Genomics **64,** 155–169 (2000)
doi:10.1006/geno.2000.6122, available online at http://www.idealibrary.com on IDEAL®

Genomic Characterization of the Lysophosphatidic Acid Receptor Gene, *lp*$_{A2}$/*Edg4,* and Identification of a Frameshift Mutation in a Previously Characterized cDNA

James J. A. Contos* and Jerold Chun*,†,[1]

*Neurosciences Graduate Program and †Biomedical Sciences Graduate Program, Department of Pharmacology, School of Medicine, University of California, San Diego, La Jolla, California 92093-0636

Received October 26, 1999; accepted January 6, 2000

To understand the regulation, evolution, and genetics of *lp*$_{A2}$/*Edg4*, a second lysophosphatidic acid receptor gene, we characterized its complete cDNA sequence, genomic structure, and chromosomal location. The full-length mouse transcript sequence was determined using rapid amplification of cDNA ends. Southern blot and restriction fragment length polymorphism segregation analyses revealed that the mouse gene was present as a single copy and located at the middle of Chromosome 8 near the mutations for myodystrophy (*myd*) and "kidney–anemia–testes" (*kat*). This region is syntenic with human chromosome 19p12, where the human genomic clone containing the *lp*$_{A2}$ gene (*EDG4*) was mapped. Sequence analysis of genomic clones demonstrated that both mouse and human transcripts were encoded by three exons, with an intron separating the coding region for transmembrane domain VI. Reverse transcriptase-PCR demonstrated that the three exons were spliced in all mouse tissues shown to express the transcript. Finally, in a comparison of all human *lp*$_{A2}$ sequences present in the database, we identified several sequence variants in multiple tumors. One such variant (a G deletion) in the initially characterized Edg4 cDNA clone (derived from an ovarian tumor) results in a frameshift mutation near the 3′ end of the coding region. In addition to increasing our understanding of the mechanisms underlying lysophosphatidic acid signaling and lysophospholipid receptor gene evolution, these results have important implications regarding the genomic targeting and oncogenic potential of *lp*$_{A2}$.

INTRODUCTION

Lysophospholipids such as lysophosphatidic acid (LPA) and sphingosine-1-phosphate (S1P) are extracellular signaling molecules produced by various cell types, including activated platelets (reviewed in Durieux, 1995; Moolenaar *et al.*, 1997). LPA and S1P activate specific G-protein-coupled receptors (GPCRs), thereby causing a wide array of cellular effects, such as increased proliferation, Rho-dependent morphological changes, and increased ion conductances in membranes (e.g., van Corven *et al.*, 1989; Ridley and Hall, 1992; Fernhout *et al.*, 1992). The first identified lysophospholipid receptor gene, *lp*$_{A1}$ (also called *vzg-1/Edg2/mrec1.3/Gpcr26*), specifically interacts with LPA (Hecht *et al.*, 1996; Fukushima *et al.*, 1998; Chun *et al.*, 1999). Upon initial searches of GenBank, two groups of similar receptor sequences were identified: those encoded by the *Edg1, H218/Agr16,* and *Edg3* orphan receptor genes, with 32–36% amino acid sequence identity to *lp*$_{A1}$, and those encoded by the *Cnr1* (CB1) and *Cnr2* (CB2) cannabinoid receptor genes, with approximately 28% amino acid identity to *lp*$_{A1}$ (Contos and Chun, 1998) The endogenous cannabinoid receptor ligands (anandamide and 2-arachidonylglycerol) are lipid molecules with structures similar to those of lysophospholipids (Felder *et al.*, 1993; Stella *et al.*, 1997). We and others determined that the *Edg1, H218/Agr16,* and *Edg3* genes encode high-affinity receptors for S1P (An *et al.*, 1997; Lee *et al.*, 1998; Zondag *et al.*, 1998; Zhang *et al.*, 1999). We thus refer to our characterized mouse clones for these genes as *lp*$_{B1}$, *lp*$_{B2}$, and *lp*$_{B3}$, respectively (Zhang *et al.*, 1999; Contos and Chun, 1998).

In the course of genomic characterization of the mouse *lp*$_{A1}$ gene, another GenBank search revealed a novel human genomic sequence that encoded a receptor with 55% identity to *lp*$_{A1}$ (Contos and Chun, 1998). Such high identity, of magnitude similar to that within the *lp*$_{B}$ subfamily or within other GPCR subfamilies that bind the same ligand, suggested that this receptor was a high-affinity LPA receptor and should be clustered in the same subfamily as *lp*$_{A1}$. We thus termed this gene *lp*$_{A2}$ (Contos and Chun, 1998). Concurrently with our studies, it was shown that overexpression of

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF218844.

[1] To whom correspondence should be addressed at Department of Pharmacology, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0636. Telephone: (619) 534-2659. Fax: (619) 822-0041. E-mail: jchun@ucsd.edu.

0888-7543/00 $35.00

the human lp_{A2} gene (called *EDG4*)[2] could potentiate LPA responsivity, but not S1P responsivity, on a cell line, supporting the hypothesis that this was a second LPA receptor gene (An *et al.*, 1998).

A characterization was undertaken to understand several aspects of the lp_{A2} gene more clearly: (1) its complete genomic structure, (2) *cis*-elements controlling its expression, (3) its chromosomal location, and (4) its evolutionary relationship to other *lp* receptor genes. Furthermore, mouse lp_{A2} genomic characterization is requisite for the generation of mice with a targeted deletion of the gene, which would allow analysis of lp_{A2} biological function in the whole animal. Previous characterization of the mouse lp_{A1} gene provided a basis for the analysis of lp_{A2} (Contos and Chun, 1998). The coding region of lp_{A1} is spread over two or three exons, depending on which of the multiple 5′ exons is found in the transcript. One intron is located within the coding region for transmembrane domain VI, an insertion site not found in other GPCR gene subfamilies, including the lp_B and *Cnr* genes, which are intronless in their coding region (Liu and Hla, 1997; Abood *et al.*, 1997; Zhang *et al.*, 1999). While the entire lp_{A1} gene is located on mouse Chromosome 4, the last exon is duplicated on Chromosome 6 in some mouse strains. With this background in mind, we isolated and characterized mouse lp_{A2} genomic clones. We found that although lp_{A2} has a conserved intron similar to that in lp_{A1}, there is only one primary exon. The two coding exons are present as single copies in *Mus musculus* and are located at the central part of mouse Chromosome 8. In addition, we determined that the previously characterized Edg4 cDNA clone (An *et al.*, 1998) has a guanine nucleotide deletion that causes a frameshift near its C-terminal coding region. This likely reflects a somatic mutation in the ovary tumor cells from which the cDNA was isolated and may have altered the function of the encoded receptor.

MATERIALS AND METHODS

5′ and 3′ rapid amplification of cDNA ends (RACE). For both 5′ and 3′ RACE, we utilized Marathon cDNA templates from different mouse tissues (Clontech). After a primary PCR, products were diluted 1:20, and 1 μL was used as template in a secondary PCR with nested primers on either end. PCRs of 50 μL consisted of 1× PCR buffer B [75 mM Tris, pH 9.0, 15 mM $(NH_4)_2SO_4$],[3] 0.25 mM each dNTP, 0.5 mM each primer, 1 μL of template, and 0.5 U *Pfu* (added after the reaction mix was overlaid with mineral oil and heated to 90°C). Reaction mixes were cycled 35× (95°C for 30 s, 60°C for 30 s, and 72°C for 3 min). Of the many lp_{A2} gene primer combinations tried, the following combinations gave clear products that could be subcloned: 5′ RACE, edg6t′/AP1 and then edg6e/AP2; 3′ RACE, edg6a/AP1 and then edg6o/AP2[4] (oligonucleotide sequences are listed in Table 1). The 5′ RACE products were subcloned into pBluescript SK(+) (Stratagene) using two enzymes whose restriction sites were at the ends of the final PCR product: *Sac*I (in edg6e) and *Xma*I (by AP2). Colonies were grown with blue/white selection and white colonies screened using PCR (edg6e/T7). 3′ RACE products were subcloned using blunt/sticky (product digested with *Not*I and 4-cutters listed below; pBS digested with *Eco*RV/*Not*I) or blunt/blunt ligations (product digested separately with *Alu*I, *Hae*III, *Bst*UI, or *Rsa*I; pBS digested with *Eco*RV). This "shotgun" technique allowed rapid sequencing of the 1.8-kb 3′ RACE product.

TABLE 1

Oligonucleotide Sequences

edg6a	5′-CGAGACCATCGGTTTCTTCTATA-3′
edg6a′	5′-CGAGACCATCGGCTTTTTCTATA-3′
edg6b	5′-CCCAGGATGATGACAACAGTCTT-3′
edg6b′	5′-CCCAGAATGATGACAACCGTCTT-3′
edg6c	5′-CTCATGTTCCACACTGGTCC-3′
edg6d	5′-AAGCCCTGCCGCAGGAACCA-3′
edg6e	5′-CTGAGCTCCAAGCCGCTGTT-3′
edg6f	5′-GACCACACTCAGCCTAGTCA-3′
edg6h	5′-TCCTCAGCCCTGGTTGGTTT-3′
edg6i4	5′-GGCCAGCCTGGTCTACCAA-3′
edg6i5	5′-GCAGAGGCAGGCAGAACTTTT-3′
edg6Te15′	5′-GCCATGGGTCTCAGTCCTGCTTCAA-3′
edg6m′	5′-GAAGAGACCTGAGAGTTTG-3′
edg6o	5′-TTTGTGGTGTGCTGGACACCG-3′
edg6t′	5′-TGTGCAGGTAGCAACCCCAGA-3′
edg6v	5′-GTGACTTGGACAGTTGCTCACGCAT-3′
edg6s	5′-AGCCCTGATCTTCCTATTCC-3′
edg6x	5′-GCTGGGCATGGGATTTCA-3′
edg6e3a	5′-AGTCTAGAGGCTCTGCAAGTGACCT-3′
edg6e3b	5′-CATCTCAGGTTTTAGGGTTT-3′
edg6e3c	5′-ACATCTAGAGATAACACAGTAA-3′
edg6KO2	5′-AGACTTCGGGAAGCAAGGTAGT-3′
edg6P1	5′-CCACTCGTGCCGCACTACCTT-3′
edg6P2	5′-GTTAAAGACGCTGCTCTTACTG-3′
edg6P4	5′-TAGTGCCACACACCTCTACTTG-3′
edg6P5′	5′-TCTTGCACATTTGTCTTTGTGCT-3′
AP1	5′-CCATCCTAATACGACTCACTATAGGGC-3′
AP2	5′-ACTCACTATAGGGCTCGAGCGGC-3′
β-actin a	5′-ACAGCTTCTTTGCAGCTCC-3′
β-actin b	5′-GGATCTTCATGAGGTAGTCTGTC-3′
lpA2e2mh1	5′-CCTACCTCTTCCTCATGTTC-3′
lpA2e3mh1	5′-TAAAGGGTGGAGTCCATCAG-3′

Genomic Southern blot analysis. Genomic DNA (20 μg) from a *M. musculus* mixed background strain (C57BL/6J × Balb/C) was 10-fold overdigested with the restriction enzymes indicated in Fig. 2. Conditions for making and probing the Southern blots are detailed elsewhere (Contos and Chun, 1998), with the exceptions listed below. Probe fragments were purified with the Qiaquick gel extraction kit (Qiagen) after PCR amplification from genomic clone templates. Labeled probes (edg6a/edg6b and edg6e3b/edg6m′) were hybridized at a concentration of 1.5×10^6 dpm/mL instead of 1.0×10^6 dpm/mL. After being air-dried, the blot was exposed at −80°C to Kodak MS film in a cassette containing two regular intensifying screens and the special HE (high-energy) screen (Kodak). The more sensitive film and intensifying screen allowed exposure times to be reduced approximately 5-fold (Southen blots shown in Fig. 2 were exposed for 40 h).

[2] The approved symbol for the lp_{A2}/*Edg4* gene is *Edg4* for mouse and *EDG4* for human.

[3] All chemical reagents used were purchased from Sigma, with the exceptions of radionucleotides (Dupont), random hexamers/T3 polymerase/T7 polymerase/RNasin (Boehringer Mannheim), sequencing reagents (USB), and others where noted. All restriction and modifying enzymes used were purchased from from New England Biolabs, with the exception of *Taq* and *Superscript* (Gibco), *Pfu* polymerase (Stratagene), and *Sequenase* (Amersham).

[4] As expected, using edg6f/AP2 as the secondary reaction for 3′ RACE yielded a robust product only ~30 bp smaller than 6o/AP2. However, this reaction also yielded a product of equal intensity that was approximately 200 bp smaller (this smaller product was not characterized). Considering that edg6f and edg6o lie adjacent to one another in the cDNA clone but are in exons 2 and 3, respectively, this may indicate a distinct third exon.

TABLE 2

Subcloned Fragments and Plasmid Names

Donor DNA	Fragment subcloned	Name	Relative location of subclone
clone 1F	6.0-kb *Xho*I/*Not*I	XN5.0	Promoter, exons 1–3
clone 1F	6.0-kb *Xho*I/*Sal*I	XS5.0	Promoter, exons 1–3
clone 3F	12-kb *Xho*I/*Not*I	XN12	Promoter, exons 1–3, 3′ sequence
PCR product	edg6a/6b	pBSedg6a/6b	Exon 2
XN5.0	4.0-kb *Apa*I/*Sal*I	AS4.0	Exon 2–exon 3
XN5.0	1.0-kb *Apa*I/*Apa*I	AA0.65	Intron 1, exon 2
XN5.0	2.4-kb *Xho*I/*Nhe*I	XNhe2.4	Promoter, exons 1–2
XN5.0	0.73-kb *Sac*I/*Sac*I	SS0.73	Intron 1

Genomic clone isolation and restriction mapping. A PCR strategy (detailed in Contos and Chun, 1998) was used to isolate λ clones from a 129/SvJ genomic library in Lambda FIX II (Stratagene). PCR with the initially designed primers (edg6a/6b) gave a specific product of the expected size (~714 bp) using mouse 129/SvJ genomic DNA as template. This product was confirmed to be lp_{A2} by T/A subcloning (see below) and sequencing. Using the edg6a/6b PCR, we screened a total of approximately 460,000 clones (46 wells × 10,000 clones/well) and isolated three positives after tertiary screens by conventional plaque hybridization.[5] The three pure phage clones were grown by liquid lysate, and DNA was isolated with the Wizard λ Prep Kit (Promega). All three clones were characterized by restriction mapping, making use of sites that flank each side of the insert (*Xba*I, *Sac*I, *Sal*I, and *Not*I). Since no *Not*I or *Sal*I sites were found in any of the inserts, we double-digested with one of these enzymes (*Not*I) to map *Eco*RI, *Hind*III, *Xho*I, or *Bam*HI sites (in addition to single digestions with each). Digested clones were electrophoresed, Southern blotted (as above), and then probed with the edg6a/6b probe. PCRs using the primers edg6a, edg6b, edg6c, edg6d, edg6e, or edg6h with T7 or T3 (both T3 and T7 flank the inserts) allowed determination of orientations and more precise relative locations of the exon amplified by edg6a/6b. Restriction maps were constructed utilizing all these data, as well as the determined sequence.

Subcloning and sequencing. Inserts from the purified λ clone DNA were subcloned for ease of larger preparations, finer restriction mapping, and the obtainment of templates that could be sequenced manually. Table 2 lists the fragments subcloned and names of the plasmids containing them. In addition to λ genomic clones, other subcloned fragments (from PCR and RACE) are listed. Subcloning of genomic fragments and 5′ RACE products was usually performed using an in-gel ligation protocol, described elsewhere (Contos and Chun, 1998). Digested 3′ RACE products were subcloned with the same protocol, except inserts were simply heat-treated, ethanol precipitated, and resuspended (in place of the insert being gel-purified). The edg6a/edg6b PCR product (amplified with regular *Taq*, which tends to add an A at the 3′ ends) was T/A subcloned by adding a T onto the ends of *Eco*RV-digested pBS (Marchuk *et al.*, 1991). Cloning vector (pBS) was almost always treated with calf intestinal alkaline phosphatase briefly after restriction digestion. Subcloned inserts were sequenced using the dideoxy method and T3, T7, or reverse primers that flank the multiple cloning site or with oligonucleotides designed to insert sequence (Sanger *et al.*, 1977).

DNA sequence analysis. Raw sequence data were read into files and assembled into contigs using the DNasis software program (Hitachi). The human genomic cosmid and Edg4 cDNA sequences were downloaded from GenBank at the NCBI Web site (http://www.ncbi.nlm.nih.gov). Searches (with parts of these sequences) were then performed using the BLAST algorithm (Altschul *et al.*, 1990) in both the nonredundant (nr) and the EST (dbEST) databases, also at the NCBI Web site. Alignments of all ESTs were made with DNasis. However, where substantial parts of the sequences became very divergent from the consensus, they were separated and aligned independently. Repetitive elements (Smit, 1996) were determined using RepeatMasker (A. F. A. Smit and P. Green; http://ftp.genome.washington.edu/RM/RepeatMasker.html). All sequences determined here were deposited with GenBank. Human sequence variations not noted in Fig. 8 included EST AA298791 (CTG CTT GT**T**), which differed from the genomic and Edg4 cDNA clone sequences (CTG CTT GT**C**), but did not change the encoded amino acid. In additon, EST AA312795 (AAT G**T**T GCT) differed from three other sequences (AAT G**C**T GCT), though the amino acid change was conservative (valine to alanine). The one mouse sequence variation found (in 3′UTR) was beween AA118482 (TCA**T**AGT) and the three following sequences: our genomic clone, our 3′ RACE product, and EST AI550109 (TCA**C**AGT).

Reverse transcriptase-PCR (RT-PCR). To generate cDNA template, total RNA from various cell lines or mouse Balb/C tissues was isolated using either a guanidinium isothiocyanate (Ausubel *et al.*, 1994) or a Trizol reagent protocol (Gibco) and quantitated spectrophotometrically. A 40-μL reaction consisting of 1× Superscript first-strand synthesis buffer, 20 U RNasin, 0.5 mM each dNTP, 10 μg RNA, 2 μL (200 pmol) diluted random hexamers, and 200 U Superscript was incubated at 23°C for 10 min, at 42°C for 60 min, at 95°C for 5 min, cooled on ice, diluted to 100 μL with H_2O, and stored at −20°C. PCR was performed under reaction conditions similar to those used in the RACE experiments, except that the annealing temperature was 56°C, *Taq* was used instead of *Pfu*, and 1 μL of diluted cDNA was used as template in 20-μL reactions. Genomic DNA (gDNA; 100 ng) was used as a control template. The number of cycles for all PCRs shown was adjusted so product was not maximized, compared to quantities obtained with excess product as template (35 cycles for the lp_{A2} reactions and 23 cycles for the β-actin reactions). Primer combinations (sequences listed in Table 2), relative locations, and expected product sizes shown in Fig. 7 were:

edg6Te15′/edg6t′	exons 1 and 2	569 bp
lpA2e2mh1/lpA2e3mh1[6]	exons 2 and 3	798 bp
actin-a/actin-b	β-actin exons	630 bp

The β-actin primers were designed to mouse sequence corresponding to the rat sequences previously used (Raff *et al.*, 1997) and do not amplify any product from pseudogenes in gDNA (a common problem with a large proportion of normalization studies performed). Additional primer combinations utilized to confirm and/or check expression were as follows:

edg6f/edg6e3c	exons 2 and 3	517 bp
edg6Te15′/edg6i5	exon 1 and intron 2	1252 bp spliced; 2089 bp unspliced
edg6Te15′/edg6i4	exon 1 and intron 2	1627 bp spliced; 2464 bp unspliced

[5] It is likely that there were at least double this number of clones present in the initial population. However, because adequate numbers of clones were being isolated on secondary and tertiary screens, all initial positives were not pursued to this level of purity.

[6] The lpA2e2mh1 and lpA2e3mh1 primers are complementary to human sequence as well as mouse sequence. Thus these primers can be used in RT-PCR from both species.

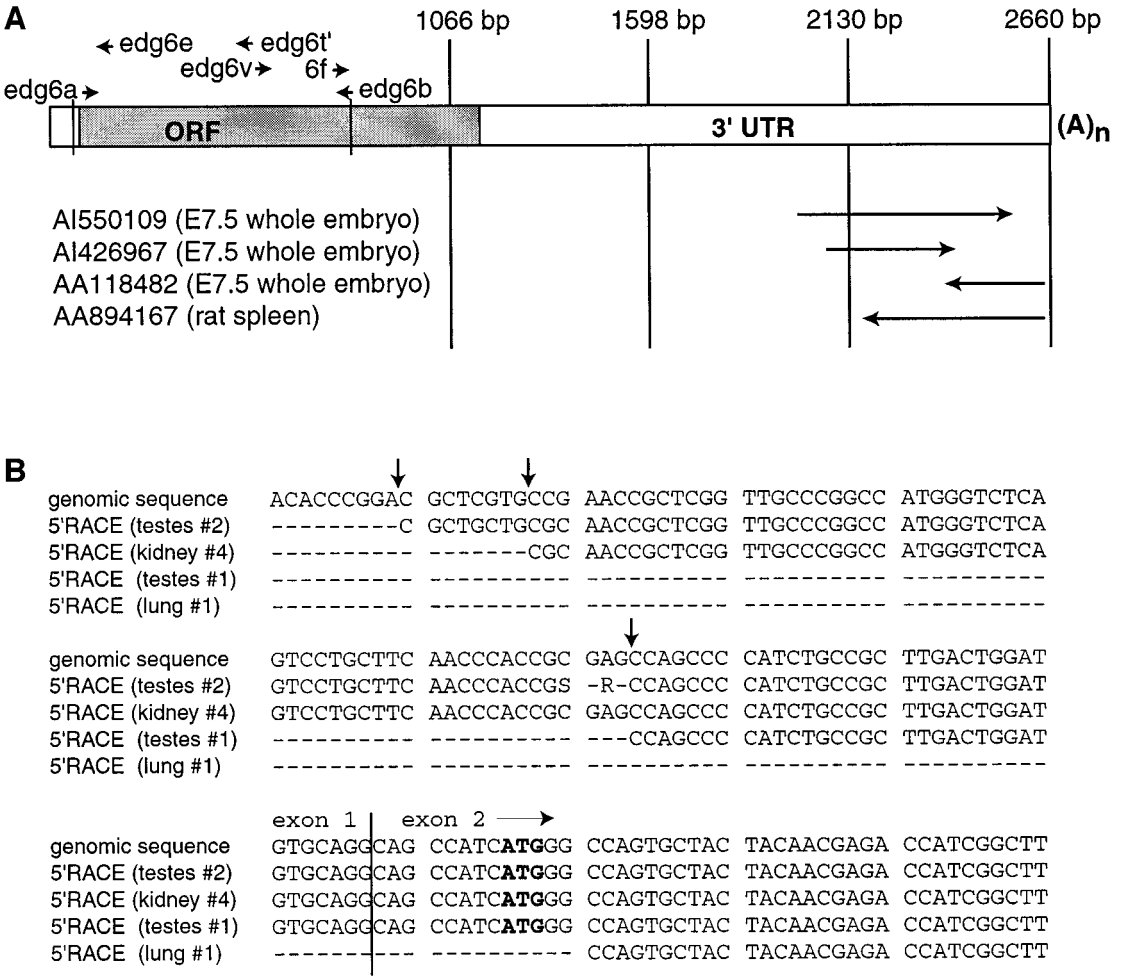

FIG. 1. Mouse lp_{A2} transcript sequence analysis. (**A**) Schematic of mouse mRNA sequence, relative locations of primers used for 5′ RACE (edg6t′/edg6e), 3′ RACE (edg6v/edg6f), or PCR screening of genomic clones (edg6a/edg6b), and relative locations of all rodent lp_{A2} EST clones. Shaded area indicates open reading frame (ORF), while the white area indicates the untranslated region (UTR). Two lines are drawn at the exon boundaries. Rodent EST clones are referred to by their GenBank accession number and the tissue from which they were isolated (in parentheses). (**B**) Aligned 5′ RACE products from various tissues. With the exception of two ambiguities in the testis 2 sample and three bases at the 5′ end, all products aligned exactly with one another and with genomic sequence (determined later).

Restriction fragment length polymorphism detection and chromosomal mapping. To find restriction fragment length polymorphisms (RFLPs), primers were used to amplify products from both *M. musculus* (C57BL6/JEi) and *M. spretus* (SPRET/Ei) genomic DNA (purchased from The Jackson Laboratory). These two strains are referred to as B and S samples, respectively. We found that the primer combination edg6e/6KO2 yielded robust product from B but not S template, while edg6t′/6KO2 yielded robust product from S but not B template. Expected product sizes of approximately the same intensity from both species were obtained with edg6f/6m′, edg6a/6b, and edg6s/edg6x. By analyzing fragment sizes produced after independent digestion of these products with 4–14 restriction enzymes, we found the following RFLPs: edg6f/6m′ (*Dpn*II, *Hin*fI, *Nla*III, *Pst*I) and edg6s/edg6x (*Alu*I, *Pst*I, *Hin*fI). The *Hin*fI RFLP in the edg6s/edg6x PCR product was selected to screen 94 individuals from each backcross panel. PCR conditions were the same as outlined above for RACE, except the final volume was 20 μL, the annealing temperature was 54°C, *Taq* was used instead of *Pfu*, and 1 μL of diluted gDNA (50 ng) was used as template. Reaction products were digested by adding 10 μL of a mixture consisting of 7.5 μL H_2O, 2 μL of 10× NEB2 (New England Biolabs restriction digest buffer 2), and 0.5 μL (5 U) *Hin*fI. Tubes were incubated at 37°C for 2 h, 6× loading buffer was added, and 20 μL was electrophoresed on a 1.4% agarose gel containing ethidium bromide (Ausubel *et al.*, 1994). The formal names of the crosses are The Jackson Laboratory interspecific backcross panels (C57BL/6J × *M. spretus*)F_1 × C57BL/6J, called Jackson BSB, and (C57BL/6JEi × SPRET/Ei)F_1 × SPRET/Ei, known as Jackson BSS (Rowe *et al.*, 1994). Raw data were submitted to The Jackson Laboratory for comparison to other markers typed to the panel. Raw data can be obtained from the Internet address http://www.jax.org/resources/documents/cmdata. Genes mapping to the same chromosomal area from other backcross screenings were determined from the Mouse Genome Database: http://www.informatics.jax.org/menus/map_menu.shtml.

RESULTS

Determination of the Full-Length Mouse lp_{A2} cDNA Sequence

We designed oligonucleotides to the previously deposited human lp_{A2} genomic sequence (cosmid 33799, GenBank Accession No. AC002306) in regions of an exon that encoded amino acid residues conserved with other members of the *lp* receptor family (Fig. 1A). PCR conditions were optimized for amplification of a single product of the correct size from mouse 129/SvJ genomic DNA. This product was confirmed to be from the lp_{A2} gene by subcloning and sequencing. Additional mouse-

specific primers were then designed for use in both 5′ and 3′ RACE experiments. Because Northern blots indicated that the lp_{A2} transcript was most abundant in testis, kidney, and lung (J. J. A. Contos and J. Chun, unpublished observation), we used cDNA from these tissues as templates in the amplifications. Unlike the lp_{A1} gene, which has multiple, divergent 5′ sequences, we found that for lp_{A2}, all four subcloned 5′ RACE products from each tissue had the same sequence, although starting at slightly different points (Fig. 1B). In 3′ RACE experiments we obtained one product, which corresponded to three mouse expressed sequence tag (EST) clones deposited with GenBank, all from E7.5 whole embryos (Fig. 1A). In addition to likely primer sequence at the end of one EST, there was only one other sequence variation between the cDNA and the ESTs (noted under Materials and Methods). Polyadenylation sites were identical in both the 3′ RACE products and the one EST clone that contained a poly(A) tail (AI426967). In addition, a rat EST clone (AA894167) appeared to add a poly(A) tail at this point. The total transcript length was 2647 bp, which combined with a 250-nt poly(A) tail (Wahle, 1995) corresponds to the approximate size of the major hybridizing RNA species observed by Northern blot (J. J. A. Contos and J. Chun, unpublished observation). An open reading frame (ORF) of 1044 bp is present in the cDNA and encodes a putative protein with 347 amino acids and 55% identity to mouse LP_{A1}. The AUG start codon (at nt 96) for this ORF lies within a Kozak consensus sequence (Fig. 1B). Unlike lp_{A1}, there are no mRNA destabilization AUUUA consensus sequences in the lp_{A2} transcript.

The lp_{A2} Gene is Present as a Single Copy in M. musculus

To determine whether the mouse lp_{A2} gene was present as a single copy in the genome and whether it was divided into multiple exons (similar to the corresponding human gene), we utilized Southern blotting. *M. musculus* C57BL/6J × Balb/C genomic DNA was digested with 16 separate restriction enzymes, electrophoresed, blotted, and probed with two regions from the mouse cDNA sequence (determined by RACE) corresponding to the separate human exons. Only a single fragment hybridized with the 5′ probe (edg6a/6b) for each restriction digest (Fig. 2A). Using a 3′ probe (edg6e3b/6m′), we also observed a single fragment hybridizing to gDNA digested with most of the restriction enzymes (Fig. 2B). Of the three lanes where two fragment sizes were observed (*Eco*RI, *Bst*XI, and *Xba*I), two (*Eco*RI and *Bst*XI) could be explained by a restriction site located within the probe. The two fragments in the *Xba*I lane may be due to an RFLP between the C57BL/6J strain and the Balb/C strain. These results indicated that the lp_{A2} gene was present as a single copy in both C57BL/6J and Balb/C strains of *M. musculus*. In addition to determining the copy number of the gene, the sizes of the restiction fragments suggested that a small intron (<4 kb) was located between the corresponding probe fragment regions. While most of the restriction enzymes resulted in the same fragment size hybridizing with the two probes, it was different for five of them (*Eco*RI, *Pst*I, *Sac*I, *Bst*XI, and *Apa*I). The *Apa*I difference is likely due to a restriction site located near the 3′ end of the edg6a/6b probe. However, because none of the other four restriction sites is located in the cDNA sequence between the two probes, we concluded that there must be intervening sequence between the probe regions in the gDNA. Overall, the Southern blot results indicated that the lp_{A2} gene is present as a single copy in *M. musculus* and that multiple exons encode the transcript sequence.

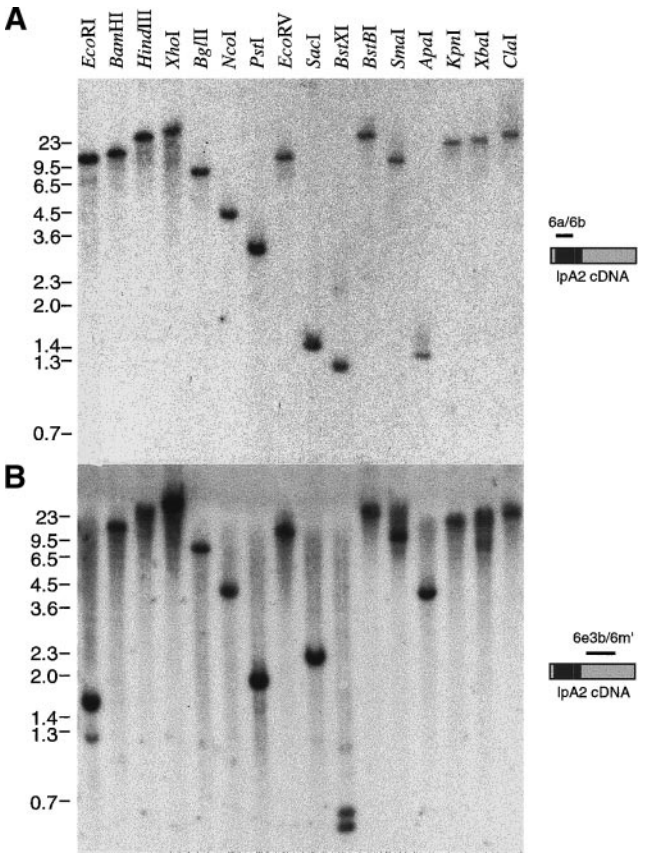

FIG. 2. Southern blot analysis of mouse lp_{A2}. C57BL/6J × Balb/C genomic DNA (20 μg per lane) was digested with the indicated restriction enzymes, electrophoresed, blotted, and hybridized with lp_{A2} probes from either the 5′ part of the ORF (**A**) or the 3′UTR (**B**). Relative locations of the probes are indicated to the right in the cDNA schematic (black indicates ORF). With the exception of *Eco*RI, *Pst*I, *Sac*I, *Bst*XI, and *Apa*I, both probes hybridized to a single fragment of the same size. In addition, two fragments were observed only with the 6e3b/edg6m′ probe in three lanes (*Eco*RI, *Bst*XI, *Xba*I).

The Mouse lp_{A2} Gene Maps to Central Chromosome 8

To map the mouse lp_{A2} gene, we screened for RFLPs between *M. musculus* C57BL/6J and *M. spretus*. PCR products were amplified from genomic DNA derived from each of these mouse species, digested with several restriction enzymes, and checked for RFLPs by agarose gel electrophoresis. A clear *Hin*fI RFLP was detected in a region of intron 2 easily amplifiable with the

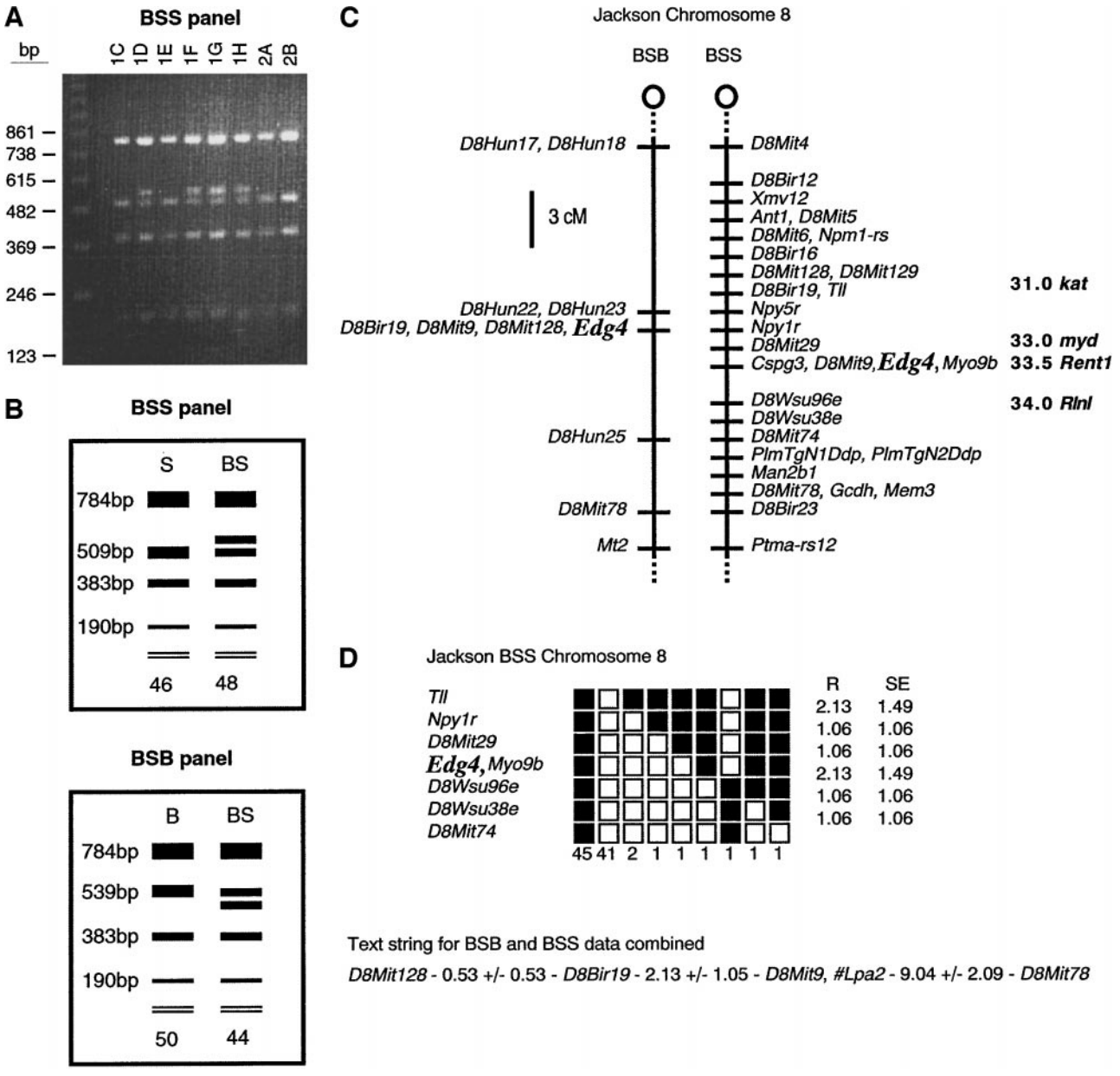

FIG. 3. RFLP screening strategy and linkage map placing the lp_{A2} gene at the central part of mouse Chromosome 8. (**A**) Clear detection of RFLP segregation in the BSS backcross panel from The Jackson Laboratory. An ethidium bromide-stained agarose gel through which *Hin*fI-digested edg6s/edg6x PCR products were electrophoresed (from samples 1C–2B in the panel) is shown. (**B**) Schematic of observed results from screening 94 samples from the BSS and BSB backcross panels. The number of each observed genotype is indicated below the lane. B, *M. musculus* C57BL/6J genotype; S, *M. spretus* genotype. (**C**) Linkage maps for BSB and BSS backcross panels. A 3-cM scale bar is shown to the left. Loci mapping to the same position are listed in alphabetical order. In addition, several genes localized to this area from other segregation analyses are indicated to the right (in boldface type) at levels corresponding to their relative mapped positions (cM positions from MGD shown). *Cspg3*, chondroitan sulfate proteoglycan 3; *kat*, kidney anemia testes; *myd*, myodystrophy; *Myo9b*, myosin IXb; *Npy1r/5r*, neuropeptide Y receptors; *Rent1*, regulator of nonsense transcripts 1; *Rlnl*, relaxin-like factor. (**D**) Haplotype from The Jackson Laboratory BSS backcross showing part of Chromosome 8 with loci linked to lp_{A2}. Loci are listed in order with the most proximal at the top. The black boxes represent the C57BL/6JEi allele, and the white boxes represent the SPRET/Ei allele. The number of animals with each haplotype is given at the bottom of each column of boxes. The percentage recombination (*R*) between adjacent loci is given to the right, with the standard error (SE) for each *R*.

primers edg6s/edg6x. This was used to screen two F_2 backcross panels (188 animals total) obtained from The Jackson Laboratory (Fig. 3A). In the panels, we observed close to a 1:1 ratio for segregation of this allele, consistent with normal Mendelian principles (Figs. 3B and 3D). This contrasts with an RFLP from exon 4 of the lp_{A1} gene, where significantly different ratios and PCR product intensities were observed, due to a duplication on Chromosome 6 in *M. spretus* (Contos and Chun, 1998). Based on the segregation pattern in this backcross panel, we placed the lp_{A2} gene at the central part of Chromosome 8 by the myosin IXb (*Myo9b*) and chondroitan sulfate proteoglycan 3 (*Cspg3*) genes (Figs. 3C and 3D). Other backcross panel mapping studies have localized the regulator of nonsense transcripts 1 (*Rent1*) gene, as well as mutations for myo-

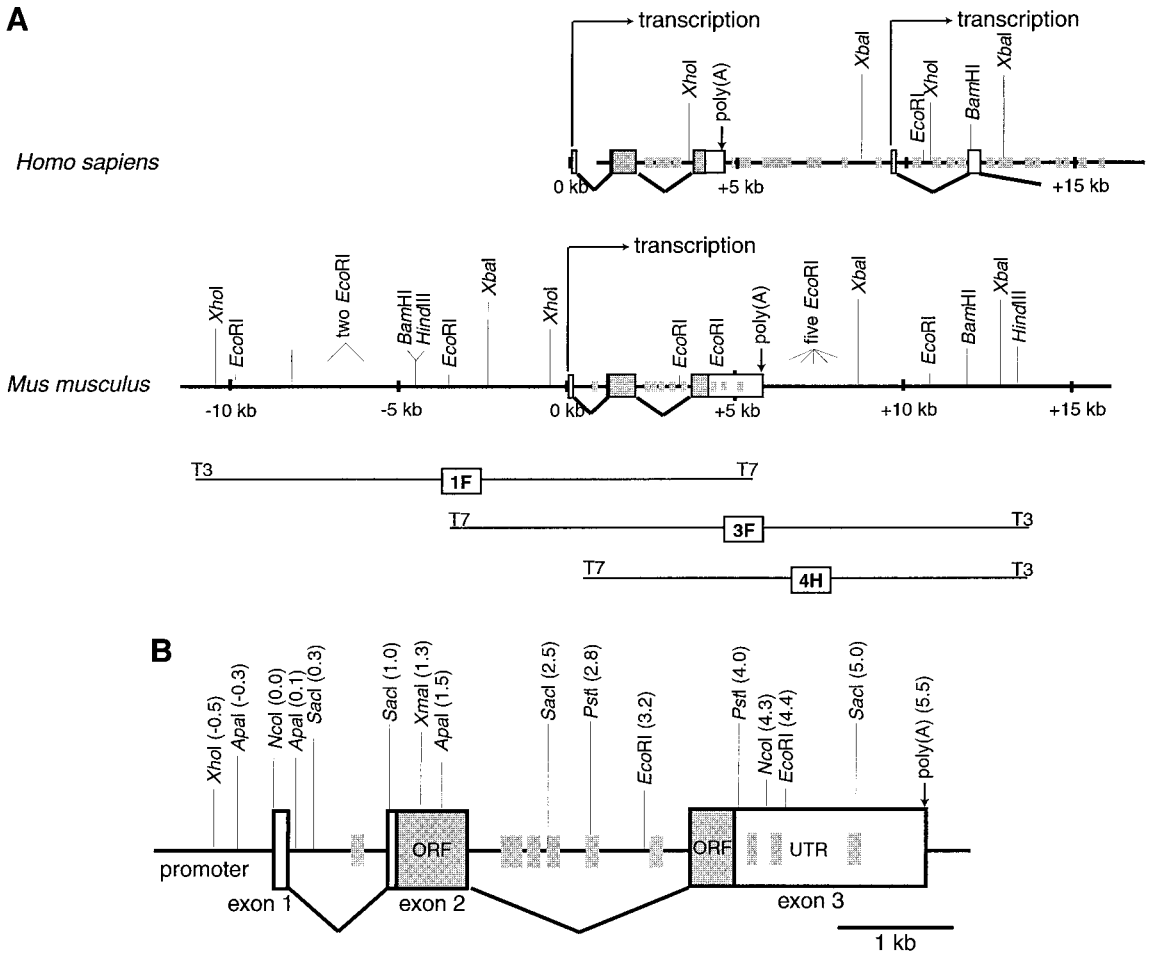

FIG. 4. Genomic organization and restriction maps for human and mouse lp_{A2} genes. (**A**) Large-scale human and mouse genomic restriction maps. Below the genomic maps are shown the relative locations of the three mouse λ genomic clones isolated. In both species, the lp_{A2} gene is divided into three exons, shown as boxes connected by triangular lines below (which indicate splicing). Additional exons found as ESTs in the database are also shown as boxes. Lighter shaded smaller boxes indicate repetitive elements. Such additional exons and repetitive elements are not shown in the mouse map beyond the sequenced region (**B**). The human genomic sequence ends 500 bp upstream of exon 2; thus the placement of the human exon 1 is presumptory. (**B**) Higher resolution map of the sequenced mouse region. Polyadenylation sites are indicated at the end of the lp_{A2} gene. Several restriction enzyme sites predicted from the Southern blot to be within intron 2 are shown (*Eco*RI, *Pst*I, *Sac*I). The 3′UTR contains three repetitive elements.

dystrophy (*myd*) and kidney–anemia–testes (*kat*), very close to this locus as well (Fig. 3C).

The Mouse and Human lp_{A2} Transcript Sequences Are Encoded by Three Exons

To determine the precise organization of the mouse lp_{A2} gene, we isolated and characterized clones from a 129/SvJ genomic DNA library. The edg6a/6b PCR was used to screen wells containing pools of 10,000 clones each. Of six initial positives, we eventually isolated three separate clones after tertiary screens. Relative locations of exons in each clone were determined with PCRs (which utilized the T7 and T3 promoter sequences flanking each end of the insert) and Southern blotting (Fig. 4A). Together with the gDNA Southern blot data presented above and sequence data discussed below, we constructed a restriction map of ~25 kb of genomic sequence that included the lp_{A2} gene (Fig. 4A). A 6.0-kb restriction fragment (XN5.0) that hybridized to both edg6a/6b and edg6e3b/6m′ was subcloned into pBluescript and sequenced entirely in both directions. A more detailed restriction map of this fragment is presented in Fig. 4B, and most of its sequence is presented in Fig. 5. By comparing the cDNA sequence to the genomic clone sequence, we determined that the mouse lp_{A2} gene was divided into three exons, which we termed 1 (97 bp), 2 (742 bp), and 3 (1808 bp). The intron separating exons 1 and 2 was 836 bp, while the intron separating exons 2 and 3 was 2002 bp. Intron boundaries corresponded to consensus donor and acceptor sequences (Fig. 6A). All coding sequence was located within exons 2 and 3, which is similar to the mrec splice-variant form of the lp_{A1} gene (Contos and Chun, 1998). Also as with lp_{A1}, the second intron separates the coding region for transmembrane domain VI and belongs to class II (i.e., inserted after the first basepair of a codon). The human lp_{A2} gene structure is also presented from analysis of a cosmid clone sequence that contains exons 2 and 3 (Fig. 4A). Both the mouse and the human genes contain exactly the same

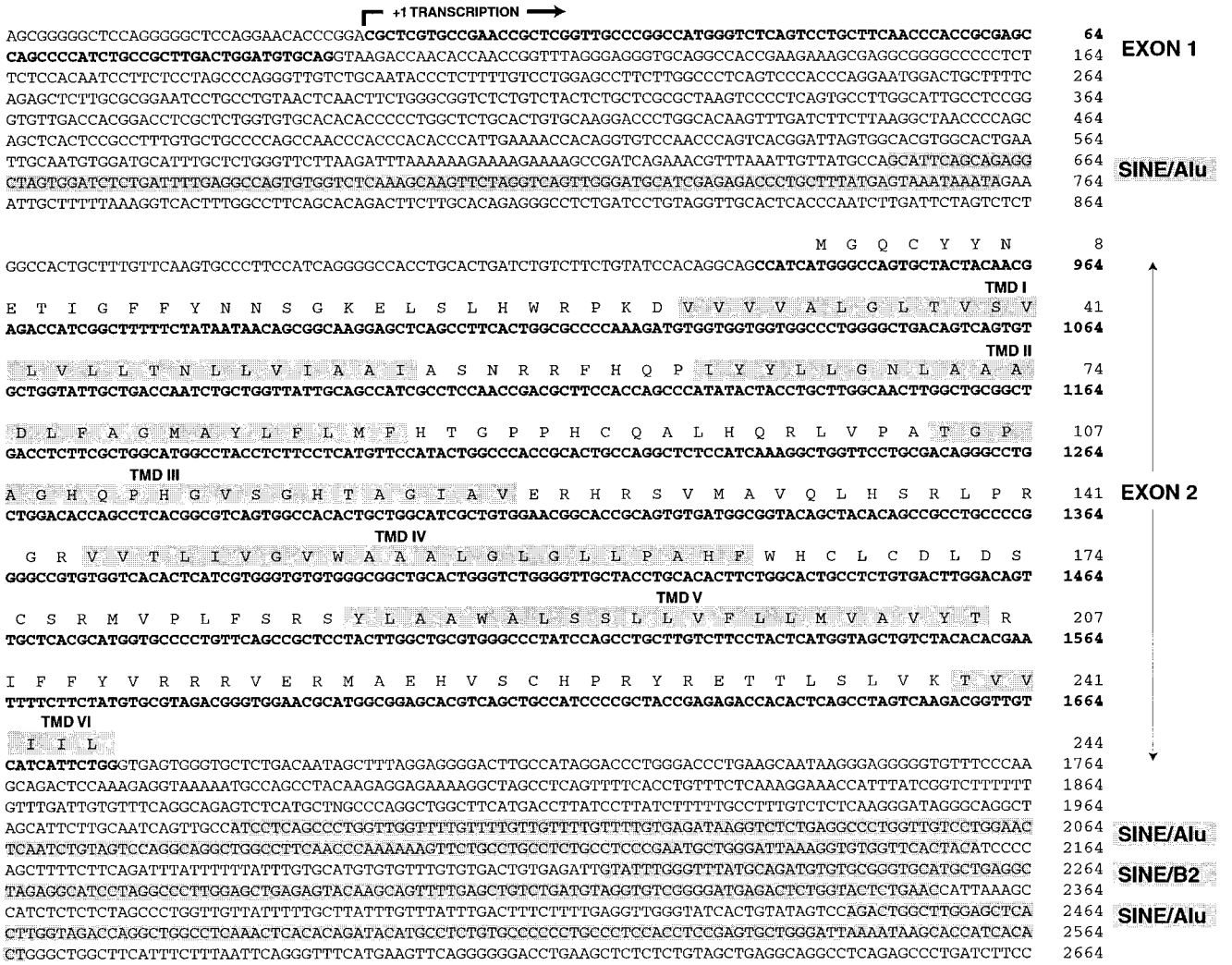

FIG. 5. Sequence of the mouse lp_{A2} gene. A total of 5588 bp of 129/SvJ genomic sequence is shown, with +1 being the most 5′ start site as determined by 5′ RACE. The lp_{A2} gene exons are in boldface type and indicated to the right of the sequence, with the ORF translated above. Putative transmembrane domains in the translation product are shaded. In addition, the polyadenylation consensus sequence is boxed, and repetitive elements are shaded and indicated to the right.

size of exon 2 and a very similarly sized intron 2 (1973 bp in human). However, polyadenylation occurs in the human clone approximately 1 kb upstream of the analogous position in mice, which explains the shorter human transcript size observed by Northern blot (An *et al.*, 1998). The mouse polyadenylation sequence, AUUAAA, like the AGUAAA in human, differs slightly from the stronger consensus AAUAAA (Fig. 6B), although both have been found to be used in other genes (Birnstiel *et al.*, 1985). Some GU clusters are found in the appropriate location relative to the polyadenylation consensus sequence (Figs. 6B and 8C). If these signal sequences are not utilized, transcription might proceed to produce the larger transcripts (~7 kb in mouse and ~8 kb in human) observed in some tissues by Northern blot (in addition to the more prevalent smaller transcript).

Analysis of lp_{A2} Exon Splicing Patterns by RT-PCR

To determine whether the proposed splicing indicated in Fig. 4 occurs in most tissues that express the lp_{A2} gene, we used RT-PCR to detect spliced transcripts. Figure 7 shows that in all mouse tissues examined that splice exons 2 and 3 together, exons 1 and 2 are also spliced together. Furthermore, in this semiquantitative assay, the relative levels of each of these PCR products were similar between tissues, demonstrating that the RNA transcript consisting of spliced exons 1, 2, and 3 is the predominant form in most cell types. Together with the 5′ RACE results, this suggests that alternate primary exons are not utilized in transcription of the mouse lp_{A2} gene, which contrasts with the the lp_{A1} gene, where multiple primary exons are found (Contos and Chun, 1998).

There is a larger mouse transcript observed by Northern blot (of approximately 7 kb), which is most visible in the embryonic brain samples, but also detectable in other adult tissues expressing large amounts of the smaller 2.8-kb transcript (J. J. A. Contos and J. Chun, unpublished observation). To determine whether this larger transcript contains intron 2 sequence, we used primers from exon 1 and intron 2 in

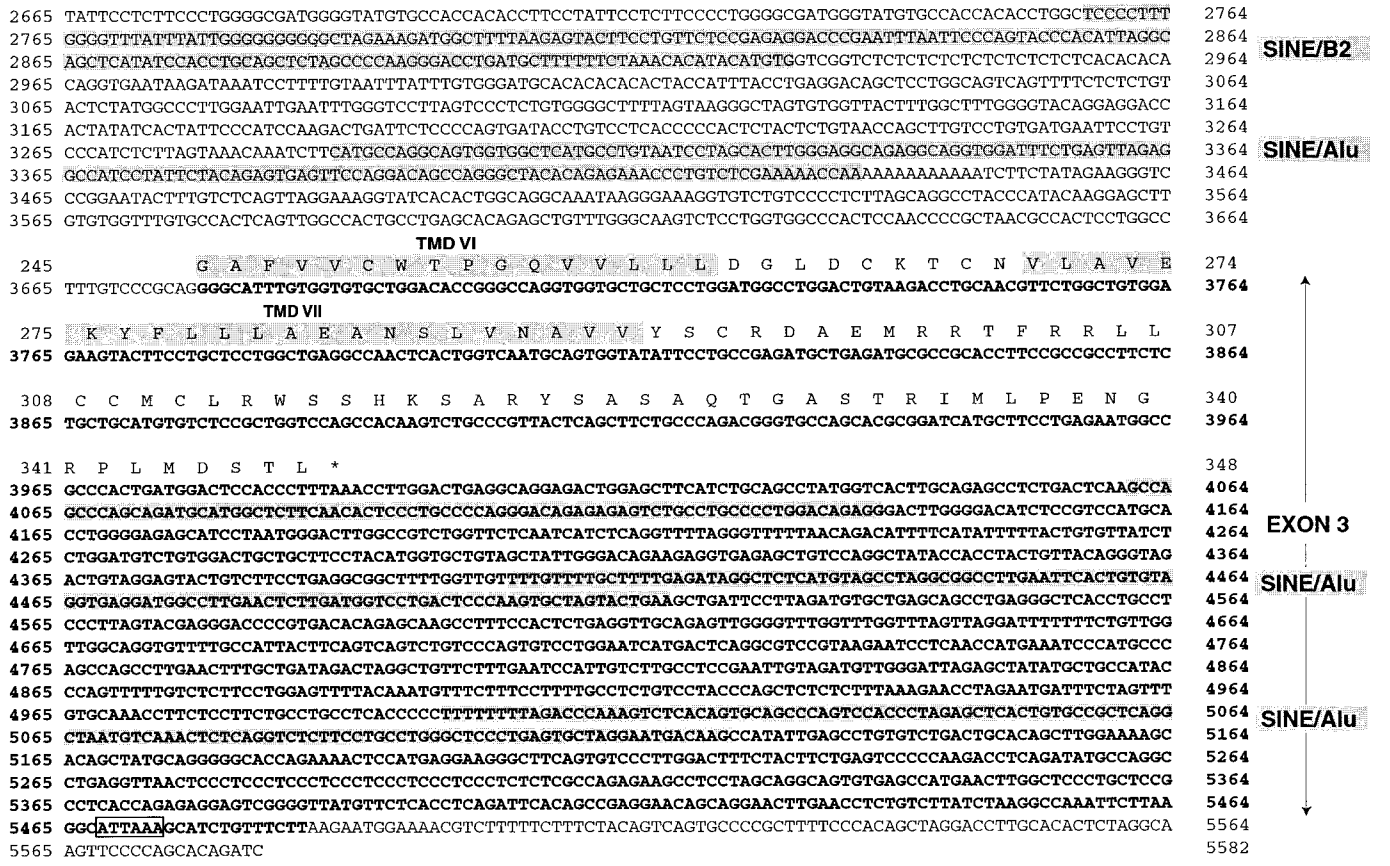

FIG. 5—*Continued*

RT-PCRs on several of these samples. We failed to detect any product from cDNA, though these primer combinations readily amplified a larger product from gDNA (data not shown). Thus, the larger transcript size is not due to a lack of splicing exons 2 and 3.

Analysis of Human lp_{A2} cDNA and EST Sequences

To complete our analysis of the human lp_{A2} gene, we searched GenBank for all current information on human lp_{A2} genomic and cDNA sequences. In addition to the genomic cosmid clone from chromosome 19p12 (deposited as part of the chromosome 19 sequencing project), there were a total of 20 ESTs, as well as the cDNA sequence "Edg4" (An *et al.*, 1998). Slightly over half (11/20) of the ESTs were from cancer cells. Alignment of all sequences demonstrated that most of the ESTs were in the 3′UTR, although several were found in the coding region (Fig. 8A). One clone from embryonic brain (T02954) had a 5′ sequence that was completely divergent from all other cDNA sequences. Aligning this sequence with the genomic clone indicated that the divergent sequence was almost exactly identical to the adjacent intron 2 sequence. This may represent reverse transcription of the hnRNA transcript before splicing. The two nearly identical prostate ESTs were best aligned by introducing a large gap approximately halfway into them, both at exactly the same point. Since these represented two separate cDNA clones from the same library (and presumably the same individual), a deletion in the tissue of origin is likely.

Smaller variations among all deposited sequences were also determined (Fig. 8B). One notable guanine nucleotide deletion was found in the Edg4 cDNA near the end of the coding region, which was not present in an EST clone or the genomic sequence from the same area. This deletion results in a frameshift mutation and a predicted protein containing 35 extra amino acids (382 instead of 347). We confirmed that there was a G deletion in the Edg4 clone (Fig. 9C) that was not observed in other clones (Figs. 9A and 9B). Two other variations in the coding region were found, but one did not change the encoded amino acid, and the other was a conservative change (valine to alanine; see Materials and Methods). A total of 37 variations were observed in the 3′UTR. Of these, 16 were present in multiple clones, which suggests true allelic diversity rather than sequencing errors. Of the differences observed in multiple clones, nearly all (15/16) were observed in the cancerous cells, and 11 of them were observed in these cell types exclusively. Clones containing poly(A) tails were aligned, demonstrating that polyadenylation primarily occurs 15 nt downstream of the AGUAAA (9/15 clones), although it was also found 16, 17, 25, and 43 nt downstream (Figs. 6B and 8C).

FIG. 6. Alignment of intron donor/acceptor and polyadenylation sequences with consensus sequences. (**A**) Mammalian consensus intron donor/acceptor sequence aligned with human and mouse lp_{A2} exon boundaries. The nearly invariant AG and GT of all such sequences are shown in boldface type, while additional residues that align with the consensus are shaded. Boxes represent sequence present in the spliced mRNA transcript. *MM, Mus musculus; HS, Homo sapiens.* (**B**) Mammalian consensus polyadenylation sequence aligned with the mouse (*MM*), rat (*RN*), and human (*HS*) lp_{A2} polyadenylation regions. Consensus and determined polyadenylation sites are indicated with arrows, and GT clusters are shown in boldface type. Sequence beyond the poly(A) site in rat was not determined.

DISCUSSION

The genomic characterization of the lp_{A2} gene reported here is necessary to understand its evolutionary origins, its relationship to other genes that have been mapped, and the relevance of sequence variations. Conserved locations of introns 1 and 2 with the lp_{A1} gene suggest that a common ancestral gene, distinct from the gene that gave rise to three lp_B genes, duplicated and diverged into the two lp_A genes. Sequence variations and frameshift mutations/deletions among multiple deposited and characterized clones indicate that previously obtained functional data need to be confirmed with unmutated clones and suggest that there may be a link between cancer and the lp_{A2} gene. Finally, by mapping the lp_{A2} gene to central mouse Chromosome 8 by several other genes and mutations, we have provided additional relevant information for understanding regulation of this genomic region and the relationship of lp_{A2} to previously mapped mutations.

The finding of a guanine nucleotide deletion in the ORF of the Edg4 cDNA relative to two other human clones (genomic and EST) is significant because this deletion changes the last few C-terminal amino acids and the predicted length of the translation product. We have sequenced our own human cDNA clone, which does not contain the deletion (data not shown). The mouse cDNA and genomic sequences predict a 347-amino-acid translation product, the same size as the human protein predicted from the presence of a G in the human cDNA. In addition, the sequence immediately surrounding the deleted G is identical in mouse clones and the human clones without the deletion. Several possibilities could explain the deletion. First, it could simply have been reported due to a sequencing error. However, this possibility has been ruled out through our manual sequencing, as well as additional sequencing by the original author (Songzhu An, pers. comm. 1999). Second, RNA editing could theoretically explain the difference, though no mechanisms have been described that insert a G into vertebrate mRNA transcripts. Third, the deletion may have arisen in that particular cDNA clone as an artifact during the cloning process. Sequencing additional lp_{A2} clones from the same library could eliminate this possibility. Fourth, the deleted G may represent an allelic variation in humans that is present in a proportion of the population. However, this is unlikely because the nucleotides and encoded amino acids surrounding the deletion are strictly conserved across mammals, suggesting that an essential function is imparted to the encoded protein. A final possibility is that the deletion was due to a somatic mutation in the ovary tumor cells from which the transcript was derived. This hypothesis could be confirmed by isolating and sequencing additional lp_{A2} clones from the same library that might also contain the deletion.

It is conceivable that the G deletion in lp_{A2} may have contributed to transformation of the original ovary cells that formed a tumor. First, the C-terminal regions of GPCRs contain multiple phosphorylation sites that are required for proper receptor desensitization after activation (Freedman and Lefkowitz, 1996). Such desensitization occurs through binding of arrestin proteins to the phosphorylated receptors. Should these events be disrupted, perhaps through steric blocking of

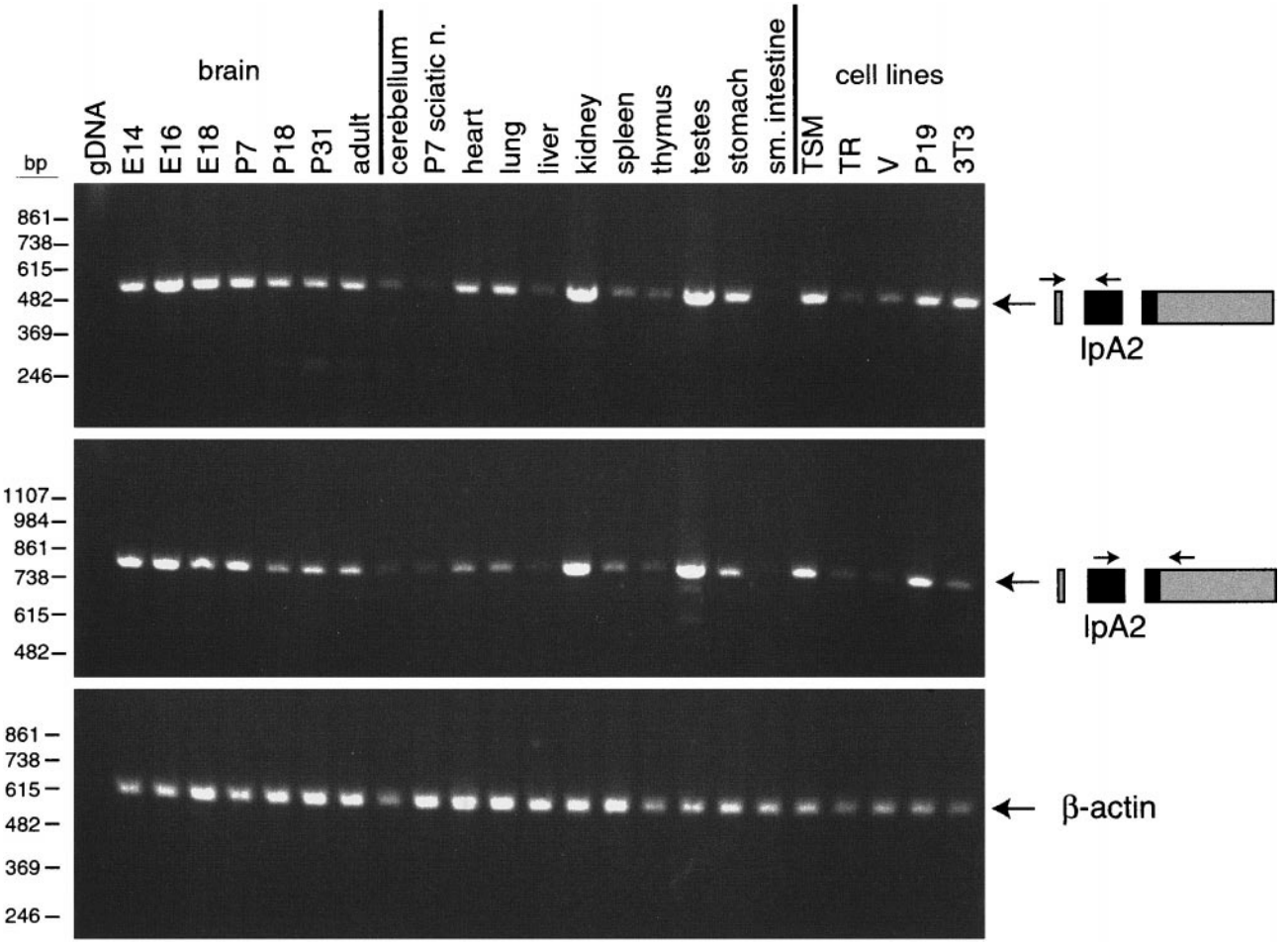

FIG. 7. RT-PCR detection of lp_{A2} gene splicing in mouse. Relative locations of primers in the lp_{A2} gene are indicated to the right (boxes indicate exons, with shaded regions representing the coding region). Above each lane is the template cDNA used, including whole brain at various developmental stages (E, embryonic day; P, postnatal day), adult cerebellum, P7 sciatic nerve, nine additional adult organs, and five mouse cell lines. As a control, the first lane shows product from genomic DNA (gDNA) template. The β-actin PCR demonstrates relative levels of cDNA template used in each sample.

arrestin binding, signaling from activated LP_{A2} receptors might be extended, leading to misregulation of target effects (e.g., proliferation). Future experiments will determine whether functional differences exist between the variant protein forms, which would allow a reevaluation of the initial functional experiments resulting from expression of the Edg4 transcript in Jurkat T cells (An *et al.*, 1998). A second line of evidence is suggestive of a role for the lp_{A2} gene in cancer. LPA has been shown to increase the invasiveness of tumor cells as well as activate both ovarian and breast cancer cells (Imamura *et al.*, 1993; Xu *et al.*, 1995; Stam *et al.*, 1998). In addition, high plasma LPA levels were almost invariably found in ovarian cancer patients, but not in any controls (Xu *et al.*, 1998). These results suggest that activation of LPA signaling pathways in ovary-derived cells can lead to cancer formation, which might occur with gain-of-function mutations.

In addition to the guanine nucleotide deletion in the coding region, other clone sequences had variations in the 3′UTR, especially near the poly(A) site. The large number of sequence variations in the 3′UTR probably reflects the fact that there were many more clones to align here. However, the fact that nearly half of the variations were found in multiple clones suggests that many of the variations represent true allelic diversity or somatic mutations, rather than cloning or sequencing errors. It has been found that 3′UTRs affect stability of transcripts, and when 3′UTRs are mutated or overexpressed, several consequences can result, including increased proliferation. For example, rearrangement in a 3′UTR of a cyclin G gene family member (CCND1/BCL1/PRAD1) leads to an over sixfold increase in the half-life of the transcript and is found in 10% of certain types of lymphomas and leukemias (Rimokh *et al.*, 1994). Furthermore, transfection of a particular 3′UTR can complement an immortalization-defective mutant of SV40 T antigen in immortalization of rat embryonic fibroblasts (Powell *et al.*, 1998). Perhaps mutations or sequence variations in the lp_{A2} transcript 3′UTR lead to loss of normal regulation of lp_{A2} gene expression. Because lp_{A2} encodes a growth factor receptor, such misregulation of expression might affect proliferation of cells.

Sequence variations near the poly(A) site may also have a role in allowing a larger transcript form (8 kb)

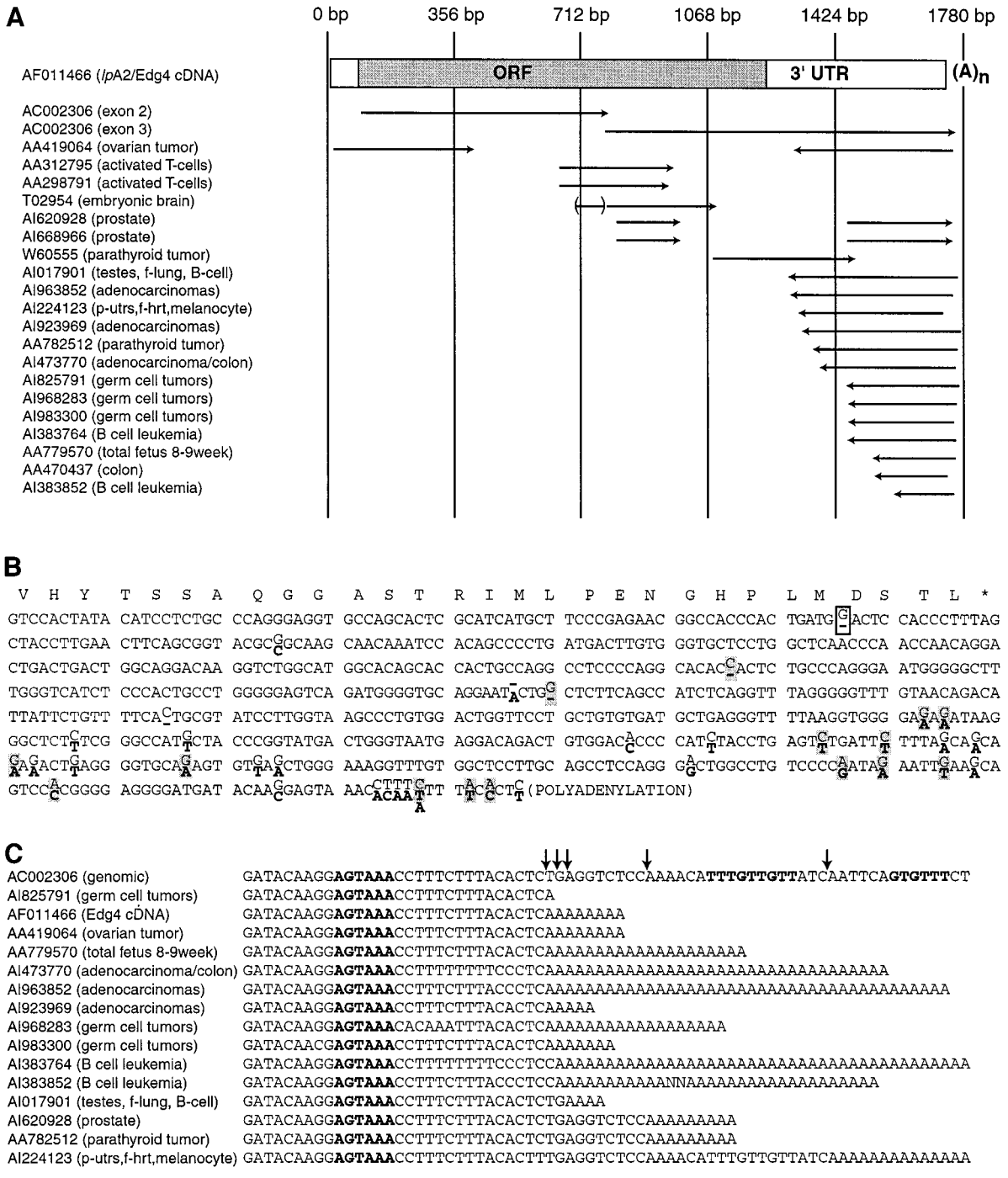

FIG. 8. Identification of mutations in human lp_{A2} cDNA clones. (**A**) All EST clones in dbEST are shown aligned with the Edg4 cDNA sequence and with the predicted exons from a deposited genomic cosmid clone at chromosome 19p12. The parentheses around EST T02954 (embryonic brain) indicate sequence that is identical to intron just 5′ of exon 3. Each of the two clones from prostate aligns in two separate areas of the cDNA, separated by ~450 bp. (**B**) Sequence variation between all deposited ESTs, cDNAs, and the genomic clone. The last part of the coding region (with predicted amino acids indicated above) along with all the 3′UTR is shown. Sequences in other parts of the cDNA were identical in all clones analyzed, except T02954 (embryonic brain), which had many out-of-frame errors, likely due to seqencing errors. Where variability exists, the most common sequence is shown above and the anomalies below (in boldface type). A dash indicates a deletion. Shaded areas indicate variations present in multiple clones. The deposited Edg4 cDNA sequence (AF011466) has a G deletion in the fourth-to-last codon (boxed). (**C**) Polyadenylation sites in ESTs and cDNAs. All clones containing poly(A) tails are shown aligned with the genomic DNA sequence. Polyadenylation consensus sequences are shown in boldface type. Arrows above the genomic DNA sequence indicate all observed polyadenylation sites. The first site is the most common (5/11 clones), followed (in order) by more 3′ sites. Note two additional sequence variations (past the most common poly(A) site) in the aligned clones.

to be generated in humans. This larger transcript is abundant in peripheral blood leukocytes and is the predominant form in most cancerous cell lines that express the gene (An *et al.*, 1998). One possibility for generation of the larger transcript is a lack of polyadenylation after the AGUAAA site, which would result

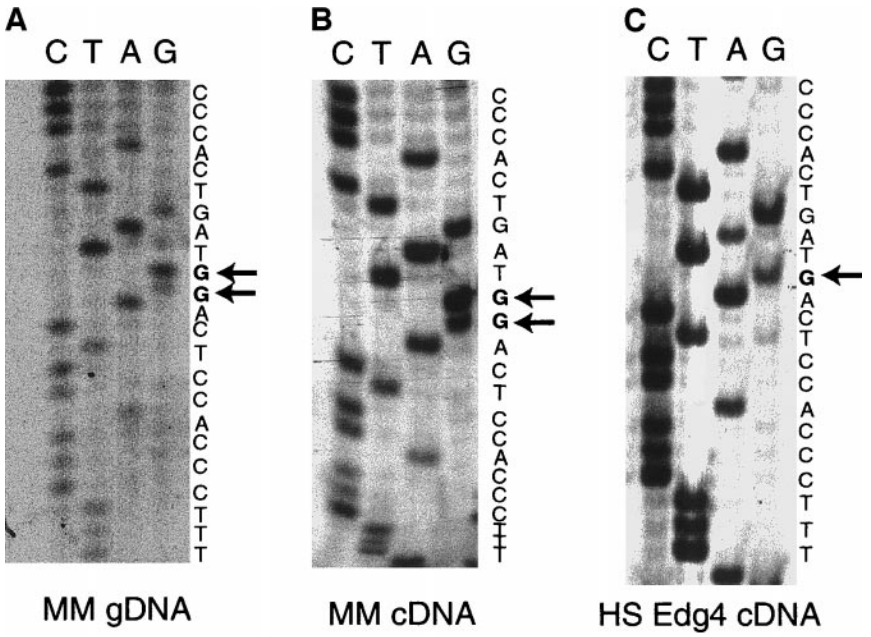

FIG. 9. Confirmation of deletion mutation in the human Edg4 cDNA clone. Sequences are shown for mouse genomic DNA (**A**), mouse cDNA (**B**), and the human Edg4 cDNA (**C**). Each sequencing reaction utilized a primer located in the 3′UTR. The complement of the actual dideoxy reactions is shown for ease of comparison to the data in Figs. 5 and 8. Arrows point to the difference between the sequences: two Gs in mouse gDNA and cDNA (which is also the sequence in human genomic and EST clones) and only one G in the Edg4 cDNA clone.

in transcription proceeding further downstream. Such a longer transcript would most likely include the downstream exons (known from ESTs) found in the human genomic cosmid because the first of these begins just 5 kb downstream of the AGUAAA poly(A) signal. A contig made from the downstream ESTs has a high level of sequence similarity to *Pbx* genes (data not shown), which are transcription factors that cooperate with HOX and Engrailed proteins to bind DNA and activate transcription (Phelan *et al.*, 1995; van Dijk *et al.*, 1995; Peltenburg and Murre, 1996, 1997). We are currently determining the full cDNA sequence of this *Pbx*-like gene and the possibility that a bicistronic mRNA, including the lp_{A2} ORF, is transcribed. *Pbx* genes have been implicated in oncogenesis ever since translocation and fusion [t(1:19) (q23;p13.3)] of *Pbx1* to the *E2A* gene were shown to precede the development of many acute lymphoblastic leukemias (Kamps *et al.*, 1990) A similar situation may be occurring with the *Pbx*-like gene downstream of lp_{A2}, because other congenital myeloid leukemias result from t(11;19) (q23;p12–p13.1) translocations (Huret *et al.*, 1993), which is where the human clone containing the *Pbx*-like gene has been mapped. Alternatively, translocations may disrupt the lp_{A2} gene, which also might lead to oncogenicity.

Our finding that the lp_{A2} gene is present as a single copy in *M. musculus* indicates that this gene should be relatively straightforward to mutagenize through homologous recombination. The presence of pseudogenes or duplicated functional genes can make isolation, targeting, and specific detection by Southern blot or PCR very difficult. With respect to copy number, the lp_{A2} gene coding exons are similar to the analogous coding exons of lp_{A1} in *M. musculus* C57BL/6J (Contos and Chun, 1998), but they differ from lp_{A1}'s exon 4 in *M. spretus*, which we have shown to be duplicated on a separate chromosome.

Genes that map to the same mouse chromosomal location as lp_{A2} include *myd*, *Cspg3*, *Rent1*, and *Myo9b*. Homozygous *myd* mice have a diffuse and progressive myopathy, with focal lesions in all skeletal muscles (Lane *et al.*, 1976). Body and organ weights are considerably less than those of littermates, and the mean lifespan of those that survive to puberty is only 17 weeks. It has been proposed that *myd* is the counterpart of human mutations causing facioscapulohumeral dystrophy, which maps to distal 4q (Mills *et al.*, 1995). These results, together with the expression pattern of lp_{A2}, indicate that lp_{A2} and *myd* are not the same gene. We found the 12-kb human *Cspg3* gene, also known as *Neurocan* (Rauch *et al.*, 1995), on a separate 19p12 human cosmid (Accession No. AC002126). Thus there appears to be syntenic conservation in linkage of *Cspg3* and lp_{A2} between human and mouse. Although it was previously proposed that *Rent1* was located at 19p13.11–p13.2 (Perlick *et al.*, 1996; full data not shown), we have found the human gene on two cosmid sequences from 19p12 (AC003972 and AC004517), again suggesting synteny with the mouse. Like *Rent1*, *Myo9b* was determined to be at 19p13.1 (Bahler *et al.*, 1997). Because this gene is not on any sequenced human genomic cosmid clone, there are no conflicting data with the 19p13.1 mapping. However, the chromosomal mapping may have had the same complications in distinguishing 19p13.1 and 19p12 as for *Rent1*. It is possible that all these genes are located at the border of the two cytogenetic bands. Interestingly, the myosin IXb protein contains a RhoGAP domain that suggests that it may be involved in LPA signaling pathways (Post *et al.*, 1998). Morphological responses to LPA-

receptor stimulation rely on Rho activation (Moolenaar *et al.*, 1997; Fukushima *et al.*, 1998), which in other experiments has been shown to cause phosphorylation of myosins through activation of a Rho kinase (Amano *et al.*, 1996). Perhaps there is a genomic regulatory link between lp_{A2} and this RhoGAP.

Another gene that maps fairly close to lp_{A2} is *kat*. The gross phenotype of homozygous *kat* mutant mice (on an RBF background), as well as homozygous kat^{2J} allelic mutant mice, includes polycystic kidney disease, anemia, and male sterility (Janaswami *et al.*, 1997). However, the homozygous *kat* phenotype is dependent on background strain (e.g., on a C57BL/6J background, the only gross phenotypes observed were a domed skull, dwarfing, and a mortality of ~50% before weaning). Interestingly, the more severe phenotypes are in organs where the lp_{A2} gene is most abundantly expressed. The *kat* gene was mapped to a small interval between the *D8Mit128* and the *D8Mit129* marker loci. Although we found no crossovers between lp_{A2} and *D8Mit128* in the BSB panel, there were four crossovers with both *D8Mit128* and *D8Mit129* in the BSS panel. We also excluded mutations in lp_{A2} exons as the cause of the *kat* phenotype through sequencing amplified exons from both *kat* and RBF strains (data not shown). Thus, mutations in the lp_{A2} gene are likely not responsible for the *kat* phenotypes.

Our genomic characterization and sequence comparison results provide a large number of hypotheses that will be interesting to test in future studies. These involve the role of 3′UTR or coding mutations in the oncogenicity of the lp_{A2} gene, the possibility that mutations in lp_{A2} are common in certain types of cancer, and the relation of lp_{A2} and LPA signaling to neighboring transcription units in the genome. Furthermore, we provide the necessary information and reagents to create lp_{A2} targeted mutations. The genomic characterization of the lp_{A2} gene presented thus provides a basis for future experiments that will definitively determine the role of lp_{A2} in both development and disease.

ACKNOWLEDGMENTS

We thank Lucy Rowe and Mary Barter for alignment of the segregation mapping results and providing the linkage maps; Songzhu An for providing the Edg4 cDNA clone; Carol Akita for sequencing one of the Edg4 clones; and K. C. Cox for copyediting the manuscript. This work was supported by the National Institute of Mental Health and by a grant from Allelix Biopharmaceuticals.

REFERENCES

Abood, M. E., Ditto, K. E., Noel, M. A., Showalter, V. M., and Tao, Q. (1997). Isolation and expression of a mouse CB1 cannabinoid receptor gene. Comparison of binding properties with those of native CB1 receptors in mouse brain and N18TG2 neuroblastoma cells. *Biochem. Pharmacol.* **53:** 207–214.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Amano, M., Ito, M., Kimura, K., Fukata, Y., Chihara, K., Nakano, T., Matsuura, Y., and Kaibuchi, K. (1996). Phosphorylation and activation of myosin by Rho-associated kinase (Rho-kinase). *J. Biol. Chem.* **271:** 20246–20249.

An, S., Bleu, T., Hallmark, O. G., and Goetzl, E. J. (1998). Characterization of a novel subtype of human G protein-coupled receptor for lysophosphatidic acid. *J. Biol. Chem.* **273:** 7906–7910.

An, S., Bleu, T., Huang, W., Hallmark, O. G., Coughlin, S. R., and Goetzl, E. J. (1997). Identification of cDNAs encoding two G protein-coupled receptors for lysosphingolipids. *FEBS Lett.* **417:** 279–282.

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," Wiley, New York.

Bahler, M., Kehrer, I., Gordon, L., Stoffler, H. E., and Olsen, A. S. (1997). Physical mapping of human myosin-IXB (MYO9B), the human orthologue of the rat myosin myr 5, to chromosome 19p13.1. *Genomics* **43:** 107–109.

Birnstiel, M. L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3′ processing: The end is in site! *Cell* **41:** 349–359.

Chun, J., Contos, J., and Munroe, D. (1999). A growing family of receptor genes for lysophosphatidic acid (LPA) and other lysophospholipids (LPs). *Cell. Biochem. Biophys.* **30:** 213–242.

Contos, J. J., and Chun, J. (1998). Complete cDNA sequence, genomic structure, and chromosomal localization of the LPA receptor gene, lpA1/vzg-1/Gpcr26. *Genomics* **51:** 364–378.

Durieux, M. E. (1995). "Lysophosphatidate Signaling: Cellular Effects and Molecular Mechanisms," Landes, Austin, TX.

Felder, C. C., Briley, E. M., Axelrod, J., Simpson, J. T., Mackie, K., and Devane, W. A. (1993). Anandamide, an endogenous cannabimimetic eicosanoid, binds to the cloned human cannabinoid receptor and stimulates receptor-mediated signal transduction. *Proc. Natl. Acad. Sci. USA* **90:** 7656–7660.

Fernhout, B. J., Dijcks, F. A., Moolenaar, W. H., and Ruigt, G. S. (1992). Lysophosphatidic acid induces inward currents in *Xenopus laevis* oocytes: Evidence for an extracellular site of action. *Eur. J. Pharmacol.* **213:** 313–315.

Freedman, N. J., and Lefkowitz, R. J. (1996). Desensitization of G protein-coupled receptors. *Rec. Prog. Horm. Res.* **51:** 319–353.

Fukushima, N., Kimura, Y., and Chun, J. (1998). A single receptor encoded by vzg-1/lpA1/edg-2 couples to G-proteins and mediates multiple cellular responses to lysophosphatidic acid (LPA). *Proc. Natl. Acad. Sci. USA* **95:** 6151–6156.

Hecht, J. H., Weiner, J. A., Post, S. R., and Chun, J. (1996). Ventricular zone gene-1 (vzg-1) encodes a lysophosphatidic acid receptor expressed in neurogenic regions of the developing cerebral cortex. *J. Cell. Biol.* **135:** 1071–1083.

Huret, J. L., Brizard, A., Slater, R., Charrin, C., Bertheas, M. F., Guilhot, F., Hahlen, K., Kroes, W., van Leeuwen, E., Schoot, E. V., Beishuizen, A., Tanzer, J., and Hagemeijer, A. (1993). Cytogenetic heterogeneity in t(11;19) acute leukemia: Clinical, hematological and cytogenetic analyses of 48 patients—Updated published cases and 16 new observations. *Leukemia* **7:** 152–160.

Imamura, F., Horai, T., Mukai, M., Shinkai, M., Sawada, M., and Adkedo, H. (1993) Induction of *in vitro* tumor cell invasion of cellular monolayers by lysophosphatidic acid or phospholipase D. *Biochem. Biophys. Res. Commun.* **193:** 497–503.

Janaswami, P. M., Birkenmeier, E. H., Cook, S. A., Rowe, L. B., Bronson, R. T., and Davisson, M. T. (1997). Identification and genetic mapping of a new polycystic kidney disease on mouse chromosome 8. *Genomics* **40:** 101–107.

Kamps, M. P., Murre, C., Sun, X. H., and Baltimore, D. (1990). A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. *Cell* **60:** 547–555.

Lane, P. W., Beamer, T. C., and Myers, D. D. (1976). Myodystrophy, a new myopathy on chromosome 8 of the mouse. *J. Hered.* **67:** 135–138.

Lee, M. J., Van Brocklyn, J. R., Thangada, S., Liu, C. H., Hand, A. R., Menzeleev, R., Spiegel, S., and Hla, T. (1998). Sphingosine-1-

phosphate as a ligand for the G protein-coupled receptor EDG-1. *Science* **279:** 1552–1555.

Liu, C. H., and Hla, T. (1997). The mouse gene for the inducible G-protein-coupled receptor edg-1. *Genomics* **43:** 15–24.

Marchuk, D., Drumm, M., Saulino, A., and Collins, F. S. (1991). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19:** 1154.

Mills, K. A., Mathews, K. D., Scherpbier-Heddema, T., Schelper, R. L., Schmalzel, R., Bailey, H. L., Nadeau, J. H., Buetow, K. H., and Murray, J. C. (1995). Genetic mapping near the myd locus on mouse chromosome 8. *Mamm. Genome* **6:** 278–280.

Moolenaar, W. H., Kranenburg, O., Postma, F. R., and Zondag, G. C. (1997). Lysophosphatidic acid: G-protein signalling and cellular responses. *Curr. Opin. Cell Biol.* **9:** 168–173.

Peltenburg, L. T., and Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* **15:** 3385–3393.

Peltenburg, L. T., and Murre, C. (1997). Specific residues in the Pbx homeodomain differentially modulate the DNA-binding activity of Hox and Engrailed proteins. *Development* **124:** 1089–1098.

Perlick, H. A., Medghalchi, S. M., Spencer, F. A., Kendzior, R. J., Jr., and Dietz, H. C. (1996). Mammalian orthologues of a yeast regulator of nonsense transcript stability. *Proc. Natl. Acad. Sci. USA* **93:** 10928–10932.

Phelan, M. L., Rambaldi, I., and Featherstone, M. S. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell. Biol.* **15:** 3989–3997.

Post, P. L., Bokoch, G. M., and Mooseker, M. S. (1998). Human myosin-IXb is a mechanochemically active motor and a GAP for rho. *J. Cell. Sci.* **111:** 941–950.

Powell, A. J., Gates, P. B., Wylie, D., Velloso, C. P., Brockes, J. P., and Jat, P. S. (1998). Immortalization of rat embryo fibroblasts by a 3′-untranslated region. *Exp. Cell Res.* **240:** 252–262.

Raff, T., van der Giet, M., Endemann, D., Wiederholt, T., and Paul, M. (1997). Design and testing of beta-actin primers for RT-PCR that do not co-amplify processed pseudogenes. *BioTechniques* **23:** 456–460.

Rauch, U., Grimpe, B., Kulbe, G., Arnold-Ammer, I., Beier, D. R., and Fassler, R. (1995). Structure and chromosomal localization of the mouse neurocan gene. *Genomics* **28:** 405–410.

Ridley, A. J., and Hall, A. (1992). The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70:** 389–399.

Rimokh, R., Berger, F., Bastard, C., Klein, B., French, M., Archimbaud, E., Rouault, J. P., Santa Lucia, B., Duret, L., Vuillaume, M., *et al.* (1994). Rearrangement of CCND1 (BCL1/PRAD1) 3′ untranslated region in mantle-cell lymphomas and t(11q13)-associated leukemias. *Blood* **83:** 3689–3696.

Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S., Thurston, S. J., and Birkenmeier, E. H. (1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5:** 253–274.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Smit, A. F. (1996). The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6:** 743–748.

Stam, J. C., Michiels, F., van der Kammen, R. A., Moolenaar, W. H., and Collard, J. G. (1998). Invasion of T-lymphoma cells: Cooperation between Rho family GTPases and lysophospholipid receptor signaling. *EMBO J.* **17:** 4066–4074.

Stella, N., Schweitzer, P., and Piomelli, D. (1997). A second endogenous cannabinoid that modulates long-term potentiation. *Nature* **388:** 773–778.

van Corven, E. J., Groenink, A., Jalink, K., Eichholtz, T., and Moolenaar, W. H. (1989). Lysophosphatidate-induced cell proliferation: Identification and dissection of signaling pathways mediated by G proteins. *Cell* **59:** 45–54.

van Dijk, M. A., Peltenburg, L. T., and Murre, C. (1995). Hox gene products modulate the DNA binding activity of Pbx1 and Pbx2. *Mech. Dev.* **52:** 99–108.

Wahle, E. (1995). Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.* **270:** 2800–2808.

Xu, Y, Fang, X. J., Casey, G., and Mills, G. B. (1995). Lysophospholipids activate ovarian and breast cancer cells. *Biochem. J.* **309:** 933–940.

Xu, Y, Shen, Z., Wiper, D. W., Wu, M., Morton, R. E., Elson, P., Kennedy, A. W., Belinson, J., Markman, M., and Casey, G. (1998). Lysophosphatidic acid as a potential biomarker for ovarian and other gynecologic cancers. *J. Am. Med. Assoc.* **280:** 719–723.

Zhang, G., Contos, J. J., Weiner, J. A., Fukushima, N., and Chun, J. (1999). Comparative analysis of three murine G-protein coupled receptors activated by sphingosine-1-phosphate. *Gene* **227:** 89–99.

Zondag, G. C., Postma, F. R., Etten, I. V., Verlaan, I., and Moolenaar, W. H. (1998). Sphingosine 1-phosphate signalling through the G-protein-coupled receptor Edg-1. *Biochem. J.* **330:** 605–609.