J. Appl. Genet. 39(2), 1998, pp. 199-204

Unidirectional orientation of twelve expressed tagged sites within 40 kb of human chromosomal region 22q13.1

Carsten PUSCH[1], Zhili WANG[2], Bruce ROE[2], Nikolaus BLIN[1]

[1]Institute of Anthropology and Human Genetics, University of Tübingen, Germany
[2]Department of Chemistry and Biochemistry, The University of Oklahoma, Norman, OK, USA

Abstract. The single copy sequence D22S16 from human chromosomal region 22q13.1 that carries a putative conserved gene, was used to probe a chromosome 22-specific cosmid library. Genomic sequencing of one positive, 40 kb long cosmid (C1155) revealed a hereto unmapped gene (a subunit of DNA-dependent RNA polymerase II, POLR2F), a SOX9-related sequence and 12 expressed sequence tags. Although not parts of one consecutive gene, all 12 ESTs and, in addition, the polymerase gene are oriented in the same transcriptional direction within the genomic sequence represented by cosmid C1155.

Key words: chromosome 22, cosmid, EST, mapping, PCR, sequencing.

Introduction

Genetic changes on the human chromosome 22 have been linked to a variety of congenital and neoplastic disorders. For a set of such disorders, specific genes were assigned, others were merely linked to chromosomal regions. One of them is the "meningioma chromosomal region" (MGCR) in 22q13.1-qter and further mapping efforts are required to elucidate the concerned genomic sequences. Although the relative density of anonymous markers on chromosome 22 has been increasing over the past years (COLLINS et al. 1995), significant gaps still remain on this map, in particular in region 22q13. One of the loci mapped therein by us (D22S16), derived from a plasmid recombinant p22hom13E. Its 2.1kb EcoRI-BamHI subfragment displayed an open reading

Received: November 1997.
Correspondence: N. BLIN, Institute of Anthropology and Human Genetics, Wilhelmstr. 27, 72074 Tübingen, Germany, E-mail: blin@uni-tuebingen.de.

frame and species conservation (GÖTTERT et al. 1989), both strong indicators for a coding region. Thus, this recombinant was used to probe extended sequences within this particular chromosomal region.

Material and methods

Cosmid libraries

Cosmid c1155 was isolated from a chromosome 22 library containing approximately 10,000 cosmids (NIZETIC et al. 1991) screened with the DNA-probe D22S16 (GÖTTERT et al. 1989). The sequencing of the complete c1155 was performed via our previously described shotgun-based approach (BODENTEICH et al. 1993) (GenBank Accession No. L48815). FASTA searches at the EMBL database (Heidelberg) with the minimum limit setting of > 94% strand-matching yielded sequence information on a set of ORFs.

Filter hybridization (DNA, RNA)

For Southern and Northern blots, standard techniques were applied as published previously (BLIN et al. 1993). The DNA probes were labeled using a random primed DNA labeling kit (Boehringer Mannheim). For expression studies, total RNA from rat, mouse and human was isolated as published (GOUGH 1988) for Northern blots to be hybridized with fragments of the cosmid DNA.

PCR

The single copy nature of the ORFs was confirmed by deducing appropriate PCR primers. The cycling reaction had a volume of 50 µL and contained 50 mM Tris-HCl (pH 9.0), 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 200 M each deoxynucleotide, 1 µM each primer, 1 µL Taq polymerase (0.5 µL enzyme, 5 U/µL plus 7 µL 1 × buffer low Mg^{2+}) and 100 ng genomic DNA as a template. After a denaturing step for 5 min at 94°C and 35 cycles per 1 min at 94°C, 1 min at diverse annealing temperatures (54-61°C) and 1 min at 72°C, a final extension for 1min at 72°C was performed. The resulting PCR-products were analyzed on 5% polyacrylamide gels (1 × TBE-buffer).

Results

A chromosome 22 specific cosmid library was screened with a single copy subclone fragment of p22hom13E, representing D22S16. This hybridization

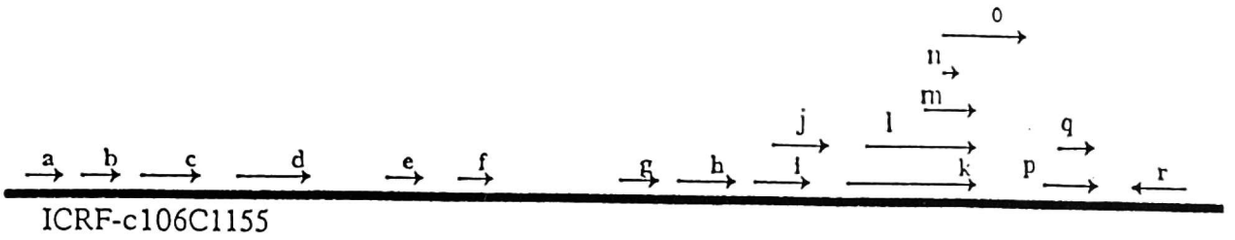

Figure 1. Linear map of C1155 showing 12 ORFs (a, g-q), the 5 exons of 14.4 kD subunit of RNA-Polymerase II (b-f) and a SOX9-related sequence (r). The arrows above indicate the direction of transcription.

The ESTs represent the following database accession numbers: a – r48447, b-f – z27113, g – z41097, h – r73550, i – h25137, j – r72919, k – d31272, l – d31002/ d31414, m – d31042, n – d31420, o – r50611, p – n32174, q – n35256, r – z46629/ s74504

resulted in a set of 12 cosmids (G0943, H0152, H0112, G9403, B0537, H0861, E0551, C1155, C0221, G0327, C0116, D0932) out of the ICRF library, being positive for the entire probe or parts of it. Only cosmids H0861, E0551 and C1155 resulted in signals that could be distinguished on Northern blots

Table 1. The set of ESTs localized in C1155, with their sequence database accession numbers and corresponding oligonucleotide primer pairs

z41097:
forward primer: TGTCCCTCTTCATTTCGACC
reverse primer: CATCATGGAAGCTCTTCAAGC
r73550:
forward primer: GAAGAGACAGGCCAGGTCAG
reverse primer: TCCTGACCTGTACATTCTGCC
r72919:
forward primer: TGATTCCTTTCCTGGAGCC
reverse primer: GAAGCACCAGGGTTGGTG
r50611:
forward primer: GACCTGTCAGCCTCTTCAGC
reverse primer: CCCCAGTTTGACTACTCTGACC
r48447:
forward primer: GGTCTCCTTCCTTTTTGATGC
reverse primer: GCATCGGTACAAAGATTTGATG
h25137:
forward primer: TGTGTCTCCTGGAGCCAAG
reverse primer: CAGTGCCTAAGGGCTAGGC
d31272/ d31002/ d31042/ d31420/ d31414:
forward primer: CATAGGAGTGGTAAGGCCTCC
reverse primer: ACTGTCCCGGCCCTAAAG
n35256/ n32174:
forward primer: ATACCACTGGGAGGCAAGG
reverse primer: CTAGCCACCCAGAGACTTGC

(data not shown). Cosmid C1155, the recombinant with the longest insert among the 3 positive candidates, was used for further investigation and its Northern signals appeared as a smear in the range from 500-600 down to 150-200 bases. Randomly subcloning the resulting EcoR I-fragments and using them to rehybridize to the Northern filters, did not permit defining a major coding region. This result indicated a dispersed distribution of expressed sequences along the entire genomic insert of C1155.

Our sequencing of the entire genomic DNA in c1155 showed regions of complete sequence identity with a previously characterized cDNA coding for one subunit of DNA-dependent RNA-polymerase II (hRPB14.4).

Further analysis of this region also revealed the location of a total of forty full length Alu-repeat sequences, one polymorphic $(CA)_{16}$-cluster, a new, SOX9- related gene sequence and twelve additional ESTs (Figure 1). Eleven out of the twelve ESTs (with sizes from 170 up to 470 bases) were present in the proximal third of the cosmid sequence. The polymerase II gene and all twelve ESTs are coded by the same DNA strand showing a surprising unidirectional orientation of transcription for genes in this region. Only the SOX9-related sequence is transcribed in the opposite direction.

To confirm the single copy character of all ESTs, PCR with a set of 6 different genomic DNAs as template and EST flanking primers (Table 1) was performed. The amplification product of each reaction was as expected, and after labeling and hybridization to Northern filters, each PCR product yielded the expected hybridization bands (data not shown).

Discussion

The sequencing of the entire genomic sequence cloned in C1155 showed regions of complete sequence identity with a cDNA of one subunit of DNA-dependent RNA-polymerase II (hRPB14.4) (ACKER et al. 1994a, b). Subsequent sequence analysis revealed the presence of several Alu-repeats and the exon/intron structure of the complete 14.4 kD subunit gene of RNA-polymerase II (PUSCH et al. 1996).

Furthermore, sequence comparison with public databases revealed information on a new, SOX9-related sequence and an additional set of 12 expressed sequence tags. The characterization of these 12 ESTs in public databases makes it highly unlikely that they would constitute parts of one complete gene. Nevertheless, all ESTs and the polymerase gene are transcribed in the same

direction within the genomic sequence of cosmid C1155. Until now, no extended genomic region in humans with such a high coding density was described to display a pronounced bias for transcriptional orientation.

Our present goal is to extend this mapping approach to obtain additional clones that should allow for filling the remaining gaps in this previously poorly charted region 22q13.1 (BLIN et al. 1993, GAYPAY et al. 1994, RILEY et al. 1994, BELL et al. 1995, COLLINS et al. 1995, SCHMITT et al. 1996).

The SOX9-related sequence is indeed the new gene SOX10 (PUSCH et al. 1998).

Acknowledgements. This project was aided by an EC Concerted Action travel support (Biomed1 B14), grants from the Deutsche Forschungsgemeinschaft (Mu 1153/1-1) and Bundesministerium für Bildung und Forschung (BEO 21/03-11212)(N.B.) and the U.S. National Institutes of Health Human Genome Project (HG00313)(B.R.).

REFERENCES

ACKER J., MATTEI M.-G., WINTZERITH M., ROECKEL N., DEPETRIS D., VIGNERON M., KEDINGER C. (1994a). Chromosomal localization of human RNA polymerase II subunit genes. Genomics 20: 496-499.

ACKER J., WINTZERITH M., VIGNERON M., KEDINGER C. (1994b). A 14.4 KDa acidic subunit of human RNA polymerase II with a putative leucine-zipper. DNA Seq. 4: 329-331.

BELL C., BUDARF M., NIEUWENHUIJSEN B., BARONSKI B., BUETOW K., CAMPBELL K., COLBERT A., COLLINS J., DALY M., DESJARDINS P., DEZWAAN T., ECKMAN B., FOOTE S., HART K., HIESTER K., van HET HOOG M., HOPPER E., KAUFMAN A., MCDERMID H., OVERTON G., REEVE M., SEARLS D., STEIN L., VALMIKI V., WATSON E., WILLIAMS S., WINSTON R., NUSSBAUM R., LANDER E., FISCHBECK K., EMANUEL B., HUDSON T. (1995). Integration of physical, breakpoint and genetic maps of chromosome 22. Hum. Mol. Genet. 4: 59-69.

BLIN N., MÜLLENBACH R., SUIJKERBUIJK R., HERZOG R., MEESE E. (1993). Cosmid mapping and locus linkage within the human chromosomal region 22q13.1. Cancer Genet. Cytogenet. 70: 108-111.

BODENTEICH A., CHISSOE S., WANG Y.F., ROE B.A. (1993). Shotgun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. In: Automated DNA Sequencing and Analysis Techniques (J.C. Venter, ed.). Academic Press, London: 42-50.

COLLINS J.E., COLE C.G., SMINK L., GARRETT C., LEVERSHA M., SODERLUND C., MASLEN G., EVERETT L., RICE K., COFFEY A., GREGORY S., GWILLIAM R., DUNHAM A., DAVIES A., HASSOCK S., TODD C., LEHRACH H., HULSEBOS J., WEISSENBACH J., MORROW B., KUCHERLAPATI R., WADEY R., SCAMBLER P., KIM U.J. (1995). A high density YAC contig map of human chromosome 22. Nature 377: 369-379.

GAYPAY G., MORISSETTE J., VIGINAL A., DIB C., FIZAMES C., MILLASSEAU P., MARC S., BERNARDI G., LATHROP M., WEISSENBACH J. (1994). The 1993-94 Genethon human genetic linkage map. Nature Genet. 7: 246-334.

GOUGH N.M. (1988). Rapid and quantitative preparation of cytoplasmic RNA from small number of cells. Anal. Biochem. 173: 93.

GÖTTERT E., METZDORF R., FÄRBER U., BLIN N. (1989). Regional localization and molecular characterization of a DNA sequence on the long arm of chromosome 22. Hum. Genet. 81: 385-387.

NIZETIC D., ZEHENTER G., MONACO A.P., GELLEN L., YOUNG B.D., LEHRACH H. (1991). Construction, arraying and high-density screening of large insert libraries of human chromosomes X and 21: Their potential use as reference libraries. Proc. Natl. Acad. Sci. USA 88: 3233-3237.

PUSCH C., HUSTERT E., PFEIFER D., SÜDBECK P., ROE B., WANG Z., BALLING R., BLIN N., SCHERER G. (1998). The SOX10/sox10 gene from human and mouse: sequence, expression and transactivation by the encoded HGM domain transcription factor. Hum. Genet., in press.

PUSCH C., WANG Z., ROE B., BLIN N. (1996). Genomic structure of RNA polymerase II small subunit (hRPB14.4) and mapping to 22q13.1 by sequence identity. Genomics 34: 440-442.

RILEY B., WILLIAMSON R., VERGNAUD G. (1994). The EUROGEM map of human chromosome 22. Europ. J. Hum. Genet. 2: 246-247.

SCHMITT H., KIM U.J., SLEPAK T., BLIN N., SIMON M., SHIZUYA H. (1996). Framework for a physical map of the human 22q13-region using Bacterial Artificial Chromosomes (BACs). Genomics 33: 9-20.