

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

3

4 Jaye Moors^{1#}, Mohanraj Krishnan^{2#}, Nick Sumpter³, Riku Takei³, Matt Bixley¹, Murray
5 Cadzow¹, Tanya J. Major¹, Amanda Phipps-Green¹, Ruth Topless¹, Marilyn Merriman¹,
6 Malcolm Rutledge¹, Ben Morgan¹, Jenna C. Carlson^{2,4}, Jerry Z. Zhang⁴, Emily M. Russell²,
7 Guangyun Sun⁵, Hong Cheng⁵, Daniel E. Weeks^{2,4}, Take Naseri^{6,7}, Muagututi‘a Sefuiva
8 Reupena⁸, Satupa‘itea Viali⁹, John Tuitele¹⁰, Nicola L. Hawley¹¹, Ranjan Deka⁵, Stephen T.
9 McGarvey⁷, Janak de Zoysa¹², Rinki Murphy¹², Nicola Dalbeth¹², Lisa Stamp¹³, Mele
10 Taumoepeau¹⁴, Frances King¹⁵, Philip Wilcox¹⁶, Sally McCormick¹, Ryan L. Minster², Tony R.
11 Merriman^{1,3*} and Megan Leask^{1,3}

- 12 1. Department of Biochemistry, University of Otago, Dunedin, New Zealand
13 2. Department of Human Genetics, Graduate School of Public Health, University of
14 Pittsburgh, Pittsburgh, PA, USA
15 3. Division of Clinical Rheumatology and Immunology, University of Alabama at
16 Birmingham, Birmingham, AL, USA
17 4. Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh,
18 Pittsburgh, PA, USA
19 5. Department of Environmental and Public Health Sciences, College of Medicine,
20 University of Cincinnati, Cincinnati, OH, USA
21 6. Ministry of Health, Apia, Samoa
22 7. International Health Institute, Department of Epidemiology, School of Public Health,
23 Brown University, Providence, RI, USA

- 24 8. Lutia i Puava 'ae Mapu i Fagalele, Apia, Samoa
- 25 9. School of Medicine, National University of Samoa, Apia, Samoa
- 26 10. Department of Public Health, Lyndon B. Johnson Tropical Medical Center, Faga'alu,
- 27 American Samoa, USA
- 28 11. Department of Chronic Disease Epidemiology, Yale School of Public Health, New
- 29 Haven, CT, USA
- 30 12. Department of Medicine, University of Auckland, Auckland, New Zealand
- 31 13. Department of Medicine, University of Otago, Christchurch, New Zealand
- 32 14. Department of Psychology, University of Otago, Dunedin, New Zealand
- 33 15. Ngāti Porou Hauora, Te Puia Springs, New Zealand
- 34 16. Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand

35 #These authors contributed equally.

36

37 Correspondence to:

38 Prof Tony Merriman

39 Division of Clinical Rheumatology and Immunology, University of Alabama at Birmingham,

40 Birmingham, AL, USA

41 Email: trmerriman@uabmc.edu

42

43

44

45

46

47 **ABSTRACT**

48 Sequencing of *CETP* in Māori and Pacific peoples identified a common (MAF ~2.4%-5.4%)
49 population-specific missense variant (rs1597000001, *CETP*:c.530C>T p.Pro177Leu) that
50 associates with higher HDL-C levels ($\hat{\beta}_{\text{HDLmeta}} = 0.236$ mmol/L [95% CI 0.211; 0.260]) and
51 lower LDL-C ($\hat{\beta}_{\text{LDLmeta}} = -0.133$ mmol/L [95% CI -0.209; -0.058]). In a subsample of the study
52 cohort (n = 11), heterozygous carriers of the population-specific variant had lower plasma *CETP*
53 activity ($P = 0.028$). Our study identifies a population-specific missense variant in *CETP* which
54 lowers *CETP* activity with an effect on HDL-C that is comparable to Mendelian *CETP* loss-of-
55 function mutations.

56

57 **Word count: 2889**

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

59

60

61

62

63

64

65

66

67

68

69 Dyslipidaemia, defined as elevated total or low-density lipoprotein cholesterol (LDL-C) levels,
70 triglycerides, and/or lower levels of high-density lipoprotein cholesterol (HDL-C), is an
71 established risk factor for metabolic diseases such as cardiovascular disease, type 2 diabetes and
72 gout¹. In gout, elevated triglycerides may contribute to increased urate levels ², and elevated very
73 low-density-lipoprotein (VLDL) associates with gout in Aotearoa New Zealand (NZ) Māori and
74 Pacific peoples (herein defined as peoples of Micronesian, Melanesian and Polynesian ethnicity)
75 ³. Dyslipidaemia is more prevalent in Māori and Pacific populations ^{4,5}, and a higher proportion
76 of Pacific nation ancestry has been associated with lower HDL-C ^{6,7}. Consistent with heritability
77 estimates of HDL-C levels (40%–60%) ^{8,9}, the prevalence of dyslipidaemia in Māori and Pacific
78 populations suggests that genetics contributes to an atherogenic lipid profile ¹⁰. While socio-
79 economic inequities contribute to the development and increased impact of metabolic diseases
80 ^{10,11}, research on population-specific genetic variants associated with metabolic traits is important
81 to provide insights into the genetic determinants of phenotypic differences between populations,
82 as well as furthering genomic justice ¹².

83
84 One locus with genetic variation that consistently associates with lipid levels in multiple
85 population groups, including Pacific peoples, is at the gene that encodes cholesteryl ester transfer
86 protein: CETP ¹³⁻¹⁶. CETP modifies lipid levels by mediating the bi-directional transfer of
87 cholesterol esters from atheroprotective HDL to atherogenic VLDL in exchange for triglycerides,
88 and vice versa. Various common polymorphisms including the *CETP* promoter variant
89 ‘–629 C/A’ (rs1800775), intronic ‘Taq1B’ variant (rs708272), missense variant p.Ile405Val
90 (rs5882) and loss-of-function variant p.Asp459Gly (rs2303790) have been associated with serum
91 lipid profiles, CETP levels and/or CETP activity, and cardiovascular disease and

92 pharmacological response to CETP inhibitors¹⁷⁻²⁰. In a GWAS for lipids in Samoans (using
93 samples analyzed in greater depth here), the *CETP* locus was identified as the most significantly
94 associated locus for HDL-C levels (rs289708, $P = 1.19 \times 10^{-11}$)¹⁵. Given the previously reported
95 associations with lipid levels and variation at *CETP* in Pacific peoples^{15,21}, we used a discovery
96 and replication study design²² to investigate whether population-specific genetic variation in
97 *CETP* could contribute to HDL-C and CETP activity in Māori and Pacific populations.

98

99 RESULTS

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

101 We extracted the gene region of *CETP* (~22 Kb) from whole-genome sequence data for 55
102 individuals of Māori and Pacific ethnicity, and identified a common missense variant in exon 6
103 of the *CETP* gene (dbSNP (build 155) rs1597000001, *CETP*:c.530C>T, p.Pro177Leu). The
104 minor allele frequency (MAF) was 3.4% in the 55 discovery genomes and 0.0036% in GnomAD
105²³. A CADD score²⁴ of 24.4 placed the variant in the top 1% of predicted deleteriousness and a
106 GERP score²⁵ of 4.4 indicated that the variant is conserved in mammals (Figure 1A, CADD and
107 GERP tracks, respectively). The p.Pro177Leu amino acid substitution caused by the
108 rs1597000001 T-allele corresponds to amino acid position 160 in the mature protein structure of
109 CETP²⁶ owing to the first 17 amino acid residues of the CETP sequence consisting of the signal
110 peptide²⁷ (Figure 1B).

111

112 rs1597000001 associates with higher HDL-C

113 We genotyped rs1597000001 in a discovery cohort of 2,270 Māori and Pacific individuals living
114 in Aotearoa NZ. Demographic and biochemical characteristics of the participants are

115 summarized in Table 1. A similar MAF to the 55 Māori and Pacific genome sequences was
116 observed (MAF = 3.5%), and the MAF ranged from 3.2% - 5.4% in the Pacific nation subgroups
117 (stratified by self-reported Pacific nation of grandparental ethnicity) (Aotearoa NZ NZ Māori
118 MAF 3.4%, Aotearoa NZ Cook Island Māori MAF 3.8%, Aotearoa NZ Samoan MAF 4.0%,
119 Aotearoa NZ Tongan MAF 4.2%, Aotearoa NZ Niuēan MAF 5.4% and Aotearoa NZ
120 Other/Mixed MAF 4.5%). People of Pukapukan ethnicity were monomorphic for the C allele of
121 rs1597000001 (Table 2). The MAF was 2.4% in the Ngāti Porou Hauora group. rs1597000001
122 did not significantly deviate from Hardy-Weinberg equilibrium (HWE) in any of the Aotearoa
123 NZ Pacific nation subgroups or the Ngāti Porou Hauora group (Table 2: $P > 0.05$, all datasets).
124
125 The mean concentration and distributions of HDL-C and LDL-C (mmol/L) were not different
126 between the Pacific nation subgroups or the Ngāti Porou Hauora group in the Aotearoa NZ
127 cohort (Table 1 and Supplementary Figure 1 (post-hoc Tukey test $p > 0.05$, all datasets)). To test
128 if rs1597000001 associates with HDL-C and LDL-C levels, we carried out a linear model for
129 association of rs1597000001 adjusted by age, sex, principal component vectors (PCs) and
130 relatedness in the Aotearoa NZ dataset (Figure 2). Significantly higher HDL-C levels were
131 evident in those with the rs1597000001 T allele, and this association was observed in all of the
132 surveyed Pacific nation subgroups (Aotearoa NZ NZ Māori ($\hat{\beta} = 0.390$ mmol/L [95% CI 0.291;
133 0.488]), Aotearoa NZ Cook Island Māori ($\hat{\beta} = 0.411$ mmol/L [95% CI 0.200; 0.622]), Aotearoa
134 NZ Samoan ($\hat{\beta} = 0.192$ mmol/L [95% CI 0.065; 0.319]), Aotearoa NZ Tongan ($\hat{\beta} = 0.340$
135 mmol/L [95% CI 0.154; 0.526]), and Aotearoa NZ Other/Mixed Pacific ($\hat{\beta} = 0.261$ mmol/L
136 [95% CI 0.108; 0.414])) with the exception of the Aotearoa NZ Niuēan subgroup where there
137 was no evidence of association ($\hat{\beta} = 0.178$ mmol/L [95% CI -0.138; 0.494]) (Figure 2A). An

138 association with HDL was also observed in the Ngāti Porou Hauora group ($\hat{\beta} = 0.260$ mmol/L
139 [95% CI 0.100; 0.421]).

140

141 A fixed-effect meta-analysis showed a significant overall association of rs1597000001 T-allele
142 with higher HDL-C ($\hat{\beta} = 0.305$ mmol/L [95% CI 0.2489; 0.362]), with no evidence of
143 heterogeneity between the sample sets ($P = 0.223$) (Figure 2A). The proportion of variance in
144 HDL-C levels explained by rs1597000001 in the Aotearoa NZ cohort was 4.5%, similar to the
145 proportion of variance explained by sex (5.0%). There was no association of rs1597000001 with
146 LDL-C (Figure 2B) in any of the Pacific nation subgroups, Ngāti Porou Hauora group or in a
147 fixed effect meta-analysis of the Aotearoa NZ cohort ($\hat{\beta} = -0.046$ mmol/L [95% CI -0.212;
148 0.120] (Figure 2B)).

149

150 We carried out association analyses in three independent Samoan cohorts (Samoa I-III) which
151 serve as replication cohorts to validate the associations we observed in the Aotearoa NZ
152 discovery cohort. Association analyses adjusted by age, sex, PCs and relatedness confirmed the
153 association between HDL-C and rs1597000001 (Samoa cohort I: $\hat{\beta} = 0.225$ mmol/L [95% CI
154 0.190; 0.260], Samoa cohort II: $\hat{\beta} = 0.193$ mmol/L [95% CI 0.138; 0.248] and Samoa cohort
155 III: $\hat{\beta} = 0.242$ mmol/L [95% CI 0.171; 0.313] (Figure 2A)) with no evidence of heterogeneity
156 between the sample sets ($P = 0.502$). Unlike the Aotearoa NZ cohort there was an association of
157 rs1597000001 with lower LDL-C in Samoa cohorts I and II (Samoa cohort I: $\hat{\beta} = -0.134$
158 mmol/L [95% CI -0.238; -0.030] and Samoa cohort II: $\hat{\beta} = -0.276$ mmol/L [95% CI -0.462; -
159 0.090] (Figure 2B)) but not Samoa III (Samoa cohort III: $\hat{\beta} = -0.077$ mmol/L [95% CI -0.312;
160 0.158] (Figure 2B)). The MAF was 4.4% in Samoa I, 4.7% in Samoa II and 4.8% in Samoa

161 III (compared to 4.0% in the Aotearoa NZ Samoan cohort) and did not deviate from HWE ($P >$
162 0.05, Samoan I-III (Table 2)). The proportion of variance of HDL-C explained by rs1597000001
163 in the Samoan I and Samoan II and Samoan III cohorts was 5.1%, 6.4% and 7.8% respectively.

164

165 A fixed-effect meta-analysis of the Aotearoa NZ Pacific nation subgroups, Ngāti Porou Hauora
166 group and Samoan I-III cohorts demonstrated the strong association of rs1597000001 with HDL-
167 C levels ($\hat{\beta}_{\text{HDLmeta}} = 0.236$ mmol/L [95% CI 0.211; 0.260]) with no heterogeneity between the
168 cohorts ($P = 0.054$). rs1597000001 associated with lower LDL-C levels in the fixed-effect meta-
169 analysis of the Aotearoa NZ and Samoan cohorts I-III ($\hat{\beta}_{\text{LDLmeta}} = -0.133$ mmol/L [95% CI -
170 0.209; -0.058]) (Figure 2B) with no heterogeneity between the cohorts ($P = 0.534$).

171

172 Next we explored what effect this variant has on HDL-C and LDL-C levels in a cohort of young
173 Pacific people (PTO cohort) before onset of metabolic disease indicators (Table 1). The mean
174 concentration of HDL-C was significantly higher in the PTO cohort compared to all Pacific
175 nation subgroups and the Ngāti Porou Hauora group in the Aotearoa NZ cohort (Table 1 and
176 Supplementary Figure 1A (post-hoc Tukey test $P = < 0.05$; all datasets)). LDL-C (mmol/L) was
177 significantly lower in the PTO cohort in comparison to all Pacific nation subgroups and the Ngāti
178 Porou Hauora group in the Aotearoa NZ cohort (Table 1 and Supplementary Figure 1B (post-hoc
179 Tukey test $P = < 0.05$; all datasets)). rs1597000001 was monomorphic for the major C allele in
180 Pacific participants without Polynesian ethnicity (i.e. of Melanesian and/or Micronesian
181 ethnicity) (Table 2). In PTO participants with Polynesian ethnicity the MAF was 4.5% and did
182 not deviate from HWE ($P > 0.05$). Association analysis (adjusted by age, sex, and grandparental
183 ethnicity) in participants of Polynesian ethnicity indicated that rs1597000001 associates with

184 higher HDL-C levels ($\hat{\beta}$ = 0.366 mmol/L [95% CI 0.221; 0.511] (Figure 3A)) and lower LDL-C
185 levels ($\hat{\beta}$ = -0.340 mmol/L [95% CI -0.663; -0.016] (Figure 3B)) consistent with the negative
186 estimates from the Samoan I and Samoan II cohorts. The proportion of variance of HDL-C
187 explained by rs1597000001 in the PTO cohort was 11.8%.

188

189 **rs1597000001 associates with decreased CETP activity**

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

195

196 **Multiple independent genetic associations are identified at CETP in Māori and Pacific** 197 **People**

198 Genetic contribution to HDL-C is polygenic, and patients with extreme HDL concentrations
199 possess both rare large-effect variants and common small-effect variants²⁸ at *CETP* that
200 contribute to HDL-C²⁹. We re-examined¹⁵ the entire *CETP* locus for association with HDL-C
201 levels in Māori and Pacific peoples in the Aotearoa NZ cohort. rs1597000001 was the maximally
202 associated variant followed by the *CETP* promoter polymorphism '-629 C/A' rs1800775 (Figure
203 5A). The rs1800775 C-allele was present in the Aotearoa NZ cohort (46.0%) at a frequency
204 similar to Europeans (48.0% in gnomAD), and exhibited weak linkage disequilibrium with
205 rs1597000001 (r^2 = 0.032). The effect size for rs1800775 C-allele ($\hat{\beta}$ = -0.090 mmol/L [95% CI
206 0.112; -0.069] was small in comparison to the effect of rs1597000001 ($\hat{\beta}$ = 0.301 mmol/L) and

207 the proportion of variance of HDL-C explained by rs1800775 in the Aotearoa NZ cohort was
208 3.3% compared to 4.5% for rs1597000001. The rs1800775 C-allele associates with the same
209 direction of effect previously observed for Europeans ($\hat{\beta} = -0.080$ mmol/L, $p = 3.7 \times 10^{-93}$)¹³ and
210 Pacific peoples ($\hat{\beta} = -0.055$ mmol/L, $P = 1.7 \times 10^{-4}$)²¹.

211
212 Given the large effect of rs1597000001 on HDL-C and the fact that rs1597000001 and
213 rs1800775 exhibit some linkage disequilibrium we carried out conditional analyses to test
214 whether the effects at rs1800775 and rs1597000001 were dependent on each other. In the
215 Aotearoa NZ cohort, the effect on HDL-C persisted for rs1597000001 ($\hat{\beta} = 0.266$ mmol/L [95%
216 CI 0.210; 0.323]) conditioned on rs1800775 genotype (Figure 5B) and for rs1800775
217 conditioned on rs1597000001 genotype (rs1800775 C-allele; $\hat{\beta} = -0.072$ mmol/L [95% CI -
218 0.094; -0.052]) (Figure 5C). However, after conditioning on rs1597000001, rs183130 was the
219 most significantly associated variant (T-allele; $\hat{\beta} = 0.087$ mmol/L [95% CI 0.062; 0.112]) at
220 *CETP* (r^2 with rs1800775 = 0.31). In the Samoan I cohort, conditioning on rs1597000001
221 resulted in two additional signals marked by rs11076175 (G-allele; $\hat{\beta} = -0.0623$ mmol/L [95% CI
222 -0.0876; -0.0369]) and rs4783961 (A-allele; $\hat{\beta} = 0.0480$ mmol/L [95% CI 0.0279; 0.0681])
223 (Supplementary Figure 2). rs4783961 exhibited linkage disequilibrium with rs1800775 ($r^2 =$
224 0.227) and rs183130 ($r^2 = 0.623$). However, rs11076175 had only weak linkage disequilibrium
225 with rs1800775 ($r^2 = 0.114$) and rs183130 ($r^2 = 0.014$).

226

227 **DISCUSSION**

228 Using whole-genome sequence data from individuals of Māori and Pacific ethnicity, we have
229 identified a Population-specific missense variant in *CETP* that strongly associates with a favorable

230 lipid profile (higher HDL-C and lower LDL-C) in a meta-analysis of four independent Pacific
231 cohorts. The positive effect on HDL-C levels of the rs1597000001 T-allele was observed in every
232 group analysed, with the exception of the Niuēan Pacific nation subgroup. This is likely a reflection
233 of the small sample size of this subgroup, limiting our detection power. However, the effect size
234 estimate was positive, which is consistent with all the other island nation subgroups in the Aotearoa
235 NZ cohort. Given the likely, but unstudied, nutritional and environmental variations across these
236 groups within Aotearoa NZ and between Aotearoa NZ and the Samoan Islands, it is noteworthy
237 that similar effect sizes were detected.

238

239 The effects of rs1597000001 on LDL-C are less conclusive. In the Samoan I and II cohorts, an
240 association with lower LDL-C was observed. However, the Aotearoa NZ, Samoan III, and PTO
241 cohorts did not show an association of rs1597000001 with LDL-C levels. The differences could
242 be due to unmeasured heterogeneity in medication use and other genetic and environmental
243 factors. Our findings that rs1597000001 associates with HDL-C consistently throughout Pacific
244 nation groups but not LDL-C corroborate data from Mendelian randomization studies of CETP
245 which indicate that CETP activity is an important causal determinant of HDL-C³⁰ levels but not
246 LDL-C levels.

247

248 Variants in *CETP* that cause loss-of-function and consequently have large effects on HDL-C
249 levels are rare in the general population and are enriched in study populations that have high
250 HDL-C levels³¹⁻³³. Conversely rs1597000001 (MAF ~ 2.4%-5.4%) is a common variant of large
251 effect in Māori and Pacific peoples, and the allele frequency was stable throughout the Pacific
252 nations surveyed here. It is notable that the variant was not detected in people of Pukapukan

253 ethnicity and Pacific peoples without Polynesian ethnicity. The most probable explanation is that
254 the absence of the variant from people of Pukapukan ethnicity is the result of a different
255 population history for this group for example bottle-necking event(s) and/or founder effects. That
256 we did not observe the variant in non-Polynesian Pacific people suggests that this is a
257 Polynesian-specific variant.

258

259 The large effect on HDL-C level that we observe for rs1597000001 is reminiscent of that
260 observed for the common hyperalphalipoproteinemia 1 (HALP1) (elevated HDL-C levels) loss-
261 of-function variant rs2303790 (p.Asp259Gly; $\hat{\beta}_{\text{HDL}} = 0.44 \text{ mmol/L}$)²⁰ located in exon 15 of
262 *CETP*. Functionally rs2303790 reduces CETP activity, likely emulating Anacetrapib inhibition
263 of CETP³⁴. Anacetrapib on average raises HDL-C levels by 1.12 mmol/L in patients with
264 atherosclerotic vascular disease³⁵. Given that rs1597000001 is located in exon 6, it is likely that
265 its functional effect on CETP is disparate to rs2303790. However, it is notable that the effect of
266 one allele of rs1597000001 on HDL-C is ~25% of that observed for Anacetrapib. Here, we
267 showed that heterozygous carriers of rs1597000001 have lower CETP activity which supports
268 our hypothesis that the rs1597000001 T-allele results in dysfunctional CETP and consequently
269 increases HDL-C.

270

271 The magnitude of effect on HDL-C of rs1597000001 is also comparable to carrier status for
272 protein truncating variants in *CETP* ($\hat{\beta} = 0.585 \text{ mmol/L}$)³⁶. We note that rs1597000001 is in very
273 close proximity to a rare truncating variant p.Trp179* which introduces a stop codon in exon 6.
274 This rare variant has only been identified in people with HALP1³¹ thus we are unable to make
275 comparisons of the estimates of effect on HDL. p.Trp179* could possibly alter CETP levels

276 through nonsense-mediated decay of the truncated mRNA. Alternatively, p.Trp179* could also
277 functionally disrupt the CETP protein. p.Trp179* and rs1597000001 (p.Pro177Leu) are located
278 in the N-terminal β -barrel domain of CETP which is predicted to penetrate the HDL surface ³⁷,
279 and is a putative entry site for cholesterol ester. Both variants are contained in the Ω 6 flap
280 (Gln¹⁵⁵ – Trp¹⁶²) which separates from the Ω 5 flap upon HDL penetration, then opens the N-
281 terminal distal end for uptake of cholesterol ester ³⁷. Specifically, Trp¹⁶² plays an important
282 anchoring role in the penetration and settling of CETP in HDL. Binding of CETP to the HDL
283 substrate occurs via hydrophobic interactions ³⁸, and it could therefore be predicted that the
284 p.Pro177Leu substitution from proline to the hydrophobic amino acid leucine at this site could
285 alter substrate binding and therefore cholesterol ester transfer activity of CETP. Alternatively,
286 proline residues commonly form bends and turns in protein structure ³⁹, thus the substitution to
287 leucine could disrupt the conformation of the CETP N-terminal domain.

288

289 Although dyslipidaemia is prevalent in people of Māori and Pacific ethnicity, our study has
290 identified a population-specific variant that associates with an improved lipid profile. It remains
291 to be seen whether this variant is also associated with lower cardiovascular event risk. Although
292 the variant has a tangible biological mechanism of effect and results in CETP deficiency,
293 understanding how this variant is modified by other common, smaller-effect variants within the
294 locus and how these variants function will be important information for more effective targeted
295 interventions with pharmaceuticals ⁴⁰. Most importantly, this variant and the accumulating
296 evidence of the presence of other Population-specific trait-associated variants ⁴¹⁻⁴⁵ emphasizes
297 the significance of population-specific variation and its influence on disease traits and risk.
298 Understanding the genetic basis of metabolic disease in Māori and Pacific populations is

299 essential for the improvement of health outcomes for these groups ¹⁰, and comprehensive
300 evaluations of genome-wide genetic variation is long overdue and critical for equity in healthcare
301 as we move towards an era of personalized medicine grounded in genetics.

302

303 **METHODS**

304 **Study cohorts**

305 Demographic and biochemical characteristics of the participants are summarized in Table 1.

306 2,272 participants of Māori and Pacific ethnicity were recruited in Aotearoa New Zealand (NZ).

307 The Aotearoa NZ cohort serves as the discovery cohort in this study. 4,309 participants of

308 Samoan ethnicity were recruited in the Independent State of Samoa and the US territory of

309 American Samoa into the Samoan I (n = 2,851), II (n = 908) and III (n = 550) cohorts which

310 serve as the replication cohorts in this study. Finally, an additional 255 young people (aged 14 –

311 25 years) identifying with at least one Pacific Island ethnicity within Oceania (i.e., Melanesia,

312 Micronesia or Polynesia) residing in Dunedin, New Zealand were recruited by the Pacific Trust

313 Otago into the Pacific Trust Otago (PTO) Cohort.

314

315 For all participants information obtained at recruitment included age, sex, height, and weight, as

316 measured by trained assessors. BMI was calculated by dividing an individual's weight by the

317 square of their height in meters. Blood biochemical measurements including lipid measurements

318 were performed at Southern Community Laboratories (Dunedin, NZ) for the Aotearoa NZ

319 participants, at Northwest Lipid Labs (Seattle, WA, USA) for the Samoan I cohort, and at Lipids

320 Research Clinic at Miriam Hospital, Brown University for the Samoan II and III cohorts.

321

322 The Aotearoa NZ cohort is the amalgamation of three separate groups. 2002 participants aged \geq
323 16 years, located primarily in Auckland and Christchurch were recruited to Genetics of Gout,
324 Diabetes and Kidney Disease in Aotearoa NZ Study⁴¹. The participants from this study were
325 separated into subgroups based on the self-reported Pacific nation of ethnicity of their
326 grandparents. Those participants who also reported some non-Pacific grandparent ethnicity were
327 grouped according to their Pacific nation ethnicity. This resulted in six sample sets: Aotearoa NZ
328 NZ Māori ($n = 814$), Aotearoa NZ Cook Island Māori ($n = 172$), Aotearoa NZ Samoan
329 ($n = 322$), Aotearoa NZ Tongan ($n = 155$), Aotearoa NZ Niuēan ($n = 37$) and an ‘Other/Mixed’
330 Aotearoa NZ Pacific group ($n = 232$), which included individuals of Tahitian ($n = 1$) and
331 Tuvaluan ($n = 5$) ethnicity, along with individuals who self-reported grandparental ethnicity from
332 more than one Pacific nation ($n = 230$). An additional 270 participants from Te Tairāwhiti (east
333 coast of the North Island, NZ) were recruited in collaboration with the Ngāti Porou Hauora
334 (Health Service) Charitable Trust. At the request of Ngāti Porou Hauora (NPH) these participants
335 were analyzed separately to the participants in the six Pacific nation subgroups. 72 participants
336 of Pukapukan ethnicity were recruited in collaboration with the Pukapukan Community of New
337 Zealand Inc. in Mangere, South Auckland, NZ.

338

339 The Samoan I cohort consists of 2,851 Samoan adults aged 22–65 years residing in the
340 Independent State of Samoa. The Samoan II cohort is a family study consisting of 908 Samoan
341 adults aged 18–88 years residing in the Independent State of Samoa and the US territory of
342 American Samoa. The Samoan III cohort consists of 550 Samoan adults aged 29–88 years
343 residing in the Independent State of Samoa and the US territory of American Samoa. With the
344 exception of the participants in the Samoan II cohort, participants were also asked about the

345 ethnicity of each of their grandparents. Participants from the Samoan I and III cohorts all
346 reported four Samoan grandparents. Although participants from the Samoan II cohort did not
347 report the ethnicity of their grandparents, in principal components analysis of ancestry (see
348 below), the participants of this cohort cluster together with participants from the Samoan III
349 cohort.

350
351 Ethical approval for the Aotearoa NZ cohort study was given by the NZ Multi-Region Ethics
352 Committee (MEC/05/10/130; MEC/10/09/092; MEC/11/04/036), the Northern Y Region Health
353 Research Ethics Committee (Ngāti Porou Hauora Charitable Trust study; NTY07/07/074), and
354 the University of Otago Human Health Ethics Committee (Pacific Trust Otago; 12/349). Ethical
355 approval for the Samoan I cohort study was given by the Health Research Committee of the
356 Samoa Ministry of Health and the institutional review board of Brown University. Ethical
357 approvals for the Samoan II and III cohort studies were given by the Health Research Committee
358 of the Samoa Ministry of Health and the Institutional Review Boards (IRBs) of the Department
359 of Health in American Samoa; of The Miriam Hospital, Providence, Rhode Island; and of
360 Brown University. The consent forms for Samoan I, II, and III cohorts were available to
361 participants in both Samoan and English. All participants provided written informed consent for
362 the collection of samples and subsequent analysis. Participants recruited in Aotearoa New
363 Zealand were asked if they would like a karakia (Māori prayer) carried out upon disposal of their
364 blood samples, if indicated this was carried out. The PTO project was guided by the University
365 of Otago Pacific Research Protocol.

366

367 **Identification of rs1597000001**

368 For genomic sequencing 2 µg of total genomic DNA from 55 Māori and Pacific individuals
369 (participants with gout in Ardea Biosciences (USA) clinical trials for urate-lowering therapy
370 ^{42,46,47}) were submitted to Kinghorn Centre for Clinical Genomics at Garvan Institute of Medical
371 Research in New South Wales, Australia for library preparation and Next Generation Sequencing
372 30x WGS (TruSeq Nano v2.5). The *CETP* gene region (based on reference transcripts from
373 Ensembl (<https://asia.ensembl.org/index.html>)) was extracted from whole genome sequence data
374 in FASTQ format aligned to the human genome (GRCh37/hg19) following implementation of
375 the Genome Analysis ToolKit (GATK) best practices using the Burrows-Wheeler Aligner ⁴⁸
376 Picardtools (<https://broadinstitute.github.io/picard/>) and GATK v3.6.0 ⁴⁹ (NeSI pipeline for
377 GATK; DOI: <http://doi.org/10.5281/zenodo.2564243>). Variants in *CETP* were annotated with
378 allele frequencies from the Genome Aggregation Database (<http://gnomad.broadinstitute.org/>) ²³
379 to identify population-specific variants, and annotated using Variant Effect Predictor (VEP)
380 (www.ensembl.org/Tools/VEP) ⁵⁰. *In silico* predictions for deleteriousness of rs1597000001 and
381 conservation at rs1597000001 were obtained using the UCSC browser (GRCh37/hg19) track
382 collection ⁵¹ for Combined Annotation Dependent Depletion (CADD) scores ²⁴ and Genomic
383 Evolutionary Rate Profiling (GERP) scores for Mammalian Alignments ²⁵ respectively. The
384 *CETP* protein structure ([10.2210/pdb2OBD/pdb](https://www.rcsb.org/pdb/entry/10.2210/pdb2OBD/pdb)) ²⁶ was obtained from The Protein Databank
385 (<https://www.rcsb.org/>) ⁵² and visualized using the PyMOL molecular graphics system (v2.3.2
386 (Shrodinger, LLC)).

387

388 **Genotyping**

389 In the Aotearoa NZ, NPH and PTO cohorts, rs1597000001 was genotyped using a custom-
390 designed TaqMan probe-set (Applied Biosystems, Foster City, CA, USA). A custom Python

391 script (snp_design; DOI: <https://doi.org/10.5281/zenodo.56250>) was used to annotate the human
392 genome build 37 reference sequence (<ftp://ftp.ensembl.org/pub/grch37>) with rs159700001 and
393 any surrounding SNPs (obtained from the NCBI dbSNP build 147 common SNP
394 list; <ftp://ftp.ncbi.nlm.nih.gov/snp>) before primer and probe design. Forward primer:
395 CGGTGCCTGGTACACACTAG; reverse primer: TGTGAACAGCTGCTTGATCCA; probe 1
396 (VIC): CTGCCTTCAGGCCTG; probe 2 (FAM): CTGCCTTCAGGCTTG. Genotyping was
397 carried out using the LightCycler 480 Real-Time Polymerase Chain Reaction System (Roche
398 Applied Science, Indianapolis, IN, USA) in 384 well plates. Additional genotypes beyond
399 rs159700001 at the *CETP* locus for the Aotearoa NZ participants were obtained from the
400 Illumina Infinium CoreExome v24 bead chip platform whole-genome genotyping data generated
401 as previously described ⁴¹. In the Samoan I cohort, participants were genotyped using the
402 Genome-Wide Human SNP Array 6.0 (Affymetrix, CA, USA) with quality control as described
403 in ⁵³. rs159700001 was not included on the array, but the genotypes from the array served as the
404 scaffold for imputation of the variant. A subset of Samoan I cohort participants ($n = 1,294$) were
405 whole-genome sequenced as part of the Trans-Omics for Precision Medicine (TOPMed) program
406 ⁵⁴, a Samoan-specific haplotype reference panel was generated from the sequences using Eagle2
407 (v.2.3.5) ⁵⁵, and additional genotypes, including those of rs159700001 were imputed based on
408 the genotype scaffold and the reference panel using Minimac4 ⁵⁶. In the Samoan II and III
409 cohorts, participants were genotyped using the Infinium Global Screening Array-24 v3.0
410 BeadChip (Illumina, CA, USA) with custom content that included rs159700001. Quality
411 control and quality assurance checks for the Samoan II and III cohorts were conducted using
412 GWASTools ⁵⁷.

413

414 **Principal components and relatedness matrices**

415 Whole-genome principal component vectors were calculated for the Aotearoa NZ cohort using
416 2,858 ancestry informative markers (as identified by Illumina) extracted from the CoreExome
417 whole-genome genotypes. Ten PCs were generated from the SmartPCA (EIGENSOFT v6.0.1))
418 ⁵⁸ program and used as covariates in the association analyses to account for population
419 stratification and cryptic relatedness. Relatedness coefficients were calculated in the Aotearoa
420 NZ dataset using the software GEMMA (v0.98.4) ⁵⁹ from 257,069 independent SNPs from the
421 CoreExome whole-genome genotyping data.

422
423 For the Samoan I, II, and III cohorts, principal components and empirical kinship coefficients
424 were calculated using genotypes from the Genome-Wide Human SNP 6.0 array (Samoan I) and
425 the Infinium Global Screening Array-24 v3.0 BeadChip (Samoan II and III). In the Samoan I
426 cohort, PCs were calculated as described in Minster *et al.* ⁵³ and the empirical kinship matrix was
427 calculated from 10,000 independent autosomal markers using OpenMendel ^{53,60,61}. In the Samoan
428 II and III cohorts, 55,640 autosomal markers were used to calculate PCs and empirical kinship
429 matrices with PC-AiR and PC-Relate, respectively ⁶², per the recommended procedure in the
430 GENESIS R/Bioconductor package ⁶³.

431

432 **Statistical analyses**

433 All association analyses described below were carried out using the R v4.0.2 software. For the
434 association analyses in the Aotearoa NZ cohort, a generalized linear mixed model-based Wald test
435 was carried out using GMMAT software (v1.3.1) ⁶⁴ to test for associations between rs1597000001
436 and the continuous variables HDL-C and LDL-C. Analyses were adjusted by sex, age, 10 PCs and

437 relatedness. A linear model was used to test for associations in the PTO cohort between
438 rs1597000001 and the continuous variables HDL-C and LDL-C using the `lm()` function in R.
439 The PTO analyses were adjusted by sex, age and self-reported ethnicity group for each
440 grandparent, owing to absence of whole-genome ancestry informative markers in this cohort. For
441 the association analyses in the Samoan I cohort, linear mixed-model regression of the phenotypes
442 on rs1597000001 was performed using the `lmekin()` function of the `coxme` R package with sex,
443 age, four PCs and relatedness as covariates. In the Samoan II and III cohorts, a mixed model
444 association testing was carried out using `lmekin()` to test for associations between rs1597000001
445 and HDL-C or LDL-C phenotypes with sex, age, first four PCs, polity (Samoa or American
446 Samoa), and relatedness as adjusting variables.

447

448 Each Pacific population sample set (Aotearoa NZ cohort: Aotearoa NZ NZ Māori, Aotearoa NZ
449 Cook Island Māori, Aotearoa NZ Samoan, Aotearoa NZ Niuēan, Aotearoa NZ Pukapukan and
450 Aotearoa NZ ‘Other /Mixed’ Pacific nations subgroups; Ngāti Porou Hauora group, Samoan
451 Cohort I, Samoan Cohort II and Samoan Cohort III; and Pacific Trust Otago) was analyzed
452 separately, and the effects combined for the Aotearoa NZ cohort and all cohorts (excluding the
453 PTO cohort) in an inverse variance-weighted fixed-effect meta-analysis using the R package
454 `meta` (v.3.0-2). Heterogeneity between sample sets was assessed using Cochran’s heterogeneity
455 (Q) statistic. The proportion of variance explained by rs1597000001 and *CETP* promoter
456 common variant rs1800775 for HDL-C was calculated using the `partial.rsq()` function in
457 the `rsq` R package (v2.2)⁶⁵ which calculates the partial R-squared value for each predictor from
458 the linear model generated by the `lm()` function in R. The β coefficient in all analyses
459 represents the estimated effect on HDL-C and LDL-C units (mmol/L) per copy of the

460 rs1597000001 T-allele. For the Aotearoa NZ analyses HDL-C and LDL-C measurements were in
461 mmol/L. In the Samoan cohorts I, II and III measurements were in mg/dl and converted to
462 mmol/L before the association analyses.

463
464 To contextualize the effect of rs1597000001 among the other variation in *CETP*, variants on the
465 Illumina Infinium CoreExome v24 bead chip platform and present in the Aotearoa NZ cohort
466 (minor allele frequency (MAF) > 0.01) were extracted from the *CETP* region (rs1597000001 ±
467 500 kb). A linear model was used to test variants for association with HDL-C using PLINK
468 (v1.90b6.10) adjusted by age, sex and 10 PCs. Linkage disequilibrium and conditional analyses
469 for rs1597000001 and *CETP* promoter common polymorphism rs1800775 in the Aotearoa NZ
470 cohort were carried out in PLINK (v1.90b6.10), adjusting by age, sex and 10 PCs calculated from
471 the Aotearoa NZ dataset. Conditional analysis was conducted in Samoan cohort I across the *CETP*
472 region with rs1597000001 modeled as an additional fixed-effect covariate. Regional association
473 plots were created using a custom R package (LocusZoom-like Plots: DOI:
474 <https://doi.org/10.5281/zenodo.5154379>) for the Aotearoa NZ cohort and using LocusZoom⁶⁶ for
475 the Samoan I cohort.

476

477 **CETP activity**

478 Individuals of the PTO cohort were recalled specifically for the purpose of providing fresh plasma
479 samples for the CETP activity assays. CETP activity in 11 participants, of whom four were
480 heterozygous for rs1597000001 and seven were homozygous for rs1597000001 C-allele was
481 assessed using the CETP activity assay kit II (F) from BioVision (K595-100) according to the
482 manufacturer's instructions. Fresh plasma (1 µL) was used for the assays which were carried out

483 in technical triplicate. Fluorescence was measured using a BMG LabTech CLARIOstar plate
484 reader and CLARIOstar software (v.5.01 R2, Firmware 1.10). Using the `stat.desc()` function
485 of the `pastecs` R package, we tested whether data within each genotypic group was normally
486 distributed. We found that both genotypic groups had skewness/2SE and kurtosis/2SE values
487 between -1 and 1, indicating normality of the data. The Shapiro-Wilks tests for skewness were also
488 non-significant. A two-samples t-test with Welch's correction was conducted to test for a
489 difference in mean CETP activity between the two genotypic groups. One technical replicate was
490 removed after it was identified as an outlier based on a significant Grubbs test for one outlier ($P =$
491 0.00065) using the `grubbs.test()` function from the `outliers` R package. Statistical analyses
492 were conducted in R v4.0.2 software with all data reported as means \pm standard error of the mean.

493

494

495 **ACKNOWLEDGMENTS**

496 The authors sincerely thank the participants for generously donating their time and information to
497 this study. The authors would like to thank J. Drake (Department of Rheumatology, Canterbury
498 District Health Board, Christchurch, New Zealand), J. de Kwant, R. Laurence, C. Franklin, and
499 M. House (all Department of Medicine, University of Auckland, Auckland, New Zealand), J. Harre
500 Hindemarsch, N. Aupouri, R. Akuhata, and C. Ford (all of Ngāti Porou Hauora Charitable Trust,
501 Te Puia Springs, New Zealand), G. Sexton (Counties Manukau District Health Board, Auckland,
502 New Zealand), and B. Vincent (Waitemata District Health Board, Auckland, New Zealand) for
503 recruitment. For the Samoan cohorts, we acknowledge the Samoan Ministry of Health, the Samoa
504 Bureau of Statistics, and the American Samoan Department of Health for their support of this
505 research. The work was funded by the National Institute of Health grants R01HL093093 (STM),

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

517

518 **AUTHOR CONTRIBUTIONS**

519

520 MM, RT, TJM and AP-G coordinated the sampling, genome sequencing and genome-wide
521 genotyping for the Aotearoa NZ cohort and 55 Aotearoa NZ genome sequences. MB and MC
522 carried out the genome sequencing data analyses in the 55 genome sequences. For the Samoan I
523 cohort NH led the field work data collection and phenotype analyses under the supervision of
524 STM. For the Samoan II and III cohorts, STM led the field work data collection and phenotype
525 analyses. GS and HC performed genotyping experiments for the Samoan I cohort, and HC
526 prepared DNA for genotyping at the Center for Inherited Disease Research for the Samoan II and
527 III cohorts and for sequencing by the TOPMed Program for the sequenced subset of the
528 Samoan I cohort, under the supervision of RD. JZZ and JCC performed genotype imputation for

529 the participants in the Samoa I cohort with guidance from RLM and DEW. JM coordinated the
530 samples and genotyping for the PTO cohort and carried out the genotyping over the Aotearoa NZ
531 cohort. JM, MK, and ML carried out the association analyses with guidance from TRM, DEW,
532 RLM and assistance from NS, RT, EMR, JCC and TJM. MR and BM carried out the CETP
533 assays supervised by SM. MSR, SV, and JT facilitated fieldwork in Samoa and American
534 Samoa. TN, MSR, SV, NH, PW, FK, MT, ND, LS, RM and JdZ contributed to the discussion of
535 the public health and cultural implications of the findings. ML wrote the manuscript with
536 guidance from TRM and contribution from all co-authors. All co-authors contributed to this
537 work, discussed the results, and critically reviewed and revised the manuscript.

538

539 **COMPETING INTERESTS**

540 The authors declare no competing interests.

541

542 **DATA AVAILABILITY**

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

548

549 **CODE AVAILABILITY**

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

552 REFERENCES

- 553 1. Kolovou, G.D., Anagnostopoulou, K.K. & Cokkinos, D.V. Pathophysiology of
554 dyslipidaemia in the metabolic syndrome. *Postgrad Med J* **81**, 358-66 (2005).
- 555 2. Rasheed, H., Hughes, K., Flynn, T.J. & Merriman, T.R. Mendelian randomization provides
556 no evidence for a causal role of serum urate in increasing serum triglyceride levels.
557 *Circulation: Cardiovascular Genetics* **7**, 830-837 (2014).
- 558 3. Rasheed, H. *et al.* The relationship of apolipoprotein B and very low density lipoprotein
559 triglyceride with hyperuricemia and gout. *Arthritis research & therapy* **16**, 1-10 (2014).
- 560 4. Gentles, D. *et al.* Serum lipid levels for a multicultural population in Auckland, New
561 Zealand: results from the Diabetes Heart and Health Survey (DHAH) 2002-2003. *N Z Med*
562 *J* **120**, U2800 (2007).
- 563 5. Beaglehole, R., Prior, I.A., Eyles, E. & Sampson, V. High density lipoprotein cholesterol
564 and other serum lipids in New Zealand Maoris. *N Z Med J* **90**, 139-42 (1979).
- 565 6. Win Tin, S.T., Kenilorea, G., Gadabu, E., Tasserei, J. & Colagiuri, R. The prevalence of
566 diabetes complications and associated risk factors in Pacific Islands countries. *Diabetes*
567 *Res Clin Pract* **103**, 114-8 (2014).
- 568 7. Sun, H. *et al.* The impact of global and local Polynesian genetic ancestry on complex traits
569 in Native Hawaiians. *PLoS Genet* **17**, e1009273 (2021).
- 570 8. Qasim, A. & Rader, D.J. Human genetics of variation in high-density lipoprotein
571 cholesterol. *Curr Atheroscler Rep* **8**, 198-205 (2006).
- 572 9. Lusis, A.J. Genetics of atherosclerosis. *Trends Genet* **28**, 267-75 (2012).
- 573 10. Merriman, T.R. & Wilcox, P.L. Cardio-metabolic disease genetic risk factors among Maori
574 and Pacific Island people in Aotearoa New Zealand: current state of knowledge and future
575 directions. *Ann Hum Biol* **45**, 202-214 (2018).
- 576 11. Faatoese, A.F. *et al.* Community screening for cardiovascular risk factors and levels of
577 treatment in a rural Maori cohort. *Aust N Z J Public Health* **35**, 517-23 (2011).
- 578 12. Reardon, J. *The postgenomic condition*, (University of Chicago Press, 2017).
- 579 13. Ridker, P.M. *et al.* Polymorphism in the CETP gene region, HDL cholesterol, and risk of
580 future myocardial infarction: Genomewide analysis among 18 245 initially healthy women
581 from the Women's Genome Health Study. *Circ Cardiovasc Genet* **2**, 26-33 (2009).
- 582 14. Mirmiran, P. *et al.* Genetic variations of cholesteryl ester transfer protein and diet
583 interactions in relation to lipid profiles and coronary heart disease: a systematic review.
584 *Nutr Metab (Lond)* **14**, 77 (2017).
- 585 15. Carlson, J.C. *et al.* Genome-wide association studies in Samoans give insight into the
586 genetic architecture of fasting serum lipid levels. *J Hum Genet* **66**, 111-121 (2021).
- 587 16. Kenny, E.E. *et al.* Increased power of mixed models facilitates association mapping of 10
588 loci for metabolic traits in an isolated population. *Hum Mol Genet* **20**, 827-39 (2011).
- 589 17. Dacet, C., Poirier, O., Cambien, F., Chapman, J. & Rouis, M. New functional promoter
590 polymorphism, CETP/-629, in cholesteryl ester transfer protein (CETP) gene related to
591 CETP mass and high density lipoprotein cholesterol levels: role of Sp1/Sp3 in
592 transcriptional regulation. *Arterioscler Thromb Vasc Biol* **20**, 507-15 (2000).
- 593 18. Marschang, P. *et al.* Plasma cholesteryl ester transfer protein concentrations predict
594 cardiovascular events in patients with coronary artery disease treated with pravastatin. *J*
595 *Intern Med* **260**, 151-9 (2006).

- 596 19. Tsai, M.Y. *et al.* Cholesteryl ester transfer protein genetic polymorphisms, HDL
597 cholesterol, and subclinical cardiovascular disease in the Multi-Ethnic Study of
598 Atherosclerosis. *Atherosclerosis* **200**, 359-67 (2008).
- 599 20. Tang, C.S. *et al.* Exome-wide association analysis reveals novel coding sequence variants
600 associated with lipid traits in Chinese. *Nat Commun* **6**, 10206 (2015).
- 601 21. Lowe, J.K. *et al.* Genome-wide association studies in an isolated founder population from
602 the Pacific Island of Kosrae. *PLoS Genet* **5**, e1000365 (2009).
- 603 22. Studies, N.-N.W.G.o.R.i.A. *et al.* Replicating genotype-phenotype associations. *Nature*
604 **447**, 655-60 (2007).
- 605 23. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in
606 141,456 humans. *Nature* **581**, 434-443 (2020).
- 607 24. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human
608 genetic variants. *Nat Genet* **46**, 310-5 (2014).
- 609 25. Davydov, E.V. *et al.* Identifying a high fraction of the human genome to be under selective
610 constraint using GERP++. *PLoS Comput Biol* **6**, e1001025 (2010).
- 611 26. Qiu, X. *et al.* Crystal structure of cholesteryl ester transfer protein reveals a long tunnel
612 and four bound lipid molecules. *Nat Struct Mol Biol* **14**, 106-13 (2007).
- 613 27. UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**,
614 D480-D489 (2021).
- 615 28. Dron, J.S. *et al.* Polygenic determinants in extremes of high-density lipoprotein cholesterol.
616 *Journal of lipid research* **58**, 2162-2170 (2017).
- 617 29. Goodrich, J.K. *et al.* Determinants of penetrance and variable expressivity in monogenic
618 metabolic conditions across 77,184 exomes. *Nat Commun* **12**, 3505 (2021).
- 619 30. Blauw, L.L. *et al.* Mendelian randomization reveals unexpected effects of CETP on the
620 lipoprotein profile. *Eur J Hum Genet* **27**, 422-431 (2019).
- 621 31. Lee, C.J. *et al.* CETP, LIPC, and SCARB1 variants in individuals with extremely high
622 high-density lipoprotein-cholesterol levels. *Scientific reports* **9**, 1-7 (2019).
- 623 32. Oldoni, F., Sinke, R.J. & Kuivenhoven, J.A. Mendelian disorders of high-density
624 lipoprotein metabolism. *Circ Res* **114**, 124-42 (2014).
- 625 33. Rosenson, R.S. *et al.* HDL and atherosclerotic cardiovascular disease: genetic insights into
626 complex biology. *Nature Reviews Cardiology* **15**, 9-19 (2018).
- 627 34. Jamal, M., Zeinali, M. & Ghaffari, M.A. A molecular dynamics investigation on the
628 inhibition mechanism of cholesteryl ester transfer protein by Anacetrapib. *Medicinal*
629 *Chemistry Research* **25**, 62-69 (2016).
- 630 35. Group, H.T.R.C. *et al.* Effects of Anacetrapib in Patients with Atherosclerotic Vascular
631 Disease. *N Engl J Med* **377**, 1217-1227 (2017).
- 632 36. Nomura, A. *et al.* Protein-truncating variants at the cholesteryl ester transfer protein gene
633 and risk for coronary heart disease. *Circulation research* **121**, 81-88 (2017).
- 634 37. Cilpa-Karhu, G., Jauhiainen, M. & Riekkola, M.L. Atomistic MD simulation reveals the
635 mechanism by which CETP penetrates into HDL enabling lipid transfer from HDL to
636 CETP. *J Lipid Res* **56**, 98-108 (2015).
- 637 38. Zhang, M. *et al.* HDL surface lipids mediate CETP binding as revealed by electron
638 microscopy and molecular dynamics simulation. *Sci Rep* **5**, 8741 (2015).
- 639 39. Morgan, A.A. & Rubenstein, E. Proline: the distribution, frequency, positioning, and
640 common functional roles of proline and polyproline sequences in the human proteome.
641 *PLoS One* **8**, e53785 (2013).

- 642 40. Merriman, T.R. Application of Genetic Epidemiology to CETP (Cholesteryl Ester Transfer
643 Protein) Concentration and Risk of Cardiovascular Disease. *Circ Genom Precis Med* **11**,
644 e002138 (2018).
- 645 41. Krishnan, M. *et al.* Discordant association of the CREBRF rs373863828 A allele with
646 increased BMI and protection from type 2 diabetes in Maori and Pacific (Polynesian)
647 people living in Aotearoa/New Zealand. *Diabetologia* **61**, 1603-1613 (2018).
- 648 42. Ji, A. *et al.* Aotearoa New Zealand Māori and Pacific population-specific gout risk variants:
649 CLNK is a separate risk gene at the SLC2A9 locus. *The Journal of Rheumatology* (2021).
- 650 43. Tanner, C. *et al.* Population-Specific Resequencing Associates the ATP-Binding Cassette
651 Subfamily C Member 4 Gene With Gout in New Zealand Māori and Pacific Men. *Arthritis
652 & Rheumatology* **69**, 1461-1469 (2017).
- 653 44. Klück, V. *et al.* Rare genetic variants in interleukin-37 link this anti-inflammatory cytokine
654 to the pathogenesis and treatment of gout. *Annals of the rheumatic diseases* **79**, 536-544
655 (2020).
- 656 45. Patel, S.G. *et al.* Potential PINK1 Founder Effect in Polynesia Causing Early-Onset
657 Parkinson's Disease. *Movement Disorders*.
- 658 46. Stamp, L.K., Merriman, T.R. & Singh, J.A. Expert opinion on emerging urate-lowering
659 therapies. *Expert Opin Emerg Drugs* **23**, 201-209 (2018).
- 660 47. Wood, R., Fermer, S., Ramachandran, S., Baumgartner, S. & Morlock, R. Patients with
661 Gout Treated with Conventional Urate-lowering Therapy: Association with Disease
662 Control, Health-related Quality of Life, and Work Productivity. *J Rheumatol* **43**, 1897-
663 1903 (2016).
- 664 48. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
665 transform. *Bioinformatics* **26**, 589-95 (2010).
- 666 49. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
667 next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
- 668 50. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
- 669 51. Navarro Gonzalez, J. *et al.* The UCSC Genome Browser database: 2021 update. *Nucleic
670 Acids Res* **49**, D1046-D1057 (2021).
- 671 52. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-42 (2000).
- 672 53. Minster, R.L. *et al.* A thrifty variant in CREBRF strongly influences body mass index in
673 Samoans. *Nature Genetics* (2016).
- 674 54. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed
675 Program. *Nature* **590**, 290-299 (2021).
- 676 55. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium
677 panel. *Nature genetics* **48**, 1443-1448 (2016).
- 678 56. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**,
679 1284-1287 (2016).
- 680 57. Gogarten, S.M. *et al.* GWASTools: an R/Bioconductor package for quality control and
681 analysis of genome-wide association studies. *Bioinformatics* **28**, 3329-31 (2012).
- 682 58. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide
683 association studies. *Nat Genet* **38**, 904-9 (2006).
- 684 59. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-
685 wide association studies. *Nat Methods* **11**, 407-9 (2014).
- 686 60. Lange, K. *et al.* Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics*
687 **29**, 1568-70 (2013).

- 688 61. Zhou, H. *et al.* OPENMENDEL: a cooperative programming project for statistical genetics.
689 *Hum Genet* **139**, 61-71 (2020).
- 690 62. Conomos, M.P., Reiner, A.P., Weir, B.S. & Thornton, T.A. Model-free Estimation of
691 Recent Genetic Relatedness. *Am J Hum Genet* **98**, 127-48 (2016).
- 692 63. Gogarten, S.M. *et al.* Genetic association testing using the GENESIS R/Bioconductor
693 package. *Bioinformatics* **35**, 5346-5348 (2019).
- 694 64. Chen, H. *et al.* Control for Population Structure and Relatedness for Binary Traits in
695 Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet* **98**, 653-66
696 (2016).
- 697 65. Zhang, D. rsq: R-Squared and Related Measures. R package version 2.2. [https://CRAN.R-](https://CRAN.R-project.org/package=rsq)
698 [project.org/package=rsq](https://CRAN.R-project.org/package=rsq). (2021).
- 699 66. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan
700 results. *Bioinformatics* **26**, 2336-7 (2010).
- 701

702 FIGURE LEGENDS

703 **Figure 1. The population-specific rs159700001 variant changes the amino acid sequence of**
704 **CETP.** A. The CADD, GERP and Multiz UCSC track alignment at rs159700001 for which the
705 substitution creates a non-synonymous amino acid residue change (p.Pro177Leu). 3D schematic
706 of p.Pro177Leu (P160L) in the mature CETP protein structure in the N-terminal barrel domain
707 (left) of CETP. The CETP protein structure ([10.2210/pdb2OBD/pdb](https://doi.org/10.2210/pdb2OBD/pdb))²⁶ was obtained from The
708 Protein Databank (<https://www.rcsb.org/>)⁵² and illustrated using PyMOL. CETP, cholesteryl
709 ester transfer protein. CADD, Combined Annotation Dependent Depletion scores. GERP,
710 Genomic Evolutionary Rate Profiling.

711

712 **Figure 2. Association analyses of rs159700001 with HDL-C and LDL-C levels in Aotearoa**
713 **NZ discovery cohort and the Samoan I, Samoan II and Samoan II replication cohorts.**
714 Forest plot of a fixed-effect meta-analysis for the association of rs159700001 with HDL-C
715 (mmol/L) (A) and LDL-C (mmol/L) (B). Associations were adjusted by age, sex, 10 PCs and
716 relatedness in the Aotearoa NZ cohort and age, sex, first four PCs and relatedness in the Samoan
717 I-III cohorts. HDL-C, high-density lipoprotein cholesterol. LDL-C, low-density lipoprotein

718 cholesterol. NZ Māori, New Zealand Māori. CI Māori, Cook Island Māori. PCs, principal
719 components.

720

721 **Figure 3. Association analyses of rs1597000001 with HDL-C and LDL-C levels in**
722 **Polynesian participants of the PTO cohort.** Forest plot of rs1597000001 association with
723 HDL-C (mmol/L) and LDL-C (mmol/L) in the PTO cohort (Polynesian ethnicity only).
724 Associations were adjusted by age, sex and grandparental ethnicity. HDL-C, high-density
725 lipoprotein cholesterol. LDL-C, low-density lipoprotein cholesterol.

726

727 **Figure 4. CETP activity by rs1597000001 genotype.** Plasma activity of CETP (pmol/hr/μL) in
728 11 participants of the PTO cohort who had the heterozygous rs1597000001 genotype (CT) versus
729 the homozygous rs1597000001 genotype for the major allele (CC). CETP, cholesteryl ester
730 transfer protein. Error bars represent standard error of the mean. A two-sample t-test with Welch's
731 correction was conducted to test for a statistically significant ($P = 0.028$) difference in mean CETP
732 activity between the two genotypic groups.

733

734 **Figure 5. Association of HDL-C levels at the *CETP* locus.** Association of HDL-C at the *CETP*
735 locus (+/- 500kb the lead variant rs1597000001) (A) and after conditioning on rs1597000001 (B)
736 and rs1800775 (C) using variants on Illumina Infinium CoreExome v24 bead chip platform for
737 genotyped participants from the Aotearoa NZ cohort. The strength of LD, as measured by the r^2 ,
738 between each variant and rs1597000001 (A and C) and rs183130 (B) is represented by the color
739 of each point according to the legend in the top right hand corner. The plot was generated using a

740 custom locus zoom R package (<https://github.com/Geeketics/LocusZooms>). CETP, cholesteryl

741 ester transfer protein. HDL-C, high-density lipoprotein cholesterol.

742

743

744 **Table 1. Baseline characteristics of the cohorts.**

| | Aotearoa NZ Discovery Cohort | | | | | | | | Samoa Replication Cohorts | | | Pacific Trust Cohort |
|-------------------------------|------------------------------|----------------|----------------|----------------|---------------|---------------|----------------|--------------------|---------------------------|---------------|--------------|----------------------|
| | NZ Māori | CI Māori | Samoa | Tongan | Niuēan | Pukapukan | Other/Mixed | Ngāti Porou Hauora | Samoa I | Samoa II | Samoa III | All |
| Participants (n) | 814 | 172 | 322 | 155 | 37 | 72 | 232 | 270 | 2851 | 908 | 550 | 255 |
| Sex (male) n (%) | 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) | 126 (49.41) |
| 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 ± 15.36 | 50.97 ± 15.22 | 44.78 ± 14.18 | 42.17 ± 14.53 | 48.78 ± 13.62 | 44.93 ± 17.2 | 41.34 ± 15.69 | 56.15 ± 13.61 | 45.09 ± 11.18 | 43.26 ± 16.23 | 43.46 ± 8.74 | 18.49 ± 2.54 |
| Height (cm) | 170.31 ± 9.9 | 169.44 ± 10.43 | 173.57 ± 9.05 | 174.48 ± 8.65 | 172.13 ± 8.18 | 165.44 ± 9.57 | 171.74 ± 8.21 | 169.38 ± 8.68 | 165.2 ± 7.8 | 166.6 ± 8.1 | 165.4 ± 7.9 | 172.11 ± 8.95 |
| Weight (kg) | 95.58 ± 23.27 | 100.26 ± 24.53 | 106.04 ± 23.74 | 107.43 ± 20.92 | 98.49 ± 18.59 | 95.6 ± 20.4 | 104.07 ± 28.05 | 100.17 ± 24.81 | 91.3 ± 18.9 | 93.1 ± 22.7 | 87.6 ± 18.3 | 87.98 ± 20.28 |
| BMI (kg/m²) | 33.02 ± 7.84 | 34.99 ± 8.31 | 35.24 ± 7.43 | 35.2 ± 6.31 | 33.09 ± 4.84 | 34.86 ± 6.63 | 35.17 ± 8.23 | 34.79 ± 8.01 | 33.5 ± 6.6 | 33.5 ± 7.8 | 32.0 ± 6.1 | 29.66 ± 6.37 |
| LDL (mmol/L) | 2.74 ± 0.96 | 2.89 ± 0.96 | 2.91 ± 1.03 | 2.85 ± 1.05 | 2.89 ± 0.94 | 2.75 ± 1.1 | 2.78 ± 0.95 | 2.95 ± 1.06 | 3.4 ± 0.86 | 3.2 ± 0.90 | 3.6 ± 0.85 | 2.36 ± 0.64 |
| HDL (mmol/L) | 1.18 ± 0.4 | 1.18 ± 0.38 | 1.12 ± 0.34 | 1.11 ± 0.37 | 1.1 ± 0.32 | 1.21 ± 0.23 | 1.15 ± 0.37 | 1.15 ± 0.3 | 1.2 ± 0.29 | 1.1 ± 0.26 | 1.0 ± 0.28 | 1.33 ± 0.32 |
| Gout n (%) | 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 ± Standard Deviation or n (%), unless otherwise stated.

748 ^aData were unavailable for lipid-lowering medications.

749 New Zealand (NZ), Cook Islands (CI)

750 **Table 2. Minor Allele Frequency and Hardy-Weinberg Equilibrium of rs159700001**

| | Aotearoa NZ Discovery Cohort | | | | | | | | Samoa Replication Cohorts | | | Pacific Trust Otago Cohort | |
|----------------------|---------------------------------|---------------|---------------|---------------|--------------|---------------|---------------|--------------------------|------------------------------|---------------|---------------|-------------------------------|-----------------------------------|
| | NZ Māori | CI Māori | Samoa | Tongan | Niuēan | Pukapukan | Other/Mixed | Ngāti Porou Hauora | Samoa I | Samoa II | Samoa III | Polynesian | Non-Polynesian Pacific peoples |
| Total (n) | 814 | 172 | 322 | 155 | 37 | 72 | 232 | 270 | 2851 | 908 | 550 | 220 | 35 |
| CC (n) (%) | 760 (93.4) | 159 (92.4) | 297 (92.2) | 143 (92.3) | 33 (89.2) | 72 (100.0) | 212 (91.4) | 257 (95.2) | 2600 (91.2) | 825 (90.9) | 497 (90.4) | 201 (91.4) | 35 (100.0) |
| CT (n) (%) | 52 (6.4) | 13 (7.6) | 24 (7.5) | 11 (7.1) | 4 (10.8) | 0 (0.0) | 19 (8.2) | 13 (4.8) | 242 (8.5) | 80 (8.8) | 53 (9.6) | 18 (8.2) | 0 (0.0) |
| TT (n) (%) | 2 (0.3) | 0 (0.0) | 1 (0.3) | 1 (0.7) | 0 (0.0) | 0 (0.0) | 1 (0.4) | 0 (0.0) | 9 (0.3) | 3 (0.3) | 0 (0.0) | 1 (0.5) | 0 (0.0) |
| 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 <i>P</i> | 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 |

751

752 Minor allele frequency (MAF), Hardy Weinberg Equilibrium (HWE), New Zealand (NZ), Cook Islands (CI).

753

754

755

756

757

758

759

760

761

762

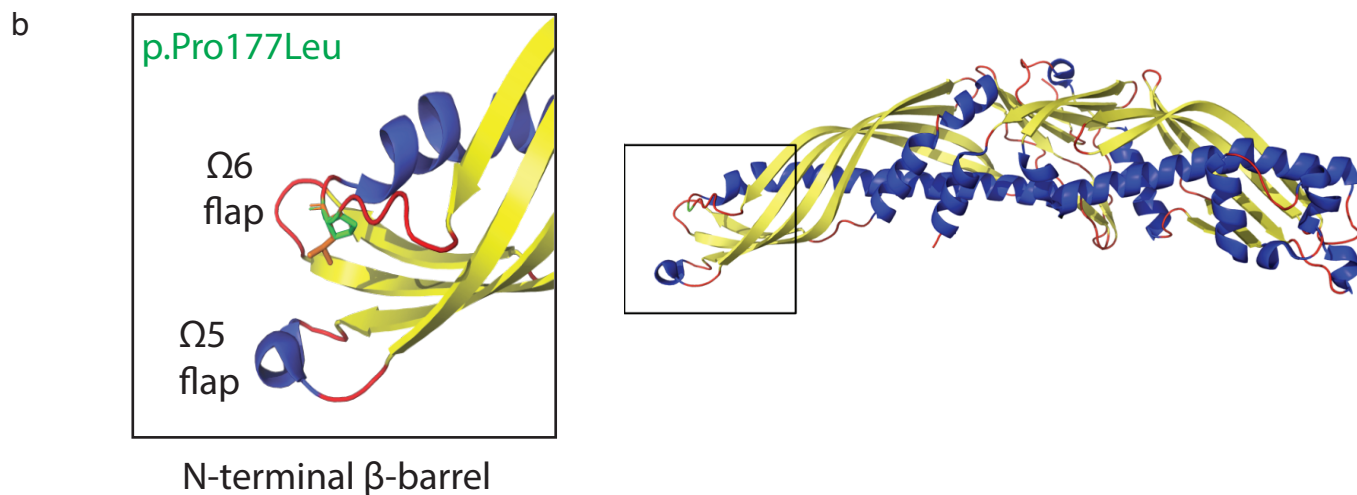
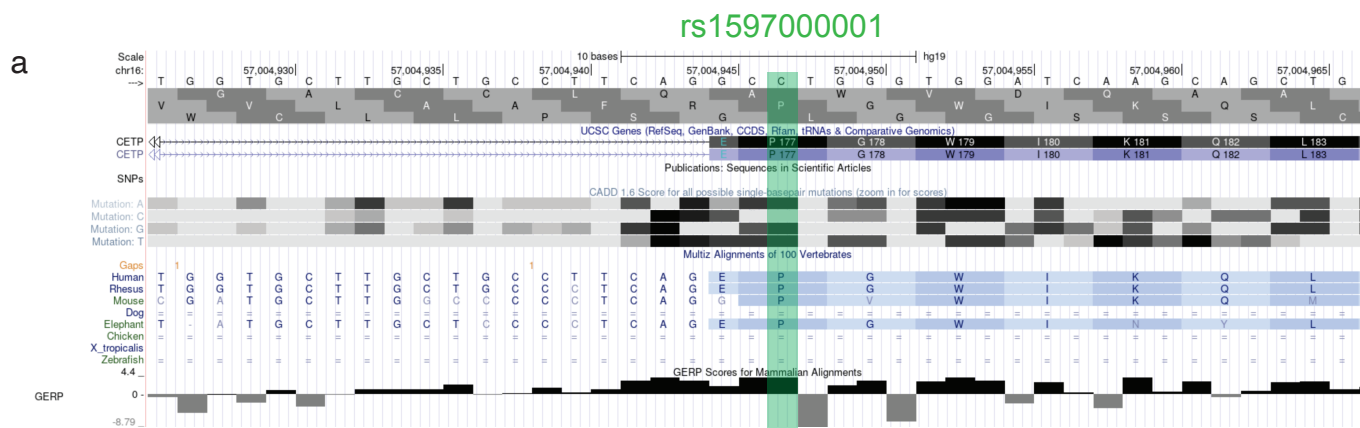
763

764

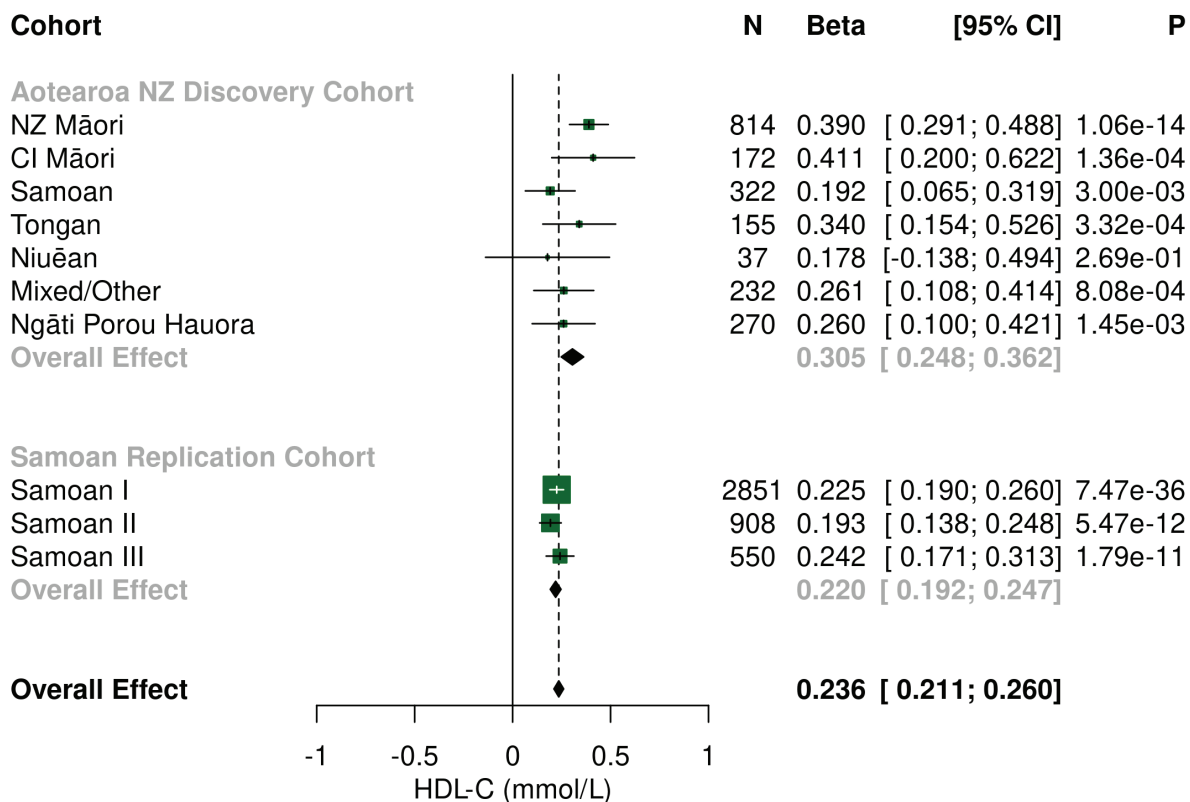
765

766

Figure 1



a



b

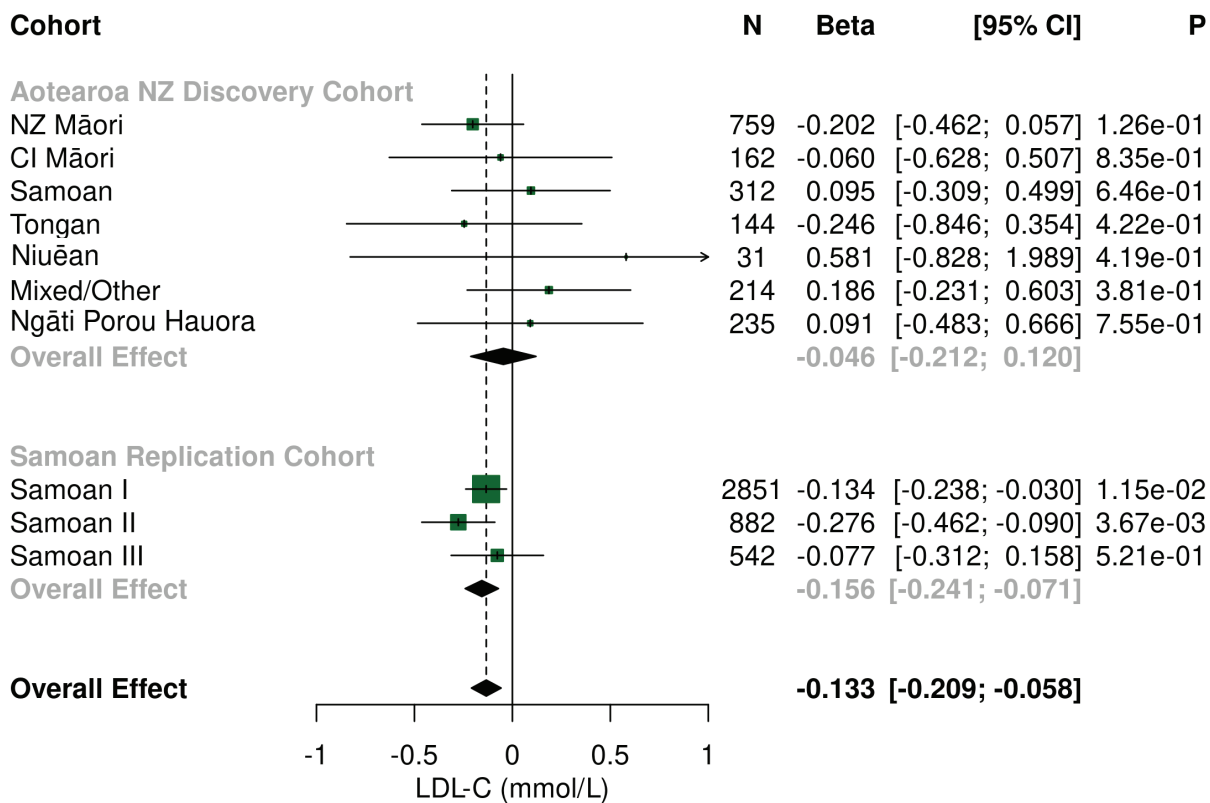
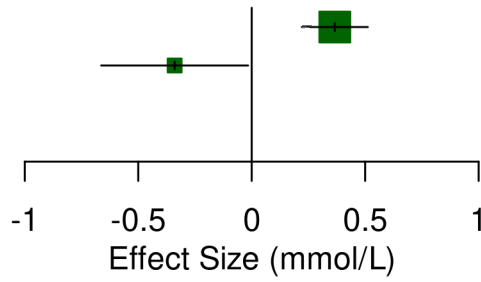


Figure 3

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. All rights reserved. No reuse allowed without permission.

PTO Cohort

HDL
LDL



| N | Beta | [95% CI] | P |
|----------|-------------|------------------|----------|
| 210 | 0.366 | [0.221; 0.511] | 7.77e-07 |
| 208 | -0.340 | [-0.663; -0.016] | 3.97e-02 |

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.

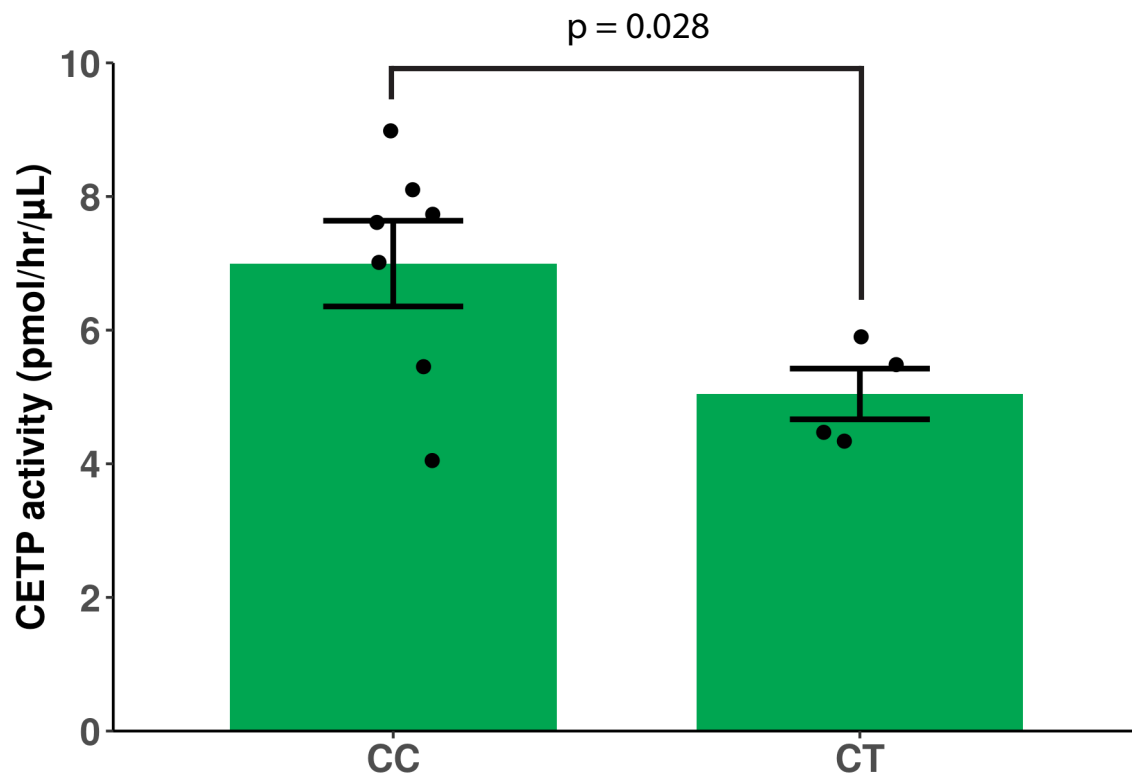
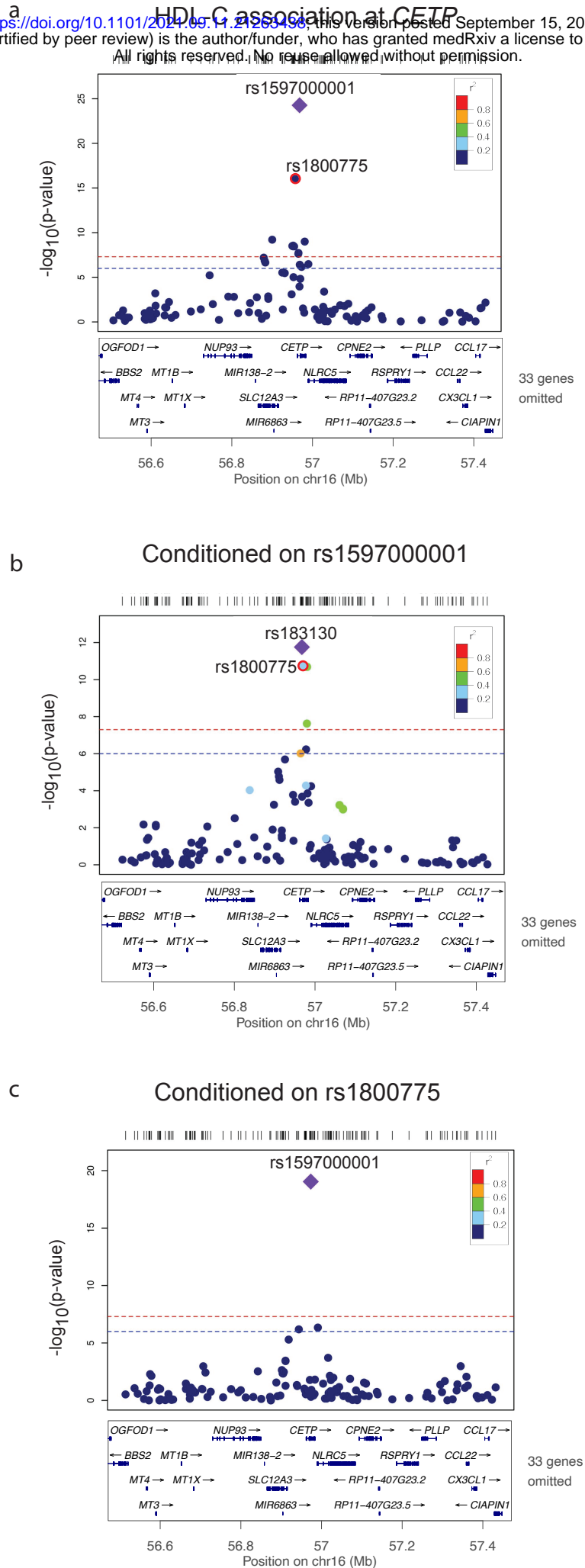
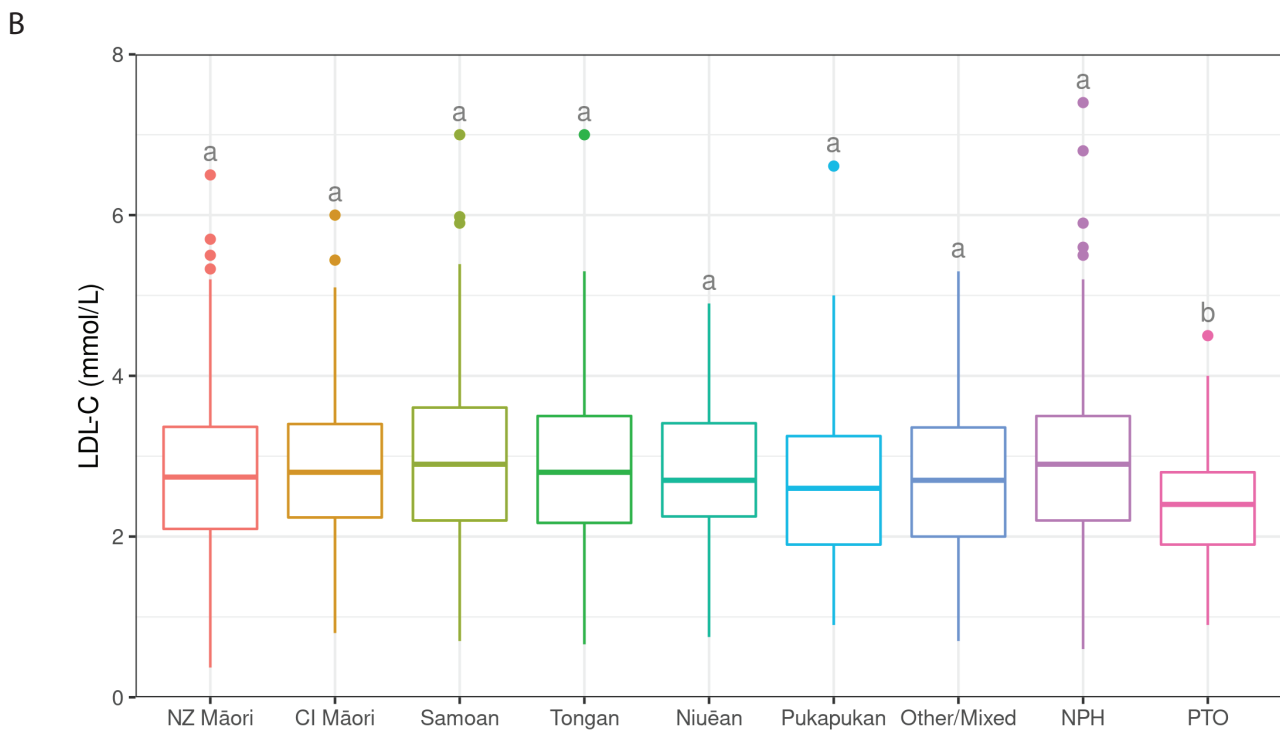
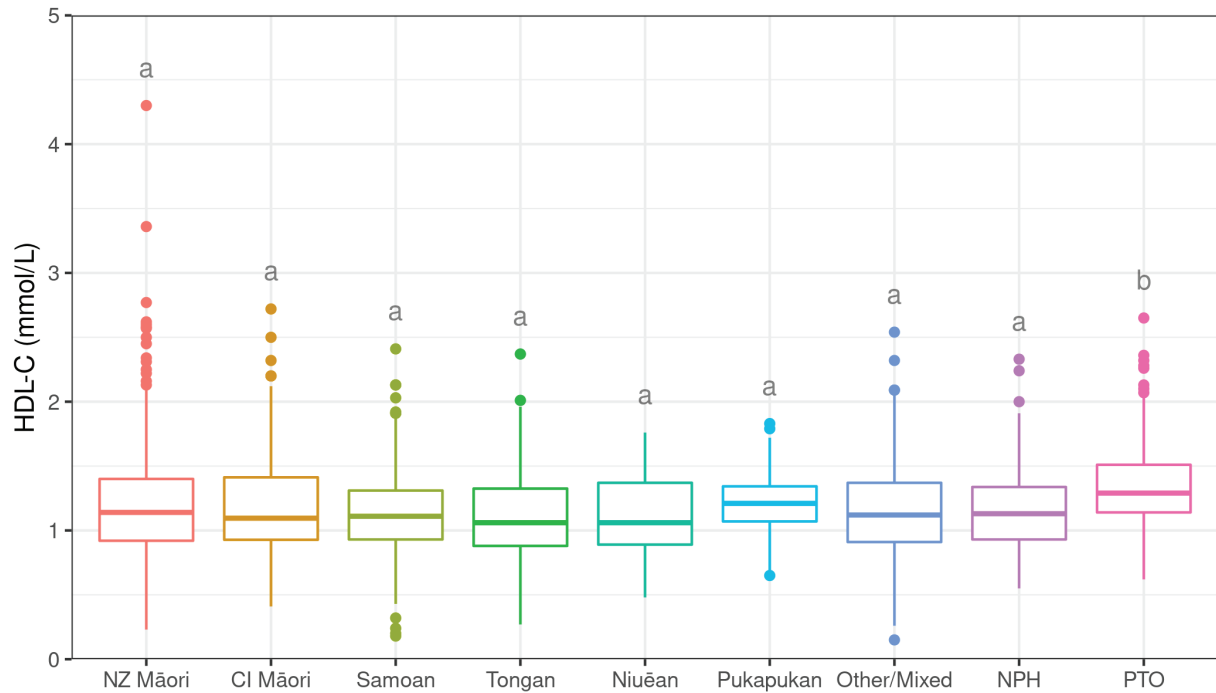


Figure 5

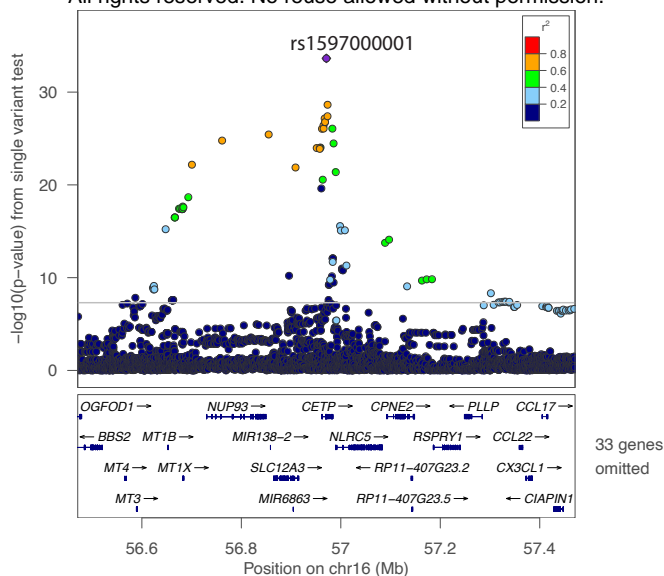
medRxiv preprint doi: <https://doi.org/10.1101/2021.09.14.21263438>; this version posted September 15, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



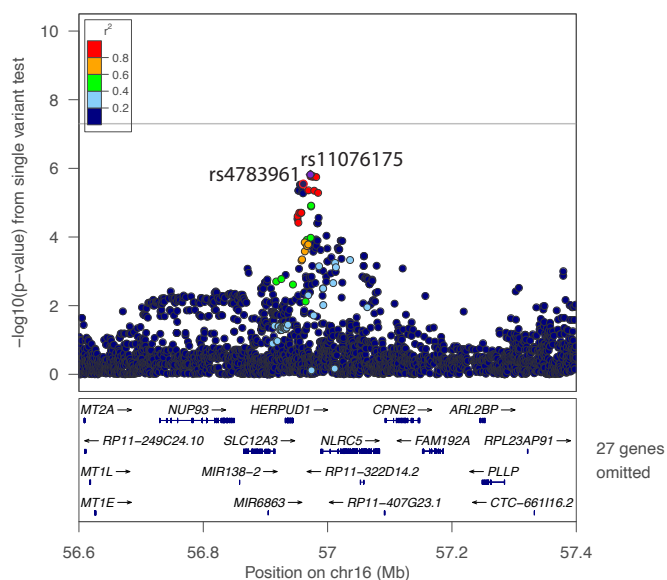


Supplementary Figure 1. HDL- and LDL-C (mmol/L) distributions in the Aotearoa New Zealand cohort separated into Pacific nation of grand-parental ethnicity and Ngāti Porou Hauora subgroups and the Pacific Trust Otago. HDL-C, High-density lipoprotein cholesterol. LDL-C, Low-density lipoprotein cholesterol. 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.

HDL-C association at CETP



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 (\pm 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 r^2 , 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.