Apolipoprotein L gene family: tissue-specific expression, splicing, promoter regions; discovery of a new gene

Philippe N. Duchateau,[1,2,*] Clive R. Pullinger,[1,*] Min H. Cho,[*] Celeste Eng,[*] and John P. Kane[†,§]

Cardiovascular Research Institute,* Department of Medicine,† and Department of Biochemistry and Biophysics,§ University of California, San Francisco, CA 94143-0130

Abstract **Previously we identified and cloned the cDNA for a new protein, apolipoprotein L (apoL), present in plasma and mainly associated with large high density lipoprotein particles. Using 5′ rapid amplification of cDNA ends, RT-PCR and comparison with three Human Genome Project and three expressed sequence tag sequences, we have characterized the gene for apoL and for three additional, highly homologous proteins that constitute a new family of proteins that display no homology with previously described apolipoproteins. The genes for all four proteins, apoL-I, apoL-II, apoL-III, and apoL-IV, are located at chromosome 22q12.1-13.1 within a 127,000-bp region. The apoL-I gene is in the opposite orientation to the other three. All four genes have TATA-less promoters, which contain putative sterol regulatory elements, suggesting that transcription of these genes may be coordinated with that of the low density lipoprotein receptor and genes in pathways involving the synthesis of triglycerides and cholesterol. The gene family has a consensus eight-exon structure with alternative splice sites that could produce as many as eight distinct gene products. The apoL-II and apoL-III genes have alternative transcriptional start sites as a result of additional 5′ exons. apoL-I, apoL-II, and apoL-III are expressed to the highest degree in the lung. Other tissues with high expression are the pancreas, prostate, spleen, liver, and placenta. Four clustered common polymorphisms, three of which altered the protein sequence, were found in apoL-I, all in linkage disequilibrium, and describing two haplotypes: the more common $Lys_{166}/Ile_{244}/Lys_{271}$ and the rarer $Glu_{166}/Met_{244}/Arg_{271}$.**—Duchateau, P. N., C. R. Pullinger, M. H. Cho, C. Eng, and J. P. Kane. **Apolipoprotein L gene family: tissue-specific expression, splicing, promoter regions; discovery of a new gene.** ***J. Lipid Res.*** **2001.** 42: **620–630.**

Supplementary key words HDL • alternative splicing • lipoprotein • chromosome 22

Epidemiological studies have shown that individuals with low levels of high density lipoproteins (HDL) in blood have a substantially increased risk of coronary artery disease (1, 2). Mechanisms by which HDL protect against atherosclerosis involve reverse cholesterol transport (3), whereby HDL acquire cholesterol from peripheral cells and facilitate its esterification and delivery to the liver, and the ability of HDL to impede the oxidation of other plasma lipoproteins (4, 5). Other protective roles may be revealed by identifying novel protein constituents of HDL (6, 7). Because ultracentrifugation causes protein dissociation and can modify HDL structures, we used a purification strategy that conserves lipoprotein integrity: selected affinity immunosorption, which isolates lipoproteins under minimally perturbing conditions (8). Using this technique, combined with two-dimensional electrophoresis, we isolated apolipoprotein L (apoL), a novel 42-kDa protein (9), together with a 38-kDa truncated form of apoL. Selected affinity isolation, using an antibody to apoL, showed the protein to be associated chiefly with large HDL particles (9). Secondary structure analysis of the protein sequence, deduced from the cDNA sequence, revealed the presence of four lipid-binding amphipathic α-helices. Computer-based homology searches revealed no significant similarity with other proteins. Furthermore, plasma levels of apoL were shown to be correlated with total plasma triglycerides and cholesterol in normolipidemic subjects, whereas triglycerides were the major independent covariate influencing apoL concentration in plasma of dislipidemic patients and those with type 2 diabetes (10).

In the present study we have determined the 5′ end of apoL (now termed apoL-I) mRNA using 5′-rapid amplification of cDNA ends (RACE). By reverse transcriptase PCR we have found an alternatively spliced variant containing an additional exon. These studies revealed the presence of apoL-I gene polymorphisms, the frequencies of which were studied among a group of control individuals. In the course of our studies, clones for three closely related cDNAs were obtained and analyzed. Using Internet-based Basic Local Alignment Search Tool (BLAST) searches and comparison with the sequences of these

Abbreviations: RACE, rapid amplification of cDNA ends; EST, expressed sequence tag; SRE, sterol regulatory element; HDL, high density lipoprotein; BLAST, Basic Local Alignment Search Tool; UTR, untranslated region; TNF, tumor necrosis factor.

[1] P. N. Duchateau and C. R. Pullinger contributed equally to this study.

[2] To whom correspondence should be addressed.

cDNAs, we have studied the structure of a family of four apoL-related genes. The locations of the promoter regions were determined. These were examined for regulatory elements that potentially control expression of the apoL genes. Sites of expression of the apoL genes were determined by establishing a competitive quantitative RT-PCR assay and by dot blotting.

METHODS

Identification of homologous genes

We compared our previously reported apoL-I cDNA sequence (9) and the additional 5′ sequence of apoL-I that we determined by 5′ RACE with GenBank entries using BLAST at http://www.ncbi.nlm.nih.gov. Primers were designed using the sequences with accession numbers Z95144 and Z82215, and PCR was carried out using human genomic DNA. These PCR products were analyzed to confirm the sequences of the intron-exon junctions predicted from a comparison of these genomic sequences and our cDNA clones. Full-length cDNA clones for apoL-I, apoL-II, apoL-III, and apoL-IV were generated by RT-PCR (see below). DNA sequencing was performed directly on PCR products or cDNA clones using the Thermo Sequenase Cycle Sequencing kit (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) with primers end-labeled with [γ-^{33}P]ATP (Amersham). The extent of the homology between the corresponding exons of the four genes was compared using the GeneWorks program (Oxford Molecular Group, Inc., Campbell, CA). Clustal alignment analysis of the peptide sequences was performed using the MacVector program (Oxford Molecular Group).

RACE assays

To determine the 5′ untranslated portion of the apoL-I mRNA, we performed conventional 5′ RACE using two primers designed from our previously reported cDNA sequence (9). Primer #453 (**Table 1**) was used to make the first strand cDNA from human pancreatic poly A$^+$ RNA using a 5′/3′ RACE kit (Boehringer Mannheim Corp., Indianapolis, IN). Primer #449 (Table 1), together with the oligo d(T)-anchor primer, was used for the RACE reaction. Nested PCR was performed using primer #404 (Table 1) and the anchor primer. PCR products were cloned into pCR-Script-SK(+) (Stratagene, La Jolla, CA) and the clones sequenced as described above. Additional primers were designed from exon 1 and exon 3 sequences, revealed by these studies, and from homologous regions in the three other genes. These were used for 5′ RACE studies, again using the Boehringer kit. The partial exon 1a sequence of apoL-III, found using primers #495 and #494 (Table 1), was used to design primers #514 and #508 (Table 1). These were then used in 5′ RACE to determine the 5′ end of exon 1a.

An antisense degenerate primer, #474 (Table 1), in exon 1 that hybridizes to apoL-I, II, and IV was used to prepare cDNA from human placental poly A$^+$ RNA using the Marathon cDNA amplification kit (Clontech Laboratories Inc., Palo Alto, CA). The cDNA was ligated to the Marathon cDNA adaptor and subjected to PCR with the adaptor primer and with either primer #501 (apoL-I- and IV-specific) or #502 (apoL-II-specific) using the Advantage cDNA PCR kit (Clontech).

Predicted NH_2-terminal protein sequences were analyzed for the presence of signal peptides and the position of cleavage sites

TABLE 1. Oligouucleotide primers used in this study

Primers For	Primer Number	Sequence	Gene	Hybridizing Region–Orientation
Tissue blot	377	5′-TCGTCGGCATGGGTCTGGC-3′	L-I	Exon 7–Sense
	385	5′-AATGTTTAGCTTCTCCTCC-3′	L-I	Exon 7–Antisense
Exon 7 polymorphism	440	5′-CACCTACCATGAAAGAGCATC-3′	L-I	Intron 6–Sense
	441	5′-CTTTACCTCACCCTCTTTATCC-3′	L-I	Exon 7–Antisense
Quantitative RT-PCR	482 (d)	5′-CCGTGCC(A/C)TCAG(G/A)C(A/G)AGCCAGAGCC-3′	L-I, II, III	Exon 7 of L-I, Exon 5 of L-II and L-III)–Sense
	483 (d)	5′-GACTTTGCCCCCTCA(T/A)G(T/C)AAGTGCTTTG-3′	L-I, II, III	Exon 7 of L-I, Exon 5 of L-II and L-III)–Antisense
	484 (d)	5′-GACTTTGCCCCCTCA(T/A)G(T/C)AAGTGCTTTG GCTTACAGGGGCCACATCC-3′	L-I	Exon 7–Antisense
Splice variants	492	5′-GCTCATCTTTCTGCTGCACATATTCGTCC-3′	L-III	Exon 5–Antisense
	493	5′-GGAAGTATTTGGTGGCCTCTTCAGTAAAGC-3′	L-III	Exon 4–Antisense
	509	5′-GATTTTGCCCGAAGTTGTGAGAC-3′	L-III	Exon 1a–Sense
Cloning	448	5′-GTGGGATCCACACAGCTCAGAACAG-3′	L-I, II, III, IV	Exon 1 of L-I, II, III, IV–Sense
	456 (d)	5′-AAAAAGT(T/C)TGCATTTTGTCCTGGCC-3′	L-I, II	Exon 7 of L-I and Exon 5 of L-II – Antisense
	485	5′-ACGTGTCTGGTTATTATATAGG-3′	L-IV	Exon 1–Sense
	488	5′-CTGGTGGCTGCACTGCTCTGG-3′	L-IV	Exon 6–Antisense
5′ RACE	453	5′-GGAGTAGCAGATTCTGTGTGC-3′	L-I	Exon 6–Antisense
	449	5′-TTGAGGATCTCCAGTATCTGTCCC-3′	L-I	Exon 5–Antisense
	404	5′-GTTTTGTTGCACCCTCGCTCCAGCTTCCTC-3′	L-I	Spanning exons 4 and 5–Antisense
	494	5′-CATAATAACCAGACACGTTCTCCAGTCTCTG-3′	L-III	Exon 1 of L-I, II, III, IV–Antisense
	495	5′-GGATCCCACCTCCAGCCGTGCATCTGCA-3′	L-III	Exon 1 of L-I, II, III, IV–Antisense
	508	5′-GGTTTCAGTCTCACAACTTCGGG-3′	L-III	Exon 1a–Antisense
	514	5′-TACCCTTTTCCCCAGACCCACAC-3′	L-III	Exon 1a–Antisense
	501 (d)	5′-GTCCTCCAG(T/C)CCCCAAGATATACC-3′	L-I, IV	Exon 1 of L-I, II, III, IV–Antisense
	502	5′-GTCCTCCGGCCTCCCAGATATACC-3′	L-II	Exon 1 of L-I, II, III, IV–Antisense
	474 (d)	5′-CTG(A/G)CAGAGACTGAGCAAGATCC-3′	L-I, II, IV	Exon 1 of L-I, II, III, IV–Antisense

After the primer number(d) indicates that the oligonucleotide primer is degenerate, and the actual degeneracy is indicated in parentheses in the sequence.

using SignalP (11) (http://130.225.67.199/services/SignalP/index.html).

Poly A$^+$ RNA dot blot

A human RNA Master Blot (Clontech) was hybridized with an apoL-I probe, strictly adhering to the recommended protocol. Each dot contained 100–500 ng of purified poly A$^+$ RNA from 1 of 50 human tissues. The individual loading had been normalized for the expression of eight housekeeping genes. A 580-bp fragment from the 5′ end of exon 7 of apoL-I was generated by PCR, using primers #377 and #385, and labeled with [α-^{32}P]dCTP by random priming using the Multiprime DNA Labelling System (Amersham). The blot was exposed to preflashed film, and the radioautogram was scanned and then analyzed on a Macintosh computer using the NIH image program (available at http://rsb.info.nih.gov/nih-image).

Determination of splicing variants by RT-PCR

After exon 1 of apoL-I and apoL-II had been identified by 5′ RACE, primer #448 was synthesized along with #483 in exon 7 of apoL-I to perform RT-PCR. These two primers were designed so that they hybridize to the homologous sequences in both apoL-I and apoL-II. Random hexamers and AMV reverse transcriptase (Boehringer Mannheim) were used to synthesize cDNA. PCR products were cloned into pCR-Script and sequenced to ascertain the apoL-I and apoL-II splicing patterns. In initial studies, many of these clones were from a third homologous gene, which we named apoL-III. To determine the relative amounts of the apoL-III splice variants in different tissues, RT-PCR products that had been synthesized were digested separately with restriction enzymes specific for apoL-I, apoL-II, or apoL-III. These were BclI, SpeI, and HaeII, respectively. The reverse transcriptase reactions were carried out at 42°C for 60 min using 1 μg of total RNA from eight tissues (Clontech), each in a total of 10 μl containing 5 μM random hexamers, 1 mM dNTPs, 12.5 units RNase inhibitor (Boehringer Mannheim), 20 units AMV reverse transcriptase (Boehringer Mannheim), 50 mM Tris-HCl, 8 mM $MgCl_2$, 30 mM KCl, and 1 mM dithiotheitol at pH 8.5. The PCR reactions were carried out in the same tubes in a total volume of 50 μl in 50 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$, 3 mM $MgCl_2$, 200 μM dNTPs, 2.5 units of Hot Tub polymerase (Amersham), and 20 pmols of each primer (#448 and #483). The tubes were heated for 5 min at 95°C and subjected to 24 cycles: 62°C for 30 s, 72°C for 1 min, and 96°C for 30 s.

Full-length apoL-IV clones were obtained by RT-PCR using the above protocol. For this purpose, two primers were designed (#485 and #488, Table 1) specific for apoL-IV from a comparison of the exon 1 and the 3′ untranslated region (UTR) sequences of the four genes.

Quantitative RT-PCR

To measure simultaneously the concentrations of apoL-I, apoL-II, and apoL-III mRNA in several tissues, two degenerate primers (#482 and #483, Table 1) that hybridize to homologous sequences in the terminal exon of each gene were used for quantitative RT-PCR. The standard RNA used in these assays was generated from an apoL-I exon 7 PCR product using T7 RNA polymerase (Boehringer Mannheim). This PCR fragment was produced from an apoL-I pCR-Script clone using a T7 primer and primer #484, which was designed to yield a product containing a 40-bp deletion. In the RT-PCR reaction, the standard RNA yields a 206-bp band. To check the reproducibility of the assay, two series of control reactions were carried out as described by Zhang and co-workers (12). In one series (T-type RT-PCR), a fixed amount of standard was used with varying amounts of total RNA. In the other (C-type RT-PCR), both the standard and total amounts of RNA were fixed and the number of cycles was varied. The basic RT-PCR reaction conditions were as described above with 25 ng of total RNA used for each tissue.

Analysis of 5′-flanking sequences

A comparison of scores generated on the WWW from the TRANSFAC database using Transcription Element Search Software (TESS) (13) (URL: http://www.cbil.upenn.edu/tess/index.html), and by MatInspector V2.2 (14) (URL: http://www.gsf.de/biodv/matinspector.html) was made to locate, in the 5′ flanking regions, regulatory elements that potentially control the expression of the apoL genes.

ApoL-I polymorphism analysis

We noticed a number of sequence discrepancies between our published apoL-I cDNA and Z82215, notably in the region corresponding to exon 7. These were investigated by amplifying this region using primers #440 and #441 by PCR, using genomic DNA from 26 unselected control subjects. PCR conditions were as described above, except that 38 cycles were performed: 94°C for 20 s, 61°C for 20 s, and 72°C for 40 s. These PCR products were directly sequenced using the Thermo Sequenase Cycle Sequencing kit (Amersham). An additional 150 genomic DNA samples were analyzed for the HindIII and NcoI RFLPs by agarose gel electrophoresis following enzyme digestion.

RESULTS

ApoL-I polymorphisms

Our previously reported apoL-I cDNA sequence, together with the additional 5′ sequence determined by 5′ RACE, was compared with GenBank entries using BLAST. These searches found three expressed sequence tags (ESTs). One consisted of exons 4 and 5 (H55286) and two contained parts of exon 7 (AA847557 and AA194154). We also found, with the exception of a few differences, a match in the genomic sequence Z82215. All but one of the differences was in exon 7, and these were examined further by PCR and sequencing from 26 individuals. As a result, a cluster of four polymorphisms was discovered within a 316-bp region (**Table 2**). Three alter the predicted protein sequence. Of these, one is a HindIII RFLP (Lys_{166}→Glu), and another an NcoI RFLP (Ile_{244}→ Met) (**Fig.1**). These were studied in additional control individuals to determine more precisely the allele frequencies (Table 2). These two are in strong linkage disequilibrium. The more limited sequencing data also show the Lys_{271}→ Arg polymorphism to be in linkage disequilibrium with both the Lys_{166}→Glu and Ile_{244}→Met polymorphisms. Thus we observed two haplotypes: the more common $Lys_{166}/Ile_{244}/Lys_{271}$ and the less common $Glu_{166}/Met_{244}/Arg_{271}$. In **Fig. 2**, the sequence of the apoL-I protein is from the $Lys_{166}/Ile_{244}/Lys_{271}$ haplotype.

Two other anomalies found when examining the sequences in the 26 individuals showed consistent agreement with the Z82215 sequence and differed from our original cDNA. Codon 171 was always ATA (Ile), not ACA (Thr), and codon 362 was GTG (Val), not GCG (Ala). These changes are reflected in Fig. 2. Conversely, with an-

TABLE 2. ApoL-I exon 7 polymorphisms and allele frequencies

cDNA	Position a.a. Residue	Z82215	Polymorphism			Screening Method
			Codon (Allele Frequency)	a.a. Residue	Genotype and Frequency	
770 (n = 168)	166	33580	**A**AG/**G**AG (0.771)/(0.229)	Lys/Glu	A/A–102/168 A/G–55/168 G/G–11/168	HindIII
976 (n = 25)	234	33786	GC**A**/GC**C** (0.700)/(0.300)	Ala/Ala	A/A–12/25 C/A–11/25 C/C–2/25	Sequencing
1006 (n = 172)	244	33816	AT**A**/AT**G** (0.779)/(0.221)	Ile/Met	A/A–107/172 A/G–54/172 G/G–11/172	NcoI
1086 (n = 26)	271	33896	**A**AG/**G**GG (0.712)/(0.288)	Lys/Arg	A/A–13/26 G/A–11/26 G/G–2/26	Sequencing

other difference in the signal peptide, we found codon 40 in exon 4 to be consistently AGA (Arg), as we originally reported, and not GGA (Gly) as in Z82215.

Discovery of a family of four apoL genes

ESTs and genomic sequences for two additional genes (apoL-II and apoL-III) with close homology to apoL-I were detected in Z82215 (apoL-II) and Z95114 (apoL-II and apoL-III). While the present studies were being completed, the sequence of the euchromatic part of human chromosome 22 was reported (15). The authors identified the presence of apoL-I sequences, except for that part of the signal peptide encoded by exons 2 and 3, and the noncoding exon 1. They also identified a second homologous gene and named it apoL-II, as we did. The regions corresponding to exons 3, 4, and the coding region of exon 5 of apoL-II were identified by the authors (15).

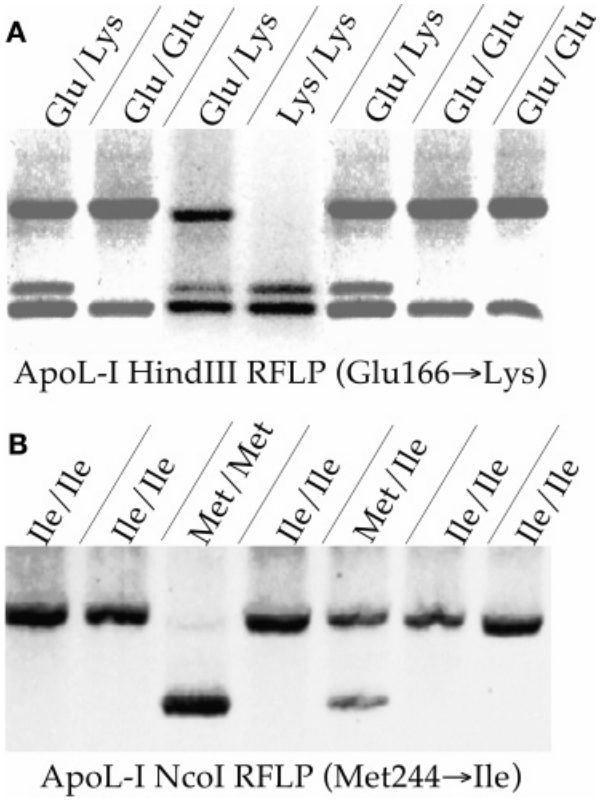

Fig. 1. Agarose gels showing the apoL-I HindIII (A) and NcoI (B) RFLPs in exon 7.

Part of the gene for a TNF-inducible protein CG12-1 (16) was also identified in the chromosome 22 report (15). The mRNA sequence for this protein corresponds to part of exon 1 and exons 2–3, 4, 5 of the apoL-III gene. This is one of the transcripts we identified, apoL-IIIα/a (**Fig. 3**), and predicts the peptide of intermediate length. We prefer to use the name apoL-III to be consistent with the nomenclature of apoL-I and apoL-II. The translational start site ATG sequences in exon 1 and exon 3 are both in a strong context and that in exon 5 in a weaker, but adequate, context, according to Kozak's rules for initiation of translation (17). Database searches revealed an EST (AC360265) containing exon 3 and parts of exons 1 and 4 of apoL-III. The apoL-III gene extends for 26 kb (**Fig. 4**).

During the analysis of these genes and apoL-I by RT-PCR, we produced clones from a fourth gene, which we have named apoL-IV. As work on Z95114 progressed and was reported, we were able to locate the gene for apoL-IV on this contig. Analysis of the apoL-IV sequence predicted two peptides, according to the splicing pattern (Fig. 3), as a result of exon 3 being present or not. These proteins, respectively, are 39.2 kDa (apoL-IVa) and 38.8 kDa (apoL-IVb). The corresponding mRNA sequences have been submitted to GenBank. The accession numbers are AY014908 and AY014915, respectively. The apoL-IV exon 2 is homologous to exon 3 of apoL-I, whereas exon 3 of apoL-IV is unique. The translational start site ATG in exon 3 is in strong context, and that in exon 2 is in an adequate context (17).

The nucleotide sequences for all of the exons described in Fig. 3 have been submitted to GenBank. The accession numbers are apoL-I, AF 323542–AF323548; apoL-II, AF324225–AF324230; apoL-III, AF324231–AF324238; and apoL-IV, AY014909- AY014914.

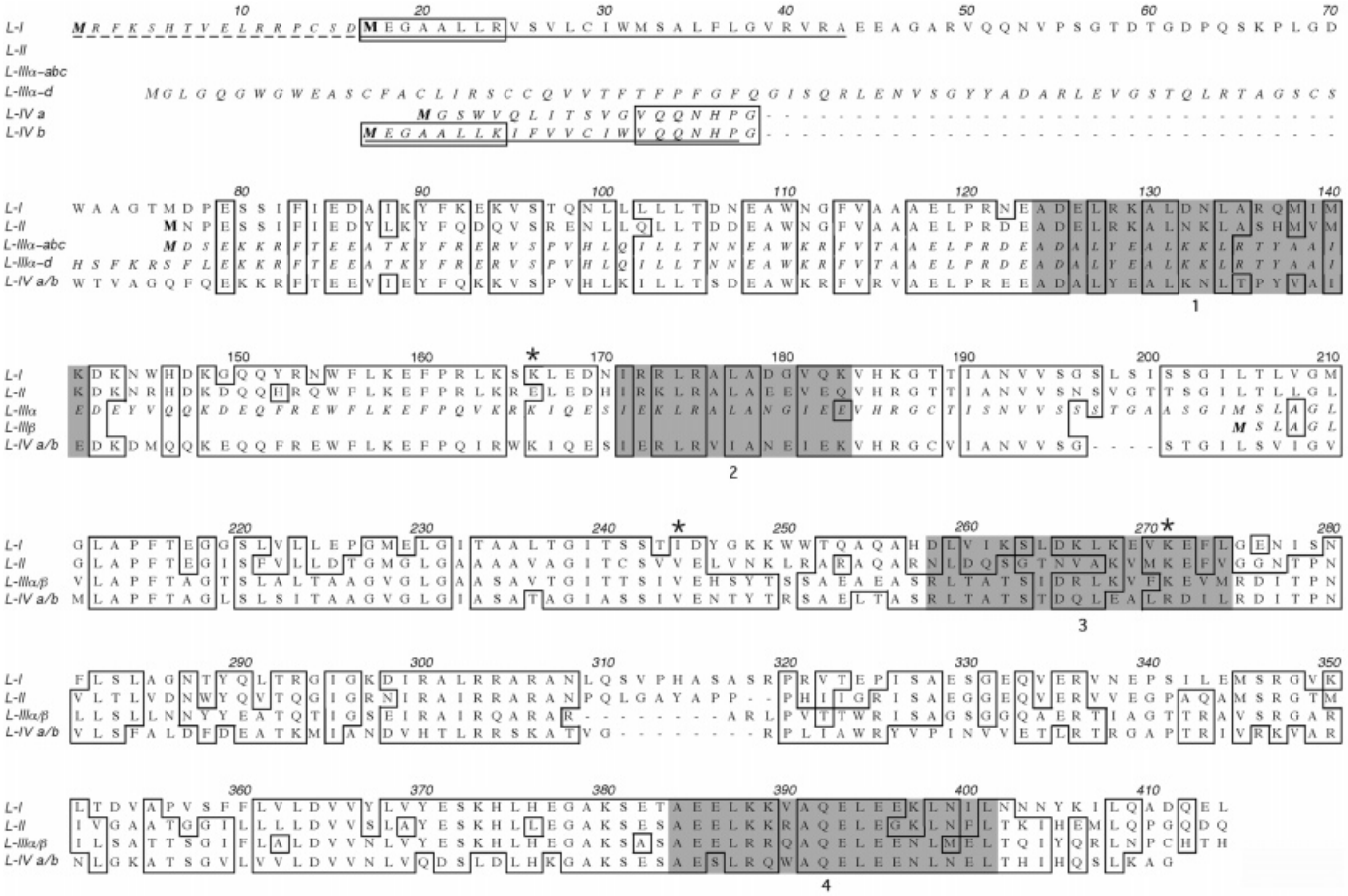

Fig. 2. Alignment of the four apoL gene family proteins. Boxed areas show similar or identical residues. The N-termini of apoL-I, apoL-III, and apoL-IV differ according to the splicing pattern. For apoL-III and apoL-IV, the three and two N-termini, respectively, are shown separately. In the case of apoL-Ib, the different N-terminus is indicated by italics. Conserved amphipathic alpha helices are shaded, and possible signal peptides are underlined. The apoL-I sequence shown represents the most common of two haplotypes. This is $Lys_{166}/Ile_{244}/Lys_{271}$, the less common being $Glu_{166}/Met_{244}/Arg_{271}$. These residues are indicated by *.

Determination of the 5′ and 3′ ends of the apoL genes by RACE

An additional 372-bp sequence was found by 5′ RACE upstream of the beginning of our original cDNA for apoL-I (9). Comparison of the complete sequence with the human chromosome 22 sequence from clone RP1-6802 (accession number Z82215) revealed that apoL-I gene spans 14 kb (Fig. 4) and comprises 7 exons (Fig. 3). The sequences flanking the predicted intron-exon boundaries for apoL-I were confirmed by PCR, and sequencing using intronic primers designed from the Z82215 sequence.

Examination of clones from the 5′ RACE experiments indicated that exon 2 was occasionally present in transcripts. This exon contains an in-frame ATG codon in an adequately strong context consistent with Kozak's rules (17). The presence of exon 2 predicts a protein (apoL-Ib, Fig. 3) with a 43-residue signal peptide (Fig. 2). The common transcript, lacking exon 2, predicts a protein (apoL-Ia, Fig. 3) with a signal peptide of 27 amino acid residues, 15 more than we originally reported (9). The accession numbers of the two apoL-I mRNA sequences (apoL-Ia and apoL-Ib) are AF323541 and AF323540, respectively.

Figure 5 shows the combined 5′ RACE results for the four apoL genes. Only the clones sequenced are indicated, and only exon 1 is shown in each case. A total of 23 clones was sequenced for apoL-I. The results indicate multiple transcription start sites for all four genes. Together with the low to moderate abundance of the transcripts, this is consistent with the failure to detect any clear bands by primer extension analysis (data not shown). Three rare clones with longer inserts were found for apoL-III and one for apoL-II. The apoL-III clones contained exon 1b and part of exon 1a (Fig. 4). Further 5′ RACE studies were performed to locate the 5′ end of apoL-III exon 1a. This exon is 1571 bp long and extends from nucleotide 150671 to 149101 (Z95114). Exon 1b is 63 bp and extends from nucleotide 146055 to 145993 (Z95114). Transcripts of apoL-III containing exons 1a and 1b have a foreshortened exon 1 of 121 bp from 145298 to 145178 (Z95114). The accession numbers of these two apoL-III mRNA sequences (apoL-IIIβ/a and apoL-IIIβ/b) are AY014906 and AY014907. Analysis of the apoL-II clone revealed exon 1a, which extends for 205 bp from 8028 to 8232 (Z82215). The accession number corresponding to this rare transcript (apoL-II) is AF324224, and that of the common one (apoL-IIα) is AF324223. Here, the apoL-II exon 1 is also

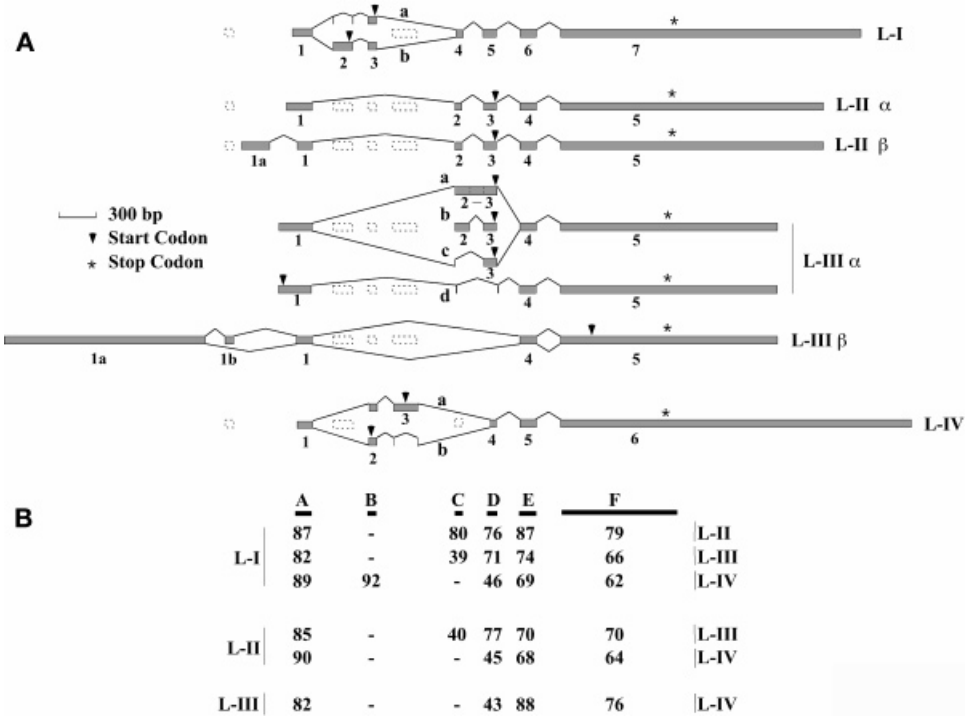

Fig. 3. Comparison of the splicing patterns of the apoL genes (A). Boxes with dotted lines represent homologous regions not found in particular transcripts. Percent similarities between homologous exon regions are shown below (B). The bars (A to F) represent the regions tested for homology.

121 bp, extending from 7857 to 7737 (Z82215), and the splice site at 7857 (Z82215) is homologous to that of apoL-III gene at 145298 (Z95114). Interestingly, this splice site consensus sequence is not conserved (see Fig. 5) in the apoL-I or apoL-IV genes, and no evidence was found for alternative promoters in these other two genes.

The 3′ UTRs of the apoL genes were investigated by a combination of 3′RACE and RT-PCR. We established that the 3′ UTR for apoL-I extended 1.5 k bp to the polyadenylation signal at 35800 (Z82215). The 3′ UTRs of apoL-II and apoL-III are somewhat shorter, each with two polyadenylation signals. For apoL-II these are 762 and 1174 bp downstream of the stop codon at 211913 (Z95114). The corresponding positions for apoL-III are 850 and 1108 bp upstream, and the stop codon is at 125711 (Z95114). There are four polyadenylation signals in the corresponding region of the apoL-IV gene. We found, by 3′ RACE, that the most downstream of these, at 174419 (Z95114), was most commonly used, yielding a 3′ UTR of 1.2 k bp.

Location of regulatory elements in the 5′ flanking regions

Table 3 shows the results of a comparative analysis of the regulatory elements immediately upstream of exon 1. Only sites with high core and matrix similarities have been listed. This region in each case contains a number of AP1 and AP4 sites. Some of these are in conserved locations in two or more of the four genes. Unlike the other three genes, apoL-II has no CAAT box. All have at least one Sp1 site and one GC box. Of interest is the large number of zinc finger binding sites (MZF1), many of which are conserved, containing a core of four or five G residues. The

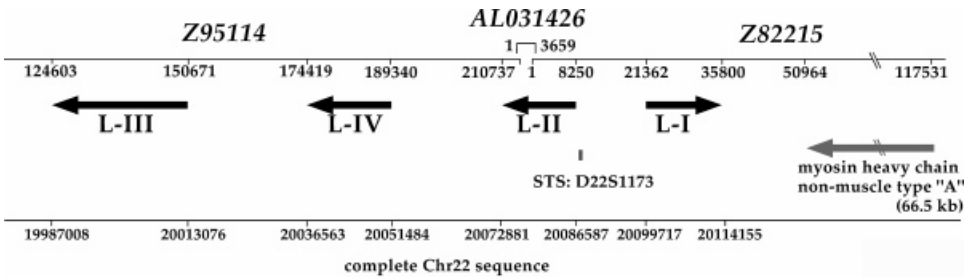

Fig. 4. Arrangement and orientation of the apoL gene family on chromosome 22. The position of the closest sequence-tagged site (D22S1173) is shown, as is that of the neighboring myosin heavy chain nonmuscle type "A" gene.

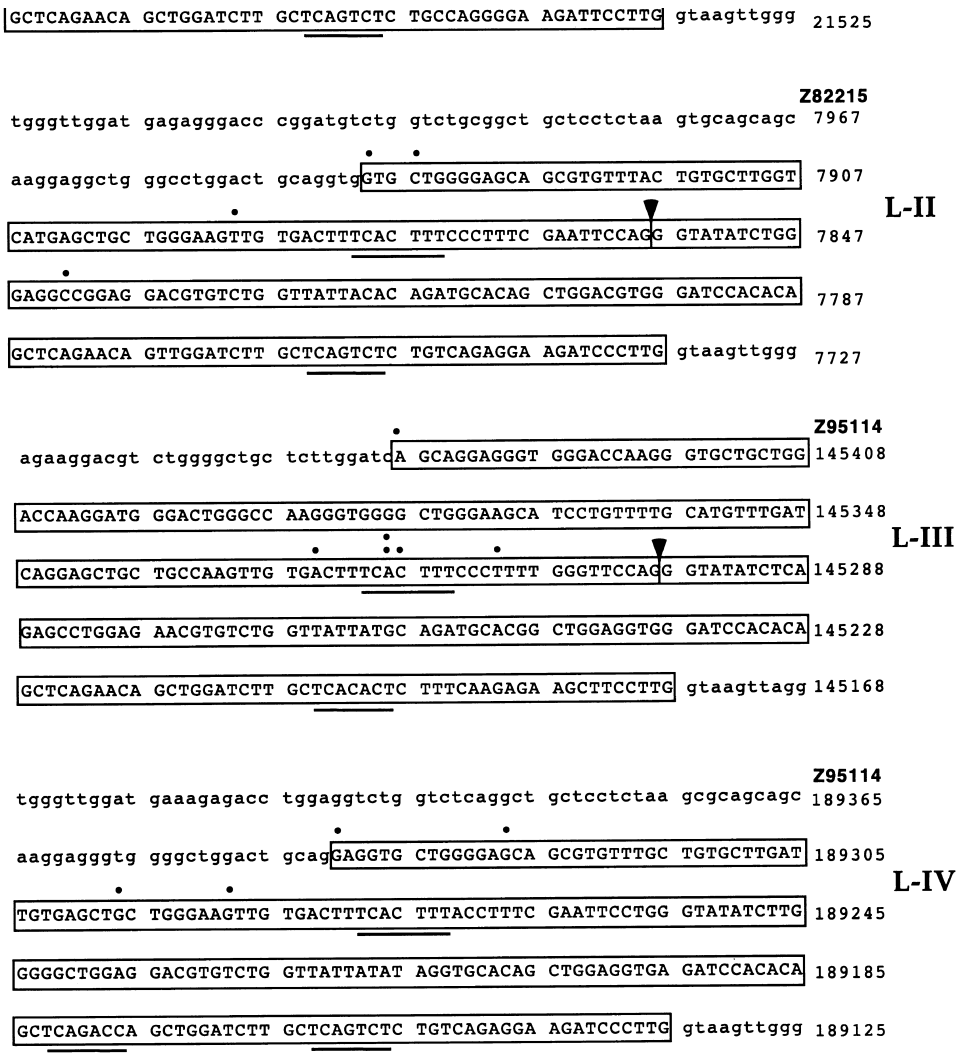

Fig. 5. The exon 1 regions of the four apoL genes as determined by 5′ RACE. The dots represent the 5′ ends of individual clones. Underlined are the Inr sites: those with a solid line are conserved in all four genes. The arrowheads in the apoL-II and apoL-III sequences indicate alternative splice sites. These sites are absent in apoL-I and apoL-IV.

promoters of a number of genes involved in lipid metabolism have sterol regulatory element binding protein (SREBP) sites. The apoL-I promoter has two, and apoL-II, apoL-III, and apoL-IV promoters each have one putative SREBP-1 binding site. In none of the four genes are TATA boxes present in the region flanking exon 1. Two conserved initiator (Inr) sequences, which are often found at or near the transcription start site in TATA-less promoters (18), are present in exon 1 in each case, plus a third in apoL-IV (Fig. 5). Thus it seems that the most commonly used promoter regions of the four apoL genes are TATA-less with multiple transcription initiation sites.

The region upstream of exon 1a of apoL-III was also examined. There is a TATA-like sequence at position −32 from position 150671 (Z95114), which was the start of our 5′ RACE product. A stronger TATA box is at −71 bp, and there is a CAAT box at −90 bp.

The role of the alternative apoL-II and apoL-III promoters is unclear, but it is apparent that they are rarely used. Hence, it seems that the apoL family of genes is, in general, transcriptionally controlled by TATA-less promoters, a characteristic usually associated with so-called housekeeping genes (19). Such genes are often ubiquitously expressed (19), as is the case with the apoL gene family. **Figure 6** shows the results from a poly A^+ dot blot hybridized with an apoL-I-specific probe. It can be seen that apoL-I is expressed in a wide variety of tissues, the only exception being fetal brain.

Analysis of splicing variants by RT-PCR

Initial clones of apoL-I and apoL-III derived by RT-PCR indicated differential splicing. The relative amount of these different transcripts was examined semiquantitatively in a number of tissues. It was not possible to design a specific PCR primer for each gene in exon 1, owing to the high homology. We decided to co-amplify transcripts from apoL-I, apoL-II, and apoL-III, using exon 1 primer #448, which is an exact match for each of these three genes, and

TABLE 3. Comparative analysis of regulatory elements of the apoL gene family

Regulatory Element	ApoL-I	ApoL-II	ApoL-III	ApoL-IV
AP1	−1034	−1038	−1036	−1041
	−1027	−974	—	—
	−897	—	−890	−904
	−571	−568	−571	—
	−544	—	—	—
	−352	−353	—	−359
	—	−328	—	—
AP4	—	−1182	—	−1106
	−944	—	−959	—
	−585	−445	—	—
	−319	−334	—	—
	−231	—	−262	−237
CAAT-box	−1052	—	−1013	−1059
	—	—	−398	—
E47	—	−963	−1180	—
	−855	−453	−872	−876
	−290	−296	−307	−297
	—	−212	—	—
GC-box	—	−420	—	—
	−221	—	−210	−227
HNF-3β	−1157	1088	−544	–
	−825	—	−539	−832
MZF1	−941	−669	—	−1190
	−928	—	−931	−935
	−615	−610	—	−859
	—	−595	—	—
	−404	−453	−420	−411
	−379	−373	−396	−386
NF-1	−1113	−932	−1026	—
	−770	—	—	−777
	−658	—	—	−664
	—	—	—	−554
	—	—	−303	−470
NF-κB	—	—	—	−857
Oct-1	—	−701	−915	—
	—	—	−500	—
RXR-α	—	−946	—	—
SP-1	−1188	−418	−1055	—
	−221	—	—	−227
SREBP-1	−1185	−955	−1200	—
	−746	—	—	−752

Only high scoring matches to the consesus matrix quences are included. The numbers refer to the nucleotide position relative to the *last* nucleotide in the principal initial exon (apoL-I: 21,515 in Z82215; apoL-II: 7,737 in Z82215; apoL-III: 145,177 in Z95114; apoL-IV: 189,134 in Z95114). Shading indicates positional similarities between the four promoter sequences.

#483, a degenerate primer in the last exon of each gene. **Figure 7** shows the results of an RT-PCR reaction using these primers. Digestion with gene-specific restriction enzymes reveals the origin of each of the six bands seen. BclI cuts only apoL-I transcripts, SpeI only apoL-II, and HaeII only apoL-III. The uncommon apoL-I transcript containing exon 2 is not detected using this technique. Only one band for apoL-II was seen (1108 bp). The sizes of the four apoL-III bands correspond with the four transcripts found by sequencing (Fig. 3). The L-IIIa band (Fig. 7) corresponds to exons 1, 2–3, 4, 5 and migrates to a position predicted to be 1273 bp. L-IIIb corresponds to exons 1, 2, 3, 4, 5 (1157 bp), L-IIIc to exons 1, 3, 4, 5 (1044 bp), and L-IIId to exons 1, 4, 5 (947 bp). The lung pattern was similar to, though more intense than, that seen with spleen (not shown). The pattern with pancreas was similar to that found with placenta, liver, prostate, and small intestine, though in these four tissues the bands were less intense. The relative abundance of the apoL-III transcripts is in the order L-IIIc > L-IIId > L-IIIa = L-IIIb. The GenBank accession numbers of the four common apoL-III transcripts are AY014902–AY014905.

The two alternative splicing variants of apoL-I (Fig. 3) give rise to two proteins that differ only in the length of their signal peptide (Fig. 2). The more common splicing pattern that excludes exon 2 gives rise to 27-residue signal peptide. The rare variant predicts one with 43 residues.

Quantitative RT-PCR and multi-tissue dot blot

To measure, quantitatively, the total amount of mRNA for each gene in a number of different tissues, we decided to use degenerate primers and determine simultaneously the abundance for apoL-I, apoL-II, and apoL-III, because of the close homology between these three genes. The primers were designed to the terminal exon in each case, so the results reflect the sum of all splice variants. At this stage we had not detected transcripts for apoL-IV, but the primers used do not amplify mRNA from this gene. **Figure 8** shows a graph of T type RT-PCR results that establishes the conditions for the quantitative competitive assay. The optimum number of cycles was determined to be 24 from the C-type RT-PCR control assays. The RT-PCR results show that apoL-I is expressed at its highest level in lung (as is apoL-II and apoL-III). Noted that the values in Fig. 8 are per ng of total RNA. For the dot blot assay (Fig. 6), which analyzes far more tissue and is not as quantitative, loading of RNA was normalized for the expression of eight housekeeping genes. This is probably the reason for the different order of expression of apoL-I. Another complicating factor may be cross-hybridization of the apoL-I probe used in the dot blot with other apoL gene transcripts, including apoL-IV. In line with northern blot analysis in our previous study, pancreatic expression of apoL-I was high (Fig. 8).

DISCUSSION

The discovery of three additional genes for members of the apoL family now reveals a family of four genes, closely arrayed on chromosome 22, that have an unusually high level of homology. The presence of splice variants allows for the elaboration of seven gene products in this family. There is considerable redundancy of splicing mechanisms with a total of 12 transcripts detected. However, the 38-kDa variant of apoL-I that is found in plasma (9) is not the result of splice variation, but is produced posttranslationally. The observation that alternate 5′ sequences are involved in the transcription of identical messages raises the question as to whether they may exert differential effects on transcription. Similarly, the finding of alternative

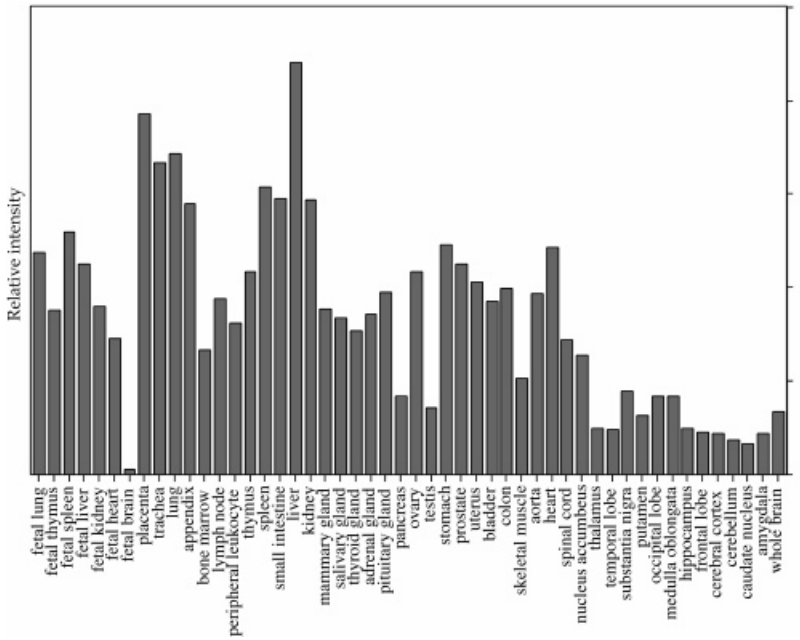

Fig. 6. Histogram showing the relative intensities obtained by scanning a dot blot containing RNA from 50 human tissues hybridized with an apoL-I probe.

lengths of signal peptides for apoL-I may have biological significance. For the case of apoL-IV, alternative splicing produces a form with a probable signal peptide and one without, suggesting that the gene product exists in both intracellular and extracellular locations. This phenomenon has been described for CETP (20, 21). Of the apoL-III transcripts, only apoL-IIIα/d predicts a protein with a possible signal peptide.

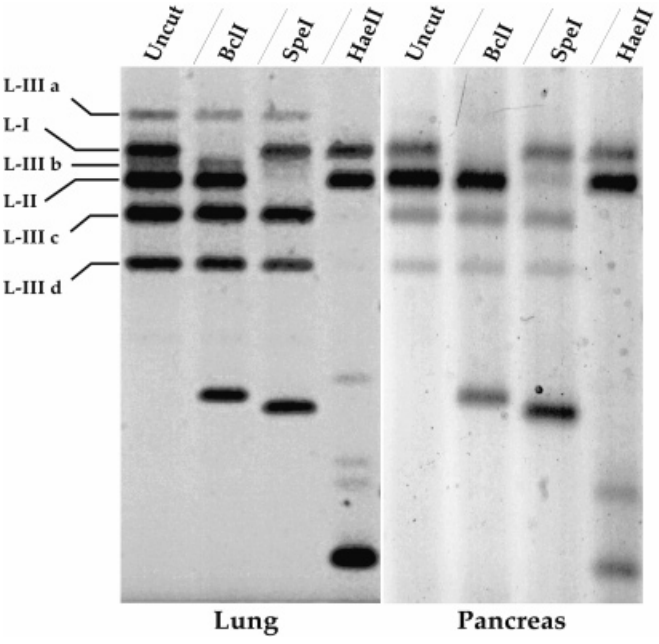

Fig. 7. Determination of the relative amounts of the splice variants in lung and pancreas. RT-PCR products were digested separately with restriction enzymes specific for apoL-I, apoL-II, or apoL-III (BclI, SpeI, and HaeII, respectively). The apoL-III splice variants (see Fig. 3) are (a) exons 1, 2–3, 4, 5; (b) exons 1, 2, 3, 4, 5; (c) exons 1, 3, 4, 5; and (d) exons 1, 4, 5.

The presence of classical amphipathic helices in all of the predicted gene products suggests that they could be involved in interaction with lipid, in the case of the selected members (apoL-I and apoL-IVb) with plasma lipoproteins. In the case of the gene products, apoL-II, apoL-III, and apoL-IVa that are likely to remain within cells, roles in the movement of lipids in the cytoplasm or binding of lipids to organelles may exist. The relatively high transcription rate for three of the members of this family in lung suggests that they could have roles in the intracellular transport and in secretion of the pulmonary surfactant. The presence of putative sterol response elements in all four genes suggests roles related to lipid synthesis or transport. Our recent observation (10) that the content of apoL-I in plasma is related to the level of triglyceride further suggests a role in the secretion or transport of lipids.

The observation that the transcription rate in endothelium for at least one of the members of this family (apoL-III) is increased 10-fold or more during exposure to TNFα (16) suggests that this protein may have a function in the inflammatory response. This response could represent the differential effect of the redundant promoters that we have found in apoL-III gene.

ApoL-I is known to be present in HDL at a concentration comparable to some other HDL apolipoproteins. Its structure suggests that it could function as a catalyst for lipid transfer. The presence of likely signal peptides in some predicted gene products of two other genes in this family suggests that, at least under some circumstances, they too will appear in plasma, in association with lipoproteins. The finding of a relationship of apoL-III to inflammation mediated by TNFα may reflect an extension of the hepatic lipoprotein response to sepsis (22). Both this response and possible functions in lipid transport or metabolism may ultimately reveal roles for these proteins in

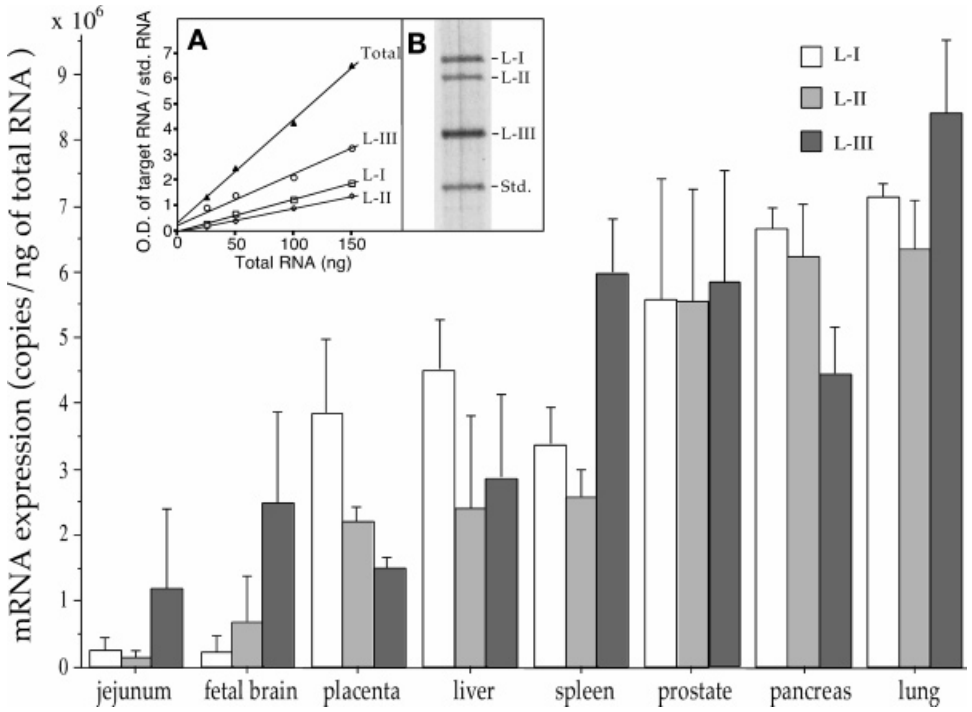

Fig. 8. Histogram showing mRNA expression in eight human tissues as determined by quantitative RT-PCR. Inset A shows the results from the T-type RT-PCR used to standardize the assay. Here, a fixed amount of standard was used with varying amounts of total RNA. Inset B shows a gel from a typical reaction.

atherosclerosis and in disorders involving pulmonary surfactant activity.

This work was supported by National Institutes of Health Grant HL 31210; by the American Heart Association, Western States Affiliates Grant-in-Aid 98-212; by a gift from the Joseph Drown Foundation; and by gifts from Donald Yellon and Donald and Susan Schleicher.

Manuscript received 3 August 2000 and in revised form 1 December 2000.

REFERENCES

1. Gordon, T., W. P. Cotelli, M. C. Hjortland, W. B. Kannel, and T. R. Dawber. 1977. High density lipoprotein as a protective factor against coronary heart disease: the Framingham Study. *Am. J. Med.* **66:** 707–714.
2. Miller, G. S., and N. F. Miller. 1975. Plasma high density lipoprotein and development of ischemic heart disease. *Lancet.* **1:** 16–19.
3. Bruce, C., R. A. Chouinard, Jr., and A. R. Tall. 1998. Plasma lipid transfer proteins, high-density lipoproteins, and reverse cholesterol transport. *Ann. Rev. Nutr.* **18:** 297–330.
4. Kunitake, S. T., M. Jarvis, R. L. Hamilton, and J. P. Kane. 1992. Binding of transition metals by apolipoprotein A-I-containing plasma lipoproteins: inhibition of oxidation of low density lipoproteins. *Proc. Natl. Acad. Sci. USA.* **89:** 6993–6997.
5. Parthasarathy, S., J. Barnett, and L. G. Fong. 1990. High-density lipoprotein inhibits the oxidative modification of low-density lipoprotein. *Biochim. Biophys. Acta.* **1044:** 275–283.
6. James, R. W., D. Hochstrasser, J-D. Tissot, M. Funk, R. Appel, F. Barja, C. Pellegrini, A. F. Muller, and D. Pometta. 1988. Protein heterogeneity of lipoprotein particles containing apolipoprotein A-I without apolipoprotein A-II and apolipoprotein A-I with apolipoprotien A-II isolated from human plasma. *J. Lipid Res.* **29:** 1557–1571.
7. Kunitake, S. T., C. T. Carilli, K. Lau, A. A. Protter, J. Naya-Vigne, and J. P. Kane. 1994. Identification of proteins associated with apolipoprotein A-I-containing lipoproteins purified by selected-affinity immunosorption. *Biochemistry.* **33:** 1988–1993.
8. McVicar, J. P., S. T. Kunitake, R. L. Hamilton, and J. P. Kane. 1984. Characteristics of human lipoproteins isolated by selected-affinity immunosorption of apolipoprotein A-I (high density lipoproteins). *Proc. Natl. Acad. Sci. USA.* **81:** 1356–1360.
9. Duchateau, P. N., C. R. Pullinger, R. E. Orellana, S. T. Kunitake, J. Naya-Vigne, P. M. O'Connor, M. J. Malloy, and J. P. Kane. 1997. Apolipoprotein L, a new human high density lipoprotein apolipoprotein expressed by the pancreas. Identification, cloning, characterization, and plasma distribution of apolipoprotein L. *J. Biol. Chem.* **272:** 25576–25582.
10. Duchateau, P. N., I. Movsesyan, S. A. Schoenhaus, S. J. Spencer, S. Mark, R. B. Jaffe, R. F. Redberg, J. P. Kane, and M. J. Malloy. 2000. Plasma apolipoprotein L levels are influenced by plasma triglyceride levels in normolipidemic, hyperlipidemic and NIDDM subjects. *J. Lipid Res.* **41:** 1231–1236**.**
11. Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10:** 1–6.
12. Zhang, J., M. Desai, S. E. Ozanne, C. Doherty, C. N. Hales, and C. D. Byrne. 1997. Two variants of quantitative reverse transcriptase PCR used to show differential expression of alpha-, beta- and gamma-fibrinogen genes in rat liver lobes. *Biochem. J.* **321:** 769–775.
13. Schug, J., and C. G. Overton. 1997. Technical Report (CBIL-TR-1997-1001-v0.0) of the Computational Biology and Informatics Laboratory; School of Medicine, University of Pennsylvania.
14. Quandt, K., K. Frech, H. Karas, E. Wingender, and T. Werner. 1995. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23:** 4878–4884.
15. Dunham, I., N. Shimizu, B. A. Roe, S. Chissoe, A. R. Hunt, J. E. Collins, R. Bruskiewich, D. M. Beare, M. Clamp, L. J. Smink, R. Ainscough, J. P. Almeida, A. Babbage, C. Bagguley, J. Bailey, K. Barlow, K. N. Bates, O. Beasley, C. P. Bird, S. Blakey, A. M. Bridgeman, D. Buck, J. Burgess, W. D. Burrill, K. P. O'Brien, et al. 1999. The DNA sequence of human chromosome 22. *Nature.* **402:** 489–495.
16. Horrevoets, A. J., R. D. Fontijn, A. J. van Zonneveld, C. J. de Vries, J. W. ten Cate, and H. Pannekoek. 1999. Vascular endothelial genes

that are responsive to tumor necrosis factor-alpha in vitro are expressed in atherosclerotic lesions, including inhibitor of apoptosis protein-1, stannin, and two novel genes. *Blood.* **93:** 3418–3431.
17. Kozak, M. 1996. Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome.* **7:** 563–574.
18. Smale, S. T. 1997. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta.* **1351:** 73–88.
19. Semenza, G. L. 1998. Transcription Factors and Human Disease. Oxford Monographs on Medical Genetics. Oxford University Press, New York and Oxford.
20. Inazu, A., E. M. Quinet, S. Wang, M. L. Brown, S. Stevenson, M. L. Barr, P. Moulin, and A. R. Tall. 1992. Alternative splicing of the mRNA encoding the human cholesteryl ester transfer protein. *Biochemistry.* **31:** 2352–2358.
21. Quinet, E., T. P. Yang, C. Marinos, and A. Tall. 1993. Inhibition of the cellular secretion of cholesteryl ester transfer protein by a variant protein formed by alternative splicing of mRNA. *J. Biol. Chem.* **268:** 16891–16894.
22. Feingold, K. R., J. L. Funk, A. H. Moser, J. K. Shigenaga, J. H. Rapp, and C. Grunfeld. 1995. Role for circulating lipoproteins in protection from endotoxin toxicity. *Infect. Immun.* **63:** 2041–2046.