The transferability of lipid loci across African, Asian and European cohorts

Nikita Telkar[1,25], Theresa Reiker[2,3,4,5,25], Robin G. Walters[6,7], Kuang Lin[7], Anders Eriksson[8], Deepti Gurdasani[2], Arthur Gilly[2,9], Lorraine Southam[2,9,10], Emmanouil Tsafantakis[11], Maria Karaleftheri[12], Janet Seeley[13,14,15], Anatoli Kamali[15], Gershim Asiki[15,16,17], Iona Y. Millwood[6,7], Michael Holmes[6,7], HuaidongDu[6,7], Yu Guo[18], Understanding Society Scientific Group[19], Meena Kumari[19], George Dedoussis[20], Liming Li[21], Zhengming Chen[7], Manjinder S. Sandhu[22], Eleftheria Zeggini[2,9,], Karoline Kuchenbaecker[2,23,24]*

[1] Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

[2] Department of Human Genetics, Wellcome Sanger Institute, Hinxton, CB10 1SA, UK

[3] Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK

[4] Swiss Tropical and Public Health Institute, Basel, Switzerland

[5] University of Basel, Basel, Switzerland

[6] Medical Research Council Population Health Research Unit (MRC PHRU), Nuffield Department of Population Health, University of Oxford, Oxford, UK

[7] Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK

[8] Department of Medical and Molecular Genetics, King's College London, London SE1 9RT, UK

[9] Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[10] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

[11] Anogia Medical Centre, Anogia, 740 51, Greece

[12] Echinos Medical Centre, Echinos, Xanthi 67300, Greece

[13] Department of Global Health and Development, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

[14] Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

[15] Medical Research[1] Council/Uganda Virus Research Institute and London School of Hygiene (MRC/UVRI and LSHTM), Uganda Research Unit, Entebbe, Uganda

[16] African Population and Health Research Center, Nairobi, Kenya

[17] Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

[18] Chinese Academy of Medical Sciences, Beijing 100730, China

[19] Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, UK

[20] Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Greece

[21] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China

[22] Department of Medicine, University of Cambridge, Cambridge CB2 0QQ, UK

[23] Division of Psychiatry, University College of London, London W1T 7NF, UK

[24] UCL Genetics Institute, University College London, London WC1E 6BT, UK

[25] These authors contributed equally to this work

* Correspondence to: k.kuchenbaecker@ucl.ac.uk

Abstract

The majority of genetic studies for cardiometabolic traits were based on samples with European ancestry. Our aim was to assess whether genetic variants associated with blood lipids, a major risk factor for CVD, are shared across different populations.

We compared genetic associations with lipids between samples from Uganda (N=6,407), China (N=21,295), Japan (N=162,255), the UK (N=9,961) and Greece (N=3,586). Using simulations, we established trans-ethnic colocalization as a method to distinguish shared from population-specific trait loci.

Genetic correlations for HDL, LDL and triglycerides between European ancestry and Asian cohorts were close to 1. A polygenic score based on established LDL-cholesterol-associated loci from European discovery samples had consistent effects on serum levels in samples from the UK, Uganda and Greek population isolates (r=0.23 to 0.28, $p<1.9x10^{-14}$). Overall, ~75% of the major lipid loci from European discovery studies displayed evidence of replication at $p<10^{-3}$, except triglyceride loci in the Ugandan samples of which only 10% replicated. Specific replicating loci were identified using trans-ethnic colocalization. Ten of the fourteen lipid loci that did not replicate in the Ugandan population had pleiotropic associations with BMI in European ancestry samples while none of the replicating loci did. While lipid associations were highly consistent across European and Asian populations, there was a lack of replication particularly for established triglyceride loci in the Ugandan population. These loci might affect lipids by modifying food intake or metabolism in an environment offering diets rich in certain nutrients. This suggests that gene-environment interactions could play an important role for the transferability of complex trait loci.

65 **Introduction**

66 Cardiovascular disease (CVD) is one of the leading causes of death worldwide[1]. As the predictive ability
67 of common variants for CVD and cardiometabolic traits improves, risk prediction in clinical settings
68 finds increasing consideration[2,3]. The foundations for this were provided by genome "white"
69 association studies: the majority of samples included in these studies were British or US-Americans
70 with European ancestry[4,5] which does not accurately representation the ethnically and ancestrally
71 diverse populations of these nations. Moreover, three quarters of CVD deaths occur in low- and
72 middle-income countries with incidences further rising[6]. Consequently, it is important to determine
73 whether cardiometabolic trait loci are transferable to other populations. We focussed on blood lipids,
74 a major cardiovascular risk factor.

75 Previous research assessed the effects of different allele frequencies and linkage disequilibrium (LD)
76 on genetic associations across ancestry groups[7]. Here we ask the fundamental question whether
77 causal variants for lipid biomarkers are shared across populations. Heterogeneity in effects of variants
78 could result from epistasis or gene-environment interactions. However, differences in LD structure
79 between populations make it difficult to compare associations between ancestry groups because the
80 observable effect of a variant depends on its correlation with the causal variant(s)[7]. Differences in
81 frequency also impact on the power to detect associations in other ancestry groups.

82 We employed several strategies which account for these effects to quantify the extent to which
83 genetic variants affecting lipid biomarkers are shared between individuals from Europe/North
84 America, Asia, and Africa. We assessed the transferability of individual signals and compared
85 association patterns across the genome using data from the African Partnership for Chronic Disease
86 Research – Uganda (APCDR-Uganda, N=6,407)[8], China Kadoorie Biobank (CKB, N=21,295)[9], the
87 Hellenic Isolated Cohorts (HELIC-MANOLIS, N=1,641 and HELIC-Pomak, N=1,945)[10,11], and the UK
88 Household Longitudinal Study (UKHLS, N=9,961)[12]. We also used summary statistics from Biobank
89 Japan (BBJ, N=162,255)[13] and the Global Lipid Genetics Consortium (European ancestry, GLGC2013
90 N=188,577, GLGC2017 N=237,050)[14,15].

91

92 **Results**

93 We assessed replication rates across established lipid-associated variants in different populations. We
94 distinguished major lipid loci, i.e. those with $p<10^{-100}$ in the largest European ancestry GWAS. In this
95 context, replication was operationalised as at least one variant from the credible set associated at
96 $p<10^{-3}$ in the target study. As a benchmark, we also assessed replication in two European ancestry
97 studies. We found evidence of replication for 76.5% of major HDL loci in these two studies (Table 1).
98 For the non-European groups replication rates ranged from 70.6 to 82.4%. Similar replication rates
99 were observed for LDL loci (61.5-76.9%). For major triglycerides (TG) loci, replication rates ranged
100 from 78.9 to 94.7%, except in APCDR-Uganda. Only 10.5% of these loci showed evidence of replication
101 in that sample. Replication rates for known loci with $p\geq10^{-100}$ in the discovery set were generally
102 moderate to low. However, Biobank Japan, the largest study, had markedly higher replication rates
103 for these loci than the other studies.

104

105 Trans-ethnic genetic correlations were estimated between the three largest studies, China Kadoorie
106 Biobank, Biobank Japan and GLGC2013. Correlations were high for each biomarker and were not
107 significantly different from 1 (Figure 1, Supplementary Table 1). We also compared associations across
108 biomarkers. This consistently showed negative genetic correlations between TG associations and HDL
109 associations, with estimates ranging from r_{gen}=-0.48 to r_{gen}=-0.86.

110

111 In order to assess patterns of sharing of risk alleles for the smaller studies, we constructed polygenic
112 scores based on the established lipid loci from discovery samples with European ancestry and
113 estimated the score associations with levels of HDL, LDL and TG in HELIC, APCDR-Uganda and also
114 UKHLS as a benchmark (Figure 2). All genetic scores were significantly associated with their respective
115 target lipid in the three European samples with largely consistent correlation coefficients and mutually
116 overlapping 95% confidence intervals (CIs) (Table 2). For HDL, LDL and TG, the estimated correlation

117 coefficients had a range of 0.27-0.28, 0.23-0.28 and 0.20-0.24, respectively. In APCDR-Uganda, the
118 strongest association was observed for LDL (r=0.28, SE=0.01, $p=1.9x10^{-107}$). The HDL association was
119 attenuated compared to the European samples (r=0.12, SE=0.01, $p=6.1x10-^{22}$). The effect of the TG
120 score was markedly weaker (r=0.06, SE=0.01, $p=4.5x10^{-7}$). We also assessed associations between a
121 given score and levels of each of the other biomarkers (Supplementary Table 2). In line with the trans-
122 ethnic genetic correlation results, we observed inverse associations between the HDL score and TG
123 levels and vice versa in all studies, except APCDR-Uganda.
124
125 Differences in LD structure, MAF and sample size make it difficult to assess replication for individual
126 loci. Therefore, we propose a new strategy to assess evidence for shared causal variants between two
127 populations: trans-ethnic colocalization. For this we re-purposed a method that was originally
128 developed for colocalization of GWAS and eQTL results: Joint Likelihood Mapping (JLIM)[16]. In order to
129 assess its performance for GWAS results from samples with different ancestry, we carried out a
130 simulation study. UK Biobank (UKB) was used as a European ancestry reference and compared to CKB
131 and APCDR-Uganda. Phenotypes were simulated. In the simulations of distinct causal variants in the
132 non-European and the reference group, the frequencies of false negatives were close to 0.05 as
133 expected (Table 3), with an almost uniform distribution of p-values (Supplementary Figure 1). The
134 power to detect shared associations was good for both populations: 73.1% for APCDR-Uganda and
135 93.1% for CKB. To investigate whether the lower power for APCDR-Uganda could be due to its smaller
136 sample size, we reran the analyses for CKB using a random subset of samples matching the sample
137 size of APCDR-Uganda. The results were similar to the ones for the full CKB set, suggesting that the
138 power of this trans-ethnic colocalization method decreases somewhat with greater genetic distance
139 between the populations that are compared.
140
141 We applied trans-ethnic colocalization for established lipid loci to each study with UKHLS as the
142 reference. There was evidence for significant (p_{jlim}<0.05) colocalization with at least one of the target

143 studies for about half of the major lipid loci (Supplementary Table 3). For several major TG loci, such
144 as *GCKR* at 2p23.3 or *LPL* at 8p21.3, strong evidence of replication in the Asian studies was observed
145 while there was no evidence of association in APCDR-Uganda (Figure 3b,c).
146 We compared major lipid loci showing evidence of replication in APCDR-Uganda with those not
147 displaying any suggestion of replication. The proximal genes of replicating loci were enriched for lipid
148 pathways including lipoprotein metabolism, lipid digestion mobilisation and transport, chylomicron-
149 mediated lipid transport and metabolism of lipids and lipoproteins. The proximal genes of the non-
150 replicating loci were enriched for several other pathways in addition to lipid metabolism, including
151 SHP2 signalling, ABV3 integrin pathway, cytokine signalling in immune system, cytokine-cytokine
152 receptor interaction and transmembrane transport of small molecules (Supplementary Figures 2 and
153 3). We also assessed the associations of these loci with BMI in samples with European ancestry using
154 publicly available summary statistics from the GIANT consortium[17] (N≥484,680) (Table 4). Ten of the
155 fourteen non-replicating lipid loci had pleiotropic associations with BMI at a Bonferroni-adjusted
156 threshold of $p<0.0024$. None of the seven replicating lipid loci were associated with BMI. We also
157 assessed four additional loci that were not significant in the trans-ethnic colocalization but displayed
158 small regional p-values in APCDR-Uganda. Out of these only APOE was significantly associated with
159 BMI ($p=4x10^{-21}$).
160
161
162 **Discussion**
163 Recent efforts to increase global diversity in genetics studies have been vital, enabling this
164 comprehensive cross-population comparison of genetic associations with blood lipids. We provide
165 evidence for extensive sharing of genetic variants that affect levels of HDL- and LDL-cholesterol and
166 triglycerides between individuals with European ancestry and samples from China, Japan and Greek
167 population isolates. There was evidence of replication for about three quarters of major HDL and LDL
168 loci (and triglyceride loci except in APCDR-Uganda). This was highly consistent across all studies.

169 Estimates of trans-ethnic genetic correlations between European, Chinese and Japanese samples were
170 close to 1. Associations of polygenic scores for LDL were not attenuated in the Ugandan population
171 compared to the UK samples. All PRS associations in the two Greek isolated populations were also
172 highly consistent with those in the UK samples.
173 Previous studies that compared the direction of effect of established loci or assessed associations of
174 polygenic scores reported differing degrees of consistency[18–29]. However, most of them were
175 conducted in American samples with diverse ancestry, had smaller sample sizes and applied a single-
176 variant look-up or PRS for a limited number of genetic variants. The high degree of consistency for
177 cholesterol biomarkers we observed also contrasts with previously reported trans-ethnic genetic
178 correlations for other traits, such as major depression, rheumatoid arthritis, or type 2 diabetes, which
179 were substantially different from 1[30,31]. In a recent application using data from individuals with
180 European and Asian ancestry from the UK and USA, the average genetic correlation across multiple
181 traits was 0.55 (SE = 0.14) for GERA and 0.54 (SE=0.18) for UK Biobank[32].
182 Differences in LD structure, MAF and sample size make it difficult to assess replication of individual
183 loci. We therefore propose a new approach: trans-ethnic colocalization. Simulations showed
184 consistent control of type I error rates as well as good power to detect associations. Colocalization
185 identified shared causal variants even at loci where none of the individual variants were associated at
186 stringent p-value thresholds. However, for many of the major lipid loci, more than one independent
187 association signal was identified in discovery GWAS[15]. When these are located in close proximity to
188 each other, they can interfere with the trans-ethnic colocalization analysis because JLIM assumes a
189 single causal variant (Figure 3d). Therefore, future work should extend this approach to accommodate
190 loci harbouring multiple causal variants. Using trans-ethnic colocalization, we showed that many
191 established loci for triglycerides did not affect levels of this biomarker in Ugandan samples. This
192 included loci associated at genome-wide significance in all the other studies, such as *GCKR* at 2p23.3
193 or *LPL* at 8p21.3. The polygenic score for triglycerides had a weak effect on measured levels in APCDR-
194 Uganda. This is unlikely to be an artefact of unreliable measurement: triglyceride levels had a heritable

195 component in this sample (SNP heritability of 0.25, SE=0.05[8]) and there were some genome-wide
196 significant associations (Supplementary Figure 6e). It is also unlikely that this can be explained purely
197 by differences in LD and MAF because they would affect the analyses of the other two biomarkers as
198 well. Instead these discrepancies could be caused by gene-environment interactions. Most of the lipid
199 loci that did not replicate in the Ugandans had pleiotropic associations with BMI in European ancestry
200 samples while none of the replicating loci were linked to BMI. It is possible that the non-replicating
201 variants affect the amount of food intake with downstream consequences for lipid levels. This might
202 require an environment offering diets that are rich in certain nutrients. While the replicating genes
203 were almost exclusively linked to pathways of lipid metabolism, the non-replicating genes were
204 involved in a diverse pathways which is in line with hypothesis. An alternative explanation could be
205 that the non-replicating loci are involved in metabolising nutrients given a particular diet that is not
206 common in Uganda with downstream consequences for weight.
207 Overall, this could suggest an important role of environmental factors in modifying which genetic
208 variants affect lipid levels. Studying the causes for discordant loci between groups has promise to
209 further elucidate the biological mechanisms of lipid regulation and other complex traits. Applying
210 genetic risk prediction within clinical settings is receiving increasing attention. Our findings
211 demonstrate that the transferability of genetic associations across different ancestry groups and
212 environmental settings should be assessed comprehensively for medically relevant traits. This is
213 important in order to maximise the potential of precision medicine to yield health benefits that are
214 widely shared within and across populations. Ongoing programs in underrepresented countries[33],
215 such as the Human Hereditary and Health in Africa Initiative[34], and programs focussing on
216 underrepresented groups, such as PAGE[35], All of Us[36], or East London Genes and Health[37], could
217 provide the basis for this.
218
219
220 **Methods**

221 *Data resources*

222 We included data from the Global Lipid Genetics Consortium (European ancestry samples only, GLGC),
223 The UK Household Longitudinal Study (UKHLS), two isolated populations from the Greece Hellenic
224 Isolated Cohorts (HELIC) , a rural West Ugandan population from the African Partnership for Chronic
225 Disease Research (APCDR-Uganda) study, China Kadoorie Biobank (CBK), and Biobank Japan (BBJ). In
226 addition, we used data from European ancestry samples from the eMERGE network to confirm
227 replication rates of known loci. Raw genotype and phenotype data were available for UKHLS, APCDR-
228 Uganda, CKB, HELIC-MANOLIS, HELIC-Pomak and eMERGE. Our analyses were based on summary
229 statistics for BBJ and GLGC. Study details are provided in Supplementary Table 3.

230 Each study underwent standard quality control. Details of the genome-wide association analyses with
231 lipid traits have been previously described for GLGC[14], BBJ[13], HELIC[10], and UKHLS[12]. The association
232 analysis for APCDR-Uganda was carried out within a mixed model framework using GEMMA[38]. Rank-
233 based inverse normal transformation was applied to the lipid biomarkers after adjusting for age and
234 gender. In China Kadoorie Biobank, lipid levels were regressed against eight principle components,
235 region, age, age^2, sex, and - for LDL and TG - fasting time^2. LDL levels were derived using the Friedewald
236 formula. After rank-based inverse normal transformation, the residuals were used as the outcomes in
237 the genetic association analyses using linear regression. In eMERGE, biomarkers were adjusted for
238 age, gender, kidney disease, statin use, type 2 diabetes status and disorders relating to growth
239 hormones. Associations were carried out within a mixed model framework using BOLT-LMM[39].
240 Manhattan plots for eMerge, UKHLS, BBJ, CKB, and APCDR-Uganda are shown in Supplementary
241 Figures 4-6.

242

243 *Established lipid loci*

244 A list of established lipid-associated loci was extracted from the latest Global Lipid Genetics
245 Consortium (GLGC2017) publication[15] reporting 444 independent variants in 250 loci

246 associated at genome-wide significance with HDL, LDL, and triglyceride levels. We excluded three LDL
247 variants where the association was not primarily driven by the samples with European ancestry. We
248 assessed evidence of replication of the loci, applied trans-ethnic colocalization and used them to
249 construct polygenic scores.
250
251 *Replication of established lipid loci*
252 We assessed evidence of replication across these established lipid variants. For loci harbouring
253 multiple signals, we only kept the most strongly associated variant. This left 170 HDL, 135 LDL and 136
254 TG variants. We distinguished major loci, i.e. those with $p<10^{-100}$ in GLGC2017. For each lead SNP we
255 identified all variants in LD ($r^2>0.6$) based on the European ancestry 1000 Genomes data. We assessed
256 whether the lead or any of the correlated variants, henceforth called credible set, displayed evidence
257 of association in the target study. We used a p-value threshold of $p<10^{-3}$. If this was not the case, we
258 tested whether there was any other variant with evidence of association within a 50Kb window. While
259 this p-value threshold might not be appropriate to provide conclusive evidence of replication for
260 individual loci, we used this to test evidence of replication across sets of loci. As a benchmark, we
261 computed the minimum p-value in 1000 random windows of 50Kb for each study. Less than 5% of
262 random windows had a minimum $p<10^{-3}$ for the non-European ancestry studies.
263
264 *Trans-ethnic genetic correlations*
265 We used the popcorn software[30] to estimate trans-ethnic genetic correlations between studies while
266 accounting for differences in LD structure. This provides an indication of the correlation of causal-
267 variant effect sizes across the genome at SNPs common in both populations. Variant LD scores were
268 estimated for ancestry-matched 1000 Genomes data for each study combination. The estimation of
269 LD scores failed for chromosome 6 for some groups. We therefore left out chromosome 6 from all
270 comparisons. Variants with imputation accuracy $r^2<0.8$ or MAF<0.01 were excluded. Popcorn did not

271 converge for any of the studies with less than 20,000 samples. Therefore, results are presented for
272 comparisons between GLGC2013, CKB and BBJ. We estimated effect rather than impact correlations.
273
274 *Polygenic scores*
275 We created polygenic scores based on the established lipid loci and assessed their associations with
276 lipid levels in UKHLS, the HELIC cohorts, and APCDR-Uganda, as it was not possible to compute trans-
277 ethnic genetic correlations for these studies. For HELIC and UKHLS, extreme values ($\mu \pm 3\,SD$, sex
278 stratified) were filtered. Age, age^2 and sex were adjusted for by regressing them on the biomarker
279 values and using the residuals as outcomes for subsequent analyses. For each biomarker in each
280 sample set, we checked normality and homoscedasticity. HDL and LDL were approximately normally
281 distributed. For TG levels, a Box Cox transformation was used to normalize the data. APCDR-Uganda
282 phenotype data were rank-based inverse normally transformed.
283 To make sure PRS were comparable across studies, we excluded variants that were absent, rare
284 (MAF<0.01) or badly imputed (r^2<0.8) in any of the studies and variants that had different alleles from
285 those in the GLGC. The variant with larger discovery p-value from each correlated pair of SNPs (r^2>0.1)
286 was also removed. This left 120, 103 and 101 variants for HDL, LDL and TG, respectively. We created
287 trait-specific weighted PRS. The β-regression coefficients from SNP-trait associations in GLGC2017[15]
288 were used as weights. All biomarkers and scores were scaled to mean=0 and standard deviation=1 for
289 each study so that the regression coefficient represent estimates of the correlation between scores
290 and biomarkers.
291 We carried out association analyses between each polygenic score and each biomarkers using a linear
292 mixed model with random polygenic effect implemented in GEMMA[38] in order to account for
293 relatedness and population structure. We used a Bonferroni correction to adjust for multiple testing
294 of three PRS with three different biomarker outcomes (p<0.05/9=0.0056).
295
296 *Trans-ethnic colocalization*

318 Differences in allele frequency, LD structure and sample size make it difficult to assess whether a given
319 GWAS hit replicates in samples with different ancestries. Therefore, we applied trans-ethnic
320 colocalization. Colocalization methods test whether the associations in two studies can be explained
321 by the same underlying signal even if the specific causal variant is unknown. The joint likelihood
322 mapping (JLIM) statistic was developed by Chun and colleagues to estimate the posterior probabilities
323 for colocalization between GWAS and eQTL signals and compare them to probabilities of distinct
324 causal variants[16]:

325
$$\Lambda = \sum_{i \in N_\theta^1(m^*)} L_1(i) \times log \frac{L_1(i)L_2(i)}{\max_{j \notin N_\theta^2(i)} L_1(i)L_2(j)}$$

i SNP
m^* lead SNP
$L_1(i)$ likelihood of SNP i being causal for trait 1
$L_2(i)$ likelihood of SNP i being causal for trait 2
$N_\theta^1(i)$, $N_\theta^2(i)$ sets of SNPs in LD with i
θ LD threshold

326

327 JLIM explicitly accounts for LD structure. Therefore, we assessed whether it is suitable for trans-ethnic
328 colocalization. For samples with summary statistics, LD scores were estimated using ancestry matched
329 samples from the 1000 Genomes Project v3. JLIM assumes only one causal variant within a region in
330 each study. We therefore used a small windows of 50Kb for each known locus to minimise the risk of
331 interference from additional association signals. Distinct causal variants were defined by separation
332 in LD space by $r^2 \geq 0.8$ from each other. We excluded loci within the major histocompatibility region
333 due to its complex LD structure. We used a significance threshold of $p<0.05$ given the evidence of
334 association of the established lipid loci in Europeans and the overall evidence for shared causal genetic
335 architecture across populations for most lipid traits from our other analyses. We compared each
336 target study to UKHLS because of their high level of homogeneity in terms of ancestry, biomarker
337 quantification and study design.

338

339 *Simulation*

340 To test the power of trans-ethnic colocalization to detect associations shared between pairs of
341 populations with different ancestry, we ran JLIM on two sets of simulated traits with realistic effect
342 size and environmental noise level. The first set of simulations used the same causal variant in both
343 populations, whereas the second set of simulations discordant causal variants, chosen randomly. We
344 sampled 10,000 randomly chosen biallelic variants with MAF>0.05 and simulated random phenotypes
345 in CKB, APCDR-Uganda und 50,000 individuals with British ancestry from UK Biobank. For each data
346 set relatives were excluded. We also sub-sampled CKB to match the smaller number of individuals in
347 APCDR-Uganda in order to test whether different performance is due to ancestry or sample size. We
348 used a simple linear model to generate the phenotype for each individual i:

349 $$y_i = \beta * (x_i - 1) + \eta_i$$

350 where y is the phenotype value, β is the effect size, x is the number of the alternate alleles carried at
351 the locus and $\eta_i \sim N(0, \sigma^2)$, where σ^2 is the variance of the environmental noise and $Cov(\eta_i, \eta_j) = 0$.
352 We used $\beta = 0.25$ and $\sigma^2 = 1$.
353
354 *Comparison of replicating with non-replicating loci*
355 We aimed to assess whether there are systematic differences between loci that are shared between
356 European ancestry samples and APCDR-Uganda and loci that are not. We identified all loci with
357 evidence of replication based on the above definition that also had significant (p<0.05) colocalization.
358 We only kept one variant per region. We contrasted them with loci where none of the evidence
359 suggested replication: p>0.05 for colocalization, no variant with a lipid association at $p<10^{-3}$ in the
360 region and the lead variant from the discovery study was not rare in APCDR-Uganda. We identified
361 the nearest protein coding gene for each locus and carried out a pathway analyses for the two sets
362 using FUMA[40]. We also assessed the associations of the lead variants with body mass index (BMI) in
363 European ancestry samples using results from a meta-analysis between the GIANT consortium and UK
364 Biobank[17]. We used a Bonferroni adjusted p-value threshold.
365

366 *Data availability*

367 The UKHLS EGA accession number is EGAD00010000918. Genotype-phenotype data access for UKHLS
368 is available by application to Metadac (www.metadac.ac.uk). eMERGE is available through dbgap
369 (study ID: phs000888.v1.p1). Summary statistics for GLGC
370 (http://csg.sph.umich.edu/abecasis/public/) and Biobank Japan (http://jenger.riken.jp/en/) are
371 publicly available. The HELIC genotype and WGS datasets have been deposited to the European
372 Genome-phenome Archive (https://www.ebi.ac.uk/ega/home): EGAD00010000518;
373 EGAD00010000522; EGAD00010000610; EGAD00001001636, EGAD00001001637. The APCDR
374 committees are responsible for curation, storage, and sharing of the APCDR-Uganda data under
375 managed access. The array and sequence data have been deposited at the European Genome-
376 phenome Archive (EGA, http://www.ebi.ac.uk/ega/, study accession number EGAS00001000545,
377 datasets EGAS00001001558 and EGAD00001001639 respectively) and can be requested through
378 datasharing@sanger.ac.uk. Requests for access to phenotype data may be directed to
379 data@apcdr.org.

380

381 **Supplemental data**

382 The supplemental data contain three figures and two tables.

383

384 **Funding and acknowledgements**

385 We would like to thank the participants and volunteers of the each of the studies. We appreciate the
386 advice received from Brielin Brown regarding the use of Popcorn and from Shamil Sunyaev and Sung
387 Chun regarding the suitability of JLIM for trans-ethnic colocalization.
388 This work was funded by the Wellcome Trust (WT098051), (212360/Z/18/Z), and the European
389 Research Council (ERC-2011-StG 280559-SEPI). The MANOLIS study is dedicated to the memory of
390 Manolis Giannakakis, 1978-2010. We thank the residents of the Pomak villages and of the
391 Mylopotamos villages for taking part. The HELIC study has been supported by many individuals who

have contributed to sample collection (including Antonis Athanasiadis, Olina Balafouti, Christina Batzaki, Georgios Daskalakis, Eleni Emmanouil, Pounar Feritoglou, Chrisoula Giannakaki, Margarita Giannakopoulou, Kiki Kaldaridou, Anastasia Kaparou, Vasiliki Kariakli, Stella Koinaki, Dimitra Kokori, Maria Konidari, Hara Koundouraki, Dimitris Koutoukidis, Vasiliki Mamakou, Eirini Mamalaki, Eirini Mpamiaki, Nilden Selim, Nesse Souloglou, Maria Tsoukana, Dimitra Tzakou, Katerina Vosdogianni, Niovi Xenaki, Eleni Zengini), data entry (Thanos Antonos, Dimitra Papagrigoriou, Betty Spiliopoulou), sample logistics (Sarah Edkins, Emma Gray), genotyping (Suzannah Bumpstead, Robert Andrews, Hannah Blackburn, Doug Simpkin, Siobhan Whitehead), research administration (Anja Kolb-Kokocinski, Carol Smee, Danielle Walker) and informatics (Kathleen Stirrups, Martin Pollard, Josh Randall).

APCDR-Uganda was funded by the Wellcome Trust, The Wellcome Trust Sanger Institute (WT098051), the UK Medical Research Council (G0901213-92157, G0801566, and MR/K013491/1), and the Medical Research Council/Uganda Virus Research Institute Uganda Research Unit on AIDS core funding. We thank the African Partnership for Chronic Disease Research (APCDR) for providing a network to support this study as well as a repository for deposition of curated data. We also thank all study participants who contributed to this study.

The UK Household Longitudinal Study was funded by grants from the Economic & Social Research Council (ES/H029745/1) and the Wellcome Trust (WT098051). UKHLS is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute.

The Understanding Society Scientific Group are Michaela Benzeval(1), Jonathan Burton(1), Nicholas Buck(1), Annette Jäckle(1), Meena Kumari(1), Heather Laurie(1), Peter Lynn(1), Stephen Pudney(1), Birgitta Rabe(1), Dieter Wolke(2)

1) Institute for Social and Economic Research

2) University of Warwick

Declaration of interests

The authors declare no competing interests.

Web resources

We have made code to run trans-ethnic colocalization using JLIM and simulations available through github: https://github.com/KarolineKuchenbaecker/TEColoc

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet Lond. Engl.* **385**, 117–171 (2015).
2. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
3. Kuchenbaecker, K. B. *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J. Natl. Cancer Inst.* **109**, (2017).
4. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
5. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nat. News* **538**, 161 (2016).
6. Bovet, P. & Paccaud, F. Cardiovascular Disease and the Changing Face of Global Public Health: A Focus on Low and Middle Income Countries. *Public Health Rev.* **33**, 397–415 (2011).
7. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
8. Heckerman, D. *et al.* Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7377–7382 (2016).

9. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* **40**, 1652–1666 (2011).
10. Gilly, A. *et al.* Cohort-wide deep whole genome sequencing and the allelic architecture of complex traits. *Nat. Commun.* **9**, 4674 (2018).
11. Gilly, A. *et al.* Very low depth whole genome sequencing in complex trait association studies. *Bioinforma. Oxf. Engl.* (2018). doi:10.1093/bioinformatics/bty1032
12. Prins, B. P. *et al.* Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Sci. Rep.* **7**, 11008 (2017).
13. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
14. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
15. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758 (2017).
16. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600–605 (2017).
17. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. *Hum. Mol. Genet.* (2018). doi:10.1093/hmg/ddy327
18. Musunuru, K. *et al.* Multi-Ethnic Analysis of Lipid-Associated Loci: The NHLBI CARe Project. *PLoS ONE* **7**, (2012).
19. Teslovich, T. M. *et al.* Biological, Clinical, and Population Relevance of 95 Loci for Blood Lipids. *Nature* **466**, 707–713 (2010).
20. Dumitrescu, L. *et al.* Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *PLoS Genet.* **7**, (2011).
21. Bryant, E. K. *et al.* A Multiethnic Replication Study of Plasma Lipoprotein Levels-Associated SNPs Identified in Recent GWAS. *PLoS ONE* **8**, (2013).

470 22. Wu, Y. *et al.* Trans-Ethnic Fine-Mapping of Lipid Loci Identifies Population-Specific Signals and
471 Allelic Heterogeneity That Increases the Trait Variance Explained. *PLoS Genet.* **9**, (2013).
472 23. Wang, Z. *et al.* Genetic associations with lipoprotein subfraction measures differ by ethnicity in
473 the multi-ethnic study of atherosclerosis (MESA). *Hum. Genet.* **136**, 715–726 (2017).
474 24. Keebler, M. E. *et al.* Association of Blood Lipids with Common DNA Sequence Variants at
475 Nineteen Genetic Loci in the Multiethnic United States National Health and Nutrition
476 Examination Survey III. *Circ. Cardiovasc. Genet.* **2**, 238–243 (2009).
477 25. Lanktree, M. B., Anand, S. S., Yusuf, S. & Hegele, R. A. Replication of genetic associations with
478 plasma lipoprotein traits in a multiethnic sample. *J. Lipid Res.* **50**, 1487–1496 (2009).
479 26. Chang, M. *et al.* Racial/ethnic variation in the association of lipid-related genetic variants with
480 blood lipids in the US adult population. *Circ. Cardiovasc. Genet.* **4**, 523–533 (2011).
481 27. Below, J. E. *et al.* Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-
482 specific effects, and tissue-specific enrichment of eQTLs. *Sci. Rep.* **6**, (2016).
483 28. Johnson, L., Zhu, J., Scott, E. R. & Wineinger, N. E. An Examination of the Relationship between
484 Lipid Levels and Associated Genetic Markers across Racial/Ethnic Populations in the Multi-Ethnic
485 Study of Atherosclerosis. *PloS One* **10**, e0126361 (2015).
486 29. Marigorta, U. M. & Navarro, A. High Trans-ethnic Replicability of GWAS Results Implies Common
487 Causal Variants. *PLoS Genet.* **9**, e1003566 (2013).
488 30. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from
489 Summary Statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
490 31. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the
491 genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
492 32. Galinsky, K. J. *et al.* Estimating cross-population genetic correlations of causal effect sizes. *Genet.*
493 *Epidemiol.* (2018). doi:10.1002/gepi.22173
494 33. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**, 175–
495 185 (2018).

496 34. Mulder, N. *et al.* H3Africa: current perspectives. *Pharmacogenomics Pers. Med.* **Volume 11**, 59–
497 66 (2018).
498 35. Carlson, C. S. *et al.* Generalization and Dilution of Association Results from European GWAS in
499 Populations of Non-European Ancestry: The PAGE Study. *PLOS Biol.* **11**, e1001661 (2013).
500 36. Program Overview - All of Us | National Institutes of Health. Available at:
501 https://allofus.nih.gov/about/about-all-us-research-program. (Accessed: 28th December 2018)
502 37. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community based
503 population genomics and health study in people of British-Bangladeshi and -Pakistani heritage.
504 *bioRxiv* (2018). doi:10.1101/426163
505 38. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies.
506 *Nat. Genet.* **44**, 821–824 (2012).
507 39. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for
508 biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
509 40. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and
510 annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
511

512 **Figures**

513 **Figure 1. Trans-ethnic genetic correlations** for associations with high-density lipoprotein (HDL), low-
514 density lipoprotein (LDL) cholesterol and triglycerides (TG). a) shows the comparison of GLGC2013
515 (European) and Biobank Japan, b) GLGC2013 and China Kadoorie Biobank and c) Biobank Japan and
516 China Kadoorie Biobank.

517 **Figure 2. Associations of polygenic scores** based on established lipid-associated loci with levels of
518 high-density lipoprotein (HDL), low-density lipoprotein (LDL) cholesterol and triglycerides (TG) in a)
519 UKHLS, b) HELIC-MANOLIS, c) HELIC-Pomak, d) APCDR-Uganda. Estimates are given as correlation
520 coefficients. Stars indicate statistically significant associations ($p<0.0056$).

521 **Figure 3. Regional association plots** for a selection of established lipid-associated loci for UKHLS,
522 Biobank Japan, APCDR-Uganda, and China Kadoorie Biobank and p-value p_{jlim} for the trans-ethnic
523 colocalization with UKHLS.

524 **Tables**

525 **Table 1. Percentage of established lipid-associated loci with evidence of replication** in each target
526 study. Results shown separately by strength of association (whether $p<10^{-100}$) in the discovery study
527 (GLGC). Only one SNP was kept for each locus with multiple associated variants in close proximity.
528 Regions were defined as 25Kb either side of the lead variant. The credible set contains the reported
529 lead variant and variants in LD (r2>0.6) with it.

P in GLGC:		**$<10^{-100}$**			**$\geq 10^{-100}$**		
Study	**Trait**	**n.s.[a]**	**Region[b]**	**Credible[c]**	**n.s. [a]**	**Region[b]**	**Credible[c]**
eMERGE	HDL	11.8	11.8	76.5	73.9	19.0	7.2
	LDL	15.4	7.7	76.9	89.3	9.0	1.6
	TG	10.5	10.5	78.9	81.2	15.4	3.4
UKHLS	HDL	5.9	17.6	76.5	81.0	13.7	5.2
	LDL	7.7	15.4	76.9	77.0	16.4	6.6
	TG	0.0	5.3	94.7	82.1	14.5	3.4
CKB	HDL	11.8	5.9	82.4	71.2	16.4	12.4
	LDL	7.7	30.8	61.5	83.6	7.4	9.0
	TG	5.3	15.8	78.9	82.9	10.3	6.8
BBJ	HDL	11.8	11.8	76.5	47.7	19.6	32.7
	LDL	7.7	30.8	61.5	64.8	10.7	24.6
	TG	5.3	10.5	84.2	55.6	12.8	31.6
UG	HDL	11.8	17.6	70.6	73.2	25.5	1.3
	LDL	23.1	7.7	69.2	73.8	24.6	1.6
	TG	42.1	47.4	10.5	79.5	17.1	3.4

530 [a] No variant in the region associated in target set at $p<10^{-3}$
531 [b] No variant in the credible set associated in the target set at $p<10^{-3}$ but an uncorrelated variant in
532 the region is associated in target set at $p<10^{-3}$
533 [c] A variant in the credible set is associated in the target set at $p<10^{-3}$

534 **Table 2: Associations of polygenic scores** based on established lipid-associated loci and respective
535 biomarkers levels in UKHLS, HELIC-MANOLIS, -Pomak, and APCDR-Uganda using a linear mixed
536 model analysis.

Trait	N	Correlation (SE[a])	P-value
UKHLS			
HDL	9706	0.284 (0.010)	8.34×10^{-165}
LDL	9767	0.273 (0.010)	8.38×10^{-155}
Triglycerides	9635	0.203 (0.010)	2.62×10^{-86}
HELIC-MANOLIS			
HDL	1186	0.276 (0.029)	8.65×10^{-20}
LDL	1186	0.230 (0.029)	1.89×10^{-14}
Triglycerides	1176	0.237 (0.030)	3.01×10^{-14}
HELIC-Pomak			
HDL	1078	0.272 (0.030)	9.67×10^{-18}
LDL	1075	0.285 (0.030)	1.35×10^{-18}
Triglycerides	1066	0.235 (0.030)	1.68×10^{-13}
APCDR-Uganda			
HDL	6407	0.121 (0.012)	6.06×10^{-22}
LDL	6407	0.280 (0.012)	1.91×10^{-107}
Triglycerides	6407	0.063 (0.013)	4.46×10^{-7}

537 [a] SE=standard error
538

539 **Table 3: 10,000 simulation runs to assess the performance of trans-ethnic colocalization.**
540 Phenotypes were simulated for CKB and APCDR-Uganda and trans-ethnic colocalization was run to
541 compare each to a reference set of 50,000 samples with British ancestry from UK Biobank.

Sample, N	Power	Type I error rate
CKB, 72,473	93.1%	5.2%
APCDR-Uganda, 4,597	73.1%	4.6%
CKB, 4,597	94.8%	4.5%

542

543 **Table 4. Association of established lipid-associated loci with body mass index** by locus replication
544 status in APCDR-Uganda. Association results are based on N≥484,680 samples from the meta-
545 analysis between GIANT and UK Biobank

Replication	Trait	rs-id	Chr	Position	Annotation	beta	SE	P-value
no	HDL	rs11755393	6	34824636	*UHRF1BP1*	-0.025	0.002	**9.8×10^{-48}**
no	HDL	rs1178979	7	72856430	*BAZ1B*	-0.010	0.002	**3.1×10^{-6}**
no	HDL	rs4731702	7	130433384	*KLF14*	0.008	0.002	**3×10^{-7}**
no	HDL	rs2954033	8	126493746	*NSMCE2*	-0.010	0.002	**6.4×10^{-8}**
no	LDL	rs4245791	2	44074431	*ABCG8*	0.002	0.002	0.22
no	LDL	rs3846662	5	74651084	*HMGCR*	0.020	0.002	**1.9×10^{-35}**
no	LDL	rs2737229	8	116648565	*TRPS1*	0.014	0.002	**1.9×10^{-15}**
no	LDL	rs635634	9	136155000	*IL6R*	0.005	0.002	0.03
no	LDL	rs2000999	16	72108093	*HPR*	0.011	0.002	**8.6×10^{-8}**
no	TG	rs1260326	2	27730940	*GCKR*	-0.011	0.002	**1.2×10^{-10}**
no	TG	rs2943641	2	227093745	*IRS1*	0.006	0.002	**5.8×10^{-4}**
no	TG	rs6905288	6	43758873	*VEGFA*	-0.010	0.002	**1.9×10^{-9}**
no	TG	rs11820589	11	116633862	*APOA5*	-0.003	0.003	0.41
no	TG	rs58542926	19	19379549	*TM6SF2*	-0.003	0.003	0.33
yes	HDL	rs643531	9	15296034	*TTC39B*	0.000	0.002	0.92
yes	HDL	rs1800588	15	58723675	*LIPC*	-0.002	0.002	0.25
yes	HDL	rs3764261	16	56993324	*CETP*	-0.002	0.002	0.39
yes	HDL	rs16942887	16	67928042	*PSKH1*	-0.005	0.003	0.06
yes	LDL	rs12740374	1	109817590	*CELSR2*	0.003	0.002	0.18
yes	LDL	rs1367117	2	21263900	*APOB*	-0.002	0.002	0.19
yes	LDL	rs6511720	19	11202306	*LDLR*	0.006	0.003	0.03
suggestive	HDL	rs4841132	8	9183596	*PPP1R3B*	0.000	0.003	0.88
suggestive	HDL	rs328	8	19819724	*LPL*	0.005	0.003	0.10
suggestive	HDL	rs769449	19	45410002	*APOE*	-0.026	0.003	**4.4×10^{-21}**
suggestive	HDL	rs386000	19	54792761	*LILRB2*	0.001	0.002	0.60

546

Figure 1

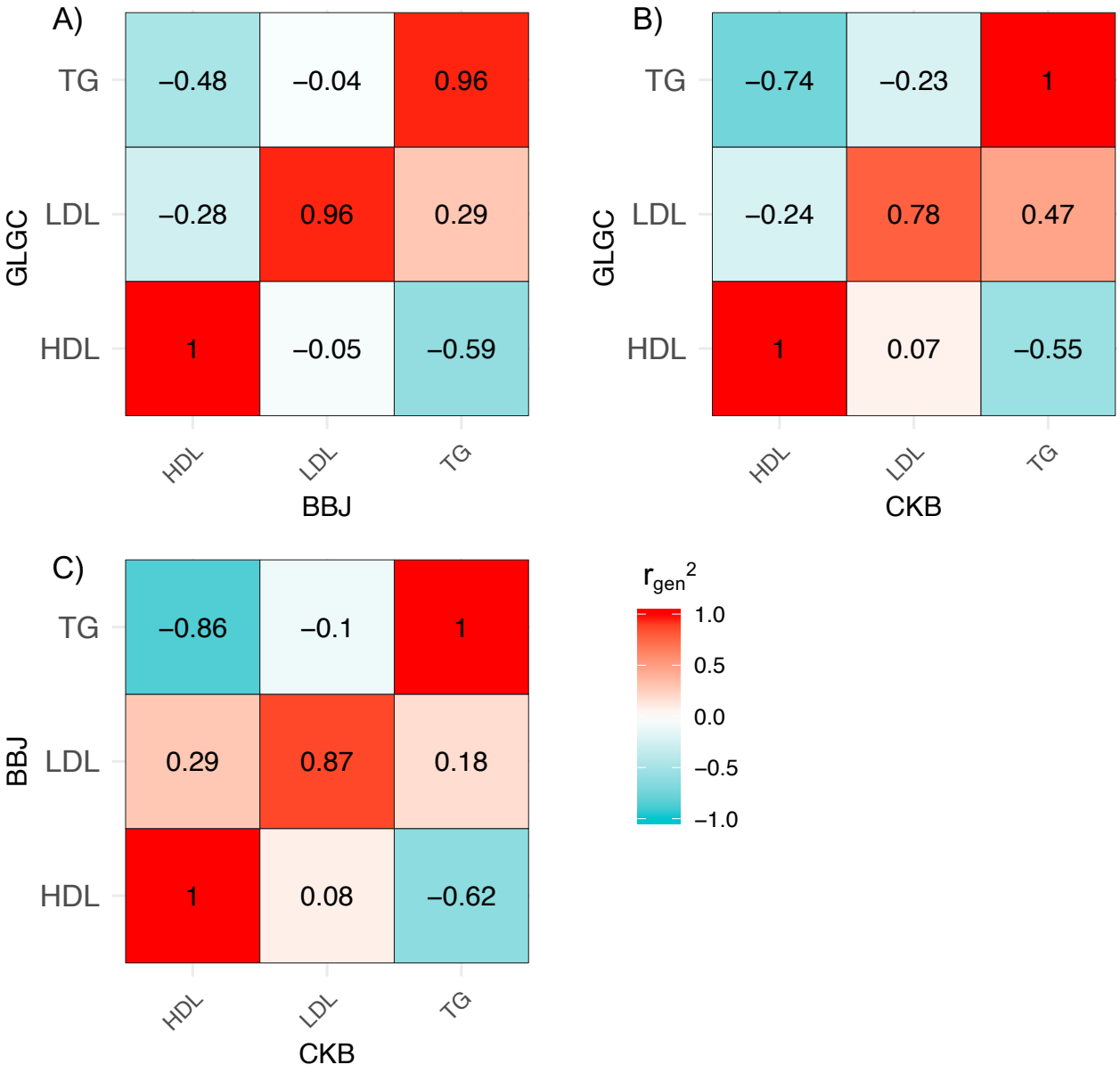
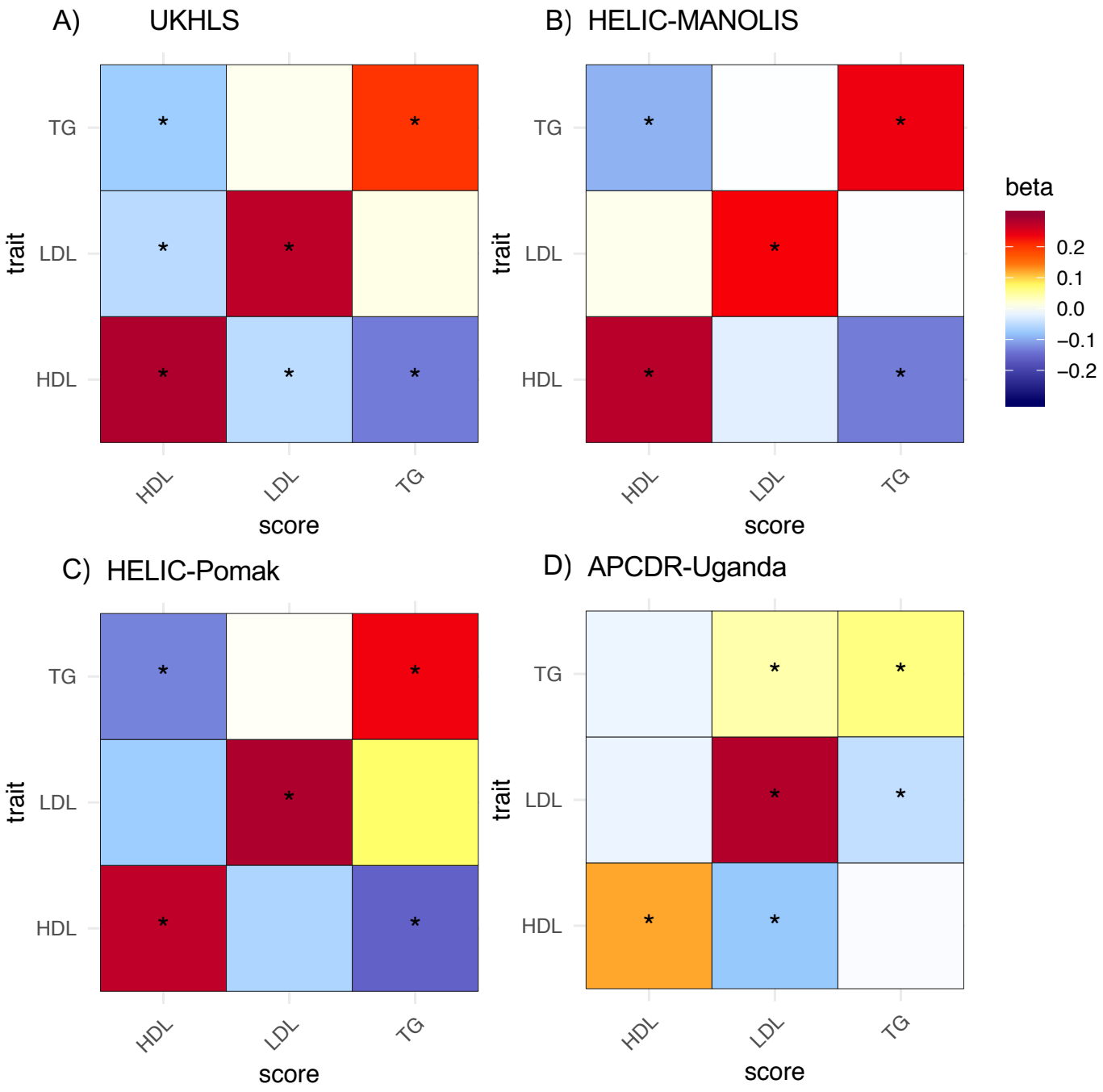
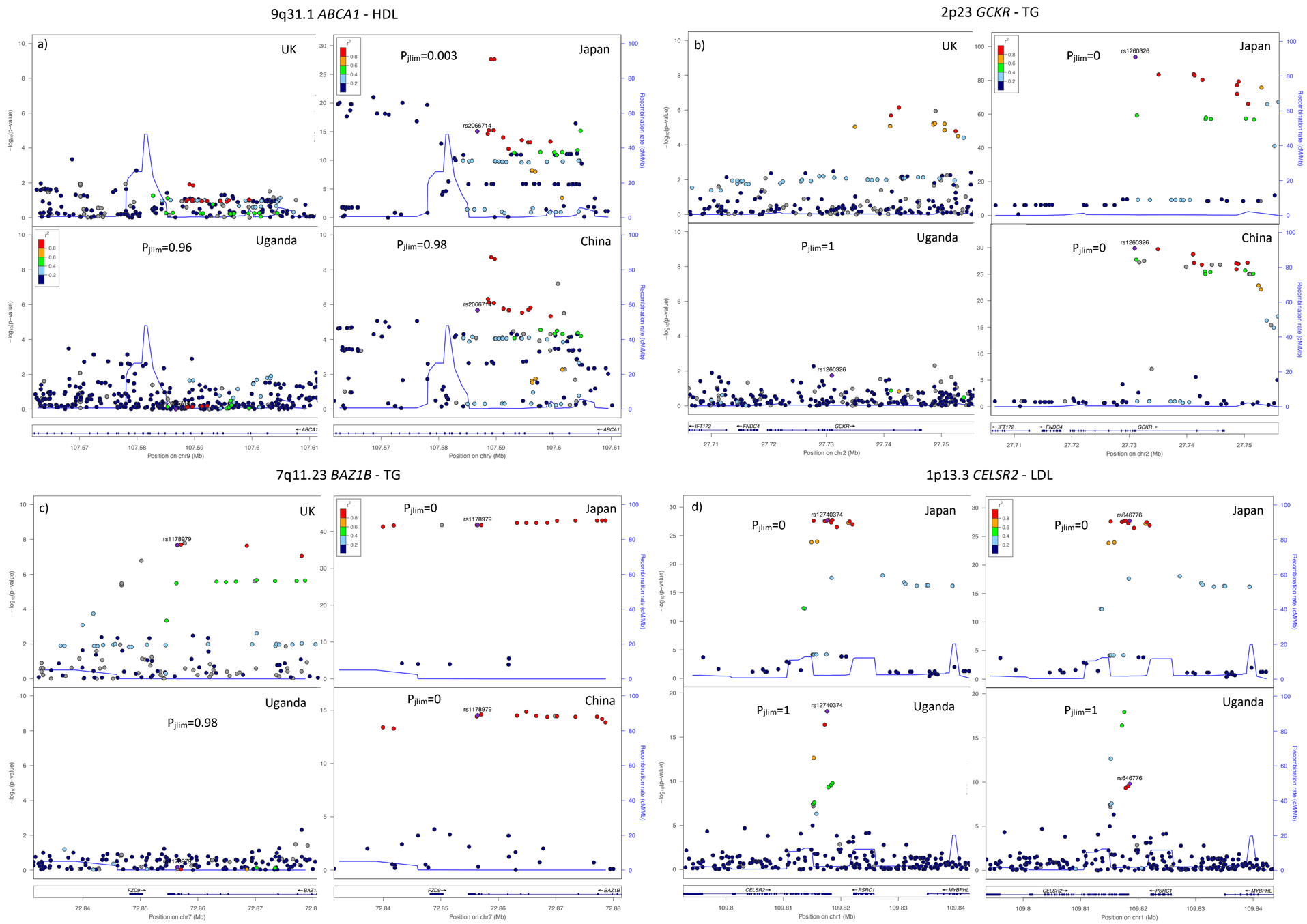

A)
TG
LDL
HDL
GLGC
−0.48
−0.04
0.96
−0.28
0.96
0.29
1
−0.05
−0.59
HDL
LDL
TG
BBJ
B)
TG
LDL
HDL
GLGC
−0.74
−0.23
1
−0.24
0.78
0.47
1
0.07
−0.55
HDL
LDL
TG
CKB
C)
TG
LDL
HDL
BBJ
−0.86
−0.1
1
0.29
0.87
0.18
1
0.08
−0.62
HDL
LDL
TG
CKB
r_{gen}^2
1.0
0.5
0.0
−0.5
−1.0

Figure 2


A) UKHLS
B) HELIC-MANOLIS
C) HELIC-Pomak
D) APCDR-Uganda
trait
TG
LDL
HDL
score
beta
0.2
0.1
0.0
−0.1
−0.2


9q31.1 ABCA1 - HDL
a)
UK
Japan
Uganda
China
Pjlim=0.003
Pjlim=0.96
Pjlim=0.98
rs2066714
2p23 GCKR - TG
b)
Pjlim=0
Pjlim=1
rs1260326
7q11.23 BAZ1B - TG
c)
rs1178979
Pjlim=0.98
1p13.3 CELSR2 - LDL
d)
rs12740374
rs646776
Pjlim=1
Position on chr9 (Mb)
Position on chr2 (Mb)
Position on chr7 (Mb)
Position on chr1 (Mb)
Recombination rate (cM/Mb)