the YACs were hybridized to a normalized infant brain (NIB) cDNA library that has been used extensively for expressed sequence tag (EST) generation. These screens yielded 12 cDNAs from the CH5SP fetal brain library and four cDNAs from the NIB library that mapped to discrete intervals within the CDCCR. Four cDNAs mapped within the minimal CDCCR deletion interval, with the remaining cDNAs being located beyond the boundaries. Only one cDNA shared sequence overlap between the CH5SP and NIB sets of clones. None of the remaining 11 CH5SP cDNAs were homologous to EST sequences, suggesting, in common with previous data on these libraries, that chromosome-specific cDNA libraries are a rich source of new expressed sequences. The single cDNA that did overlap with the NIB library contained two copies of a sequence motif shared with thrombospondin, properdin, and several complement proteins. This motif is usually present in adhesive proteins, and appears to mediate cell–cell or cell–substrate interactions. This new thrombospondin-like gene, and the other three cDNAs that map within the CDCCR, represent candidate genes for the cri-du-chat contiguous gene deletion syndrome.

[The sequence data described in this paper have been submitted to GenBank under accession nos. U52827–U52840.]

Cri-du-chat is a human deletion syndrome characterized by large deletions of the short arm of chromosome 5. The clinical features of cri-du-chat include a characteristic high-pitched cat cry, severe mental retardation, microcephaly, hypertelorism, and epicanthic folds (Niebuhr 1978a). Although the size of the deletions varies considerably, karyotypic analysis has revealed that a critical region at 5p15 is deleted from one copy of chromosome 5 in all patients (Niebuhr 1978b; Overhauser et al. 1986).

When the phenotypes from ~50 cri-du-chat patients were compared with their chromosomal breakpoints, two distinct regions were resolved (Overhauser et al. 1994; Gersh et al. 1995). The hallmark cat–cry phenotype mapped distal to the remaining features of cri-du-chat, which mapped to 5p15.2. The latter region has been designated the cri-du-chat critical region (CDCCR), and previously we have reported the derivation of a framework transcription map of this interval (Simmons et al. 1995). This was accomplished using direct cDNA selection with 30 cosmids that were detected by sequence-tagged sites (STSs) within the CDCCR. Subsequently, we constructed a yeast artificial chromosome (YAC) contig of ~3 Mb covering the CDCCR that can be minimally spanned by two Centre d'Etude du Polymorphisme Humain (CEPH) mega-YACs (Goodart et al. 1994). Nine novel cDNAs from our selections with the framework cosmids were mapped to discrete intervals within this YAC contig. These represented the first and, to date, only transcription units to be placed within the CDCCR. However, the molecular dissection of this complex disorder and large genomic region will require a much more detailed gene map and extended cDNA clones, rather than the relatively short cDNAs that have been placed within it to date. In this report we

[3]Corresponding author.
E-MAIL lovett@ryburn.swmed.edu; FAX (214) 648-1666.

7:118–127 ©1997 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/97 $5.00

have specifically targeted longer cDNA sequences by using the two large YACs that span the interval as probes against various arrayed cDNA libraries.

The identification of candidate genes is often the rate-limiting step in the positional cloning of genes that cause human genetic disorders. To facilitate this step, we have used direct cDNA selection with an entire human chromosome to generate chromosome-specific cDNAs. These resources are comprised of a complex set of cDNAs and have a broad spectrum of potential applications (Del Mastro et al. 1995). In a previous report from our group, chromosome 5 genomic DNA was used in conjunction with direct cDNA selection to construct prototype chromosome 5-specific (CH5SP) cDNA libraries from five different tissue sources (Del Mastro et al. 1995). The genomic target for these selections was a single pool of 24,768 cosmid clones from a chromosome 5-specific genomic DNA library (Longmire et al. 1993). The resulting cDNA libraries contain mRNAs that are normally low in abundance at relatively high levels as a result of quasi-normalization of cDNA abundance (Weissman 1987; Lovett et al. 1991; Parimoo et al. 1991). To rapidly isolate additional cDNAs from the cri-du-chat critical region, two YACs, representing ~3 Mb of genomic DNA sequence, were used to directly screen high-density arrays of clones from a fetal brain CH5SP cDNA library. The two YACs were also used to screen high-density arrays of clones from the arrayed normalized infant brain (NIB) cDNA library that is being used to derive expressed sequence tags (ESTs) by the Merck/Washington University EST project (Lennon et al. 1996). The NIB cDNA library was constructed by a reassociation technique using a directionally cloned infant brain cDNA library (Soares et al. 1994), whereas the chromosome-specific cDNAs were normalized against genomic DNA. The parallel screening of the two sets of brain-derived cDNA libraries, which have been quasi-normalized using different techniques, allowed us to conduct an indirect comparison of the two resources.

RESULTS

Hybridization of Cri-du-chat YACs to cDNA Resources

The CEPH mega-YACs 938G6 and 904G7, which span the CDCCR (Fig. 1), were purified from the host yeast chromosomes using pulsed-field gel electrophoresis. Purified YAC DNAs were radiolabeled and hybridized to filters containing a high-density array of 4608 clones from the fetal brain CH5SP cDNA library (Del Mastro et al. 1995). Although our standard repeat suppression procedure for YACs is performed for 2 hr, this led to insufficient repeat suppression of the CDCCR YACs (A. Simmons, unpubl.). To obtain consistent blocking of repetitive sequences, the blocking time was extended to 15 hr. This confirms several reports suggesting that chromosome 5p15 has an unusually high concentration of low or medium copy number repetitive sequences (Sargent et al. 1994; Simmons et al. 1995; Thompson et al. 1995). An example of the hybridization of the YACs to a single filter of arrayed clones from the CH5SP fetal brain cDNA library is shown in Figure 2. A small number of clones, <1% (n = 28), show strong duplicate signals that are independent of the YAC that is hybridized. These artifactual positives, indicated with open arrows, are homologous to sequences within the YAC cloning vector. Examples of the weak duplicate positives (n = 56) that were analyzed further are indicated with closed arrows in Figure 2. In a parallel set of experiments, the two mega-YACs were radiolabeled, repeat-suppressed, and hybridized to filters containing 27,648 arrayed clones from the NIB cDNA library. Thirty-four duplicate positives were identified, picked, and further characterized.

Sequence Analysis of Positive Clones

The 56 positive clones obtained from the CH5SP library were sequenced and compared with DNA sequences deposited in GenBank (Altschul et al. 1990). Twenty-two of the clones (39%) were novel by sequence analysis. Two identical clones (4%), designated CSA1, were homologous to two ESTs deposited in GenBank. The remaining 32 clones (57%) were homologous to LINE1 repeats or other low/medium copy number repeats (i.e., MER, THE, OFR). This is in contrast to other data obtained by screening this resource (R.G. Del Mastro and M. Lovett, in prep.), where we generally observe <5% of the clones containing repetitive elements. This appears to be attributable to the aforementioned presence of an unusually high concentration of medium and low copy number repetitive elements within the CDCCR. Although these clones were eliminated from further analysis because of the technical problems in working with repeats, they may indeed represent genes from the CDCCR with repetitive elements in their untranslated regions.

The 3′ ends of the 34 positive NIB clones were sequenced and analyzed. Seven of the positives (21%) contained repetitive elements. Three (9%) of the clones were identical to the DAP-1 gene de-

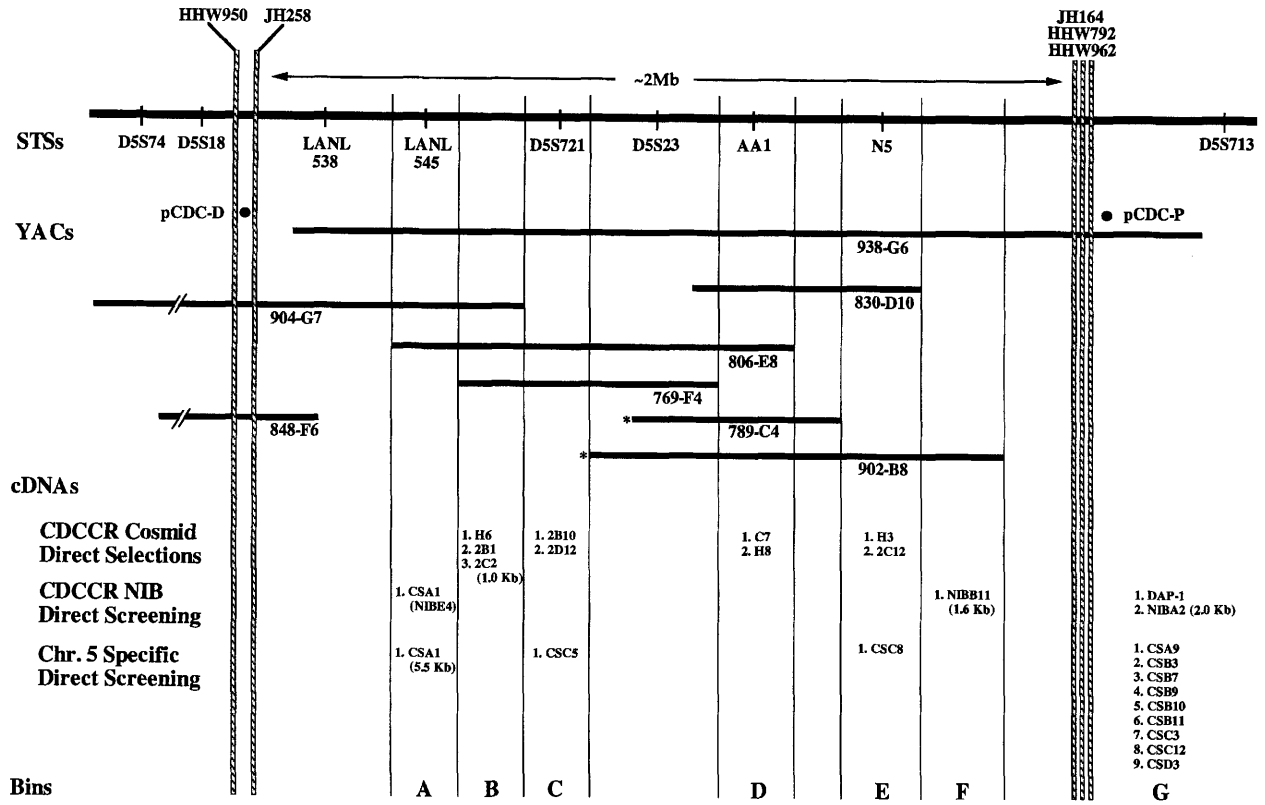

Figure 1 Relative placement of the cDNAs isolated within the CDCCR physical map. The hatched vertical bars represent the chromosomal breakpoints from the five cri-du-chat patients that define the ~2-Mb deletion interval (Goodart et al. 1994). Thin vertical bars divide the CDCCR into seven discrete bins containing cDNAs. The bins, arbitrarily labeled A–G, are based on unique overlap patterns within the YACs and somatic cell hybrids that define the CDCCR. The relative placement of the 15 cDNAs obtained from the CH5SP and NIB screening, as well as nine cDNAs that were isolated previously using direct cDNA selection with 30 framework cosmids (Simmons et al. 1995), has been indicated. With one exception, the approximate size of cDNAs ⩾1 kb is indicated in parenthesis. Clone NIBE4 is indicated in parenthesis to represent a NIB cDNA that was identified by hybridization, but later assembled into a ~5.4 kb contig with CSA1.

scribed previously (Deiss et al. 1995). The DAP-1 gene was initially identified in a screen for suppressers of γ interferon-induced apoptotic cell death. Previously it had been mapped to the general vicinity of the CDCCR (5p15.2) by our group through sequencing and mapping of randomly picked clones from chromosome 5-specific cDNA libraries (Del Mastro et al. 1995). Subsequently, DAP-1 was shown to map to CEPH mega-YAC 938G6 (Berry et al. 1995). All of the remaining NIB-derived clones (n = 24, 70%) were homologous to ESTs deposited in GenBank. This high percentage of EST homologies is not surprising because the NIB library has been used extensively for EST generation.

Mapping and Redundancy Analysis

Removal of the repetitive and ribosomal clones yielded 24 and 27 clones, respectively, from the CH5SP and NIB resources. PCR primers were derived from these sequences and confirmed to map to the CDCCR by PCR analysis of somatic cell hybrids and YACs from the deleted interval. Of the 24 clones from the CH5SP library, 17 mapped to discrete intervals within the physical map of the CDCCR. Similarly, 16 of the 27 clones obtained from the NIB screening were confirmed to map to the CDCCR. The cDNAs that mapped within the CDCCR were separated into bins corresponding to regions defined by unique patterns of YAC and/or somatic cell hybrid overlap (Fig. 1). With the exception of NIBA2, the cDNAs mapped unambiguously within the YACs and somatic cell hybrids. Primers derived from both the 5′ and 3′ ends of NIBA2 unambiguously mapped to only YAC 938G6. However, these primers amplified mouse and hamster genomic

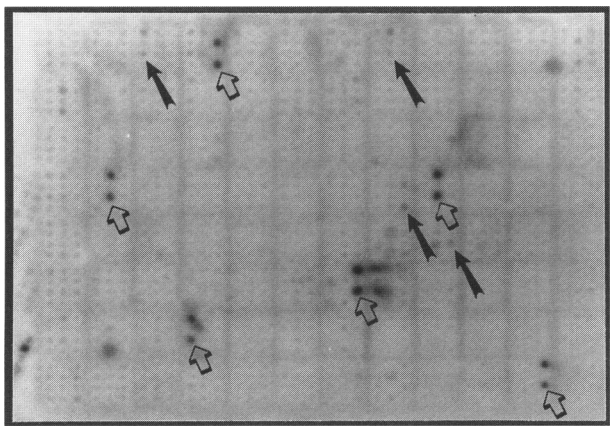

Figure 2 Representative example of hybridization of CDCCR YACs 938G6 and 904G7 to a filter of arrayed clones from the CH5SP cDNA library. Each filter contains 576 clones, gridded twice, to provide duplicate positive signals when hybridized with radiolabeled probes. Six intensely hybridizing clones, indicated by open arrows, are artifacts that are homologous to sequences within the YAC cloning vector. These clones have been cataloged and removed from further analysis. Examples of the weak duplicate positives that were analyzed further are indicated with solid arrows.

DNA and thus could not be unequivocally positioned within the intervals established by the hybrid panel. All other cDNAs that mapped only to YAC 938G6 mapped within bin G, and on that basis cDNA NIBA2 was arbitrarily placed in this interval. The YAC contig in Figure 1 also shows the placement of nine cDNAs that we isolated previously using direct cDNA selection with 30 framework cosmids from the CDCCR (Simmons et al. 1995).

Sequence alignments of the positive cDNAs merged the 17 CH5SP cDNAs into 12 overlapping sequences and the 16 NIB cDNAs into four overlapping cDNAs. At this point in our analysis, there was no sequence overlap between the two sets of cDNAs. To investigate this relationship further, each positive clone obtained from one of the libraries was screened against the other library to determine whether undetected overlap existed. PCR primer pairs from the 12 CH5SP cDNAs were used to determine whether these clones were present in a pool of the NIB cDNA library (comprised of 50,000 clones). It should be noted that this number is in excess of the ~40,000 cDNAs arrayed from the NIB library for EST generation and screening (i.e., Research Genetics). Only one of the cDNAs tested, CSA1, was present in the NIB library. The four cDNAs from the NIB screening were hybridized to high-density arrays of 4608 clones from each of five CH5SP cDNA libraries. These five libraries were derived from different tissue sources including fetal brain (Del Mastro et al. 1995). Of the four NIB cDNAs tested, only the DAP-1 gene was present in the CH5SP libraries. DAP-1 was present in the HeLa and placental CH5SP cDNA libraries, but was not present in the fetal brain CH5SP library that was used for direct screening. No sequence overlap was observed between the 16 novel cDNAs described here and the nine cDNAs that we isolated previously using direct cDNA selection with 30 framework cosmids from the CDCCR (Simmons et al. 1995).

Northern Blot Analysis and cDNA Extension

To evaluate expression profiles and determine transcript lengths, the three CH5SP and two NIB cDNAs that mapped within the CDCCR (Fig. 1) were analyzed by hybridization to Northern blots. Three of these five cDNAs were expressed at detectable levels by Northern blot analysis. The two remaining cDNAs were derived from the CH5SP cDNA library and their lack of detectable signal on Northern blots supports previous observations that chromosome-specific libraries contain mRNAs that are normally low in abundance (Del Mastro et al. 1995). One cDNA, NIBB11, detected two transcripts of ~2 kb and 5 kb (Fig. 3). Although NIBB11 was moderately expressed in heart, skeletal muscle, pancreas, and prostate, weak expression was visible in several other tissues examined. The two remaining cDNAs, CSA1 and NIBE4, displayed almost identical expression profiles. Both cDNAs detected two transcripts of 10 kb and 8 kb, although CSA1 also detected two shorter transcripts in testis (Fig. 3). Both cDNAs were highly expressed in colon, heart, and kidney, although moderate or weak expression was observed in most of the tissues examined. Because CSA1 and NIBE4 mapped to the same interval and had very similar expression profiles, all combinations of PCR primer pairs were tested to determine whether they could be linked together (see Methods). Results from cDNA linking experiments and sequence analysis of ESTs deposited within the database were used to join and extend this cDNA to a final size of ~5.4 kb. Thus cDNAs CSA1 and NIBE4 represent different parts of the same gene. To summarize these studies: The ~5.4-kb extended CSA1 cDNA was identical to six randomly sequenced ESTs, representing sequence analysis of the 5′ and 3′ ends from three separate clones (Fig. 4A). Interestingly, two of the three randomly sequenced clones obtained from the EST database were the result of mispriming from an AT-rich region of the cDNA

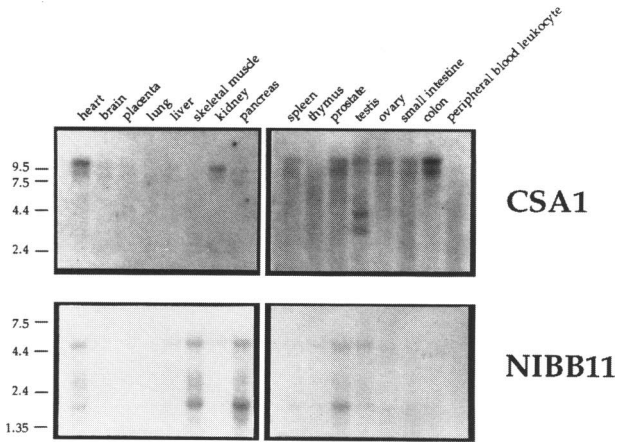

Figure 3 Expression profiles of cDNAs CSA1 and NIBB11. Clontech MTN blots I and II were hybridized with radiolabeled probes derived from cDNAs CSA1 and NIBB11. (*Left*) The positions in kilobases of RNA markers. CSA1 detected two transcripts of 10 and 8 kb, although two lower molecular mass bands were present in testis. High levels of expression were observed in colon, heart, and kidney, although moderate or weak expression was observed in most of the tissues examined. Two transcripts of ~2 and 5 kb were observed for NIBB11. Although NIBB11 was moderately expressed in heart, skeletal muscle, pancreas, and prostate, weak expression was visible in several other tissues examined.

rather than the true poly(A) tail. It has been estimated that ~13% of the clones in the NIB cDNA library are the result of such aberrant priming (Soares et al. 1994). The final size of the cDNAs ⩾1 kb has been indicated in Figure 1. The final cDNAs for DAP-1, NIBA2, NIBB11, and CSA1 are >95% identical to 53, 27, 13, and six randomly sequenced ESTs, respectively. With the exception of DAP-1 and CSA1, all of the remaining CH5SP and NIB cDNA clones (n = 13) were novel by DNA sequence analysis.

Sequence analysis of the CSA1 cDNA revealed similarity with several members of the thrombospondin gene family. Specifically, CSA1 has two copies of a highly conserved ~60 amino acid type 1 repeat present in thrombospondins 1 and 2 (TSP1 and TSP2) (Dixit et al. 1986; Kobayashi et al. 1986; Lawler and Hynes 1986; LaBell and Byers 1993). Additional proteins that contain type 1 thrombospondin repeats (TSR) include properdin, f-spondin, and the C6-, C7-, C8a-, C8b-, and C9-terminal components of complement (Goundis and Reid 1988; Robson et al. 1988; Bornstein and Sage 1994). In contrast to CSA1, TSP1 and TSP2 contain three copies of the type 1 TSR. An amino acid alignment of the TSR of CSA1 with the TSR of human TSP1 and TSP2 is shown in Figure 4B. Two functions of the type 1 TSR have been proposed: (1) binding of CD36 through the CSVTCG sequence (Rich et al. 1990; Catimel et al. 1992), and (2) binding cell surface sulfate proteoglycans and sulfatides through the WSXW motif (Guo et al. 1992). Although the latter WSXW motif is conserved in both copies of the CSA1 type 1 TSR, the former motif is not as conserved (Fig. 4B). A putative role for CSA1 in cell–cell or cell–substrate interactions can be postulated based upon these observed similarities with thrombospondins.

DISCUSSION

A detailed transcription map of the CDCCR is an essential step toward understanding the molecular basis of the cri-du-chat syndrome. In this study we used two YACs that span the CDCCR to screen cDNAs directly from two independently constructed brain-derived cDNA resources. Sequencing, redundancy, and mapping analysis of the positive clones yielded 15 unique cDNAs that map to discrete bins in a YAC contig of the CDCCR. The 15 cDNAs can be separated into four distinct sets. The first set consists of two NIB clones that are homologous to EST sequences of unknown function. The second set consists of 11 novel cDNAs identified from the CH5SP fetal brain library. The third set is represented by the known gene DAP-1, which was identified in the NIB cDNA library. The fourth set is represented by CSA1, which has a region of sequence similarity with several members of the thrombospondin gene family, and appears to be a new thrombospondin-like gene. When extended to ~5.4 kb, CSA1 was the only clone that shared sequence overlap between the two sets of clones obtained from the CH5SP and NIB libraries. Although the presence of only a single overlapping cDNA in our study is surprising, three explanations for the lack of sequence overlap can be presented. First, the lack of sequence overlap might be anticipated because the two cDNA resources were derived from temporally and morphologically different brain samples. An alternative, and less likely, explanation for this lack of overlap is that the two different methods by which these two libraries were constructed resulted in different skewing of the sequence representation in the final cDNA population. However, the extensive number of EST sequences that have been generated from the NIB library suggests that the sequence complexity of this library is quite high. Likewise, our sequence analysis of the CH5SP resources indi-

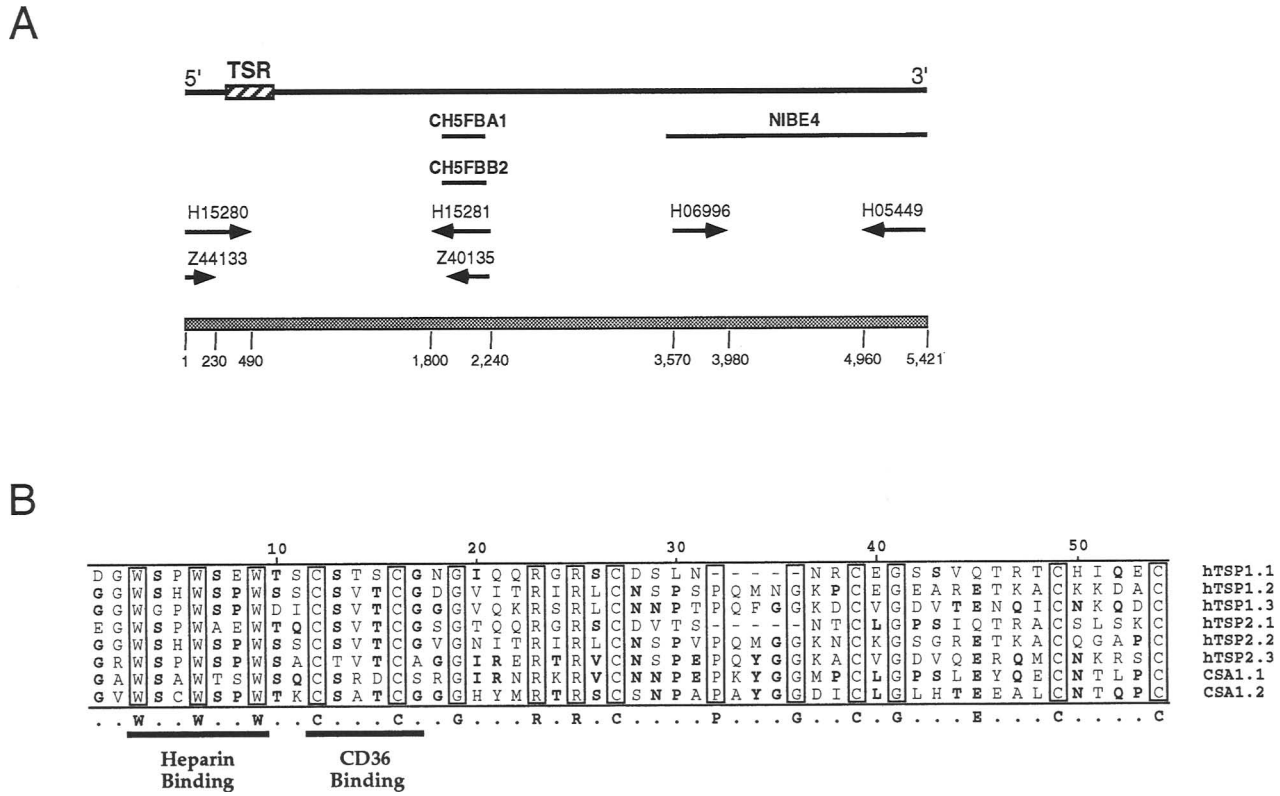

Figure 4 (*A*) Placement of CH5SP, NIB, and EST cDNAs within the ~5.4-kb CSA1 contig. Three cDNAs, representing different portions of the CSA1 gene, were obtained. First, two cDNA clones, designated CH5FBA1 and CH5FBB2, were identified by screening the CH5SP fetal brain cDNA library with YACs from the CDCCR. Second, cDNA NIBE4 was identified by screening the NIB resource with YACs from the CDCCR. The relative placement of six ESTs, representing the 5′ and 3′ sequences from three cDNA clones, is also indicated. The EST clones are designated by their GenBank accession numbers. The hatched bar, also indicated by TSR, represents the region of amino acid similarity with type 1 TSR present in TSP1 and TSP2. (*B*) Amino acid alignments of the type 1 TSRs present in cDNA CSA1, TSP1, and TSP2. The names of the sequences, followed by the repeat unit, are indicated to the *right* of the alignments. The three TSRs from TSP1 and TSP2 are shown along with the two TSRs from CSA1. The residues conserved between all sequences, including six conserved cysteines, are boxed and displayed as a consensus sequence. Residues in TSP1 or TSP2 that are identical to amino acids present in either repeat unit of CSA1 are indicated in bold. Binding sites for CD36 (Rich et al. 1990; Catimel et al. 1992) and heparin (Guo et al. 1992) are indicated with bold horizontal bars. Amino acid sequences were obtained from sp:P07996 (TSP1) and sp:P35442 (TSP2). Alignment was performed by eye using the work of Bornstein (1992) as an example.

cates that the sequence complexity in these libraries is high, and that a significant lack of overlap with ESTs does exist (Del Mastro et al. 1995). Third, because the CH5SP library was constructed from a random primed fetal brain cDNA pool, whereas the NIB cDNA library was constructed from an oligo(dT)-primed infant brain cDNA library, the low sequence overlap may represent cDNAs where the CH5SP clone is greater than ~1.5 kb from the poly(A) tail. This latter explanation appears unlikely because the cDNAs that we mapped within the CDCCR map to distinct physical areas and show different sized transcripts on Northern blots.

An unusually high number of low or medium copy number repetitive elements were obtained when the CDCCR YACs were used to screen the cDNA libraries. Random picking and sequencing of cDNAs from these resources indicates that ~10% of the cDNAs are repetitive, a number that is not significantly different than that found in conventional cDNA libraries (Del Mastro et al. 1995). The high concentration of repeats in the CDCCR is consistent with the cytogenetic location of the region within a Giemsa (G) dark staining band. Dark staining G bands are believed to be rich in LINE-1 elements and to generally contain a low number of expressed se-

quences (Gardiner 1995). Furthermore, a detailed analysis of the spinal muscular atrophy (SMA) region in 5q12-q13 has led to the identification of several families of chromosome 5-specific repeats and expressed pseudogenes present on both the short and long arms of chromosome 5 (Sargent et al. 1994; Simmons et al. 1995; Thompson et al. 1995), at least some of which lie close to the CDCCR. It is tempting to speculate that the unusually high occurrence of deletions present in the cri-du-chat region is attributable to aberrant recombination or rearrangements within this dense region of sequence repeats.

Although the two CDCCR YACs used for the hybridization span the CDCCR, the majority of the cDNAs obtained from the screening mapped outside the CDCCR. Three explanations can be proposed to account for this observation. First, nonuniform labeling of the YACs might lead to an uneven distribution of hybridization signal and biased cDNA isolation. Second, there may be a nonrandom distribution of genes within this region of the genome, with the majority of genes clustering outside the CDCCR. Third, because the YAC contig of the CDCCR was constructed by STS content mapping, rather than YAC end walking, a large part of YAC 938G6 may extend outside of the CDCCR interval. Thus, the cDNA distribution may merely reflect a random distribution over a wider region. Work in progress on targeted direct selections performed with purified YACs 938G6 and 904G7 also shows a clustering of cDNAs outside of the CDCCR (A. Simmons, unpubl.), supporting either of the latter two possibilities.

Four novel cDNAs have been placed within the CDCCR and are candidate genes for causing the phenotypes observed in the cri-du-chat syndrome. Three of these cDNAs are novel by DNA sequence analysis, and are of unknown function. They join the previously mapped nine cDNAs as genes that will require extensive functional characterization. The fourth cDNA (CSA1) contains two copies of a type 1 TSR and appears to encode a ubiquitously expressed thrombospondin-like protein. Thrombospondins mediate a diverse group of processes, including cell adhesion, migration, and proliferation (for review, see Frazier 1991; Bornstein and Sage 1994; Gardiner 1995), and CSA1 may also regulate one or more of these processes. Mutations in another member of the thrombospondin gene family, cartilage oligomeric matrix protein (COMP), have been shown to cause pseudoachrondroplasia, an autosomal dominant skeletal dysplasia (Briggs et al. 1995; Hecht et al. 1995). Interestingly, a locus for chrondrocalcinosis, an autosomal dominant condition associated with arthritis, has been mapped recently to chromosome 5p (Hughes et al. 1995), but appears to map proximal to CDCCR. It is possible, given the diverse roles that thrombospondins play in cellular processes, that haploinsufficiency for CSA1 might lead to some aspects of the cri-du-chat phenotypes. However, more detailed analysis of CSA1 and the other candidate cDNAs that we have isolated from this region, including their roles during normal development as well as gene dosage and imprinting studies, will be required to correlate the genotypic and phenotypic features of the cri-du-chat syndrome.

Unraveling the genetic basis of cri-du-chat will also require a detailed characterization of the entire critical region, including a comprehensive analysis of all the genes that are deleted. Although a total of 24 nonoverlapping cDNAs have now been identified and mapped to discrete intervals within a YAC contig of the CDCCR, current estimates of gene density suggest that ~100 genes may reside within the ~2-Mb deleted interval. To construct a more detailed transcription map of the CDCCR, targeted direct selections are currently in progress using several different cDNA sources, including cDNAs from various developmental stages and complex tissue sources, using the two YACs that span the critical region as the genomic target. These data, when combined with genomic DNA sequence information from the region, will provide the framework for detailed investigations of the molecular basis of cri-du-chat.

METHODS

Construction of cDNA Libraries

The construction and analysis of the chromosome 5-specific cDNA libraries has been described previously (Del Mastro et al. 1995). Briefly, purified DNA from the LANL chromosome 5-specific cosmid library (n = 24,768 clones) was biotinylated *en masse* and hybridized in solution to a fetal brain cDNA pool (Simmons et al. 1995). The biotinylated cosmids plus cognate cDNAs were then captured using streptavidin-coated paramagnetic beads. After washing, the cDNAs were eluted, PCR amplified, and passed through a second round of selection. The cDNA eluted from the second round of selection were PCR amplified, annealed to the vector pAMP10 (BRL), and transformed into DH5α library efficiency competent cells (BRL). Recombinant clones (n = 4608) were arrayed into forty-eight 96-well microtiter plates for further analysis. Purified plasmid DNA from the normalized infant brain cDNA library was kindly provided by Dr. M.B. Soares (Soares et al. 1994). Approximately 10 ng of DNA was electroporated into electroMAX DH5α cells (BRL) and 27,648 recombinants were arrayed into 96-well microtiter plates. A Biomek 1000 workstation was used to generate high-density nylon grids of the cDNA libraries (Olsen et al. 1993). The grids were stamped on Hybond N+

(Amersham) and processed as recommended by the manufacturer. The average size of clones from the CH5SP fetal brain and NIB cDNA libraries are 0.5 and 1.5 kb, respectively.

Screening of cDNA Libraries

YAC plugs were prepared as described (Carle and Olsen 1984). YACs 938G6 and 904G7 were size fractionated by pulsed-field gel electrophoresis (CHEF mapper, Bio-Rad), and a prep-a-gene kit (Bio-Rad) was used to purify DNA from slices of the agarose containing the YAC. Fifty nanograms of purified DNA from each YAC was pooled and labeled using nick translation and random primed labeling kits (Boehringer Mannheim) with three radiolabeled nucleotides (dCTP, dATP, and dGTP) and cold dTTP. After a 2-hr incubation at 4°C or 37°C, respectively, the reactions were pooled and purified by Sephadex G-50 chromatography. Blocking was performed overnight at 65°C in 500 µl of 1× hybridization solution (Lovett 1994) with 1 mg of sheared human placental DNA (Sigma) and 200 µg of Cot-1 DNA (BRL). The filters were blocked overnight at 65°C in 50 ml of 6× SSC, 0.1% Blotto, 0.5% SDS, 1 mM EDTA (pH 8.0), 10 mM Na phosphate buffer (pH 7.2), 0.1 mg/ml of sonicated calf thymus DNA, and 0.1 mg/ml of sonicated human placental DNA. We have observed recently an improved signal-to-noise ratio using high concentrations of SDS in hybridization solutions (Church and Gilbert 1984). Hybridization was performed overnight at 65°C in the same solution with ~2 × 10^8 cpm of the labeled and blocked probes. Filters were washed three times, 20 min each, in 1× SSC, 0.1% SDS at 65°C. Autoradiography was performed for 2–3 days at −70°C with Biomax film (Kodak).

DNA Sequence Analysis

Vector primers flanking the insert were used to amplify cDNA clones by PCR for direct sequencing. Plasmid inserts were amplified using standard M13For (5′-CACGACGTTGTAAAACGACG-3~) and M13Rev (5′-ATTTCACACAGGAAACAGCT-3′) vector primers present in both the pAMP10 (CH5SP) and L-BA (NIB) vectors. Reactions were performed in a Perkin Elmer Cetus 9600 thermal cycler as follows: 30 cycles of 30 sec at 94°C, 30 sec at 60°C, and 30 sec at 72°C. Reaction conditions were: 1× Perkin Elmer Cetus PCR buffer, 0.25 mM dNTPs, 1.0 µM each primer, and 1 unit *Taq* polymerase (Perkin Elmer Cetus) in a total volume of 25 µl. PCR products were separated by agarose gel electrophoresis and the excised fragments were purified using QIAquick spin columns (Qiagen). Sequence analysis of the purified template was performed using dye-labeled primers [M13(-21) and M13RP] on an ABI 373 automated sequencer (Applied Biosystems, ABI). Where necessary, additional sequencing was performed by primer walking using dye terminator kits (ABI). Sequence assembly was performed using Sequencher 3.0 (Gene Codes). Sequences were analyzed using the BLASTN and BLASTX e-mail servers and databases at National Center for Biotechnology Information (NCBI) (Altschul et al. 1990). Crosswise PCR using primers derived from CH5FBA1 (~0.5 kb) and NIBE4 (~1.8 kb) led to the identification of an ~1.3 kb junction fragment (Fig. 4A). This junction fragment was sequenced and combined with the sequences of CH5FBA1 and NIBE4 to form a ~3.6 kb composite cDNA, designated CSA1. The 5′ end of the ~3.6-kb CSA1 cDNA was found to overlap with the 3′ ends of three ESTs deposited in GenBank. These ESTs were the result of mispriming from an AT-rich region of the cDNA rather than the true poly(A) tail. The corresponding 5′ ends of these ESTs were derived from the database, and primers from these, in combination with primers from CSA1, were used to close the gap between these sequences. The resulting fragment was directly sequenced to obtain a total length of ~5.4 kb for the entire extended cDNA. The 15 novel sequences have been deposited in GenBank under accession nos. U52827–U52840.

Mapping of cDNA Fragments

PCR reactions were performed as described above with 100 ng of purified YAC (Philippsen et al. 1991) or somatic cell hybrid DNA (Sambrook et al. 1989). The YACs and hybrids that define the CDCCR have been characterized previously (Goodart et al. 1994; Overhauser et al. 1994). PCR primer pairs used for mapping are as follows: CSA1 5′ (CACCAACCCGTGTCCTGAG and CATTTCGATTCTCTGTCTTCC), CSA1 3′ (NIBE4) (CACTGACAGTTTACAGGAAG and GTTTCACCAAAGCTACATGTC), CSA9 (CACCTGTACATCACGCCGC and GGTGTGATGCATGGGTGCTG), CSB3 (GGTCCACACCAATGGACGG and TCCACAGTGTCCCTTAACAC), CSB7 (AGATCCTTAGTGGCTCACTG and GTTGGTTCAAGGTGAGCTCTA), CSB9 (AGAGCTCTGCTAGCGTCGTG and GACGAGCTCACCACAACCAG), CSB10 (GACATTAGTTCCATGAGCCA and ATGTGAGCTAAGCTCTCTGC), CSB11 (GGAGCTGCAGACAGGTGACT and AATCTGGCTTCCATCTCTCC), CSC3 (GAACTCTGCACATGAGGCTG and ACATTGTCTACTGGCAGAGC), CSC5 (AATGTTCCAAATGCCTCCAA and GAAACTAGGAAATACGCAGTC), CSC8 (CTCTCAGCTTCAACCTTGCT and TGCTGGCACTATCTAGCTCC), CSC12 (GTTATGGCTCTCACAGAGCA and GTTTGAGGAAGGAGCAGTGT), CSD3 (AAAACACTTCACCTCTCAGG and GGAGATCACAATGACTGATT), NIBA2 5′ (GCATCTCAGGTCCTCAACAG and CTTGCTGAGCGGTTGTTGG), NIBA2 3′ (CATCTAAAATTTTAGGCAGTGAA and CTAAACAAGGCATTCAGCAG), and NIBB11 (CTTCTCTAAGACCACTCATAC and TAGGTGTATCCTAAAGAAGAG). All oligonucleotides are listed 5′ to 3′.

Redundancy and Northern Blot Analysis

The NIB cDNA pool used for the PCR was prepared by plating ~50,000 clones on LB agar containing ampicillin (100 µg/ml) and growing the bacteria overnight at 37°C. Colonies were then scraped off the agar and DNA was prepared by use of the alkaline lysis method (Sambrook et al. 1989). PCR amplification was performed using the primer pairs and conditions described above. Hybridization of the four NIB cDNAs was performed under the following conditions: 6× SSC, 0.5% SDS, 1 mM EDTA (pH 8.0), 10 mM Na phosphate buffer (pH 7.2), 0.1% Blotto, and 0.1 mg/ml of sonicated salmon sperm DNA. Hybridization was performed overnight at 65°C and two 65°C washes were performed, 15 min each, in 0.1% SDS, 0.1× SSC. Probes corresponding to the four NIB cDNAs were generated by PCR amplification of the cDNA inserts using flanking M13 vector primers as described above. The PCR products were separated by agarose gel electrophoresis, purified using QIAquick spin columns (Qiagen), and labeled using Rediprime Kits (Amersham). DAP-1 represented 0.17% ($n = 8$) and 0.07% ($n = 3$) of the CH5SP HeLa and placental cDNA libraries, respectively. Northern blots MTN I and II (Clontech)

were hybridized and washed as recommended by the manufacturer.

ACKNOWLEDGMENTS

We are grateful to Dr. Anne Bowcock for her critical reading of the manuscript. We thank M.B. Soares for generously providing an aliquot of purified DNA from his NIB library. This work was supported by grants HG00368 to M.L. and HG00236 to J.O. from the National Institutes of Health. A.S. was supported by a National Institutes of Health training grant in genomic science HG00038.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Meyer, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Berry, R., T.J. Stevens, N.A. Walter, A.S. Wilcox, T. Rubano, J.A. Hopkins, J. Weber, R. Goold, M.B. Soares, and J.M. Sikela. 1995. Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nature Genet.* **10:** 415–423.

Bornstein, P. 1992. Thrombospondins: Structure and regulation of expression. *FASEB J.* **6:** 3290–3299.

Bornstein, P. and E.H. Sage. 1994. Thrombospondins. *Methods Enzymol.* **245:** 62–85.

Briggs, M.D., S.M. Hoffman, L.M. King, A.S. Olsen, H. Mohrenweiser, J.G. Leroy, G.R. Mortier, D.L. Rimoin, R.S. Lachman, E.S. Gaines, J.A. Cekleniak, R.G. Knowlton, and D.H. Cohn. 1995. Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. *Nature Genet.* **10:** 330–336.

Carle, G.F. and M.V. Olson. 1984. Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nucleic Acids Res.* **12:** 5647–5664.

Catimel, B., L. Leung, H. el Ghissasi, N. Mercier, and J. McGregor. 1992. Human platelet glycoprotein IIIb binds to thrombospondin fragments bearing the C-terminal region, and/or the type I repeats (CSVTCG motif), but not to the N-terminal heparin-binding region. *Biochem. J.* **284:** 231–236.

Church, G.M. and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81:** 1991–1995.

Deiss, L.P., E. Feinstein, H. Berissi, O. Cohen, and A. Kimchi. 1995. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. *Genes & Dev.* **9:** 15–30.

Del Mastro, R.G., L. Wang, A.D. Simmons, T.D. Gallardo, G.A. Clines, J.A. Ashley, C.J. Hilliard, J.J. Wasmuth, J.D. McPherson, and M. Lovett. 1995. Human chromosome-specific cDNA libraries: New tools for gene identification and genome annotation. *Genome Res.* **5:** 185–194.

Dixit, V.M., S.W. Hennessy, G.A. Grant, P. Rotwein, and W.A. Frazier. 1986. Characterization of a cDNA encoding the heparin and collagen binding domains of human thrombospondin. *Proc. Natl. Acad. Sci.* **83:** 5449–5453.

Frazier, W.A. 1991. Thrombospondins. *Curr. Opin. Cell Biol.* **3:** 792–799.

Gardiner, K. 1995. Human genome organization. *Curr. Opin. Genet. Dev.* **5:** 315–322.

Gersh, M., S.A. Goodart, L.M. Pasztor, D.J. Harris, L. Weiss, and J. Overhauser. 1995. Evidence for a distinct region causing a cat-like cry in patients with 5p deletions. *Am. J. Hum. Genet.* **56:** 1404–1410.

Goodart, S.A., A.D. Simmons, D. Grady, K. Rojas, R.K. Moyzis, M. Lovett, and J. Overhauser. 1994. A yeast artificial chromosome contig of the critical region for cri-du-chat syndrome. *Genomics* **24:** 63–68.

Goundis, D. and K.B. Reid. 1988. Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs. *Nature* **335:** 82–85.

Guo, N.H., H.C. Krutzsch, E. Negre, T. Vogel, D.A. Blake, and D.D. Roberts. 1992. Heparin- and sulfatide-binding peptides from the type I repeats of human thrombospondin promote melanoma cell adhesion. *Proc. Natl. Acad. Sci.* **89:** 3040–3044.

Hecht, J.T., L.D. Nelson, E. Crowder, Y. Wang, F.F. Elder, W.R. Harrison, C.A. Francomano, C.K. Prange, G.G. Lennon, M. Deere, and J. Lawler. 1995. Mutations in exon 17B of cartilage oligomeric matrix protein (COMP) cause pseudoachondroplasia. *Nature Genet.* **10:** 325–329.

Hughes, A.E., D. McGibbon, E. Woodward, J. Dixey, and M. Doherty. 1995. Localisation of a gene for chondrocalcinosis to chromosome 5p. *Hum. Mol. Genet.* **4:** 1225–1228.

Kobayashi, S., F. Eden-McCutchan, P. Framson, and P. Bornstein. 1986. Partial amino acid sequence of human thrombospondin as determined by analysis of cDNA clones: Homology to malarial circumsporozoite proteins. *Biochemistry* **25:** 8418–8425.

LaBell, T.L. and P.H. Byers. 1993. Sequence and characterization of the complete human thrombospondin 2 cDNA: Potential regulatory role for the 3′ untranslated region. *Genomics* **17:** 225–229.

Lawler, J. and R.O. Hynes. 1986. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J. Cell Biol.* **103:** 1635–1648.

Lennon, G., C. Auffray, M. Polymeropoulos, and M.B.

Soares. 1996. The I.M.A.G.E. consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33:** 151–152.

Longmire, J.L., N.C. Brown, L.J. Meincke, M.L. Campbell, K.L. Albright, J.J. Fawcett, E.W. Campbell, R.K. Moyzis, C.E. Hildebrand, G.A. Evans, and L.L. Deaven. 1993. Construction and characterization of partial digest DNA libraries made from flow-sorted human chromosome 16. *Genet. Anal. Tech. Applic.* **10:** 69–76.

Lovett, M. 1994. Direct selection of cDNAs using genomic contigs. In *Current protocols in human genetics* (ed. J. Seidman), pp. 6.3.1–6.3.15. Wiley Interscience, New York, NY.

Lovett, M., J. Kere, and L.M. Hinton. 1991. Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci.* **88:** 9628–9632.

Niebuhr, E. 1978a. The cri du chat syndrome. *Hum. Genet.* **44:** 227–275.

———. 1978b. Cytologic observations in 35 individuals with a 5p-karyotype. *Hum. Genet.* **42:** 143–156.

Olsen, A.S., J. Combs, E. Garcia, J. Elliott, C. Amemiya, P. deJong, and G. Threadgill. 1993. Automated production of high-density cosmid and YAC colony filters using a robotic workstation. *BioTechniques* **14:** 116–117.

Overhauser, J., A.L. Beaudet, and J.J. Wasmuth. 1986. A fine structure physical map of the short arm of chromosome 5. *Am. J. Hum. Genet.* **39:** 562–572.

Overhauser, J., X. Huang, M. Gersh, W. Wilson, J. McMahon, U. Bengtsson, K. Rojas, M. Meyer, and J.J. Wasmuth. 1994. Molecular and phenotypic mapping of the short arm of chromosome 5: Sublocalization of the critical region for the cri-du-chat syndrome. *Hum. Mol. Genet.* **3:** 247–252.

Parimoo, S., R.P. Sankhavaram, S. Hridayabhiranjan, D.D. Chaplin, and S.M. Weissman. 1991. cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci.* **88:** 9623–9627.

Philippsen, P., A. Stotz, and C. Scherf. 1991. DNA of *Saccharomyces cerevisiae*. In *Guide to yeast genetics and molecular biology* (ed. C. Guthrie and G.R. Fink), pp. 169–182. Academic Press, New York, NY.

Rich, K.A., F.W. George, J.L. Law, and W.J. Martin. 1990. Cell-adhesive motif in region II of malarial circumsporozoite protein. *Science* **249:** 1574–1577.

Robson, K.J., J.R. Hall, M.W. Jennings, T.J. Harris, K. Marsh, C.I. Newbold, V.E. Tate, and D.J. Weatherall. 1988. A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* **335:** 79–82.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sargent, C.A., I.J. Chalmers, M. Leversha, and N.A. Affara. 1994. A rearrangement on chromosome 5 of an expressed human beta-glucuronidase pseudogene. *Mamm. Genome* **5:** 791–796.

Simmons, A.D., S.A. Goodart, T.D. Gallardo, J. Overhauser, and M. Lovett. 1995. Five novel genes from the cri-du-chat critical region isolated by direct selection. *Hum. Mol. Genet.* **4:** 295–302.

Soares, M.B., M.F. Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci.* **91:** 9228–9232.

Thompson, T.G., C.J. DiDonato, L.R. Simard, S.E. Ingraham, A.H. Burghes, T.O. Crawford, C. Rochette, J.R. Mendell, and J.J. Wasmuth. 1995. A novel cDNA detects homozygous microdeletions in greater than 50% of type I spinal muscular atrophy patients. *Nature Genet.* **9:** 56–62.

Weissman, S.M. 1987. Molecular genetic techniques for mapping the human genome. *Mol. Biol. Med.* **4:** 133–143.

Received September 16, 1996; accepted in revised form December 17, 1996.