An apolipoprotein identified through comparative sequence analysis influences triglyceride levels in humans and mice.

Len A. Pennacchio, Michael Olivier, Jaroslav A. Hubacek, Jonathan C. Cohen, David R. Cox, Jean-Charles Fruchart, Ronald M. Krauss, and Edward M. Rubin*

L. A. Pennacchio, R. M. Krauss and E. M. Rubin, Genome Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA.

M. Olivier, D. R. Cox, Stanford Human Genome Center, Dept. of Genetics, Stanford University School of Medicine, 975 California Avenue, Palo Alto, CA, 94304, USA.

J. A. Hubacek and J. C. Cohen, Center for Human Nutrition and McDermott, Center for Human Growth and Development, UT Southwestern Medical Center, Dallas TX, 75390-9052, USA.

J. -C. Fruchart, 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.

current address:
Michael Olivier, Human and Molecular Genetics Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI, 53226, USA.

*To whom correspondence should be addressed. Email: EMRubin@lbl.gov

ABSTRACT

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 three-fold decrease in plasma triglyceride concentrations; conversely, knockout mice lacking *apoAV* had a four-fold increase in plasma triglycerides. 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 contribute to severe hyper-triglyceridemia *(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 kilobasepairs (kbp) of orthologous mouse DNA and compared the mouse and human sequences (Figure 1) *(18)*. Based on extended inter-species sequence conservation approximately 30 kbp proximal to the *apoAI/CIII/AIV* gene cluster, we identified a genomic interval that contained a putative apolipoprotein-like gene (*apoAV*) (Figure 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 1,107 basepair (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 similar genomic structure 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 N-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 (Figure 2A-B). Transcripts approximately 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 poly-adenylation *(21, 22)*.

To assess the function of apoAV, we generated mice over-expressing human *apoAV* as well as mice lacking *apoAV* through standard mouse transgenic and gene knockout technologies (Figure 2C-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 using a 26 kbp *Xho*I fragment predicted to contain only human *apoAV* and this genomic transgene was expressed in liver, similar to the endogenous gene (Figure 2C). These transgenic mice had approximately three-fold lower levels of plasma triglyceride when compared with control littermates (0.32±0.11 (S.D.) mg/ml versus 0.90±0.29; T-test $p<0.0001$) (Figure 3A). Similar data were obtained from a second independent founder line (data not shown). *ApoAV* knockout mice were generated by deleting the three exons predicted to encode apoAV (Figure 2D). Despite the lack of *apoAV* transcript (Figure 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 approximately four-fold higher levels of plasma triglyceride when compared with wild-type littermates (1.53±0.77 (S.D.) mg/ml versus 0.37±0.12; T-test $p<0.001$) (Figure 3B). Characterization of lipoprotein particles by fast protein liquid chromotography (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 compared with controls (see supplemental data). VLDL levels in a heterozygous knockout mouse were intermediate

between the homozygous knockout and control mouse. The FPLC peak VLDL elution volumes were similar in all animals, indicating comparable VLDL particle size *(27)*.

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 a*poAV* knockouts in our study displayed an approximately 400% increase in plasma triglycerides compared to the 30% decrease noted in *apoCIII* knockouts, while *apoAV* transgenics showed decreased triglyceride levels compared to 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 two-fold lower triglycerides than the previously described *apoCIII* knockout mice indicate *(10)* that changes in apoCIII alone can not explain the entire effect of apoAV. In addition to *apoCIII*, the over-expression of several human apolipoprotein transgenes has been shown to increase triglyceride levels in mice *(8, 9, 28-31)*, while only the *apoAV* transgene leads to decreased triglycerides suggesting a novel mechanism behind this effect.

The observation of significant lipid abnormalities in mice over-expressing 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 *(32)* (Figure 1A). Four markers with relatively high minor allele frequencies (>8%) were obtained. Three of the SNPs were separated by three kbp within *apoAV* (SNP1-3), while the fourth SNP (SNP4) was located ~11 kbp upstream of the gene

(Figure 1A). These markers were scored in approximately 500 random unrelated normo-lipidemic Caucasian individuals who had been phenotyped for numerous lipid parameters before and after consumption of high- and low-fat diets *(33)*. We found significant associations between both plasma triglyceride levels and VLDL mass and the three neighboring SNPs (SNPs1-3) within *apoAV* but not with the distant upstream SNP4 (Figure 1A, 4A). Specifically, the minor allele of each of these SNPs (SNPs1-3) was associated with higher triglyceride levels independent of diet. Independent analysis of each of these SNPs (SNP1-3) revealed plasma triglyceride levels were 20-30% higher in individuals having one minor allele compared to individuals homozygous for the major allele (Figure 4A). Analysis of SNP allele frequencies in more than 1,000 chromosomes revealed that the three neighboring SNPs (SNPs1-3) in *apoAV* were in significant linkage disequilibrium that does not extend to SNP4 (located ~11kb upstream of *apoAV*) (Figure 4B). This finding supports the existence of a common haplotype in the *apoAV* region influencing plasma triglyceride levels (Figure 1A, 4B). Furthermore, studies in this population found no significant association of triglyceride levels with a *Sst*1 polymorphism in *apoCIII* (located ~40 kbp upstream of *apoAV*) (Figure 1A) which has been previously associated with severe hyper-triglyceridemia *(15, 16, 34)* (see supplemental data). This finding indicate the *apoCIII Sst*1 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 *(35)*. 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 (SNPs1-2) (Figure 4A-B). In the second study, we examined the allele frequencies for SNP3 in an unrelated group of Caucasians stratified according to plasma triglyceride levels (Figure 4C). The two groups represented 161 individuals with triglyceride levels in the top tenth-percentile and 298 individuals

from the bottom tenth-percentile. A significant over-representation of the heterozygous genotype was found in individuals with high- compared to low-plasma triglyceride levels (21.7% versus 6.7%, respectively), thereby validating the effect in a second cohort. When the cohort was stratified based on gender, an even more pronounced over-representation of the heterozygous genotype was found in males with high- compared to 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 human/mouse sequence comparison, 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. While 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 hyper-triglyceridemia 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, *American Journal of Public Health* **41**, 279-286 (1951).

2. N. Duverger, et al., *Science* **273**, 966-8 (1996).

3. R. M. Krauss, *Am J Med* **105**, 58S-62S (1998).

4. E. M. Rubin, A. Tall, *Nature* **407**, 265-9 (2000).

5. A. J. Lusis, *Nature* **407**, 233-41 (2000).

6. A. S. Plump, et al., *Cell* **71**, 343-53 (1992).

7. G. A. Bruns, S. K. Karathanasis, J. L. Breslow, *Arteriosclerosis* **4**, 97-102 (1984).

8. Y. Huang, et al., *J Biol Chem* **273**, 26388-93 (1998).

9. Y. Ito, N. Azrolan, A. O'Connell, A. Walsh, J. L. Breslow, *Science* **249**, 790-3 (1990).

10. N. Maeda, et al., *J Biol Chem* **269**, 23610-6 (1994).

11. L. Masucci-Magoulas, et al., *Science* **275**, 391-4 (1997).

12. E. M. Rubin, R. M. Krauss, E. A. Spangler, J. G. Verstuyft, S. M. Clift, *Nature* **353**, 265-7 (1991).

13. H. Allayee, et al., *Am J Hum Genet* **63**, 577-85 (1998).

14. A. P. Wojciechowski, et al., *Nature* **349**, 161-4 (1991).

15. M. R. Hayden, et al., *Am J Hum Genet* **40**, 421-30 (1987).

16. M. Dammerman, L. A. Sandkuijl, J. L. Halaas, W. Chung, J. L. Breslow, *Proc Natl Acad Sci U S A* **90**, 4562-6 (1993).

17. Electronic homology searches with human *apoAI*, *apoCIII*, and *apoAIV* mRNA sequences using the BLAST algorithm *(36)* identified a genomic bacterial artificial chomosome (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 using the polymerase chain reaction (PCR) with mouse primers: apoAI-F1-5'-GAGGATGTGGAGCTCTACCGC-3' and apoAI-R1-5'-

CTGTGTGCGCAGAGAGTCTACG-3') (RPCI-23, BACPAC Resources, Children's Hospital Oakland Research Institute; http://www.chori.org/bacpac/ *(37)*. Positive clone RPCI-23-175F2 was identified, randomly sheared, sub-cloned and sequenced to approximately six-fold coverage (GenBank accession number AF401201) *(38, 39)*. Human and mouse sequence comparisons were performed as previously described and are available at http://pga.lbl.gov *(40)*.

19. A. M. Gotto, Jr., H. J. Pownall, R. J. Havel, *Methods Enzymol* **128**, 3-41 (1986).

20. Protein analyses were performed using the web-based Predict-Protein package, COILS *(41)*, and SignalP *(42)* (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**, 685-90 (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, *Nat Genet* **9**, 424-31 (1995).

24. C. Paszty, et al., *Nat Genet* **11**, 33-39 (1995).

25. Restriction enzyme predictions for human genomic sequence (Genbank accession number AC007707) indicated the entire human *apoAV* gene, but not neighboring genes, was contained within a 26 kbp *Xho*I DNA fragment (corresponding to approximately 1-27 kbp in Figure 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 pulse-field gel electrophoresis. The 26 kbp *Xho*I DNA fragment containing human *apoAV* was purified using QiaEx II gel purification (Qiagen, Valencia, CA), adjusted to a final concentration of ~1 ng/ml and micro-injected into fertilized FVB inbred mouse eggs using standard procedures *(23)*. Two founder transgenic mice were identified as determined by PCR amplification using primers hApoAV-intrn-F1-5'-CCCGCTGCAGTCCCCAGAAT-3' and hApoAV-intrn-R1-5'-CAGGGTCGAGGGCTCTTGTCCT-3'. Each

founder line was expanded by breeding to isogenic FVB strain mice (The Jackson Laboratory, Bar Harbor, MN).

The targeting construct to delete mouse *apoAV* was built 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 using primers containing introduced 5' restriction sites for *Xba*I and *Eco*RI, respectively: mAV-XbaI-F1-5'-TGACTCTAGATACCCTTGGTCCCATGTTCCAGAT-3' and mAV-EcoRI-R1-5'-CATTGAATTCGACAAGAGAAAGACGGGGCTCAAG-3'. The resulting 4.2 kbp PCR product was cloned into pXL-Topo (Invitrogen, Carlsbad, CA), DNA prepared by standard alkaline lysis (Qiagen, Valencia, CA) and digested with *EcoR*I according to the manufacturer's recommendations (New England Biolabs, Beverly, MA). A 4.2 kbp *EcoR*I fragment was gel-purified and cloned into the *EcoR*I site of the pPN2T vector to yield pPN2T-Arm1 *(24)*. Clones were PCR screened for inserts using the above described primers and positive clones were sequenced for proper orientation. The second homology arm was PCR-amplified using primers mAV-NotI-F4-5'-TATGACTGCGGCCGCCACCAATCCCACATCTAAGCATCT-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. 129/SvJ ES cells (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 FIAU for 8 days. Individual clones were isolated, expanded and screened by Southern blot analysis. The external 3' probe was amplified by PCR using primers mApoAV-3'probe-F2-5'-CTTGAGGATGGGCATCAGCTGTAT-3' and mApoAV-3'probe-R2-5'-GCTCACTAACAGCGCTCTTGCCT-3'. Targeted clones were injected into C57BL/6 blastocysts and chimeric males were bred to C57BL/6

females (The Jackson Laboratory, Bar Harbor, MN). Agouti offspring were tested for germline transmission of the targeted allele by PCR using primers specific to the neomycin gene (NeoF1-5'-CTTTTTGTCAAGACCGACCTG-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'-ACAGTTGGAGCAAAGGCGTGAT-3' and mApoAV-R2-5'-CTTGCTCGAAGCTGCCTTTCAG-3').

26. Blood samples were collected after a 5-hour fast by retro-orbital bleeding using heparinized micro-hematocrit tubes. Total cholesterol and triglyceride concentrations were measured using enzymatic methods on a Gilford System 3500 analyzer (Gilford Instruments, Oberlin, OH) *(43)*.

27. E. L. Gong, L. J. Stoltfus, C. M. Brion, D. Murugesh, E. M. Rubin, *J Biol Chem* **271**, 5984-87 (1996).

28. E. Boisfer, et al., *J Biol Chem* **274**, 11564-72 (1999).

29. N. S. Shachter, et al., *J Clin Invest* **93**, 1683-90 (1994).

30. N. S. Shachter, et al., *J Clin Invest* **98**, 846-55 (1996).

31. C. M. Allan, J. M. Taylor, *J Lipid Res* **37**, 1510-8 (1996).

32. For the entire genomic sequence of *apoAV*, overlapping sequence-tagged sites (STSs) of 400-498 bp in size were designed and tested using PCR-amplification on human genomic DNA as previously described *(44)*. 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 sequencing reactions with purified PCR products were performed using Big Dye Terminator chemistry and forward or reverse primers in separate sequencing reactions (Applied Biosystems, Foster City, CA). Reactions were analyzed using a 3700 Sequence Analyzer

(Applied Biosystems). Sequence traces were automatically analyzed using PhredPhrap and Polyphred *(45, 46)*. For SNPs identified through this analysis, PCR Invader assays (Third WaveTechnologies, Madison, WI) were designed and tested on 90 samples from the Polymorphism Discovery Resource panel (PDR90) *(47)*. Successful assays were subsequently used to analyze samples from our study. Genotypes were assigned automatically by cluster analysis (Olivier et al., in preparation). Differences among genotypes were analyzed by one way ANOVA using STATVIEW 4.1 software (Abacus Concepts, Inc, Berkeley, CA). SNPs1-4 are available in dbSNP under accession numbers ss3199913, ss3199914, ss3199915 and ss3199916, respectively.

33. Subjects were a combined subset of 501 healthy, nonsmoking Caucasian individuals aged >20 years (429 men, 72 women) who had participated in previous dietary intervention protocols *(48, 49)* (Krauss et al. unpublished). All subjects had been free of chronic disease during the previous five 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/L (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 EO Lawrence Berkeley National Laboratory, University of California, Berkeley, and participated in a medical interview. Fasting blood samples were obtained on their usual diets, and after 4-6 weeks of consuming diets containing high fat (35-46% energy) and low fat (20-24% energy) *(48, 49)*. Plasma lipid and lipoprotein measurements were performed as previously described *(48, 49)*. In addition, on the high and low fat diets, total lipoprotein mass was measured by analytic ultra-centrifugation *(48, 49)*.

34. 393 of the 501 individuals in the original study were genotyped by PCR-amplification for the *Sst*1 polymorphism as previously described *(16)* (see supplemental data).

35. To genotype the C/T SNP3 polymorphisms upstream of *apoAV*, oligonucleotides AV6-F-5'-GATTGATTCAAGATGCATTTAGGAC-3' and AV6-R-5'-CCCCAGGAACTGGAGCGAAA***T***T 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/l KCl, 10 mmol/l Tris (pH 8.3), 1.5 mmol/l $MgCl_2$, 0.2 mmol/l of each dNTP, 1 U of Taq DNA polymerase and 200 pmol/l of each primer. DNA was amplified using the following conditions: initial denaturation of 96°C for 2 min, followed by 32 cycles of 94°C for 15 sec, 55°C for 30 sec and 72°C for 30 sec, and a final step at 72°C for 3 min. 20 μls of PCR product were digested with 10 U of *Mse*I (New England Biolab) at 37°C for 3 h. The PCR products were size-fractionated on 3% agarose gels, stained with ethidium bromide and visualized on a UV transilluminator.

36. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J Mol Biol* **215**, 403-10 (1990).

37. K. Osoegawa, et al., *Genome Res* **10**, 116-28 (2000).

38. I. Dubchak, et al., *Genome Res* **10**, 1304-6 (2000).

39. G. G. Loots, et al., *Science* **288**, 136-40 (2000).

40. C. Mayor, et al., *Bioinformatics* **16**, 1046-7 (2000).

41. A. Lupas, M. Van Dyke, J. Stock, *Science* **252**, 1162-4 (1991).

42. H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, *Protein Eng* **10**, 1-6 (1997).

43. C. C. Allain, L. S. Poon, C. S. Chan, W. Richmond, P. C. Fu, *Clin Chem* **20**, 470-5 (1974).

44. E. M. Beasley, R. M. Myers, D. R. Cox, L. C. Lazzeroni, *PCR Applications* (Academic Press, San Diego, CA, 1999).

45. D. A. Nickerson, V. O. Tobe, S. L. Taylor, *Nucleic Acids Res* **25**, 2745-51 (1997).

46. B. Ewing, P. Green, *Genome Res* **8**, 186-94 (1998).

47. C. A. Mein, et al., *Genome Res* **10**, 330-43 (2000).

48. R. M. Krauss, D. M. Dreon, *Am J Clin Nutr* **62**, 478S-487S (1995).

49. D. M. Dreon, H. A. Fernstrom, P. T. Williams, R. M. Krauss, *Arterioscler Thromb Vasc Biol* **17**, 707-14 (1997).

50. Animals were sacrificed and tissues harvested for either total RNA isolation using the RNAeasy-midi protocol (Qiagen, Valencia, CA) or for polyA mRNA isolation using the FastTrack 2.0 system (Invitrogen, Carlsbad, CA). Approximately 10 μg of total RNA or 2 μg of polyA 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, Austin, TX). Probe templates were generated by PCR amplification of liver cDNA using degenerate primers degApoAV-F2-5'-GCGCGTGGTGGGRGAAGACA-3' and degApoAV-R2-TCGCGCAGCTGGTCCAGGTT-3'. Filters were washed in 2X saline sodium citrate at room temperature for 20 minutes and in 0.1X SSC at 42º C for 20 minutes, followed by autoradiography visualization.

51. R. C. Lewontin, *Genetics* **120**, 849-52 (1988).

52. Acknowledgments. 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, Veena Afzal and Xinli 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 in part by the National Dairy Promotion and Research Board and administered in cooperation with the National Dairy Council and NIH Program Project grant HL-18574 from the National Heart, Lung, and Blood Institute through the U.S. Department of Energy under contract no. DE-AC03-76SF00098 (RMK), the NIH, NHLBI, Programs for Genomic Application Grant HL66681 and performed under Department of Energy Contract DE-AC0376SF00098, University of California (EMR), HL-53917 (JCC), 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 (LAP).

FIGURE LEGENDS

FIGURE 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 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 50-100% identity).

FIGURE 2: *ApoAV* expression in humans and wild-type, transgenic and knockout mice *(50)*. **(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), including: 1, heart; 2, brain; 3, spleen; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; and 8, testis. **(B)** A human *apoAV* cDNA probe was hybridized to a RNA blot containing eight human tissues (Clontech, Palo Alto, California), including: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; and 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 follow: 1,5 transgenic liver; 2,6 transgenic intestine; 3,7 wild-type liver; 4,8 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 stems cells were identified using an external 3' probe which detects a 17 kb *EcoR*I fragment

wild-type allele and a 10 kb *EcoR*I fragment upon targeting (data not shown). **(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 was examined by ethidium bromide staining.

FIGURE 3: Plasma triglyceride and cholesterol levels for *apoAV* transgenic and knockout mice on standard chow diet. **(A)** Human *apoAV* transgenic mice compared to isogenic FVB strain control littermates (n=48 for transgenics; n=44 for controls; student t-test *p<0.0001 for transgenic versus control). **(B)** Mice lacking *apoAV* compared to mixed 129Sv/C57Bl6 strain controls littermates (n=13 for wild-type, +/+; n=22 for heterozygotes, +/-; n=10 for homozygous knockouts, -/-; student 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 to controls (data not shown).

FIGURE 4: Human *apoAV* polymorphisms and lipid association data. **(A)** Plasma lipid concentrations for a given genotype for 4 neighboring SNPs (SNPs1-4). 501 individuals were genotyped and the number of successfully scored individuals is indicated. 1,1=homozygous for the major allele; 1,2=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. Due to the small number of individuals, these data were excluded from the analysis. All sites were found to be in Hardy-Weinberg equilibrium (data not shown). The minor allele frequency for each SNP (SNPs1-4) 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, data not shown). **(B)** Pair-wise measure of linkage disequilibrium (|D'|) was calculated for all combinations of SNPs as previously described *(51)*. 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 using a similar analysis (p=0.044).