Haplotypes in the *APOA1-C3-A4-A5* Gene Cluster affect Plasma Lipids in both Humans and Baboons

Qian-fei Wang[1], Xin Liu[2], Jeff O'Connell[3], Ze Peng[1], Ronald M. Krauss[1,4], David L. Rainwater[5,6], John L. VandeBerg[5,6], Edward. M. Rubin[1,7], Jan-Fang Cheng[1,7], and Len A. Pennacchio[1,7,*]

[1]Genome Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[3]Division of Endocrinology, Diabetes, and Nutrition University of Maryland School of Medicine, Baltimore, MD, USA

[4]Children's Hospital Oakland Research Institute, Oakland, CA, USA

[5]Department of Genetics and [6]Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, USA

[7]USA Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, USA

To whom correspondence should be address:

Dr. Len A. Pennacchio, Department of Genome Sciences, MS 84-171, One Cyclotron Road, Lawrence Berkeley National Laboratory, Berkeley, CA. 94720. Email: LAPennacchio@lbl.gov. Phone: (510) 486-7498. Fax: (510) 486-4229.

Abstract

Genetic studies in non-human primates serve as a potential strategy for identifying genomic intervals where polymorphisms impact upon human disease-related phenotypes. It remains unclear, however, whether independently arising polymorphisms in orthologous regions of non-human primates leads to similar variation in a quantitative trait found in both species. To explore this paradigm, we studied a baboon apolipoprotein gene cluster (*APOA1/C3/A4/A5*) for which the human gene orthologs have well-established roles in influencing plasma HDL-cholesterol and triglyceride concentrations. Our extensive polymorphism analysis of this 68 kb gene cluster in 96 pedigreed baboons identified several haplotype blocks each with limited diversity, consistent with haplotype findings in humans. To determine whether baboons, like humans, also have particular haplotypes associated with lipid phenotypes, we genotyped 634 well characterized baboons using 16 haplotype tagging SNPs. Genetic analysis of single SNPs, as well as haplotypes, revealed an association of *APOA5* and *APOC3* variants with HDL-cholesterol and triglyceride concentrations, respectively. Thus, independent variation in orthologous genomic intervals does associate with similar quantitative lipid traits in both species, supporting the possibility of uncovering human QTL genes in a highly controlled non-human primate model.

Introduction

Defining the genetic basis for quantitative traits in humans is essential for our better understanding of the vast biological differences found between individuals, including disease susceptibility. Cross-species approaches offer a logical means for leveraging the similarities between organisms to identify genomic intervals where variation may have an impact upon phenotypes in common across species. While quantitative trait loci (QTLs) in humans and rodents have occasionally been mapped to large orthologous intervals (1, 2), it is rarely possible to determine if the same gene is involved due to phenotypic and genomic differences between these organisms.

In contrast to rodents, non-human primates are expected to share a larger number of phenotypic traits and associated genetic etiologies with humans due to their close evolutionary relationship and physiology. However, while wide-spread success has been obtained at identifying the molecular defects in single gene diseases, human genetic studies of complex traits has proved extremely difficult due to the number of genes involved, the small effects of individual polymorphisms, the complicated nature of gene:gene and gene:environment interactions, and uncontrolled environmental factors (3, 4).

Baboons (*Papio hamadryas*) are a well-studied non-human primate model with which it is possible to control mating and environment (exercise, life-style, diet, etc), thereby alleviating several confounding factors found in human studies (5). Through the recent

development of a baboon linkage map (6), several lipoprotein and hypertension QTLs have been physically mapped of which some overlap with QTLs mapped to orthologous genomic intervals in humans (7, 8). To date, however, no fine-scale orthologous sequence analysis has been performed for any quantitative trait to define whether independently arising polymorphisms in the identical genomic interval are responsible for similar phenotypic differences found in both species.

The apolipoprotein gene cluster (*APOA1/C3/A4/A5*) on human chromosome 11q23 is among the best characterized regions of the genome for its association with plasma lipid levels, a quantitative trait. A wealth of human mutation and genetic association studies demonstrate that sequence variants in *APOC3* and *APOA5* are associated with plasma HDL-cholesterol and/or triglyceride concentrations (9-17). In addition, over-expression as well as deletion of these genes provide confirmation that each plays an important role in lipid homeostasis (10, 18, 19).

We examined the orthologous baboon gene cluster in a large pedigreed population to explore whether independent mutation events at the *APOA1/C3/A4/A5* gene cluster might account for quantitative differences in lipid phenotypes in a second primate species. Our goal was to examine the genetic architecture at this site and to determine if polymorphisms within the baboon cluster also influence plasma lipid levels similar to that found in humans. In these studies we found baboons also display limited haplotype diversity and show genetic associations between common variants in this apolipoprotein cluster and HDL-cholesterol and triglyceride concentrations. These studies highlight the

utility of the environmentally controlled baboon model to uncover genomic intervals where polymorphisms can also impact upon phenotypes relevant to human disease.

Results

SNP Discovery

To systematically characterize linkage disequilibrium and haplotype structure in the baboon *APOA1/C3/A4/A5* gene cluster, we performed direct DNA sequencing across the interval in 24 baboons selected from 15 presumably independent pedigrees. We analyzed 68 kb of sequence containing the entire gene cluster plus flanking sequence and detected 381 genetic variation sites (8 in coding exons, 79 in introns, and 6 in untranslated regions). Of them, 363 are single nucleotide polymorphisms (SNPs) and 18 (or 4.7%) are insertion/deletion (in/dels) polymorphisms (Table 1). In the human homologous region, 182 SNPs and 4 in/dels were reported in NCBI dbSNP build 116. As expected, none of the 381 baboon and 186 human polymorphisms were shared between the two primate species. As shown in table 1, the baboon SNPs average 1 per 187 bp across the gene cluster; with 1 in 276 bp in the UTR region, 1 in 626 bp in the coding region, and 1 in 177 bp in the intron and intergenic regions. The average minor allele frequency of the total number SNPs in 24 baboons is 17.3%. The average minor allele frequency is 18.0%, 19.3% and 17.2% in the UTR region, the coding region and in the intron and intergenic regions, respectively. After removing SNPs that were not in Hardy-Weinberg equilibrium, we selected 248 SNPs that had a minor allele frequency of more than 8% for subsequent analysis. These SNPs spanned a total of 68 kb of sequence on baboon chromosome 14, resulting in an average distance of 274 bp between neighboring SNPs.

Selection of representative SNPs across the apolipoprotein gene cluster

In an effort to reduce the numbers of SNPs needed for future genotyping, we analyzed the extent of linkage disequilibrium across these variants in the apolipoprotein gene cluster. Since r^2 is an appropriate statistic in determining whether an assayed SNP can detect an unassayed SNP, we calculated r^2 for all pairs of SNPs using unphased data from 24 unrelated baboons and thereby clustered SNPs with highly correlated alleles (with r^2 >0.8) (20, 21) (Figure 1). Within each cluster, several SNPs were selected to represent the correlated group of SNPs. In total, 35 SNPs were selected as displayed in Figure 1.

Development of the baboon haplotype block structure and selection of tagging SNPs based on common haplotypes

To determine the haplotype structure of the baboon apolipoprotein gene cluster, we genotyped the 35 selected SNPs in 96 pedigreed baboons. We used PHASE to infer haplotypes across the gene cluster for each individual and to identify blocks and common haplotypes within each block (Figure 2). Three haplotype blocks were identified in the region (Figure 2), with *APOA1*, *APOC3/APOA4*, and *APOA5* falling into separate blocks. Each LD block was on average 11 kb in length and was comprised of 4 common haplotypes which accounted for 75-85% of all chromosomes. Based on these results, 16 SNPs were selected as haplotype tagging SNPs that suffice to determine the haplotypes accounting for 82%-91% of the chromosomes within each haplotype block (Figure 2).

Association analysis of single nucleotide polymorphisms and haplotypes with baboon lipoprotein phenotypes

We genotyped the 16 haplotype tagging SNPs in a study population that consisted of 634 baboons. We performed single marker QTDT tests on all 16 SNPs and found six associated with five lipoprotein phenotypes that had an asymptotic P-value < 0.01. When permutation tests (m=1,000) were performed on all 6 markers and 5 traits, the only significant association was between SNP 13 from *APOA5* intron 2 and HDL-cholesterol on the chow diet (Table 3).

Since analysis of haplotypes can have stronger statistical power compared to that of single marker polymorphisms, we performed two-locus, three-locus and four-locus haplotype analysis on the *APOA1*, *APOC3,* and *APOA5* genes in the baboon cluster. Lipid traits included in the test were triglycerides and HDL-cholesterol since they are known to be associated with the orthologous gene cluster in humans. Haplotype analysis confirmed the result from single SNP association (Table 3) (empirical P=0.014), further supporting that the *APOA5* variant is associated with HDL-cholesterol levels (Table 4). In addition, a four-locus haplotype within the *APOA1/C3* block showed significant association with serum TG levels (empirical P=0.002, table 4), an effect that was not observed when individual markers from this haplotype were tested.

Discussion

In this study we sought to systematically examine the baboon polymorphism architecture and it's relationship to phenotypic data in a region orthologous to humans which has been biologically well characterized. The baboon interval was selected based on orthology to the extensively studied human apolipoprotein gene cluster on chromosome 11q23 which has been repeatedly implicated in contributing to inter-individual differences in plasma lipid levels. Our primary goal was to study variation at this locus and to determine if baboon variants within this region are also associated with quantitative differences in plasma lipid levels.

In addition, this study supports a higher SNP density in the baboon population relative to humans. While numerous studies indicate that human SNPs are estimated to be spaced at approximately one every 300 bp in the human population (minimum minor allele frequency of 1%) (22-24), we found that baboons display one SNP every 187 bp (minimum minor allele frequency of 2%), and that these SNP alleles display strong patterns of linkage disequilibrium, similar to findings in humans. Our original direct sequencing analysis identified 363 SNPs in 24 baboons which collapsed into 35 representative SNPs based on the exclusion of redundant SNPs displaying high levels of linkage disequilibrium. Further characterization of 35 SNPs by genotyping in 96 baboons revealed that the baboon *APOA1/C3/A4/A5* gene cluster forms block structures with each block exhibiting limited haplotype diversity. Specifically, *APOA1*, *APOC3/APOA4,* and *APOA5* fell into three separate blocks with an average block size of

11 kb. However, the exact location of haplotype blocks varied depending on the SNP density, SNP frequency and the block definition used (data not shown). While it is generally not practical to compare block distribution between studies (25), a similar complexity of haplotype pattern in the human *APOA1/C3/A4/A5* region have recently been reported (26). Our goal, nevertheless, was to capture the general SNP architecture in the region where we found limited diversity and this information proved useful to facilitate our subsequent genetic association analysis.

To study genetic variants in the baboon *APOA1/C3/A4/A5* gene cluster and their possible contribution to plasma lipid concentrations, we used the haplotype structure in the *APOA1/C3/A4/A5* gene cluster region to select 16 haplotype tagging SNPs (ht SNPs) for genotyping in 634 baboons which had undergone a dietary challenge. Single marker QTDT as well as haplotype analysis supported an association of variation in the *APOA5* gene with HDL-cholesterol levels (empirical P value=0.014). In humans associations between *APOA5* variants and HDL-cholesterol levels have been reported (11-13), though we did not observe associations with *APOA5* variants and triglycerides despite reports in humans (10, 27-31). In addition, a second association was identified between an *APOC3* baboon haplotype and triglyceride concentrations. This is entirely consistent with a large number of human studies which have shown strong genetic associations between *APOC3* variants and triglyceride levels (9, 14-17) Thus, independently arising polymorphisms in humans and baboons both can contribute to quantitative differences in similar plasma traits, but the effects are not completely overlapping.

A major assumption underlying studies using non-human primates (and other mammals) to help identify loci contributing to traits that are directly relevant to those found in humans is that independent mutations have occurred at orthologous loci across species and that these polymorphisms have similar effects on a phenotype. In addition to possible physiological differences, this independent origin of polymorphisms in humans and baboons most likely explains the incomplete overlap between genetic variants and plasma phenotypes and additional large studies are needed to clarify the extent of which polymorphisms in orthologous human/baboon genes account for similar quantitative phenotypic variability in both species.

One significant advantage of genetic studies in baboons versus humans is the potential for stringent control of environmental factors which are likely to profoundly impact upon complex phenotypes. In addition, matings in baboons can be planned to specifically facilitate the ability to detect, characterize, and localize genes that influence phenotypic variation. As QTL mapping has led to the identification of many genes underlying polygenic traits, it is becoming a promising approach in the genetic dissection of mammalian complex traits (1). With the recent development of a baboon linkage map, current efforts are focused on linkage and segregation analysis with the ultimately goal of uncovering genomic intervals affecting baboon, as well as human, biology. Our results support that the baboon is a valuable resource for uncovering genetic loci of relevance to human phenotypes and that similar fine-scale studies aimed at refining genetically mapped QTLs in non-human primates has the potential to uncover novel genetic contributors to a wide spectrum of shared human traits.

Materials and Methods

Pedigreed baboons

The pedigreed baboons (*Papio hamadryas*) used in this study have been previously described (32) and represent the genetic diversity arising from ~212 founders. The animals were maintained at the Southwest Foundation for Biomedical Research, a facility certified by the Association for Assessment and accreditation of Laboratory Animal Care International. Animals were subjected to a previously-described dietary challenge protocol (32, 33). Briefly, animals were fed three diets contrasting in levels of fat and cholesterol: basal diet was low in cholesterol (0.03 mg/kcal) and fat (4% of calories), LCHF was high in fat (40% of calories) but no cholesterol was added, and HCHF diet was the high fat diet with high levels of cholesterol (1.7 mg/kcal). Serum samples taken at the end of each dietary regime were stored as single-use aliquots protected from oxidation and dessication (34). Experimental protocols were supervised by a veterinarian and approved by the Institutional Animal Care and Use Committee at Southwest Foundation for Biomedical Research.

SNP identification

A baboon bacterial artificial chromosome (BAC) containing the *APOA1/C3/A4/A5* gene cluster was isolated by filter hybridization using human probes (http://bacpac.chori.org/baboon41bac.htm). The sequence of one positive BAC (RPCI-41:109F19 GenBank Accession # AC145521) was determined by fluorescent dye-terminator sequencing and was assembled using the Phred-Phrap-Consed assembly suite

(35-37). This baboon BAC sequence served as the reference for subsequent polymorphism detection in additional baboon individuals. Briefly, primers were designed to amplify 2 to 3 Kb of overlapping genomic DNA in a 68 Kb region containing the *APOA1/C3/A4/A5* gene cluster. PCR was performed using 2ng/μl of genomic DNA, 100 uM of dNTPs, 140 nM of primers, eLONGase and buffer (Invitrogen) for 35 cycle at 94X for 30 sec, 60˚C for 30 sec, and 68˚C for 2 min. DNA samples amplified from 24 unrelated baboons were sequenced on both strands by BigDye terminators (Applied Biosystems) with custom primers designed every 300 bp. Sequence reads were assembled using Phred-Phrap-Consed suite (35-37) and nucleotide polymorphisms were determined by PolyPhred (38) and then confirmed by visual inspection.

SNP genotyping

Genotyping of pedigreed baboons was carried out using a commercially available primer extension technique, SNaPshot (Applied Biosystems). Five SNPs were genotyped simultaneously in the same reaction. Equal amount (50 ng each) of PCR amplified genomic DNA fragments containing 5 surveyed SNPs were pooled and treated with 5U of SAP (shrimp alkaline phosphatase) and 2U of Exo I at 37˚C for 1 hour in a 20 μl reaction to remove the residual primers and free nucleotides. The SAP/Exo I reaction was quenched at 75˚C for 15 min. Sequences adjacent to the SNP sites were used to design SNaPshot primers, and 5 primers with a similar Tm but different length of tail sequences were combined in the same SNaPshot reaction. The SNaPshot reactions contain 3 ul of the pooled PCR products, 0.2 uM each of the 5 SNaPshot primers, 5 μl of the SNaPshot Multiplx Reaction mix (Applied Biosystems), and ddH2O to adjust to a 10

ul final volume. The SNaPshot reactions were performed using the thermal cycling condition recommended by the manufacturer, and then treated with 1U of CIP (calf intestinal phosphatase) at 37˚C for 1 hour and quenched at 75˚C for 15min. 1 μl of the SNaPshot samples were then mixed with the same volume of the GeneScan-120 LIZ size standard (Applied Biosystems), diluted with 18 μl of Hi-Di formamide, and denatured at 95˚C for 5 min before loading to the ABI Prism 3700 DNA Analyzer. Gel images were analyzed using the GeneScan Analysis Software version 3.7 and the polymorphic nucleotides were called using the Genotyper Software version 3.7. All nucleotide calls were inspected visually by two individuals to ensure accuracy. Thirty-five SNPs in 96 baboons and 16 SNPs in 634 baboons were genotyped using SNaPshot (see results). To independently evaluate the data quality, we randomly selected 24 baboons from the study population and acquired genotypes at all SNP locations used in Snapshot assays by direct sequencing. In the data analysis, we only included SNPs which have a minimum of 95% consistency between the genotypes from the two methods.

Analysis of LD and haplotypes

The genotypes of 24 unrelated baboons were used to calculate r^2 value using the VG2 software (39, 40) (http://pga.gs.washington.edu/VG2.html) and haploview (http://www-genome.wi.mit.edu/personal/jcbarret/haplo/docs.html). To select representative SNPs from each r^2 clustering group, we chose a minimum of one SNP to cover each cluster group and SNPs located in the proximity of a gene were preferred over those in intergenic or repetitive DNA regions.

Ninety-six pedigreed baboons were genotyped using SNaPshot (see above). Genotypes in violation with Mendelian expectations (0.18%) were excluded from further analysis. We treated 96 pedigreed baboons as population samples and used the PHASE program (version 2.0, (41)) to infer haplotypes . In two cases Mendelian inheritance was not found based on inferred haplotypes and these data were excluded from further analysis. The haplotypes were then imported into haploblockFinder (version 0.6, (42) http://cgi.uc.edu/cgi-bin/kzhang/haploBlockFinder.cgi) for block structure determination. The haplotypes of 16 SNPs for the 634 baboons were estimated using the program PHASE.

Although redundant SNPs were removed based on our r^2 clustering analysis ($r^2 > 0.8$) (see above and results), strong linkage disequilibrium(LD) still existed among some of the selected 31(in 96 baboons) and 16 (in 634 baboons) SNPs (data not shown). As serious concern has been raised as to use programs assuming no LD to perform pedigree haplotype inference (43), we did not use GENEHUNTER or SIMWALK2 despite such programs consider family relationship in haplotype estimation. We were unable to take advantage of other analytical tools allowing for LD for nuclear families ((44), PHASE-Phamily analysis at http://archimedes.well.ox.ac.uk/pise/phamily-simple.html) based on our extended baboons pedigree which is structured with inbreedings and multiple mates for a single male.

Genetic association

Both single point and multipoint (haplotype) analyses were performed using the QTDT program (version 2.3.0, (45)). P-values <0.05 were considered statistically significant. For single marker analysis, 16 markers (see table 2) and 8 lipoprotein measurements on 3 different diets as well as triglyceride levels on chow diet (see "Pedigreed baboons" and "Lipoprotein measurements") were tested.

To perform haplotype-based analysis, two- , three- or four locus haplotypes in the *APOA1/C3* and *APOA5* region (table 4) were recoded as single genotypes. *APOA4* haplotypes were not examined since there was no association of genetic variants in this gene with lipoprotein traits. For this analysis, traits only included HDL-Cholesterol and triglycerides since those were the two measurements which consistently showed association with *APOA1/C3* or *APOA5* variants in human. Haplotypes with a frequency < 0.5% were omitted from further analysis. Haplotypes were determined by the PHASE program (version 2.0, (41)) for each individual and the threshold for phase and allele certainty was set at 95%. If a haplotype contains SNP(s) which had uncertain phase or allele (0.8% of cases), it was treated as missing data in the recoding. Mendelian inconsistencies (< 0.4%) from resulted genotypes were resolved by removal of the genotype in question.

Lipoprotein measurements

Measured lipid and apolipoprotein phenotypes at SFBR included concentrations of HDL and LDL (i.e., nonHDL) cholesterol and several apolipoproteins: ApoAI, ApoB and ApoE (8, 32). Lipoprotein size distribution phenotypes were based on gradient gel

electrophoresis and Sudan black B staining as previously described (46). Triglyceride levels on the basal diet were measured at Children's Hospital Oakland Research Institute.

Acknowledgements

This work was supported in part by the NIH-NHLBI Programs for Genomic Application Grant HL66681, Grant HL071954A through the U.S. Department of Energy under contract no. DE-AC03-76SF00098, P01-HL28972, and P51-RR13086.

References:

1. Korstanje, R. and Paigen, B. (2002) From QTL to gene: the harvest begins. *Nat Genet,* **31,** 235-6.
2. Sugiyama, F., Churchill, G.A., Higgins, D.C., Johns, C., Makaritsis, K.P., Gavras, H. and Paigen, B. (2001) Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics,* **71,** 70-7.
3. Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet,* **33 Suppl,** 228-37.
4. Risch, N.J. (2000) Searching for genetic determinants in the new millennium. *Nature,* **405,** 847-56.
5. Rogers, J. and Hixson, J.E. (1997) Baboons as an animal model for genetic studies of common human disease. *Am J Hum Genet,* **61,** 489-93.
6. Rogers, J., Mahaney, M.C., Witte, S.M., Nair, S., Newman, D., Wedel, S., Rodriguez, L.A., Rice, K.S., Slifer, S.H., Perelygin, A. *et al.* (2000) A genetic linkage map of the baboon (Papio hamadryas) genome based on human microsatellite polymorphisms. *Genomics,* **67,** 237-47.
7. Kammerer, C.M., Cox, L.A., Mahaney, M.C., Rogers, J. and Shade, R.E. (2001) Sodium-lithium countertransport activity is linked to chromosome 5 in baboons. *Hypertension,* **37,** 398-402.
8. Rainwater, D.L., Kammerer, C.M., Mahaney, M.C., Rogers, J., Cox, L.A., Schneider, J.L. and VandeBerg, J.L. (2003) Localization of genes that control LDL size fractions in baboons. *Atherosclerosis,* **168,** 15-22.
9. Groenendijk, M., Cantor, R.M., de Bruin, T.W. and Dallinga-Thie, G.M. (2001) The apoAI-CIII-AIV gene cluster. *Atherosclerosis,* **157,** 1-11.

10. Pennacchio, L.A., Olivier, M., Hubacek, J.A., Cohen, J.C., Cox, D.R., Fruchart, J.C., Krauss, R.M. and Rubin, E.M. (2001) An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science,* **294,** 169-73.
11. Aouizerat, B.E., Kulkarni, M., Heilbron, D., Drown, D., Raskin, S., Pullinger, C.R., Malloy, M.J. and Kane, J.P. (2003) Genetic analysis of a polymorphism in the human apoA-V gene: effect on plasma lipids. *J Lipid Res,* **44,** 1167-73.
12. Evans, D., Buchwald, A. and Beil, F.U. (2003) The single nucleotide polymorphism -1131T>C in the apolipoprotein A5 (APOA5) gene is associated with elevated triglycerides in patients with hyperlipidemia. *J Mol Med,* **81,** 645-54.
13. Lai, C.Q., Tai, E.S., Tan, C.E., Cutter, J., Chew, S.K., Zhu, Y.P., Adiconis, X. and Ordovas, J.M. (2003) The apolipoprotein A5 locus is a strong determinant of plasma triglyceride concentrations across ethnic groups in Singapore. *J Lipid Res*.
14. Tas, S. (1989) Strong association of a single nucleotide substitution in the 3'-untranslated region of the apolipoprotein-CIII gene with common hypertriglyceridemia in Arabs. *Clin Chem,* **35,** 256-9.
15. Dammerman, M., Sandkuijl, L.A., Halaas, J.L., Chung, W. and Breslow, J.L. (1993) An apolipoprotein CIII haplotype protective against hypertriglyceridemia is specified by promoter and 3' untranslated region polymorphisms. *Proc Natl Acad Sci U S A,* **90,** 4562-6.
16. Zeng, Q., Dammerman, M., Takada, Y., Matsunaga, A., Breslow, J.L. and Sasaki, J. (1995) An apolipoprotein CIII marker associated with hypertriglyceridemia in Caucasians also confers increased risk in a west Japanese population. *Hum Genet,* **95,** 371-5.
17. Surguchov, A.P., Page, G.P., Smith, L., Patsch, W. and Boerwinkle, E. (1996) Polymorphic markers in apolipoprotein C-III gene flanking regions and hypertriglyceridemia. *Arterioscler Thromb Vasc Biol,* **16,** 941-7.
18. Plump, A.S., Azrolan, N., Odaka, H., Wu, L., Jiang, X., Tall, A., Eisenberg, S. and Breslow, J.L. (1997) ApoA-I knockout mice: characterization of HDL metabolism in homozygotes and identification of a post-RNA mechanism of apoA-I up-regulation in heterozygotes. *J Lipid Res,* **38,** 1033-47.
19. Maeda, N., Li, H., Lee, D., Oliver, P., Quarfordt, S.H. and Osada, J. (1994) Targeted disruption of the apolipoprotein C-III gene in mice results in hypotriglyceridemia and protection from postprandial hypertriglyceridemia. *J Biol Chem,* **269,** 23610-6.
20. Mackelprang, R., Carlson, C.S., Subrahmanyan, L., Livingston, R.J., Eberle, M.A. and Nickerson, D.A. (2002) Sequence variation in the human T-cell receptor loci. *Immunol Rev,* **190,** 26-39.
21. Carlson, C.S., Eberle, M.A., Rieder, M.J., Smith, J.D., Kruglyak, L. and Nickerson, D.A. (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat Genet,* **33,** 518-21.
22. Reich, D.E., Gabriel, S.B. and Altshuler, D. (2003) Quality and completeness of SNP databases. *Nat Genet,* **33,** 457-8.

23. Kruglyak, L. and Nickerson, D.A. (2001) Variation is the spice of life. *Nat Genet,* **27,** 234-6.
24. Gibbs, R.A. and Belmont, J.W. and Hardenbol, P. and Willis, T.D. and Yu, F. and Yang, H. and Ch'ang, L.Y. and Huang, W. and Liu, B. and Shen, Y. *et al.* (2003) The International HapMap Project. *Nature,* **426,** 789-96.
25. Cardon, L.R. and Abecasis, G.R. (2003) Using haplotype blocks to map human complex trait loci. *Trends Genet,* **19,** 135-40.
26. Olivier, M., Wang, X., Cole, R., Gau, B., Kim, J., Rubin, E.M. and Pennacchio, L.A. (2004) Haplotype analysis of the apolipoprotein gene cluster on human chromosome 11. *Genomics*.
27. Pennacchio, L.A., Olivier, M., Hubacek, J.A., Krauss, R.M., Rubin, E.M. and Cohen, J.C. (2002) Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. *Hum Mol Genet,* **11,** 3031-8.
28. Endo, K., Yanagi, H., Araki, J., Hirano, C., Yamakawa-Kobayashi, K. and Tomura, S. (2002) Association found between the promoter region polymorphism in the apolipoprotein A-V gene and the serum triglyceride level in Japanese schoolchildren. *Hum Genet,* **111,** 570-2.
29. Nabika, T., Nasreen, S., Kobayashi, S. and Masuda, J. (2002) The genetic effect of the apoprotein AV gene on the serum triglyceride level in Japanese. *Atherosclerosis,* **165,** 201-4.
30. Ribalta, J., Figuera, L., Fernandez-Ballart, J., Vilella, E., Castro Cabezas, M., Masana, L. and Joven, J. (2002) Newly identified apolipoprotein AV gene predisposes to high plasma triglycerides in familial combined hyperlipidemia. *Clin Chem,* **48,** 1597-600.
31. Talmud, P.J., Hawe, E., Martin, S., Olivier, M., Miller, G.J., Rubin, E.M., Pennacchio, L.A. and Humphries, S.E. (2002) Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. *Hum Mol Genet,* **11,** 3039-46.
32. Rainwater, D.L., Kammerer, C.M., Carey, K.D., Dyke, B., VandeBerg, J.F., Shelledy, W.R., Moore, P.H., Jr., Mahaney, M.C., McGill, H.C., Jr. and VandeBerg, J.L. (2002) Genetic determination of HDL variation and response to diet in baboons. *Atherosclerosis,* **161,** 335-43.
33. Rainwater, D.L., Kammerer, C.M., Hixson, J.E., Carey, K.D., Rice, K.S., Dyke, B., VandeBerg, J.F., Slifer, S.H., Atwood, L.D., McGill, H.C., Jr. *et al.* (1998) Two major loci control variation in beta-lipoprotein cholesterol and response to dietary fat and cholesterol in baboons. *Arterioscler Thromb Vasc Biol,* **18,** 1061-8.
34. Cheng, M.L., Woodford, S.C., Hilburn, J.L. and VandeBerg, J.L. (1986) A novel system for storage of sera frozen in small aliquots. *J Biochem Biophys Methods,* **13,** 47-51.
35. Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res,* **8,** 175-85.
36. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res,* **8,** 186-94.
37. Gordon, D., Abajian, C. and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res,* **8,** 195-202.

38. Nickerson, D.A., Tobe, V.O. and Taylor, S.L. (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res,* **25,** 2745-51.
39. Nickerson, D.A., Taylor, S.L., Weiss, K.M., Clark, A.G., Hutchinson, R.G., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E. and Sing, C.F. (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet,* **19,** 233-40.
40. Rieder, M.J., Taylor, S.L., Clark, A.G. and Nickerson, D.A. (1999) Sequence variation in the human angiotensin converting enzyme. *Nat Genet,* **22,** 59-62.
41. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet,* **68,** 978-89.
42. Zhang, K. and Jin, L. (2003) HaploBlockFinder: haplotype block analyses. *Bioinformatics,* **19,** 1300-1.
43. Schaid, D.J., McDonnell, S.K., Wang, L., Cunningham, J.M. and Thibodeau, S.N. (2002) Caution on pedigree haplotype inference with software that assumes linkage equilibrium. *Am J Hum Genet,* **71,** 992-5.
44. Rohde, K. and Fuerst, R. (2001) Haplotyping and estimation of haplotype frequencies for closely linked biallelic multilocus genetic phenotypes including nuclear family information. *Hum Mutat,* **17,** 289-95.
45. Abecasis, G.R., Cardon, L.R. and Cookson, W.O. (2000) A general test of association for quantitative traits in nuclear families. *Am J Hum Genet,* **66,** 279-92.
46. Rainwater, D.L., Moore, P.H., Jr., Shelledy, W.R., Dyer, T.D. and Slifer, S.H. (1997) Characterization of a composite gradient gel for the electrophoretic separation of lipoproteins. *J Lipid Res,* **38,** 1261-6.

Figure legend:

Figure 1: r^2 clustering of 248 SNPs in the baboon gene cluster region. SNPs are clustered such that highly correlated sites (with r^2 >0.8) are placed near one another. *: indicates SNPs which are chosen as representative SNPs used in genotyping 96 baboons for haplotype development. Both visual genotype and clustering were done using VG2 program developed in University of Washington.

Figure 2: Haplotype block and Common Haplotypes in the APOA1/C3/A4/A5 gene cluster region

Genotype data of 31 SNPs for 96 pedigreed baboons were used to identify haplotype blocks in the gene cluster region using HaploblockFinder (see materials and methods). A scale was shown on the top for the gene cluster examined. All genes in this region are depicted by black horizontal block arrows. Gene names are given above the arrow. The approximate location of each SNP is indicated by the vertical lines.

The blocks, represented by gray squares, were generated using the chromosome coverage definition with the threshold set at 0.8. The corresponding haplotypes within each block are depicted by the diagrams in the lower panel. The numbers 1, 2, 3, or 4 represent the bases A, T, C, G, respectively. Lines between haplotype blocks represent connections found in the samples. The 16 SNPs used for genotyping the 634 baboons in association study were marked by vertical arrows. The percentage of chromosomes being covered by listed haplotypes was also shown for each block.

Table 1: Summary of SNPs identified in the 68 Kb of baboon *APOA1/C3/A4/A5* gene cluster

	SNP density (1 per bp)				Minor allele frequency (%)			
Total SNPs (in/dels*)	Average	UTR	coding	intron/intergenic	Average	UTR	coding	intron/intergenic
363 (18*)	187	276	626	177	17.3	18.0	19.3	17.2

Table 2: SNPs used for genotyping 635 pedigreed baboons.

Marker	position*	minor allele frequency	No. of chromosomes	Heterozygosity	HWE P-value
SNP 1	6007	0.3477	1228	0.4540	0.593
SNP 2	6619	0.3179	1164	0.4340	0.174
SNP 3	9077	0.4221	1156	0.4883	0.718
SNP 4	10816	0.0976	1188	0.1764	0.941
SNP 5	16987	0.1707	1066	0.2834	0.287
SNP 6	21564	0.2425	1204	0.3677	0.392
SNP 7	24109	0.1391	1186	0.2397	0.781
SNP 8	25935	0.4619	1182	0.4975	0.993
SNP 9	37839	0.0826	1198	0.1517	0.995
SNP 10	49685	0.1906	1186	0.3087	0.233
SNP 11	54530	0.1112	1232	0.1978	0.045
SNP 12	55532	0.4119	1146	0.4849	0.905
SNP 13	55905	0.1305	1180	0.2271	0.249
SNP 14	56390	0.1137	1170	0.2017	0.057
SNP 15	57686	0.1072	1194	0.1916	0.836
SNP 16	57963	0.3232	1154	0.4379	0.234

* SNP position refers to its relative position in the 68 Kb consensus sequences spanning the baboon gene cluster.

Table 3: QTDT tests of 16 markers within the *APOA1/C3/A4/A5* gene cluster for association with 9 traits (P < 0.01)

Traits	Markers	P value	
		asymptotic	empirical
HDL-cholesterol (chow)	SNP 13	0.0067	0.014
LDL PPD NM (chow)	SNP 3	0.0085	>0.1
	SNP 8	0.0044	>0.1
	SNP 11	0.0098	>0.1
Median diam BetaLP NM (chow)	SNP 10	0.0022	>0.1
LDL PPD NM (LCHF)	SNP 4	0.0091	>0.1
Median diam BetaLP NM (LCHF)	SNP 4	0.0077	>0.1

*: 16 markers are listed in table 2. See "Lipoprotein measurements" in "materials and methods" for traits tested. No significant association of any marker for TG was found.

Table 4: QTDT tests of ApoA1, ApoC3 and ApoA5 Haplotypes for Association with TG and HDL-cholesterol

Traits	Marker SNPs[a]	P value[b]		
		Global	Asymptotic	Empirical
TG	3 4 5 6			
	2 SNP/ A1[c]	>0.1	>0.1	>0.1
	2 SNP/ C3	0.0142	0.0039	0.056
	4 SNP/ A1-C3	0.0215	0.0008	0.002
HDL-Cholesterol	12 13 14			
	2 SNP/ A5	0.0778	0.023	0.046
	2 SNP/ A5	0.0623	0.019	0.033
	3 SNP/ A5	>0.1	0.034	0.065

a: SNP numbering is the same as in table 2.
b: All P values refer to haplotypes showing most significant association. Unreported haplotypes had P > 0.1 in all cases. Global, asymptotic and empirical p values all refer to the same haplotype.
c: A1is a abbreviation for ApoA1, C3 for ApoC3, A5 for ApoA5

Fig. 1 r2 clustering of 248 SNPs in the baboon APOA1/C3/A4/A5 gene cluster region

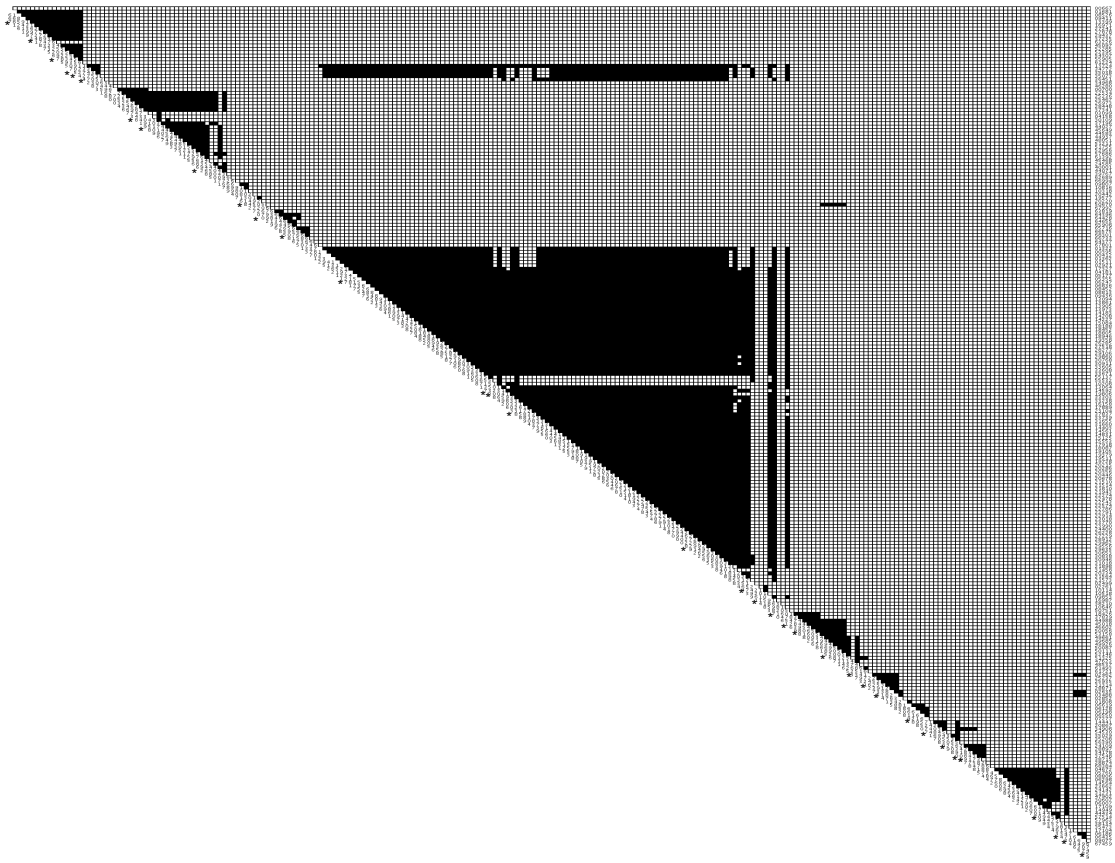

Figure 2: Haplotype blocks and Common Haplotypes in the APOA1/C3/A4/A5 gene cluster region