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

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

[1] Division of Biosciences, 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] Anogia Medical Centre, Anogia, 740 51, Greece

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

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

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

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

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

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

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

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

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

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

[19] Institute of Translational Genomics, Helmholtz Zentrum München, D-85764 Neuherberg, Germany

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

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

[+]These authors contributed equally

*email: k.kuchenbaecker@ucl.ac.uk

Abstract

The under-representation of non-European samples in genome-wide association studies could ultimately restrict who benefits from medical advances through genomic science. Our aim was therefore to address the fundamental question whether causal variants for blood lipids are shared across populations.

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 (correlation coefficient r=0.23 to 0.28 per LDL standard deviation, $p<1.9x10^{-14}$). Trans-ethnic genetic correlations between European ancestry, Chinese and Japanese cohorts did not differ significantly from 1 for HDL, LDL and triglycerides. In each study, >60% of major lipid loci displayed evidence of replication with one exception. There was evidence for an effect on serum levels in the Ugandan samples for only 10% of major triglyceride loci. The PRS was only weakly associated in this group (r=0.06, SE=0.013). We establish trans-ethnic colocalization as a method to distinguish shared from population-specific trait loci.

Our results provide evidence for high levels of consistency of genetic associations for cholesterol biomarkers across populations. However, we also demonstrate that the degree of shared causal genetic architecture can be population-, trait- and locus-specific. Efforts to implement genetic risk prediction in clinical settings should account for this.

Introduction

As the predictive ability of common variants for complex traits improves, risk prediction in clinical settings finds increasing consideration[1,2]. However, individuals with non-European ancestry are strongly underrepresented in genome-wide association studies (GWAS), with 79% of participants classified as European and only 3% as African descent[3,4]. Consequently, it is important to determine the transferability of existing findings based on samples with European ancestry. Previous research focused on the effects of different allele frequencies and linkage disequilibrium (LD)[5]. Here we ask the fundamental question whether causal variants for lipid traits are shared across populations. Heterogeneity in effects of variants could result from epistasis or gene-environment interactions. As the observable association of a variant depends on its correlation with the causal variant(s), differences in LD structure between populations make it difficult to compare GWAS results directly[5]. Differences in frequency also impact on the power to detect associations in other ancestry groups.

We employed several strategies to account for these effects and quantify the extent to which genetic variants affecting lipid biomarkers are shared between individuals of European, Asian, and African descent. We assessed the transferability of individual signals and compared association patterns across the genome using data from the African Partnership for Chronic Disease Research – Uganda (APCDR, N=6,407)[6], China Kadoorie Biobank (N=21,295)[7], Biobank Japan (N=162,255)[8], the Global Lipid Genetics Consortium (European ancestry, GLGC2013 N=188,577, GLGC2017 N=237,050)[9,10], the Hellenic Isolated Cohorts (HELIC-MANOLIS, N=1,641 and HELIC-Pomak, N=1,945)[11,12], and the UK Household Longitudinal Study (UKHLS, N=9,961)[13].

Results

We assessed replication rates across established lipid-associated variants in different populations. We distinguished major lipid loci, i.e. those with $p<10^{-100}$ in the largest European ancestry GWAS. Replication was operationalised as at least one variant from the credible set being associated at $p<10^{-3}$ in the target study. As a benchmark, we also assessed replication in two European ancestry studies.

We found evidence of replication for 76.5% of major HDL loci in these two studies (Table 1). For the non-European groups replication rates ranged from 70.6 to 82.4%. Similar replication rates were observed for LDL loci (61.5-76.9%). For major triglycerides (TG) loci, replication rates ranged from 78.9 to 94.7%, with one exception. Only 10.5% of these loci showed evidence of replication in APCDR-Uganda. Replication rates for known loci with $p \geq 10^{-100}$ in the discovery set were generally low. However, Biobank Japan, the largest study, had markedly higher replication rates for these loci than the other studies. Of note, up to 30.8% of the loci that did not replicate contained a variant within 50kb which was associated at $p<10^{-3}$ in the target study, possibly suggesting the presence of independent causal variants.

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

In order to assess patterns of sharing of risk alleles for the smaller studies, we constructed polygenic scores based on the established lipid loci from samples with European ancestry and estimated the score associations with levels of HDL, LDL and TG in HELIC, APCDR-Uganda and also UKHLS as a benchmark (Figure 2). All genetic scores were significantly associated with their respective target lipid in the three European samples with largely consistent correlation coefficients and mutually overlapping 95% confidence intervals (CIs) (Table 2). For HDL, LDL and TG, the estimated correlation coefficients had a range of 0.27-0.28, 0.23-0.28 and 0.20-0.24, respectively. In APCDR-Uganda, the strongest association was observed for LDL (r=0.28, SE=0.01, $p=1.9 \times 10^{-107}$). The HDL association was attenuated compared to the European samples (r=0.12, SE=0.01, $p=6.1 \times 10^{-22}$). The effect of the TG score was markedly weaker (r=0.06, SE=0.01, $p=4.5 \times 10^{-7}$). We also assessed associations between a given score and levels of each of the other biomarkers (Supplementary Table 2). In line with the trans-

ethnic genetic correlation results, we observed inverse associations between the HDL score and TG levels and vice versa in all studies, except APCDR-Uganda.

Differences in LD structure, MAF and sample size make it difficult to assess replication for individual loci. Therefore, we propose a new strategy to assess evidence for shared causal variants. We carried out trans-ethnic colocalization using the JLIM model[14] for each study, comparing it to UKHLS. There was evidence for significant (p_{jlim}<0.05) colocalization with at least one of the target studies for about half of the major lipid loci (Table 3). For example, the 9q31.1 *ABCA1* locus for HDL displayed evidence of a shared causal variant between Biobank Japan and UKHLS (p_{jlim}=0.003) while there was no evidence of association in APCDR-Uganda (p_{jlim}=0.96) (Figure 3a). For several major TG loci, such as *GCKR* at 2p23.3 or *LPL* at 8p21.3, strong evidence of replication in the Asian studies was observed while there was no evidence of association in APCDR-Uganda (Figure 3b,c).

Discussion

Recent efforts to increase global diversity in genetics studies have been vital, enabling this first comprehensive cross-population comparison of genetic associations with blood lipids, covering individual loci as well as patterns across the genome. We provide evidence for extensive sharing of genetic variants regulating levels of high- and low-density lipoprotein cholesterol between individuals with European ancestry and samples from China, Japan, Uganda and Greek population isolates. There was evidence of replication for about three quarters of major HDL and LDL loci. This was highly consistent across all studies. Estimates of trans-ethnic genetic correlations between European, Chinese and Japanese samples were close to 1. Associations of polygenic risk scores for LDL were not attenuated in Ugandans compared to a UK sample. The PRS associations in the two Greek isolated populations were also highly consistent with those in the UK samples. It is important to note, however, that high genetic correlations or consistent PRS associations do not imply replication of all individual loci.

Previous studies that compared the direction of effect of established loci or assessed associations of polygenic scores reported differing degrees of consistency[15–25]. However, most of them were conducted in American samples with diverse ancestry and had smaller sample sizes. The high degree of consistency for cholesterol biomarkers we observe also contrasts with previously reported trans-ethnic genetic correlations for traits such as major depression, rheumatoid arthritis, or type 2 diabetes, which were substantially different from 1[26,27]. In a recent application using data from individuals with European and Asian ancestry from the UK and USA, the average genetic correlation across multiple traits was 0.55 (SE = 0.14) for GERA and 0.54 (SE=0.18) for UK Biobank[28]. The degree of overlap in causal genetic architecture might be population- and trait-specific. In fact, we provide evidence that this can be the case even for closely linked biomarkers involved in fundamental metabolic processes such as blood lipids. We show that many established loci for triglycerides did not affect levels of this biomarker in Ugandan samples. This includes major loci that were associated at genome-wide significance in all the other studies, such as *GCKR* at 2p23.3 or *LPL* at 8p21.3. The polygenic score for triglycerides had a weak effect on measured levels in APCDR-Uganda. This is unlikely to be an artefact of unreliable measurement: triglyceride levels have a heritable component in this sample (SNP heritability of 0.25, SE=0.05[6]) and there are some variants that are associated at genome-wide significance (Supplementary Figure 3e). It is also unlikely that this can be explained purely by differences in LD and MAF because they would affect the analyses of the other two biomarkers as well. The extensive overlap in genetic risk factors for HDL and LDL biomarkers implies significant benefits from combining data across ancestry groups to empower locus discovery and fine-mapping. However, we demonstrate that the degree of overlap can be population-, trait- and locus-specific. For other traits, it is therefore important to systematically assess the genetic architecture in the groups of interest first. The lack of shared causal variants at major loci for TG between Europeans and Ugandans might be a consequence of gene-environment interactions, for example involving dietary factors.

Differences in LD structure, MAF and sample size make it difficult to assess replication of individual loci by comparing associations of single variants. We therefore propose a new approach: trans-ethnic colocalization. It identified shared causal variants even at loci where none of the individual variants

were associated at stringent p-value thresholds. Studying the causes for discordant loci between African, Asian and European populations has promise to further elucidate the biological mechanisms of lipid regulation. Of note, for many of the major lipid loci, more than one independent association signal has been identified in discovery GWAS[10]. When these are located in close proximity to each other, they can interfere with the trans-ethnic colocalization analysis because JLIM assumes a single causal variant (Figure 3d). Therefore, future work should extend this approach to accommodate loci harbouring multiple causal variants.

Applying genetic risk prediction within clinical settings is receiving increasing attention, highlighting the importance of determining the transferability of findings across populations at a time of rising incidence of cardiovascular disease in many low- and middle-income nations. Ongoing programs in under-represented countries[29], such as the Human Hereditary and Health in Africa Initiative[30], and programs focussing on under-represented groups, such as PAGE[31], the All of Us Research program[32], or East London Genes and Health[33], will provide the basis for further insights into the transferability of genetic results across many traits, ancestry groups and environments.

Methods

Data resources

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

Each study underwent standard quality control. Details of the genome-wide association analyses with lipid traits have been previously described for GLGC[9], BBJ[8], HELIC[11], and UKHLS[13]. The association

analysis for APCDR-Uganda was carried out within a mixed model framework using GEMMA[34]. Rank-based inverse normal transformation was applied to the lipid biomarkers after adjusting for age and gender. In China Kadoorie Biobank, lipid levels were regressed against eight principle components, region, age, age^2, sex, and - for LDL and TG - fasting time and fasting time^2. LDL levels were derived using the Friedewald formula. After rank-based inverse normal transformation, the residuals were used as the outcomes in the genetic association analyses using linear regression. In eMERGE, biomarkers were adjusted for age, gender, kidney disease, statin use, type 2 diabetes status and disorders relating to growth hormones. Associations were carried out within a mixed model framework using BOLT-LMM[35]. Manhattan plots for eMerge, UKHLS, BBJ, CKB, and APCDR-Uganda are shown in Supplementary Figures 1-3.

Established lipid loci

A list of established lipid-associated loci was extracted from a 2017 Global Lipid Genetics Consortium (GLGC) publication[10] reporting 444 independent variants in 250 loci genome-wide significantly associated with HDL, LDL, and triglyceride levels. We excluded three LDL variants where the association was not primarily driven by the samples with European ancestry. We assessed evidence of replication of the loci, applied trans-ethnic colocalization and used them to construct polygenic risk scores.

Replication of established lipid loci

We assessed evidence of replication across these established lipid variants. For loci harbouring multiple signals, we only kept the most strongly associated variant. This left 170 HDL, 135 LDL and 136 TG variants. We distinguished major loci, i.e. those with $p<10^{-100}$ in GLGC2017. For each lead SNP we identified all variants in LD ($r^2>0.6$) based on the European ancestry 1000 Genomes data. We assessed whether the lead or any of the correlated variants, henceforth called credible set, displayed evidence of association in the target study. We used a p-value threshold of $p<10^{-3}$. If this was not the case, we tested whether there was any other variant with evidence of association within a 50Kb window. While this p-value threshold might not be appropriate to provide conclusive evidence of replication for

individual loci, we used this to test evidence of replication across sets of loci. As a benchmark, we computed the minimum p-value in 1000 random windows of 50Kb for each study. Less than 5% of random windows has a minimum $p<10^{-3}$ for the non-European ancestry studies and UKHLS. In eMerge 10.9% for windows there was a variant with $p<10^{-3}$. However, we maintained the $p<10^{-3}$ as the threshold for replication because our focus was on the non-European samples.

Trans-ethnic genetic correlations

We used the popcorn software[26] to estimate trans-ethnic genetic correlations between studies while accounting for differences in LD structure. This provides an indication of the correlation of causal-variant effect sizes across the genome at SNPs common in both populations. Variant LD scores were estimated for ancestry-matched 1000 Genomes data for each study combination. The estimation of LD scores failed for chromosome 6 for some groups. We therefore left out chromosome 6 from all comparisons. Variants with imputation accuracy $r^2<0.8$ or MAF<0.01 were excluded. Popcorn did not converge for any of the studies with less than 20,000 samples. Therefore, results are presented for comparisons between GLGC2013, CKB and BBJ. We estimated effect rather than impact correlations.

Polygenic risk scores

We created polygenic risk scores based on the established lipid loci and assessed their associations with lipid levels in UKHLS, the HELIC cohorts, and APCDR-Uganda, as it was not possible to compute trans-ethnic genetic correlations for these studies. For HELIC and UKHLS, extreme values ($\mu \pm 3\,SD$, sex stratified) were filtered. Age, age^2 and sex were adjusted for by regressing them on the biomarker values and using the residuals as outcomes for subsequent analyses. For each biomarker in each sample set, we checked normality and homoscedasticity. HDL and LDL were approximately normally distributed. For TG levels, a Box Cox transformation was used to normalize the data. APCDR-Uganda phenotype data were rank-based inverse normally transformed.

To make sure PRS were comparable across studies, we excluded variants that were absent, rare (MAF<0.01) or badly imputed ($r^2<0.8$) in any of the studies and variants that had different alleles from

those in the GLGC. The variant with larger discovery p-value from each correlated pair of SNPs (r^2>0.1) was also removed. This left 120, 103 and 101 variants for HDL, LDL and TG, respectively. We created trait-specific weighted PRS. The β-regression coefficients from SNP-trait associations in GLGC2017[10] were used as weights. All biomarkers and scores were scaled to mean=0 and standard deviation=1 for each study so that the regression coefficient represent estimates of the correlation between scores and biomarkers.

We carried out association analyses between each polygenic score and each biomarkers using a linear mixed model with random polygenic effect implemented in GEMMA[34] in order to account for relatedness and population structure. We used a Bonferroni correction to adjust for multiple testing of three PRS with three different biomarker outcomes (0.05/9=0.0056).

Trans-ethnic colocalization

Differences in allele frequency, LD structure and sample size make it difficult to assess whether a given GWAS hit replicates in samples with different ancestries. Therefore, we applied trans-ethnic colocalization. Colocalization methods test whether the associations in two studies can be explained by the same underlying signal even if the specific causal variant is unknown. The joint likelihood mapping (JLIM) statistic was developed by Chun and colleagues to estimate the posterior probabilities for colocalization between GWAS and eQTL signals and compare them to probabilities of distinct causal variants[14]:

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

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

JLIM explicitly accounts for LD structure. Therefore, we assessed whether it is suitable for trans-ethnic colocalization. For samples with summary statistics, LD scores were estimated using ancestry matched

samples from the 1000 Genomes Project v3. JLIM assumes only one causal variant within a region in each study. We therefore used a small windows of 50Kb for each known locus to minimise the risk of interference from additional association signals. Distinct causal variants were defined by separation in LD space by $r^2 \geq 0.8$ from each other. We excluded loci within the major histocompatibility region due to its complex LD structure. We used a significance threshold of $p<0.05$ given the evidence of association of the established lipid loci in Europeans and the overall evidence for shared causal genetic architecture across populations for most lipid traits from our other analyses. We compared each target study to UKHLS because of their high level of homogeneity in terms of ancestry, biomarker quantification and study design.

Data availability

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

Acknowledgements

We would like to thank the participants and volunteers of the each of the studies. We appreciate the advice received from Brielin Brown regarding the use of Popcorn and from Shamil Sunyaev and Sung Chun regarding the suitability of JLIM for trans-ethnic colocalization.

The Understanding Society Scientific Group includes the following members: Michaela Benzeval, Jonathan Burton, Nicholas Buck, Annette Jäckle, Meena Kumari, Heather Laurie, Peter Lynn, Stephen Pudney, Birgitta Rabe, Dieter Wolke

Author contributions

KK conceived this project and supervised the work. KK, NT and TR carried out the analyses. KK and TR wrote the manuscript. HELIC: EZ and GD are the principle investigators, AG and LS carried out the quality control, MK and ET were involved in data collection. APCDR-Uganda: MS, DG, GA, JS, AK were involved in collecting and preparing data as well as leading the study. China Kadoorie Biobank: ZG and LL are the principle investigators; RW is the genomics lead; RW, IM and YG were involved in data collection, IM prepared the phenotypes, RW and KL carried out quality control. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

References

1. 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).
2. 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).

3. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
4. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nat. News* **538**, 161 (2016).
5. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
6. 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).
7. 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).
8. 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).
9. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
10. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758 (2017).
11. Gilly, A. *et al.* Cohort-wide deep whole genome sequencing and the allelic architecture of complex traits. *Nat. Commun.* **9**, 4674 (2018).
12. Gilly, A. *et al.* Very low depth whole genome sequencing in complex trait association studies. *Bioinforma. Oxf. Engl.* (2018). doi:10.1093/bioinformatics/bty1032
13. 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).
14. 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).
15. Musunuru, K. *et al.* Multi-Ethnic Analysis of Lipid-Associated Loci: The NHLBI CARe Project. *PLoS ONE* **7**, (2012).
16. Teslovich, T. M. *et al.* Biological, Clinical, and Population Relevance of 95 Loci for Blood Lipids. *Nature* **466**, 707–713 (2010).

17. 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).
18. Bryant, E. K. *et al.* A Multiethnic Replication Study of Plasma Lipoprotein Levels-Associated SNPs Identified in Recent GWAS. *PLoS ONE* **8**, (2013).
19. Wu, Y. *et al.* Trans-Ethnic Fine-Mapping of Lipid Loci Identifies Population-Specific Signals and Allelic Heterogeneity That Increases the Trait Variance Explained. *PLoS Genet.* **9**, (2013).
20. Wang, Z. *et al.* Genetic associations with lipoprotein subfraction measures differ by ethnicity in the multi-ethnic study of atherosclerosis (MESA). *Hum. Genet.* **136**, 715–726 (2017).
21. Keebler, M. E. *et al.* Association of Blood Lipids with Common DNA Sequence Variants at Nineteen Genetic Loci in the Multiethnic United States National Health and Nutrition Examination Survey III. *Circ. Cardiovasc. Genet.* **2**, 238–243 (2009).
22. Lanktree, M. B., Anand, S. S., Yusuf, S. & Hegele, R. A. Replication of genetic associations with plasma lipoprotein traits in a multiethnic sample. *J. Lipid Res.* **50**, 1487–1496 (2009).
23. Chang, M. *et al.* Racial/ethnic variation in the association of lipid-related genetic variants with blood lipids in the US adult population. *Circ. Cardiovasc. Genet.* **4**, 523–533 (2011).
24. Below, J. E. *et al.* Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. *Sci. Rep.* **6**, (2016).
25. Johnson, L., Zhu, J., Scott, E. R. & Wineinger, N. E. An Examination of the Relationship between Lipid Levels and Associated Genetic Markers across Racial/Ethnic Populations in the Multi-Ethnic Study of Atherosclerosis. *PloS One* **10**, e0126361 (2015).
26. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
27. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
28. Galinsky, K. J. *et al.* Estimating cross-population genetic correlations of causal effect sizes. *Genet. Epidemiol.* (2018). doi:10.1002/gepi.22173

29. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**, 175–185 (2018).
30. Mulder, N. *et al.* H3Africa: current perspectives. *Pharmacogenomics Pers. Med.* **Volume 11**, 59–66 (2018).
31. Carlson, C. S. *et al.* Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLOS Biol.* **11**, e1001661 (2013).
32. Program Overview - All of Us | National Institutes of Health. Available at: https://allofus.nih.gov/about/about-all-us-research-program. (Accessed: 28th December 2018)
33. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community based population genomics and health study in people of British-Bangladeshi and -Pakistani heritage. *bioRxiv* (2018). doi:10.1101/426163
34. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
35. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).

Figure legends

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

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

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

Tables

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

p in GLGC:		**<10^{-100}**			**≥10^{-100}**		
study	**trait**	**n.s.***	**region#**	**credible^**	**n.s.***	**region#**	**credible^**
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

* no variant in the region associated in target set at p<10^{-3}

no variant in the credible set associated in the target set at p<10^{-3} but an uncorrelated variant in the region is associated in target set at p<10^{-3}

^ a variant in the credible set is associated in the target set at p<10^{-3}

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

trait	N	correlation (SE*)	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}

* SE=standard error

Table 3. P-value for the trans-ethnic colocalization based on the JLIM model for established lipid-associated loci in UKHLS, China Kadoorie Biobank (CKB), Biobank Japan (BBJ) and Uganda (APCDR).

					GLGC		replication*				JLIM p-value#		
rs-id	**chr**	**position**	**near gene**	**MAF**	**p-value**	**multi$**	**UKHLS**	**CKB**	**BBJ**	**APCDR**	**CKB**	**BBJ**	**APCDR**
HDL													
rs4660293	1	40028180	PABPC4	0.21	6.1E-36	distant	uc	ns	cor	ns	0.8	0.014	NA
rs11755393	6	34824636	UHRF1BP1	0.36	4.2E-23	distant	ns	ns	cor	ns	0.035	0	NA
rs1178979	7	72856430	BAZ1B	0.18	1.3E-26	near	uc	uc	cor	ns	0	0	NA
rs4731702	7	130433384	SNX27	0.46	1.2E-35	distant	ns	cor	cor	ns	0.14	0.003	NA
rs4841132	8	9183596	NECAP2	0.9	1.0E-123	no	ns	ns	ns	cor	0.99	0.24	0.16
rs328	8	19819724	LPL	0.098	1.7E-316	near	cor	cor	cor	cor	0.21	0.005	0.62
rs2954033	8	126493746	AMPD1	0.72	3.0E-61	near	cor	cor	cor	ns	0.94	0.002	NA
rs643531	9	15296034	TTC39B	0.88	3.8E-42	no	ns	ns	uc	uc	NA	NA	NA
rs2066714	9	107586753	ABCA1	0.15	3.6E-31	near	uc	cor	cor	uc	0.97	0.007	0.94
rs1883025	9	107664301	ABCA1	0.26	2.1E-118	near	uc	cor	cor	uc	0.84	0.8	0.97
rs2792751	10	113940329	GPAM	0.73	3.8E-21	near	ns	ns	cor	uc	0.002	0.002	NA
rs7350481	11	116586283	C1orf158	0.91	3.2E-100	distant	cor	cor	cor	uc	0	0	0.99
rs964184	11	116648917	PRAMEF2	0.85	2.6E-217	near	cor	cor	cor	uc	1	1	1
rs10468017	15	58678512	SNX27	0.27	1.8E-306	near	cor	cor	cor	cor	1	0.98	0.99
rs1800588	15	58723675	SNX27	0.24	0	distant	cor	cor	cor	cor	0.007	0.009	0.017
rs247616	16	56989590	OR10K2	0.31	0	near	cor	cor	cor	cor	0	0	0
rs3764261	16	56993324	OR10K2	0.31	0	near	cor	cor	cor	cor	0	0	0
rs34065661	16	56995935	CETP	0.005	5.6E-103	near	uc	uc	uc	cor	0	0	0
rs16942887	16	67928042	PSKH1	0.13	9.8E-93	near	ns	ns	cor	cor	0.96	0.29	0.025
rs72836561	17	41926126	CD300LG	0.028	8.1E-111	no	uc	uc	uc	ns	NA	NA	NA
rs7241918	18	47160953	MTHFR	0.85	1.2E-104	distant	cor	cor	cor	ns	0.16	1	1
rs116843064	19	8429323	ANGPTL4	0.02	4.8E-146	near	uc	ns	ns	uc	NA	NA	NA
rs769449	19	45410002	APOE	0.11	6.9E-129	near	cor	cor	cor	cor	0.009	0.02	0.95
rs386000	19	54792761	SLC45A3	0.22	1.1E-41	distant	cor	uc	cor	cor	0	1	0.71
LDL													
rs11591147	1	55505647	PCSK9	0.015	0.0	near	uc	uc	uc	uc	0.94	0.64	0.92
rs12740374	1	109817590	CELSR2	0.22	0.0	near	cor	cor	cor	cor	0	0	0

rs1367117	2	21263900	APOB	0.28	3.6E-278	near	cor	cor	cor	cor	1	1	0
rs541041	2	21294975	TNFSF4	0.81	1.3E-287	distant	cor	uc	uc	cor	1	1	NA
rs4245791	2	44074431	ABCG8	0.72	1.7E-120	near	uc	ns	ns	ns	NA	0.81	0.99
rs3846662	5	74651084	HMGCR	0.48	3.3E-128	near	cor	cor	cor	ns	0.002	0	1
rs2737229	8	116648565	TRPS1	0.34	8.9E-15	distant	cor	ns	cor	ns	0.015	0.025	NA
rs635634	9	136155000	IL6R	0.19	4.9E-109	near	ns	cor	cor	ns	0.96	0.97	NA
rs2000999	16	72108093	HPR	0.2	4.0E-71	distant	cor	cor	cor	ns	1	0	1
rs6511720	19	11202306	LDLR	0.11	0.0	near	cor	uc	uc	cor	0	NA	0
rs28399654	19	45316588	BCAM	0.027	7.5E-232	distant	cor	uc	uc	cor	1	1	NA
rs7412	19	45412079	APOE	0.075	0.0E+00	near	cor	cor	cor	cor	0	0	0
Triglycerides													
rs10889353	1	63118196	DOCK7	0.33	6.4E-170	no	cor	cor	cor	uc	0	0	0.88
rs676210	2	21231524	APOB	0.26	4.9E-118	near	cor	uc	cor	uc	1	1	1
rs1260326	2	27730940	GCKR	0.63	0.0	near	cor	cor	cor	ns	0	NA	NA
rs2943641	2	227093745	ALDH4A1	0.66	4.9E-33	no	ns	ns	cor	ns	0.006	NA	1
rs6905288	6	43758873	SH2D5	0.59	9.0E-35	near	cor	ns	cor	ns	0	0	NA
rs1178979	7	72856430	BAZ1B	0.18	1.5E-179	near	cor	cor	cor	ns	0	0	NA
rs35332062	7	73012042	MLXIPL	0.12	5.2E-205	distant	cor	cor	cor	uc	0.99	1	NA
rs326	8	19819439	LPL	0.3	0.0	near	cor	cor	cor	uc	0.91	0.91	0.72
rs2954029	8	126490972	AMPD1	0.45	8.3E-205	near	cor	cor	cor	ns	0	0	1
rs1883025	9	107664301	ABCA1	0.26	1.2E-13	no	uc	ns	ns	ns	1	0.001	NA
rs7350481	11	116586283	C1orf158	0.91	0.0	distant	cor	cor	cor	uc	0	0	0
rs11820589	11	116633862	BUD13	0.066	4.4E-133	near	cor	uc	uc	ns	0	1	1
rs2075291	11	116661392	APOA5	0.003	5.7E-65	near	uc	cor	cor	ns	NA	1	NA
rs10047462	11	116722041	SIK3	0.86	9.9E-180	near	cor	cor	cor	uc	0	0	1
rs247616	16	56989590	OR10K2	0.31	2.4E-38	near	ns	ns	cor	cor	0.014	0.024	0.78
rs116843064	19	8429323	ANGPTL4	0.02	4.2E-175	near	uc	ns	ns	ns	NA	NA	NA
rs58542926	19	19379549	TM6SF2	0.074	3.7E-125	no	cor	cor	cor	ns	NA	NA	NA
rs439401	19	45414451	ATP13A2	0.63	2.7E-168	near	cor	cor	cor	uc	0.009	0.53	1

$ indicates whether multiple independent hits have been reported within 50kb ("near") or 1Mb ("distant")

* indicates whether any variant from the credible set ("cor") or any uncorrelated variant within 50kb ("uc") is associated with the target biomarker at $p<10^{-3}$ in each of the target studies

p-value from the JLIM trans-ethnic colocalization analysis using UKHLS as the comparison set

Figures

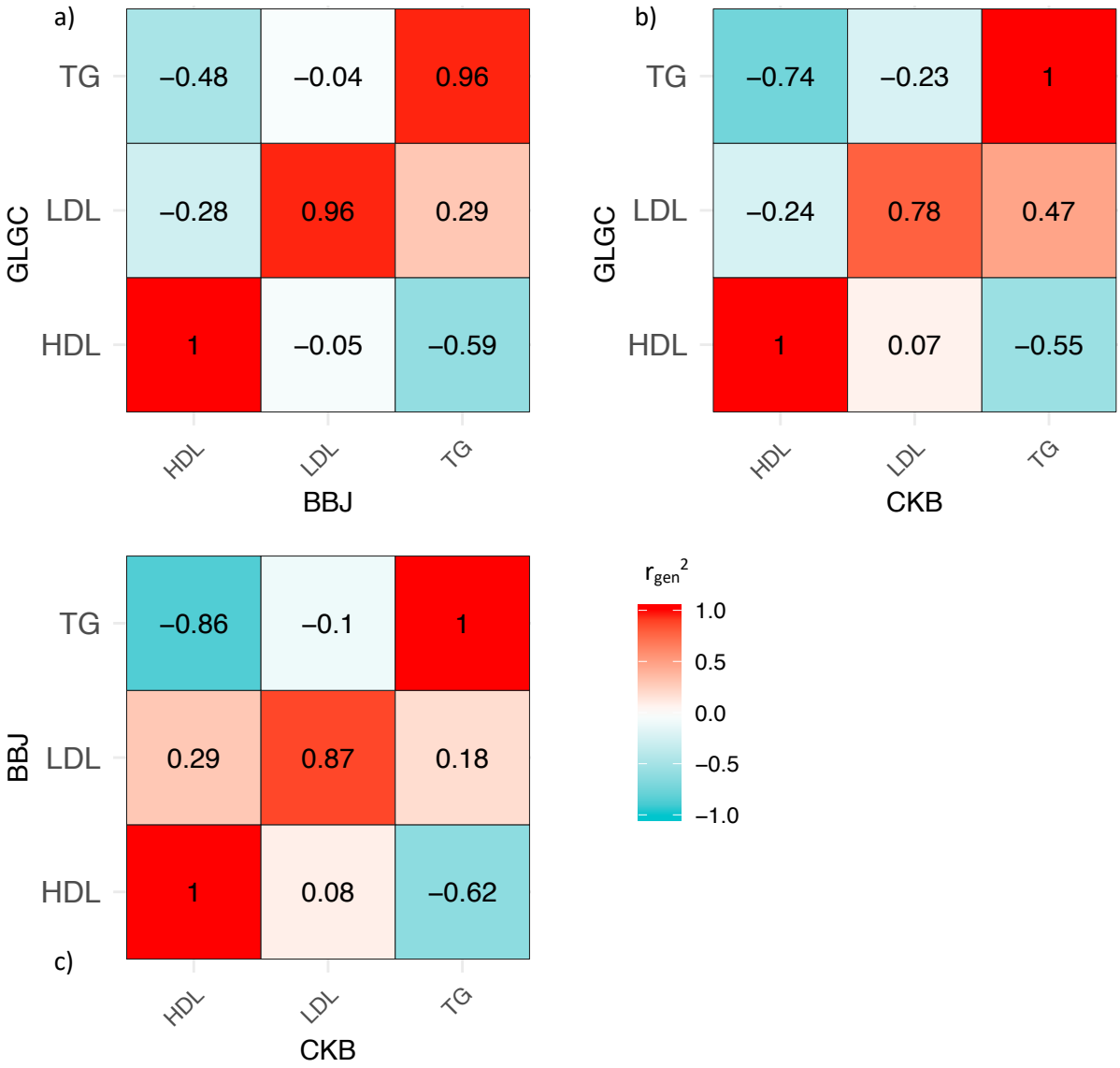


Figure 1. Trans-ethnic genetic correlations for associations with high-density lipoprotein (HDL), low-density lipoprotein (LDL) cholesterol and triglycerides (TG). a) shows the comparison of GLGC2013 (European) and Biobank Japan, b) GLGC2013 and China Kadoorie Biobank and c) Biobank Japan and China Kadoorie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

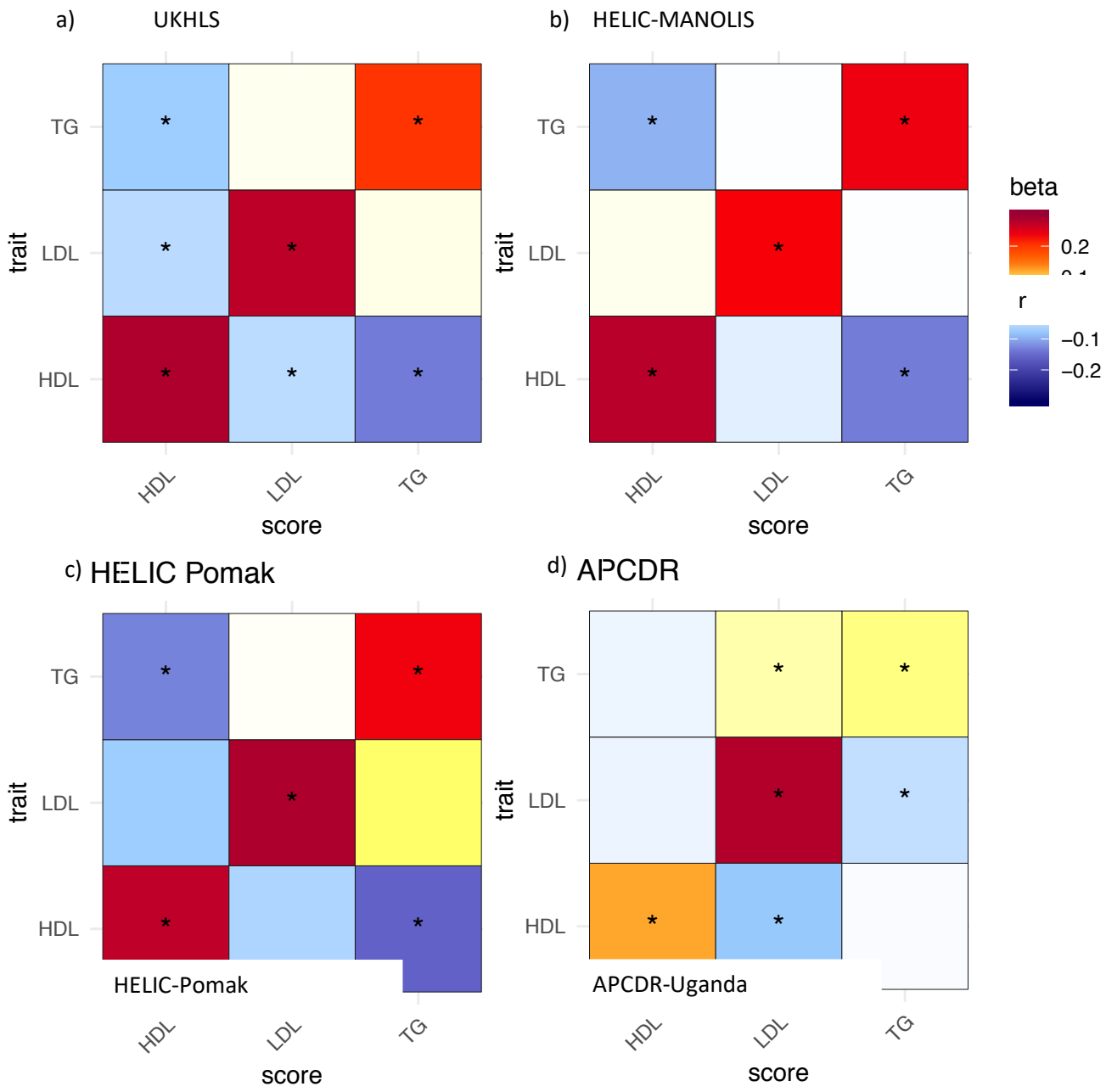
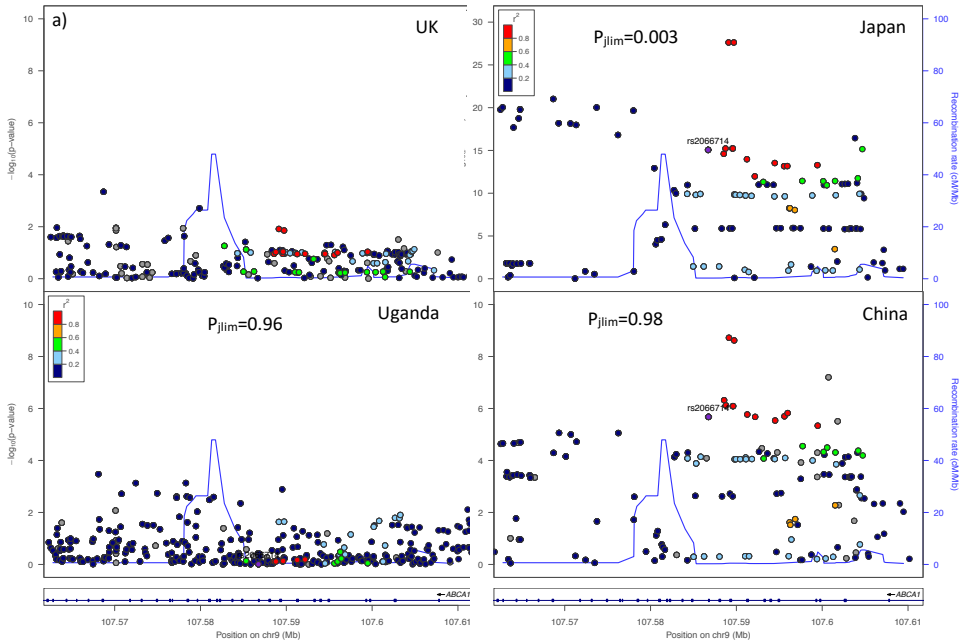
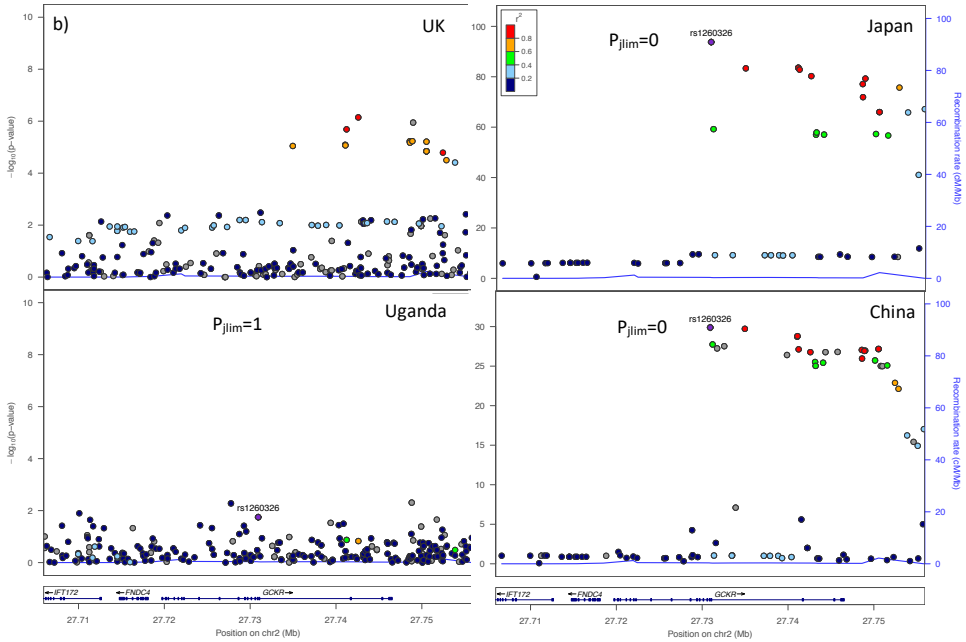


Figure 2. Associations of polygenic scores based on established lipid-associated loci with levels of high-density lipoprotein (HDL), low-density lipoprotein (LDL) cholesterol and triglycerides (TG) in a) UKHLS, b) HELIC-MANOLIS, c) HELIC-Pomak, d) APCDR-Uganda. Estimates are given as correlation coefficients.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

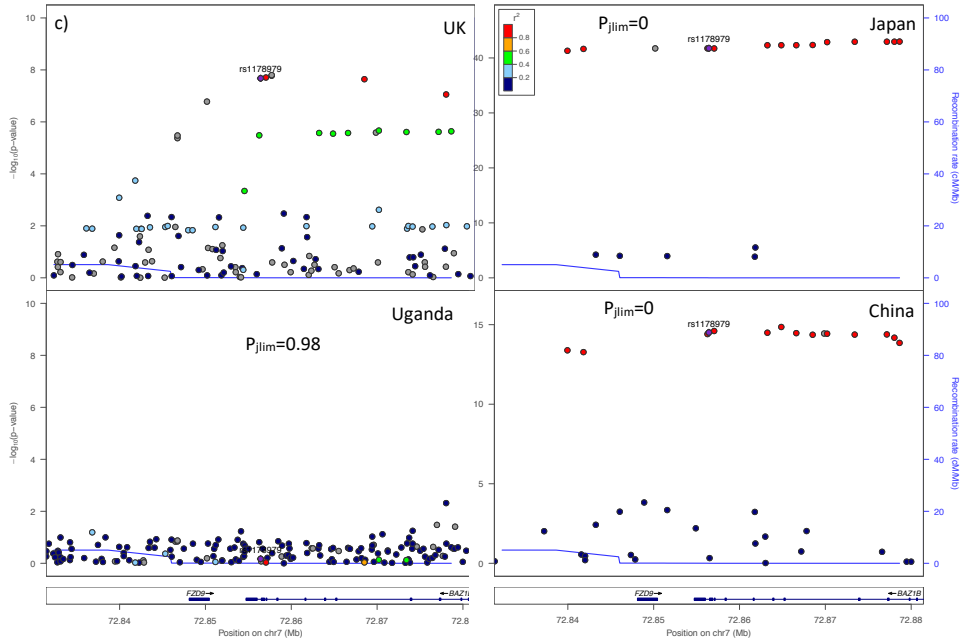
9q31.1 *ABCA1* - HDL



2p23 *GCKR* - TG



7q11.23 *BAZ1B* - TG



1p13.3 *CELSR2* - LDL

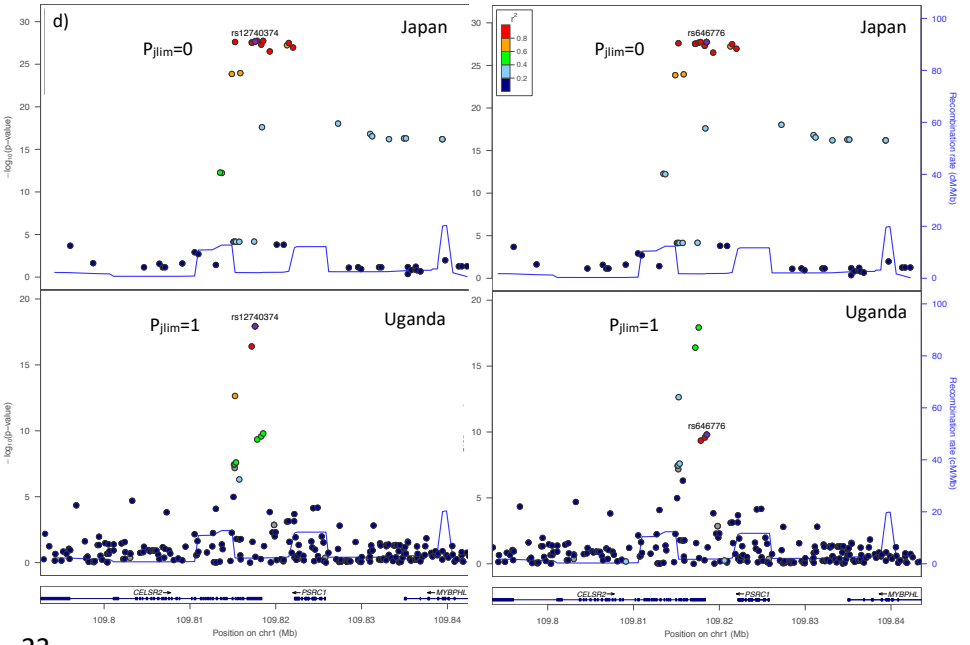


Figure 3. Regional association plots for a selection of established lipid-associated loci for UKHLS, Biobank Japan, APCDR-Uganda, and China Kadoorie Biobank and p-value p_{jlim} for the trans-ethnic colocalization with UKHLS.