BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **231,** 628–634 (1997)
ARTICLE NO. RC976152

Genomic Organization, Sequence, and Chromosomal Localization of the Human Helix–Loop–Helix Id1 Gene[1]

Jan O. Nehlin,* Eiji Hara,*[,2] Wen-Lin Kuo,*† Colin Collins,*† and Judith Campisi*[,3]
*Department of Cancer Biology and †Molecular Cytogenetics Resource, Berkeley National Laboratory, University of California, Berkeley, California 94720

Received December 16, 1996

The helix–loop–helix protein Id-1 regulates growth and differentiation in many mammalian cells. In human fibroblasts, Id1 and Id1′, a putative splicing variant, are cell cycle regulated, essential for proliferation, repressed by senescence, and overexpressed by some tumor cells. To better understand Id1, we determined the complete sequence, transcriptional start, and localization of the human Id1 gene. Human Id1 has two exons (426 bp and 42 bp), separated by an intron (239 bp). Id1′ results from failure to splice the intron, which encodes 7 amino acids prior to a stop codon. Thus, Id-1 and Id1′ proteins differ only at the extreme C-terminus. Id1 transcription initiated 96 bp upstream of the initiation AUG; 2 kb of upstream sequence stimulated transcription of a reporter gene. Human Id1 maps to chromosome 20 at q11, very close to the centromere but outside the amplicons frequently found in human cancers. © 1997 Academic Press

Id-1 belongs to a small family of helix-loop-helix (HLH) proteins that inhibit transcription factors of the basic HLH (bHLH) family. bHLH transcription factors bind DNA as homo- or heterodimers, and the HLH domain serves as the dimerization interface. Because the basic regions of bHLH proteins comprise half a DNA binding motif, dimerization is required for DNA binding and subsequent transactivation (see 1). In contrast to bHLH proteins, Id proteins contain only a HLH domain and lack a basic region. Thus, heterodimers between bHLH and Id proteins cannot bind DNA or transactivate transcription (2-5).

bHLH proteins belong to a large multi-gene family. Some bHLH genes are widely expressed, whereas others are restricted to only one or a few cell types. By contrast, Id genes are generally expressed in multiple tissues, albeit with some tissue-specific differences in levels. Moreover, only four Id genes (Id1, Id2, Id3 and Id4) have been identified in humans and rodents. One of these, Id1, has two forms that have been proposed to derive from alternative splicing. Thus, there may be five Id proteins that regulate the activity of bHLH transcription factors in diverse mammalian tissues (2-10).

Id1 was first identified in myoblasts, where it prevented the E12/E47 bHLH proteins from dimerizing with MyoD, a muscle-specific bHLH transcription factor (2). Signals that induce myoblasts to differentiate repress Id1 expression, and constitutive Id1 expression prevents myoblast differentiation (11). Id1 is widely expressed among mammalian tissues. Consistent with the findings in muscle, Id1 expression declines when many cell types differentiate, and constitutive Id1 expression blocks differentiation in myeloid and B cell precursors and mammary epithelial cells (12-14). Id1 is mitogen-inducible in human fibroblasts, and this induction is essential for cells to enter the S phase of the cell cycle (8,15). In addition, Id1 is repressed in senescent cells, and overexpressed by some tumor cells (8). In most cell types, the bHLH proteins targeted by Id1 are not yet known.

Because Id1 appears to be so important for the growth, differentiation and senescence of mammalian cells, we characterized a genomic clone containing the human Id1 gene and several kb of 5′ and 3′ flanking sequence. We report here the organization and sequence of the gene, its transcription start site and chromosomal localization, and the activity of 2 kb of promoter sequence human fibroblasts.

MATERIALS AND METHODS

Genomic library screening and sequencing. A genomic library was made by digesting DNA from human placenta with Sau3AI and sub-

[1] The nucleotide sequences reported in this paper were submitted to the GenBank Database Library under Accession Number U57645.

[2] Present address: Imperial Cancer Research Fund, Molecular Oncology Laboratory, 44 Lincoln's Inn Fields, London, WC2A 3PX, United Kingdom.
[3] Corresponding author: Mailstop 70A-1118, Berkeley National Laboratory, Berkeley, CA 94720. Fax: (510) 486-4475. E-mail: JCAMPISI@LBL.GOV.

0006-291X/97 $25.00

Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.

cloning the fragments into the BamHI site in the λEMBLIII vector. The library (1.5×10^6 phage) was screened with the human Id1 cDNA (8). A phage clone containing an 18 kb insert was isolated, digested with SacI, and analyzed by Southern hybridization. A 7.5 kb fragment containing Id1 sequence was subcloned into the SacI site of pBluescriptII SK (Stratagene, La Jolla) in both orientations generating plasmids pBS-IdP-1 (3′ to 5′) and pBS-IdP-15 (5′ to 3′). The latter plasmid was cut with BamHI, generating pBS-IdP15B, and PstI, generating pBS-IdP15P. T3 and T7 primers were used to sequence areas near the polylinkers in pBS-IdP15B and pBS-IdP15P. Further sequence was obtained by subcloning fragments into pUC118. The open reading frame and untranslated regions were sequenced in both orientations. Where necessary, overlapping sequences were obtained using appropriate primers.

Primer extension. The Id1 transcription start site was determined by primer extension using a commercial kit (Promega; Madison WI). Briefly, 10 pmol each of control (kanamycin) or Id1 (3′TGATTCTTGGCGACTGGCTG) primers were labeled with 30 μCi γ^{32}P-ATP (3000 Ci/mmol). The primers were annealed to 2 ng control (kanamycin) RNA or 10 μg total RNA from proliferating WI-38 human fibroblasts, respectively, at 58°C in 40 μl, extended by AMV reverse transcriptase, and analyzed on 8% denaturing polyacrylamide gels, as recommended by the supplier. Labeled Hinf I-digested φX174 DNA was run on the same gel and served as size markers.

PAC library screening. Filters containing an arrayed human PAC library (Genome Systems, Inc., GSI, St. Louis, Missouri) were prehybridized and hybridized with the human Id1 open reading frame (BspHI - EcoNI fragment) labeled with ^{32}P-dCTP by random priming (Rediprime; Amersham Corp.) (16). Briefly, each filter was hybridized at 68°C for 16 h in 10 ml hybridization solution containing 5×10^6 cpm of probe. The filters were washed in 2× SSPE, 0.5% SDS for 5 min at room temperature (RT), 2× SSPE, 0.1% SDS for 15 min at RT, 0.1× SSPE, 0.5% SDS for 30 min at 37°C, 0.1× SSPE, 0.5% SDS for 1 h at °68 C and 0.1× SSPE for 5 min at RT. While still damp, the filters were wrapped in Saran wrap and exposed to film for 3 days. One positive clone was identified, and bacteria containing the corresponding PAC was obtained from GSI (PAC-145-I17, control number 7337). PAC DNA was isolated by alkaline/lysis (Protocol 1, GSI) for Southern analysis and FISH.

Southern analysis. DNA from PAC clone 145-I17 (20 μg) or plasmid pBS-IdP-1 (1 μg) was digested with the indicated enzymes, precipitated and dissolved in water. Digested DNA was separated by electrophoresis, transferred to a membrane, and hybridized overnight at 68° C to a ^{32}P-labeled BspHI - EcoNI Id1 cDNA fragment (16). The membrane was washed as described for screening the PAC library.

Fluorescence in situ hybridization (FISH). PAC clone 145-I17 was localized on human chromosomes by FISH, as described (17). Briefly, PAC DNA was labeled with digoxigenin-dUTP (Boehringer-Mannheim) by nick translation (BRL Life Sciences Technology) to yield 300-800 bp fragments. The probe was hybridized to slides of normal human metaphase spreads, and was detected by FITC-labeled anti-digoxigenin antibody (Boehringer-Mannheim). The FITC signal was localized by DAPI-banding and fractional length measurements, using an imaging system.

Promoter activity. Human fetal lung fibroblasts (WI-38 cells) were transfected with pGL3, a promoter-less luciferase reporter vector (Promega, Inc.), or pId1SBs-luc, containing the Id1 5′ region (SacI - BspHI fragment) (Fig. 1) upstream of the luciferase gene in pGL3, by electroporation using positive (CMV-luc) and normalization (CMV-Bgal) control vectors, as previously described (18). Transfected cells were allowed to attach to 60 mm dishes for 9-12 h in 10% serum, then were washed and shifted to fresh medium containing 10% (growing cells) or 0.2% (quiescent cells) serum. Lysates were prepared 60-94 h later. Some quiescent cells were stimulated with serum for 2 h before lysis. Lysates were assayed for luciferase and β-galactosidase and normalized for transfection efficiency as described (18).

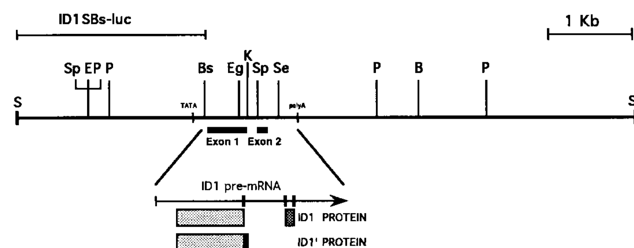

FIG. 1. Organization of the human Id1/Id1′ gene. The map shows the 7.5 kb SacI fragment cloned into pBS-IdP-15, indicating positions of the TATA box, polyadenylation sites, and restriction enzyme cutting sites for SacI (S), SphI (Sp), EcoRI (E), PstI (P), BspHI (Bs), EagI (Eg), KpnI (K), SpeI (Se), and BamHI (B). Shown below the map are the exons (small solid bars) and the intron, the Id1/Id1′ pre-mRNA, and the Id1 and Id1′ proteins (large solid bars) which differ at the C-terminus due to differential splicing of the intron. The SacI-BspHI fragment used to drive luciferase expression (Table 1) is indicated above the map.

RESULTS

Sequence and genomic organization of human Id-1. The human Id1 cDNA (8) was used to isolate an Id1-hybridizing, 7.5 kb SacI fragment from a human genomic library. Cutting sites for several restriction enzymes were mapped on this fragment (Fig. 1), and smaller fragments were subcloned for sequencing. The sequence of the entire 7.5 kb SacI fragment was determined by double stranded sequencing of overlapping, smaller fragments (Fig. 2).

The SacI fragment contained the complete open reading frames of the Id1 and Id1′ cDNAs isolated from human fibroblasts (8) (Figs. 1 and 2). It also contained 2.3 kb of DNA upstream of the translation initiation codon, and 4.5 kb of DNA downstream of the translation stop codon. Thus, the 7.5 kb Sac I fragment contained the complete human Id1 gene plus several kb of 5′ and 3′ untranslated sequence.

The sequence analysis showed that there is a classic TATA box located 130 bp upstream of the translation initiation codon, and a consensus transcription initiator sequence (19) located about 20 bp downstream of this TATA box (Fig. 2). Primer extension analysis using mRNA from normal human fibroblasts showed that the transcriptional start site is located 96 bp upstream of the initiation AUG (Fig. 3). About 50 bp downstream of this transcription start site, the partial 5′ untranslated sequence reported in the human Id1 cDNA (8) is present. Taken together, these data suggest that the human Id1 gene initiates transcription from a classic TATA-containing promoter, generating a 96 bp 5′ untranslated region in the mRNA.

The Id1 coding sequence comprises two exons: a 426 bp 5′ exon, which contains the HLH region, and a 42 bp 3′ exon. These exons are separated by a single 239 bp intron (Fig. 2). From the sequence of the intron, it appears that Id1′ is generated by skipping the 5′ splice

```
***GAGCTCCTTTCTTTAGAGTTGTGAAAAAGATACAGAAGTTGATCAAGCCGGGTGACAGAGTAAGACCCTGTATCAT   -2134
TATTATTTTATTTTGAAACAGAGTCCCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCATAATCTCGGCTCACTACAACC   -2055
TCCGCCTCCTGGGTTCCAGCGAGTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGGTATTACAGGTGCGCGCCACCACACC   -1976
TGGCTAATTTTTCATTTTTAGTAGAGACGGGATTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCTCGTGATC   -1897
TGCCCACCACGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCCGGCCCCCTGTATCCTTTTTTTTTTT   -1818
TTTTTTTTGGAGACGAGTCTCACTCTGTCGCCAGGCTTGAGTGCAGTGGTGCGATCTCGGCTCACTGCAACCTCCGCCT   -1739
CCCAGGTTCAAGCAATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACAGGCGAACGCTACCATGCCCGGCTAATT   -1660
TTTGTATTTTTAGTAGAGATGGGGTTTCACCATATTGGTCAGGCTGGTCTCGAACTCCTGACCTCAGGCAATCCACCCG   -1581
CCTCGGCCTTCCAAATTGTTGGGATTACAGGCGTGAGCCACCGCGCAGCCCTTCATTTTTTAAAATAAGAAATAAATGA   -1502
AAAAAGATGATCAAGACAAACTCTTCATACAGTGCCCGCTGTACAAAGTAAGAAATTAATAATGGTCACGTTTGTGATT   -1423
ATTATTTAAACACCAAGCAGATATTAAATATTAGGGCTCAGAGAGGGCATGCGACCCGCCTGAATTCACTCAGCTGCAG   -1344
AGCTGGAAAGAGAACTCAGGCCTTTTTCCCCACGCTGGAAGGGGTAGCTGGGGCGGAGAGGTGGGGGGAGCGGTGAAGA   -1265
AACCCCAAGCGGCCCAAGCTGTGGGTCTGGGTTGGGAGACTCGCAGGTGTGGGGCGGGGAGGTAAGGTGACCCTTGCTC   -1186
AGCGACCGCGCCCGCAAGAAACGCATTCCCAGGCCTCCCGCCCGGGGTCTGCAGGTGACGGGCTGGGGGGAGCACGGGA   -1107
ACTACCTAGACCAGTTTGTCGTCTCCATGGCGACCGCCCGCGCGGCGCCAGCCTGACAGTCCGTCCGGGTTTTATGAAT   -1028
GGGTGACGTCACAGGCCTGGCGTCTAACGGTCTGAGCCGCTGGTTCAGACGCTGACACAGACCGGCCCGGGAAGGGAGG   -949
GGGGAGACTGTATCTCGCAGCTGCCGCGCCGTGGGAGGGAGACCCTGCTCTGAGGTCTTTGAGAAGAAAATTTAAAAAG   -870
CAGCCAAAAATGGGAAAAAACATTAAAAAATCACGAACTGTTGCAGGTTCAGGAAATTTTTGCAAGGAGCTGCAAATTC   -791
AAGGTGGAATCGAATGCAGCCTCACTCCACTGCGCTCTATCTAGTTCACTTCCCAGCCACCCAGCCCCAAACTTACTAG   -712
ACTTTCCCGAATTAATTGCTCCCACCCGGGAGGGATCTGGGTAGGCCCTCCGGGTCTCAGGAACACGAACAGCAACATT   -633
ATTTAGGAATTGAGAAAGCCCAGGGGAGCAGATGTAAAGGAGCGCCCGCTTTAAATTTCGGTCCTGATATTTGTAGATC   -554
AGGGTCCGAGAAGCATCTTCCAAGAGGGTCCGGAAACCCAGGCTCTCTGGGCTATCCAGAGCCAGCCGCTTCTCCGTTC   -475
GGGCCTCAATTTCTCTCATCTGTGAAATGGAGCTGGAGAAGTGAGAAAGTGAATATGGGAAAAACTGATTCCTGAGGTC   -396
TTTTCATTATAAGGCAATGCCTGCTTAGCTTCCTTGCCTCCTTCCAGAGGAGCCCAGTGCGGCCCTCGCCTCTACCCGC   -317
TGCCCTGCCCGGTGCCCTGCCACCGACCCACCCTTGCTGTTCTGAAACCCGGGTCGTCTTTCCACACTGCGAGCAGGCA   -238
CTAGACGAGCAGGAGGCTGGACCTAGGAGCGCGGGTCACGCCCCCATGCCGCCCATTGGCTGCTTTTGAACGTTCTGAG   -159
CCCGCCCCTCCGGGGGCCGTGGCGTGTTTATAAAAGACAAGCTGTGGCTCCGCACTCTCATTCCACGTTCTTAACTGTT   -80
CCATTTTCCGTATCTGCTTCGGGCTTCCACCTCATTTTTTTCGCTTTGCCCATTCTGTTTCAGCCAGTCGCCAAGAATC   -1
ATG  AAA  GTC  GCC  AGT  GGC  AGC  ACC  GCC  ACC  GCC  GCC  GCG  GGC  CCC  AGC  TGC  GCG  CTG  AAG    60
GCC  GGC  AAG  ACA  GCG  AGC  GGT  GCG  GGC  GAG  GTG  GTG  CGC  TGT  CTG  TCT  GAG  CAG  AGC  GTG    120
GCC  ATC  TCG  CGC  TGC  GCC  GGG  GGC  GCC  GGG  GCG  CGC  CTG  CCT  GCC  CTG  CTG  GAC  GAG  CAG    180
CAG  GTA  AAC  GTG  CTG  CTC  TAC  GAC  ATG  AAC  GGC  TGT  TAC  TCA  CGC  CTC  AAG  GAG  CTG  GTG    240
CCC  ACC  CTG  CCC  CAG  AAC  CGC  AAG  GTG  AGC  AAG  GTG  GAG  ATT  CTC  CAG  CAC  GTC  ATC  GAC    300
TAC  ATC  AGG  GAC  CTT  CAG  TTG  GAG  CTG  AAC  TCG  GAA  TCC  GAA  GTT  GGA  ACC  CCC  GGG  GGC    360
CGA  GGG  CTG  CCG  GTC  CGG  GCT  CCG  CTC  AGC  ACC  CTC  AAC  GGC  GAG  ATC  AGC  GCC  CTG  ACG    420
GCC  GAG  GTG  AGA  TCC  AGA  TCC  GAC  CAC  TAG                                                  450
ATCATCCTTATACCGACGGGGAAACGGAGGCCAGAGAGGGCGTGGGCGCTTGCACCACTTCCGTCCCATCCTTGCGGGT   529
ACCTGGCTATGCGGGGGTGCCTAAGGAGCCTGGAAAAAGCGCTCCCCCGTCGTGCTTCCTGGGGAAGGGGGCGTTCGCT   608
GCGCTCGGAGCGGCGTCCCTTCCAACCCGCCGGTCTCATTTCTTCTCGTTTTCACAG                         665
GCG  GCA  TGC  GTT  CCT  GCG  GAC  GAT  CGC  ATC  TTG  TGT  CGC  TGA                        707
AGCGCCTCCCCCAGGGACCGGCGGACCCCAGCCATCCAGGGGGCAAGAGGAATTACGTGCTCTGTGGGTCTCCCCCAAC   786
GCGCCTCGCCGGATCTGAGGGAGAACAAGACCGATCGGCGGCCACTGCGCCCTTAACTGCATCCAGCCTGGGGCTGAGG   865
CTGAGGCACTGGCGAGGAGAGGGCGCTCCTCTCTGCACACCTACTAGTCACCAGAGACTTTAGGGGGTGGGATTCCACT   944
CGTGTGTTTCTATTTTTTGAAAAGCAGACATTTTAAAAAATGGTCACGTTTGGTGCTTCTCAGATTTCTGAGGAAATTG   1023
CTTTGTATTGTATATTACAATGATCACCGACTGAAAATATTGTTTTACAATAGTTCTGTGGGGCTGTTTTTTTGTTATT   1102
AAACAAATAATTTAGATGGTGGTAAAGTTGTAGTGACTTCTTGGGTTGAGGGGAGGGGCCTGGGCTGGGGGTTGACCCC   1181
TGACCCTCTTCACAGCTGGTTCTGGGAGGAGAATTGGAGGGCTACATCTGGATTGTTGCTCTTACCGGCCTGAATGAGT   1260
GTTTCCGGTGTCTTTAAAGTCTATTCTTCCCCCACCATCCCTGCCCATTTCCCTGCTCCCCAGTGTTATAAAACTAAAT   1339
ACTCCAGACCTCTGGGGTTTCTCACAGTGATCATTGGGCCATCTAGATCACCTGGGGAGGTTGTTAAAATGCAGATGAA   1418
CCTGCCAGGGCCCCCTCTGTATTGGGAGTATTGAGCAAGCCATTTAAGTTGGAGAACCTGGCCGCTGGACCACCCACTT   1497
CCTTGCCAAGATGTGGATAAAGTGAGACCCAGTGGAAAGGGACTTGCTTAGGGCCATAGCTTGTGAGAGGCAGTGATTC   1576
TATCACCCCATTCCCCAGTTGGAGGTGCCTCCCCCAACAGCCTACCTCTGGGTTCCAGCAGGCAGCCCTGGGGCTCCTG   1655
GCTCATCCAAGTGGGTGAGCAAGTTGGCAGGGCCAGGCTGGGCACAGCTGCCGAGACACCTGCTCTGGGTCCAGACTCG   1734
GCCTCCTCACTCTGGCCTGGCTCTGTCGGCAGTTCAAACAAATCTCCCCCACCAACCCCGGGGGTTCTTGGGTTAGGAA   1813
TTTGGAGGAGTGGAACGGATTTTCATTAACCCAGTTCCTGAGCAGGAAAAATATGTCATTGGCAAGGGTCTAAGAATCT   1892
TTCCTGGACATGTATGTGGATTTCTACTCAGGGGAGCATGTATGAGTGTCTACGCCTGTGTATGTAACAGGAGTAAGCT   1971
GATCTTCCCCCACTACTCCCACAGTGGGGATGGGCCAACTGTTCACATATAGGCTTGGGGGAGGGGGTGTAAGTAGGAG   2050
TATGCTTGGGGGTGTGTGCAGAAGGGCAAATGAGAGCACATGTGTTCTGCAGATAAGTGTGGAACCTCCAGCTGTTTGG   2129
GCTTGTGTGTTCTGCCTAAATGCATTGTAGGCATCTCTGCATGGGTCTGGCATGATGGATGGGTTTGTGTGTCTGGGGG   2208
CCACAGAACCACACTGTGGGTTTCTGTGTGTATAACTGTTCTATAGATTGGTTGGCATTCCTGGGCCTCTGCGGGTCTG   2287
GAGTGAAGGTGGGGACAGCTGTGTTTGGGCAAACCCTTCCTCAAAAGTTCACGTGGCAACTCCCTGGGCTATCGTTAGG   2366
TCCTGGGGGCACCAACCCTCAAGGATTGCTGCCAGGTAAACCCCCCTGTATTACATATGGAGAAACTGAGTCCCAGGAA   2445
TGACAAAGAACTTGTCCCAGACAGGTAAGGGCTATTAATATATTTTGTCAATAAACAATAGAATAAAATCAATAGCCCT   2524
TAGCACCAGAAGCCTGGCTCTGCTCAACTGAACCCTTTACAAAAACGATTTCATGGATCC                      2584
```

FIG. 2. Sequence of the human Id1/Id1′ gene. The untranslated sequence is given with no spaces; translated sequence is organized into triplets. Position 1 indicates the first base of the translation initiation codon (underlined). In the 5′ noncoding region, the EGR1 consensus binding site is double underlined, the putative TATA box is underlined, and the transcription start site at position −96 is underlined and in bold. In the coding sequence, the HLH domains are underlined, and the 8 triplets in the intron that encode the C-terminal 7 amino acids and translation stop codon of Id1′ are double underlined. The translation stop codon of Id1 is also underlined. In the 3′ noncoding region, the two putative polyadenylation signals are underlined, as is the last base (G) before the poly-A tail in the Id1 mRNA.

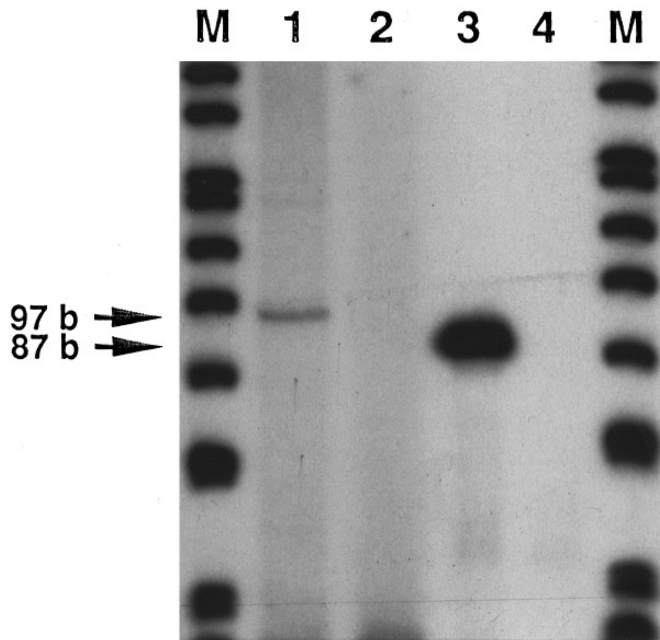

FIG. 3. Primer extension mapping of the Id1/Id1′ transcription start site. The appropriate primers were annealed to total RNA from WI-38 cells or control RNA, extended, and the extension product analyzed, as described in Materials and Methods. Lane 1: RNA from WI-38 cells and Id1 primer. The 5′ nucleotide of the primer is complementary to the A of the initiation codon. Thus, an extension product of 97 b corresponds to a transcription start site 96 bp upstream of the initiation codon. Lane 2: no RNA and Id1 primer. Lane 3: control RNA and control primer. The extended product is 87 b. Lane 4: no RNA and control primer. M = size markers.

donor signal that follows the first exon. This skipping adds 24 bp of translatable sequence from the intron to the first exon. These 24 bp encode the 7 C-terminal amino acids that are unique to Id1′, followed by a stop codon. Thus, the Id1 and Id1′ proteins are identical over the amino acids coded by the first exon, which includes the HLH domain. Id1 and Id1′ differ only at the extreme C-terminus: splicing generates Id1, which contains 13 C-terminal amino acids encoded by the second exon, whereas failure to splice generates Id1′, which contains 7 different C-terminal amino acids derived from the intron (Figs. 1 and 2).

Approximately 400 bp downstream of the translation stop codon in the second exon, there are two consensus poly-adenylation sequences that are separated by a single base. Thus, the Id1 and Id1′ mRNAs have approximately 400 bp and 700 bp, respectively, of 3′ untranslated sequences.

Activity of human Id-1 5′ upstream sequences in human fibroblasts. Id1 mRNA is undetectable in quiescent human fibroblasts, but is induced about 20-fold within 2 h of serum stimulation; it then declines as cells progress through G1, rising again before S phase (8). The Id1 5′ region contains several potential transcription factor binding sites, some of which are listed in Table 1. To determine whether this region contained a functional promoter, and whether Id1 was transcriptionally regulated, we linked 2.2 kb of the 5′ region (SacI - BspHI fragment; see Fig. 1) to a luciferase reporter gene. The resulting plasmid, pId-1SB-luc, was transiently transfected into human fibroblasts and luciferase activity was determined (Table 2).

In proliferating cells, normalized luciferase activity driven by pId-1SB-luc was 20- to 40-fold higher than activity driven by the promoter-less control vector (pGL3). This suggests that 2.2 kb of the Id1 5′ region contains elements sufficient to direct transcription. However, pId-1SB-luc was only slightly less active (about 2-fold) in quiescent cells, compared to proliferating cells; in addition, activity was similar in proliferating and serum-stimulated cells (Table 1). Thus, 2.2 kb of the Id1 5′ region did not confer regulated reporter activity that quantitatively resembled the growth-dependent changes in mRNA (8). These data raise several possibilities about the control of Id1 gene expression.

Id-1 is on the long arm of human chromosome 20, near the centromere. To determine the chromosomal localization of the human Id1 gene, we isolated a P1 artificial chromosome (PAC) clone suitable for fluorescence in situ hybridization (FISH). An arrayed PAC library of human DNA was screened using the Id1 coding sequence as a probe, and one PAC clone (145-I17) hybridized strongly. Digestion with four restriction enzyme combinations and Southern analysis showed that the Id1-hybridizing fragments in clone 145-I17 were identical to those from a similarly digested plasmid containing the 7.5 kb SacI fragment (Fig. 4). We conclude that PAC clone 145-I17 contains the human Id1 gene, and therefore is suitable for FISH.

TABLE 1

Potential Transcription Factor Binding Sites in the 5′ Region of Human Id1

Transcription factor	Position(s)
C/EBP	−249; −613; −782; −802; −1906; −2033; −2284
CREB	−1017; −1594
E47	−1213
ELK/ETS	−515; −762
EGR	−1064
MYB	−472; −998
OCT1	−383; −1418; −2224
SP1	−146; −151; −723; −1206; −1286; −1937

The human Id1 sequence upstream of the TATA box (position −130 in Fig. 2) was analyzed using TFSEARCH and the TRANSFAC MATRIX TABLE database. The search was limited to vertebrate transcription factor binding sites and scores of 85 or better. Position(s) refers to those indicated in Fig. 2. The Table lists only a subset of potential binding sites, those that are relatively common and/or are likely to be important mediators of growth-regulated expression in fibroblasts.

PAC clone 145-117 hybridized to chromosome 20q11, determined by FISH analysis of normal human metaphase chromosomes (Figs. 5). The fractional length of the hybridization signal from the p-terminus (FLpter) of chromosome 20 was 0.500 ± 0.008 (mean ± SE), based on measurements from 10 independent chromosomes. Thus, Id1 localizes to human chromosome 20q11, very near the centromere (Fig. 6). Our measurements place Id1 centromeric to Bcl-X and E2F1 (Fig. 6), which are located at FLpter 0.539 and 0.541, respectively (17). Id1 maps outside the region of chromosome 20q13 that is amplified in 9.6% of human breast tumors (20), and outside the region at 20q11 that is amplified in 6.2% of human tumors (21). Interphase FISH showed that Id-1 was not amplified in BT474 cells (data not shown), which harbor the amplicon at 20q13 (20).

DISCUSSION

Id1 is a negative regulator of bHLH transcription factors, the first of four such genes to be cloned (2). Id1 was first identified as a negative regulator of myogenic bHLH transcription factors. Subsequently, it was shown to be widely expressed among embryonic and adult mammalian tissues, and to control the growth and/or differentiation of several cell types (8,9,13,18,22).

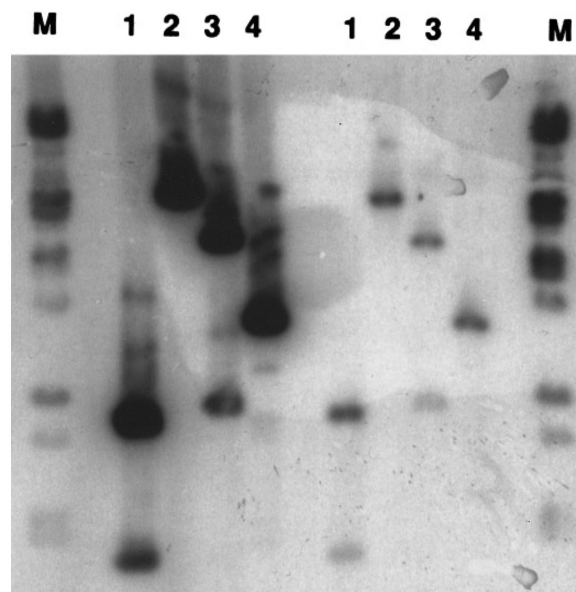

FIG. 4. Southern analysis of plasmid pBS-IdP-1 and PAC clone 145-I17 for Id1 coding sequences. DNA was digested with the following restriction enzymes: PstI and BspHI (lanes 1); SacI (lanes 2); SacI and BspHI (lanes 3); PstI (lanes 4). The left-hand set of lanes 1–4 is plasmid pBS-IdP-1 (1 μg); the right-hand set of lanes 1–4 is PAC clone 145-I17 (20 μg); lanes marked M are DNA markers. The digested DNA was separated, transferred to a membrane, and the membrane was probed with the human Id1 coding sequence, as described in Materials and Methods.

TABLE 2

Promoter Activity of 2.2 kb of 5′ Upstream Sequence of Human Id1

Cell condition	Vector	Relative normalized activity
Growing[a]	pGL-3	1
Growing[a]	pCMV-luc	2920 [940-6280]
Growing[b]	pId1SB-luc	41 [31-52]
Quiescent[c]	pId1SBs-luc	22 [9-40]
Stimulated, 2 h[d]	pId1SBs-luc	39 [20-63]

Proliferating cells were electroporated with 1 μg of reporter vector and 0.5 μg of normalization vector (CMV-Bgal) as described (18). pGL3 is a negative control (no promoter); pCMV-luc, is a positive control (human cytomegalovirus early promoter); pId1SBs-luc contains the human Id1 promoter and 2 kb of 5′ sequence (Fig. 1). Electroporated cells were allowed to attach for 9-12 h in 10% serum, with several medium changes to remove cell debris, and then were given 10% (Growing) or 0.2% (Quiescent) serum for 60-94 h. Quiescent cells were given 10% or 20% serum for 2 h (Stimulated). Lysates were prepared and assayed for luciferase and β-galactosidase as described (18). For each sample, luciferase activity was normalized for β-galactosidase activity. The normalized values for CMV-luc and Id1SBs-luc were divided by the normalized value for pGL-3 to obtain a Relative Normalized Activity. Within an experiment, each transfection was done in duplicate. The range of values obtained for multiple experiments are in brackets.

[a] Average of 4 experiments; lysates were prepared 85-94 h after electroporation.

[b] Average of 3 experiments; lysates were prepared 90-94 h after electroporation.

[c] Average of 5 experiments; cells recovered from electroporation for 9-12 h, then were shifted to 0.2% serum for 60-90 h before lysates were prepared.

[d] Average of 3 experiments; cells recovered from for 9-10 h, were shifted to 0.2% serum for 85-90 h, and were stimulated with 10% or 20% serum for 2 h.

In human fibroblasts, two Id1 cDNAs appeared to derive from alternate splicing (8). Our sequence shows how these cDNAs arise. Id1 derives from standard joining of the two exons, after intron removal by splicing. Id1′, by contrast, derives from failure to splice the intron, the first 24 bp of which specifies 7 amino acids followed by a stop codon. Because the first exon, which is shared by Id1 and Id1′, is much larger (426 bp) than the second (42 bp), the Id1 and Id1′ proteins are identical except at the extreme C-terminus. The 13 C-terminal amino acids of Id1 derive from the second exon, whereas the 7 C-terminal amino acids of Id1′ derive from the intron. The Id1 and Id1′ mRNAs also differ in size. This difference, and possible heterogeneity in polyadenylation, may explain why Id1 transcripts frequently appear as a diffuse band on northern blots. The functional significance of the differential splicing of human Id1 is not yet known. In rat cardiac muscle, alternative splicing inserts 214 bp into the Id1 coding region, which alters its dimerization properties (9). In the human gene, however, Id1 and Id1′ have identical HLH regions. However, Id1, but not Id1′, comple-

mented a defective viral oncogene (SV40 T antigen) which cannot otherwise stimulate senescent fibroblasts to initiate DNA synthesis (18). We presume this difference in biological activity resides in the C-terminus of the Id1 and Id1′ proteins.

The human Id1 5′ region contains sequences sufficient to drive transcription of a reporter gene. This region has many potential transcription factor binding sites, but we do not yet know which are functional. Because it did not confer strong growth-regulated expression on a transiently transfected reporter, Id1 regulation may require additional 5′ sequences, or intronic or 3′ untranslated sequences. In the murine Id1 gene, an enhancer needed for expression in B cells resides in the 3′ untranslated region (23). It is also possible that the Id1 5′ region contains elements that either oblate repression or diminish induction in transient transfection assays. In the murine Id1 promoter, sequences >1.5 kb upstream of the transcription start site reduced induction by serum in transient transfections, and a site that binds the early growth response transcription factor Egr1 was critical for induction (24). Human Id1 contains a similarly positioned Egr1 binding site (Fig. 2; Table 1). In addition, integration or association with a nuclear matrix may be critical for growth-regulated expression. Finally, Id1 mRNA levels may be posttranscriptionally regulated in human fibroblasts.

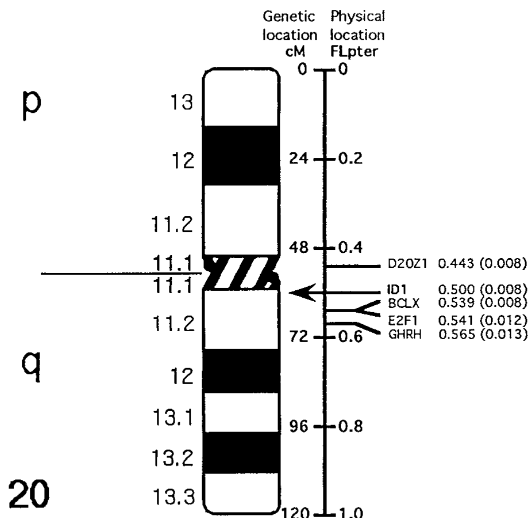

FIG. 6. Position of Id1 relative to other markers on human chromosome 20. Id1 localizes to the q arm at FLpter 0.500 near the centromeric marker D2021 at FLpter 0.443. Telomeric to Id1 are the genes Bcl-X (0.539), E2F1 (0.541), and GHRH (0.565).

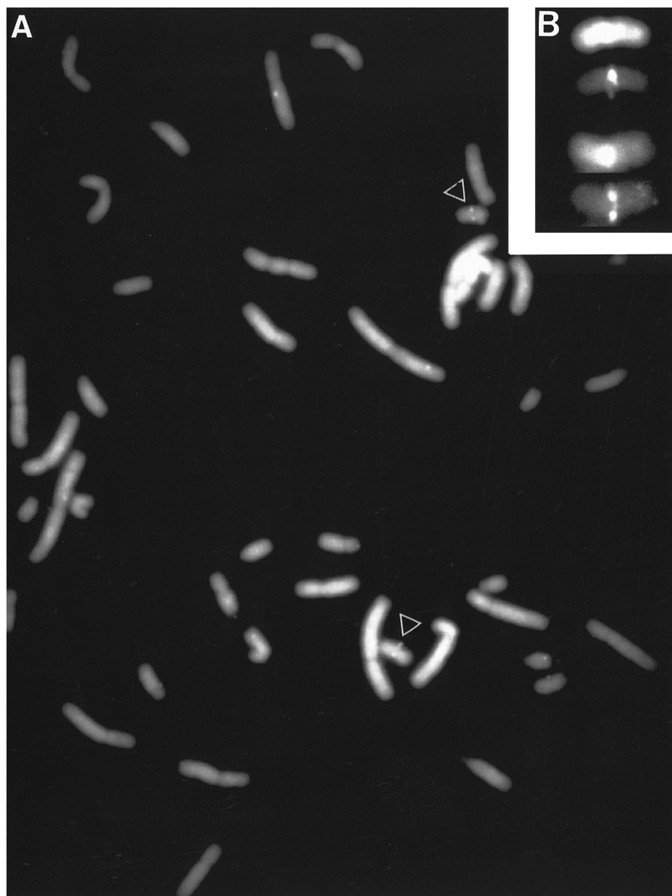

FIG. 5. Localization of Id1 to human chromosome 20q11. (A) FISH of PAC clone 145-I17 to human metaphase chromosomes. The small arrowheads indicate the FISH signal on each of the two chromosomes 20. (B) Higher magnification of four chromosomes 20 from two additional metaphase spreads, showing the centromeric FISH signal.

Localization of Id1 to human chromosome 20q11 places it well outside the well-characterized amplicons that are present in a significant proportion of human breast cancers (20,21). This localization refines that reported earlier (25). It was somewhat disappointing, however, because we have shown that repression of Id1 is critical for normal differentiation of breast epithelial cells in culture, and ectopic Id1 expression induces an invasive phenotype in these cells (13). However, our recent data suggest that failure to repress Id in response to growth arrest or differentiation signals, rather than Id1 overexpression, correlates with the invasive potential of human breast cancer cells (Lin, Nehlin, Campisi and Desprez, unpublished).

Repression of Id1 expression is important for senescence in normal human cells (18). Although Id2 was shown to bind and neutralize growth suppression by the retinoblastoma tumor suppressor (pRb) in human tumor cells (26), Id1 -- but not Id2 -- bypassed the requirement for pRb inactivation when senescent human fibroblasts were stimulated by the viral oncogene T antigen (18). In this regard, amplification of chromosome 20q was recently implicated in the immortalization of human uroepithelial cells that express the viral E7 gene, which inactivates pRb (27). Thus, Id1 may cooperate with cellular or viral genes to inactivate func-

tions of pRb required for the finite replicative life span of normal human cells (18).

In summary, we provide the complete sequence of the human Id1 gene, including >2 kb of 5′ sequence and >3 kb of 3′ sequence. The sequence explains the origins of the Id1 and Id1′ isoforms that have been described. We also show that the human Id1 gene is localized very near the centromere at 20q11. The cloning of the human Id1 flanking sequences, including those present in the 80–100 kb PAC clone, will facilitate understanding how Id1 expression is regulated during normal and abnormal growth and differentiation.

ACKNOWLEDGMENTS

We thank Drs. Claudia Lin and Pierre Desprez for helpful discussions. This work was supported by National Institute on Aging grants AG11658 and AG09909, and the Department of Energy under contract DE-AC0376SF00098 to the University of California. J.O.N. was supported in part by a long-term EMBO fellowship.

REFERENCES

1. Baxevanis, A. D., and Vinson, C. R. (1993) *Curr. Opin. Genet. & Dev.* **3,** 278–285.
2. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990) *Cell* **61,** 49–59.
3. Sun, X. H., Copeland, N. G., Jenkins, N. A., and Baltimore, D. (1991) *Molec. Cell. Biol.* **11,** 5603–5611.
4. Christy, B. A., Sanders, L. K., Lau, L. F., Copeland, N. G., Jenkins, N. A., and Nathans, D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1815–1819.
5. Reichmann, V., van Cruchten, I., and Stablitzky, F. (1994). *Nucl. Acids Res.* **22,** 749–755.
6. Biggs, J., Murphy, E. V., and Israel, M. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1512–1526.
7. Deed, R. W., Bianchi, S. M., Atherton, G. T., Johnston, D., Santibanez-Koref, M., Murphy, J. J., and Norton, J. D. (1993) *Oncogene* **8,** 599–607.
8. Hara, E., Yamaguchi, T., Nojima, H., Ide, T., Campisi, J., Okayama, H., and Oda, K. (1994). *J. Biol. Chem.* **269,** 2139–2145.
9. Springhorn, J. P., Singh, K., Kelly, R. A., and Smith, T. W. (1994) *J. Biol. Chem.* **269,** 5132–5136.
10. Pagliuca, A., Bartoli, P. C., Saccone, S., Della Valle, G., and Lania, L. (1995) *Genomics* **27,** 200–203.
11. Jen, Y., Weintraub, H., and Benezra, R. (1992) *Genes & Dev.* **6,** 1466–1479.
12. Kreider, B. L., Benezra, R., Rovera, G., and Kadesch, T. (1992) *Science* **255,** 1700–1702.
13. Desprez, P. Y., Hara, E., Bissell, M., and Campisi, J. (1995) *Molec. Cell. Bio.* **15,** 3398–3404.
14. Sun, X. H. (1994) *Cell* **79,** 893–900.
15. Barone, M. V., Pepperkok, R., Peverali, F. A., and Philipson, L. (1994) *Proc. Nat'l Acad. Sci. USA* **91,** 4985–4988.
16. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994) Current Protocols in Molecular Biology, John Wiley & Sons, New York.
17. Stokke, T., Collins, C., Kuo, W-L., Kowbel, D., Shadravan, F., Tanner, M., Kallioniemi, A., Kallioniemi, O-P., Pinkel, D., Deaven, L., and Gray, J. W. (1995) *Genomics* **26,** 134–137.
18. Hara, E., Uzman, J. A., Dimri, G. P., Nehlin, J. O., Testori, A., and Campisi, J. (1996) *Dev. Genetics* **18,** 161–172.
19. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S. T. (1994) *Molec. Cell. Bio.* **14,** 116–127.
20. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W-L., Waldman, F. M., Isola, J. J., Gray, J. W., and Kallioniemi, O. P. (1994) *Cancer Res.* **54,** 4257–4260.
21. Tanner, M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjarvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., and Kallioniemi, O. P. (1996) *Cancer Res.* **56,** 3441–3445.
22. Evans, S. M., and O'Brien, T. X. (1993) *Dev. Biol.* **159,** 485–499.
23. Saisanit, S., and Sun, X. H. (1995). *Mol. Cell. Bio.* **15,** 1513–1521.
24. Tournay, O., and Benezra, R. (1996) *Mol. Cell. Bio.* **16,** 2418–2430.
25. Matthew, S., Chen, W., Murty, V., Benezra, R., and Chaganti, R. S. K. (1995) *Genomics* **30,** 385–387.
26. Lasorella, A., Iavarone, A., and Israel, M. (1996) *Mol. Cell. Bio.* **16,** 2570–2578.
27. Reznikoff, C. A., Belair, C., Saveelieva, E., Zhai, Y., Pfeifer, K., Yeager, T., Thompson, K. J., Devries, S., Bindley, C., Newton, M. A., Sekhon, G., and Waldman, F. (1994). *Genes & Dev.* **8,** 2227–2240.