Hum Genet (1998) 102:282–288

ORIGINAL INVESTIGATION

Jean-Louis Blouin · Geneviève Duriaux Saïl
Michel Guipponi · Colette Rossier
Marie-Pierre Pappasavas · Stylianos E. Antonarakis

Isolation of the human BACH1 transcription regulator gene, which maps to chromosome 21q22.1

Received: 26 September 1997 / Accepted: 10 October 1997

Abstract In order to contribute to the development of the transcriptional map of chromosome 21, we performed exon trapping using cosmid clones mapped in the region 21q22.1–22.2 and identified a number of potential exons. One of the trapped exons (Genbank No. AF026200) showed a strong homology with the mouse Bach1 gene (Genbank No. D86603), a transcription factor regulating gene expression. We then isolated the full-length coding region of the human BACH1 gene using expressed sequence tags, reverse transcription-polymerase chain reaction and rapid amplification of cDNA ends. The predicted BACH1 protein contains 736 amino acids and is 88% identical to its mouse homolog. It contains basic leucine zipper and BTB-zinc finger domains (which are directly involved in DNA binding for transcription regulation). The BACH1 gene maps in a relatively gene-poor region on 21q22.1 in yeast artificial chromosome 814c1 of the collection of Chumakov et al. Northern blot analysis revealed that it is expressed as an mRNA species of approximately 5.8 kb in all 16 adult and 4 fetal tissues examined; an additional mRNA species of 2.8 kb was observed in adult testis. The contribution of the BACH1 gene to the pathophysiology of trisomy or monosomy 21 is unknown. In addition, no monogenic disorders associated with mutations in the BACH1 gene have yet been identified.

J.-L. Blouin · S. E. Antonarakis (✉)
Division of Medical Genetics,
Centre Médical Universitaire,
1 rue Michel Servet,
CH-1211 Genève, Switzerland
Tel.: +41-22-702-5707; Fax: +41-22-702-5706;
e-mail: Stylianos.Antonarakis@medecine.unige.ch

G. Duriaux Saïl · M. Guipponi · C. Rossier
M.-P. Pappasavas · S. E. Antonarakis
Laboratory of Human Molecular Genetics,
Department of Genetics and Microbiology,
University of Geneva Medical School,
Geneva, Switzerland

Introduction

Down syndrome is one of the most common chromosomal abnormalities in liveborn infants. It affects 1 in 600–1000 newborns worldwide (Epstein 1995) and is caused by trisomy 21. Human chromosome 21 (HC21) is the smallest autosome and is estimated to contain 600–1000 genes (reviewed in Antonarakis 1993). The cloning and characterization of genes on HC21 is a necessary step for the understanding of the molecular basis of Down syndrome, partial monosomy 21, monogenic disorders that map on this chromosome, and HC21 loci that predispose to multigenic complex phenotypes. The genetic linkage map (McInnis et al. 1993; Antonarakis et al. 1995), yeast artificial chromosome (YAC) and cosmid-based physical maps of HC21 are well developed (Chumakov et al. 1992; Nizetic et al. 1994; Soeda et al. 1995; Osoegawa et al. 1996) and have provided the necessary infrastructure to localize newly identified genes and loci on this chromosome. The transcription (genic) map of HC21 and the cloning and characterization of its genes have been progressing rapidly (Cheng et al. 1994; Peterson et al. 1994; Antonarakis et al. 1995; Lucente et al. 1995; Tassone et al. 1995; Yaspo et al. 1995; Chen et al. 1996a). However, only about 75 of the genes on HC21 have been cloned to date [Genome DataBase (GDB) (http:// gdbwww.gdb.org) and SWISS-PROT (http://expasy.hcuge. ch/cgi-bin/lists?humchr21.txt)].

In order to contribute to the development of the genic map of HC21 and the cloning of genes that may be responsible for Down syndrome, monogenic disorders and microdeletion syndromes of HC21, we have used exon trapping (Buckler et al. 1991) from the HC21-specific cosmid LL21NC02-Q library (Soeda et al. 1995) to identify potential exons of genes on this chromosome (Chen et al. 1996a) and thereafter to clone their corresponding genes (for example see Chen et al. 1995a, b; 1996a, b). In addition, we have also performed region-specific exon-trapping experiments from cloned DNA from within a gene-rich 21q22.3 region, the formerly called Down syndrome chromosomal region (DSCR; Lucente et al. 1995), and

more recently the relatively gene-poor (Gardiner 1996) region on 21q22.1 between APP and D21S17. As a result of the exons identified in this last region, here we report the cloning of BACH1, the human homolog of the murine Bach1 gene, which encodes a potential transcription regulator protein with two recognizable domains: a cap'n'collar type of basic leucine zipper (CNC-bZip) and a broad complex-tramtrack-bric-a-brac (BTB) type of zinc finger domain. The mouse Bach1 protein a transcription factor that, in heterodimerization with MafK, activates or represses transcription (Oyake et al. 1996). The human BACH1 gene maps on 21q22.1 and is expressed in all 20 tissues studied.

Materials and methods

Selection of cosmids

Cosmids used were from the HC21 genomic library LL21NC02-Q provided by Pieter de Jong (Soeda et al. 1995). Cosmid clones were selected by hybridization of colony grid filters probed with DNAs of CEPH yeast artificial chromosomes (YACs) (Chumakov et al. 1992) located between APP and D21S17 on chromosome 21q21–21q22.2; the probe from the YAC insert was labeled by random priming after purification by pulsed field gel electrophoresis. Cosmid clones cross-hybridizing with mouse genomic sequences or with ribosomal sequences were previously eliminated as described in Chen et al. (1996a, b)

Exon trapping

Exon trapping (Buckler et al. 1991) from the selected cosmids from the HC21-specific library LL21NC02-Q that hybridized to YACs between markers APP and D21S17 in 21q22.1 was performed as described (Chen et al. 1996a; Blouin et al. 1997) DNA from pools of 12 cosmids was cleaved with *Pst*I and the resulting fragments were subcloned in vector pSPL3. Recombinant subclone DNAs were then transfected using lipofectACE (GIBCO-BRL) into COS 7 cells. After 2 days of culture in DMEM supplemented with 10% fetal calf serum, cells were harvested and RNA isolated using TRIZOL reagents. Trapped sequences were recovered after reverse transcription-polymerase chain reaction (RT-PCR) using SA2 and SD6 oligonucleotide primers. After elimination of vector-derived products by *Bst*XI cleavage, spliced sequences were subcloned into the pAMP10 vector using Uracyl DNA Glycosylase (GIBCO-BRL). Nonrecombinant vector clones, vector self-splicing products, and cosmid vector sequences were eliminated by appropriate hybridizations (Chen et al. 1996a). Trapped sequences of different sizes were then subjected to nucleotide sequencing using the ABI373A automated nucleotide sequencer. Sequence homology analysis was performed using BLAST programs (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/). One trapped sequence, jla60g5 (Genbank No. AF026200), showed strong homology to the mouse Bach1 gene and was used for isolation of the full coding region of the human homolog as described below.

Cloning and sequencing of human BACH1 cDNA

The 5' end and coding region up to nucleotide 664 was obtained from sequencing of IMAGE clone 115234, corresponding to expressed sequence tags (ESTs) yd87e06.r1 and yd87e06.s1. This EST was identified using BLASTN database searches with the mouse Bach1 sequence. To obtain the 3' end of the coding sequence of human BACH1, we performed a 3' rapid amplification of cDNA ends (RACE) on human heart poly(A)$^+$ mRNA using a cDNA amplification kit (Clontech, K1802-1). Two rounds of PCR in thermocyclers MJR PTC-100 or Biometra UNO1 were performed using first oligonucleotide primers jla60g5A (5'ACTGCCATTCAATGCACAAC3') and AP1 and subsequently primer AP2 in a semi-nested reaction. The middle one-third of the coding sequence (between the EST sequence yd87e06 and the trapped exon) has been isolated by PCR from a human heart cDNA using a primer in the previously identified 5 EST (yd87co3C, 5'GAATCCTGCTTTCAGTTTCTG3') and a primer in exon jla60g5 (jla60g5B, 5'CAGCTTCTCA-ATTTCTGATTC3'). Semi-nested PCR was then performed using the primers yd87co3B (5'TGGACTCCACTGCAGACC-AGC3') and jla60g5B. All nucleotide sequences were determined in both orientations using standard protocols in an ABI373 automated sequencer.

Northern blot analysis

Northern blot analysis was performed using random priming of a PCR-amplified 1.4 (1.384)-kb segment of the BACH1 cDNA as a probe between nucleotides 635–2019 by using oligonucleotide primers yd87co3B (5'TGGACTCCACTGC-AGACCAGC3') and jla60g5B. Hybridization to a multiple-tissue Northern blot containing 2 μg of poly(A)$^+$ mRNA from 16 human adult and 4 fetal tissues (Clontech filters 7760-1, 7759-1, 7756-1) was performed according to the manufacturer's instructions. Equal loading was confirmed using the -actin probe as a control.

Mapping to HC21

The mapping position of BACH1 on HC21 was determined using several methods. First, by PCR amplification from genomic DNA from mouse-human monochromosomal somatic hybrids (NIGMS2 panel) (Drwinga et al. 1993) and specific somatic cell hybrids containing defined portions of HC21 (Patterson et al. 1993) using oligonucleotide primers jla60g5A and jla60g5B. Second, by hybridization of PCR-amplified labeled exon probe jla60g5 on the HC21-specific cosmid library LL21NC02-Q. Positive cosmids were confirmed by PCR amplification. Furthermore, selected YACs from the Chumakov et al. (1992) HC21 contig were used for PCR amplification using the trapped exon-specific oligonucleotide primers. Third, PCR amplification from 93 human-hamster radiation somatic cell hybrids of the Genebridge 4 panel (provided by the UK Human Gene Mapping Project Resource Centre) using primers jla60g5A and jla60g5B were performed and the data were submitted to the Whitehead/MIT genome mapping web site (http://www-genome.wi.mit.edu/) for incorporation in the radiation hybrid framework map.

Sequence analysis

Nucleotide and predicted amino acid sequences were analyzed with BLAST programs at the NCBI (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST) (Altschul et al. 1990). Alignment of sequences was performed using PC/GENE software (developed by A. Bairoch) (Intelligenetics) and at the T.I.G.E.M. site with the CAP sequence assembly program (http://www.tigem.it/ASSEMBLY/assemble.html). Protein sequence analysis for domain recognition was performed using the Expasy tool package (http://www.expasy.ch.www/tools.html). Sequence homology alignments were performed with Clustalw and subsequently edited using the Genedoc multiple sequence alignment software.

Results

A trapped exon with homology to mouse Bach1 cDNA

Using exon trapping on cosmids that mapped in the interval between markers APP and D21S17 on 21q22.1 we have identified a number of potential exons (J. L. B. and S. E. A., manuscript in preparation). The predicted amino acid sequence of one trapped exon of 207 bp, (jla60g5; Genbank No. AF026200) obtained from a pool of 12 cosmids (Q82F8, Q82H8, Q82B7, Q86C3, Q86B2, Q88D9, Q88A11, Q89B9, Q89G8, Q90F10, Q90D1, Q94A9, Q97D11) showed excellent homology with the mouse Bach1 protein sequences (Genbank No. D86603, $P = 6.0e^{-55}$, 84% identity, Oyake et al. 1996). We subsequently decided to clone the corresponding human full-length coding cDNA, since it was likely that exon jla60g5 was part of a human gene similar to the murine transcription regulator Bach1.

Cloning of the human BACH1 cDNA

The nucleotide sequence of the full-length coding region of the human BACH1 cDNA was determined by the use of (i) three human ESTs homologous to the mouse gene (yd87e06.s1, Genbank No. T86439, $P = 1.0^{-90}$; yd87e06.r1, Genbank No. T86529, $P = 1.7e^{-22}$; and zr94e05.r1, Genbank No. AA215308, $P = 4.6e^{-23}$); (ii) an amplicon of 1.4 kb from a heart cDNA library using primers yd87c03C and jla60g5B and nested primer yd87c03B in the EST and trapped exon (see Materials and methods); (iii) the 0.7-kb product of 3 RACE PCR from the heart

Fig. 1 Nucleotide and predicted protein sequences of the human BACH1 gene (Genbank No. AF026199). The initiation methionine and the termination codons are in *bold*. Trapped exon jla60g5 comprises nucleotides 1813 to 2019 and is *double-lined above*. The CNC-type bZIP domain is from amino acids 562 to 624. The basic portion (amino acids 562–577) of the CNC-type bZIP domain is *boxed* and the leucine zipper is shown as *gray stripes*. The BTB zinc finger domain (amino acids 16–122) is indicated as a *gray-shaded region*. Clone IMAGE 115234, containing expressed sequence tags yd87e06.r1 and yd87e06.s1, is *lined above*. Oligonucleotide primers yd87co3C, yd87co3B, jla60g5A, and jla60g5B are in *underlined bold* in that order

```
GTCTACTCAGCCCGGTGGCTGTCGCGCGTGGAATCGCGTAAGAAAAGCCGAGTTTGTGGCTGGGGAGAGAAGGCCACCGTGCTGAGCTGGATTTAGCGAAGACTGGTTTTGGGGACCGGA  120
GAGCCCAGGACTCCCTTTGTTGGAGTTTTGCCCACGCGTTGTAATTAAGCCTCGCACAATATGGTTGATGATAATTAGAAGCATGCTTTCCACTGAACTTCCCGACAACATTTGTTATGC  240

AGAATGTCTCTGAGTGAGAACTCGGTTTTTGCCTATGAATCTTCTGTGCATAGCACCAATGTTTTACTCAGCCTTAATGACCAGCGGAAGAAAGATGTGCTGTGCGATGTCACCATCTTT  360
   M  S  L  S  E  N  S  V  F  A  Y  E  S  S  V  H  S  T  N  V  L  L  S  L  N  D  Q  R  K  K  D  V  L  C  D  V  T  I  F   39

GTGGAAGGACAGCGGTTCCGCGCTCACCGGTCCGTGCTGGCGGCATGCAGCAGTTACTTCCACTCAAGAATCGTAGGCCAGGCTGATGGAGAGCTGAACATTACTCTTCCAGAAGAGGTG  480
V  E  G  Q  R  F  R  A  H  R  S  V  L  A  A  C  S  S  Y  F  H  S  R  I  V  G  Q  A  D  G  E  L  N  I  T  L  P  E  E  V   79

ACAGTTAAAGGATTTGAACCTTTAATTCAGTTTGCCTACACTGCTAAACTGATTTTAAGTAAAGAGAATGTGGATGAAGTGTGCAAATGTGTGGAGTTTTTAAGTGTACATAATATTGAG  600
T  V  K  G  F  E  P  L  I  Q  F  A  Y  T  A  K  L  I  L  S  K  E  N  V  D  E  V  C  K  C  V  E  F  L  S  V  H  N  I  E  119

GAATCCTGCTTTCAGTTTCTGAAATTTAAGTTTTTGGACTCCACTGCAGACCAGCAAGAATGCCCCCGCAAAAAATGCTTTTCATCACACTGTCAGAAAACAGACCTTAAACTTACACTT  720
E  S  C  F  Q  F  L  K  F  K  F  L  D  S  T  A  D  Q  Q  E  C  P  R  K  K  C  F  S  S  H  C  Q  K  T  D  L  K  L  T  L  159

TTGGACCAGAGGGATCTAGAAACTGATGAAGTGGGGGAATTTCTGGAAAATAAAAATGTTCAGACTCCTCAGTGTAAACTCCGCAGGTATCAAGGAAATGCAAAAGCCTCACCTCCTCTA  840
L  D  Q  R  D  L  E  T  D  E  V  G  E  F  L  E  N  K  N  V  Q  T  P  Q  C  K  L  R  R  Y  Q  G  N  A  K  A  S  P  P  L  199

CAAGACAGTGCCAGTCAGACATATGAGTCCATGTGCTTAGAGAAGGATGCTGCTCTGGCCTTGCCTTCTTTATGCCCCAAATACAGAAAATTCCAAAAAGCATTTGGAACTGACAGAGTC  960
Q  D  S  A  S  Q  T  Y  E  S  M  C  L  E  K  D  A  A  L  A  L  P  S  L  C  P  K  Y  R  K  F  Q  K  A  F  G  T  D  R  V  239

CGTACTGGGGAATCTAGTGTCAAAGACATTCATGCTTCTGTTCAGCCGAATGAAAGGTCTGAAAATGAATGCCTGGGAGGAGTCCCGGAGTGTAGAGATTTGCAGGTGATGTTAAAATGT  1080
R  T  G  E  S  S  V  K  D  I  H  A  S  V  Q  P  N  E  R  S  E  N  E  C  L  G  G  V  P  E  C  R  D  L  Q  V  M  L  K  C  279

GACGAAAGTAAATTAGCAATGGAACCTGAAGAAACGAAGAAAGATCCTGCTTCTCAGTGCCCAACTGAAAAATCAGAAGTGACTCCTTTCCCCCACAATTCTTCCATAGACCCTCATGGA  1200
D  E  S  K  L  A  M  E  P  E  E  T  K  K  D  P  A  S  Q  C  P  T  E  K  S  E  V  T  P  F  P  H  N  S  S  I  D  P  H  G  319

CTTTATTCTTTGTCTCTTTTACACACATATGACCAATATGGTGACTTGAATTTTGCTGGTATGCAAAACACAACAGTGTTAACAGAAAAGCCTTTGTCAGGTACAGACGTCCAAGAAAAA  1320
L  Y  S  L  S  L  L  H  T  Y  D  Q  Y  G  D  L  N  F  A  G  M  Q  N  T  T  V  L  T  E  K  P  L  S  G  T  D  V  Q  E  K  359

ACATTTGGTGAAAGTCAGGATTTACCTTTGAAATCCGACTTGGGCACCAGGGAAGATAGTAGTGTTGCATCTAGTGATAGGAGTAGTGTGGAGCGAGAAGTGGCAGAACACTTAGCAAAA  1440
T  F  G  E  S  Q  D  L  P  L  K  S  D  L  G  T  R  E  D  S  S  V  A  S  S  D  R  S  S  V  E  R  E  V  A  E  H  L  A  K  399

GGGTTCTGGAGTGACATTTGCAGCACGGACACTCCTTGCCAAATGCAGTTATCACCTGCTGTGGCCAAAGATGGCTCAGAACAGATCTCACAGAAACGGTCTGAGTGTCCGTGGTTAGGT  1560
G  F  W  S  D  I  C  S  T  D  T  P  C  Q  M  Q  L  S  P  A  V  A  K  D  G  S  E  Q  I  S  Q  K  R  S  E  C  P  W  L  G  439

ATCAGGATTAGTGAGAGCCCGGAACCAGGTCAAAGGACTTTCACAACATTAAGTTCTGTCAACTGCCCTTTTATAAGTACTCTGAGTACTGAAGGCTGTTCAAGCAATTTGGAAATTGGA  1680
I  R  I  S  E  S  P  E  P  G  Q  R  T  F  T  T  L  S  S  V  N  C  P  F  I  S  T  L  S  T  E  G  C  S  S  N  L  E  I  G  479

AACGATGATTATGTTTCAGAACCCCAGCAAGAACCTTGCCCATATGCTTGTGTCATTAGCTTGGGAGACGACTCTGAGACGGACACCGAAGGAGACAGTGAATCCTGTTCAGCCAGAGAA  1800
N  D  D  Y  V  S  E  P  Q  Q  E  P  C  P  Y  A  C  V  I  S  L  G  D  D  S  E  T  D  T  E  G  D  S  E  S  C  S  A  R  E  519

CAAGAATGTGAGGTAAAACTGCCATTCAATGCACAACGGATAATTTCACTGTCTCGAAATGATTTTCAGTCCTTGTTGAAAATGCACAAGCTTACTCCAGAACAGCTGGATTGTATCCAT  1920
Q  E  C  E  V  K  L  P  F  N  A  Q  R  I  I  S  L  S  R  N  D  F  Q  S  L  L  K  M  H  K  L  T  P  E  Q  L  D  C  I  H  559

GATATTCGAAGAAGAAGTAAAAACAGAATTGCTGCACAGCGCTGTCGCAAGAGAAAACTTGACTGTATACAGAATCTTGAATCAGAAATTGAGAAGCTGCAAAGTGAAAAGGAGAGCTTG  2040
D  I  R  R  R  S  K  N  R  I  A  A  Q  R  C  R  K  R  K  L  D  C  I  Q  N  L  E  S  E  I  E  K  L  Q  S  E  K  E  S  L  599

TTGAAGGAAAGAGATCACATTTTGTCAACTCTGGGTGAGACAAAGCAGAACCTAACTGGACTTTGCCAGAAAGTTTGTAAAGAAGCAGCTCTGAGTCAAGAACAAATACAGATACTCGCC  2160
L  K  E  R  D  H  I  L  S  T  L  G  E  T  K  Q  N  L  T  G  L  C  Q  K  V  C  K  E  A  A  L  S  Q  E  Q  I  Q  I  L  A  639

AAGTACTCAGCTGCAGATTGCCCACTTTCATTTTTAATTTCTGAAAAAGATAAAAGTACTCCTGATGGTGAACTGGCGTTACCATCAATTTTCAGTTTATCTGACCGGCCTCCAGCAGTG  2280
K  Y  S  A  A  D  C  P  L  S  F  L  I  S  E  K  D  K  S  T  P  D  G  E  L  A  L  P  S  I  F  S  L  S  D  R  P  P  A  V  679

CTGCCTCCCTGTGCCAGAGGAAACAGTGAGCCTGGCTACGCGCGAGGGCAGGAGTCCCAGCAGATGTCCACAGCCACCTCTGAGCAAGCTGGGCCTGCGGAACAGTGTCGTCAGAGTGGT  2400
L  P  P  C  A  R  G  N  S  E  P  G  Y  A  R  G  Q  E  S  Q  Q  M  S  T  A  T  S  E  Q  A  G  P  A  E  Q  C  R  Q  S  G  719

GGGATCTCAGATTTCTGTCAGCAGATGACTGATAAATGTACTACTGATGAGTAAACTTGCATTCACTTCCTTCAAACC  2478
G  I  S  D  F  C  Q  Q  M  T  D  K  C  T  T  D  E  *                              736
```

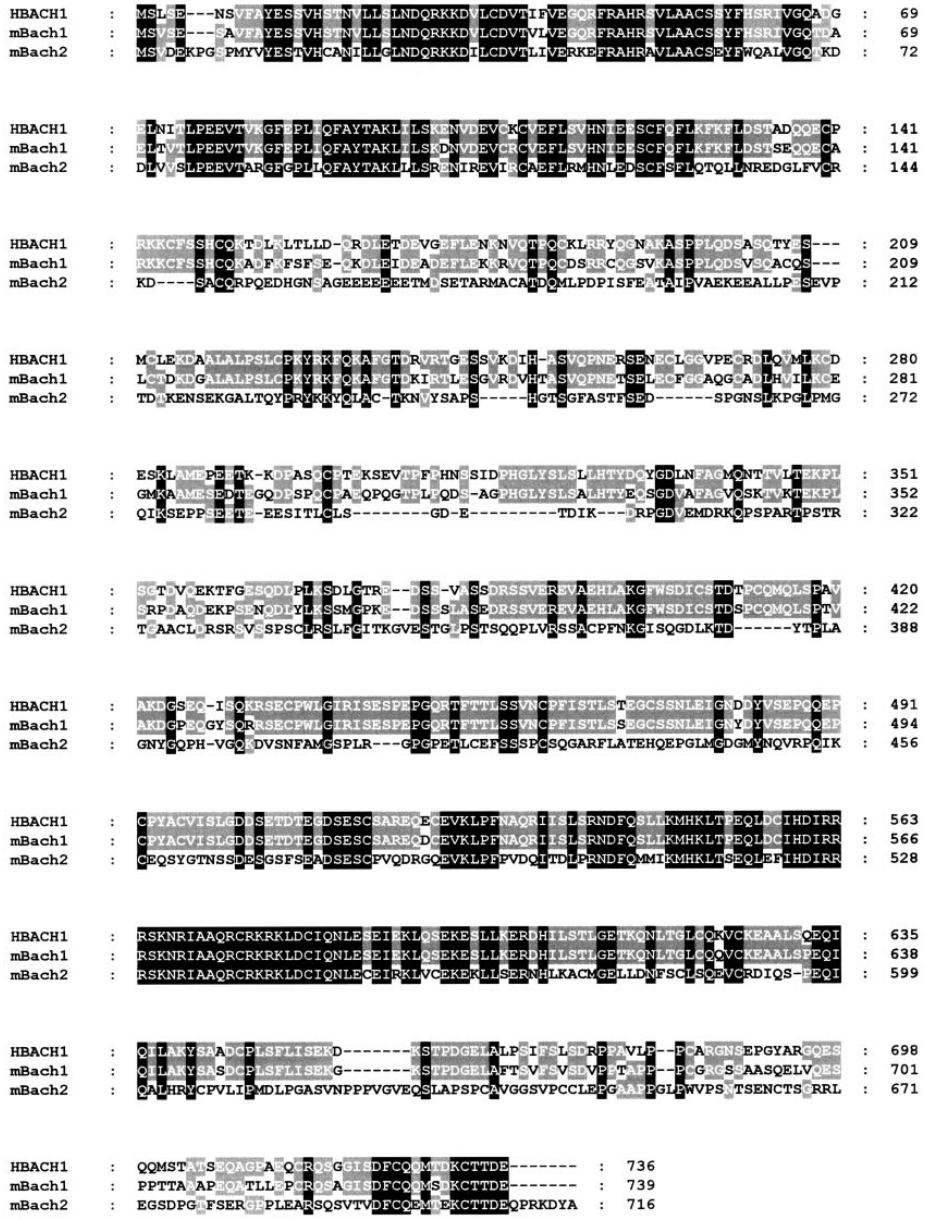

Fig. 2 Multiple sequence alignment of the predicted human BACH1 (Genbank No. AF026199), mouse Bach1 (Genbank No. D86603) and mouse Bach2 (Genbank No. D86604) proteins generated with the Genedoc program. *Dark shading* indicates identity among the three proteins; *gray shading* indicates identity between two of the sequences. Conservative amino acid substitutions are not shaded

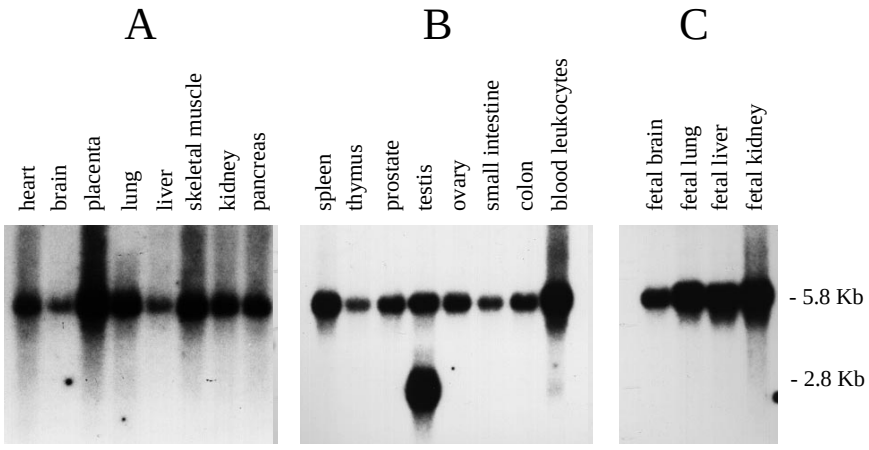

Fig. 3A–C Expression pattern of human BACH1 on a selection of poly(A)$^+$ mRNAs isolated from 16 adult (**A**, **B**) and 4 fetal (**C**) tissues. The Northern blots (Clontech filters 7760–1, 7759–1, 7756–1) were probed with a polymerase chain reaction (PCR)-amplified product of 1.4 kb of the BACH1 cDNA between nucleotides 635–2019. A single 5.8-kb mRNA species was observed in all tissues. In addition, testis displays a second mRNA species of 2.8 kb. The same filters were hybridized with an actin probe (data not shown); the amount of mRNA loaded in all lanes is approximately equal

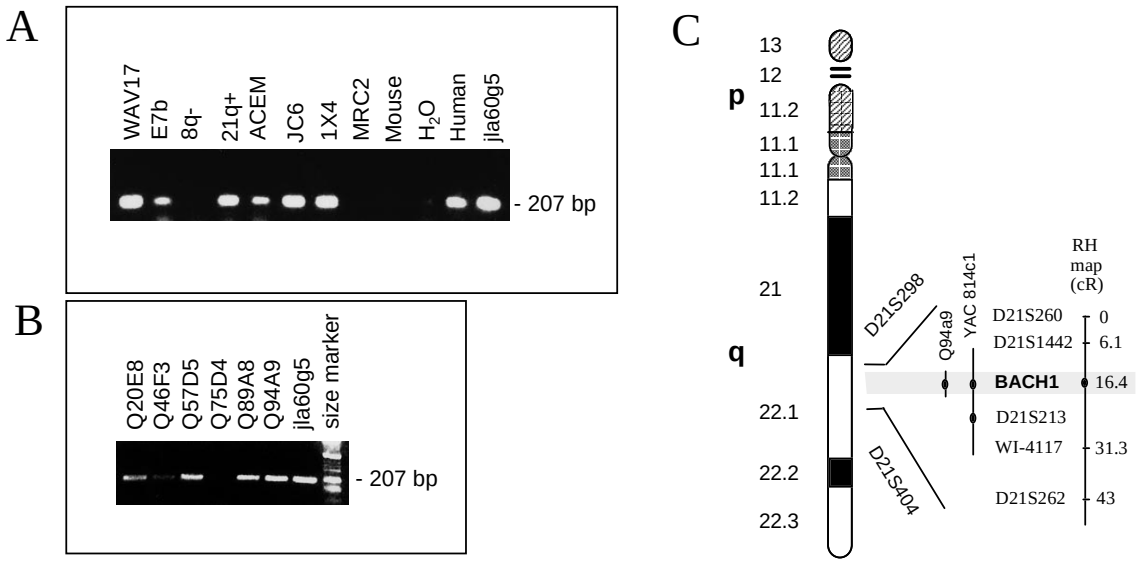

Fig 4A–C Mapping of the human BACH1 gene. PCR amplification of the 207-bp exon jla60g5 of the human BACH1 on DNA from **A** selected HC21 somatic cell hybrids and **B** cosmids from the HC21-specific LL21NC02-Q library. **C** Schematic representation of the mapping position of the human BACH1 gene. Somatic cell hybrids localize this gene between markers D21S298 and D21S404 in 21q22.1. In this interval the positive cosmid Q94A9 and yeast artificial chromosome (YAC) 814c1 are shown. The YAC is also positive for additional markers, including D21S213. The orientation of BACH1 and D21S213 relative to the telomere has not been determined and is shown arbitrarily. BACH1 is also incorporated between markers D21S1442 and WI-4117 in the Radiation Hybrid (RH) framework map using the Genebridge 4 panel. The distances in the partial RH map shown are in centiRays

cDNA library (see Materials and methods for oligonucleotides).

Figure 1 shows the nucleotide and predicted amino acid sequence of human BACH1 (Genbank No. AF026199). The methionine codon for translation initiation starting at nucleotide 244 corresponds to the mouse initiation codon and is within the context of a Kozak (1987) consensus sequence. The open reading frame extends for 2208 nucleotides and the predicted protein contains 736 amino acids versus 739 such residues in the mouse. At the nucleotide level there is 76% identity between the human BACH1 gene and its mouse homolog; the identity of their predicted amino acid sequences is 88% (Fig. 2)

Similar to mouse Bach1, the predicted protein sequence of human BACH1 contained the following recognizable domains: a CNC-type bZIP domain extending from amino acids 562–624. The basic portion of the domain is at amino acids 562–577 and the leucine zipper motif is included in the region of amino acids 578–624 (http://www.expasy.ch/sprot/prosite.html) (Fig. 1). The bZIP motif is usually found in DNA-binding proteins that function as transcription regulators. The leucine zipper that generally follows the basic domain is defined as a periodic repetition of leucine residues at every seventh position. This motif is involved in the homo- or heterodimerization of the protein. The CNC-type of basic leucine zipper is a subtype of this domain found in some proteins, including human FOS and JUN (Van Straaten et al. 1983; Schuette et al. 1989)

The second recognizable feature is the BTB domain between amino acids 16–122 (Fig. 1). This domain, which contains a C2H2-type zinc finger motif (Prosite No. PDOC00028), is found in several proteins, including human transcription termination factor l-interacting peptide, BCL-6 (B-cell lymphoma 6 protein), Kruppel-related zinc finger protein and MIZ-1 among others (Baron et al. 1993; Kawamata et al. 1994; Bellefroid et al. 1989; Sauer and Jackle 1993, respectively). This domain is evolutionarily conserved from *Drosophila* to mammals and is involved in protein-protein interactions (Zollman et al. 1994).

Northern blot analysis

In order to study the expression pattern of the human BACH1 gene, we used Northern blot analysis. A partial cDNA probe (from nucleotides 635–2019 of Fig. 1) was hybridized against poly(A)$^+$ mRNA from 16 adult and 4 fetal tissues (MTN blots from Clontech). An mRNA species of 5.8 kb was observed in every tissue examined (Fig. 3). The level of expression varies among the different tissues. In testis, in addition to the 5.8-kb species, an abundant mRNA species of approximately 2.8 kb was observed. This additional mRNA species may result from the use of different 3’ or 5’untranslated regions (UTRs), alternative splicing or a cross-hybridizing sequence. The cDNA sequence shown in Fig. 1 does not obviously contain all the 5’ or 3’ UTR sequence.

The human BACH1 maps to 21q22.1

The specific localization of BACH1 to chromosome 21q22.1 was determined by several experiments (Fig. 4).

PCR amplification of the 207-bp trapped exon sequence jla60g5 from monochromosomal somatic cell hybrids specific (NIGMS2) (Dwringa et al. 1993) and from defined portions of HC21 (Patterson et al. 1993) was performed. From the monochromosomal hybrids, only that with HC21 resulted in amplification of the specific product. The partial HC21-containing hybrids that resulted in the specific amplification were WAV17, 153e7b, 21q+, ACEM, 1X4 and JC6. Furthermore, no amplification was obtained from hybrids 8q-, and MRC-2G (Fig. 4A). These data indicate that BACH1 maps to 21q22.1 between D21S298 and D21S404 (Fig. 4C).

PCR amplification of the same 207 bp using jla60g5 exon primers on DNA of somatic cell hybrids from the Genebridge 4 mapping panel was used to incorporate BACH1 in the radiation hybrid map. The gene was placed in the interval between markers D21S1442 and WI-4117, 10.3 cR distal to D21S1442 (Fig. 4C); this interval is located on 21q22.1 on contig wc21.1 (http://www-genome.wi.mit.edu/).

Hybridization of HC21-specific cosmid and YAC library filters using a partial cDNA as a probe revealed several positive cosmids, Q94A9, Q20E8, Q46F3, Q57D5, and Q89A8 and one positive YAC, 814c1. These positive clones were then confirmed by PCR amplification using primers jla60g5A and jla60g5B (Fig. 4B). Cosmid Q94A9 was included in the cosmid pool DNA for the exon-trapping experiment from which exon jla60g5 was isolated. The results of all mapping experiments are summarized schematically in Fig. 4C.

Discussion

In this paper we describe the cloning, chromosomal mapping and initial characterization of a novel gene that maps on human chromosome 21q22.1, and encodes a polypeptide that is the human homolog of the mouse Bach1 protein.

Similar to its human homolog, Bach1 contains CNC-type bZIP and BTB domains. These domains have been described in proteins that modulate transcription. The mouse Bach1 (and its homologous Bach2 protein) functions as a transcription repressor in fibroblasts and transcription activator in cultured erythroid cells (Oyake et al. 1996). Bach1 probably forms heterodimers with the MafK proteins and the complex binds in vitro to NF-E2 binding sites (Oyake et al. 1996). It is likely that the function of the human BACH1 protein is similar to that of its mouse homolog.

Are there any monogenic disorders associated with BACH1? There are no human monogenic disorders that have been mapped to the interval of the mapping position of BACH1. In addition no mouse phenotype has been associated with mutations in mouse Bach1. The generation of mice with targeted disruption of the mouse Bach1 gene will enhance our understanding of the function of this gene.

The mapping location of BACH1 is within the APP to SOD1 region, which has been associated with several features of monosomy 21 (Chettouh et al. 1995; Orti et al. 1997). The phenotypes associated with this partial monosomy 21 were facial anomalies, growth retardation, arthrogryposis-like syndrome, hypertonia, heart defects and mental retardation. It is therefore possible that haploinsufficiency for BACH1 is involved in one of these phenotypes.

Is the overexpression of three copies of BACH1 involved in one of the phenotypes of Down syndrome? BACH1 maps outside the so-called DSCR (between markers D21S17 and ETS2), triplication of which is associated with many phenotypes of Down syndrome (Delabar et al. 1993; McCormick et al. 1989). However the existence of a single DSCR has been challenged since rare patients wih proximal trisomy 21 that does not include the D21S17–ETS2 region display some of the phenotypes of Down syndrome (Korenberg et al. 1994). BACH1 maps within the region of this proximal partial trisomy 21.

The study of transgenic mice that overexpress BACH1 may contribute to the understanding of potential involvement of this gene in the pathogenesis of Down syndrome. A mouse model (Ts65Dn) with partial trisomy 16 (which corresponds to a partial human trisomy 21 from APP to MX1) has been made that demonstrates impaired performance in learning tasks (Reeves et al. 1995). The synteny between mouse and human genes suggests that BACH1 is included in the triplicated part of mouse chromosome 16 in these animals and therefore its potential overexpression in different tissues and stages of development could contribute to its phenotype.

Acknowledgements We thank H. S. Scott, H. M. Chen, R. Chrast, M. Lalioti and M. Barman for advice and/or help in specific experiments, P. deJong for providing the HC21-specific cosmid library LL21NC02 (the construction of this library was sponsored by the US Department of Energy), and D. Patterson for the HC21-specific somatic cell hybrids. This study was supported by grants from the Swiss FNRS 31-40500-94 from the European Union PL-930015 and funds from the University and Cantonal Hospital of Geneva.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Antonarakis SE (1993) Human chromosome 21: genome mapping and exploration circa 1993. Trends Genet 9:142–148

Antonarakis SE, Patterson D, Broeckhoven C van, Shimizu N, Gardiner K, Delabar J, Korenberg J, Reeves R (1995) Report of the committee on the genetic constitution of chromosome 21. In: Cutucchia AJ (ed) Human Gene Mapping 1994. Johns Hopkins University Press, pp 732–766

Baron BW, Nucifora G, McCabe N, Espinosa RD, Le Beau MM, McKeithan TW (1993) Identification of the gene associated with the recurring chromosomal translocations t(3;14)(q27;q32) and t(3;22)(q27;q11) in B-cell lymphomas. Proc Natl Acad Sci USA 90:5262–5266

Bellefroid EJ, Lecocq PJ, Benhida A, Poncelet DA, Belayew A, Martial JA (1989) The human genome contains hundreds of genes coding for finger proteins of the Kruppel type. DNA 8:377–387

Blouin JL, Duriaux Saïl G, Antonarakis SE (1997) A specific chromosome 21q22.1 exon trapping for detection of expressed sequences within a gene-rich region. Med Genet Suppl 9: 4.185

Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE (1991) Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 88:4005–4009

Chen H, Morris MA, Rossier C, Blouin JL, Antonarakis SE (1995a) Cloning of the cDNA for the human ATP synthase OSCP subunit (ATP5O) by exon trapping and mapping to chromosome 21q22.1–q22.2. Genomics 28:470–476

Chen HM, Chrast R, Rossier C, Gos A, Antonarakis SE, Kudoh J, Yamaki A, Shindoh N, Meada H, Minoshima S, Shimizu N (1995b) The human homolog of the *Drosophila* SIM (single-minded) gene maps in the Down syndrome critical region. Nat Genet 10:9–10

Chen H, Chrast R, Rossier C, Morris MA, Lalioti MD, Antonarakis SE (1996a) Cloning of 559 potential exons of genes of human chromosome 21 by exon trapping. Genome Res 6:747–760

Chen H, Rossier C, Antonarakis SE (1996b) Cloning of a human homolog of the *Drosophila* enhancer of zeste gene (EZH2) that maps to chromosome 21q22.2. Genomics 38:30–37

Cheng JF, Boyartchuk V, Zhu YW (1994) Isolation and mapping of human chromosome 21 cDNA - progress in constructing a chromosome 21 expression map. Genomics 23:75–84

Chettouh Z, Croquette MF, Delobel B, Gilgenkrants S, Leonard C, Maunoury C, Prieur M, Rethore MO, Sinet PM, Chery M, et al. (1995) Molecular mapping of 21 features associated with partial monosomy 21: involvement of the APP-SOD1 region. Am J Hum Genet 57:62–71

Chumakov I, Rigault P, Guillou S, Ougen P, Billaut A, Guasconi G, Gervy P, LeGall I, Soularue P, Grinas L, Bougueleret L, Bellanne-Chantelot C, Lacroix B, Barillot E, Gesnouin P, Pook S, Vaysseix G, Frelat G, Schmitz A, Sambucy JL, Bosch A, Estivill X, Weissenbach J, Vignal A, Riethman H, Cox D, Patterson D, Gardiner K, Hattori M, Sakaki Y, Ichikawa H, Ohki M, Paslier D Le, Heilig R, Antonarakis SE, Cohen D (1992) A continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359:380–386

Delabar JM, Theophile D, Rahmani Z, Chettouh Z, Blouin JL, Prieur M, Noel B, Sinet PM (1993) Molecular mapping of twenty-four features of Down syndrome on chromosome 21. Eur J Hum Genet 1:114–124

Drwinga HL, Toji LH, Kim CH, Greene AE, Mulivor RA (1993) NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. Genomics 16:311–314

Epstein CJ (1995) Epilogue: toward the twenty-first century with Down syndrome – a personal view of how far we have come and of how far we can reasonably expect to go. Prog Clin Biol Res 393:241–246

Gardiner K (1996) Base composition and gene distribution: critical patterns in mammalian genome organization. Trends Genet 12:519–524

Kawamata N, Miki T, Fukuda T, Hirosawa S, Aoki N (1994) The organization of the BCL6 gene. Leukemia 8:1327–1330

Korenberg JR, Chen XN, Schipper R, Sun Z, Gonsky R, Gerwehr S, Carpenter N, Daumer D, Dignan P, Disteche C, Graham JM, Hugdins L, McGillivray B, Miyazaki K, Ogasawara N, Park JP, Pagon R, Pueschel S, Sack G, Say B, Schuffenhauer S, Soukup S, Yamanaka T (1994) Down syndrome phenotypes – the consequences of chromosomal imbalance. Proc Natl Acad Sci USA 91:4997–5001

Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Lucente D, Chen HM, Shea D, Samec SN, Rutter M, Chrast R, Rossier C, Buckler A, Antonarakis SE, McCormick MK (1995) Localization of 102 exons to a 2.5 Mb region involved in Down syndrome. Hum Mol Genet 4:1305–1311

McCormick MK, Schinzel A, Petersen MB, Stetten G, Driscoll DJ, Cantu ES, Tranebjaerg L, Mikkelsen M, Watkins PC, Antonarakis SE (1989) Molecular genetic approach to the characterization of the "Down syndrome" region on chromosome 21. Genomics 5:325–331

McInnis MG, Chakravarti A, Blaschak J, Petersen MB, Sharma V, Avramopoulos D, Blouin JL, König U, Brahe C, Cox Matise T, Warren AC, Talbot CC Jr, Broeckhoven C Van, Litt M, Antonarakis SE (1993) A linkage map of human chromosome 21: 43 PCR markers at average intervals of 2.5 cM. Genomics 16:562–571

Nizetic D, Gellen L, Hamvas RMJ, Mott R, Grigoriev A, Vatcheva R, Zehetner G, Yaspo ML, Dutriaux A, Lopes C, Delabar JM, Broeckhoven C Van, Potier MC, Lehrach H (1994) An integrated YAC-overlap and "cosmid-pocket" map of the human chromosome 21. Hum Mol Genet 3:759–770

Orti R, Mégarbane A, Maunoury C, Broeckhoven C Van, Sinet PM, Delabar JM (1997) High-resolution physical mapping of a 6.7-Mb YAC contig spanning a region critical for the monosomy 21 phenotype in 21q21.3–q22.1. Genomics 43:25–33

Osoegawa K, Susukida R, Okano S, Kudoh J, Minoshima S, Shimizu N, Jong PJ de, Groet J, Ives J, Lehrach H, Nizetic D, Soeda E (1996) An integrated map with cosmid/PAC contigs of a 4-Mb Down syndrome critical region. Genomics 32:375–387

Oyake T, Itoh K, Motohashi H, Hayashi N, Hoshino H, Nishizawa M, Yamamoto M, Igarashi K (1996) Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. Mol Cell Biol 16:6083–6095

Patterson D, Rahmani Z, Donaldson D, Gardiner K, Jones C (1993) Physical mapping of chromosome 21. Prog Clin Biol Res 384:33–50

Peterson A, Patil N, Robbins C, Wang L, Cox DR, Myers RM (1994) A transcript map of the Down syndrome critical region on chromosome 21. Hum Mol Genet 3:1735–1742

Reeves RH, Irving NG, Moran TH, Wohn A, Kitt C, Sisodia SS, Schmidt C, Bronson RT, Davisson MT (1995) A mouse model for Down syndrome exhibits learning and behaviour deficits. Nat Genet 11:177–184

Sauer F, Jackle H (1993) Dimerization and the control of transcription by Kruppel. Nature 364:454–457

Schuette J, Viallet J, Nau M, Segal S, Fedorko J, Minna J (1989) Jun-b inhibits and c-fos stimulates the transforming and transactivating activities of c-jun. Cell 22:987–997

Soeda E, Hou DX, Osoegawa K, Atshuchi Y, Yamagata T, Shimokawa T, Kishida H, Soeda E, Okano S, Chumakov I, Cohen D, Raff M, Gardiner K, Graw SL, Patterson D, Jong PD, Ashworth LK, Slezak T, Carrano AV (1995) Cosmid assembly and anchoring to human chromosome 21. Genomics 25:73–84

Straaten F Van, Muller R, Curran T, Beveren C Van, Verma IM (1983) Complete nucleotide sequence of a human c-onc gene: deduced amino acid sequence of the human c-fos protein. Proc Natl Acad Sci USA 80:3183–3187

Tassone F, Xu H, Burkin H, Weissman S, Gardiner K (1995) cDNA selection from 10 Mb of chromosome 21 DNA: efficiency in transcriptional mapping and reflections of genome organization. Hum Mol Genet 4:1509–1518

Yaspo ML, Gellen L, Mott R, Korn B, Nizetic D, Poustka AM, Lehrach H (1995) Model for a transcript map of human chromosome 21: isolation of new coding sequences from exon and enriched cDNA libraries. Hum Mol Genet 4:1291–1304

Zollman S, Godt D, Privé GG, Couderc JL, Laski FA (1994) The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. Proc Natl Acad Sci USA 91:10717–10721