An Apolipoprotein Influencing Triglycerides in Humans and Mice Revealed by Comparative Sequencing

Len A. Pennacchio,[1] Michael Olivier,[2*] Jaroslav A. Hubacek,[3] Jonathan C. Cohen,[3] David R. Cox,[2] Jean-Charles Fruchart,[4] Ronald M. Krauss,[1] Edward M. Rubin[1†]

Comparison of genomic DNA sequences from human and mouse revealed a new apolipoprotein (APO) gene (*APOAV*) located proximal to the well-characterized *APOAI/CIII/AIV* gene cluster on human 11q23. Mice expressing a human *APOAV* transgene showed a decrease in plasma triglyceride concentrations to one-third of those in control mice; conversely, knockout mice lacking *Apoav* had four times as much plasma triglycerides as controls. In humans, single nucleotide polymorphisms (SNPs) across the *APOAV* locus were found to be significantly associated with plasma triglyceride levels in two independent studies. These findings indicate that APOAV is an important determinant of plasma triglyceride levels, a major risk factor for coronary artery disease.

Plasma lipid levels are a major determinant of cardiovascular disease susceptibility (*1*). Members of the apolipoprotein gene family have been shown to play a significant role in determining an organism's lipid profile, with alterations in the level or structure of these molecules leading to abnormal lipid levels and atherosclerosis susceptibility (*2–6*). The apolipoprotein gene cluster (*APOAI/CIII/AIV*) on human 11q23 (*7*) is a well-studied region known to influence plasma lipid parameters in humans. Defined mutations in this cluster dramatically affect plasma lipid profiles in both humans and mice (*2*, *8–12*), and common sequence polymorphisms in this interval have been implicated as contributing to severe hypertriglyceridemia (*13–16*).

Genome sequencing efforts produced finished sequence throughout the human *APOAI/CIII/AIV* region, thereby providing a resource to better understand the genomic structure of this locus (*17*). To facilitate the identification of evolutionarily conserved sequences with potential function near this cluster, we determined the sequence of ~200 kilobase pairs (kbp) of orthologous mouse DNA and compared the mouse and human sequences (Fig. 1) (*18*). On the basis of extended interspecies sequence

[1]Genome Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, 975 California Avenue, Palo Alto, CA 94304, USA. [3]Center for Human Nutrition and McDermott, Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX 75390–9052, USA. [4]Department of Atherosclerosis–INSERM U545, Institut Pasteur de Lille, 1, rue du Professeur Calmette, 59019 Lille cedex, France and Faculté de Pharmacie, University of Lille, 59006 Lille cedex, France.

*Present address: Human and Molecular Genetics Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA.
†To whom correspondence should be addressed. E-mail: EMRubin@lbl.gov

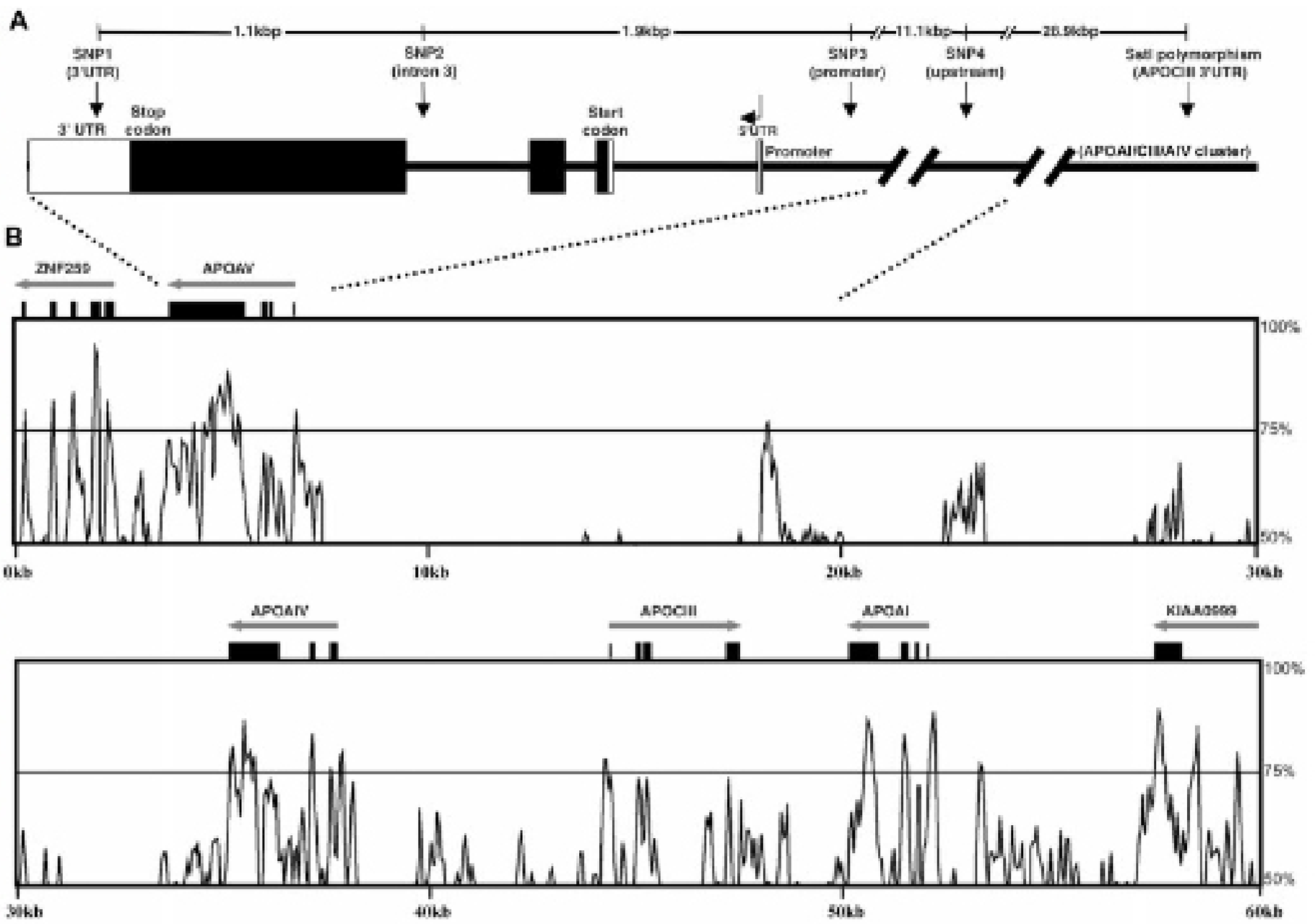

Fig. 1. Human and mouse comparative sequence analysis of the *APOAI/CIII/AIV* gene cluster. (A) A schematic of the genomic organization of human *APOAV* and the relative SNP positions (arrows). *APOAV* exons are shown with solid boxes and the distance between each SNP is indicated above the line. The predicted transcription start site is depicted by a bent arrow and the relative position of the promoter and the start and stop codons are shown. (B) In each panel, 30 kbp of contiguous human sequence is illustrated horizontally. Above each panel arrows correspond to known genes and their orientation with each exon depicted by a box (gene names are indicated above each arrow). The VISTA graphical plot displays the level of homology between human and the orthologous mouse sequence. Human sequence is represented on the *x* axis and the percent similarity with the mouse sequence is plotted on the *y* axis (ranging from 50 to 100% identity).

conservation about 30 kbp proximal to the *APOAI/CIII/AIV* gene cluster, we identified a genomic interval that contained a putative apolipoprotein-like gene (*APOAV*) (Fig. 1B). The presence of publicly available mouse expressed-sequence tags (ESTs) matching the mouse genomic sequence suggested that the interval was transcribed. The annotation of mouse ESTs on the mouse genomic sequence identified four exons containing a 1107–base pair (bp) open reading frame. The predicted 368–amino acid sequence showed significant homology to various known apolipoproteins, with the strongest similarity to mouse Apoaiv (24% identity and 49% similarity). Examination of the orthologous human genomic sequence indicated a genomic structure similar to the mouse region and predicted an open reading frame encoding a 366–amino acid protein with high sequence homology to mouse Apoav (71% identity and 78% similarity), as well as human APOAIV (27% identity, 48% similarity). Protein structure analyses predicted several amphipathic helical domains and an NH_2-terminal signal peptide in both human and mouse APOAV, characteristic features of lipid-binding apolipoproteins (*19*, *20*). To determine the expression pattern of *APOAV*, we hybridized Northern blots containing mRNA from several different human and mouse tissues with *APOAV* cDNA probes from human and mouse, respectively (Fig. 2, A and B). Transcripts about 1.3 and 1.9 kilobases (kb) in length were identified predominantly in liver tissue from both species. The full-length sequences of mouse cDNAs indicated the two transcripts in mice are likely the result of alternative polyadenylation (*21*, *22*).

To assess the function of APOAV, we generated mice overexpressing human *APOAV* as well as mice lacking *Apoav*, through standard mouse transgenic and gene knockout technologies (Fig. 2, C to E) (*23*–*25*). Upon comparing these two groups, we observed dramatic, but opposite effects on plasma triglyceride levels (*26*). Human *APOAV* transgenic mice were created by using a 26-kbp Xho I fragment predicted to contain only human *APOAV*, and this genomic transgene was expressed in liver, as is the endogenous gene (Fig. 2C). These transgenic mice had levels of plasma triglyceride that were about one-third of those of control littermates [0.32 ± 0.11 (S.D.) mg/ml versus 0.90 ± 0.29; *t* test, $P < 0.0001$] (Fig. 3A). Similar data were obtained from a second independent founder line (*27*). *Apoav* knockout mice were generated by deleting the three exons predicted to encode Apoav (Fig. 2D). Despite the lack of *Apoav* transcript (Fig. 2E), mice homozygous for the deletion were born at the expected Mendelian rate and appeared normal. In contrast to the decreased triglyceride levels noted in *APOAV* transgenics, *Apoav* knockout mice had about four times as much plasma triglyceride as their wild-type littermates [1.53 ± 0.77 (SD) mg/ml versus 0.37 ± 0.12; *t* test, $P < 0.001$) (Fig. 3B). Characterization of lipoprotein particles by fast-protein liquid chromatography (FPLC) and gradient-gel electrophoresis (GGE) revealed that levels of very low density lipoprotein (VLDL) particles were increased in the homozygous knockout mice and decreased in the transgenic mice com-

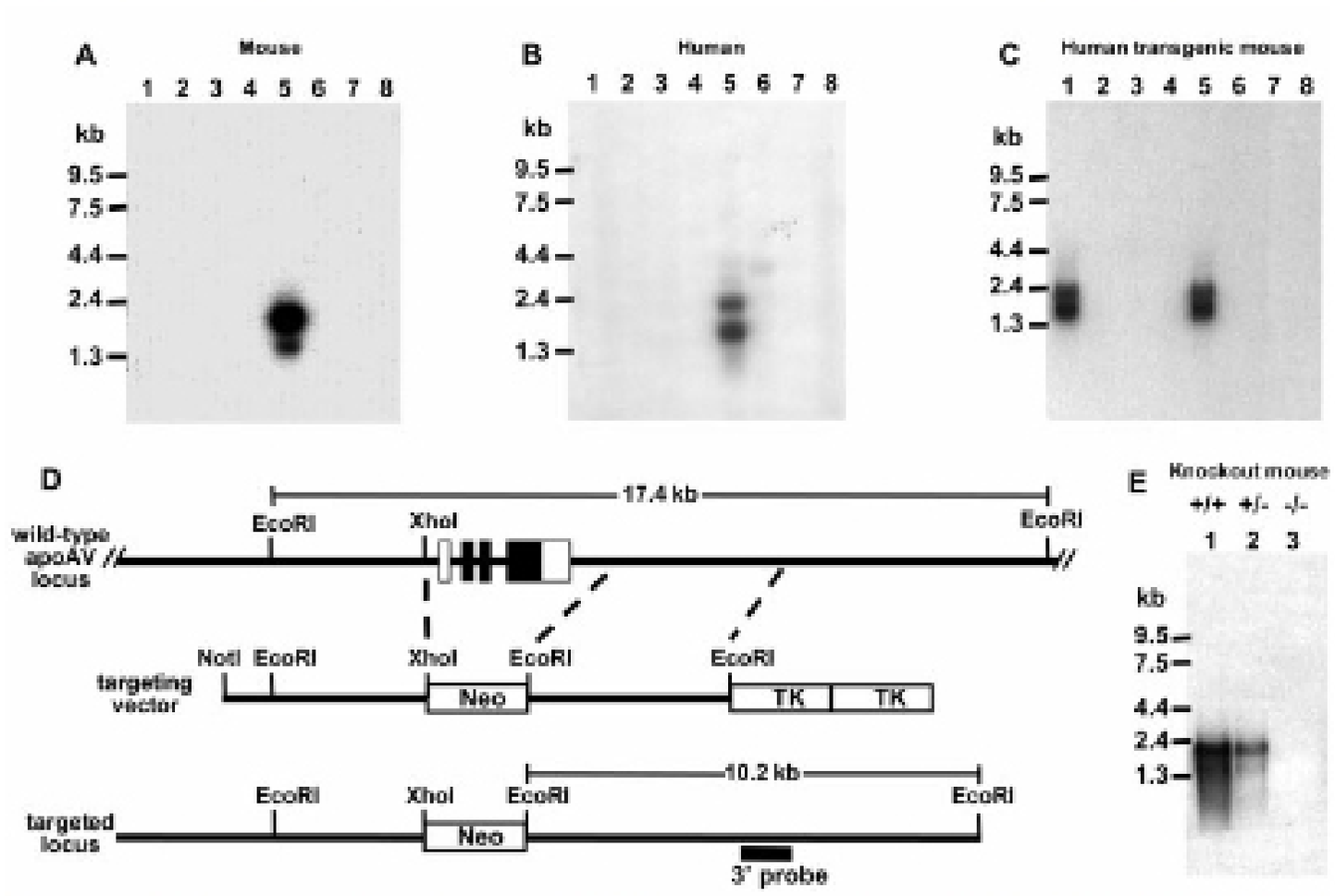

Fig. 2. *APOAV* expression in humans and wild-type, transgenic, and knockout mice (*52*). (**A**) A mouse *Apoav* cDNA probe was hybridized to a multi-tissue RNA blot from wild-type mice. Each lane contained one of eight mouse tissues (Clontech, Palo Alto, California): 1, heart; 2, brain; 3, spleen; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; or 8, testis. (**B**) A human *APOAV* cDNA probe was hybridized to an RNA blot containing eight human tissues (Clontech,): 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; or 8, pancreas. (**C**) A human-specific *APOAV* cDNA probe was hybridized to total RNA blots from human *APOAV* transgenic mice and controls. Lane assignments are as follows: 1 and 5, transgenic liver; 2 and 6, transgenic intestine; 3 and 7, wild-type liver; and 4 and 8, wild-type intestine. (**D**) A diagram of the targeting construct used to generate *Apoav*-deficient mice. Homology arms were designed to delete the coding exons of the gene (depicted by black boxes). Properly targeted embryonic stem cells were identified by using an external 3′ probe, which detects a 17-kb Eco RI fragment wild-type allele and a 10-kb Eco RI fragment upon targeting (*27*). (**E**) Northern blot analysis of various genotype mice following the *Apoav* targeting event. Each lane contains liver mRNA from a wild-type (lane 1), heterozygous (lane 2), and homozygous knockout mouse (lane 3). To confirm similar amounts of RNA were loaded per lane, duplicate gels were examined by ethidium bromide staining.

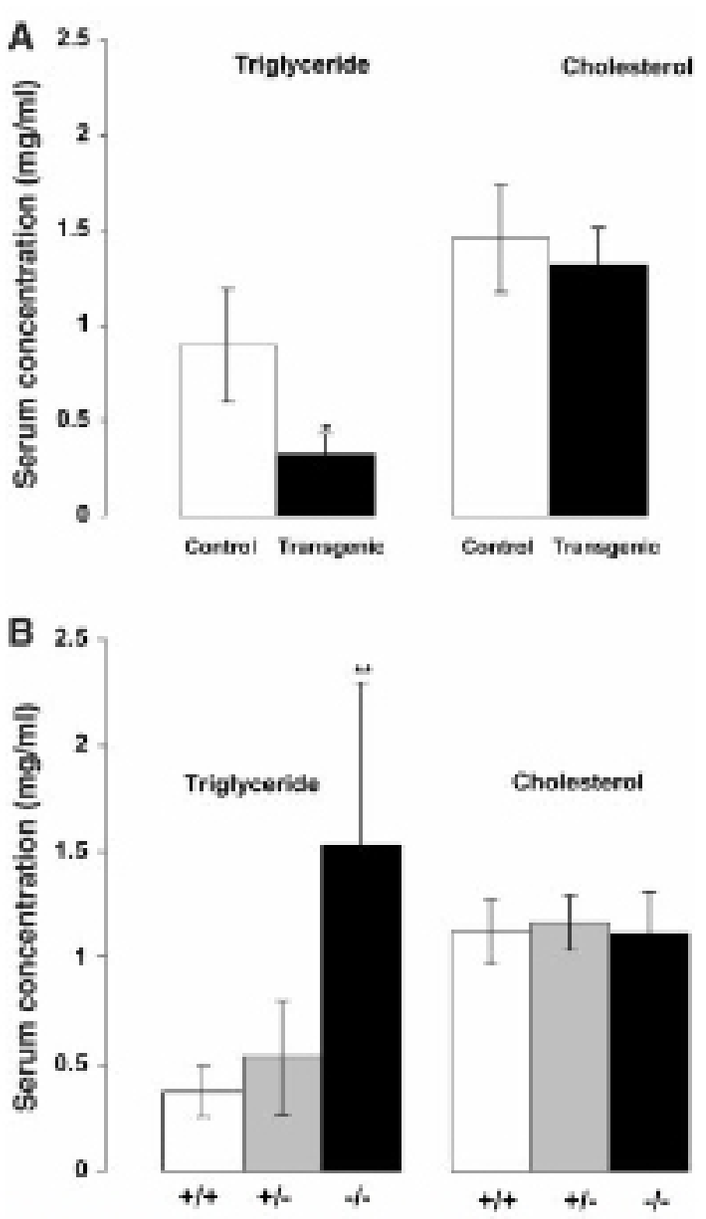

Fig. 3. Plasma triglyceride and cholesterol levels for *APOAV* transgenic and knockout mice on standard chow diet. (**A**) Human *APOAV* transgenic mice compared with isogenic FVB strain control littermates ($n = 48$ for transgenics; $n = 44$ for controls; Student's *t* test $^*P < 0.0001$ for transgenic versus control). (**B**) Mice lacking *Apoav* compared with mixed 129Sv/C57BL6 strain controls littermates ($n = 13$ for wild-type, +/+; $n = 22$ for heterozygotes, +/–; $n = 10$ for homozygous knockouts, –/–; Student's *t* test, $^{**}P < 0.001$ for wild-type versus knockout). Error bars correspond to the standard deviation for both graphs. No differences were found in HDL-cholesterol levels in transgenic or knockout mice compared with controls (*27*).

pared with controls (*28*). VLDL levels in a heterozygous knockout mouse were intermediate between the homozygous knockout and control mouse. VLDL peak particle size as assessed by GGE and FPLC peak elution volume was similar in all animals (*29*). Analysis of FPLC elution volumes demonstrated mouse Apoav immunoreactivity in VLDL and HLD fractions.

The observed changes in plasma triglyceride levels in *Apoav* knockout and transgenic mice were directly opposite those previously reported in *Apociii* knockout and transgenic mice (*9*, *10*). The *Apoav* knockouts in our study displayed about a 400% increase in plasma triglycerides compared with the 30% decrease noted in *Apociii* knockouts, whereas *APOAV* transgenics showed decreased triglyceride levels compared with the increase reported in *APOCIII* transgenics. Accordingly, we examined the effect of altered *APOAV* expression on Apociii levels. Differences were found in apociii protein but not transcript levels in both *APOAV* transgenic and knockout animals; Apociii levels were increased ~90% in *Apoav* knockouts and decreased ~40% in *APOAV* transgenics. Because alterations in *APOAV* expression lead to changes in Apociii protein levels, the effect on triglycerides we observed may be mediated through Apociii. The fact that *APOAV* transgenic mice have one-half the triglycerides that the previously described *Apociii* knockout mice have indicates (*10*) that changes in Apociii alone cannot explain the entire effect of *APOAV*. In addition to *APOCIII*, the overexpression of several human apolipoprotein transgenes has been shown to increase triglyceride levels in mice (*8*, *9*, *30*–*33*), whereas only the *APOAV* transgene leads to decreased triglycerides, suggesting another mechanism behind this effect.

The observation of significant lipid abnormalities in mice overexpressing and lacking *Apoav* led us to explore the relationship between DNA sequence polymorphisms in the gene and plasma lipid levels in humans. To serve as genetic markers for association studies, we identified single nucleotide polymorphisms (SNPs) across and surrounding the human *APOAV* locus (*34*) (Fig. 1A). Four markers with relatively high minor allele frequencies (>8%) were obtained. Three of the SNPs were separated by 3 kbp within *APOAV* (SNP1 to SNP3); the fourth SNP (SNP4) was located ~11 kbp upstream of the gene (Fig. 1A). These markers were scored in about 500 random unrelated normolipidemic Caucasian individuals who had been phenotyped for numerous lipid parameters before and after consumption of high- and low-fat diets (*35*). We found significant associations between both plasma triglyceride levels and VLDL mass and the three neighboring SNPs (SNPs 1 to 3) within *APOAV* but not with the distant upstream SNP4 (Figs. 1A and 4A). Specifically, the minor allele of each of these SNPs (SNPs 1 to 3) was associated with higher triglyceride levels independent of diet. Independent analysis of each of these SNPs (SNPs 1 to 3) revealed plasma triglyceride levels were 20 to 30% higher in individuals having one minor allele compared with individuals homozygous for the major allele (Fig. 4A). Analysis of SNP allele frequencies in more than

A

Associations of *APOAV* genotypes with plasma lipid parameters (mg/dL ± SEM)													
Genotype	**SNP1 1,1**	**SNP1 1,2**	**p**	**SNP2 1,1**	**SNP2 1,2**	**p**	**SNP3 1,1**	**SNP3 1,2**	**p**	**SNP4 1,1**	**SNP4 1,2**	**SNP4 2,2**	**p**
n	364	72		358	67		354	72		172	213	55	
Triglyceride													
Random diet	119.3±3.8	144.4 ± 11.3	0.02	119.8±3.8	150.7±12.0	0.003	116.5±3.6	151.7±11.2	0.0002	126.8±5.8	124.6±5.6	117.4±9.8	0.7
High fat diet	98.6±3.2	129.8±12.1	0.0006	98.4±3.2	134.9±13.0	<0.0001	96.5±2.9	132.9±12.1	<0.0001	107.9±5.4	110.2±7.7	91.7±5.6	0.39
Low fat diet	127.0±4.4	171.7±21.9	0.0015	128.5±4.6	175.9±23.4	0.0012	126.1±4.5	174.7±21.8	0.0006	138.7±7.4	137.6±9.1	131.5±13.3	0.92
VLDL mass													
High fat diet	73.9±3.3	108.0±11.8	0.0002	73.7±3.3	112.8±12.7	<0.0001	70.6±3.2	113.7±11.8	<0.0001	82.1±5.7	82.6±6.5	71.7±6.8	0.67
Low fat diet	116.5±4.9	152.9±16.8	0.0065	112.9±5.0	155.6±18.1	0.0071	116.0±5.0	155.2±16.9	0.0037	122.8±7.7	127.5±7.8	123.3±15.5	0.91
LDL-cholesterol													
High fat diet	122.0±1.8	122.5±3.9	0.90	122.7±1.8	122.2±4.1	0.91	123.5±1.9	122.7±4.2	0.86	121.8±2.5	121.4±2.4	126.2±4.6	0.65
Low fat diet	111.7±1.6	110.5±4.0	0.76	112.5±1.6	109.8±4.2	0.51	112.8±4.2	109.9±4.6	0.50	110.8±2.3	111.3±2.2	115.7±4.4	0.39
HDL-cholesterol													
High fat diet	48.6±0.7	46.8±1.5	0.26	48.4±0.7	47.7±1.5	0.66	48.7±0.7	47.7±1.6	0.86	48.1±0.9	48.3±0.9	49.2±1.9	0.83
Low fat diet	41.6±0.5	40.5±1.3	0.40	41.6±0.5	40.8±1.3	0.52	41.8±0.5	40.6±1.4	0.37	41.5±0.8	41.2±0.7	42.6±1.4	0.66

B

	SNP1	SNP2	SNP3	SNP4
SNP1		0.870	1.000	0.219
SNP2	0.870		1.000	0.328
SNP3	1.000	1.000		0.308
SNP4	0.219	0.328	0.308	

C

Group	n	BMI	Age	TG	SNP3 1,1	SNP3 1,2	SNP3 2,2	p value
Total TG>90%	161	29.0±4.8	49.2±16.3	340.6±249.4	125 (77.8)	35 (21.7)	1 (0.8)	>0.0001
Total TG<10%	298	25.7±4.1	48.9±13.7	54.8±13.4	278 (93.3)	20 (6.7)	0	
Male TG>90%	97	28.9±3.7	50.1±16.6	338.4±155.3	67 (69.1)	29 (29.9)	1 (1.0)	>0.0001
Male TG<10%	192	26.5±3.8	54.4±15.6	57.4±12.3	184 (95.8)	8 (4.2)	0	
Female TG>90%	64	29.3±6.0	47.8±15.8	347.1±349.6	58 (90.6)	6 (9.4)	0	0.69
Female TG<10%	106	24.1±4.2	42.4±12.8	50.0±13.8	94 (88.7)	12 (11.3)	0	

Fig. 4. Human *APOAV* polymorphisms and lipid association data. (**A**) Plasma lipid concentrations for a given genotype for 4 neighboring SNPs (SNPs 1 to 4). Individuals ($n = 501$) were genotyped, and the number of successfully scored individuals is indicated. Notation: 1,1 is homozygous for the major allele; 1,2 is heterozygous for the major and minor alleles. Three individuals were homozygous for the SNP3 minor allele and had a mean plasma triglyceride level of 210 ± 155 mg/dL. Because of the small number of individuals, these data were excluded from the analysis. All sites were found to be in Hardy-Weinberg equilibrium (*53*). The minor allele frequency for each SNP (SNPs 1 to 4) was 9.1, 8.4, 9.2 and 36.3%, respectively. Not shown is the lack of association between each of the four SNPs and IDL-, LDL-, HDL-mass, ApoAI, and ApoB levels [$P > 0.05$, (*54*)]. (**B**) Pair-wise measure of linkage disequilibrium (ID'I) was calculated for all combinations of SNPs as previously described (*55*). A ID'I value of 1 indicates complete linkage disequilibrium between two markers. (**C**) A summary of SNP3 genotyping data from an independent set of individuals stratified based on triglyceride levels. *P* values were determined by chi-square analysis. BMI, body mass index; TG, plasma triglyceride level (mg/dl ± SD). Similar analysis stratifying the original population did result in statistically significant differences in the genotype distribution when we used a similar analysis ($P = 0.044$).