

# Transcript Mapping in a 46-kb Sequenced Region at the Core of 12q13.3 Amplification in Human Cancers

Abdel G. Elkahlon,<sup>\*,1</sup> David B. Krizman,<sup>\*,1</sup> Zhili Wang,<sup>†</sup> Tiffany A. Hofmann,<sup>\*</sup>  
Bruce Roe,<sup>†</sup> and Paul S. Meltzer<sup>\*,2</sup>

<sup>\*</sup>Laboratory of Cancer Genetics, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892-4470; and <sup>†</sup>Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019-0390

Received December 6, 1996; accepted March 19, 1997

**We used a combination of sequence analysis and exon trapping in an effort to determine the complete transcript map for a cosmid (6E5) derived from 12q13.3, a region of DNA sequence amplification in human cancers. This cosmid, previously known to contain three genes (*CDK4*, *SAS*, and *OS9*), was sequenced, and that information was used for computer-assisted analysis. In addition, 6E5 was subjected to both internal and 3'-terminal exon-trapping protocols, and the results of these studies were used to guide cDNA cloning experiments. These studies demonstrate that this cosmid is derived from a remarkably gene-dense region and add two new transcripts (KIAA0167 and 6E5.2) to the list of sequences that are expressed in tumors bearing amplification of this region.** © 1997 Academic Press

## INTRODUCTION

DNA amplification in human cancers is an important manifestation of genomic instability. A large body of evidence suggests that emergence of amplification-bearing tumor clones is consequent to the growth advantage provided by specific overexpressed amplified genes. Cytologically, amplified DNA may exist either extrachromosomally as double minutes or intrachromosomally as homogeneously staining regions. Recent studies using molecular cytogenetic techniques, such as comparative genomic hybridization, fluorescence *in situ* hybridization, and chromosome microdissection, have demonstrated that domains of amplification can be structurally complex and include segments of DNA from several regions (Kallioniemi *et al.*, 1993; Guan *et al.*, 1994; Tanner *et al.*, 1996). These observations

clearly demonstrate the importance of high-resolution amplicon mapping in human cancers. With this knowledge and the availability of improved genome maps, it is possible to define the core regions consistently amplified in multiple tumors. These regions must still be examined in detail to determine their genetic content.

Sequences derived from chromosome 12q13–q15 contribute to complex amplifications frequently observed in sarcomas and brain tumors. Physical mapping studies have demonstrated that 12q13–q15 amplicons contain two separate core regions, one containing *MDM2* and the other containing *CDK4* (Elkahlon *et al.*, 1996b; Berner *et al.*, 1996; Reifemberger *et al.*, 1996). On the normal chromosome, these regions are separated by over 5 Mb, and markers in the intervening segment are usually not amplified. Most often, the *CDK4* and *MDM2* regions are coamplified, but individual tumors have been described with independent amplification of only one of these regions. Currently, because of their roles in regulating the cell cycle (*MDM2* through p53 and *CDK4* through pRb) these genes are considered the most probable target genes for 12q13–q15 amplification. However, because additional genes map close to *CDK4* (Krauter *et al.*, 1995; Elkahlon *et al.*, 1996a), we have continued to characterize the *CDK4* core region to identify all the genes within it.

An approximately 100-kb cosmid contig covering the *CDK4* core region has been developed by screening of a chromosome 12 cosmid library with a chromosome microdissection probe (Elkahlon *et al.*, 1996a). Within this contig, cosmid 6E5 contains *CDK4* and two additional genes of unknown function (*SAS* and *OS9*). To assist in gene identification, the complete sequence of cosmid 6E5 was determined and scanned for genes by computational methods. In parallel, 6E5 DNA was also subjected to exon trapping. This information was then used to guide cDNA cloning and expression analysis. Our results, which have resulted in the assignment of two novel transcripts to the *CDK4* region, also allow comparison of the capabilities and efficiencies of these methods for mammalian transcript mapping. We found

Sequence data from this article have been deposited with the GenBank Data Library under Accession No. U81031.

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed at NHGRI/NIH, Building 49, Room 4A10, 49 Convent Drive, Bethesda, Maryland 20892-4470. Telephone: (301) 402-2023. Fax: (301) 402-2040. E-mail: pmeltzer@nhgri.nih.gov.

that although each individual method was partially successful in identifying known or suspected novel transcribed sequence, the combination of these methodologies provides a significantly more comprehensive approach. Our results demonstrate that 6E5 is derived from a remarkably gene-dense region, and its amplification necessarily affects multiple genes.

## MATERIALS AND METHODS

**Sequencing and sequence analysis.** A shotgun library prepared from 6E5 cosmid DNA was sequenced using the dideoxynucleotide method as previously described (Bodenteich *et al.*, 1994). The GenBank Accession No. is U81031. The sequence was analyzed by the GRAIL2 and BLAST programs (Uberbacher and Mural, 1991; Altschul *et al.*, 1990). PowerBlast software was obtained from the National Center for Biotechnology information (ftp: ncbi.nlm.nih.gov cd/pub/sim2/PowerBlast).

**Exon trapping.** Both internal and 3'-terminal exon-trapping protocols were followed as previously described (Buckler *et al.*, 1991; Krizman and Berget, 1993; Church *et al.*, 1994; Krizman *et al.*, 1994). Briefly, cosmid DNA was prepared by the alkaline lysis method and digested with *EcoRI*, *BamHI*, and *BglII* restriction enzymes. Digested DNA was ligated to the 3'-terminal exon-trapping vector pTAG4 that was prepared as previously described for use in the direct ligation/transfection procedure unique to this protocol. The same digested DNA was used to prepare a plasmid library in the internal exon-trapping vector pSPL3, and recombinant plasmid DNA was transfected into Cos7 cells. Individual subclones were sequenced using the ABI dye-primer system according to the manufacturer's recommendations and were analyzed on an ABI 377 automatic sequencer.

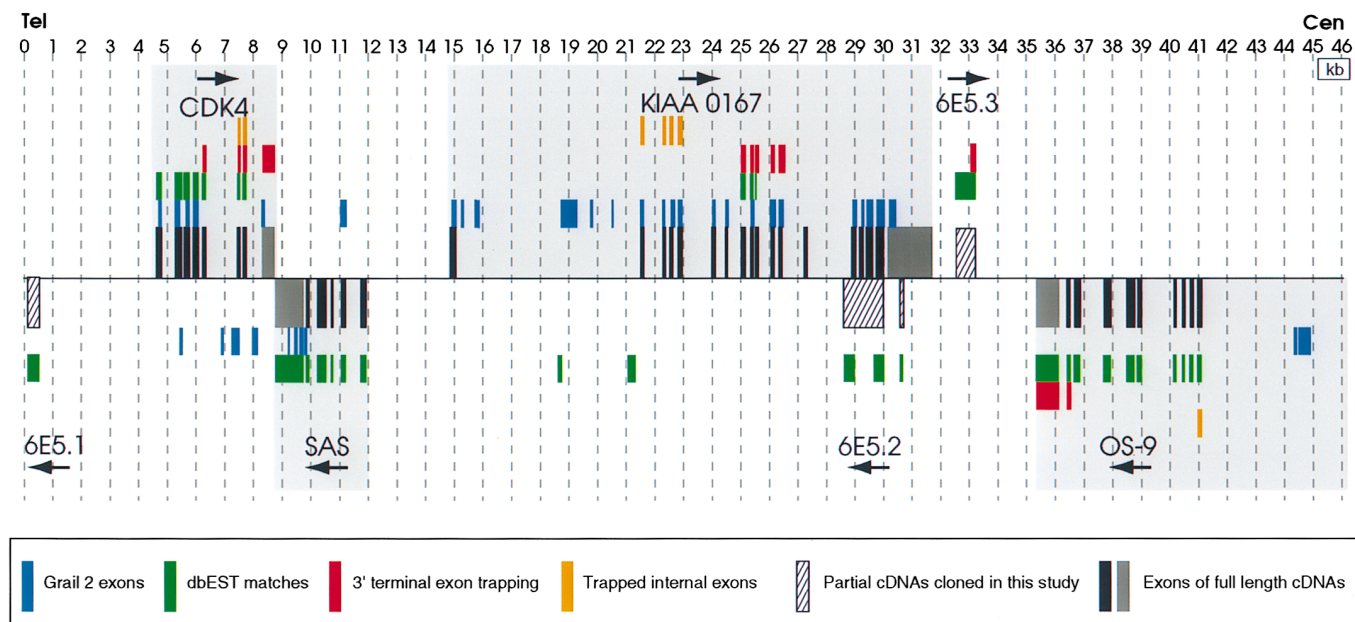
**cDNA library screening and expression analysis.** cDNAs representing ESTs were either obtained from the I.M.A.G.E. Consortium (Lennon *et al.*, 1996) or selected from a GeneTrapper fetal brain

cDNA library according to manufacturer's recommendations (Life Technologies, Inc.). The oligonucleotides used for library screening were CAGGTAACAAGTGGGGAACC (for EST T88828), CTTGGT-GTTGGGATTTGTTGTGAG (for EST M79186), and GAAAGCATG-GCTCAGAAGCTCTAACAC (for EST T84977). RNA extraction and Northern analysis were as previously described (Su *et al.*, 1996). Multiple tissue Northern blots and normal human muscle RNA were obtained from Clontech. NGP cells were provided by Garrett Brodeur (Children's Hospital of Philadelphia). OsA-CL and RMS-13 cells were provided by Tom Look (St. Jude Children's Research Hospital).

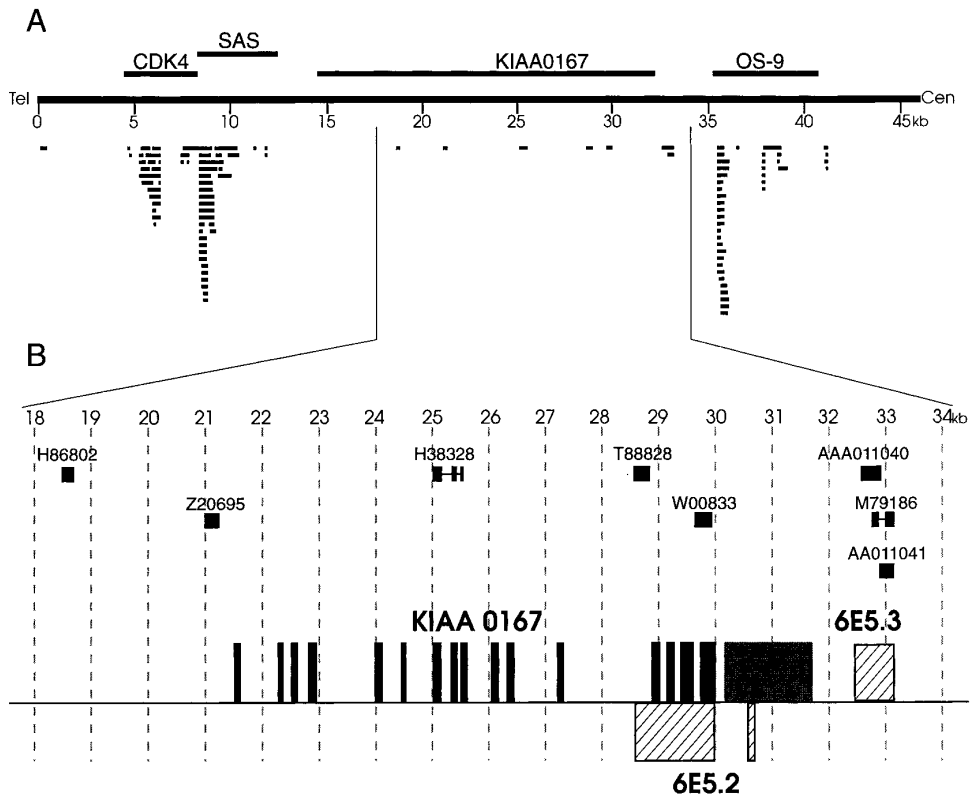
## RESULTS

### Database Searching

The sequence of cosmid 6E5 consists of 45,976 bp with a 53% G+C composition. Only 11.5% of the sequence is composed of repetitive elements. Based on previous mapping analysis, 6E5 was known to contain *CDK4*, *SAS*, and *OS9* (Elkahloun *et al.*, 1996a). The cosmid sequence is consistent with the previously reported genomic structures of *CDK4* and *SAS* and provides that information for most of *OS9*. Figure 1 illustrates that the 8 exons of *CDK4* span the region from base 4597 to 8732 and are transcribed toward the centromere (Zuo *et al.*, 1996). The 6 exons of *SAS* fall immediately centromeric to *CDK4* (base 8750–11929) and are transcribed toward the telomere (Jankowski *et al.*, 1995). The 3'-terminal exons of *CDK4* and *SAS* are separated by only 17 nucleotides. The *OS9* gene is found at the centromeric end of 6E5 terminating at base 35,350 with 10 exons present including the 3'-terminal exon (Su *et al.*, 1996) (Fig. 1). Based on the



**FIG. 1.** Transcript map of cosmid 6E5 based on computer-assisted DNA sequence analysis, exon trapping, and cDNA cloning. Confirmed exons of sequences represented in GenBank by full-length cDNAs are depicted in black (internal exons) and gray (3'-terminal exons). cDNAs isolated during this study are crosshatched. GRAIL2 predictions are blue, exact matches with dbEST entries are green, exons trapped by 3'-terminal exon trapping are red, and exons trapped by internal exon trapping are orange. The direction of transcription of each gene is indicated by an arrow. Features labeled above the line are oriented toward the centromere, and those below the line are oriented toward the telomere.



**FIG. 2.** Output of PowerBlast search of dbEST with the 6E5 sequence. **(A)** Overview of dbEST matches with gene positions indicated above the line and each matching dbEST entry indicated by a bar below the line. **(B)** Enlarged view of the middle section of the cosmid, which contains eight "singleton" ESTs. Orientation of ESTs is not specified. The scale is in kilobases. Sequences aligned with 6E5 in this diagram were identical with the cosmid sequence allowing for the error rate of single-pass EST sequencing.

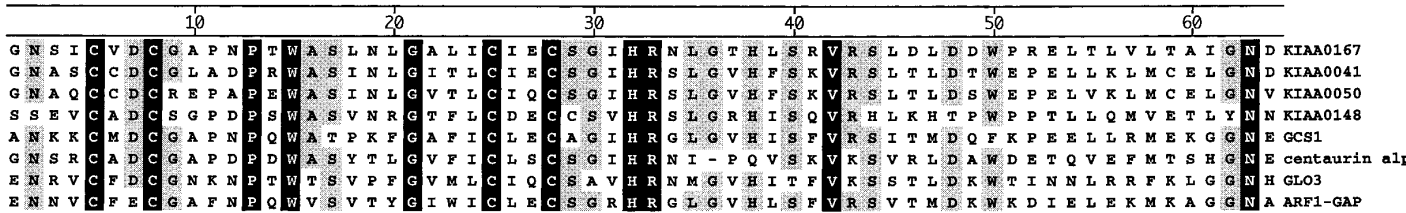
full-length cDNA sequence of *OS9*, it can be concluded that the promoter and the 5' 664 nucleotides of the *OS9* cDNA are located centromeric to 6E5. Considering the positions of these three genes, a small region from base 1 to 4597 and a larger region from base 11,900 to 35,350 contain no genes previously assigned to this cosmid.

PowerBlast, which masks repeats and provides a graphical output of BLAST searches, was used to search dbEST for the presence of ESTs that map to 6E5. The PowerBlast output is illustrated in Fig. 2, and those sequences present in dbEST are also shown in green in Fig. 1. All three known genes were represented by multiple ESTs from several tissues, suggesting that these genes are widely expressed at relatively high levels, and all 24 exons of these genes were represented in dbEST. When ESTs contained multiple exons, their splicing pattern was consistent with the previously determined gene structure. In addition to the three previously known genes, PowerBlast search of GenBank identified a full-length anonymous cDNA designated KIAA0167 (Accession No. D79989) isolated from myeloblast RNA (Nomura *et al.*, 1994) and not previously assigned to this region. This gene consists of 18 exons distributed from position 14,821 to 31,777 transcribed toward the centromere (Fig. 1). BLASTX analysis of this genomic region and BLASTP analysis

of the predicted peptide encoded by KIAA0167 identify a strong similarity in exon 16 to a highly conserved zinc finger-like motif found in several proteins from a wide variety of eukaryotic organisms (Fig. 3).

PowerBlast also identified several additional ESTs in the segments of 6E5 that lack previously known genes and that do not correspond to any of the four genes for which there are full-length cDNA sequences (Fig. 2). In the region telomeric to *CDK4*, there is a single EST (T84977) at positions 57 to 440. This EST was derived from the 5' end of a cDNA and most likely represents a gene transcribed in a telomeric direction and extending beyond 6E5. Seven ESTs were identified in the large segment from 11,900 to 35,350, five of which are nonoverlapping (discussed in detail below). The first is found at position 18,603 and the last occurs at position 32,821. Although this region is relatively poorly represented in dbEST, the presence of these ESTs suggests that additional transcribed sequences exist in this region of the cosmid.

The 6E5 sequence was also analyzed by GRAIL2 both to examine the ability of this program to identify the known genes and to scan for potentially novel transcribed sequences (Fig. 1). GRAIL2 identified 4/8 exons from *CDK4* and 15/18 exons from KIAA0167. In contrast, GRAIL2 located 0/10 exons from *OS9* and only 1/6 exons of *SAS*. Most of the exons identified by



**FIG. 3.** Multiple sequence alignment of the zinc finger-like motif of KIAA0167 with seven related cDNAs. Fully conserved amino acids are shaded in black, and residues present in at least six of eight sequences are shaded in gray. Multiple sequence alignment was generated with DNASTAR software. KIAA0041 (GenBank D26069), KIAA0050 (GenBank D30758), and KIAA0148 (GenBank D63482) are anonymous human myeloblast cDNAs. GCS1 (GenBank L24125) and GLO3 (GenBank X79514) are *Saccharomyces cerevisiae* genes. Centaurin alpha (GenBank U51013) and ARF1-GAP (GenBank U35776) are rat cDNAs.

GRAIL2 were not demarcated at the correct splice-site nucleotides. Fourteen additional exons were predicted, which do not correspond to any of the four genes whose full-length cDNAs are represented in GenBank.

#### Exon Trapping of 6E5 DNA

The search for transcribed sequences was extended by exon-trapping experiments (Fig. 1). 3'-terminal exons were trapped from *Eco*RI, *Bam*HI, and *Bg*II restriction digests of 6E5 using trapping vector pTAG4 for the direct ligation/transfection protocol as previously described (Krizman and Berget, 1993; Krizman *et al.*, 1994). The RT/PCR product, which demonstrated multiple bands on agarose gel electrophoresis ranging in size from 300 to 1200 bp (data not shown), was subcloned into a plasmid vector. Sixteen subclones containing inserts of varying sizes were sequenced. Analysis of these sequences showed that they were derived from three genes, *CDK4* (8/16 clones), KIAA0167 (1/16 clones), and *OS9* (2/16 clones), with some subclones containing several upstream internal exons spliced to the 3'-terminal exon. No clone containing *SAS* sequences was identified. The clone containing KIAA0167 sequences contained exons 8–12 and utilized a polyadenylation signal in intron 12. An additional 3' trapping product represented by a single clone (DKG2) was transcribed toward the centromere, terminating at position 33,214. Background in the 3'-terminal exon trapping consisted of trapped cosmid vector (2/16 clones), 1 clone containing an *Alu* repeat (derived from bases 14,140 to 14,195), and 1 clone derived entirely from the trapping vector.

Internal exon trapping was performed on the same restriction digests used for terminal exon trapping (Buckler *et al.*, 1991; Krizman *et al.*, 1994). RT/PCR results demonstrated many intensely staining bands on an ethidium bromide/agarose gel ranging from 110 to 700 bp in size (data not shown). The RT/PCR product was size-fractionated, fragments greater than 177 bp were cloned, and 16 clones containing inserts of varying size were sequenced. These results demonstrated that 7/16 clones corresponded to KIAA0167 (exons 2,3,4, and 5), 2/16 clones corresponded to *CDK4* (exons 6 and 7), and 1/16 represented *OS9*. There were no exons recovered from *SAS* or from any of the ESTs that do

not overlap these four genes. Altogether, 10/16 clones were derived from known internal exons. False positives included 3/16 subclones containing an *Alu* repeat at base 42,830 to 42,880 and 3/16 subclones derived from the trapping vector.

#### cDNA Cloning and Northern Analysis

To clarify the significance of the ESTs that were poorly represented in dbEST, we used a combination of RT/PCR, cDNA cloning, and Northern blot analysis. Greater attention was given to those that were expressed in tumor cell lines with 6E5 amplification. In the small telomeric segment of 6E5 with no known exons, PowerBlast identified a single 5' EST (T84977). An oligonucleotide based on this EST was used to screen a fetal brain cDNA library, and a 1.8-kb cDNA was isolated. The 5' sequence of this cDNA was identical with the EST and 6E5 genomic sequence, but the 3' end was not present in 6E5, suggesting that this EST represents a fragment of a gene, which we have designated 6E5.1, with its terminus located telomeric to 6E5. However, no transcript was identified using this probe on Northern blots of cells carrying amplification of 6E5 (data not shown), and further attempts to isolate a full-length cDNA were not pursued.

The region between *SAS* and *OS9* contains one full-length cDNA (KIAA0167) and several ESTs (Fig. 2B). In addition, a 3'-terminal exon-trapping product (clone DKG2) was recovered from this region. Considering these sequences from telomere to centromere, the first encountered is a 3' EST (H86802). This sequence aligns with a contiguous region of the first intron of KIAA0167 and is present once in dbEST. Transcriptional orientation is opposite to that of KIAA0167. Although described as a 3' EST, the sequence lacks a polyadenylation signal and is very GC rich. The next EST encountered (Z20695) also falls in the first intron of KIAA0167 and, as in the previous case, is unspliced and represented only once in dbEST. Its orientation is not described. Probes for both of these segments of 6E5 failed to identify highly expressed transcripts by Northern hybridization to RNA from cells carrying amplification of this region (data not shown), and the significance of these two ESTs is indeterminate.

The next EST encountered (H38328) represents the

correctly spliced exons 8, 9, and 10 of KIAA0167. No ESTs are present until intron 13 of KIAA0167 where a single 3' EST (T88828) is found oriented opposite to KIAA0167. This sequence is unspliced and extends into exon 14 of KIAA0167. Because of this overlap, an oligonucleotide corresponding to T88828 but falling outside the overlap with KIAA0167 was utilized to screen a cDNA library. We identified and sequenced a 1.3-kb cDNA clone (AB9), the 3' end of which overlaps with T88828 (Fig. 2B). This clone contains a single splice and is transcribed in the orientation opposite to that of KIAA0167. At its 5' end, AB9 overlaps an additional EST (W00833) which contains the same splice. Transcriptional orientation of this clone is also opposite to that of KIAA0167. When hybridized to a Northern blot, the AB9 cDNA identifies a 1.3-kb transcript designated 6E5.2, which is strongly expressed in cell lines amplified for this sequence and is distinct in size from the transcript identified by a probe for exons 2 through 13 of KIAA0167 (Fig. 4). However, BLASTX analysis of this region of 6E5 in the orientation of 6E5.2 does not identify any significant protein similarities.

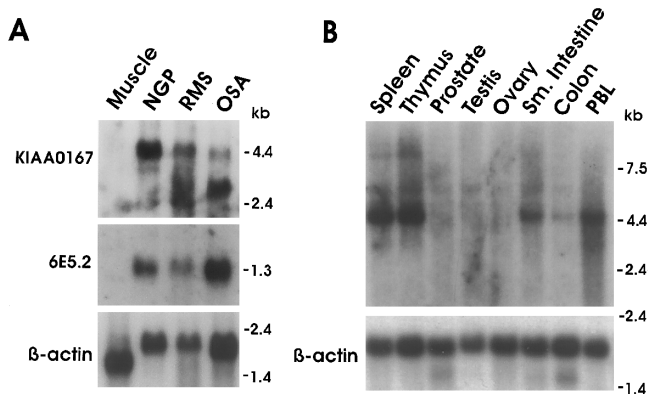
Three ESTs are identified between the terminal exons of KIAA0167 and *OS9* (Fig. 2B). These two clones, ESTs A011040, A11041 (representing opposite ends of the same clone), and M79186, overlap but differ in that M79186 contains a splice absent from A01140. The 3' termini of these ESTs fall at precisely the same point as the clone DKG2 isolated by 3'-terminal exon trapping. In addition, we utilized sequences from M79186 to isolate an 850-bp cDNA (AEK14), which overlaps

A011040 and also lacks the splice present in M79186. This transcript was designated 6E5.3 (Fig. 2B). Like 6E5.2, 6E5.3 generates no strong similarities on BLASTX searches. Northern hybridization of AEK14 failed to detect a transcript in cells carrying amplification of this sequence or in normal muscle, spleen, or pancreas (data not shown).

## DISCUSSION

In the process of characterizing a region of gene amplification in human cancer, we were presented with an opportunity to directly compare and contrast various DNA-based transcript mapping methodologies. This cosmid is derived from a remarkably gene-dense region of low repeat content organized so that adjacent genes are transcribed from opposite strands with very little intergenic DNA. We found that PowerBlast analysis of dbEST was the most useful single tool to apply to the genomic sequence. Using PowerBlast we were able to identify all 24 exons of the three previously known genes with multiple ESTs. Remarkably, no gene was identified by any method that was not at least partially represented in dbEST. GRAIL2 predicted 21/42 exons of the four genes with full-length cDNAs. Many sequences marked by GRAIL2 (including some scored as excellent) remained unconfirmed by other methods, leaving considerable uncertainty as to their significance.

In general, the significance of "singleton" entries in dbEST is difficult to assess (Boguski *et al.*, 1993). When multiple ESTs with the same sequence are present, it can be assumed with confidence that they represent authentic genes. Unique dbEST entries pose a greater problem and might be due to cloning artifacts of various types. There were nine such ESTs present in 6E5 (derived from eight cDNA clones). Two (H86802 and Z20695) fall in the first intron of KIAA0167 and could not be linked to any of the other transcripts. One corresponded to KIAA0167, one to 6E5.1, two to 6E5.2, and three to 6E5.3. Because we were able to isolate cDNAs corresponding to three of these singletons (6E5.1, 6E5.2, and 6E5.3), it is likely that they do indeed represent transcribed sequences. However, little more can be concluded regarding their significance since none exhibits high similarity to any known protein, and all except 6E5.2 are expressed at low levels even when amplified. Full characterization of these three transcripts will require the isolation of full-length cDNAs and more extensive surveys of their expression. Most likely, their low representation in dbEST is related to low mRNA abundance in the tissues used for EST library construction. It is also conceivable that, in some cases, singleton ESTs represent a form of noise in the transcriptional machinery. The utilization of cryptic promoters might produce nonfunctional transcripts that appear in dbEST as low-frequency entries. Distinguishing these possibilities is impossible without experimental characterization of the relevant cDNAs. In



**FIG. 4.** (A) Northern blot analysis of KIAA0167 and 6E5.2 in normal muscle RNA and in three cell lines (NGP neuroblastoma, RMS13 rhabdomyosarcoma, and OSA osteosarcoma) previously demonstrated to contain amplification of 6E5 (Elkahloun *et al.*, 1996b). Both transcripts are readily detected in the 12q-amplified cell lines but not in normal muscle. Multiple bands are detected with the KIAA0167 probe (an exon 2–13 RT/PCR product) suggestive of alternative processing. The larger band corresponds closely in size to the 3.95-kb KIAA0167 cDNA. The 1.3-kb transcript recognized by the 6E5.2 probe (cDNA AB9) corresponds in size to the largest cDNA isolated for this transcript. (B) Multiple tissue Northern blot analysis of KIAA0167 demonstrates a tissue-specific pattern of gene expression with higher levels of expression in spleen, thymus, small intestine, and peripheral blood leukocytes. Both blots were normalized for loading by hybridization with a  $\beta$ -actin probe.

the present case, we were primarily interested in searching for transcripts that were overexpressed in cells with amplification of this region, and we were able to exclude high levels of expression associated with gene amplification for all the singleton ESTs except those that correspond to 6E5.2 and KIAA0167.

With regard to the laboratory methods used in this study, 3'-terminal exon trapping correctly identified two previously known genes (*CDK4* and *OS9*). This method also identified KIAA0167, but utilization of a polyadenylation signal in intron 12 provided misleading localization of the 3' end of this gene. In addition, 8 internal exons were trapped simultaneously with these 3'-terminal exons. A 3'-terminal exon (DKG2) corresponding to 6E5.3 was also isolated, providing independent confirmation of a functional exon in this region. However, 6E5.3 was minimally expressed in the cells that were tested, and its significance is uncertain. Conceivably it represents an alternative 3' terminus for KIAA0167 in certain tissues. In this regard, Northern blot analysis with the upstream KIAA0167 probe demonstrated multiple bands suggestive of alternate splicing or termination (Fig. 4). Internal exon-trapping products represented 7/38 internal exons adding 5 exons to those identified by 3'-terminal trapping. The additional steps of cDNA cloning, RT/PCR, and Northern blot analysis were necessary to confirm and integrate the results of database searching and exon trapping.

In relation to the phenomenon of 12q13.3 amplification in cancers, the region of 6E5 that contains KIAA0167 and 6E5.2 is of the greatest interest. Both of these cDNAs are expressed in the 12q-amplified cell lines that we studied, and neither had been previously recognized as falling within the 12q13.3 amplicon. The presence of 6E5.2 overlapping the highly conserved region of KIAA0167 and transcribed in the opposite orientation is remarkable. The 6E5.2 transcript was abundant in all the tumor cell lines with amplification of this region. However, there is no apparent similarity of AB9 to any known protein. In contrast, the overlapping region of KIAA0167 contains a highly conserved zinc finger-like domain (Fig. 3) and is expressed in a highly tissue-specific fashion. Although several of the sequences containing this motif are anonymous, it was first described in the *Saccharomyces cerevisiae* protein GCS1, which is essential for the G0 to G1 transition in yeast (Ireland *et al.*, 1994). Functional analysis of the KIAA0167 protein will be necessary to determine the potential effect of its presence in tumor cells.

In light of our results, several conclusions are possible in relation to the phenomenon of gene amplification in cancer in general and to 12q13.3 amplification in particular. When a gene-dense region such as this undergoes increase in copy number, it is inevitable that multiple genes will be coamplified. Five transcripts (*CDK4*, *SAS*, KIAA0167, 6E5.2, and *OS9*) are expressed when the genomic segment corresponding to 6E5 is amplified. Although several additional se-

quences were identified by their presence in dbEST or by exon trapping, none of these are highly expressed when amplified. These observations emphasize the complexity of the disturbance of gene expression associated with gene amplification and the importance of supplementing sequence-based analysis of amplified regions with expression data.

## ACKNOWLEDGMENTS

This work was supported in part by NCHGR HG00313 to Bruce Roe. The assistance of Darryl Leja with the illustrations is gratefully acknowledged.

## REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Berner, J. M., Forus, A., Elkahlon, A., Meltzer, P. S., Fodstad, O., and Mykkelbost, O. (1996). Separate amplified regions encompassing *CDK4* and *MDM2* in human sarcomas. *Genes Chromosomes Cancer* **17**: 254–259.
- Bodenteich, A., Chissoe, S., Wang, Y. F., and Roe, B. A. (1994). Shotgun cloning as the strategy of choice to generate templates for high throughput dideoxynucleotide sequencing. In "Automated DNA Sequencing and Analysis Techniques" (M. D. Adams, C. Fields, and J. C. Venter, Eds.), pp. 42–50, Academic Press, London.
- Boguski, M. S., Lowe, T. M., and Tolstoshev, C. M. (1993). dbEST—Database for "expressed sequence tags." *Nature Genet.* **4**: 332–333.
- Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88**: 4005–4009.
- Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A., and Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6**: 98–105.
- Elkahlon, A. G., Meltzer, P. S., Guan, X. Y., McNinch, J. S., Trent, J. M., and de Jong, P. J. (1996a). Isolation of a cosmid sublibrary for a region of chromosome 12 frequently amplified in human cancers using a complex chromosome microdissection probe. *Genomics* **31**: 343–347.
- Elkahlon, A. G., Bittner, M., Hoskins, K., Gemmill, R., and Meltzer, P. S. (1996b). Molecular cytogenetic characterization and physical mapping of 12q13–15 amplification in human cancers. *Genes Chromosomes Cancer* **17**: 205–214.
- Guan, X. Y., Meltzer, P. S., Dalton, W. S., and Trent, J. M. (1994). Identification of cryptic sites of DNA sequence amplification in human breast cancer by chromosome microdissection. *Nature Genet.* **8**: 155–161.
- Ireland, L. S., Johnston, G. C., Drobot, M. A., Dhillon, N., DeMaggio, A. J., Hoekstra, M. F., and Singer, R. A. (1994). A member of a novel family of yeast "Zn-finger" proteins mediates the transition from stationary phase to cell proliferation. *EMBO J.* **13**: 3812–3821.
- Jankowski, S. A., DeJong, P., and Meltzer, P. S. (1995). Genomic structure of *SAS*, a member of the transmembrane 4 superfamily amplified in human sarcomas. *Genomics* **25**: 501–506.
- Kallioniemi, O. P., Kallioniemi, A., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. (1993). Comparative genomic hybridization: A rapid new method for detecting and mapping DNA amplification in tumors. *Semin. Cancer Biol.* **4**: 41–46.
- Krauter, K., Montgomery, K., Yoon, S. J., LeBlanc-Straceski, J., Renault, B., Marondel, I., Herdman, V., Cupelli, L., Banks, A., Lie-

- man, J., Menninger, J., Bray-Ward, P., Nadkarni, P., Weissenbach, J., Le Paslier, D., Rigault, P., Chumakov, I., Cohen, D., Miller, P., Ward, D., and Kucherlapati, R. (1995). A second-generation YAC contig map of human chromosome 12. *Nature* **377**: 321–333.
- Krizman, D. B., and Berget, S. M. (1993). Efficient selection of 3'-terminal exons from vertebrate DNA. *Nucleic Acids Res.* **21**: 5198–5202.
- Krizman, D. B., Nisson, P. E., and Watkins, P. C. (1994). Isolation of exons from cloned DNA by exon trapping. In "Protocols in Human Genetics" (A. Boyle, Ed.), Greene Publ. Wiley, New York.
- Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996). The I.M.A.G.E. Consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33**: 151–152.
- Nomura, N., Nagase, T., Miyajama, N., Sazuka, T., Tanaka, A., Sato, S., Seki, N., Kawarabayasi, Y., Ishikawa, K., and Tabata, S. (1994). Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041–KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* **1**: 223–229.
- Reifenberger, G., Ichimura, K., Reifenberger, J., Elkahoulou, A. G., Meltzer, P. S., and Collins, V. P. (1996). Refined mapping of 12q13–q15 amplicons in human malignant gliomas suggests *CDK4*, *SAS* and *MDM2* as independent amplification targets. *Cancer Res.* **56**: 5141–5145.
- Su, Y. A., Hutter, C. M., Trent, J. M., and Meltzer, P. S. (1996). Complete sequence analysis of a gene (OS-9) ubiquitously expressed in human tissues and amplified in sarcomas. *Mol. Carcinogen.* **15**: 270–275.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjarvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi, O. P. (1996). Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res.* **56**: 3441–3445.
- Uberbacher, E. C., and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor–neural network approach. *Proc. Natl. Acad. Sci. USA* **88**: 11261–11265.
- Zuo, L., Weger, J., Yang, Q., Goldstein, A. M., Tucker, M. A., Walker, G. J., Hayward, N., and Dracopoli, N. C. (1996). Germline mutations in the p16INK4a binding domain of *CDK4* in familial melanoma. *Nature Genet.* **12**: 97–99.