# A population-specific missense variant rs1597000001 in CETP promotes a favorable lipid profile and reduces CETP activity. 

Jaye Moors ${ }^{1 \#}$, Mohanraj Krishnan ${ }^{2 \#}$, Nick Sumpter ${ }^{3}$, Riku Takei ${ }^{3}$, Matt Bixley ${ }^{1}$, Murray

Cadzow ${ }^{1}$, Tanya J. Major ${ }^{1}$, Amanda Phipps-Green ${ }^{1}$, Ruth Topless ${ }^{1}$, Marilyn Merriman ${ }^{1}$, Malcolm Rutledge ${ }^{1}$, Ben Morgan ${ }^{1}$, Jenna C. Carlson ${ }^{2,4}$, Jerry Z. Zhang ${ }^{4}$, Emily M. Russell ${ }^{2}$, Guangyun Sun ${ }^{5}$, Hong Cheng ${ }^{5}$, Daniel E. Weeks ${ }^{2,4}$, Take Naseri ${ }^{6,7}$, Muagututi‘a Sefuiva Reupena ${ }^{8}$, Satupa‘itea Viali ${ }^{9}$, John Tuitele ${ }^{10}$, Nicola L. Hawley ${ }^{11}$, Ranjan Deka ${ }^{5}$, Stephen T. McGarvey ${ }^{7}$, Janak de Zoysa ${ }^{12}$, Rinki Murphy ${ }^{12}$, Nicola Dalbeth ${ }^{12}$, Lisa Stamp ${ }^{13}$, Mele Taumoepeau ${ }^{14}$, Frances King ${ }^{15}$, Philip Wilcox ${ }^{16}$, Sally McCormick ${ }^{1}$, Ryan L. Minster ${ }^{2}$, Tony R. Merriman ${ }^{1,3^{*}}$ and Megan Leask ${ }^{1,3}$

1. Department of Biochemistry, University of Otago, Dunedin, New Zealand
2. Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
3. Division of Clinical Rheumatology and Immunology, University of Alabama at Birmingham, Birmingham, AL, USA
4. Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA
5. Department of Environmental and Public Health Sciences, College of Medicine, University of Cincinnati, Cincinnati, OH, USA
6. Ministry of Health, Apia, Samoa
7. International Health Institute, Department of Epidemiology, School of Public Health, Brown University, Providence, RI, USA
8. Lutia i Puava 'ae Mapu i Fagalele, Apia, Samoa
9. School of Medicine, National University of Samoa, Apia, Samoa
10. Department of Public Health, Lyndon B. Johnson Tropical Medical Center, Faga‘alu, American Samoa, USA
11. Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA
12. Department of Medicine, University of Auckland, Auckland, New Zealand
13. Department of Medicine, University of Otago, Christchurch, New Zealand
14. Department of Psychology, University of Otago, Dunedin, New Zealand
15. Ngāti Porou Hauora, Te Puia Springs, New Zealand
16. Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand \#These authors contributed equally.

Correspondence to:
Prof Tony Merriman
Division of Clinical Rheumatology and Immunology, University of Alabama at Birmingham, Birmingham, AL, USA

Email: trmerriman@uabmc.edu




#### Abstract

Sequencing of CETP in Māori and Pacific peoples identified a common (MAF $\sim 2.4 \%-5.4 \%$ ) population-specific missense variant (rs1597000001, CETP:c.530C $>$ T p.Pro177Leu) that associates with higher HDL-C levels $\left(\widehat{\beta}_{H D L m e t a}=0.236 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI} 0.211 ; 0.260]\right)$ and lower LDL-C ( $\widehat{\beta}_{\text {LDLmeta }}=-0.133 \mathrm{mmol} / \mathrm{L}[95 \%$ CI $\left.-0.209 ;-0.058]\right)$. In a subsample of the study cohort ( $\mathrm{n}=11$ ), heterozygous carriers of the population-specific variant had lower plasma CETP activity $(P=0.028)$. Our study identifies a population-specific missense variant in $C E T P$ which lowers CETP activity with an effect on HDL-C that is comparable to Mendelian CETP loss-offunction mutations.


## Word count: 2889

## Key words: CETP, HDL-C, Māori, Pacific, Polynesia, LDL-C

Dyslipidaemia, defined as elevated total or low-density lipoprotein cholesterol (LDL-C) levels, triglycerides, and/or lower levels of high-density lipoprotein cholesterol (HDL-C), is an established risk factor for metabolic diseases such as cardiovascular disease, type 2 diabetes and gout ${ }^{1}$. In gout, elevated triglycerides may contribute to increased urate levels ${ }^{2}$, and elevated very low-density-lipoprotein (VLDL) associates with gout in Aotearoa New Zealand (NZ) Māori and Pacific peoples (herein defined as peoples of Micronesian, Melanesian and Polynesian ethnicity) ${ }^{3}$. Dyslipidaemia is more prevalent in Māori and Pacific populations ${ }^{4,5}$, and a higher proportion of Pacific nation ancestry has been associated with lower HDL-C ${ }^{6,7}$. Consistent with heritability estimates of HDL-C levels $(40 \%-60 \%)^{8,9}$, the prevalence of dyslipidaemia in Māori and Pacific populations suggests that genetics contributes to an atherogenic lipid profile ${ }^{10}$. While socioeconomic inequities contribute to the development and increased impact of metabolic diseases ${ }^{10,11}$, research on population-specific genetic variants associated with metabolic traits is important to provide insights into the genetic determinants of phenotypic differences between populations, as well as furthering genomic justice ${ }^{12}$.

One locus with genetic variation that consistently associates with lipid levels in multiple population groups, including Pacific peoples, is at the gene that encodes cholesteryl ester transfer protein: CETP ${ }^{13-16}$. CETP modifies lipid levels by mediating the bi-directional transfer of cholesterol esters from atheroprotective HDL to atherogenic VLDL in exchange for triglycerides, and vice versa. Various common polymorphisms including the CETP promoter variant ‘-629 C/A' (rs 1800775), intronic ‘Taq1B’ variant (rs708272), missense variant p.Ile405Val (rs5882) and loss-of-function variant p.Asp459Gly (rs2303790) have been associated with serum lipid profiles, CETP levels and/or CETP activity, and cardiovascular disease and
pharmacological response to CETP inhibitors ${ }^{17-20}$. In a GWAS for lipids in Samoans (using samples analyzed in greater depth here), the CETP locus was identified as the most significantly associated locus for HDL-C levels (rs289708, $P=1.19 \times 10^{-11}$ ) ${ }^{15}$. Given the previously reported associations with lipid levels and variation at CETP in Pacific peoples ${ }^{15,21}$, we used a discovery and replication study design ${ }^{22}$ to investigate whether population-specific genetic variation in CETP could contribute to HDL-C and CETP activity in Māori and Pacific populations.

## RESULTS

## Discovery of a novel CETP missense variant rs1597000001 in Māori and Pacific people

We extracted the gene region of $\operatorname{CETP}(\sim 22 \mathrm{~Kb})$ from whole-genome sequence data for 55 individuals of Māori and Pacific ethnicity, and identified a common missense variant in exon 6 of the CETP gene (dbSNP (build 155) rs1597000001, CETP:c.530C $>$ T, p.Pro177Leu). The minor allele frequency (MAF) was $3.4 \%$ in the 55 discovery genomes and $0.0036 \%$ in GnomAD ${ }^{23}$. A CADD score ${ }^{24}$ of 24.4 placed the variant in the top $1 \%$ of predicted deleteriousness and a GERP score ${ }^{25}$ of 4.4 indicated that the variant is conserved in mammals (Figure 1A, CADD and GERP tracks, respectively). The p.Pro177Leu amino acid substitution caused by the rs 1597000001 T -allele corresponds to amino acid position 160 in the mature protein structure of CETP ${ }^{26}$ owing to the first 17 amino acid residues of the CETP sequence consisting of the signal peptide ${ }^{27}$ (Figure 1B).

## rs1597000001 associates with higher HDL-C

We genotyped rs1597000001 in a discovery cohort of 2,270 Māori and Pacific individuals living in Aotearoa NZ. Demographic and biochemical characteristics of the participants are
summarized in Table 1. A similar MAF to the 55 Māori and Pacific genome sequences was observed $(\mathrm{MAF}=3.5 \%)$, and the MAF ranged from $3.2 \%-5.4 \%$ in the Pacific nation subgroups (stratified by self-reported Pacific nation of grandparental ethnicity) (Aotearoa NZ NZ Māori MAF 3.4\%, Aotearoa NZ Cook Island Māori MAF 3.8\%, Aotearoa NZ Samoan MAF 4.0\%, Aotearoa NZ Tongan MAF 4.2\%, Aotearoa NZ Niuēan MAF 5.4\% and Aotearoa NZ Other/Mixed MAF 4.5\%). People of Pukapukan ethnicity were monomorphic for the C allele of rs 1597000001 (Table 2). The MAF was $2.4 \%$ in the Ngāti Porou Hauora group. rs 1597000001 did not significantly deviate from Hardy-Weinberg equilibrium (HWE) in any of the Aotearoa NZ Pacific nation subgroups or the Ngāti Porou Hauora group (Table 2: $P>0.05$, all datasets).

The mean concentration and distributions of HDL-C and LDL-C ( $\mathrm{mmol} / \mathrm{L}$ ) were not different between the Pacific nation subgroups or the Ngāti Porou Hauora group in the Aotearoa NZ cohort (Table 1 and Supplementary Figure 1 (post-hoc Tukey test $\mathrm{p}>0.05$, all datasets)). To test if rs 1597000001 associates with HDL-C and LDL-C levels, we carried out a linear model for association of rs 1597000001 adjusted by age, sex, principal component vectors (PCs) and relatedness in the Aotearoa NZ dataset (Figure 2). Significantly higher HDL-C levels were evident in those with the rs 1597000001 T allele, and this association was observed in all of the surveyed Pacific nation subgroups (Aotearoa NZ NZ Māori ( $\widehat{\beta}=0.390 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI} 0.291$; $0.488]$ ), Aotearoa NZ Cook Island Māori ( $\widehat{\beta}=0.411 \mathrm{mmol} / \mathrm{L}$ [ $95 \%$ CI $0.200 ; 0.622]$ ), Aotearoa NZ Samoan $(\widehat{\beta}=0.192 \mathrm{mmol} / \mathrm{L}[95 \%$ CI $0.065 ; 0.319])$, Aotearoa NZ Tongan $(\widehat{\beta}=0.340$ $\mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI} 0.154 ; 0.526])$, and Aotearoa NZ Other/Mixed Pacific ( $\widehat{\beta}=0.261 \mathrm{mmol} / \mathrm{L}$ [ $95 \%$ CI $0.108 ; 0.414])$ ) with the exception of the Aotearoa NZ Niuēan subgroup where there was no evidence of association ( $\widehat{\beta}=0.178 \mathrm{mmol} / \mathrm{L}[95 \%$ CI $-0.138 ; 0.494]$ ) (Figure 2A). An
association with HDL was also observed in the Ngāti Porou Hauora group ( $\widehat{\beta}=0.260 \mathrm{mmol} / \mathrm{L}$ [95\% CI 0.100; 0.421]).

A fixed-effect meta-analysis showed a significant overall association of rs1597000001 T-allele with higher HDL-C ( $\widehat{\beta}=0.305 \mathrm{mmol} / \mathrm{L}$ [ $95 \%$ CI $0.2489 ; 0.362$ ]), with no evidence of heterogeneity between the sample sets $(P=0.223)$ (Figure 2A). The proportion of variance in HDL-C levels explained by rs 1597000001 in the Aotearoa NZ cohort was $4.5 \%$, similar to the proportion of variance explained by sex (5.0\%). There was no association of rs 1597000001 with LDL-C (Figure 2B) in any of the Pacific nation subgroups, Ngāti Porou Hauora group or in a fixed effect meta-analysis of the Aotearoa NZ cohort $(\widehat{\beta}=-0.046 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI}-0.212$; $0.120]$ (Figure 2B)).

We carried out association analyses in three independent Samoan cohorts (Samoan I-III) which serve as replication cohorts to validate the associations we observed in the Aotearoa NZ discovery cohort. Association analyses adjusted by age, sex, PCs and relatedness confirmed the association between HDL-C and rs 1597000001 (Samoan cohort I: $\widehat{\beta}=0.225 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI}$ 0.190; 0.260], Samoan cohort II: $\widehat{\beta}=0.193 \mathrm{mmol} / \mathrm{L}$ [ $95 \%$ CI $0.138 ; 0.248$ ] and Samoan cohort III: $\widehat{\beta}=0.242 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI} 0.171 ; 0.313]$ (Figure 2A)) with no evidence of heterogeneity between the sample sets $(P=0.502)$. Unlike the Aotearoa NZ cohort there was an association of rs 1597000001 with lower LDL-C in Samoan cohorts I and II (Samoan cohort I: $\widehat{\beta}=-0.134$ $\mathrm{mmol} / \mathrm{L}[95 \%$ CI -0.238; -0.030] and Samoan cohort II: $\widehat{\beta}=-0.276 \mathrm{mmol} / \mathrm{L}[95 \%$ CI $-0.462 ;-$ 0.090 ] (Figure 2B)) but not Samoan III (Samoan cohort III: $\widehat{\beta}=-0.077 \mathrm{mmol} / \mathrm{L}$ [95\% CI -0.312 ; 0.158] (Figure 2B)). The MAF was $4.4 \%$ in Samoan I, $4.7 \%$ in Samoan II and $4.8 \%$ in Samoan

III (compared to $4.0 \%$ in the Aotearoa NZ Samoan cohort) and did not deviate from HWE ( $P>$ 0.05, Samoan I-III (Table 2)). The proportion of variance of HDL-C explained by rs 1597000001 in the Samoan I and Samoan II and Samoan III cohorts was $5.1 \%, 6.4 \%$ and $7.8 \%$ respectively.

A fixed-effect meta-analysis of the Aotearoa NZ Pacific nation subgroups, Ngāti Porou Hauora group and Samoan I-III cohorts demonstrated the strong association of rs 1597000001 with HDLC levels $\left(\widehat{\beta}_{\text {HDLmeta }}=0.236 \mathrm{mmol} / \mathrm{L}[95 \%\right.$ CI $\left.0.211 ; 0.260]\right)$ with no heterogeneity between the cohorts $(P=0.054)$. rs 1597000001 associated with lower LDL-C levels in the fixed-effect metaanalysis of the Aotearoa NZ and Samoan cohorts I-III ( $\widehat{\beta}_{\text {LDLmeta }}=-0.133 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI}-$ 0.209 ; -0.058$]$ ) (Figure 2B) with no heterogeneity between the cohorts $(P=0.534)$.

Next we explored what effect this variant has on HDL-C and LDL-C levels in a cohort of young Pacific people (PTO cohort) before onset of metabolic disease indicators (Table 1). The mean concentration of HDL-C was significantly higher in the PTO cohort compared to all Pacific nation subgroups and the Ngāti Porou Hauora group in the Aotearoa NZ cohort (Table 1 and Supplementary Figure 1A (post-hoc Tukey test $P=<0.05$; all datasets)). LDL-C ( $\mathrm{mmol} / \mathrm{L}$ ) was significantly lower in the PTO cohort in comparison to all Pacific nation subgroups and the Ngāti Porou Hauora group in the Aotearoa NZ cohort (Table 1 and Supplementary Figure 1B (post-hoc Tukey test $P=<0.05$; all datasets)). rs 1597000001 was monomorphic for the major C allele in Pacific participants without Polynesian ethnicity (i.e. of Melanesian and/or Micronesian ethnicity) (Table 2). In PTO participants with Polynesian ethnicity the MAF was $4.5 \%$ and did not deviate from HWE $(P>0.05)$. Association analysis (adjusted by age, sex, and grandparental ethnicity) in participants of Polynesian ethnicity indicated that rs1597000001 associates with
higher HDL-C levels ( $\widehat{\beta}=0.366 \mathrm{mmol} / \mathrm{L}$ [ $95 \%$ CI $0.221 ; 0.511$ ] (Figure 3A)) and lower LDL-C levels $(\widehat{\beta}=-0.340 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI}-0.663 ;-0.016]$ (Figure 3B)) consistent with the negative estimates from the Samoan I and Samoan II cohorts. The proportion of variance of HDL-C explained by rs 1597000001 in the PTO cohort was $11.8 \%$.

## rs1597000001 associates with decreased CETP activity

To test the hypothesis that rs1597000001 alters CETP function, we carried out CETP activity assays in serum from participants of the Pacific Trust Otago cohort who were heterozygous for rs1597000001 $(\mathrm{n}=4)$ and homozygous for rs 1597000001 C allele $(\mathrm{n}=7)$. The rs1597000001 Tallele was significantly associated with lower activity of CETP in comparison to the homozygous carriers of the major C allele (unpaired t -test with Welch's correction $\mathrm{p}=0.028$ ) (Figure 4).

## Multiple independent genetic associations are identified at CETP in Māori and Pacific

## People

Genetic contribution to HDL-C is polygenic, and patients with extreme HDL concentrations possess both rare large-effect variants and common small-effect variants ${ }^{28}$ at CETP that contribute to HDL-C ${ }^{29}$. We re-examined ${ }^{15}$ the entire CETP locus for association with HDL-C levels in Māori and Pacific peoples in the Aotearoa NZ cohort. rs 1597000001 was the maximally associated variant followed by the CETP promoter polymorphism '-629 C/A' rs1800775 (Figure 5A). The rs 1800775 C-allele was present in the Aotearoa NZ cohort (46.0\%) at a frequency similar to Europeans (48.0\% in gnomAD), and exhibited weak linkage disequilibrium with rs1597000001 $\left(r^{2}=0.032\right)$. The effect size for rs1800775 C-allele $(\widehat{\beta}=-0.090 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI}$ 0.112; -0.069] was small in comparison to the effect of rs $1597000001(\widehat{\beta}=0.301 \mathrm{mmol} / \mathrm{L})$ and
the proportion of variance of HDL-C explained by rs 1800775 in the Aotearoa NZ cohort was $3.3 \%$ compared to $4.5 \%$ for rs 1597000001 . The rs 1800775 C-allele associates with the same direction of effect previously observed for Europeans $\left(\widehat{\beta}=-0.080 \mathrm{mmol} / \mathrm{L}, \mathrm{p}=3.7 \times 10^{-93}\right)^{13}$ and Pacific peoples $\left(\widehat{\beta}=-0.055 \mathrm{mmol} / \mathrm{L}, P=1.7 \times 10^{-4}\right)^{21}$.

Given the large effect of rs1597000001 on HDL-C and the fact that rs1597000001 and rs1800775 exhibit some linkage disequilibrium we carried out conditional analyses to test whether the effects at rs1800775 and rs 1597000001 were dependent on each other. In the Aotearoa NZ cohort, the effect on HDL-C persisted for rs 1597000001 ( $\widehat{\beta}=0.266 \mathrm{mmol} / \mathrm{L}[95 \%$ CI $0.210 ; 0.323]$ ) conditioned on rs1800775 genotype (Figure 5B) and for rs1800775 conditioned on rs 1597000001 genotype (rs1800775 C-allele; $\widehat{\beta}=-0.072 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI}-$ 0.094; -0.052]) (Figure 5C). However, after conditioning on rs1597000001, rs 183130 was the most significantly associated variant (T-allele; $\widehat{\beta}=0.087 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI} 0.062 ; 0.112]$ ) at CETP ( $\mathrm{r}^{2}$ with rs1800775 $=0.31$ ). In the Samoan I cohort, conditioning on rs 1597000001 resulted in two additional signals marked by rs11076175 (G-allele; $\widehat{\beta}=-0.0623 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI}$ $-0.0876 ;-0.0369]$ ) and rs4783961 (A-allele; $\widehat{\beta}=0.0480 \mathrm{mmol} / \mathrm{L}[95 \% \mathrm{CI} 0.0279 ; 0.0681]$ ) (Supplementary Figure 2). rs4783961 exhibited linkage disequilibrium with rs1800775 ( $\mathrm{r}^{2}=$ $0.227)$ and $\mathrm{rs} 183130\left(\mathrm{r}^{2}=0.623\right)$. However, rs 11076175 had only weak linkage disequilibrium with rs1800775 $\left(\mathrm{r}^{2}=0.114\right)$ and $\mathrm{rs} 183130\left(\mathrm{r}^{2}=0.014\right)$.

## DISCUSSION

Using whole-genome sequence data from individuals of Māori and Pacific ethnicity, we have identified a Population-specific missense variant in CETP that strongly associates with a favorable
lipid profile (higher HDL-C and lower LDL-C) in a meta-analysis of four independent Pacific cohorts. The positive effect on HDL-C levels of the rs 1597000001 T-allele was observed in every group analysed, with the exception of the Niuēan Pacific nation subgroup. This is likely a reflection of the small sample size of this subgroup, limiting our detection power. However, the effect size estimate was positive, which is consistent with all the other island nation subgroups in the Aotearoa NZ cohort. Given the likely, but unstudied, nutritional and environmental variations across these groups within Aotearoa NZ and between Aotearoa NZ and the Samoan Islands, it is noteworthy that similar effect sizes were detected.

The effects of rs 1597000001 on LDL-C are less conclusive. In the Samoan I and II cohorts, an association with lower LDL-C was observed. However, the Aotearoa NZ, Samoan III, and PTO cohorts did not show an association of rs 1597000001 with LDL-C levels. The differences could be due to unmeasured heterogeneity in medication use and other genetic and environmental factors. Our findings that rs 1597000001 associates with HDL-C consistently throughout Pacific nation groups but not LDL-C corroborate data from Mendelian randomization studies of CETP which indicate that CETP activity is an important causal determinant of HDL-C ${ }^{30}$ levels but not LDL-C levels.

Variants in CETP that cause loss-of-function and consequently have large effects on HDL-C levels are rare in the general population and are enriched in study populations that have high HDL-C levels ${ }^{31-33}$. Conversely rs 1597000001 (MAF $\sim 2.4 \%-5.4 \%$ ) is a common variant of large effect in Māori and Pacific peoples, and the allele frequency was stable throughout the Pacific nations surveyed here. It is notable that the variant was not detected in people of Pukapukan
ethnicity and Pacific peoples without Polynesian ethnicity. The most probable explanation is that the absence of the variant from people of Pukapukan ethnicity is the result of a different population history for this group for example bottle-necking event(s) and/or founder effects. That we did not observe the variant in non-Polynesian Pacific people suggests that this is a Polynesian-specific variant.

The large effect on HDL-C level that we observe for rs 1597000001 is reminiscent of that observed for the common hyperalphalipoproteinemia 1 (HALP1) (elevated HDL-C levels) loss-of-function variant rs2303790 (p.Asp259Gly; $\left.\hat{\beta}_{\mathrm{HDL}}=0.44 \mathrm{mmol} / \mathrm{L}\right)^{20}$ located in exon 15 of CETP. Functionally rs2303790 reduces CETP activity, likely emulating Anacetrapib inhibition of CETP ${ }^{34}$. Anacetrapib on average raises HDL-C levels by $1.12 \mathrm{mmol} / \mathrm{L}$ in patients with atherosclerotic vascular disease ${ }^{35}$. Given that rs 1597000001 is located in exon 6 , it is likely that its functional effect on CETP is disparate to rs2303790. However, it is notable that the effect of one allele of rs 1597000001 on HDL-C is $\sim 25 \%$ of that observed for Anacetrapib. Here, we showed that heterozygous carriers of rs1597000001 have lower CETP activity which supports our hypothesis that the rs 1597000001 T -allele results in dysfunctional CETP and consequently increases HDL-C.

The magnitude of effect on HDL-C of rs1597000001 is also comparable to carrier status for protein truncating variants in $\operatorname{CETP}(\widehat{\beta}=0.585 \mathrm{mmol} / \mathrm{L}){ }^{36}$. We note that rs 1597000001 is in very close proximity to a rare truncating variant p.Trp 179* which introduces a stop codon in exon 6 . This rare variant has only been identified in people with HALP1 ${ }^{31}$ thus we are unable to make comparisons of the estimates of effect on HDL. p.Trp179* could possibly alter CETP levels
through nonsense-mediated decay of the truncated mRNA. Alternatively, p.Trp179* could also functionally disrupt the CETP protein. p.Trp179* and rs1597000001 (p.Pro177Leu) are located in the $N$-terminal $\beta$-barrel domain of CETP which is predicted to penetrate the HDL surface ${ }^{37}$, and is a putative entry site for cholesterol ester. Both variants are contained in the $\Omega 6$ flap $\left(G \ln { }^{155}-\operatorname{Tr}{ }^{162}\right.$ ) which separates from the $\Omega 5$ flap upon HDL penetration, then opens the N terminal distal end for uptake of cholesterol ester ${ }^{37}$. Specifically, Trp ${ }^{162}$ plays an important anchoring role in the penetration and settling of CETP in HDL. Binding of CETP to the HDL substrate occurs via hydrophobic interactions ${ }^{38}$, and it could therefore be predicted that the p.Pro177Leu substitution from proline to the hydrophobic amino acid leucine at this site could alter substrate binding and therefore cholesterol ester transfer activity of CETP. Alternatively, proline residues commonly form bends and turns in protein structure ${ }^{39}$, thus the substitution to leucine could disrupt the conformation of the CETP N-terminal domain.

Although dyslipidaemia is prevalent in people of Māori and Pacific ethnicity, our study has identified a population-specific variant that associates with an improved lipid profile. It remains to be seen whether this variant is also associated with lower cardiovascular event risk. Although the variant has a tangible biological mechanism of effect and results in CETP deficiency, understanding how this variant is modified by other common, smaller-effect variants within the locus and how these variants function will be important information for more effective targeted interventions with pharmaceuticals ${ }^{40}$. Most importantly, this variant and the accumulating evidence of the presence of other Population-specific trait-associated variants ${ }^{41-45}$ emphasizes the significance of population-specific variation and its influence on disease traits and risk. Understanding the genetic basis of metabolic disease in Māori and Pacific populations is
essential for the improvement of health outcomes for these groups ${ }^{10}$, and comprehensive evaluations of genome-wide genetic variation is long overdue and critical for equity in healthcare as we move towards an era of personalized medicine grounded in genetics.

## METHODS

## Study cohorts

Demographic and biochemical characteristics of the participants are summarized in Table 1. 2,272 participants of Māori and Pacific ethnicity were recruited in Aotearoa New Zealand (NZ). The Aotearoa NZ cohort serves as the discovery cohort in this study. 4,309 participants of Samoan ethnicity were recruited in the Independent State of Samoa and the US territory of American Samoa into the Samoan I $(\mathrm{n}=2,851)$, II $(\mathrm{n}=908)$ and III $(\mathrm{n}=550)$ cohorts which serve as the replication cohorts in this study. Finally, an additional 255 young people (aged 14 25 years) identifying with at least one Pacific Island ethnicity within Oceania (i.e., Melanesia, Micronesia or Polynesia) residing in Dunedin, New Zealand were recruited by the Pacific Trust Otago into the Pacific Trust Otago (PTO) Cohort.

For all participants information obtained at recruitment included age, sex, height, and weight, as measured by trained assessors. BMI was calculated by dividing an individual's weight by the square of their height in meters. Blood biochemical measurements including lipid measurements were performed at Southern Community Laboratories (Dunedin, NZ) for the Aotearoa NZ participants, at Northwest Lipid Labs (Seattle, WA, USA) for the Samoan I cohort, and at Lipids Research Clinic at Miriam Hospital, Brown University for the Samoan II and III cohorts.

The Aotearoa NZ cohort is the amalgamation of three separate groups. 2002 participants aged $\geq$ 16 years, located primarily in Auckland and Christchurch were recruited to Genetics of Gout, Diabetes and Kidney Disease in Aotearoa NZ Study ${ }^{41}$. The participants from this study were separated into subgroups based on the self-reported Pacific nation of ethnicity of their grandparents. Those participants who also reported some non-Pacific grandparent ethnicity were grouped according to their Pacific nation ethnicity. This resulted in six sample sets: Aotearoa NZ NZ Māori $(n=814)$, Aotearoa NZ Cook Island Māori $(n=172)$, Aotearoa NZ Samoan $(n=322)$, Aotearoa NZ Tongan $(n=155)$, Aotearoa NZ Niuēan $(n=37)$ and an 'Other/Mixed' Aotearoa NZ Pacific group ( $n=232$ ), which included individuals of Tahitian $(n=1)$ and Tuvaluan $(n=5)$ ethnicity, along with individuals who self-reported grandparental ethnicity from more than one Pacific nation $(n=230)$. An additional 270 participants from Te Tairāwhiti (east coast of the North Island, NZ) were recruited in collaboration with the Ngāti Porou Hauora (Health Service) Charitable Trust. At the request of Ngāti Porou Hauora (NPH) these participants were analyzed separately to the participants in the six Pacific nation subgroups. 72 participants of Pukapukan ethnicity were recruited in collaboration with the Pukapukan Community of New Zealand Inc. in Mangere, South Auckland, NZ.

The Samoan I cohort consists of 2,851 Samoan adults aged 22-65 years residing in the Independent State of Samoa. The Samoan II cohort is a family study consisting of 908 Samoan adults aged 18-88 years residing in the Independent State of Samoa and the US territory of American Samoa. The Samoan III cohort consists of 550 Samoan adults aged 29-88 years residing in the Independent State of Samoa and the US territory of American Samoa. With the exception of the participants in the Samoan II cohort, participants were also asked about the
ethnicity of each of their grandparents. Participants from the Samoan I and III cohorts all reported four Samoan grandparents. Although participants from the Samoan II cohort did not report the ethnicity of their grandparents, in principal components analysis of ancestry (see below), the participants of this cohort cluster together with participants from the Samoan III cohort.

Ethical approval for the Aotearoa NZ cohort study was given by the NZ Multi-Region Ethics Committee (MEC/05/10/130; MEC/10/09/092; MEC/11/04/036), the Northern Y Region Health Research Ethics Committee (Ngāti Porou Hauora Charitable Trust study; NTY07/07/074), and the University of Otago Human Health Ethics Committee (Pacific Trust Otago; 12/349). Ethical approval for the Samoan I cohort study was given by the Health Research Committee of the Samoa Ministry of Health and the institutional review board of Brown University. Ethical approvals for the Samoan II and III cohort studies were given by the Health Research Committee of the Samoa Ministry of Health and the Institutional Review Boards (IRBs) of the Department of Health in American Samoan; of The Miriam Hospital, Providence, Rhode Island; and of Brown University. The consent forms for Samoan I, II, and III cohorts were available to participants in both Samoan and English. All participants provided written informed consent for the collection of samples and subsequent analysis. Participants recruited in Aotearoa New Zealand were asked if they would like a karakia (Māori prayer) carried out upon disposal of their blood samples, if indicated this was carried out. The PTO project was guided by the University of Otago Pacific Research Protocol.

## Identification of rs1597000001

For genomic sequencing $2 \mu \mathrm{~g}$ of total genomic DNA from 55 Māori and Pacific individuals (participants with gout in Ardea Biosciences (USA) clinical trials for urate-lowering therapy 42,46,47) were submitted to Kinghorn Centre for Clinical Genomics at Garvan Institute of Medical Research in New South Wales, Australia for library preparation and Next Generation Sequencing 30x WGS (TruSeq Nano v2.5). The CETP gene region (based on reference transcripts from Ensembl (https://asia.ensembl.org/index.html)) was extracted from whole genome sequence data in FASTQ format aligned to the human genome (GRCh37/hg19) following implementation of the Genome Analysis ToolKit (GATK) best practices using the Burrows-Wheeler Aligner ${ }^{48}$ Picardtools (https://broadinstitute.github.io/picard) and GATK v3.6.0 ${ }^{49}$ (NeSI pipeline for GATK; DOI: http://doi.org/10.5281/zenodo.2564243). Variants in CETP were annotated with allele frequencies from the Genome Aggregation Database (http://gnomad.broadinstitute.org/) ${ }^{23}$ to identify population-specific variants, and annotated using Variant Effect Predictor (VEP) (www.ensembl.org/Tools/VEP) ${ }^{50}$. In silico predictions for deleteriousness of rs 1597000001 and conservation at rs1597000001 were obtained using the UCSC browser (GRCh37/hg19) track collection ${ }^{51}$ for Combined Annotation Dependent Depletion (CADD) scores ${ }^{24}$ and Genomic Evolutionary Rate Profiling (GERP) scores for Mammalian Alignments ${ }^{25}$ respectively. The CETP protein structure ( $10.2210 / \mathrm{pdb} 2 \mathrm{OBD} / \mathrm{pdb}){ }^{26}$ was obtained from The Protein Databank (https://www.rcsb.org/) ${ }^{52}$ and visualized using the PyMOL molecular graphics system (v2.3.2 (Shrodinger, LLC)).

## Genotyping

In the Aotearoa NZ, NPH and PTO cohorts, rs1597000001 was genotyped using a customdesigned TaqMan probe-set (Applied Biosystems, Foster City, CA, USA). A custom Python
script (snp_design; DOI: https://doi.org/10.5281/zenodo.56250) was used to annotate the human genome build 37 reference sequence (ftp://ftp.ensembl.org/pub/grch37) with rs1597000001 and any surrounding SNPs (obtained from the NCBI dbSNP build 147 common SNP list; $\mathrm{ftp}: / / \mathrm{ftp} . n c b i . n l m . n i h . g o v / \mathrm{snp})$ before primer and probe design. Forward primer:

CGGTGCCTGGTACACACTAG; reverse primer: TGTGAACAGCTGCTTGATCCA; probe 1 (VIC): CTGCCTTCAGGCCTG; probe 2 (FAM): CTGCCTTCAGGCTTG. Genotyping was carried out using the LightCycler 480 Real-Time Polymerase Chain Reaction System (Roche Applied Science, Indianapolis, IN, USA) in 384 well plates. Additional genotypes beyond rs159700001 at the CETP locus for the Aotearoa NZ participants were obtained from the Illumina Infinium CoreExome v24 bead chip platform whole-genome genotyping data generated as previously described ${ }^{41}$. In the Samoan I cohort, participants were genotyped using the Genome-Wide Human SNP Array 6.0 (Affymetrix, CA, USA) with quality control as described in ${ }^{53}$. rs 1597000001 was not included on the array, but the genotypes from the array served as the scaffold for imputation of the variant. A subset of Samoan I cohort participants ( $n=1,294$ ) were whole-genome sequenced as part of the Trans-Omics for Precision Medicine (TOPMed) program ${ }^{54}$, a Samoan-specific haplotype reference panel was generated from the sequences using Eagle2 (v.2.3.5) ${ }^{55}$, and additional genotypes, including those of rs1597000001 were imputed based on the genotype scaffold and the reference panel using Minimac4 ${ }^{56}$. In the Samoan II and III cohorts, participants were genotyped using the Infinium Global Screening Array-24 v3.0 BeadChip (Illumina, CA, USA) with custom content that included rs1597000001. Quality control and quality assurance checks for the Samoan II and III cohorts were conducted using GWASTools ${ }^{57}$.

## Principal components and relatedness matrices

Whole-genome principal component vectors were calculated for the Aotearoa NZ cohort using 2,858 ancestry informative markers (as identified by Illumina) extracted from the CoreExome whole-genome genotypes. Ten PCs were generated from the SmartPCA (EIGENSOFT v6.0.1)) ${ }^{58}$ program and used as covariates in the association analyses to account for population stratification and cryptic relatedness. Relatedness coefficients were calculated in the Aotearoa NZ dataset using the software GEMMA (v0.98.4) ${ }^{59}$ from 257,069 independent SNPs from the CoreExome whole-genome genotyping data.

For the Samoan I, II, and III cohorts, principal components and empirical kinship coefficients were calculated using genotypes from the Genome-Wide Human SNP 6.0 array (Samoan I) and the Infinium Global Screening Array-24 v3.0 BeadChip (Samoan II and III). In the Samoan I cohort, PCs were calculated as described in Minster et al. ${ }^{53}$ and the empirical kinship matrix was calculated from 10,000 independent autosomal markers using OpenMendel ${ }^{53,60,61}$. In the Samoan II and III cohorts, 55,640 autosomal markers were used to calculate PCs and empirical kinship matrices with PC-AiR and PC-Relate, respectively ${ }^{62}$, per the recommended procedure in the GENESIS R/Bioconductor package ${ }^{63}$.

## Statistical analyses

All association analyses described below were carried out using the R v4.0.2 software. For the association analyses in the Aotearoa NZ cohort, a generalized linear mixed model-based Wald test was carried out using GMMAT software (v1.3.1) ${ }^{64}$ to test for associations between rs 1597000001 and the continuous variables HDL-C and LDL-C. Analyses were adjusted by sex, age, 10 PCs and
relatedness. A linear model was used to test for associations in the PTO cohort between rs 1597000001 and the continuous variables HDL-C and LDL-C using the lm() function in R. The PTO analyses were adjusted by sex, age and self-reported ethnicity group for each grandparent, owing to absence of whole-genome ancestry informative markers in this cohort. For the association analyses in the Samoan I cohort, linear mixed-model regression of the phenotypes on rs1597000001 was performed using the lmekin () function of the coxme R package with sex, age, four PCs and relatedness as covariates. In the Samoan II and III cohorts, a mixed model association testing was carried out using lmekin () to test for associations between rs 1597000001 and HDL-C or LDL-C phenotypes with sex, age, first four PCs, polity (Samoa or American Samoa), and relatedness as adjusting variables.

Each Pacific population sample set (Aotearoa NZ cohort: Aotearoa NZ NZ Māori, Aotearoa NZ Cook Island Māori, Aotearoa NZ Samoan, Aotearoa NZ Niuēan, Aotearoa NZ Pukapukan and Aotearoa NZ 'Other /Mixed' Pacific nations subgroups; Ngāti Porou Hauora group, Samoan Cohort I, Samoan Cohort II and Samoan Cohort III; and Pacific Trust Otago) was analyzed separately, and the effects combined for the Aotearoa NZ cohort and all cohorts (excluding the PTO cohort) in an inverse variance-weighted fixed-effect meta-analysis using the R package meta (v.3.0-2). Heterogeneity between sample sets was assessed using Cochran's heterogeneity (Q) statistic. The proportion of variance explained by rs 1597000001 and CETP promoter common variant rs1800775 for HDL-C was calculated using the partial.rsq() function in the rsq R package (v2.2) ${ }^{65}$ which calculates the partial R-squared value for each predictor from the linear model generated by the $\operatorname{lm}()$ function in $R$. The $\beta$ coefficient in all analyses represents the estimated effect on HDL-C and LDL-C units (mmol/L) per copy of the
rs 1597000001 T-allele. For the Aotearoa NZ analyses HDL-C and LDL-C measurements were in $\mathrm{mmol} / \mathrm{L}$. In the Samoan cohorts I, II and III measurements were in $\mathrm{mg} / \mathrm{dl}$ and converted to $\mathrm{mmol} / \mathrm{L}$ before the association analyses.

To contextualize the effect of rs1597000001 among the other variation in CETP, variants on the Illumina Infinium CoreExome v24 bead chip platform and present in the Aotearoa NZ cohort (minor allele frequency $(\mathrm{MAF})>0.01)$ were extracted from the CETP region (rs1597000001 $\pm$ 500 kb ). A linear model was used to test variants for association with HDL-C using PLINK (v1.90b6.10) adjusted by age, sex and 10 PCs. Linkage disequilibrium and conditional analyses for rs1597000001 and CETP promoter common polymorphism rs 1800775 in the Aotearoa NZ cohort were carried out in PLINK (v1.90b6.10), adjusting by age, sex and 10 PCs calculated from the Aotearoa NZ dataset. Conditional analysis was conducted in Samoan cohort I across the CETP region with rs 1597000001 modeled as an additional fixed-effect covariate. Regional association plots were created using a custom R package (LocusZoom-like Plots: DOI: https://doi.org/10.5281/zenodo.5154379) for the Aotearoa NZ cohort and using LocusZoom ${ }^{66}$ for the Samoan I cohort.

## CETP activity

Individuals of the PTO cohort were recalled specifically for the purpose of providing fresh plasma samples for the CETP activity assays. CETP activity in 11 participants, of whom four were heterozygous for rs1597000001 and seven were homozygous for rs 1597000001 C -allele was assessed using the CETP activity assay kit II (F) from BioVision (K595-100) according to the manufacturer's instructions. Fresh plasma ( $1 \mu \mathrm{~L}$ ) was used for the assays which were carried out
in technical triplicate. Fluorescence was measured using a BMG LabTech CLARIOstar plate reader and CLARIOstar software (v.5.01 R2, Firmware 1.10). Using the stat. desc () function of the pastecs R package, we tested whether data within each genotypic group was normally distributed. We found that both genotypic groups had skewness/2SE and kurtosis/2SE values between -1 and 1 , indicating normality of the data. The Shapiro-Wilks tests for skewness were also non-significant. A two-samples t-test with Welch's correction was conducted to test for a difference in mean CETP activity between the two genotypic groups. One technical replicate was removed after it was identified as an outlier based on a significant Grubbs test for one outlier ( $P=$ 0.00065 ) using the grubbs.test () function from the outliers R package. Statistical analyses were conducted in R v4.0.2 software with all data reported as means $\pm$ standard error of the mean.

## ACKNOWLEDGMENTS

The authors sincerely thank the participants for generously donating their time and information to this study. The authors would like to thank J. Drake (Department of Rheumatology, Canterbury District Health Board, Christchurch, New Zealand), J. de Kwant, R. Laurence, C. Franklin, and M. House (all Department of Medicine, University of Auckland, Auckland, New Zealand), J. Harre Hindemarsh, N. Aupouri, R. Akuhata, and C. Ford (all of Ngāti Porou Hauora Charitable Trust, Te Puia Springs, New Zealand), G. Sexton (Counties Manukau District Health Board, Auckland, New Zealand), and B. Vincent (Waitemata District Health Board, Auckland, New Zealand) for recruitment. For the Samoan cohorts, 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. The work was funded by the National Institute of Health grants R01HL093093 (STM),

R01HL133040 (RLM), R01AG09375 (STM), R01HL52611 (MI Kamboh), R01DK59642 (STM), and R01DK55406 (RD). Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). Genome sequencing for the "NHLBI TOPMed: Genome-wide Association Study of Adiposity in Samoans" (phs000972.v4.p1) was performed at the Northwest Genomics Center (HHSN268201100037C) and the New York Genome Center (HHSN268201500016C). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I).

## AUTHOR CONTRIBUTIONS

MM, RT, TJM and AP-G coordinated the sampling, genome sequencing and genome-wide genotyping for the Aotearoa NZ cohort and 55 Aotearoa NZ genome sequences. MB and MC carried out the genome sequencing data analyses in the 55 genome sequences. For the Samoan I cohort NH led the field work data collection and phenotype analyses under the supervision of STM. For the Samoan II and III cohorts, STM led the field work data collection and phenotype analyses. GS and HC performed genotyping experiments for the Samoan I cohort, and HC prepared DNA for genotyping at the Center for Inherited Disease Research for the Samoan II and III cohorts and for sequencing by the TOPMed Program for the sequenced subset of the Samoan I cohort, under the supervision of RD. JZZ and JCC performed genotype imputation for
the participants in the Samoa I cohort with guidance from RLM and DEW. JM coordinated the samples and genotyping for the PTO cohort and carried out the genotyping over the Aotearoa NZ cohort. JM, MK, and ML carried out the association analyses with guidance from TRM, DEW, RLM and assistance from NS, RT, EMR, JCC and TJM. MR and BM carried out the CETP assays supervised by SM. MSR, SV, and JT facilitated fieldwork in Samoa and American Samoa. TN, MSR, SV, NH, PW, FK, MT, ND, LS, RM and JdZ contributed to the discussion of the public health and cultural implications of the findings. ML wrote the manuscript with guidance from TRM and contribution from all co-authors. All co-authors contributed to this work, discussed the results, and critically reviewed and revised the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## DATA AVAILABILITY

The data from the Aotearoa NZ and PTO cohorts are not publicly available owing to consent restrictions but can be requested from the corresponding author under an appropriate arrangement. Samoan cohort I data are available from dbGaP (accession number:
phs000914.v1.p1). Samoan II and III cohort data (recruited in 2002-2003 and 1990-1995, respectively) are not available as participants were not consented for data sharing.

## CODE AVAILABILITY

All software used in the analyses were open source and described in the Methods. Code written for the analyses are available in GitHub: https://github.com/MerrimanLab/CETP-Project

## REFERENCES

1. Kolovou, G.D., Anagnostopoulou, K.K. \& Cokkinos, D.V. Pathophysiology of dyslipidaemia in the metabolic syndrome. Postgrad Med J 81, 358-66 (2005).
2. Rasheed, H., Hughes, K., Flynn, T.J. \& Merriman, T.R. Mendelian randomization provides no evidence for a causal role of serum urate in increasing serum triglyceride levels. Circulation: Cardiovascular Genetics 7, 830-837 (2014).
3. Rasheed, H. et al. The relationship of apolipoprotein B and very low density lipoprotein triglyceride with hyperuricemia and gout. Arthritis research \& therapy 16, 1-10 (2014).
4. Gentles, D. et al. Serum lipid levels for a multicultural population in Auckland, New Zealand: results from the Diabetes Heart and Health Survey (DHAH) 2002-2003. N Z Med $J$ 120, U2800 (2007).
5. Beaglehole, R., Prior, I.A., Eyles, E. \& Sampson, V. High density lipoprotein cholesterol and other serum lipids in New Zealand Maoris. N Z Med J 90, 139-42 (1979).
6. Win Tin, S.T., Kenilorea, G., Gadabu, E., Tasserei, J. \& Colagiuri, R. The prevalence of diabetes complications and associated risk factors in Pacific Islands countries. Diabetes Res Clin Pract 103, 114-8 (2014).
7. Sun, H. et al. The impact of global and local Polynesian genetic ancestry on complex traits in Native Hawaiians. PLoS Genet 17, e1009273 (2021).
8. Qasim, A. \& Rader, D.J. Human genetics of variation in high-density lipoprotein cholesterol. Curr Atheroscler Rep 8, 198-205 (2006).
9. Lusis, A.J. Genetics of atherosclerosis. Trends Genet 28, 267-75 (2012).
10. Merriman, T.R. \& Wilcox, P.L. Cardio-metabolic disease genetic risk factors among Maori and Pacific Island people in Aotearoa New Zealand: current state of knowledge and future directions. Ann Hum Biol 45, 202-214 (2018).
11. Faatoese, A.F. et al. Community screening for cardiovascular risk factors and levels of treatment in a rural Maori cohort. Aust N Z J Public Health 35, 517-23 (2011).
12. Reardon, J. The postgenomic condition, (University of Chicago Press, 2017).
13. Ridker, P.M. et al. Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18245 initially healthy women from the Women's Genome Health Study. Circ Cardiovasc Genet 2, 26-33 (2009).
14. Mirmiran, P. et al. Genetic variations of cholesteryl ester transfer protein and diet interactions in relation to lipid profiles and coronary heart disease: a systematic review. Nutr Metab (Lond) 14, 77 (2017).
15. Carlson, J.C. et al. Genome-wide association studies in Samoans give insight into the genetic architecture of fasting serum lipid levels. J Hum Genet 66, 111-121 (2021).
16. Kenny, E.E. et al. Increased power of mixed models facilitates association mapping of 10 loci for metabolic traits in an isolated population. Hum Mol Genet 20, 827-39 (2011).
17. Dachet, C., Poirier, O., Cambien, F., Chapman, J. \& Rouis, M. New functional promoter polymorphism, CETP/-629, in cholesteryl ester transfer protein (CETP) gene related to CETP mass and high density lipoprotein cholesterol levels: role of $\mathrm{Sp} 1 / \mathrm{Sp} 3$ in transcriptional regulation. Arterioscler Thromb Vasc Biol 20, 507-15 (2000).
18. Marschang, P. et al. Plasma cholesteryl ester transfer protein concentrations predict cardiovascular events in patients with coronary artery disease treated with pravastatin. $J$ Intern Med 260, 151-9 (2006).
19. Tsai, M.Y. et al. Cholesteryl ester transfer protein genetic polymorphisms, HDL cholesterol, and subclinical cardiovascular disease in the Multi-Ethnic Study of Atherosclerosis. Atherosclerosis 200, 359-67 (2008).
20. Tang, C.S. et al. Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. Nat Comтии 6, 10206 (2015).
21. Lowe, J.K. et al. Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. PLoS Genet 5, e1000365 (2009).
22. Studies, N.-N.W.G.o.R.i.A. et al. Replicating genotype-phenotype associations. Nature 447, 655-60 (2007).
23. Karczewski, K.J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434-443 (2020).
24. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46, 310-5 (2014).
25. Davydov, E.V. et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 6, e1001025 (2010).
26. Qiu, X. et al. Crystal structure of cholesteryl ester transfer protein reveals a long tunnel and four bound lipid molecules. Nat Struct Mol Biol 14, 106-13 (2007).
27. UniProt, C. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49, D480-D489 (2021).
28. Dron, J.S. et al. Polygenic determinants in extremes of high-density lipoprotein cholesterol. Journal of lipid research 58, 2162-2170 (2017).
29. Goodrich, J.K. et al. Determinants of penetrance and variable expressivity in monogenic metabolic conditions across 77,184 exomes. Nat Comтип 12, 3505 (2021).
30. Blauw, L.L. et al. Mendelian randomization reveals unexpected effects of CETP on the lipoprotein profile. Eur J Hum Genet 27, 422-431 (2019).
31. Lee, C.J. et al. CETP, LIPC, and SCARB1 variants in individuals with extremely high high-density lipoprotein-cholesterol levels. Scientific reports 9, 1-7 (2019).
32. Oldoni, F., Sinke, R.J. \& Kuivenhoven, J.A. Mendelian disorders of high-density lipoprotein metabolism. Circ Res 114, 124-42 (2014).
33. Rosenson, R.S. et al. HDL and atherosclerotic cardiovascular disease: genetic insights into complex biology. Nature Reviews Cardiology 15, 9-19 (2018).
34. Jamalan, M., Zeinali, M. \& Ghaffari, M.A. A molecular dynamics investigation on the inhibition mechanism of cholesteryl ester transfer protein by Anacetrapib. Medicinal Chemistry Research 25, 62-69 (2016).
35. Group, H.T.R.C. et al. Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease. N Engl J Med 377, 1217-1227 (2017).
36. Nomura, A. et al. Protein-truncating variants at the cholesteryl ester transfer protein gene and risk for coronary heart disease. Circulation research 121, 81-88 (2017).
37. Cilpa-Karhu, G., Jauhiainen, M. \& Riekkola, M.L. Atomistic MD simulation reveals the mechanism by which CETP penetrates into HDL enabling lipid transfer from HDL to CETP. J Lipid Res 56, 98-108 (2015).
38. Zhang, M. et al. HDL surface lipids mediate CETP binding as revealed by electron microscopy and molecular dynamics simulation. Sci Rep 5, 8741 (2015).
39. Morgan, A.A. \& Rubenstein, E. Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. PLoS One 8, e53785 (2013).
40. Merriman, T.R. Application of Genetic Epidemiology to CETP (Cholesteryl Ester Transfer Protein) Concentration and Risk of Cardiovascular Disease. Circ Genom Precis Med 11, e002138 (2018).
41. Krishnan, M. et al. Discordant association of the CREBRF rs373863828 A allele with increased BMI and protection from type 2 diabetes in Maori and Pacific (Polynesian) people living in Aotearoa/New Zealand. Diabetologia 61, 1603-1613 (2018).
42. Ji, A. et al. Aotearoa New Zealand Māori and Pacific population-specific gout risk variants: CLNK is a separate risk gene at the SLC2A9 locus. The Journal of Rheumatology (2021).
43. Tanner, C. et al. Population-Specific Resequencing Associates the ATP-Binding Cassette Subfamily C Member 4 Gene With Gout in New Zealand Māori and Pacific Men. Arthritis \& Rheumatology 69, 1461-1469 (2017).
44. Klück, V. et al. Rare genetic variants in interleukin-37 link this anti-inflammatory cytokine to the pathogenesis and treatment of gout. Annals of the rheumatic diseases 79, 536-544 (2020).
45. Patel, S.G. et al. Potential PINK1 Founder Effect in Polynesia Causing Early-Onset Parkinson's Disease. Movement Disorders.
46. Stamp, L.K., Merriman, T.R. \& Singh, J.A. Expert opinion on emerging urate-lowering therapies. Expert Opin Emerg Drugs 23, 201-209 (2018).
47. Wood, R., Fermer, S., Ramachandran, S., Baumgartner, S. \& Morlock, R. Patients with Gout Treated with Conventional Urate-lowering Therapy: Association with Disease Control, Health-related Quality of Life, and Work Productivity. J Rheumatol 43, 18971903 (2016).
48. Li, H. \& Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-95 (2010).
49. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-303 (2010).
50. McLaren, W. et al. The Ensembl Variant Effect Predictor. Genome Biol 17, 122 (2016).
51. Navarro Gonzalez, J. et al. The UCSC Genome Browser database: 2021 update. Nucleic Acids Res 49, D1046-D1057 (2021).
52. Berman, H.M. et al. The Protein Data Bank. Nucleic Acids Res 28, 235-42 (2000).
53. Minster, R.L. et al. A thrifty variant in CREBRF strongly influences body mass index in Samoans. Nature Genetics (2016).
54. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590, 290-299 (2021).
55. Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nature genetics 48, 1443-1448 (2016).
56. Das, S. et al. Next-generation genotype imputation service and methods. Nat Genet 48, 1284-1287 (2016).
57. Gogarten, S.M. et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics 28, 3329-31 (2012).
58. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904-9 (2006).
59. Zhou, X. \& Stephens, M. Efficient multivariate linear mixed model algorithms for genomewide association studies. Nat Methods 11, 407-9 (2014).
60. Lange, K. et al. Mendel: the Swiss army knife of genetic analysis programs. Bioinformatics 29, 1568-70 (2013).
61. Zhou, H. et al. OPENMENDEL: a cooperative programming project for statistical genetics. Hum Genet 139, 61-71 (2020).
62. Conomos, M.P., Reiner, A.P., Weir, B.S. \& Thornton, T.A. Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet 98, 127-48 (2016).
63. Gogarten, S.M. et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics 35, 5346-5348 (2019).
64. Chen, H. et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. Am J Hum Genet 98, 653-66 (2016).
65. Zhang, D. rsq: R-Squared and Related Measures. R package version 2.2. https://CRAN.Rproject.org/package=rsq. (2021).
66. Pruim, R.J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336-7 (2010).

## FIGURE LEGENDS

Figure 1. The population-specific rs 1597000001 variant changes the amino acid sequence of CETP. A. The CADD, GERP and Multiz UCSC track alignment at rs 1597000001 for which the substitution creates a non-synonymous amino acid residue change (p.Pro177Leu). 3D schematic of p .Pro177Leu (P160L) in the mature CETP protein structure in the N -terminal barrel domain (left) of CETP. The CETP protein structure ( $10.2210 / \mathrm{pdb} 2 \mathrm{OBD} / \mathrm{pdb})^{26}$ was obtained from The Protein Databank (https://www.rcsb.org/) ${ }^{52}$ and illustrated using PyMOL. CETP, cholesteryl ester transfer protein. CADD, Combined Annotation Dependent Depletion scores. GERP, Genomic Evolutionary Rate Profiling.

## Figure 2. Association analyses of rs1597000001 with HDL-C and LDL-C levels in Aotearoa

## NZ discovery cohort and the Samoan I, Samoan II and Samoan II replication cohorts.

Forest plot of a fixed-effect meta-analysis for the association of rs 1597000001 with HDL-C $(\mathrm{mmol} / \mathrm{L})(\mathbf{A})$ and LDL-C $(\mathrm{mmol} / \mathrm{L})(\mathbf{B})$. Associations were adjusted by age, sex, 10 PCs and relatedness in the Aotearoa NZ cohort and age, sex, first four PCs and relatedness in the Samoan I-III cohorts. HDL-C, high-density lipoprotein cholesterol. LDL-C, low-density lipoprotein
cholesterol. NZ Māori, New Zealand Māori. CI Māori, Cook Island Māori. PCs, principal components.

Figure 3. Association analyses of rs 1597000001 with HDL-C and LDL-C levels in Polynesian participants of the PTO cohort. Forest plot of rs1597000001 association with HDL-C ( $\mathrm{mmol} / \mathrm{L}$ ) and LDL-C ( $\mathrm{mmol} / \mathrm{L}$ ) in the PTO cohort (Polynesian ethnicity only). Associations were adjusted by age, sex and grandparental ethnicity. HDL-C, high-density lipoprotein cholesterol. LDL-C, low-density lipoprotein cholesterol.

Figure 4. CETP activity by rs1597000001 genotype. Plasma activity of CETP ( $\mathrm{pmol} / \mathrm{hr} / \mu \mathrm{L}$ ) in 11 participants of the PTO cohort who had the heterozygous rs1597000001 genotype (CT) versus the homozygous rs1597000001 genotype for the major allele (CC). CETP, cholesteryl ester transfer protein. Error bars represent standard error of the mean. A two-sample t-test with Welch's correction was conducted to test for a statistically significant $(P=0.028)$ difference in mean CETP activity between the two genotypic groups.

Figure 5. Association of HDL-C levels at the CETP locus. Association of HDL-C at the CETP locus (+/-500kb the lead variant rs1597000001) (A) and after conditioning on rs1597000001 (B) and rs1800775 (C) using variants on Illumina Infinium CoreExome v24 bead chip platform for genotyped participants from the Aotearoa NZ cohort. The strength of LD, as measured by the $\mathrm{r}^{2}$, between each variant and rs1597000001 (A and C) and rs183130 (B) is represented by the color of each point according to the legend in the top right hand corner. The plot was generated using a
medRxiv preprint doi: https://doi.org/10.1101/2021.09.11.21263438; this version posted September 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.
custom locus zoom R package (https://github.com/Geeketics/LocusZooms). CETP, cholesteryl ester transfer protein. HDL-C, high-density lipoprotein cholesterol.

744 Table 1. Baseline characteristics of the cohorts.

|  | Aotearoa NZ Discovery Cohort |  |  |  |  |  |  |  | Samoan Replication Cohorts |  |  | Pacific Trust Cohort All |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | NZ Māori | CI Māori | Samoan | Tongan | Niuēan | Pukapukan | Other/Mixed | Ngāti Porou Hauora | Samoan I | Samoan II | Samoan III |  |
| Participants (n) | 814 | 172 | 322 | 155 | 37 | 72 | 232 | 270 | 2851 | 908 | 550 | 255 |
| $\begin{aligned} & \text { Sex (male) } \\ & \mathrm{n}(\%) \end{aligned}$ | 459 (56.39) | 105 (61.05) | 235 (72.98) | 117 (75.48) | 28 (75.68) | 32 (44.44) | 143 (61.64) | 186 (68.89) | 1146 (40.2) | 406 (44.7) | 252 (45.8) | $\begin{aligned} & 126 \\ & (49.41) \end{aligned}$ |
| Age Range | 17-85 | 18-88 | 18-81 | 18-79 | 19-75 | 18-84 | 17-88 | 18-94 | 22-65 | 18-88 | 29-88 | 14-25 |
| Age Distribution | $50.92 \pm 15.36$ | $50.97 \pm 15.22$ | $44.78 \pm 14.18$ | $42.17 \pm 14.53$ | $48.78 \pm 13.62$ | $44.93 \pm 17.2$ | $41.34 \pm 15.69$ | $56.15 \pm 13.61$ | $45.09 \pm 11.18$ | $43.26 \pm 16.23$ | $43.46 \pm 8.74$ | $\begin{aligned} & 18.49 \pm \\ & 2.54 \end{aligned}$ |
| Height (cm) | $170.31 \pm 9.9$ | $169.44 \pm 10.43$ | $173.57 \pm 9.05$ | $174.48 \pm 8.65$ | $172.13 \pm 8.18$ | $165.44 \pm 9.57$ | $171.74 \pm 8.21$ | $169.38 \pm 8.68$ | $165.2 \pm 7.8$ | $166.6 \pm 8.1$ | $165.4 \pm 7.9$ | $\begin{aligned} & 172.11 \pm \\ & 8.95 \end{aligned}$ |
| Weight (kg) | $95.58 \pm 23.27$ | $100.26 \pm 24.53$ | $106.04 \pm 23.74$ | $107.43 \pm 20.92$ | $98.49 \pm 18.59$ | $95.6 \pm 20.4$ | $104.07 \pm 28.05$ | $100.17 \pm 24.81$ | $91.3 \pm 18.9$ | $93.1 \pm 22.7$ | $87.6 \pm 18.3$ | $\begin{aligned} & 87.98 \pm \\ & 20.28 \end{aligned}$ |
| $\begin{aligned} & \text { BMI } \\ & (\mathrm{kg} / \mathrm{m} 2) \end{aligned}$ | $33.02 \pm 7.84$ | $34.99 \pm 8.31$ | $35.24 \pm 7.43$ | $35.2 \pm 6.31$ | $33.09 \pm 4.84$ | $34.86 \pm 6.63$ | $35.17 \pm 8.23$ | $34.79 \pm 8.01$ | $33.5 \pm 6.6$ | $33.5 \pm 7.8$ | $32.0 \pm 6.1$ | $\begin{aligned} & 29.66 \pm \\ & 6.37 \end{aligned}$ |
| $\begin{aligned} & \mathrm{LDL} \\ & (\mathrm{mmol} / \mathrm{L}) \end{aligned}$ | $2.74 \pm 0.96$ | $2.89 \pm 0.96$ | $2.91 \pm 1.03$ | $2.85 \pm 1.05$ | $2.89 \pm 0.94$ | $2.75 \pm 1.1$ | $2.78 \pm 0.95$ | $2.95 \pm 1.06$ | $3.4 \pm 0.86)$ | $3.2 \pm 0.90$ | $3.6 \pm 0.85$ | $\begin{aligned} & 2.36 \pm \\ & 0.64 \end{aligned}$ |
| HDL <br> ( $\mathrm{mmol} / \mathrm{L}$ ) | $1.18 \pm 0.4$ | $1.18 \pm 0.38$ | $1.12 \pm 0.34$ | $1.11 \pm 0.37$ | $1.1 \pm 0.32$ | $1.21 \pm 0.23$ | $1.15 \pm 0.37$ | $1.15 \pm 0.3$ | $1.2 \pm 0.29)$ | $1.1 \pm 0.26$ | $1.0 \pm 0.28$ | $\begin{aligned} & 1.33 \pm \\ & 0.32 \end{aligned}$ |
| $\begin{aligned} & \text { Gout } \\ & \mathrm{n}(\%) \end{aligned}$ | 356 (43.95) | 91 (52.91) | 178 (55.97) | 79 (50.97) | 24 (64.86) | 15 (20.83) | 102 (43.97) | 170 (62.96) | - | - | - | 0 (0.0) |
| Type 2 Diabetes n (\%) | 200 (27.47) | 53 (33.97) | 61 (21.79) | 35 (27.13) | 7 (24.14) | 19 (27.14) | 48 (22.64) | 62 (23.66) | 508 (17.8) | 87 (10.1) | - | 0 (0.0) |
| 745 |  |  |  |  |  |  |  |  |  |  |  |  |
| 746 | The proportions with type 2 diabetes and gout are calculated from individuals without missing data. |  |  |  |  |  |  |  |  |  |  |  |
| 747 | Data are mean $\pm$ Standard Deviation or n (\%), unless otherwise stated. |  |  |  |  |  |  |  |  |  |  |  |
| 748 | ${ }^{\text {a }}$ Data were unavailable for lipid-lowering medications. |  |  |  |  |  |  |  |  |  |  |  |
| 749 | New Zealand (NZ), Cook Islands (CI) |  |  |  |  |  |  |  |  |  |  |  |

Table 2. Minor Allele Frequency and Hardy-Weinberg Equilibrium of rs1597000001

|  | $\begin{gathered} \text { Aotearoa NZ } \\ \text { Discovery Cohort } \\ \hline \end{gathered}$ |  |  |  |  |  |  |  | SamoanReplication Cohorts |  |  | Pacific Trust Otago Cohort |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\begin{gathered} \text { NZ } \\ \text { Māori } \end{gathered}$ | $\underset{\text { Māori }}{\text { CI }}$ | Samoan | Tongan | Niuēan | Pukapukan | Other/Mixed | $\begin{gathered} \hline \text { Ngāti } \\ \text { Porou } \\ \text { Hauora } \end{gathered}$ | Samoan I | Samoan II | Samoan III | Polynesian | Non-Polynesian <br> Pacific peoples |
| Total (n) | 814 | 172 | 322 | 155 | 37 | 72 | 232 | 270 | 2851 | 908 | 550 | 220 | 35 |
| $\begin{gathered} \text { CC (n) } \\ (\%) \end{gathered}$ | $\begin{array}{\|c\|} \hline 760 \\ (93.4) \end{array}$ | $\begin{gathered} 159 \\ (92.4) \end{gathered}$ | $\begin{gathered} 297 \\ (92.2) \end{gathered}$ | $\begin{gathered} 143 \\ (92.3) \end{gathered}$ | $\begin{gathered} 33 \\ (89.2) \end{gathered}$ | $\begin{gathered} 72 \\ (100.0) \end{gathered}$ | $\begin{gathered} 212 \\ (91.4) \end{gathered}$ | $\begin{gathered} 257 \\ (95.2) \end{gathered}$ | $\begin{aligned} & 2600 \\ & (91.2) \end{aligned}$ | $\begin{gathered} 825 \\ (90.9) \end{gathered}$ | $\begin{gathered} 497 \\ (90.4) \end{gathered}$ | $\begin{gathered} 201 \\ (91.4) \end{gathered}$ | $\begin{gathered} 35 \\ (100.0) \end{gathered}$ |
| $\begin{gathered} \text { CT (n) } \\ (\%) \end{gathered}$ | $\begin{array}{\|c\|} \hline 52 \\ (6.4) \end{array}$ | $\begin{gathered} 13 \\ (7.6) \end{gathered}$ | $\begin{gathered} 24 \\ (7.5) \end{gathered}$ | $\begin{gathered} 11 \\ (7.1) \end{gathered}$ | $\begin{gathered} 4 \\ (10.8) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ | $\begin{gathered} 19 \\ (8.2) \end{gathered}$ | $\begin{gathered} 13 \\ (4.8) \end{gathered}$ | $\begin{aligned} & 242 \\ & (8.5) \end{aligned}$ | $\begin{gathered} 80 \\ (8.8) \end{gathered}$ | $\begin{gathered} 53 \\ (9.6) \end{gathered}$ | $\begin{gathered} 18 \\ (8.2) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ |
| $\begin{gathered} \text { TT (n) } \\ (\%) \end{gathered}$ | $\begin{gathered} 2 \\ (0.3) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ | $\begin{gathered} 1 \\ (0.3) \end{gathered}$ | $\begin{gathered} 1 \\ (0.7) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ | $\begin{gathered} 1 \\ (0.4) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ | $\begin{gathered} 9 \\ (0.3) \end{gathered}$ | $\begin{gathered} 3 \\ (0.3) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ | $\begin{gathered} 1 \\ (0.5) \end{gathered}$ | $\begin{gathered} 0 \\ (0.0) \end{gathered}$ |
| MAF | 0.034 | 0.039 | 0.040 | 0.042 | 0.054 | 0.000 | 0.045 | 0.024 | 0.046 | 0.047 | 0.048 | 0.045 | 0.000 |
| HWE $P$ | 0.244 | 1.000 | 0.409 | 0.231 | 1.000 | 1.000 | 0.377 | 1.00 | 0.191 | 0.448 | 0.629 | 0.364 | 1.000 |

Figure 1

$N$-terminal $\beta$-barrel
medRxiv preprint doi: https://doi.org/10.1101/2021.09.11.21263438; this version posted September 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the authon/fugdef, Q 140 has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.
a

## Cohort

Aotearoa NZ Discovery Cohort
NZ Māori
CI Māori
Samoan
Tongan
Niuēan
Mixed/Other
Ngāti Porou Hauora
Overall Effect

Samoan Replication Cohort
Samoan I
Samoan II
Samoan III
Overall Effect

Overall Effect

N Beta $\quad[95 \% \mathrm{Cl}] \quad \mathbf{P}$
8140.390 [ 0.291; 0.488] 1.06e-14
1720.411 [ $0.200 ; 0.622] 1.36 \mathrm{e}-04$ 3220.192 [ $0.065 ; 0.319] 3.00 \mathrm{e}-03$ 1550.340 [ $0.154 ; 0.526] 3.32 e-04$ 370.178 [-0.138; 0.494] 2.69e-01 2320.261 [ $0.108 ; 0.414]$ 8.08e-04 2700.260 [ $0.100 ; 0.421] 1.45 \mathrm{e}-03$ 0.305 [ 0.248; 0.362]

```
2851 0.225 [ 0.190; 0.260] 7.47e-36
908 0.193 [0.138;0.248] 5.47e-12
550 0.242 [0.171; 0.313] 1.79e-11
    0.220 [ 0.192; 0.247]
```

0.236 [ $0.211 ; ~ 0.260]$
b

Cohort


N Beta $\quad[95 \% \mathrm{Cl}] \quad \mathbf{P}$

| 759 | -0.202 | $[-0.462 ; 0.057]$ | $1.26 \mathrm{e}-01$ |
| :---: | ---: | :---: | :---: |
| 162 | -0.060 | $[-0.628 ; 0.507]$ | $8.35 \mathrm{e}-01$ |
| 312 | 0.095 | $[-0.309 ;$ | $0.499]$ |
| 144 | -0.246 | $[-0.846 ; 0.354]$ | $4.22 \mathrm{e}-01$ |
| 31 | 0.581 | $[-0.828 ; 1.989] 4.19 \mathrm{e}-01$ |  |
| 214 | 0.186 | $[-0.231 ;$ | $0.603] 3.81 \mathrm{e}-01$ |
| 235 | 0.091 | $[-0.483 ; 0.666] 7.55 \mathrm{e}-01$ |  |
|  | -0.046 | $[-0.212 ; 0.120]$ |  |

$2851-0.134[-0.238 ;-0.030] 1.15 e-02$
$882-0.276$ [-0.462; -0.090] 3.67e-03
$542-0.077$ [-0.312; 0.158] 5.21e-01 -0.156 [-0.241; -0.071]
-0.133 [-0.209; -0.058]

Figure 3
medRxiv preprint doi: https://doi.org/10.1101/2021.09.11.21263438; this version posted September 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity
PTO Cohort All rights reserved. No reuse allowed without permission.

HDL
LDL


Figure 4
medRxiv preprint doi: https://doi.org/10.1101/2021.09.11.21263438; this version posted September 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.


Figure 5
 preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity Alll rights reserved. No rewsenallow

b Conditioned on rs1597000001


C
Conditioned on rs1800775



B


Supplementary Figure 1. HDL- and LDL-C (mmol/L) distributions in the Aotearoa New Zealand cohort separated into Pacific nation of grand-parental ethnicity and Ngāti Porou Hauora subgroups and the Pacific Trust Otago. HDL-C, High-density lipoprotein cholesterol. LDL-C, Low-density lipoprotein cholesterol. NZ Māori, New Zealand Māori. CI Māori, Cook Island Māori. NPH, Ngāti Porou Hauora, PTO, Pacific Trust Otago. Datasets that do not share a letter are statistically significantly different ( p $<0.05$ ) as assessed by a post-hoc tukey test.
 medRxiv preprint doi: https://doi.org/10. All rights reserved. No reuse allowed without permission.


## Conditioned on rs1597000001



Supplementary Figure 2. Association of HDL-C levels at the CETP locus in the Samoan I cohort. Association of HDL-C at the CETP locus (+/- 500kb the lead variant rs1597000001) (A) and after conditioning on rs1597000001 (B) using using imputed data for participants from the Samoan cohort I. The strength of LD, as measured by the r2, between each variant and rs1597000001 (A) and rs11076175 (B) is represented by the color of each point according to the legend in the top right hand corner. The plot was generated using LocusZoom (Pruim et al. 2010). CETP, cholesteryl ester transfer protein. HDL-C, high-density lipoprotein cholesterol.

