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).
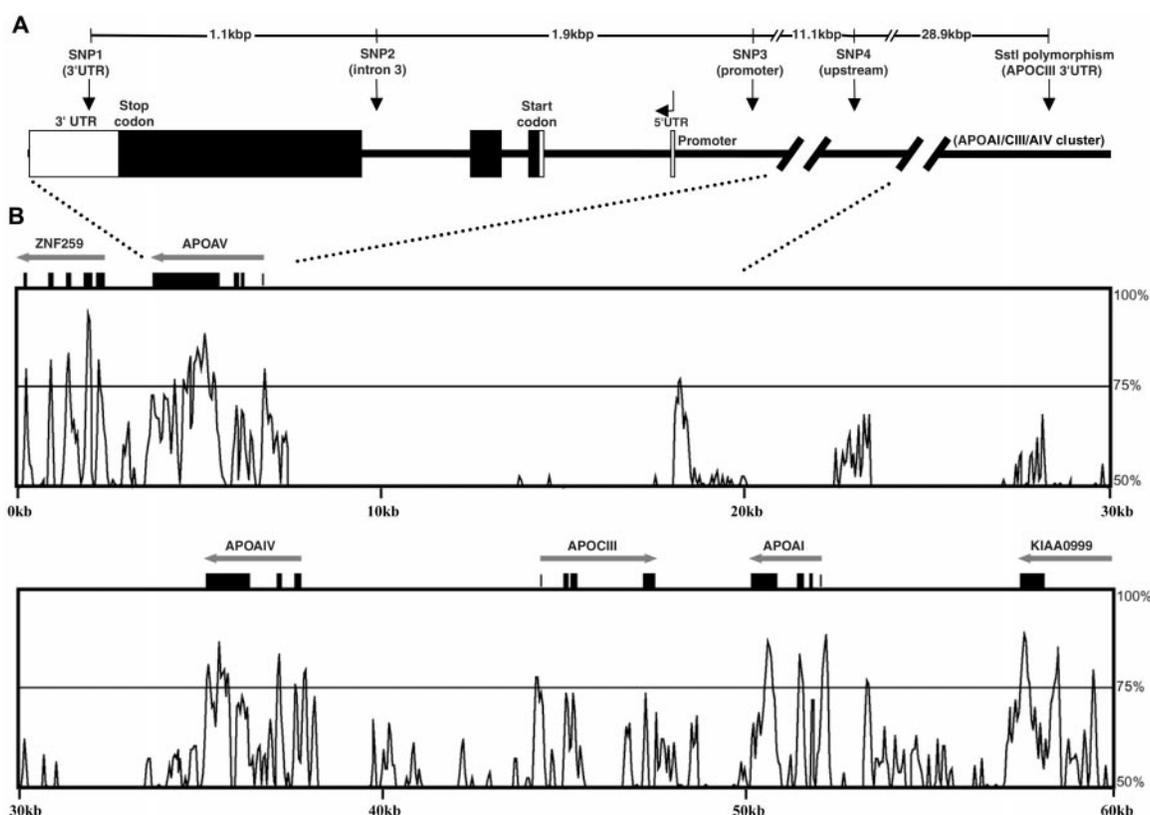
conservation about 30 kbp proximal to the *APOAI/CIII/AIV* gene cluster, we identified a genomic interval that contained a putative apolipoprotein-like gene (*APOAV*) (Fig. 1B). The presence of publicly available mouse expressed-sequence tags (ESTs) matching the mouse genomic sequence suggested that 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-
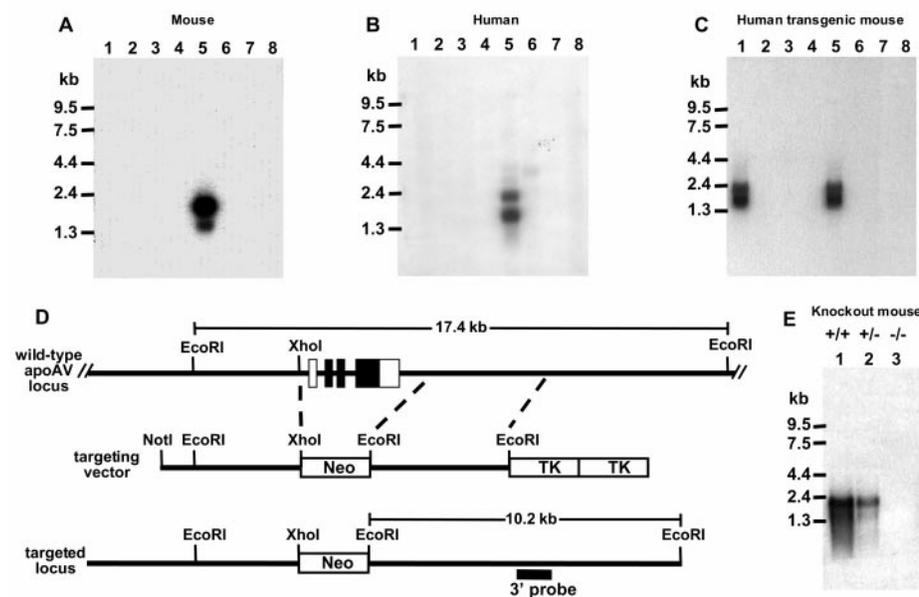
Fig. 2. *APOAV* expression in humans and wild-type, transgenic, and knockout mice (*52*). (**A**) A mouse *Apoav* cDNA probe was hybridized to a multi-tissue RNA blot from wild-type mice. Each lane contained one of eight mouse tissues (Clontech, Palo Alto, California): 1, heart; 2, brain; 3, spleen; 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.
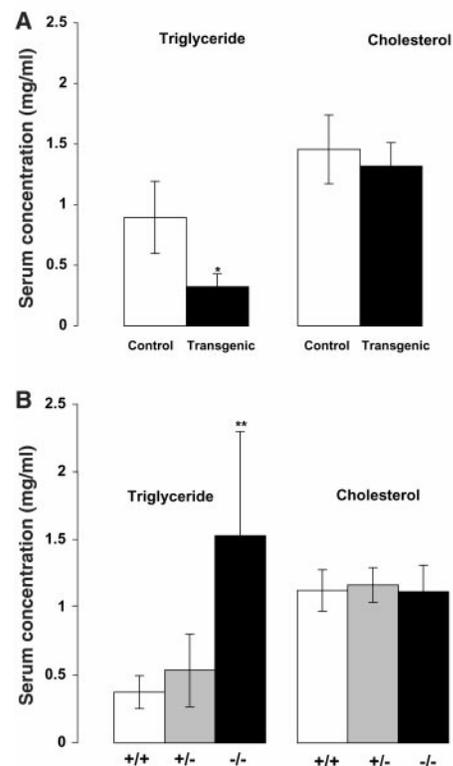
Fig. 3. Plasma triglyceride and cholesterol levels for *APOAV* transgenic and knockout mice on standard chow diet. (**A**) Human *APOAV* transgenic mice compared with isogenic FVB strain control littermates (*n* = 48 for transgenics; *n* = 44 for controls; Student's *t* test **P* < 0.0001 for transgenic 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 lev-

els 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	0.59
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±4.6	49.2±16.3	340.6±249.4	125 (77.6)	35 (21.7)	1 (0.6)	>0.0001
Total TG<10%	298	25.7±4.1	49.9±15.7	54.6±13.4	278 (93.3)	20 (6.7)	0	
Male TG>90%	97	28.9±3.7	50.1±16.6	336.4±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 (|D′|) was calculated for all combinations of SNPs as previously described (*55*). A |D′| 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).

1000 chromosomes revealed that the three neighboring SNPs (SNPs 1 to 3) in *APOAV* were in significant linkage disequilibrium that does not extend to SNP4 (located ~11kb upstream of *APOAV*) (Fig. 4B). This finding supports the existence of a common haplotype in the *APOAV* region influencing plasma triglyceride levels (Figs. 1A and 4B). Furthermore, studies in this population found no significant association of triglyceride levels with an Sst I polymorphism in *APOCIII* (located ~40 kbp upstream of *APOAV*) (Fig. 1A) which has been previously associated with hypertriglyceridemia (*15, 16, 28, 36*). This finding indicates that the *APOCIII* Sst I polymorphism is not a marker for the metabolic effect defined by the *APOAV* haplotype.

Genetic association studies have frequently proved difficult to reproduce. Therefore, we performed a second human association study with one SNP (SNP3) in an independently ascertained cohort using a different experimental design (*37*). SNP3 was chosen for genotyping in this study based on its strong association in our first study and its apparent complete linkage disequilibrium with the other two associated SNPs (SNPs 1 and 2) (Fig. 4, A and B). In the second study, we examined the allele frequencies for SNP3 in an unrelated group of Caucasians stratified according to plasma triglyceride levels (Fig. 4C). The two groups represented 161 individuals with triglyceride levels in the top 10th percentile and 298 individuals from the bottom 10th percentile. A significant overrepresentation of the heterozygous genotype was found in individuals with high compared with low plasma triglyceride levels (21.7% versus 6.7%, respectively), thereby validating the association of *APOAV* polymorphisms and triglyceride levels in a second cohort. When the cohort was stratified based on gender, an even more pronounced overrepresentation of the heterozygous genotype was found in males with high compared with low plasma triglyceride levels (29.9% versus 4.2%, respectively).

Despite the previous availability of sequence in the human *APOAI/CIII/AIV* genomic interval, we only recently were directed to a novel gene (*APOAV*) by comparison of human and mouse sequence, illustrating the power of comparative sequence analysis to prioritize potential functional regions of the genome. *APOAV* represents a fourth member of the clinically important apolipoprotein gene cluster on human 11q23. Our human and mouse data, both when taken independently and combined, indicate an important role for APOAV in plasma triglyceride homeostasis. Although previous data have associated the *APOCIII* locus with extremely high plasma triglyceride levels in humans, our study indicates that the *APOAV* genomic interval represents an independent influence on this important lipid parameter in the general population. These results suggest the possible use of *APOAV* polymorphisms as prognostic indicators for hypertriglyceridemia susceptibility and the focus on APOAV modulation as a potential strategy to reduce this known cardiovascular disease risk factor.

References and Notes

1. T. R. Dawber, G. F. Meadors, F. E. J. Moore, *Am. J. Publ. Health* **41**, 279 (1951).
2. N. Duverger *et al.*, *Science* **273**, 966 (1996).
3. R. M. Krauss, *Am. J. Med.* **105**, 58S (1998).
4. E. M. Rubin, A. Tall, *Nature* **407**, 265 (2000).
5. A. J. Lusis, *Nature* **407**, 233 (2000).
6. A. S. Plump *et al.*, *Cell* **71**, 343 (1992).
7. G. A. Bruns, S. K. Karathanasis, J. L. Breslow, *Arteriosclerosis* **4**, 97 (1984).
8. Y. Huang *et al.*, *J. Biol. Chem.* **273**, 26388 (1998).
9. Y. Ito, N. Azrolan, A. O'Connell, A. Walsh, J. L. Breslow, *Science* **249**, 790 (1990).
10. N. Maeda *et al.*, *J. Biol. Chem.* **269**, 23610 (1994).
11. L. Masucci-Magoulas *et al.*, *Science* **275**, 391 (1997).
12. E. M. Rubin, R. M. Krauss, E. A. Spangler, J. G. Verstuyft, S. M. Clift, *Nature* **353**, 265 (1991).
13. H. Allayee *et al.*, *Am. J. Hum. Genet.* **63**, 577 (1998).
14. A. P. Wojciechowski *et al.*, *Nature* **349**, 161 (1991).
15. M. R. Hayden *et al.*, *Am. J. Hum. Genet.* **40**, 421 (1987).
16. M. Dammerman, L. A. Sandkuijl, J. L. Halaas, W. Chung, J. L. Breslow, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4562 (1993).
17. Electronic homology searches with human *APOAI*, *APOCIII*, and *APOAIV* mRNA sequences with the BLAST algorithm (*38*) identified a genomic bacterial artificial chromosome (BAC) clone containing the complete *APOAI/CIII/AIV* gene cluster (GenBank accession number AC007707).
18. Orthologous mouse genomic DNA was isolated from a pooled BAC library by using the polymerase chain reaction (PCR) with mouse primers: apoAI-F1-5'-GAG-GATGTGGAGCTCTACCGC-3' and apoAI-R1-5'-CTGT-GTGCGCAGAGAGTCTACG-3' [RPCI-23, BACPAC Resources, Children's Hospital Oakland Research Institute; http://www.chori.org/bacpac/ (*39*)]. Positive clone RPCI-23-175F2 was identified, randomly sheared, subcloned and sequenced to about sixfold coverage (GenBank accession number AF401201) (*40, 41*). Human and mouse sequence comparisons were performed as previously described and are available at http://pga.lbl.gov (*42*).
19. A. M. Gotto Jr., H. J. Pownall, R. J. Havel, *Methods Enzymol.* **128**, 3 (1986).
20. Protein analyses were performed by using the Web-based PredictProtein package, COILS (*43*), and SignalP (*44*) (http://www.ch.embnet.org/software/COILS_form.html; http://www.embl-heidelberg.de/predictprotein/predictprotein.html; http://www.cbs.dtu.dk/services/SignalP).
21. J. Kawai *et al.*, *Nature* **409**, 685 (2001).
22. The mouse *Apoav* cDNA sequences are available under GenBank accession numbers AK004936 and AK004903.
23. K. A. Frazer, G. Narla, J. L. Zhang, E. M. Rubin, *Nature Genet.* **9**, 424 (1995).
24. C. Paszty *et al.*, *Nature Genet.* **11**, 33 (1995).
25. Restriction enzyme predictions for human genomic sequence (GenBank accession number AC007707) indicated that the entire human *APOAV* gene, but not neighboring genes, was contained within a 26-kbp Xho I DNA fragment (corresponding to about 1 to 27 kbp in Fig. 1B). BAC DNA corresponding to the clone sequenced from this region was prepared by standard alkaline lysis with a chromatography column (Qiagen, Valencia, CA), digested with the restriction enzyme Xho I and separated in 1% agarose by pulsed-field gel electrophoresis. The 26-kbp Xho I DNA fragment containing human *apoAV* was purified by using QiaEx II gel purification (Qiagen), adjusted to a final concentration of ~1 ng/ml and microinjected into fertilized FVB inbred mouse eggs by using standard procedures (*23*). Two founder transgenic mice were identified as determined by PCR (polymer chain reaction) amplification using primers hApoAV-intrn-F1-5'-CCCGCTGCAGTCCCCA-GAAT-3' and hApoAV- intrn-R1-5'-CAGGGTCGAG-GGCTCT TGTCCT-3'. Each founder line was expanded by breeding to isogenic FVB strain mice (The Jackson Laboratory, Bar Harbor, ME).

The targeting construct to delete mouse *Apoav* was built by using PCR products amplified from BAC-RPCI-23-175F2 DNA (BACPAC Resources, Children's Hospital Oakland Research Institute) (*18*). The first homology arm was PCR-amplified by using primers containing introduced 5' restriction sites for Xba I and Eco RI, respectively: mAV-Xba I-F1-5'-TGACTCTAGATACCC-TTGGTCCCATGTTCCAGAT-3' and mAV-Eco RI-R1-5'-CATTGAATTCGACAAGAGAAAGACGGGGCTCAAG-3'. The resulting 4.2-kbp PCR product was cloned into pXL-Topo (Invitrogen), DNA prepared by standard alkaline lysis (Qiagen) and digested with Eco RI according to the manufacturer's recommendations (New England Biolabs, Beverly, MA). A 4.2-kbp Eco RI fragment was gel-purified and cloned into the Eco RI site of the pPN2T vector to yield pPN2T-Arm1 (*24*). Clones were screened by PCR for inserts using the above-described primers, and positive clones were sequenced for proper orientation. The second homology arm was PCR-amplified with primers mAV-NotI-F4-5'-TATGACTGCGGCCGC-CACCAATCCCACATCTAAGCATCT-3', containing an introduced 5' Not I restriction site, and mAV-XhoI-R3-5'-GCTCGGTTCTGGGCACAGAGA-3'. The resulting 5.3-kbp PCR product containing an endogenous internal Xho I restriction site was digested with Not I and Xho I to yield a 5.1-kbp fragment, which was directionally cloned into the Xho I and Not I sites of the pPN2T-Arm1 vector to yield final vector pPN2T-apoAV-KO. Embryonic stem cells from the 129/SvJ strain (Incyte Genomics, Palo Alto, CA) were electroporated with 20 μg of the Not I linearized targeting construct and subsequently selected in 200 μg/ml G418 and 0.5 μg/ml 1-2'-deoxy-2'-fluoro-β-D-arabinofuranosyl-5-iodouracil for 8 days. Individual clones were isolated, expanded, and screened by Southern blot analysis. The external 3' probe was amplified by PCR with primers mApoAV-3'probe-F2-5'-CTTGAGGATGGGCATCAGCT-GTAT-3' and mApoAV-3'probe-R2-5'-GCTCACTAA-CAGCGCTCTTGCCT-3'. Targeted clones were injected into C57BL/6 blastocysts and chimeric males were bred to C57BL/6 females (Jackson Laboratory). Agouti offspring were tested for germ line transmission of the targeted allele by PCR with primers specific to the neomycin gene (NeoF1-5'-CTTTTTGTCAAGACCGAC-CTG-3' and NeoR1-5'-AATATCACGGGTAGCCAACGC-3') and heterozygous animals were intercrossed to obtain homozygous deletion animals for the mouse *Apoav* locus. Offspring were genotyped with PCR primers designed to the neomycin gene (described above) and with primers contained within the *Apoav* deleted interval (mApoAV-F2-5'-ACAGTTGGAGCAAAGGCGT-GAT-3' and mApoAV-R2-5'-CTTGCTCGAAGCTGC-CTTTCAG-3').
26. Blood samples were collected after a 5-hour fast by retro-orbital bleeding by using heparinized microhematocrit tubes. Total cholesterol and triglyceride concentrations were measured by using enzymatic methods on a Gilford System 3500 analyzer (Gilford Instruments, Oberlin, OH) (*45*).
27. L. A. Pennacchio, E. M. Rubin, unpublished observations.
28. Supplementary material is available on *Science* Online at www.sciencemag.org/cgi/content/full/294/5540/169/DC1.
29. E. L. Gong, L. J. Stoltfus, C. M. Brion, D. Murugesh, E. M. Rubin, *J. Biol. Chem.* **271**, 5984 (1996).
30. E. Boisfer *et al.*, *J. Biol. Chem.* **274**, 11564 (1999).
31. N. S. Shachter *et al.*, *J. Clin. Invest.* **93**, 1683 (1994).
32. N. S. Shachter *et al.*, *J. Clin. Invest.* **98**, 846 (1996).
33. C. M. Allan, J. M. Taylor, *J. Lipid Res.* **37**, 1510 (1996).
34. For the entire genomic sequence of *APOAV*, overlapping sequence-tagged sites (STSs) of 400 to 498 bp size were designed and tested by using PCR-amplification on human genomic DNA as previously described (*46*). Only primer pairs that resulted in a single PCR product of expected size were used for subsequent amplifications. For SNP discovery, STSs were PCR-amplified from eight DNA samples (16 chromosomes) of the Polymorphism Discovery Resource panel (PDR08, Coriell Cell Repository, Camden, NJ), and products were purified through Millipore plates according to the manufacturer's recommendations (Millipore, Bedford, MA). Subsequent se-

quencing reactions with purified PCR products were performed by using Big Dye Terminator chemistry and forward or reverse primers in separate sequencing reactions (Applied Biosystems, Foster City, CA). Reactions were analyzed by using a 3700 Sequence Analyzer (Applied Biosystems). Sequence traces were automatically analyzed by using PhredPhrap and Polyphred (47, 48). For SNPs identified through this analysis, PCR Invader assays (Third Wave Technologies, Madison, WI) were designed and tested on 90 samples from the Polymorphism Discovery Resource panel (PDR90) (49). Successful assays were subsequently used to analyze samples from our study. Genotypes were assigned automatically by cluster analysis (M. Olivier et al., in preparation). Differences among genotypes were analyzed by one-way ANOVA using STATVIEW 4.1 software (Abacus Concepts, Inc., Berkeley, CA). SNPs 1 to 4 are available in dbSNP under accession numbers ss3199913, ss3199914, ss3199915, and ss3199916, respectively.

35. Subjects were a combined subset of 501 healthy, non-smoking Caucasian individuals aged >20 years (429 men, 72 women) who had participated in previous dietary intervention protocols (50, 51) (R. M. Krauss et al., unpublished data). All subjects had been free of chronic disease during the previous 5 years and were not taking medication likely to interfere with lipid metabolism. In addition, they were required to have plasma total cholesterol concentrations <6.74 mmol/liter (260 mg/dl), triacylglycerol <5.65 mmol/l (500 mg/dl), resting blood pressure <160/105 mm Hg, and body weight <130% of ideal. Each participant signed a consent form approved by the Committee for the Protection of Human Subjects at E. O. Lawrence Berkeley National Laboratory, University of California, Berkeley, and participated in a medical interview. Fasting blood samples were obtained from subjects eating their usual diets, and after 4 to 6 weeks of consuming diets containing high fat (35 to 46% energy) and low fat (20 to 24% energy) (50, 51). Plasma lipid and lipoprotein measurements were performed as previously described (50, 51). In addition, on the high- and low-fat diets, total lipoprotein mass was measured by analytic ultracentrifugation (50, 51).
36. Of the 501 individuals in the original study, 388 were successfully genotyped by PCR amplification for the Sst I polymorphism as previously described (16, 28).
37. To genotype the C/T SNP3 polymorphisms upstream of APOAV, oligonucleotides AV6-F-5′-GAT TGAT TCAA-GATGCAT T TAGGAC-3′ and AV6-R-5′-CCCC-AGGAACTGGAGCGAAATT were used to amplify a 187-bp fragment from genomic DNA. The penultimate base in AV6-R was changed to T to create a Mse I site (TTAA) in the common allele. The PCR reactions were performed in 20 μl volumes containing 50 mmol/liter KCl, 10 mmol/liter tris (pH 8.3), 1.5 mmol/liter MgCl$_2$, 0.2 mmol/liter of each dNTP, 1 U of Taq DNA polymerase, and 200 pmol/liter of each primer. DNA was amplified under the following conditions: initial denaturation of 96°C for 2 min, followed by 32 cycles of 94°C for 15 s, 55°C for 30 s, and 72°C for 30 s, and a final step at 72°C for 3 min. PCR product (20 μl) was digested with 10 U of Mse I (New England Biolabs) at 37°C for 3 hours. The PCR products were size-fractionated on 3% agarose gels, stained with ethidium bromide, and visualized on an ultraviolet transilluminator.
38. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, J. Mol. Biol. 215, 403 (1990).
39. K. Osoegawa et al., Genome Res. 10, 116 (2000).
40. I. Dubchak et al., Genome Res. 10, 1304 (2000).
41. G. G. Loots et al., Science 288, 136 (2000).
42. C. Mayor et al., Bioinformatics 16, 1046 (2000).
43. A. Lupas, M. Van Dyke, J. Stock, Science 252, 1162 (1991).
44. H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Protein Eng. 10, 1 (1997).
45. C. C. Allain, L. S. Poon, C. S. Chan, W. Richmond, P. C. Fu, Clin. Chem. 20, 470 (1974).
46. E. M. Beasley, R. M. Myers, D. R. Cox, L. C. Lazzeroni, PCR Applications (Academic Press, San Diego, CA, 1999).
47. D. A. Nickerson, V. O. Tobe, S. L. Taylor, Nucleic Acids Res. 25, 2745 (1997).
48. B. Ewing, P. Green, Genome Res. 8, 186 (1998).
49. C. A. Mein et al., Genome Res. 10, 330 (2000).
50. R. M. Krauss, D. M. Dreon, Am. J. Clin. Nutr. 62, 478S (1995).
51. D. M. Dreon, H. A. Fernstrom, P. T. Williams, R. M. Krauss, Arterioscler. Thromb. Vasc. Biol. 17, 707 (1997).
52. Animals were killed, and tissues were harvested for either total RNA isolation by using the RNAeasy-midi protocol (Qiagen) or for poly(A)$^+$ mRNA isolation by using the FastTrack 2.0 system (Invitrogen, Carlsbad, CA). About 10 μg of total RNA or 2 μg of poly(A)$^+$ mRNA were separated in 1.0% agarose by gel electrophoresis and the RNA was transferred to a charged nylon membrane (Ambion, Austin, TX). The RNA blots were hybridized with [α-^{32}P]dCTP random-primed apoAV probes in ULTRAhyb buffer (Ambion). Probe templates were generated by PCR amplification of liver cDNA with degenerate primers degApoAV-F2-5′-GCGCGTGGTGGGRGAAGACA-3′ and degApoAV-R2-TCGCGCAGCTGGTCCAGGT T-3′. Filters were washed in 2× saline sodium citrate at room temperature for 20 min and in 0.1× SSC at 42°C for 20 min, followed by autoradiography visualization.
53. M. Olivier, unpublished observations.
54. R. M. Krauss, unpublished observations.
55. R. C. Lewontin, Genetics 120, 849 (1988).
56. We thank H. Hobbs, J. Fruchart, A. Plump, C. Prange-Pennacchio and members of the Rubin laboratory for thoughtful discussions; E. Gong, K. Houston, K. Lewis, W. Dean, J.-F. Cheng, I. Dubchak, J. Schwartz, V. Afzal, and X. Yang for technical support; V. Bustos, K. Sheppard, D. Zierten, A. de Witte, R. Freudenberg, J. Bushard, A. Almendras, and A. Indap for assistance with sequencing and genotyping; and P. Blanche, L. Holl, and J. Orr for performing lipoprotein measurements. This work was supported by the National Dairy Promotion and Research Board in cooperation with the National Dairy Council and NIH-NHBLI grant HL-18574 (R.M.K., E.M.R.), the NIH-NHLBI Programs for Genomic Application Grant HL66681 (E.M.R.), through the U.S. Department of Energy under contract no. De-AC03-76SF00098 (E.M.R.), HL-53917 (J.C.C.), and an appointment to the Alexander Hollaender Distinguished Postdoctoral Fellowship Program sponsored by the U.S. Department of Energy, Office of Biological and Environmental Research, and administered by the Oak Ridge Institute for Science and Education (L.A.P.).

30 July 2001; accepted 4 September 2001

Phosphorylation-Dependent Ubiquitination of Cyclin E by the SCF^Fbw7 Ubiquitin Ligase

Deanna M. Koepp,[1,2,3] Laura K. Schaefer,[1,2,3]* Xin Ye,[1]* Khandan Keyomarsi,[4] Claire Chu,[1] J. Wade Harper,[1] Stephen J. Elledge[1,2,3]†

Cyclin E binds and activates the cyclin-dependent kinase Cdk2 and catalyzes the transition from the G$_1$ phase to the S phase of the cell cycle. The amount of cyclin E protein present in the cell is tightly controlled by ubiquitin-mediated proteolysis. Here we identify the ubiquitin ligase responsible for cyclin E ubiquitination as SCF^Fbw7 and demonstrate that it is functionally conserved in yeast, flies, and mammals. Fbw7 associates specifically with phosphorylated cyclin E, and SCF^Fbw7 catalyzes cyclin E ubiquitination in vitro. Depletion of Fbw7 leads to accumulation and stabilization of cyclin E in vivo in human and Drosophila melanogaster cells. Multiple F-box proteins contribute to cyclin E stability in yeast, suggesting an overlap in SCF E3 ligase specificity that allows combinatorial control of cyclin E degradation.

Passage through the cell cycle is controlled by the activity of cyclin-dependent kinases (CDKs) (1). Cyclin E is the regulatory subunit of Cdk2 and controls the G$_1$ to S phase transition, which is rate-limiting for proliferation. Cyclin E is tightly regulated by ubiquitin-mediated proteolysis, which requires phosphorylation on Thr^{380} and Cdk2 activation (2–4). Failure to properly regulate cyclin E accumulation can lead to accelerated S phase entry (5), genetic instability (6), and tumorigenesis (7). Elucidating the mecha-

nism controlling cyclin E destruction has important implications for understanding control of cell proliferation during development and its subversion by tumorigenesis.

The formation of polyubiquitin-protein conjugates, which are recognized and destroyed by the 26S proteasome, involves three components that participate in a cascade of ubiquitin transfer reactions: a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), and a specificity factor (E3) called a ubiquitin ligase (8). E3s control the specificity of target protein selection and therefore are key to controlling individual target protein abundance.

The SCF (Skp1/Cullin/F-box protein) comprises a large family of modular E3s that control ubiquitination of many substrates in a phosphorylation-dependent manner (9). SCF complexes contain four subunits: Skp1, Cul1 (Cdc53), Rbx1, and an F-box–containing pro-