Genome-wide association studies in Samoans give insight into the genetic etiology of fasting serum lipid levels.

Short title: Carlson, GWAS of Lipids in Samoans

Jenna C. Carlson, PhD[1]; Daniel E. Weeks, PhD[1,2]; Nicola L. Hawley[3], PhD; Guangyun Sun, PhD[4]; Hong Cheng, BS[4]; Take Naseri, MBBS, MPH[5]; Muagututi'a Sefuiva Reupena, MA[6]; Ranjan Deka, PhD[4]*; Stephen T. McGarvey, PhD, MPH[7,8]*; Ryan L. Minster, PhD, MSIS[1]*

1. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA.
2. Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA.
3. Department of Epidemiology, School of Public Health, Yale University, New Haven, CT, USA.
4. Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH, USA.
5. Ministry of Health, Government of Samoa, Apia, Samoa.
6. Bureau of Statistics, Government of Samoa, Apia, Samoa.
7. International Health Institute and Department of Epidemiology, School of Public Health, Brown University, Providence, RI, USA.
8. Department of Anthropology, Brown University, Providence, RI, USA.

* joint senior authors

Corresponding Author:

Ryan L. Minster, PhD, MSIS

Department of Human Genetics

Graduate School of Public Health

University of Pittsburgh

Pittsburgh, PA 15261, USA.

Email: rminster@pitt.edu

Phone: 412-624-6928.

Fax: 412-624-3020.

Journal Subject Codes: Lipids and Cholesterol; Genetic, Association Studies; Genetics

ABSTRACT

Background: The current understanding of the genetic architecture of lipids have largely come from genome-wide association studies. To date, few studies have examined the genetic architecture of lipids in Polynesians, and none have in Samoans, whose unique population history, including many population bottlenecks, may provide insight into the biological foundations of variation in lipid levels.

Methods and Results: Here we performed a genome-wide association study of four fasting serum lipid levels: total cholesterol (TC), high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides (TG) in a sample of 2,849 Samoans, with validation genotyping for associations in a replication cohort comprising 1,798 Samoans and American Samoans. We identified multiple genome-wide significant associations ($P < 5 \times 10^{-8}$) previously seen in other populations – *APOA1* with TG, *CETP* with HDL, and *APOE* with TC and LDL – and several suggestive associations ($P < 1 \times 10^{-5}$), including an association of variants downstream of *MGAT1* and *RAB21* with HDL. For variants near *APOE* and *APOA1*, we observed different association signals than what has been previously reported in non-Polynesian populations.

Conclusions: The association with several known lipid loci combined with the newly-identified associations with variants near *MGAT1* and *RAB21* suggest that while some of the genetic architecture of lipids is shared between Samoans and other populations, part of the genetic etiology may be Polynesian-specific.

Key words: complex trait genetics; founder population genetics; lipids; genome-wide association study; Polynesia

INTRODUCTION

The Samoan Islands, comprising both the U.S. Territory of American Samoa (American Samoan) and the Independent State of Samoa (Samoa), have experienced a rise in prevalence of cardiovascular disease and other non-communicable diseases in the last 30 years partly due to economic modernization, rapid urbanization, and lifestyle changes such as increased caloric intake and sedentary behavior[1–3]. Our 2010 population-based survey in Samoa, which gathered the "discovery cohort" studied further here, found that many Samoans are at elevated risk of cardiovascular disease based on known risk factors—increased total cholesterol (TC), low-density lipoprotein cholesterol (LDL), and triglycerides (TG), as well as decreased high-density lipoprotein (HDL)[4–6]—with 47% of the 2,938 adult Samoans examined in the study having elevated TC (≥ 5.2 mmol/L), 88% of men and 91% of women having elevated LDL (> 2.59 mmol/L), and 43% of women having low HDL (< 1.29 mmol/L)[1].

Our current understanding of the genetic component of serum lipid level variation has been largely due to genome-wide association studies (GWAS)[7]. The Global Lipids Genetics Consortium found strong evidence for 157 loci associated with one or more of these traits using a sample of 188,577 individuals of European, East Asian, South Asian, and African ancestry[8]. However, few GWAS of serum lipid levels have been conducted in Pacific Islanders[9,10] and to our knowledge only one has included a small number of Polynesians[11]. Previous studies have estimated the heritability of serum lipid levels in Samoans, ranging from 16% for HDL to 23% for TG, and have identified genetic susceptibility loci via linkage analysis[12], warranting further study of the genetic

architecture of serum lipid levels in Samoans. Samoans are a genetically-isolated founder population, with unique evolutionary history, making them particularly useful in genomic studies[13,14]. Thus, genomic studies of serum lipid levels could reveal novel lipid-altering loci specific to Pacific Islander populations, as well as highlight susceptibility loci shared with global populations.

Here we report the results of a GWAS of fasting TC, LDL, HDL, and TG in up to 2,849 individuals from independent Samoa followed by replication in up to 1,798 individuals from independent Samoa and American Samoa, as part of the ongoing GWAS of adiposity-related traits in the Samoan Islands[1]. We identified multiple genome-wide significant associations previously seen in other populations – *APOA1* with TG, *CETP* with HDL, and *APOE* with TC and LDL – and several suggestive associations, including an association between variants downstream of *MGAT1* and *RAB21* with HDL.

METHODS

Discovery Cohort and Genotyping

The discovery cohort data are available from dbGaP (accession number: phs000914.v1.p1). The discovery cohort of 2,849 individuals is drawn from a population-based sample recruited from Samoa in 2010 (Table 1). The sample selection, data collection methods, and phenotyping, including the laboratory assays for serum lipid and lipoprotein levels, have been previously reported[1,13]. Briefly, serum lipid levels were derived from fasting whole blood samples collected after a minimum 10-hour overnight fast. Genotyping was performed using Genome-Wide Human SNP 6.0

arrays (Affymetrix). Extensive quality control was conducted on the basis of a pipeline developed by Laurie et al[15]. Additional details for sample genotyping and genotype quality control are described in Minster et al[13]. This study was approved by the institutional review board of Brown University and the Health Research Committee of the Samoa Ministry of Health. All participants gave informed consent. Imputation was not performed in this study because prior experience with this population using extant imputation panels showed that the resulting imputed genotypes did not correlate well with observed genotypes[13].

Replication Cohort and Genotyping

The replication cohort of 1,798 individuals contains two sample sets recruited from Samoa and American Samoa (Table 1). Although there is substantial economic variation across the two polities, with American Samoans generally having a higher standard of living, the Samoans from both territories form a single socio-cultural unit with frequent exchange of mates; genetically, they represent a single homogenous population[3,16]. The first sample set, referred to as the 1990–95 replication sample set, contains 716 unrelated individuals derived from a longitudinal study of adiposity and cardiovascular disease risk factors among adults from American Samoa and Samoa (Table 1). Detailed descriptions of the sampling and recruitment have been reported previously[17–19]. Briefly, participants were recruited from 46 villages and worksites in American Samoa in 1990 and 9 villages in Samoa (Western Samoa, at the time of recruitment) in 1991 and followed up four years later in 1994 and 1995, respectively. All participants were, at baseline, free of self-reported history of heart disease,

hypertension, or diabetes. This study was approved by the institutional review board of the Miriam Hospital, Providence, RI. All participants gave informed consent. The second sample set, referred to as the 2002–03 replication sample set, contains 1,082 individuals from American Samoa and Samoa and was drawn from an extended family-based genetic linkage analysis of cardiometabolic traits (Table 1)[12,20–23]. Probands and relatives were unselected for obesity or related phenotypes. The recruitment process and criteria used for inclusion in this study have been described in detail previously[21,23]. This study was approved by the institutional review board of Brown University and research ethics review committees in both Samoa and American Samoa. All participants gave informed consent. Imputation was not performed in these studies for the same rationale as the discovery cohort above, but also because genome-wide marker data was not available for the samples in these studies.

In both replication sample sets, blood samples were collected in the morning after a minimum of 10 hours fasting, from which serum lipid levels were derived using assay methods published previously[12,17]. Genotyping of variants selected for validation in the replication cohort (described below) was performed using custom-designed TaqMan OpenArray Real-Time PCR assays (Applied Biosystems). SNPs that could not be genotyped using OpenArray assays were genotyped individually using TaqMan SNP Genotyping assays (Applied Biosystems). Eight variants could not be genotyped due to technical difficulties.

Statistical Analyses

Prior to association analyses, residuals were generated for all 4 lipid traits. First, traits were transformed to normality with the Box–Cox power transformation; secondly, model selection was performed using step-wise linear regression with covariates previously associated with serum lipid levels: age, age^2, sex, log-transformed BMI, fasting glucose, smoking status, farming status (as a measure of physical activity), and interactions between age, age^2, and sex. All models included age and sex as covariates. The final TC model additionally adjusted for age^2, age × sex, and age^2 × sex; the final LDL and TG models additionally adjusted for age^2 and age^2 × sex; the final HDL model adjusted only for age and sex. Preliminary associations were performed without consideration of hypolipidemic medication use as it was not measured; however, participants did self-report use of heart disease medication. Sensitivity analysis revealed that this self-reported use of medication to treat heart disease was significantly associated with TC and LDL (results not shown); individuals reporting such medication use (n = 17) were excluded from analyses. However, these individuals were improperly included in preliminary analyses, affecting the prioritization of validation genotyping efforts. All results presented are those of the corrected analyses, removing the individuals with heart disease medication use.

Association between lipid residuals and autosomal genotypes of 659,492 SNPs with minor allele frequency (MAF) ≥ 0.05 and Hardy-Weinberg Equilibrium (HWE) test P value ≥ 5×10^{-5} was assessed using linear mixed modelling in GenABEL, including previously-derived empirical kinship estimates to adjust for subject relatedness[13,24]. The

association between X-chromosome genotypes and the lipid phenotypes were calculated in GenABEL, without adjustment using the empirical kinship estimates. Sensitivity analyses were performed by excluding BMI and fasting glucose from the covariates, marginally changing the resulting association *P* values but not altering them by more than one order of magnitude nor changing the set of loci that were most significant (results not shown). Genomic inflation due to population stratification and cryptic relatedness was assessed by estimating λ_{GC} using the lower 90% of the *P* value distribution[25]. GWAS *P* values in the discovery cohort (P_G) were compared to a threshold for genome-wide significance of $P_G < 5 \times 10^{-8}$ and a suggestive association threshold of $P_G < 1 \times 10^{-5}$. Statistical power to detect signals at these thresholds was calculated using the Genetic Power Calculator[26].

Gene-set enrichment analysis with MAGENTA was also performed to identify any biological pathways enriched for discovery association signals[27]. Briefly, gene scores were obtained from the most significant *P* value among SNPs located within each gene using the association results from each lipid GWAS. Genes scores were adjusted for confounding factors including gene size, number of variants, and linkage disequilibrium-related properties by using step-wise multiple linear regression. The 95th percentile of all gene scores was used as the enrichment cutoff for each trait[28]. Gene-set enrichment *P* values were obtained for highly ranked gene scores. Gene sets were obtained from Gene Ontology (April 2010), pathway information from the Ingenuity (June 2008) and KEGG (June 2010), and biological processes and molecular function from PANTHER (January 2010).

For each of the lipid traits, the INRICH program[29] was used to test for enrichment of known genes (as constructed from Teslovich et al.[30] and Willer et al[8]). INRICH tests if more known genes are contained in associated intervals than expected by chance, using permutation based on 1 million replicates to generate experiment-wide empirical *P* values. For each lipid trait, we defined the associated intervals as 100 kb intervals centered on the most significant SNP within association peaks with $P_G < 1 \times 10^{-4}$.

We selected 21 regions demonstrating at least suggestive association for association validation in the replication cohort. However, an additional 10 regions which should have selected for follow-up were not because they were selected based on earlier analyses that included participants taking heart disease medication. The variant from each locus with smallest *P* value across the four lipid scans (defined as 1 Mb windows surrounding the peak SNP) or a proxy SNP was selected as representative of the locus for replication genotyping.

Statistical association was measured in the 1990–95 and 2002–03 replication sample sets independently and results were combined using meta-analysis (see below). Association analyses for both sample sets were performed using GenABEL[31] in R[32], adjusting for polity (American Samoa or Samoa); the 2002-03 sample set was additionally adjusted using expected kinship, as derived from familial pedigree information[33].

Prior to meta-analysis, quality control was performed using EasyQC to check for strand and allele frequency consistency[34]. *P* value-based meta-analysis using sample sizes as weights was performed using METAL[35] to generate two *P* values: one for the meta-analysis of the two replication cohorts (P_R) and one for the replication cohorts and discovery sample together (P_{GR}). Resulting meta-analysis signals were evaluated based on genome-wide significance and suggestive thresholds (as described above) and by the contribution of the replication sample to the signal. Effect directions for meta-analysis results of peak SNPs were qualitatively compared to those of previously reported lead SNPs.

For ease of reference, any locus identified here with a corresponding signal within 1 mega base pairs (Mb) in a prior lipid study is referred to by the previously prescribed locus name[8,10]; for loci not previously associated with lipid traits, the symbol of the gene nearest the peak SNP in the locus or the hyphen-separated symbols of the nearest two genes is used as the locus label.

RESULTS

The demographic, anthropometric, and biochemical characteristics of the 2,849 participants composing the discovery cohort for this GWAS of serum lipids levels and the 1,798 participants composing the replication cohort are presented in Table 1. A detailed description of the discovery cohort and its trends compared to the historical sample sets making up the replication cohort has been previously reported[1]. Briefly, the average age was similar for all cohorts; average BMI was higher among women

compared to men and in American Samoa compared to Samoa; average BMI for men and women in Samoa was higher in more recent studies; average lipid levels are largely similar across cohorts with minor exceptions.

We assessed 659,492 unique genome-wide markers for association with 4 traits—TC, HDL, LDL and TG—in up to 2,849 Samoans in the discovery cohort. Relatedness within the discovery cohort was well-controlled using the empirical kinship coefficients; λ_{GC} ranged between 1.03 and 1.07 for the four lipid traits (Supplementary Figures S2, S6, S10 and S14). We observed 38 genome-wide suggestive or significant associations across 31 loci from the 4 GWAS (Figure1; Table 2; Supplementary Figures S1, S5, S9 and S13; Supplementary Tables S1, S7, S10 and S13).

Genome-wide significant association was observed in the discovery cohort between all traits and markers near *APOE*: TC and rs4420638 ($P_G = 2.67 \times 10^{-16}$, Figure 2E); HDL and rs4420638 ($P_G = 9.07 \times 10^{-9}$, Figure 2F); LDL and rs1160985 ($P_G = 2.61 \times 10^{-20}$; Figure 2G); and TG and rs4420638 ($P_G = 7.44 \times 10^{-10}$, Figure 2H). Additionally, HDL was associated with markers near *CETP* (rs289708, $P_G = 1.19 \times 10^{-11}$), and TG, with *APOA1* (rs6589566, $P_G = 3.98 \times 10^{-18}$, Figure 2C). Suggestive association was observed between lipid levels and markers at an additional 28 loci, including the *MGAT1* and *RAB21* loci and HDL (Figure 2A,D), and *APOA1* with TC (rs3741298, $P_G = 1.63 \times 10^{-7}$, Figure 2B). We had 80% power at $\alpha = 1 \times 10^{-5}$ and $\alpha = 5 \times 10^{-8}$ to detect SNPs that account for 1.0% and 1.5%, respectively, of the residual variance in a phenotype.

Gene-set enrichment analysis with MAGENTA highlighted, at a < 5% false-discovery rate (FDR), several lipid homeostasis pathways and gene ontologies for HDL and TG (Supplementary Tables S6 and S14). Four gene sets were below the FDR for both HDL and TG: HDL particle remodeling, reverse cholesterol transport, cholesterol efflux, and phospholipid efflux. An additional 12 gene sets were implicated for HDL and 3 gene sets for TG. The HDL particle remodeling and reverse cholesterol transport gene sets had significant enrichment for TC (Supplementary Table S3), and a single gene set was implicated with LDL, the amylase pathway (Supplementary Table S11). All 4 traits had significant enrichment for known TC, HDL, LDL, and TG loci using the INRICH method (Supplementary Tables S5, S9, S12, and S15).

Validation of peak SNPs was attempted for 21 loci. At loci with multiple associated variants, the most significant variant was chosen as representative of the locus. For some loci, the exclusion of participants using self-reported heart disease medication resulted in a different peak SNP. Thus, for the *APOE* locus rs1160985 was genotyped instead of rs4420638 ($P_G = 2.67 \times 10^{-16}$ for TC, $P_G = 9.07\times 10^{-9}$ for HDL, and $P_G = 7.44 \times 10^{-10}$ for TG); for the *APOA1* locus rs964184 was genotyped instead of rs3741298 ($P_G = 1.63 \times 10^{-7}$ for TC) or rs6589566 ($P_G = 3.98 \times 10^{-18}$ for TG); for the *MGAT1* locus rs1038143 was genotyped instead of rs249356 ($P_G = 1.06 \times 10^{-6}$ for HDL); for the *APOB* locus rs754523 was genotyped instead was rs1469513 ($P_G = 2.71 \times 10^{-6}$ for LDL).

We successfully genotyped the peak SNP, or a proxy SNP, in the replication cohorts for 15 loci. Two loci (*APOA1* with TG, $P_{GR} = 1.81 \times 10^{-29}$; *APOE* with TC, $P_{GR} = 4.29 \times 10^{-21}$, and LDL, $P_{GR} = 1.53 \times 10^{-27}$) demonstrated genome-wide significant associations in the discovery-replication meta-analysis (Table 2 and Supplementary Tables S1, S7, S10 and S13). An additional 4 associations demonstrated evidence of replication with consistent directions of effect and suggestive joint P_{GR} values (*GCKR* with TG, $P_{GR} = 5.62 \times 10^{-8}$; *MGAT1* with HDL, $P_{GR} = 2.91 \times 10^{-7}$; *APOA1* with TC, $P_{GR} = 1.72 \times 10^{-6}$; *RAB21* with HDL, $P_{GR} = 5.92 \times 10^{-7}$). Three associations had suggestive joint P_{GR} values driven by the discovery associations only (*APOB* with LDL, $P_{GR} = 5.81 \times 10^{-6}$; *LIPC* with HDL, $P_{GR} = 9.15 \times 10^{-7}$; *CDH4* with HDL, $P_{GR} = 8.77 \times 10^{-6}$); associations at *APOB* and *CDH4* had consistent directions of effect. Among the remaining loci with at least suggestive association in the discovery sample, but not in the discovery-replication meta-analysis, consistent effect directions were also seen for TC and *APOB* and *ZHX2*; LDL and *ALG10* and *CPNE8* (Table 2 and Supplementary Tables S1, S7, S10 and S13).

We compared the directions of effect to those previously reported in Willer et al.[8] and Teslovich et al.[30] for *APOB*, *GCKR*, *APOA1*, *LIPC*, *LIPG*, and *APOE* (i.e., genome-wide suggestive loci that have been previously associated with lipid traits). We observed a consistent direction of effect for the representative SNP for all associations except for *GCKR* and TG, *APOA1* and TG, and *LIPG* and HDL (Table 2).

The effect allele frequencies in the two samples—discovery and replication—were largely similar for each of the 15 successfully genotyped SNPs (Supplementary Tables S1, S7, S10 and S13). However, many loci had markedly different effect allele frequencies between Samoans and other 1000 Genomes populations (Table 2). For example, compared to 1000 Genomes populations, there were higher effect allele frequencies (EAFs) in Samoans for rs964184 near *APOA1* (0.440 vs. < 0.277) and rs1160985 near *APOE* (0.724 vs. < 0.659), and lower EAF for rs1038143 near *MGAT1* (0.691 vs. > 0.852).

DISCUSSION

In this study, we examined four measures of fasting lipid levels—TC, HDL, LDL and TG—for associations with 659,492 SNPs from a genome-wide array in a discovery cohort of 2,849 Samoans, with follow-up genotyping of significant and suggestive findings in a replication cohort comprising 1,798 Samoans from Samoa and American Samoa. Thirty-one loci had at least suggestive evidence of association with one or more lipid traits in the discovery cohort, of which 8 have been reported to be associated with lipid levels previously: *APOB, GCKR, MGAT1, APOA1*, *LIPC*, *CETP, LIPG, and APOE*[8,10,30] although the direction of effect was not consistent with previous results for all loci. Enrichment analyses highlighted known lipid metabolism gene sets and previously associated lipid loci.

We observed a difference in the architecture of the statistical association signals between the 4 lipid traits and variants near *APOE* (Figure 2E-H). The peak SNP for TC and LDL was rs1160985, an intronic variant in *TOMM40* upstream of *APOE*; whereas the peak SNP for HDL and TG was rs4420638, an intergenic variant downstream of *APOC1* and *APOE*. rs1160985 demonstrated evidence of replication for TC and LDL but not for HDL and TG, consistent with the discovery findings. These markers are in low linkage disequilibrium (r^2 = 0.093) and may represent distinct association signals. While this could support a shared genetic etiology for TC and LDL and for HDL and TG in Polynesians, this study was not positioned to adequately capture the association signal present at this locus. Future studies with sequencing or imputation of the 19q13.2 region will be necessary to dissect the genetic architecture of *APOE* and lipid levels in Polynesians.

Similarly, the structure of the association with *APOA1* (Figure 2B,C) was not identical to what has been previously reported. The minor allele of rs964184, a variant downstream of *APOA1*, has previously been associated with higher TC, LDL, and TG, and lower HDL[8]; in this study, we also found a positive association between this variant and TG, but the opposite effect in TC. While the effect with TC was not genome-wide significant, we did observe a consistent direction of effect between discovery and replication.

We did not observe suggestive or genome-wide significant association with several loci which have figured prominently in multiple lipid GWAS (eg, *LPL*, *LDLR*, *CILP2*, *FADS1*/*2*/*3*, *ANGPTL3*, *SORT1*, *PPP1R3B*, *MIXIPL*, *HNF4A*, *PCSK9*, *GALNT2*,

HMGCR) either because we lacked sufficient power to detect their effects, the effects are negligible in Samoans, or the allele frequencies of associated variants are different enough in Samoans to hinder effect detection. However, it is important to note that this study was not designed to evaluate the effect of known lipids loci in Samoans, nor were previously-associated loci examined specifically.

We detected and replicated a suggestive association between HDL and a variant on 5q35.3 (Figure 2A). While the peak SNP lies within an intron of *BTNL8*, the variant selected for follow-up genotyping is intergenic, downstream of *MGAT1*. A suggestive association between variants near *MGAT1* and HDL in a GWAS in the Micronesian population of Kosrae has been previously reported[10]. Although there is no evidence of association between *MGAT1* and lipids as reported in prior studies of non-Pacific Islanders, variation near *MGAT1* has been associated with BMI, serum fatty acid levels and composition, and glucose response in Europeans[36–38]. The encoded MGAT enzyme plays a major role in the absorption of dietary fat in the intestine[39]. Due to the lower frequency of the HDL-associated risk variant observed in Samoans compared to 1000 Genomes populations, it is plausible that variation near *MGAT1* may have a unique role in or a stronger effect on the lipid metabolism of Pacific Islanders.

We also detected and replicated a novel suggestive association between HDL and a variant downstream of *RAB21* (Figure 2D). Unlike the variant downstream of *MGAT1*, we observed similar allele frequencies between Samoans and 1000 Genomes populations (ie, < 20% difference between Samoans and another 1000 Genomes

population) in the HDL-associated variant downstream of *RAB21*. This region, 12q21, was previously seen in linkage analysis with both univariate and bivariate scans of TC and LDL[12]. The individuals included in this linkage analysis are also included in our 2002–2003 replication cohort, however, they do not appear to be driving the association signal near *RAB21* (Supplemental Table S7). Variation near *RAB21* has been previously associated with obesity[40]. *RAB21* belongs to the family of monomeric GTPases involved in control of cellular membrane trafficking and is involved in the targeted trafficking of integrins and the regulation of cell adhesion and migration[41,42].

This study is limited in drawing conclusions about the genetic architecture of lipids in Samoans, as replication genotyping was unavailable for many loci and due to the lack of genome-wide imputation. Future studies, evaluating the evidence of association between associations seen here in separate cohorts as well fine-mapping loci with genotype imputation (given the availability of a relevant reference panel), are necessary to fully evaluate the genetic architecture of lipids in Samoans.

This is the first GWAS of lipid phenotypes in Samoans, and we observed association with many known lipid loci, which was further supported by the gene-set enrichment analysis highlighting lipid metabolism gene sets. However, the surprising association results with *APOE* and *APOA1*, coupled with evidence of Pacific-Islander-specific associations with *MGAT1* and *RAB21* suggest that only some of the genetic architecture of lipids is shared between Samoans and other populations. Given this

evidence for a differential genetic etiology of lipids in Samoans, further investigation and fine-mapping of lipid loci, especially that across multiple ethnicities, is warranted.

ACKNOWLEDGEMENTS

The authors would like to thank the Samoan participants of the study, local village authorities, and the many Samoan and other field workers over the years. We acknowledge the Samoan Ministry of Health, the Samoa Bureau of Statistics, and the American Samoan Department of Health for their support of this research. We give particular thanks to two research assistants, Melania Selu and Vaimoana Lupematasila, who contributed to the 2010 recruitment and continue to assist us in our work in Samoa.

FUNDING SOURCES

This work was funded by NIH grants R01-HL133040 (R.L.M), R01-HL093093 (S.T.M.), R01-AG09375 (S.T.M.), R01-DK59642 (S.T.M.), P30-ES006096 (S.M. Ho), and R01-DK55406 (R.D.), and by Brown University student research funds. Genotyping was performed in the Core Genotyping Laboratory at the University of Cincinnati, funded by NIH grant P30-ES006096 (S.M. Ho).

DISCLOSURE

None.

REFERENCES

1. Hawley NL, Minster RL, Weeks DE, Viali S, Reupena MS, Sun G, et al. Prevalence of adiposity and associated cardiometabolic risk factors in the samoan genome-wide association study. *Am J Hum Biol* [Internet]. 2014 [cited 2017 Oct 13];26:491–501. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24799123
2. Keighley E., McGarvey ST, Quested C, Mccuddin C, Viali S, Maga UA. Nutrition and health in modernizing Samoans: temporal trends and adaptive perspectives [Internet]. In: Ohtsuka R, Ulijaszek SJ, editors. Health Change in the Asia-Pacific Region: Biocultural and Epidemiological Approaches. Cambridge, NY: Cambridge University Press; 2007. p. 147–191.Available from: http://media.eurekalert.org/aaasnewsroom/MCM/FIL_000000000615/Temporal_trends_in_Samoans_in_nutrition_book_chapter_2007.pdf
3. McGarvey ST. Cardiovascular disease (CVD) risk factors in Samoa and American Samoa, 1990-95. *Pac Health Dialog* [Internet]. 2001;8:157–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12017817
4. Kannel WB, Castelli WP, Gordon T, McNamara PM. Serum cholesterol, lipoproteins, and the risk of coronary heart disease. The Framingham study. *Ann Intern Med* [Internet]. 1971;74:1–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/5539274
5. Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med* [Internet]. 1977;62:707–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/193398
6. Roeters van Lennep J, Westerveld HT, Erkelens DW, van der Wall EE. Risk factors for coronary heart disease: implications of gender. *Cardiovasc Res* [Internet]. 2002 [cited 2018 May 15];53:538–549. Available from: https://academic.oup.com/cardiovascres/article-lookup/doi/10.1016/S0008-6363(01)00388-1
7. Lange LA, Willer CJ, Rich SS. Recent developments in genome and exome-wide analyses of plasma lipids. *Curr Opin Lipidol* [Internet]. 2015;26:96–102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25692345
8. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet* [Internet]. 2013;45:1274–1283. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24097068
9. Burkhardt R, Kenny EE, Lowe JK, Birkeland A, Josowitz R, Noel M, et al. Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. *Arterioscler Thromb Vasc Biol* [Internet]. 2008;28:2078–2084. Available from: http://atvb.ahajournals.org/content/28/11/2078.full.pdf
10. Lowe JK, Maller JB, Pe'er I, Neale BM, Salit J, Kenny EE, et al. Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. *PLoS Genet* [Internet]. 2009 [cited 2018 May 31];5:e1000365. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2628735/pdf/pgen.1000365.pdf

11. Dumitrescu L, Carty CL, Taylor K, Schumacher FR, Hindorff LA, Ambite JL, et al. Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study. *PLoS Genet* [Internet]. 2011;7:e1002138. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3128106/pdf/pgen.1002138.pdf
12. Åberg K, Dai F, Sun G, Keighley E, Indugula SR, Bausserman L, et al. A genome-wide linkage scan identifies multiple chromosomal regions influencing serum lipid levels in the population on the Samoan islands. *J Lipid Res* [Internet]. 2008;49:2169–2178. Available from: http://www.jlr.org/lookup/doi/10.1194/jlr.M800194-JLR200
13. Minster RL, Hawley NL, Su C-T, Sun G, Kershaw EE, Cheng H, et al. A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat Genet* [Internet]. 2016 [cited 2017 Oct 13];48:1049–1054. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27455349
14. Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. *Genome Biol* [Internet]. 2008 [cited 2018 Apr 16];9:109. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18771588
15. Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* [Internet]. 2010;34:591–602. Available from: http://dx.doi.org/10.1002/gepi.20516
16. Deka R, Mc Garvey ST, Ferrell RE, Kamboh MI, Yu LM, Aston CE, et al. Genetic characterization of American and Western Samoans. *Hum Biol* [Internet]. 1994;66:805–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8001911
17. McGarvey ST, Levinson PD, Bausser-Man L, Galanis DJ, Hornick CA. Population change in adult obesity and blood lipids in American Samoa from 1976–1978 to 1990. *Am J Hum Biol* [Internet]. 1993;5:17–30. Available from: http://dx.doi.org/10.1002/ajhb.1310050106
18. Chin-Hong P V, McGarvey ST. Lifestyle incongruity and adult blood pressure in Western Samoa. *Psychosom Med* [Internet]. 1996;58:131–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8849629
19. Galanis DJ, McGarvey ST, Quested C, Sio B, Afele-Fa'amuli SA. Dietary intake of modernizing Samoans: implications for risk of cardiovascular disease. *J Am Diet Assoc* [Internet]. 1999;99:184–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9972185
20. Åberg K, Dai F, Sun G, Keighley ED, Indugula SR, Roberts ST, et al. Susceptibility Loci for Adiposity Phenotypes on 8p, 9p, and 16q in American Samoa and Samoa. *Obesity* [Internet]. 2009;17:518–524. Available from: http://doi.wiley.com/10.1038/oby.2008.558
21. Dai F, Keighley ED, Sun G, Indugula SR, Roberts ST, Åberg K, et al. Genome-wide scan for adiposity-related phenotypes in adults from American Samoa. *Int J Obes* [Internet]. 2007;31:1832–1842. Available from: http://www.nature.com/articles/0803675
22. Åberg K, Dai F, Viali S, Tuitele J, Sun G, Indugula SR, et al. Suggestive linkage detected for blood pressure related traits on 2q and 22q in the population on the Samoan islands. *BMC Med Genet* [Internet]. 2009;10:107. Available from:

http://bmcmedgenet.biomedcentral.com/articles/10.1186/1471-2350-10-107
23. Dai F, Sun G, Åberg K, Keighley ED, Indugula SR, Roberts ST, et al. A Whole Genome Linkage Scan Identifies Multiple Chromosomal Regions Influencing Adiposity-Related Traits among Samoans. *Ann Hum Genet* [Internet]. 2008;72:780–792. Available from: http://doi.wiley.com/10.1111/j.1469-1809.2008.00462.x
24. Chen W-M, Abecasis GR. Family-Based Association Tests for Genomewide Association Scans. *Am J Hum Genet* [Internet]. 2007;81:913–926. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929707638695
25. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* [Internet]. 1999;55:997–1004. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11315092
26. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* [Internet]. 2003;19:149–150. Available from: http://bioinformatics.oxfordjournals.org/content/19/1/149.full.pdf
27. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank* [Internet]. 2015 [cited 2017 Oct 23];13:311–319. Available from: http://online.liebertpub.com/doi/10.1089/bio.2015.0032
28. Segrè A V, Groop L, Mootha VK, Daly MJ, Altshuler D. Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. *PLoS Genet* [Internet]. 2010;6:e1001058. Available from: http://dx.plos.org/10.1371/journal.pgen.1001058
29. Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* [Internet]. 2012 [cited 2018 Jul 13];28:1797–1799. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts191
30. Teslovich TM, Musunuru K, Smith A V., Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* [Internet]. 2010 [cited 2018 Jul 25];466:707–713. Available from: http://www.nature.com/articles/nature09270
31. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* [Internet]. 2007;23:1294–1296. Available from: http://bioinformatics.oxfordjournals.org/content/23/10/1294.abstract
32. R Core Team. R: A language and environment for statistical computing. [Internet]. 2014;Available from: http://www.r-project.org
33. Therneau T, Atkinson E, Sinnwell J, Schaid D, McDonnell S. kinship2: Pedigree functions. 2014;
34. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* [Internet]. 2014;9:1192–1212. Available from: http://www.nature.com/nprot/journal/v9/n5/pdf/nprot.2014.071.pdf
35. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of

genomewide association scans. *Bioinformatics* [Internet]. 2010;26:2190–2191. Available from: http://bioinformatics.oxfordjournals.org/content/26/17/2190.abstract

36. Johansson Å, Marroni F, Hayward C, Franklin CS, Kirichenko A V., Jonasson I, et al. Linkage and Genome-wide Association Analysis of Obesity-related Phenotypes: Association of Weight With the MGAT1 Gene. *Obesity* [Internet]. 2010 [cited 2018 Mar 1];18:803–808. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19851299
37. Jacobsson JA, Rask-Andersen M, Risérus U, Moschonis G, Koumpitski A, Chrousos GP, et al. Genetic variants near the MGAT1 gene are associated with body weight, BMI and fatty acid metabolism among adults and children. *Int J Obes* [Internet]. 2012 [cited 2018 Mar 1];36:119–129. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21304485
38. Del-Aguila JL, Beitelshees AL, Cooper-DeHoff RM, Chapman AB, Gums JG, Bailey K, et al. Genome-wide association analyses suggest NELL1 influences adverse metabolic response to HCTZ in African Americans. *Pharmacogenomics J* [Internet]. 2014 [cited 2018 May 31];14:35–40. Available from: http://www.nature.com/articles/tpj20133
39. Yen C-LE, Stone SJ, Cases S, Zhou P, Farese R V. Identification of a gene encoding MGAT1, a monoacylglycerol acyltransferase. *Proc Natl Acad Sci* [Internet]. 2002 [cited 2018 Mar 1];99:8512–8517. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12077311
40. Turcot V, Lu Y, Highland HM, Schurmann C, Justice AE, Fine RS, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* [Internet]. 2018 [cited 2018 Jun 1];50:26–41. Available from: http://www.nature.com/articles/s41588-017-0011-x
41. Pellinen T, Ivaska J. Integrin traffic. *J Cell Sci* [Internet]. 2006 [cited 2018 Jun 1];119:3723–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16959902
42. Pellinen T, Tuomi S, Arjonen A, Wolf M, Edgren H, Meyer H, et al. Integrin trafficking regulated by Rab21 is necessary for cytokinesis. *Dev Cell* [Internet]. 2008 [cited 2018 Jun 1];15:371–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18804435
43. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* [Internet]. 2010;26:2336–2337. Available from: http://bioinformatics.oxfordjournals.org/content/26/18/2336.full.pdf

TABLES

Table 1. Demographic, anthropometric and blood biochemistry statistics of the genotyped discovery and replication cohorts.

	2010 Discovery Cohort		2002–2003 Replication Sample Set		1994–1995 Replication Sample Set		2002–2003 Replication Sample Set		1994–1995 Replication Sample Set	
	Samoa, n = 2849		Samoa, n = 490		Samoa, n = 468		American Samoa, n = 592		American Samoa, n = 248	
	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
	n = 1703	n = 1146	n = 245	n = 245	n = 243	n = 225	n = 337	n = 255	n = 137	n = 111
age (years)	44.8 (11.10)	45.6 (11.10)	44 (17.00)	40.7 (16.30)	43.3 (8.60)	43.2 (8.90)	43 (16.00)	43.2 (16.60)	43.3 (9.10)	45.5 (10.60)
BMI (kg/m^2)	34.8 (6.70)	31.3 (5.90)	33.2 (7.70)	28.8 (5.40)	31.9 (5.70)	28.8 (5.00)	36.5 (8.40)	33.4 (7.60)	37.4 (6.90)	34.9 (6.00)
total cholesterol (mg/dL)	199.3 (36.10)	200.3 (38.70)	202.3 (35.90)	195.5 (40.60)	199.1 (37.20)	197.4 (35.00)	187.1 (38.60)	189.6 (37.80)	197.8 (31.20)	194.1 (31.00)
HDL (mg/dL)	46.5 (10.80)	43.7 (11.20)	47.1 (10.30)	46.3 (11.20)	43.4 (11.00)	42.5 (11.40)	40.9 (8.80)	40 (8.80)	34.8 (7.60)	32.8 (8.30)
LDL (mg/dL)	130 (32.60)	129.6 (35.30)	130.3 (32.80)	128.6 (37.50)	140.2 (34.80)	133.5 (33.00)	118.5 (33.70)	118.9 (35.10)	136.6 (32.40)	136 (37.50)
trigylcerides (mg/dL)	114.9 (80.50)	139.3 (113.10)	110.9 (58.90)	120.1 (91.30)	85.6 (72.80)	102.8 (97.20)	130.8 (78.10)	200.2 (206.50)	118.9 (63.40)	168.2 (109.00)
fasting glucose (mg/dL)	107.6 (56.20)	107.4 (54.80)	102.2 (46.40)	98.6 (49.80)	90.9 (38.00)	85.3 (18.90)	104.2 (50.30)	108.8 (54.70)	107.3 (49.80)	125.2 (61.70)

Table 2. Suggestive loci and replication genotyping

Total Cholesterol

Locus	SNP	Chr	BP	EA	OA	Dir	p_G	p_R	p_{GR}	Known gene	Traits	SAM	EAS	SAS	EUR	AMR	AFR
APOB	rs754523	2	21311691	G	A	+++	*6.25E-06*	0.178	1.20E-05	*APOB*	C,H,L,T	0.247	0.265	0.140	0.309	0.306	0.201
PDE4D	rs7711093	5	59593138	G	A	++−	*3.01E-06*	0.747	5.37E-04			0.506	0.643	0.553	0.856	0.775	0.852
LUCAT1	rs10072084	5	90539203	C	T	+··	*9.48E-06*					0.541	0.559	0.422	0.232	0.429	0.822
*FILIP1***	rs2951921	6	76165524	T	C	+··	*9.04E-07*					0.073	0.022	0.063	0.015	0.030	0.293
ZHX2	rs7841763	8	123971081	T	C	+++	*4.82E-06*	0.631	1.03E-04			0.043	0.023	0.146	0.102	0.058	0.169
APOA1	rs964184*	11	116648917	C	G	+++	5.37E-05	0.009	*1.72E-06*	*APOA1*	C,H,L,T	0.560	0.760	0.771	0.838	0.723	0.779
SIRT2	rs10405150	19	39387919	C	T	+··	*6.34E-06*					0.056	0.147	0.144	0.081	0.117	0.774
ZNF283	rs16976816	19	44339377	G	A	+··	*9.78E-06*					0.977	0.970	0.994	0.987	0.976	0.864
APOE	rs1160985*	19	45403412	C	T	+++	**2.13E-13**	**3.36E-09**	**4.29E-21**	*APOE*	C,H,L,T	0.724	0.659	0.590	0.554	0.447	0.378

HDL

Locus	SNP	Chr	BP	EA	OA	Dir	p_G	p_R	p_{GR}	Known gene	Traits	SAM	EAS	SAS	EUR	AMR	AFR
STON1-GTF2A1L	rs6739536	2	48831901	A	T	+··	*1.58E-06*					0.238	0.324	0.163	0.083	0.159	0.391
MGAT1	rs1038143*	5	180213878	G	A	+++	*3.72E-06*	0.016	*2.91E-07*			0.691	0.890	0.863	0.852	0.919	0.987
AKAP7	rs3777486	6	131584648	C	T	+−+	*3.09E-06*	0.808	4.37E-04			0.024	0.091	0.102	0.142	0.208	0.051
CSMD1	rs1626142	8	4345284	A	G	+··	*7.67E-06*					0.389	0.600	0.565	0.702	0.608	0.346
RAB21	rs328733	12	72197574	C	T	+++	*2.57E-06*	0.036	*5.92E-07*			0.212	0.374	0.300	0.131	0.262	0.389
ZNF10	rs2292029	12	133734113	C	T	+··	*4.05E-06*					0.821	0.888	0.746	0.756	0.844	0.992

HS6ST3	rs16953620	13	97508453	G	A	+··	*8.48E-06*										
LIPC	rs10438284	15	58629424	G	A	+−+	*4.00E-07*	0.133	*9.15E-07*	*LIPC*	C,H,T	0.714	0.674	0.827	0.728	0.801	0.953
CETP	rs289708	16	57038162	A	G	+··	**1.19E-11**			*CETP*	C,H,L,T	0.095	0.186	0.200	0.140	0.177	0.403
LIPG	rs16950739	18	47138509	C	T	+··	*1.07E-07*			*LIPG*	C,H	0.981	0.943	0.810	0.943	0.860	0.996
APOE	rs1160985*	19	45403412	T	C	+−+	0.003	0.342	0.004	*APOE*	C,H,L,T	0.276	0.341	0.410	0.446	0.553	0.623
CDH4	rs817687	20	59753355	T	C	+++	*2.31E-06*	0.237	*8.77E-06*			0.014	0.076	0.055	0.033	0.146	0.390

LDL

Locus	SNP	Chr	BP	EA	OA	Dir	p_G	p_R	p_{GR}	Known gene	Traits	SAM	EAS	SAS	EUR	AMR	AFR
APOB	rs754523*	2	21311691	G	A	+++	*3.25E-06*	0.158	*5.81E-06*	*APOB*	C,H,L,T	0.247	0.265	0.140	0.309	0.306	0.201
KALRN	rs6789134	3	123942339	G	A	++−	*3.22E-06*	0.925	1.96E-04			0.078	0.161	0.120	0.046	0.118	0.212
ZHX2	rs7841763	8	123971081	T	C	++−	*1.80E-06*	0.557	3.78E-05			0.043	0.023	0.146	0.102	0.058	0.169
SH2D4B	rs10509415	10	82473065	A	C	+··	*7.96E-06*					0.706	0.510	0.735	0.758	0.808	0.708
ALG10	rs3912355	12	34079616	C	T	+++	*2.12E-06*	0.544	3.97E-05			0.855	0.872	0.651	0.605	0.732	0.863
ALG10B	rs10880642	12	38554152	A	G	+··	*5.56E-06*					0.802	0.715	0.555	0.481	0.532	0.299
CPNE8	rs11169807	12	39244161	C	T	+++	*4.77E-06*	0.418	4.13E-05			0.794	0.644	0.536	0.505	0.408	0.898
LINC02408	rs17104016	12	67969929	A	T	+··	*9.29E-06*					0.574	0.734	0.924	0.853	0.906	0.838
LINC00922	rs254371	16	65943650	T	C	+··	*9.04E-06*					0.555	0.698	0.650	0.599	0.693	0.884
ZNF283	rs16976816	19	44339377	G	A	+··	*1.78E-06*					0.977	0.970	0.994	0.987	0.976	0.864
APOE	rs1160985	19	45403412	C	T	+++	**2.61E-20**	**5.07E-09**	**1.53E-27**	*APOE*	C,H,L,T	0.724	0.659	0.590	0.554	0.447	0.378

Triglycerides

Locus	SNP	Chr	BP	EA	OA	Dir	p_G	p_R	p_{GR}	Known gene	Traits	SAM	EAS	SAS	EUR	AMR	AFR
GCKR	rs780094	2	27741237	C	T	+++	*9.84E-07*	0.01	*5.62E-08*	*GCKR*	C,T	0.666	0.524	0.802	0.590	0.640	0.868
CD200	rs2399416	3	112059213	G	A	++−	*5.12E-06*	0.668	9.29E-04			0.979	0.852	0.860	0.607	0.728	0.899
SPIN1	rs7861888	9	90886340	G	A	+ · ·	*4.24E-06*					0.294	0.281	0.103	0.080	0.159	0.011
APOA1	rs964184*	11	116648917	G	C	+++	**2.37E-17**	**8.97E-14**	**1.81E-29**	*APOA1*	C,H,L,T	0.440	0.240	0.229	0.162	0.277	0.221
KIRREL3	rs3018434	11	126805881	A	G	+ · ·	*4.16E-06*					0.084	0.197	0.080	0.134	0.143	0.026
APOE	rs1160985*	19	45403412	C	T	+−+	0.312	0.836	0.507	*APOE*	C,H,L,T	0.724	0.659	0.590	0.554	0.447	0.378

EA = effect allele
OA = other allele (reference allele)
Dir = sign of β_{SNP} in the regression model
p_G = discovery cohort GWAS *p*-value
p_R = replication cohort *p*-value
p_{GR} = joint discovery and replication cohorts *p*-value
Known gene = Loci observed in Teslovich et al. 2010 or Willer et al. 2013
Traits = Traits locus associated with in Teslovich et al. 2010 or Willer et al. 2013
C = TC
H = HDL
L = LDL
T = TG
SAM = Samoan allele frequency (AF)
EAS = East Asian AF
SAS = South Asian AF
EUR = European AF
AMR = Admixed American AF
AFR = African AF

*The SNP genotyped in the replication population was not the peak SNP at this locus for this trait.

FIGURES

Figure 1. Manhattan plots for GWAS of four lipid traits in discovery cohort of 2,849 Samoans

The dashed and solid lines denote genome-wide suggestive and genome-wide significant *P* value thresholds ($P < 1 \times 10^{-5}$ and $P < 5 \times 10^{-8}$, respectively). Peaks are labeled with the candidate gene or closest gene in the region if they have at least suggestive association in the discovery cohort for at least one trait and demonstrate evidence of replication or have been previously associated.

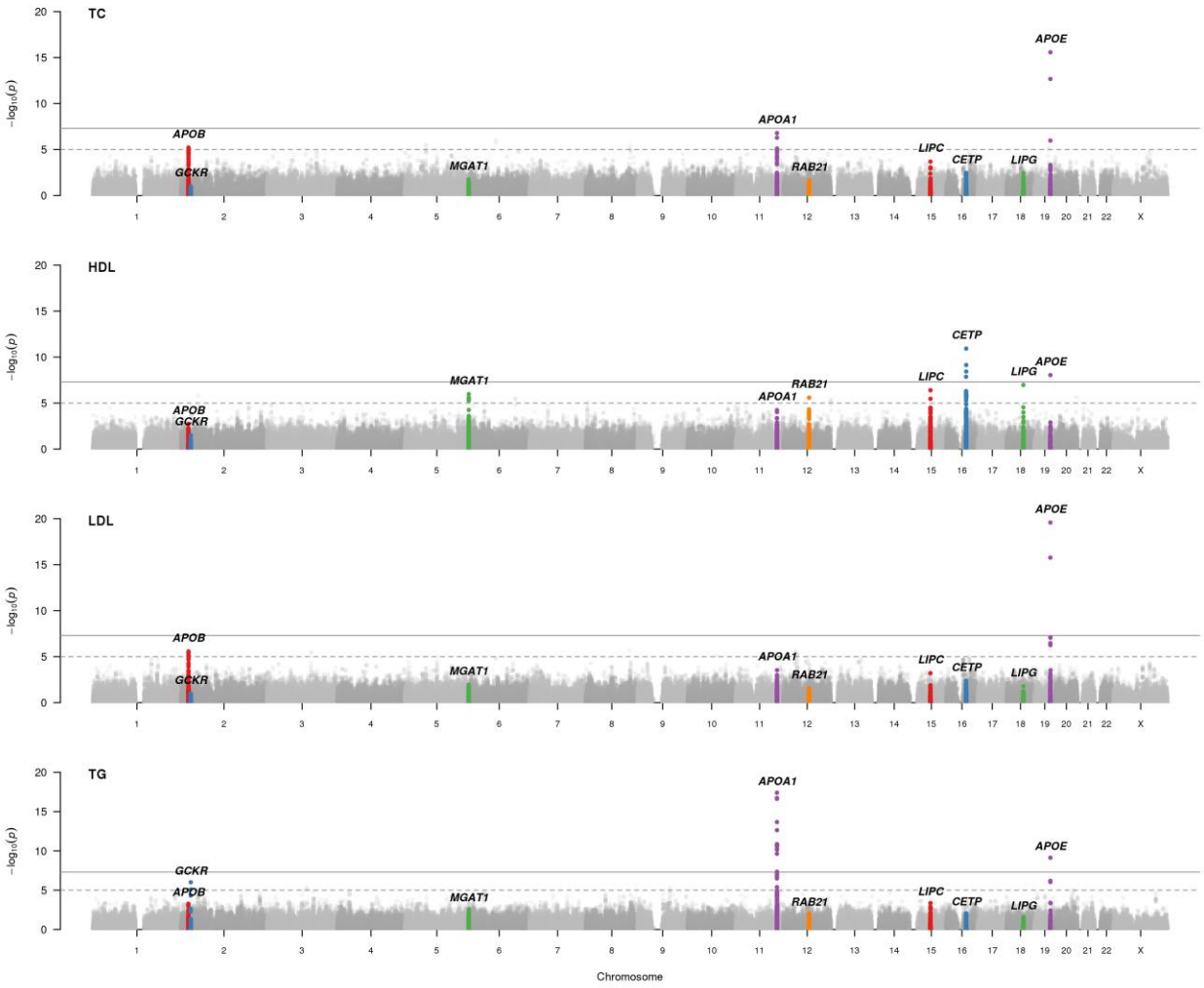
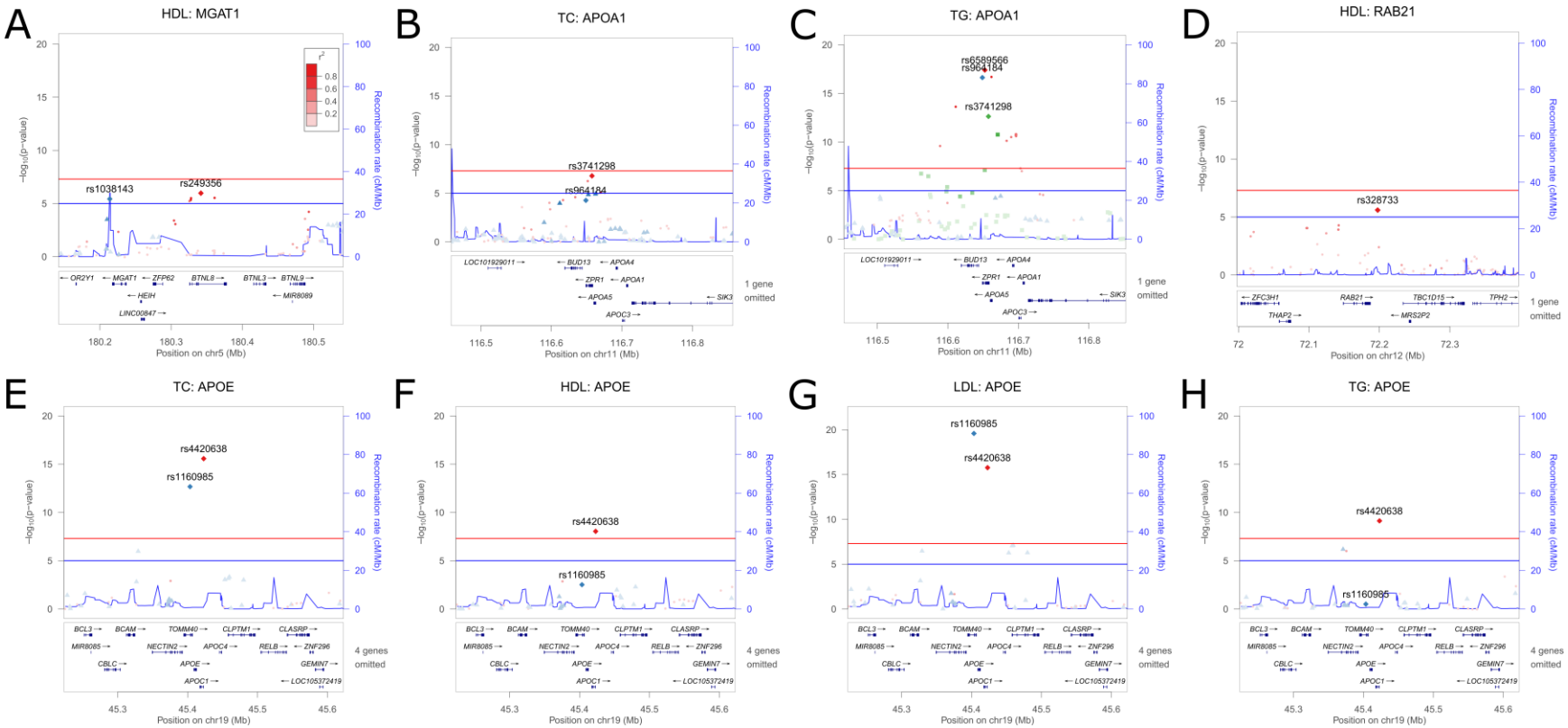
Figure 2. Regional association plots for selected loci.
Regional association plots generated in LocusZoom[43] showing –log10(*P* values) for SNPs in the (A) *MGAT1* locus and HDL, (B) *APOA1* locus and TC, (C) *APOA1* locus and TG, (D) *RAB21* locus and HDL, and the *APOE* locus and (E) TC, (F) HDL, (G) LDL, and (H) TG. Points are color coded within each plot according to pairwise linkage disequilibrium (r^2) with the labeled SNPs.