BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **243,** 317–325 (1998)
ARTICLE NO. RC988092

A Novel Extracellular Domain Variant of the Human Integrin α7 Subunit Generated by Alternative Intron Splicing[1]

Euphemia Leung,* Siew Pheng Lim,*,[2] Randy Berg,* Yi Yang,* Jian Ni,†
Shu-xia Wang,† and Geoffrey W. Krissansen*,[3]

**Department of Molecular Medicine, School of Medicine and Health Sciences, University of Auckland, Auckland, New Zealand; and †Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, Maryland 20850*

Received November 26, 1997

The integrin $\alpha7\beta1$ laminin receptor, which is expressed on replicating myoblasts, and upregulated during myogenic differentiation, is involved in cell adhesion and communication between muscle cells and the extracellular matrix. It is a major cell-surface substrate in skeletal muscle cells for the cell-surface, arginine-specific, ADP-ribosyltransferase. Both the extracellular and cytoplasmic domains of the mouse $\alpha7$ subunit undergo alternative splicing during development, generating differentially expressed variants with presumably unique ligand-binding and signalling properties. Here human cDNA clones isolated from a fetal heart λgt10 cDNA library encoded the complete sequence of the $\alpha7$ subunit and hybridised to a single major 4.4 kb $\alpha7$ subunit transcript abundantly expressed in human skeletal muscle, moderately expressed in heart, and weakly expressed in most other tissues. One clone out of four contained a novel 225-nucleotide in-frame deletion corresponding to 75 amino acids in the C-terminal region of the extracellular domain. The variant, whose expression appears to be tissue-specific, is created by alternative splicing at sites flanking an intron in the $\alpha7$ gene. A related mouse form was identified in P19 embryonal carcinoma cells. Deletion of the spliced region, which either contains or is in very close proximity to the major ADP-ribosylation site of the $\alpha7$ subunit, may serve to modulate the effects of ADP-ribosylation, or alternatively molecular associations, and receptor-ligand affinity.

The integrin $\alpha7$ subunit associates with the integrin $\beta1$ subunit, forming the $\alpha7\beta1$ laminin receptor expressed on replicating myoblasts [1]. It may induce changes in the shape and mobility of myoblasts, and facilitate their localisation at the laminin-rich sites of secondary fiber formation [2-6]. With myoblast terminal differentiation there is a stable increase in the expression of $\alpha7$ prior to cell fusion. $\alpha7$ localizes between adult fibers and the surrounding matrix [7], and at myotendinous junctions which stabilize the muscle [6].

The integrin $\alpha7$ subunit is highly similar to the $\alpha3$ and $\alpha6$ subunits, forming a subgroup of integrins which bind the E8 fragment of laminin [8, 9]. Members of this subgroup contain an array of variant forms with alternative extracellular and intracellular domains. There are two different extracellular domain isoforms of mouse $\alpha7$, designated X1 and X2, which are derived from the mutually exclusive alternative splicing of the region between the III and IV homology repeat domains near the putative ligand-binding site [10]. The isoforms are present in equal amounts in mouse skeletal myoblasts and adult heart, but only the X2 variant is expressed in mature skeletal muscle. The mouse $\alpha7$ cytoplasmic domain is also alternatively spliced, giving rise to A and B variants that are similar to corresponding $\alpha3$ and $\alpha6$ variants [10, 11]. The $\alpha7$ B mRNA was detected in all tissues and cell types tested, whereas the $\alpha7$ A variant was exclusively detectable in skeletal muscle, and following the differentiation of myoblast cells. The rat also has A, B, and C variants, where the A and C variants are induced upon terminal myogenic differentiation, and in contrast the expression of the B form diminishes upon differentiation [6]. Another potential mechanism for modulating $\alpha7$ integrin-mediated signalling pathways involves ADP-ribosylation of the $\alpha7$ subunit by the cell-surface, glycosylphosphatidylinositol-anchored, arginine-specific ADP-ribosyltransferase [12]. This enzyme can modify at

[1] Sequence data from this article have been deposited with the GenBank Database under Accession No. AF032108.

[2] Current address: Institute of Molecular and Cell Biology, 15 Lower Kent Ridge Road, Singapore 119078.

[3] Corresponding author. Fax: 0064 9 373 7492. E.mail: gw.krissansen@auckland.ac.nz.

Abbreviations: DMSO, dimethysulfoxide; EST, expressed sequence tag; PCR, polymerase chain reaction.

0006-291X/98 $25.00

least two sites in the extracellular domain of the integrin $\alpha 7$ subunit in skeletal myotubes.

To date, preliminary sequence analysis of fragments of human $\alpha 7$ cDNA has detected only the X2 extracellular, and B intracellular domain variants [6, 10]. Here we report the complete sequence for human $\alpha 7$, and the identification of a novel extracellular domain variant lacking a region either containing or in very close proximity to the major ADP-ribosylation site.

MATERIALS AND METHODS

Screening of cDNA library and sequencing. A human fetal heart 5′-stretch λgt10 cDNA library was screened with the 5′ 600 bp EcoRI/EcoR I fragment (nucleotide positions 1322 to 1886) of the cDNA insert of the clone HOEA062, and subsequently with a 240 bp PCR-generated extreme 5′-fragment of clone FH5 by using standard procedures. The latter probe was isolated by using the M13 forward primer, and the anti-sense primer $\alpha 7$-720L (5′-CTTCCATTCCCC-ACCATCCAACTC-3′, nucleotide positions 424 to 447). DNA sequences were determined by cycle sequencing using an Applied Biosystems 373A automated DNA sequenator (The Centre for Gene Technology, School of Biological Sciences, University of Auckland, Auckland, New Zealand). The $\alpha 7$ subunit sequence was obtained on both strands using a combination of universal M13 primers, and sequence-specific primers.

Northern analysis. MTN filters (Clontech) were screened with the ^{32}P-labelled 600 bp EcoRI/EcoR I fragment of the insert of clone HOEA062. The conditions of hybridization were 1% SDS, 2 × SSC (2 × SSC is 0.3 M NaCl, 30 mM sodium citrate, pH 7), 10% (w/v) dextran sulphate, 100 μg/ml denatured salmon sperm DNA, and 50% (v/v) deionized formamide at 42°C. Filters were washed twice in 0.1 × SSC, 0.1% SDS at 50°C for 30 min, and autoradiographed.

PCR amplification. Clones isolated from the fetal heart cDNA library were screened by PCR to identify 5′ sequences, by using the primer $\alpha 7$-720L and the λgt10 forward and reverse primers. DNA extracted from human cDNA libraries, and reverse-transcribed human small intestine and fetal heart mRNAs (Clontech) were subjected to 30 cycles of PCR amplification to identify alternative splice variants, with each cycle composed of a denaturation step at 95°C for 30 secs, annealing at 60°C for 60 secs, and extension at 72°C for 60 secs. The PCR reaction included 50 pmol of each of the primers $\alpha 7$-1820U (nucleotide positions 1536-1559; 5′-CCTGAGCCGTAA-CCTGGAAGAACC-3′), and $\alpha 7$-91r2 (nucleotide positions 2296-2332; 5′-CAGCTCAATGAAGACACGGGC-3′); 200 μM deoxytriphosphate (Boehringer Mannheim, Mannheim, Germany), and 2.5 units of Taq DNA polymerase (Boehringer Mannheim). PCR amplification was also carried out using cDNA prepared from total RNA isolated [13] from the mouse embryonal carcinoma cell line P19 [14] (from American Type Culture Collection) as above, but using the mouse $\alpha 7$-specific primers m$\alpha 7$-S (nucleotide positions 2008-2026; 5′-CCCACT-GTGGCTCCCATCC-3′) and m$\alpha 7$-R (nucleotide positions 2390-2409; 5′-ATGAGAGGCGCTGTCATTGG-3′), which span a region corresponding to the deleted human $\alpha 7$ sequence. P19 cells, cultured in alpha-MEM, 10 % FCS, and additives, had been differentiated with either 1% DMSO, or 1 nM retinoic acid for 5 days, followed by a further 5 days in the absence of differentiating agent.

RESULTS

Cloning of Human α7 cDNA and Identification of a Novel Splice Variant

A homology search [15] of the human EST database of Human Genome Sciences Inc. and The Institute for Genomic Research [16, 17], identified a human integrin $\alpha 7$ subunit cDNA clone HOEA062 from a human osteoblast cDNA library (Fig. 1A). The HOEA062 clone was sequenced on both strands, and alignment with the mouse and rat sequences revealed it encoded the C-terminal half of the protein. In order to isolate the missing 5′-half, a λgt10 cDNA library prepared from fetal heart was screened with a 600 bp 5′-fragment of clone HOEA062, and more than 60 positive clones identified. The longest clone, FH7.5, from a screen of sixteen positives was near full-length, but was missing sequence encoding the first 70 amino acid residues. Surprisingly this clone encoded a novel variant form containing a 225-nucleotide in-frame deletion corresponding to 75 amino acids in the C-terminal region of the extracellular domain. This same region was included in a conventional form of the $\alpha 7$ subunit, represented by clone FH7.6. The 5′-end of human $\alpha 7$ was isolated by first rescreening the fetal heart library with a 240 bp extreme 5′ fragment of FH7.5, and then performing 5′ anchor PCR with λ forward and reverse primers, and an antisense $\alpha 7$-specific primer, on all positive clones. By this approach clone FH7.4 was identified, which encoded the N-terminus, and 161 bp of 5′-untranslated sequence.

Structure of the Human α7 Subunit

The nucleotide and deduced amino acid sequence of the complete $\alpha 7$ subunit derived from the composite cDNA is shown in Fig. 1B. The 4061 bp sequence has one open reading frame extending from nucleotide position -101 to a termination codon TAG at nucleotide position 3313. There are 489 bp of 3′-untranslated sequence containing a ATTAAA polyadenylation signal site, followed 15 bases further by a poly(A) stretch. The predicted molecular weight of an unglycosylated form of the mature protein is 121 kDa. The presumptive translational start codon at nucleotide position -101 is unusual like the mouse $\alpha 7$ sequence in being flanked by sequence that does not resemble a classical Kozak consensus, PurNNAUGPur [18]. As mentioned previously [10] the unusual context of such start codons is rare, and may play a role in modulating protein production [19]. It is followed by a predominantly hydrophobic signal peptide stretch of 33 amino acids, a predominantly hydrophilic extracellular domain of 1004 amino acids, a 23-amino acid transmembrane domain, and a cytoplasmic domain of 77 amino acid residues. There are multiple potential dibasic cleavage sites RK, RR, RR, KR, RRRRE, KKK (positions 265, 501, 570, 722, 878, 909 respectively) in the extracellular domain; five potential N-linked glycosylation sites (positions 53, 709, 912, 948, 968); three potential calcium binding sites which conform to the consensus DXD/NXD/NGXXD; and an LDV sequence (position 871) (Fig. 1B). Inspection of the sequences of clones encoding repeat domains III and IV revealed that the

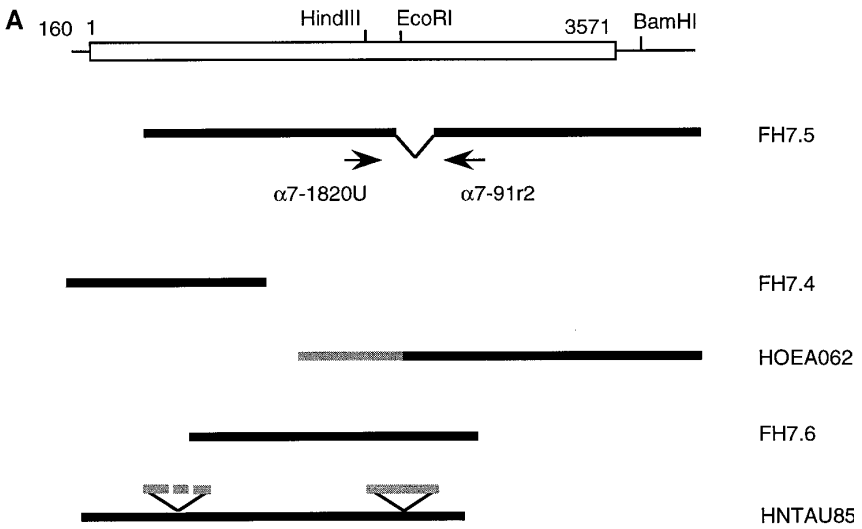

FIG. 1. Nucleotide and deduced amino acid sequence of the human $\alpha 7$ subunit. (A) Schematic representation of the $\alpha 7$ cDNA clones. The relative positions of the $\alpha 7$ cDNA clones are depicted against a composite diagram of the full-length $\alpha 7$ cDNA. The open boxed region denotes the coding sequence, and the thin lines represent either regions that are spliced in/out, or 3′ and 5′-untranslated regions. Intronic regions are marked by shaded lines above the cDNA, where a dashed line indicates that the intron has not been completely sequenced. Arrows refer to the positions of PCR primers. The 5′-end of the clone HOEA62 contained sequence (shaded) unrelated to an integrin. (B) Nucleotide and deduced amino acid sequences of the $\alpha 7$ subunit. The numbers in the left margin refer to nucleotide and amino acid positions, respectively. Nucleotide position 1 has been assigned to the first nucleotide encoding the first amino acid of the mature sequence. Five potential sites for N-linked glycosylation are indicated by dots. The putative signal peptide, and transmembrane domains are underlined. The ×2 region, between the III and IV homology repeat domains, that is alternatively spliced in the mouse $\alpha 7$ subunit is overlined, as is a putative polyadenylation signal site. Three potential calcium binding sites in the amino-terminal half of the molecule are enclosed by boxing. Consensus sequences for potential dibasic protease cleavage sites, and a potential integrin LDV binding site are underlined. The sequence that is shaded and italicized illustrates the region that is deleted in the novel $\alpha 7$ splice variant. The region containing the major ADP-ribosylation site is bracketed. The positions of intron-exon splice sites identified by partial sequencing of the HNTAU85 clone are indicated by vertical arrows.

intervening sequences between the latter domains for all the clones corresponded to the alternatively spliced mouse exon X2. The cytoplasmic domain is highly similar to the mouse $\alpha 3$ [20], human $\alpha 6$ [20-22], and mouse [10] and rat [6] $\alpha 7$ B cytoplasmic variants, indicating that it is the B cytoplasmic isoform. Features of the cytoplasmic domain of the human B isoform have already been discussed in detail [6]. Briefly, it contains serine/threonine protein kinase homology regions [AVKIL(P)R, and LGXXGXXGXG]; three unique DXHP repeats; a receptor-like protein tyrosine phosphatase homology region (TVPQYHAVKIPREDR); and a potential actin-binding site (QQFKEEK).

The Novel Extracellular Variant of the Human α7 Subunit Is Generated by Alternative Intron Splicing

The 225-nucleotide in-frame deletion corresponding to 75 amino acids (positions 625 to 699) in the C-terminal region of the extracellular domain of $\alpha 7$ (clone FH7.5), has no notable features that might provide a clue to a particular function. Isolation of a new EST clone, HNTAU85, provided insights into how this variant form of the $\alpha 7$ subunit arises. The HNTAU85 clone was a partially processed hnRNA containing intronic sequences including a 475 bp intron within the novel alternatively spliced region. Splicing at alternative sites flanking this intron appears to be the mechanism that gives rise to the $\alpha 7$ subunit variant represented by clone FH7.5. The alternative splice sites flanking the intron are sited 32 bp from the 5′-end and 193 bp from the 3′-end of the intron (Fig. 2A). To confirm the presence of this previously unidentified $\alpha 7$ mRNA variant, PCR analysis was carried out on reverse-transcribed human fetal heart and small intestine mRNA, and mRNA prepared from mouse P19 embryonal carcinoma cells, using primers that span the deleted region. When using fetal heart cDNA, a major PCR product of 796 bp, and a minor species of 571 bp, were generated that are of the expected sizes of the unspliced and spliced versions of the $\alpha 7$ subunit, respectively; given the positions of the PCR primers in the human $\alpha 7$ cDNA (Fig. 2B). In contrast only the unspliced 796 bp band was detected with small intestine cDNA. Similar results were obtained using P19 cells, with the expected fragment size of 400 bp for the unspliced version, and a 250 bp spliced version (Fig. 2C). The relative proportions of the species was maintained in P19 cells that had been forced to differentiate into cardiac and skeletal muscle cells in the presence of DMSO and low doses of retinoic acid. The above PCR products hybridised to a human $\alpha 7$ cDNA probe, thereby verifying their authenticity.

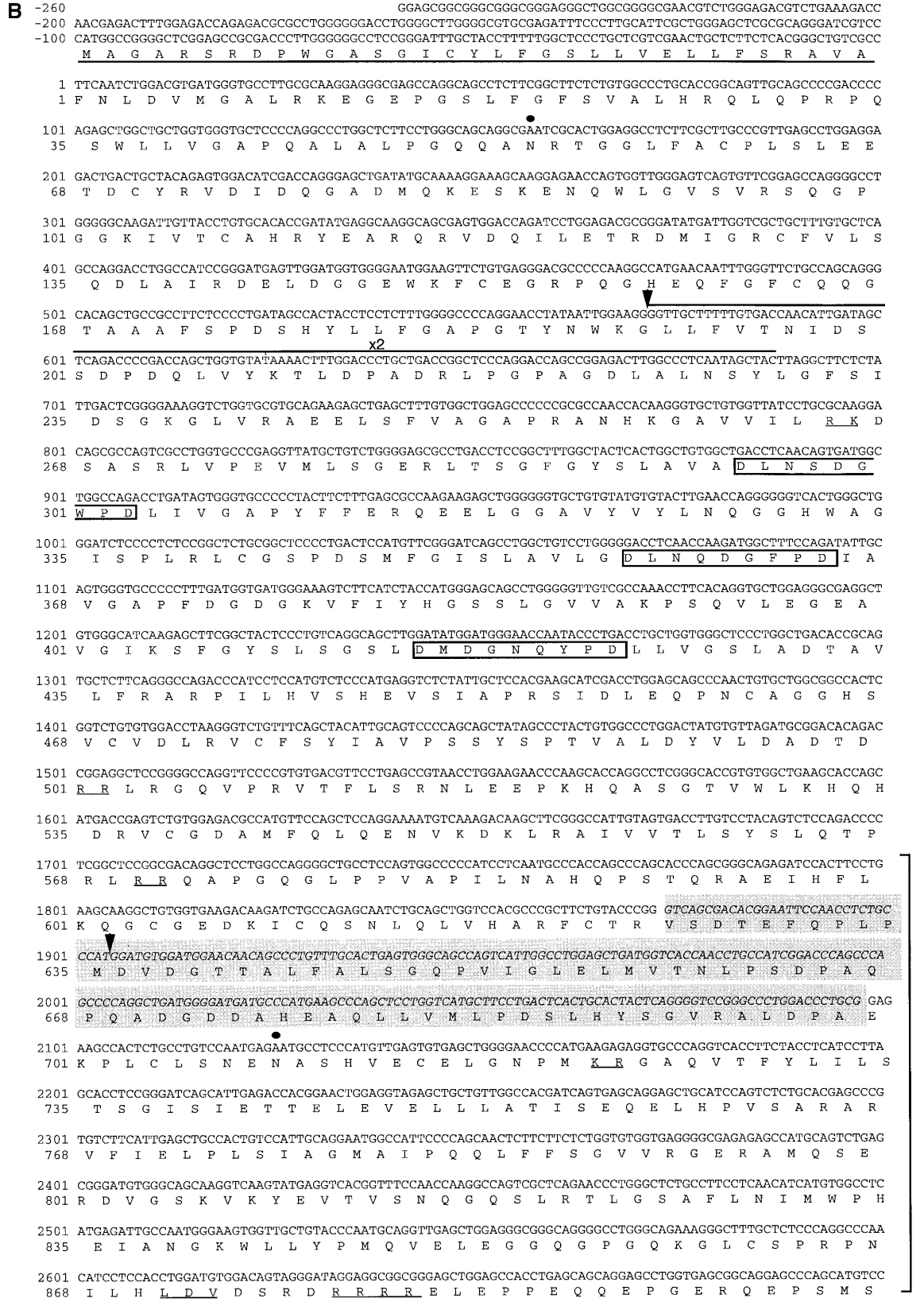

FIG. 1—*Continued*

```
2701 GGTGGCCAGTGTCCTCTGCTGAGAAGAAGAAAAACATCACCCTGGACTGCGCCCGGGGCACGGCCAACTGTGTGGTGTTCAGCTGCCCACTCTACAGCT
 901 W  W  P  V  S  S  A  E  K  K  K  N  I  T  L  D  C  A  R  G  T  A  N  C  V  V  F  S  C  P  L  Y  S  F

2801 TTGACCGCGCGGCTGTGCTGCATGTCTGGGGCCGTCTCTGGAACAGCACCTTTCTGGAGGAGTACTCAGCTGTGAAGTCCCTGGAAGTGATTGTCCGGGC
 935   D  R  A  A  V  L  H  V  W  G  R  L  W  N  S  T  F  L  E  E  Y  S  A  V  K  S  L  E  V  I  V  R  A

2901 CAACATCACAGTGAAGTCCTCCATAAAGAACTTGATGCTCCGAGATGCCTCCACAGTGATCCCAGTGATGGTATACTTGGACCCCATGGCTGTGGTGGCA
 968  N  I  T  V  K  S  S  I  K  N  L  M  L  R  D  A  S  T  V  I  P  V  M  V  Y  L  D  P  M  A  V  V  A

3001 GAAGGAGTGCCCTGGTGGGTCATCCTCCTGGCTGTACTGGCTGGGCTGCTGGTGCTAGCACTGCTGGTGCTGCTCCTGTGGAAGATGGGATTCTTCAAAC
1001 E  G  V  P  W  W  V  I  L  L  A  V  L  A  G  L  L  V  L  A  L  L  V  L  L  L  W  K  M  G  F  F  K  R

3101 GGGCGAAGCACCCCGAGGCCACCGTGCCCCAGTACCATGCGGTGAAGATTCCTCGGGAAGACCGACAGCAGTTCAAGGAGGAGAAGACGGGCACCATCCT
1035   A  K  H  P  E  A  T  V  P  Q  Y  H  A  V  K  I  P  R  E  D  R  Q  Q  F  K  E  E  K  T  G  T  I  L

3201 GAGGAACAACTGGGGCAGCCCCCGGCGGGAGGGCCCGGATGCACACCCCATCCTGGCTGCTGACGGGCATCCCGAGCTGGGCCCCGATGGGCATCCAGGG
1068  R  N  N  W  G  S  P  R  R  E  G  P  D  A  H  P  I  L  A  A  D  G  H  P  E  L  G  P  D  G  H  P  G

3301 CCAGGCACCGCCTAGGTTCCCATGTCCCAGCCTGGCCTGTGGCTGCCCTCCATCCCTTCCCCAGAGATGGCTCCTTGGGATGAAGAGGGTAGAGTGGGCT
1101 P  G  T  A

3401 GCTGGTGTCGCATCAAGATTTGGCAGGATCGGCTTCCTCAGGGGCACAGACCTCTCCCACCCACAAGAACTCCTCCCACCCAACTTCCCCTTAGAGTGCT
3501 GTGAGATGAGAGTGGGTAAATCAGGGACAGGGCCATGGGGTAGGGTGAGAAGGGCAGGGGTGTCCTGATGCAAAGGTGGGGAGAAGGGATCCTAATCCCT
3601 TCCTCTCCCATTCACCCTGTGTAACAGGACCCCAAGGACCTGCCTCCCCGGAAGTGCCTTAACCTAGAGGGTCGGGGAGGAGGTTGTGTCACTGACTCAG
3701 GCTGCTCCTTCTCTAGTTTCCCCTCTCATCTGACCTTAGTTTGCTGCCATCAGTCTAGTGGTTTCGTGGTTTCGTCTATTTATTAAAAAATATTTGAGAA
3801 CAAAAAAAAAAAAAAAAAA
```

FIG. 1—*Continued*

Similarity between Species Homologues of α7

The human sequence was aligned with the sequences of mouse, and rat α7, so regions stringently conserved or altered during evolution could be revealed. The amino acid sequences of mouse and rat α7 are 86 and 81 % similar, respectively, to the putative human homologue (Fig.3A). There are two sequences in human and mouse α7 that differ dramatically from the rat homologue, encompassing residues 50 to 62, and 192 to 224. The first

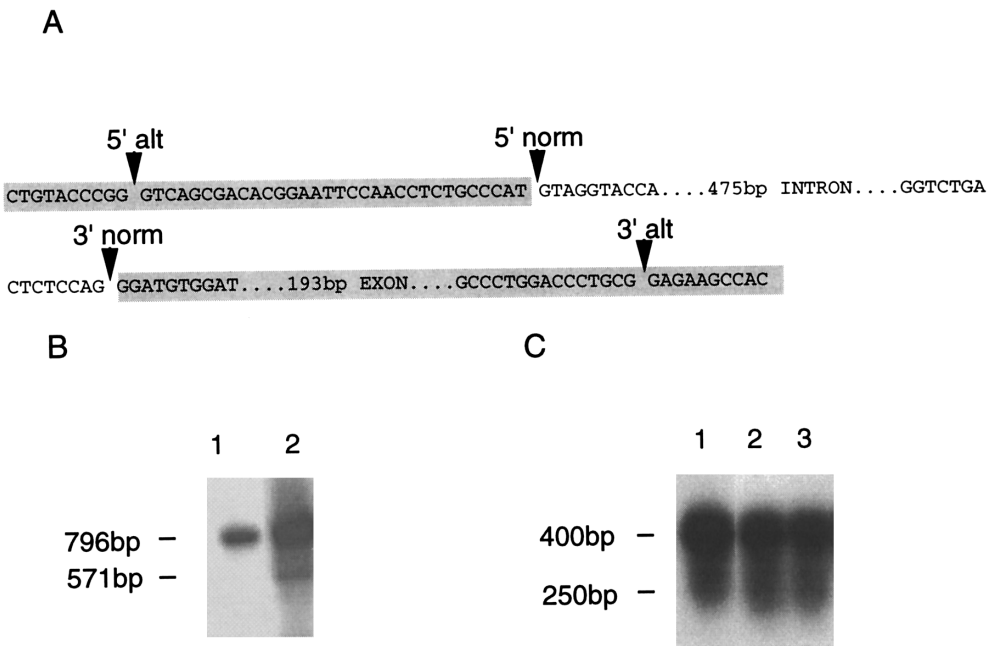

FIG. 2. Detection of the alternative α7 splice variant by PCR amplification of cDNA prepared from human fetal heart, and mouse P19 cells. (A) Alternative splicing at sites flanking an intron in the α7 gene. Exonic sequences are shaded, and vertical arrows denote the splice sites. Both the alternative (alt) and conventional (norm) 5′ splice donor sites agree with consensus splice donor sequences. The conventional 3′ acceptor splice site agrees with consensus splice acceptor sequences, but is not particularly pyrimidine-rich. In contrast the flanking alternative 3′ acceptor site appears cryptic. (B) PCR amplification of the α7 variant from human cDNA. Human small intestine (lane 1), and fetal heart (lane 2) mRNA was reverse-transcribed, and amplified with the α7 primers α7-1820U and α7-91r2 that straddle the deleted extracellular region. The PCR products were separated by agarose gel electrophoresis, transferred to Gene Screen Plus filters, and hybridised with a ^{32}P-labelled cloned PCR product derived by PCR amplification from clone FH7.6 using the primers α7-1820u and α7-91r2. Sizes of the PCR products are indicated in the left-hand margin. (C) PCR amplification of cDNA from mouse P19 cells. Total RNA from P19 cells that had been untreated (lane 1), or induced to differentiate into cardiac and skeletal muscle with DMSO (lane 2), and retinoic acid (lane 3) was reverse transcribed, and amplified with the primers mα7-S, and mα7-R that straddle a region corresponding to that deleted in human α7. PCR products were resolved and hybridised with a human α7 cDNA probe as above.

A

```
                                                  I
human   FNLDVMGALR KEGEPGSLFG FSVALHRQLQ PRPQSWLLVG APQAL.ALPG QQANRTGGLF ACPLSLEETD   69
mouse   FNLDVMGAIR KEGEPGSLFG FSVALHRQLQ PRPQSWLLVG APQAL.ALPG QQANRTGGLF ACPLSLEETD   69
rat     FNLDVMGAIR KEGEPGSLFG FSVALHRQLQ PRPQSWLLVG APQALDSYPD SRQIAHGRPL CLSLSLEETD   70
                                                 II
human   CYRVDIDQGA DMQKESKENQ WLGVSVRSQG PGGKIVTCAH RYEARQRVDQ ILETRDMIGR CFVLSQDLAI  139
mouse   CYRVDIDRGA NVQKESKENQ WLGVSVRSQG AGGKIVTCAH RYESRQRVDQ ALETRDVIGR CFVLSQDLAI  139
rat     CYRVDIDRGA NVQKESKENQ WLGVSVRPRE SGGKVVTCAH RYESRQEVDQ VLETRDVIGR CFVLSQDLAI  140
                                                III
human   RDELDGGEWK FCEGRPQGHE QFGFCQQGTA AAFSPDSHYL LFGAPGTYNW KGLLFVT...  ..NIDSSDPD  204
mouse   RDELDGGEWK FCEGRPQGHE QFGFCQQGTA ATFSPDSHYL VFGAPGTYNW KGLLFVT...  ..NIDSSDPD  204
rat     RDELDGGEWK FCEGRPQGHE QFGFCQQGTA ATFSPDSHYL IFGAPGTYNW KGTARVELCA  QGSSDLAQVD  210
                                                 IV
human   QLVYKTLDPA DRLPGPAGDL ALNSYLGFSI DSGKGLVRAE ELSFVAGAPR ANHKGAVVIL RKDSASRLVP  274
mouse   QLVYKTLDPA DRLTGPAGDL TLNSYLGFSI DSGKGLMRSE ELSFVAGAPR ANHKGAVVIL RKDSATRLIP  274
rat     DGPYEAGGEK DQDPRPS.PV PANSYLGFSI DSGKGLMRSE ELSFVAGAPR ANHKGAVVIL RKDSASRLIP  279
                                                  V
human   EVMLSGERLT SGFGYSLAVA DLNSDGWPDL IVGAPYFFER QEELGGAVYV YLNQGGHWAG ISPLRLCGSP  344
mouse   EVVLSGERLT SGFGYSLAVT DLNNDGWADL IVGAPYFFER QEELGGAVYV YMNQGGHWAD ISPLRICGSP  344
rat     EVVLSGERLT SGFGYSLAVT DLNSDGWADL IVGAPYFFER QEELGGAVYV YMNQGGHWAD ISPLRLCGSP  349
                    VI
human   DSMFGISLAV LGDLNQDGFP DIAVGAPFDG DGKVFIYHGS SLGVVAKPSQ VLEGEAVGIK SFGYSLSGSL  414
mouse   DSMFGISLAV LGDLNQDGFP DIAVGAPFDG DGKVFIYHGS SLGVVVKPSQ VLEGEAVGIK SFGYSLSGGL  414
rat     HSMFGISLAV LGDLNQDGFP DIAVGAPFDG DGKVFIYHGS SLGVVTKPSQ VLEGEAVGIK SFGYSLSGGL  419
                   VII
human   DMDGNQYPDL LVGSLADTAV LFRARPILHV SHEVSIAPRS IDLEQPNCAG GHSVCVDLRV CFSYIAVPSS  484
mouse   DVDGNHYPDL LVGSLADTAA LFRARPVLHV SQEIFIDPRA IDLEQPNCAD GRLVCVDIKI CFSYVAVPSS  484
rat     DVDGNHYPDL LVGSLADTAA LFRARPVLHV SQEIFIDPRA IDLEQPNCAD GRLVCVHVKV CFSYVAVPSS  489

human   YSPTVALDYV LDADTDRRLR GQVPRVTFLS RNLEEPKHQA SGTVWLKHQH DRVCGDAMFQ LQENVKDKLR  554
mouse   YSPSVALDYM LDGDTDRRLR GQVPRVTFLS RGLDDLRHQS SGTVWLKHQH DRVCGDTVFQ LQENVKDKLR  554
rat     YSPIVVLDYV LDGDTDRRLR GQAPRVTFPG RGPDDLKHQS SGTVSLKHQH DRVCGDTCVP AAGKRKDKLR  559

human   AIVVTLSYSL QTPRLRRQAP GQGLPPVAPI LNAHQPSTQR AEIHFLKQGC GEDKICQSNL QLVHARFCTR  624
mouse   AIVVTLSYGL RTPPLRRQAP GQELPTVAPI LNAHQPSTQR TEIHFLKQGC GQDKICQSNL QLERYQFCSR  624
rat     AIVVTLSYGL QTPRLRRQAP DQGLPLVAGI LNAHQPSTQR TEIHFLKQGC GDDKICQSNL QLVQAQFCSR  629

human   VSDTEFQPLP MDVDGTTALF ALSGQPVIGL ELMVTNLPSD PAQPQADGDD AHEAQLLVML PDSLHYSGVR  694
mouse   ISDTEFQALP MDLDGRTALF ALSGQPFIGL ELTVTNLPSD PSRPQADGDD AHEAQLLVTL PASLRYSGVR  694
rat     ISDTEFQALP MDLDG.TALF AHGGQPFIGL ELTVTNLPSD PARPQADGDD AHEAQLLATL PASLRYSGVR  698

human   ALDPAEKPLC LSNENASHVE CELGNPMKRG AQVTFYLILS TSGISIETTE LEVELLLATI SEQELHPVSA  764
mouse   ALDSVEKPLC LSNDSASHVE CELGNPMKRG AQVTFYLILS TSGITIETTE LEVKLLLATI SEQELDPVSV  764
rat     TLDSVEKPLC LSNENASHVE CELGNPMKRG TQVTFYLILS TSGITIETTE LKVELLLATI SEQDLHPVSV  768

human   RARVFIELPL SIAGMAIPQQ LFFSGVVRGE RAMQSERDVG SKVKYEVTVS NQGQSLRTLG SAFLNIMWPH  834
mouse   RAHVFIELPL SISGVATPQQ LFFSGEVKGE SAMRSERELG RKVKYEVTVS NQGQSLNTLG SANLNIMWPH  834
rat     RAHVFIELPL SISGVATPQQ LFFSGKVKGE SAMRSEWDEG SKVKYEVTVS NQGQSLNTLG SAFLNIMWPH  838

human   EIANGKWLLY PMQVELEGGQ GPGQKGLCSP RPNILHLDVD SRDRRRRELE PPEQQEPGER QEPSMSWWPV  904
mouse   EIANGKWLLY PMRVELEGGQ GPGKRGICSP RPNILQLDVD SRDRRRRELG QPEPQEPPEK VEPSTSWWPV  904
rat     EIANGKWLLY PMRVELEGGQ GPEKKGICSP RPNILHLDVD SRDRRRRELG QPEPQEPPEK VEPSTSWWPV  908

human   SSAEKKKNIT LDCARGTANC VVFSCPLYSF DRAAVLHVWG RLWNSTFLEE YSAVKSLEVI VRANITVKSS  974
mouse   SSAE.KRNMT LDCPR.TAKC VVFSCPLYSF DRAAVLHVWG RLWNSTFLEE YMAVKSLEVI VRANITVKSS  972
rat     SSAE.KRNVT LDCP.GTAKC VVFSCPLYSF DRAAVLHVWG RLWNSTFLEE YMSVKSLEVI VRANITVKSS  976

human   IKNLMLRDAS TVIPVMVYLD PMAVVAEGVP WWVILLAVLA GLLVLALLVL LLWKMGFFKR AKHPEATVPQ 1044
mouse   IKNLLLRDAS TVIPVMVYLD PMAVVVEGVP WWVILLGVLA GLLVLALLVL LLWKLGFFKR AKHPEATVPQ 1042
rat     IKNLLLRDAS TVIPVMVYLD PVAVVAEGVP WWVILLAVLA GLLVLALLVL LLWKLGFFKR AKHPEATVPQ 1046

human   YHAVKIPRED RQQFKEEKTG TILRNNWGSP RREGPDAHPI LAADGHPELG PDGHPGPGTA 1104
mouse   YHAVKIPRED RQQFKEEKTG TIQRSNWGNS QWEGSDAHPI LAADWHPELG PDGHPVPATA 1102
rat     YHAVKILRED RQQFKEEKTG TIQRSNWGNS QWEGSDAHPI LAADWHPELG PDGHPVSVTA 1106
```

FIG. 3. Comparison of the deduced amino acid sequence and structure of human $\alpha 7$ with its rodent homologues. (A) The deduced amino acid sequences of human, mouse [10], and rat $\alpha 7$ [7] have been aligned. Identical amino acids are in bold. The seven conserved repeat domains are overlined and numbered, and the transmembrane domain is overlined. The cytoplasmic domains compared for all the species homologues are the B variants. (B) Schematic diagram of the structural features of the human, mouse, and rat $\alpha 7$ subunits. The alternatively spliced region between repeat domains III and IV is denoted by diagonal shading. The transmembrane domain (TM) is in solid boxing. Numeration of amino acid positions starts with the N-terminal amino acid of the mature polypeptide, and for the mouse homologue is with respect to the $\times 2$ variant.

may have been created by a sequencing error in the rat homologue placing part of the sequence out-of-frame, whereas the second appears authentic but has no notable features. A schematic comparison of the rodent splice variants with the human $\alpha 7$ subunit is shown in Fig. 3B. As mentioned previously, only the X2 region be-

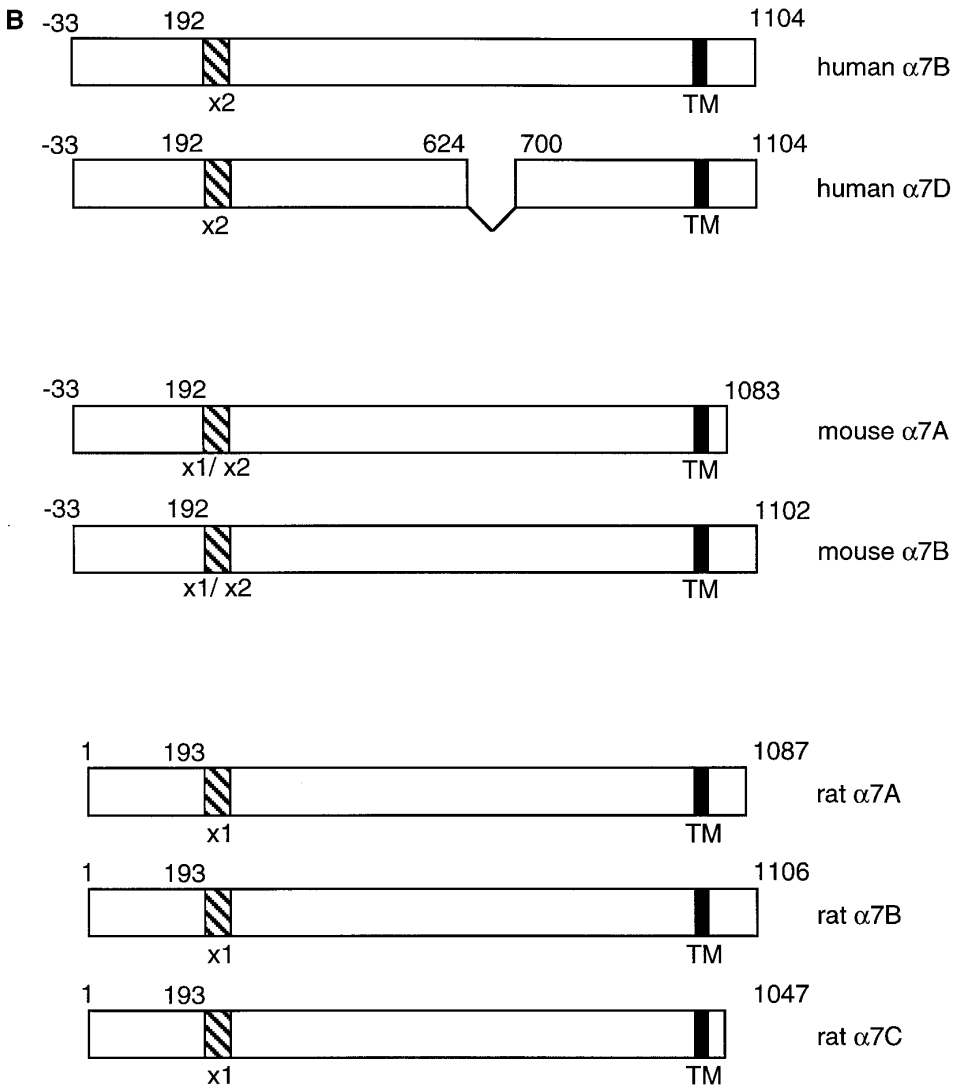

FIG. 3—*Continued*

tween the homology repeat domains III and IV was identified in the human $\alpha 7$ cDNA clones, whereas alternatively spliced X1 and X2 domains have been identified in mouse. Only the B form of the cytoplasmic domain was identified whereas rat has forms A-C.

Tissue Distribution

The distribution of $\alpha 7$ mRNA transcripts was assessed by screening Multiple Human Tissue northern blots (MTN and MTNII, Clontech) (Fig. 4). Blots were probed with the insert of clone HOEA062, revealing a single major 4.4 kb $\alpha 7$ subunit transcript abundantly expressed in skeletal muscle, moderately expressed in heart, small intestine, colon, ovary and prostate, and weakly expressed in lung and testis. Transcripts encoding the deletion variant reported here, as well as other published variants, are predicted to show little size variation, and will not be distinguishable by Northern analysis.

DISCUSSION

We have isolated and characterised cDNAs encoding the complete sequence of the human integrin $\alpha 7$ subunit, and identified a novel extracellular domain variant that has not been previously reported. The variant transcript arises from alternative splicing at sites flanking an intron in the $\alpha 7$ gene, resulting in the deletion of exonic sequence corresponding to 75 amino acids in the C-terminal region of the extracellular domain. Splicing occuring at alternative splice sites within exons is rare, and was first realised in the thalassemia syndromes [23]. The existence of the novel variant, which could be PCR amplified from fetal heart mRNA,

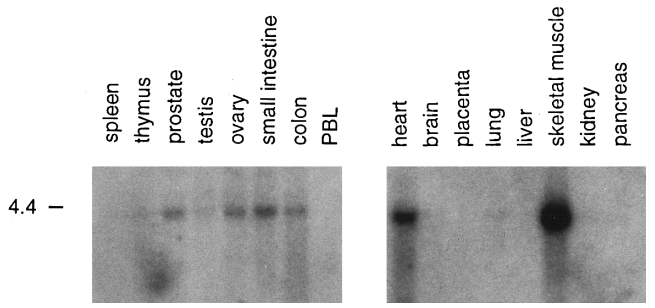

FIG. 4. Analysis of the distribution of $\alpha 7$ subunit transcripts. Multiple human tissue Northern blots (MTN and MTNII, Clontech; 2 μg mRNA per lane) were hybridized with the insert of EST clone HOEA062. The size of the major $\alpha 7$ mRNA transcript calculated with reference to RNA size markers is shown in the left-hand margin.

was not anticipated since only a single 4.4 kb transcript was detected in various tissues by Northern analysis. A difference of 225 bp would not be detectable by this analysis. The data suggest that the splice variant is a minor species, but this notion will need to be confirmed by more accurate methods to quantitate the two species. The variant form was not amplified from small intestine mRNA, suggesting that its expression may be tissue-specific.

A related $\alpha 7$ variant was identified in mouse cells, by PCR amplification of the corresponding region from cDNA prepared from P19 cells. The positions of the alternative splice sites may vary slightly in mouse, since the PCR products representing the two splice variants were closer in size than expected. Neither the overall abundance of $\alpha 7$ transcripts, nor the relative levels of the two species of mRNA were altered when the pluripotential P19 cells were forced to differentiate into cardiac and skeletal muscle. This may suggest that the alternative splice variant is unlikely to be regulated independently of the normal $\alpha 7$ subunit transcript during early mouse development, but this will need to be confirmed by additional studies.

Splicing out of sequence encoding the region encompassing amino acid residues 625 to 699 may have physiological importance, since the major site of ADP-ribosylation of the mouse integrin $\alpha 7$ subunit has been localised to a region encompassing residues 575 to 886 [12]. Hence the deleted region either contains or is in close proximity to the ADP-ribosylation site. The effects of ADP-ribosylation on the functions of $\alpha 7$ integrins remains to be determined. It did not seem to affect the interaction of $\alpha 7$ with its partner subunit $\beta 1$, or the interaction of $\alpha 7\beta 1$ with the cytoskeleton, or with laminin [12]. The αIIb subunit is the only other subunit to be spliced in a C-terminal region of the extracellular domain [24]. In this case a 102 bp exon is either spliced in or out, corresponding to a region only 10 amino acid residues preceding the transmembrane domain.

The overall structure of the human $\alpha 7$ subunit is conserved with its rodent homologues, including three potential extracellular calcium and cytoplasmic actin-binding sites, and serine/threonine kinase and protein tyrosine phosphatase homology regions. Different cytoplasmic variants of the rodent homologues of the $\alpha 7$ subunit have been identified, which presumably possess different signalling properties [6]. Both clones containing 3′-regions sequenced in the present study encoded an identical cytoplasmic domain, similar in sequence to the cytoplasmic domain B variants of the $\alpha 3$ and $\alpha 6$ subunits. There are two different extracellular domain isoforms of mouse $\alpha 7$, designated $\times 1$ and $\times 2$ [10], whereas only the $\times 2$ human variant was identified in the present study.

The results of early reports suggested that the $\alpha 7$ subunit might have a restricted distribution on muscle and certain melanoma cell lines [1, 8, 9], however more recently an RT-PCR study revealed expression of transcripts for the mouse cytoplasmic B variant in every tissue and cell line tested [11]. In the present study most tissues expressed $\alpha 7$ transcripts, with the strongest expression in heart and skeletal muscle. However, transcripts could not be detected in liver, placenta, peripheral blood lymphocytes, or spleen but this may reflect the insensitivity of Northern analysis. Very widespread expression in eighteen different tissue and cell types was revealed by a search of the HGS EST database.

In conclusion, the identification of a novel extracellular splice variant of the $\alpha 7$ subunit continues the trend for members of the $\alpha 3$, $\alpha 6$ and $\alpha 7$ integrin subgroup, whose structures, and potential functions, are altered by alternative splicing of their extracellular domains. The positioning of the major site of ADP-ribosylation either within, or in close proximity to the alternatively spliced region in the extracellular domain of the $\alpha 7$ subunit, suggests these two structural modifications could possibly be functionally related.

ACKNOWLEDGMENTS

This work was supported by grants from the Health Research Council of New Zealand, The Wellcome Trust (UK), the Lottery Grants Board of New Zealand, the Auckland Medical Research Foundation, the Cancer Society of New Zealand Inc., the National Heart Foundation of New Zealand, the Auckland Asthma Society (Mackie Estates), and the Multiple Sclerosis Society of New Zealand. G.W.K. is a recipient of a Wellcome Senior Research Fellowship in Medical Science in New Zealand. The initial EST clone used in this study was discovered as part of a joint collaboration between scientists at The Institute for Genomic Research (Rockville, MD, USA), and at Human Genome Sciences Inc. (Rockville, MD, USA). Drs Leung and Lim participated equally in this study.

REFERENCES

1. Von der Mark, H., Durr, J., Sonnenberg, A., von der Mark, K., Deutzmann, R., and Goodman, S. L. (1991) *J. Biol. Chem.* **266,** 23593–23601.

2. George-Weinstein, M., Foster, R. F., Gerhart, J. V., and Kaufman, S. J. (1993) *Dev. Biol.* **156,** 209–229.
3. Foster, R. F., Thompson, J. M., and Kaufman, S. J. (1987) *Dev. Biol.* **122,** 11–20.
4. Goodman, S. L., Risse, G., and von der Mark, K. (1989) *J. Cell Biol.* **109,** 799–809.
5. Ocalan, M., Goodman, S. L., Kuhl, U., Hauschka, S. D., and von der Mark, K. (1988) *Dev. Biol.* **125,** 158–167.
6. Song, W. K., Wang, W., Sato, H., Bielser, D. A., and Kaufman, S. J. (1993) *J. Cell Sci.* **106,** 1139–1152.
7. Song, W. K., Wang, W., Foster, R. F., Bielser, D. A., and Kaufman, S. J. (1992) *J. Cell Biol.* **117,** 643–657.
8. Song, W. K., Wang, W., Foster, R. F., Bielser, D. A., and Kaufman, S. J. (1992) *J. Cell Biol.* **115,** 843–850.
9. Kramer, R. H., Vu, M. P., Cheng, Y. F., Ramos, D. M., Timpl, R., and Waleh, N. (1991) *Cell Regul.* **2,** 805–817.
10. Ziober, B. L., Vu, M. P., Waleh, N., Crawford, J., Lin, C-S., and Kramer, R. H. (1993) *J. Biol. Chem.* **268,** 26773–26783.
11. Collo, G., Starr, L., and Quaranta, V. (1993) *J. Biol. Chem.* **268,** 19019–19024.
12. Zolkiewska, A., and Moss, J. (1995) *J. Biol. Chem.* **270,** 9227–9233.
13. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
14. McBurney, M. W., Jones-Villeneuve, E. M., Edwards, M. K., and Anderson, P. J. (1982) *Nature* **299,** 165–167.
15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
16. Adams, M. D., Kerlavage, A. R., Flelschman, R. D., Fuldner, R. A. et al. (1995) *Nature* **377,** Suppl., 3–20.
17. Feng, G. S., Ouyang, Y. B., Hu, D. P., Shi, Z. Q., Gentz, R., and Ni, J. (1996) *J. Biol. Chem.* **271,** 12129–12132.
18. Kozak, M. (1993) *Microbiol. Rev.* **47,** 1–4.
19. Kozak, M. (1991) *J. Cell Biol.* **115,** 887–903.
20. Tamura, R. N., Cooper, H. M., Collo, G., and Quaranta, V. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10183–10187.
21. Hogervorst, F., Kuikman, I., Van Kessel, A. G., and Sonnenberg, A. (1991) *Eur. J. Biochem.* **199,** 425–433.
22. Tamura, R. N., Rozza, C., Starr, L., Chambers, J., Reichardt, L. F., Cooper, H. M., and Quaranta, V. (1990) *J. Cell Biol.* **111,** 1593–1604.
23. Orkin, S. (1987) *in* The Molecular Basis of Blood Diseases (Stanatoyannopoulos, G. *et al.*, Eds.), pp. 106–126, Philadelphia, Saunders.
24. Bray, P. F., Leung, C. S.-I., and Shuman, M. A. (1990) *J. Biol. Chem.* **265,** 9587–9590.