

Variants in *LPA* are associated with familial hypercholesterolaemia: whole genome sequencing analysis in the 100 000 Genomes Project

Martin Bird ¹, Antoine Rimbert ², Alan Michael Pittman ³, Steve Eric Humphries ⁴, and Marta Futema ^{1,4*}

¹Experimental Cardiology, Cardiovascular and Genomics Research Institute, School of Health & Medical Sciences, City St George's, University of London, Cranmer Terrace, Tooting, SW17 0RE London, UK; ²Nantes Université, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France; ³Genomics, Cardiovascular and Genomics Research Institute, School of Health & Medical Sciences, City St George's, University of London, London, UK; and ⁴Institute of Cardiovascular Science, Faculty of Population Health, University College London, 5 University Street, WC1E 6JF London, UK

Received 11 July 2024; revised 18 October 2024; accepted 6 November 2024; online publish-ahead-of-print 12 November 2024

Aims

Familial hypercholesterolaemia (FH) is an inherited disease of high LDL cholesterol (LDL-C) caused by defects in *LDLR*, *APOB*, *APOE*, and *PCSK9* genes. A pathogenic variant cannot be found in ~60% of clinical FH patients. Using whole genome sequencing (WGS), we examined genetic determinants of FH.

Methods and results

Whole genome sequencing data generated by the 100 000 Genomes Project (100KGP) included 536 FH patients diagnosed using the FH Simon–Broome criteria. Rare variants in known FH genes were analysed. Genome-wide association study between 443 FH variant-negative unrelated FH cases and 77 275 control participants of the 100KGP was run using high-coverage WGS data. Polygenic risk scores for LDL-C (LDL PRS) and lipoprotein(a) (Lp(a) PRS) were computed. An FH-causing variant was found in 17.4% of FH cases. Genome-wide association study identified the *LPA* gene locus being significantly associated ($P < 1 \times 10^{-8}$). Familial hypercholesterolaemia variant-negative participants had higher LDL and Lp(a) PRSs in comparison with the controls ($P < 1.0 \times 10^{-16}$ and $P < 4.09 \times 10^{-6}$, respectively). Similar associations were found in the monogenic FH with both LDL and Lp(a) PRSs being higher than in controls ($P < 4.03 \times 10^{-4}$ and $P < 3.01 \times 10^{-3}$, respectively). High LDL PRS was observed in 36.4% of FH variant-negative cases, whereas high Lp(a) PRS in 18.5%, with 7.0% having both high LDL and Lp(a) PRSs.

Conclusion

This genome-wide analysis of monogenic and polygenic FH causes confirms a complex and heterogeneous architecture of hypercholesterolaemia, with the *LPA* gene playing a significant role. Both Lp(a) and LDL-C should be measured for precision FH diagnosis. Specific therapies to lower Lp(a) should be targeted to those who will benefit most.

Lay summary

Familial hypercholesterolaemia (FH) is a genetic condition that causes very high levels of 'bad' cholesterol (LDL-C), which increases the risk of heart disease. This study used advanced genetic testing (whole genome sequencing) to better understand the genetic causes behind high cholesterol in people with FH.

Key findings

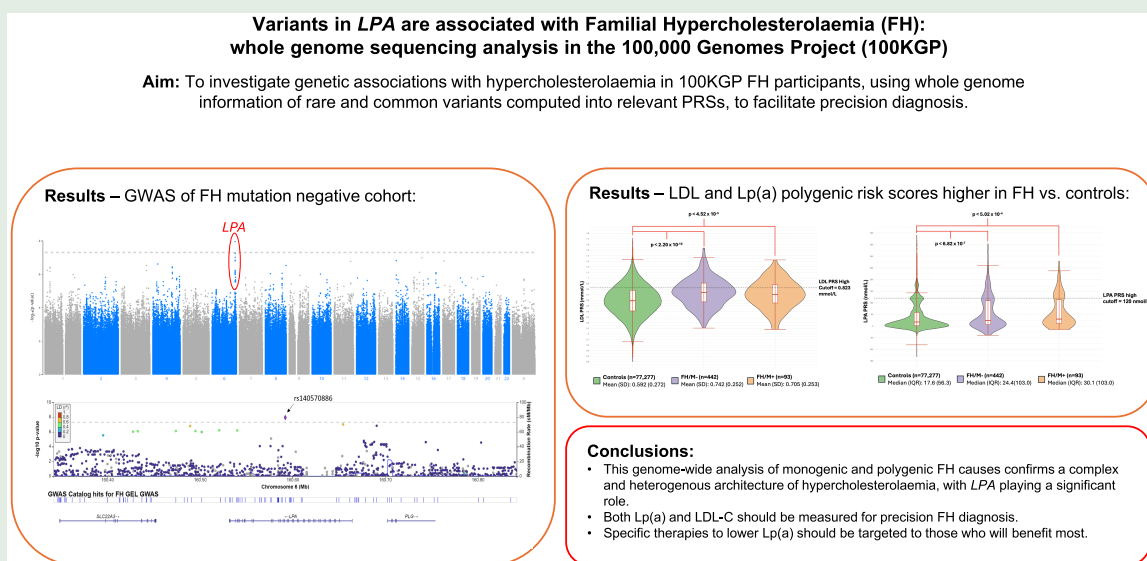
- FH is caused by both rare and common genetic changes. Many small changes in DNA combine to increase a person's overall risk, which is measured through what is called a 'polygenic risk score'. Polygenic score for LDL-C and Lp(a) is higher in FH individuals.
- The gene (*LPA*) that determines the blood amount of lipoprotein(a), or Lp(a), plays an important role in FH-affected individuals. High levels of Lp(a) are a known risk factor for heart disease, separate from LDL-C. This means that Lp(a) levels should be checked in FH patients to better manage their condition and prevent heart disease.

* Corresponding author. Tel: +0044 (0)20 8725 5587, Email: mfutema@sul.ac.uk

© The Author(s) 2024. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Graphical Abstract



Keywords

Familial hypercholesterolaemia • Whole genome sequencing • Lp(a) • *LPA* • Genome wide association study

Introduction

Familial hypercholesterolaemia (FH) is characterized by high concentrations of LDL cholesterol (LDL-C) and risk of premature coronary heart disease (CHD). The disease is caused by pathogenic variants mainly in the *LDLR*, *APOB*, and *PCSK9* genes that lead to a defective clearance of circulatory LDL-C, exposing the affected individual to a lifelong burden of LDL-C.¹ More recently a single variant in the *APOE* gene (p.Leu167del) has been also associated with FH.¹ A recessive form of inherited hypercholesterolaemia is also known due to homozygous or compound heterozygous pathogenic variants in *LDLRAP1*.² Familial hypercholesterolaemia-causing variants in the aforementioned genes are found at a frequency of 1 in 250 individuals (95% confidence interval, 1:345–1:192) in heterogeneous populations,³ with a similar frequency observed in individuals of South Asian and African ancestries.⁴ Since FH variant carriership predisposes individuals to premature CHD,⁵ early diagnosis and aggressive lipid-lowering therapy are key for an effective prevention of cardiovascular events in FH patients.⁶

The great majority of FH-causing variants are located in the *LDLR* gene. In the UK, a single pathogenic variant in the *APOB* gene, the p.Arg3527Gln, can be found in about 10% of FH patients, with other causes in *PCSK9* and *APOE* accounting for about 1–2% of FH-causing variants.⁷ Sequencing of the relevant FH genes is required to identify the causative variant, which then can be used in the cost-effective cascade testing within the family of the index patient.⁸ The detection rate in individuals with a clinical diagnosis of definite FH (DFH) may be as high as 70% while it is much lower (20–30%) in those with a clinical diagnosis of possible FH (PFH). However, since in most clinic cohorts there are twice as many PFH as DFH patients, overall, a pathogenic or likely pathogenic variant in one of the aforementioned genes is usually found in only 20–40% of patients sent for genetic testing.⁹ More recent studies suggest the FH variant detection rate is decreasing to around 15% in the Netherlands¹⁰ and was 17% in a UK survey of over 9500 patient samples in the UK Genomic Laboratory hubs.¹¹ Our previous attempt

employing whole exome sequencing of variant-negative FH patients has highlighted a few potential novel causes,¹² one of which, the *RBM25* gene, has been shown to be involved in the posttranscriptional regulation of LDL-receptor.¹³ Recently, using the Bayesian genetic testing for rare variants in whole genomes, the *RAB35* gene has been associated with FH.¹⁴ Although it is possible that some of the variant-negative FH individuals have a novel genetic variant that affect the LDL-C uptake, such cases are expected to be very rare; therefore, large-scale sequencing studies are required to identify any potential novel signals.

Individuals diagnosed with FH based on the clinical criteria, for example, the Simon–Broome diagnostic criteria, but with a negative result of an FH genetic test, have been shown to have significantly higher LDL-C polygenic risk score (PRS) than non-FH patient cohorts, and higher, although to a lesser extent, than those with monogenic FH,^{15,16} indicating a complex aetiology of the disease. Although PRS should not be interpreted as a diagnostic test in isolation of other risk factors, the high burden of LDL-C-associated variants (LDL PRS \geq 80th percentile) has been shown to significantly increase the risk of premature atherosclerotic cardiovascular disease when compared with individuals with low PRS (<20th percentile).¹⁵ Moreover, studies that employed CHD-PRS suggested that individuals with PRS at the top end of its distribution were predisposed to the same risk of developing CHD as those with monogenic FH.^{17,18}

Recent studies of variant-negative FH patients observed higher concentrations of lipoprotein(a) (Lp(a)) in comparison with those with monogenic FH.^{19,20} Lipoprotein(a) concentrations in plasma are strongly genetically determined,²¹ and high concentrations of Lp(a) have been demonstrated to be an independent causal risk factor for CHD²² also associated with the severity of coronary artery plaque.²³ Lipoprotein(a) concentration threshold of 120 nmol/L has been shown to discriminate well patients at risk of CHD.²⁴ Lipoprotein(a) is formed in the liver by covalent bonding of apolipoprotein(a) to apolipoprotein B-100 forming an LDL-like particle.²⁵ Standard LDL-C measurement assays give the composite concentration of cholesterol contained within LDL, Lp(a),

and intermediate-density lipoprotein particles; therefore, if Lp(a) concentration is not determined in a patient, this can lead to an inaccurate diagnosis of FH in those with high Lp(a) concentrations, when diagnosis is mainly based on LDL-C levels, without genetic confirmation. The mean genetic score for a previously computed Lp(a) PRS²⁶ has been recently shown to be significantly higher in FH variant-negative individuals in comparison with monogenic FH, suggesting imprecise diagnosis in about a quarter of FH variant-negative patients from Austria.²⁷

The 100 000 Genomes Project (100KGP) was launched in the UK in 2013 to investigate the role of whole genome sequencing (WGS) in the National Healthcare Service (NHS) setting.²⁸ A cohort of FH patients has been enrolled as part of the Rare Diseases Cardiovascular domain. In this study, we investigated genetic associations with hypercholesterolaemia in 100KGP FH participants, using whole genome information of rare and common variants computed into relevant PRSs, to facilitate precision diagnosis.

Methods

100 000 Genomes Project

The 100KGP, launched in the UK in 2013, generated WGS data to provide molecular diagnosis to participants with rare diseases and cancers.²⁸ The 100KGP was approved by East of England–Cambridge Central Research Ethics Committee ref:20/EE/0035. Only participants who provided written informed consent for their data to be used for research were included in the analyses. The current project (RR123) has been approved by the Genomics England Clinical Interpretation Partnership (GeCIP) cardiovascular domain committee.

100 000 Genomes Project familial hypercholesterolaemia cohort

A cohort of 467 probands and 69 affected relatives were identified as having PFH or DFH according to the clinical Simon–Broome diagnostic criteria, as recommended by the National Institute for Health and Care Excellence.¹ Patients presenting with LDL-C > 4.9 mmol/L and family history of myocardial infarction or severe hypercholesterolaemia were included in the Rare Diseases: Cardiovascular domain of the 100KGP as FH patients. All recruiting clinicians were affiliated with one of the Genomic Medical Centres, and all were experts in the field of lipidology with extensive experience in the identification and management of individuals with FH. Study participants were assigned Human Phenotype Ontology (HPO) terms at the time of recruitment to 100KGP. The most commonly assigned HPO terms to the FH cohort participants are summarized in [Supplementary material online, Table S1](#). The participants had not received a genetic diagnosis after going through usual care provided by the NHS. This is because standard genetic diagnostic tests came back negative, or none were available at the time of recruitment to 100KGP. Data on whether or not a recruited individual had had a genetic test was not collected.

Whole genome sequencing data

The DNA sample library preparation was done using the Illumina TruSeq DNA polymerase chain reaction (PCR)–free assay, and WGS was performed on a HiSeq 2500 sequencing platform, as previously described,²⁸ with more details shown in the [Supplementary Methods](#). An aggregate multi-sample VCF (AggV2) was generated by Genomics England (<https://re-docs.genomicsengland.co.uk/aggv2/>), which comprised variant call data for 78 195 germline genomes aligned to human genome GRCh38. Details about the Genomics England WGS data structure and site quality control have been published.²⁸ Definition of the PASS variant quality is shown in [Supplementary material online, Table S2](#). Variants were annotated using the Ensembl Variant Effect Predictor (VEP v99). All further analyses were performed within the Genomics England Research Environment.

Genetic ancestry

Genetic ancestry of 100KGP participants was estimated using principal component analysis (PCA) as described in the [Supplementary Methods](#).

Rare variant analysis

Sequencing data for Tier 1 (*LDLR*, *APOB*, *PCSK9*, *APOE*, and *LDLRAP1*) and Tier 2 (*ABCG5*, *ABCG8*, *CYP27A1*, and *LIPA*) FH genes were extracted from the AggV2 data file (genes' coordinates are shown in [Supplementary material online, Table S3](#)). Variants, with PASS quality, were filtered by gnomAD v3 minor allele frequency (MAF) < 0.001 (maximum across the ancestry subgroups), which is higher than the most common single FH-causing variant (the *APOB* p.Arg3527Gln, global gnomAD MAF = 0.0004), to remove likely benign variants. Filtered variants were interpreted using the American College of Medical Genetics and Genomics criteria.²⁹ Adapted criteria as defined by the ClinGen consortium were used for variants in *LDLR*.³⁰

Structural variant analysis

Structural variant (SV) calls were generated using MANTA. Putative copy number variations and SVs were intersected with the coordinates of the GRCh38 genome, using bedtools (v2.19.1). Variants overlapping with the *LDLR* gene sequence were analysed.

Genome wide association analysis

An aggregate file (AggV2) of all variants from the data release version 10 of the 100KGP, which was built on 78 195 whole germline genomes, was used for genome-wide association study (GWAS). The cohort selection is summarized in [Supplementary material online, Figure S1](#). Cases included unrelated FH variant-negative participants ($n = 443$). Controls ($n = 77\,275$) were selected after excluding participants who were assigned one or more HPO terms related to FH phenotype (hypercholesterolaemia, hyperlipidaemia, hyperlipoproteinaemia, increased LDL-C concentration, myocardial infarction, and premature coronary artery atherosclerosis). Genome-wide association study was performed using Scalable and Accurate Implementation of GEneralized mixed model (SAIGE) in a case/control manner (binary trait). The covariates were age, age², sex, age \times sex, and the 10 first principal components. SAIGE automatically accounts for sample relatedness and case–control imbalances. Further details on variant QC are shown in the [Supplementary Methods](#).

Polygenic risk score analysis: LDL polygenic risk score

We calculated and applied a published and validated polygenic risk score for LDL-C (LDL PRS) (<http://www.pgscatalog.org/>, PGS Catalog ID PGS000115)¹⁵ for FH mutation-positive cases, FH variant-negative cases, and controls. LDL PRS comprised 223 single-nucleotide variants and was developed from a GWAS (GWAS Catalog ID: GCST006612) consisting of 297 626 individuals (72.4% European, 19.3% African, and 8.3% Hispanic or Latin American).¹⁵ The cohort selection is summarized in [Supplementary material online, Figure S2A](#). Genotype data for calculating LDL PRS were taken from 100KGP WGS data using BCFtools v1.11-GCC-8.30. For each sample, LDL PRS was calculated by summing the multiple of each variants' inherited alleles by its effect size. Deciles of the LDL PRS distribution were calculated based on the 100KGP control samples ($n = 77\,660$). LDL PRS over the 8th decile or the top quintile was considered as high.

Polygenic risk score analysis: lipoprotein(a) polygenic risk score

A previously published locus-based polygenic risk score for Lp(a) (Lp(a) PRS) (<http://www.pgscatalog.org/>, PGS Catalog ID PGS000667)²⁶ was applied and calculated for FH mutation-positive cases, FH variant-negative cases, and controls. Lipoprotein(a) PRS comprised of 43 single-nucleotide variants and was developed from a GWAS consisting of 48 333 individuals

(75% European, 8.3% African, 8.3% East Asian, and 8.3% South Asian). The cohort selection is summarized in [Supplementary material online, Figure S2B](#). Genotype data for calculating Lp(a) PRS was taken from 100KGP WGS data using BCFtools v1.11-GCC-8.30. For each sample, Lp(a) PRS was calculated by summing the multiple of each variants' inherited alleles by its effect size. Lipoprotein(a) PRS higher than 120 (corresponding to >120 nmol/L) was considered as high.

Statistical analyses

Data was analysed using R v4.0.2. For comparison of LDL PRS case groups vs. control where data were normally distributed, data were analysed using an unpaired t-test. Comparison of Lp(a) PRS case groups vs. control where data were not normally distributed, reflecting the distribution of Lp(a) concentrations in a general population, was analysed using a Kruskal–Wallis *H* test. Statistical significance was considered at $P < 0.05$.

Results

Genetic ancestry structure

The majority of FH participants (84.86%) were of European genetic ancestry, followed by South Asian (5.42%), African (1.68%), East Asian (1.31%) and Admixed American (0.37%), with 6.36% being unassigned. Plotted PCs are shown in [Supplementary material online, Figure S3](#).

Likely pathogenic and pathogenic variants in Tier1 familial hypercholesterolaemia genes

Rare variant analysis of single-nucleotide variants and small deletions and insertions in Tier 1 FH genes identified 40 different likely pathogenic or pathogenic *LDLR* variants (see [Supplementary material online, Table S4](#)). These were found in 49 probands. Two *APOB* FH variants, including the single most common FH cause p.Arg3527Gln and the rare p.Arg3527Trp affecting the same amino acid, were found in six probands (see [Supplementary material online, Table S4](#)). The *APOE* gene was analysed for the previously published³ single FH variant p.Leu167del, which was found in 10 probands (see [Supplementary material online, Table S4](#)). The variant was significantly more frequent in the FH cohort when compared to non-FH participants [MAF in cases = 0.01 (10 out of 467 individuals) vs. in controls = 8×10^{-5} (12 out of 77 275 individuals), Fisher's exact test $P < 3.54 \times 10^{-17}$]. Taken together, variants in Tier1 genes explained 13.9% ($n = 65$) of the index FH cases ($n = 467$) and 15.5% ($n = 83$) of the complete FH cohort ($n = 536$), which includes affected relatives.

Eight *LDLR* variants classified as variants of uncertain significance (VUS) were found in an additional eight probands (see [Supplementary material online, Table S5](#)). Four variants in the *APOB* gene were also classified as VUS (see [Supplementary material online, Table S5](#)). The 'likely benign' *APOB* variant p.Pro994Leu was significantly enriched in the FH cohort (8 out of 467 individuals) in comparison with the non-FH 100KGP controls (155 out of 77 275 individuals) (MAF in cases = 0.009 vs. 0.001 in controls, Fisher's exact test $P = 7.53 \times 10^{-6}$), which suggest that it is a disease-modifying variant.

Structural variants in *LDLR*

Analysis of large structural variants and copy number variants in the *LDLR* gene identified four variants predicted to have a pathogenic effect (see [Supplementary material online, Table S6](#)). These were identified in four FH probands and five relatives, explaining 0.86% of the index

cases and 1.68% of the whole cohort. Whole genome sequencing data enabled the mapping of the variants' break points, as illustrated in [Supplementary material online, Figure S4](#).

Rare variants in Tier 2 genes

Variants in Tier 2 genes were filtered initially using a MAF < 0.001 cut-off. The next level of filtering included selecting variants most likely to affect function, i.e. missense, splicing, and loss-of-function variants. Several variants were identified (see [Supplementary material online, Table S7](#)); however, since most of the Tier 2 genes have been shown to be associated with the recessive pattern of FH inheritance, only the homozygous p.Trp361Ter variant, located in the *ABCG8* gene, found in one proband, was likely to explain the FH-like phenotype. Homozygote or compound heterozygote variants in *ABCG8* can cause sitosterolaemia³¹; however, we were unable to confirm levels of plant sterols in the patient.

Genome wide association study

Genome-wide association study analysis was performed using AggV2 WGS data containing 8 913 388 PASS quality and MAF > 0.1% variants from 433 participants without a likely pathogenic or pathogenic FH variant identified in Tier 1 and 2 genes (i.e. FH cases) and 77 275 100KGP control participants. Carriers of VUS were included in the analysis as FH variant-negative participants. The quantile-quantile plot of the GWAS *P* values is shown in [Supplementary material online, Figure S5](#), with $\lambda = 1.14$. The analysis identified one genome-wide significant signal at the *LPA* locus ($P < 1 \times 10^{-8}$) ([Figure 1](#)). The top associated SNP, rs140570886, has been previously associated with coronary artery disease,^{32,33} Lp(a),³⁴ and LDL-C.³⁵ The variant was found in 31 unrelated FH variant-negative participants, of whom two were homozygous. In addition, two probands and two relatives of mutation-positive FH participants were found to be heterozygote for the variant (no homozygotes were found). There was no association of the rs140570886 SNP with mutation-positive FH participants when compared with controls (MAF = 0.0163 vs. controls 0.0138, $P = 0.92$) (see LocusZoom plot in [Supplementary material online, Figure S6](#)). Therefore, 6.5% of the 100KGP FH cohort (or 7.0% of the FH variant-negative cohort) had at least one copy of the GWAS *LPA* variant. Summary statistics of the GWAS results are available via the link in [Supplementary Results](#).

LDL polygenic risk score

Calculation of the PRS for LDL-C in controls, monogenic FH, and FH variant-negative individuals was performed using WGS-derived genotype data. The LDL PRS was distributed normally among all three groups ([Figure 2](#)). The mean LDL PRS was highest in FH variant-negative participants (mean = 0.742, SD = ± 0.252), followed by monogenic FH (mean = 0.705, SD = ± 0.253), and when compared with controls (mean = 0.592, SD = ± 0.272), both were significantly higher ($P < 2.2 \times 10^{-16}$ and $P < 4.52 \times 10^{-5}$, respectively). The difference in LDL PRS between FH variant-negative and positive participants was not significant ($P < 0.64$). Considering the top quintile of the LDL PRS distribution as high, based on the controls (see [Supplementary material online, Table S8](#)), 161 (36.4%) FH variant-negative individuals had high LDL PRS > 0.823 compared to 15 448 controls (20.0%) (Fisher's exact test, $P = 1.38 \times 10^{-15}$).

Lipoprotein(a) polygenic risk score

Calculation of the Lp(a) PRS in controls, monogenic FH, and FH variant-negative individuals was performed using WGS-derived genotype data. Lipoprotein(a) PRS was found to be non-normally distributed among

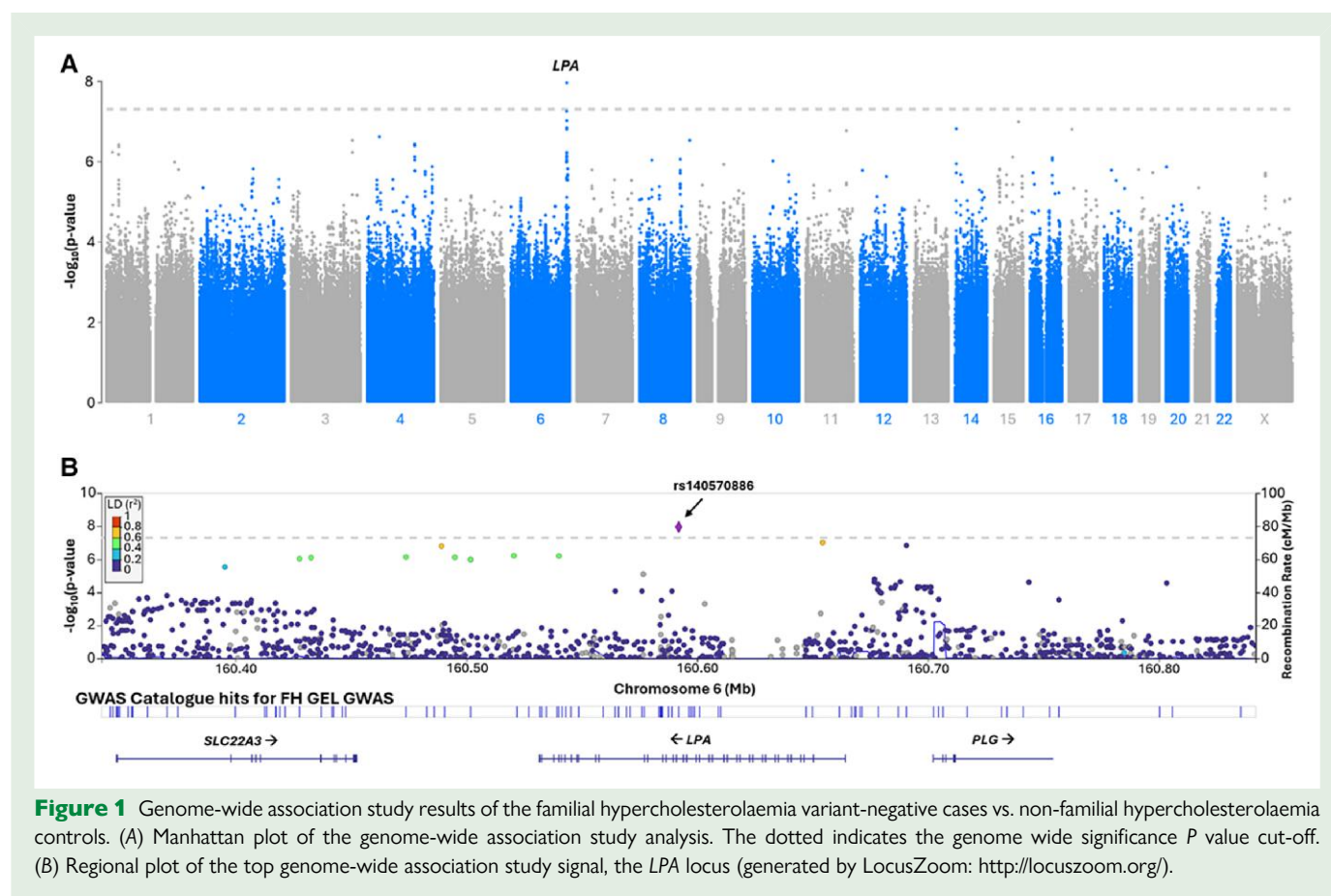


Figure 1 Genome-wide association study results of the familial hypercholesterolaemia variant-negative cases vs. non-familial hypercholesterolaemia controls. (A) Manhattan plot of the genome-wide association study analysis. The dotted indicates the genome wide significance P value cut-off. (B) Regional plot of the top genome-wide association study signal, the *LPA* locus (generated by LocusZoom: <http://locuszoom.org/>).

all three groups (Figure 3), reflecting the usual distribution of Lp(a) concentrations. We observed differences in the Lp(a) PRS distributions between the major genetic ancestry groups of 100KGP participants (see [Supplementary material online, Figure S7](#)), however applied a universal threshold of 120 nmol/L as high to all participants, since elevated Lp(a) has been strongly associated with CHD risk among multiple ancestry groups.³⁶ The Lp(a) PRS was significantly higher in both FH variant-negative (median = 24.4, IQR = 103) and monogenic FH (median = 30.1, IQR = 103) participants when compared with controls (median = 17.6, IQR = 56.3; $P < 6.82 \times 10^{-7}$ and $P < 5.02 \times 10^{-4}$, respectively). There was no significant difference of Lp(a) PRS between FH monogenic and FH variant-negative participants ($P = 0.34$). Deciles of the Lp(a) PRS distribution in the control cohort are shown in [Supplementary material online, Table S9](#). Of the FH variant-negative cases, 83 (15.5%) had Lp(a) PRS > 120 nmol/L, while among the 100KGP controls, 8206 (10.6%) individuals had elevated Lp(a) PRS. The median Lp(a) PRS for FH variant-negative individuals with high Lp(a) PRS (median = 171, IQR = 55.8) was significantly higher than the same subset of high Lp(a) PRS in controls (median = 143, IQR = 51.8) ($P < 3.16 \times 10^{-3}$).

In summary (see [Supplementary material online, Figure S8](#)), rare variant analysis of WGS data from 535 FH participants identified a monogenic cause of the disease in 93 (17.4%) individuals (74 probands and 19 relatives). We observed a high LDL PRS in 161 variant-negative FH participants (36.4% of variant-negative cases) indicating a polygenic component of hypercholesterolaemia, which in combination with other risk factors might predispose patients to high risk of CHD. High Lp(a) PRS predicting significantly elevated concentrations of Lp(a)

was found in 83 FH variant-negative participants (18.7%). Of the FH variant-negative cases, 31 (7.0%) had both high LDL and Lp(a) PRSs. Similarly, a polygenic component of hypercholesterolaemia and high Lp(a) was also observed in those with monogenic FH (see [Supplementary material online, Figure S8](#)). The majority of FH variant-negative cases ($n = 229$, 51.8%) remained without a plausible genetic diagnosis and are subject of a future investigation.

Discussion

Using WGS data, we examined the genetic causes of FH in 100KGP participants. Our analysis identified 48 different likely pathogenic and pathogenic variants in known FH Tier 1 and Tier 2 genes, providing genetic diagnosis of monogenic FH to 93 (17.4%) participants of the 100KGP FH cohort. Whole genome sequencing data analysis also allowed us to map the break points of the *LDLR* structural variants, which was limited using previous approaches, such as the multiplex ligation-dependent probe amplification assay or whole exome sequencing. This relatively high number of unique variants confirms the high heterogeneity of FH causes in the UK.

The FH variant detection rate in the 100KGP FH cohort was lower than the average 30–40% yield of FH genetic testing,⁹ but similar to the 17% detection rate in over 9500 samples sent for genetic testing to the UK Diagnostic laboratories in recent years.¹¹ This is because of the selective recruitment criteria that specified to include individuals who did not have access to a genetic diagnostic test or had negative diagnostic tests that did not include genome sequencing. The 100KGP genomic data are fully anonymized and unidentifiable; therefore, we were unable

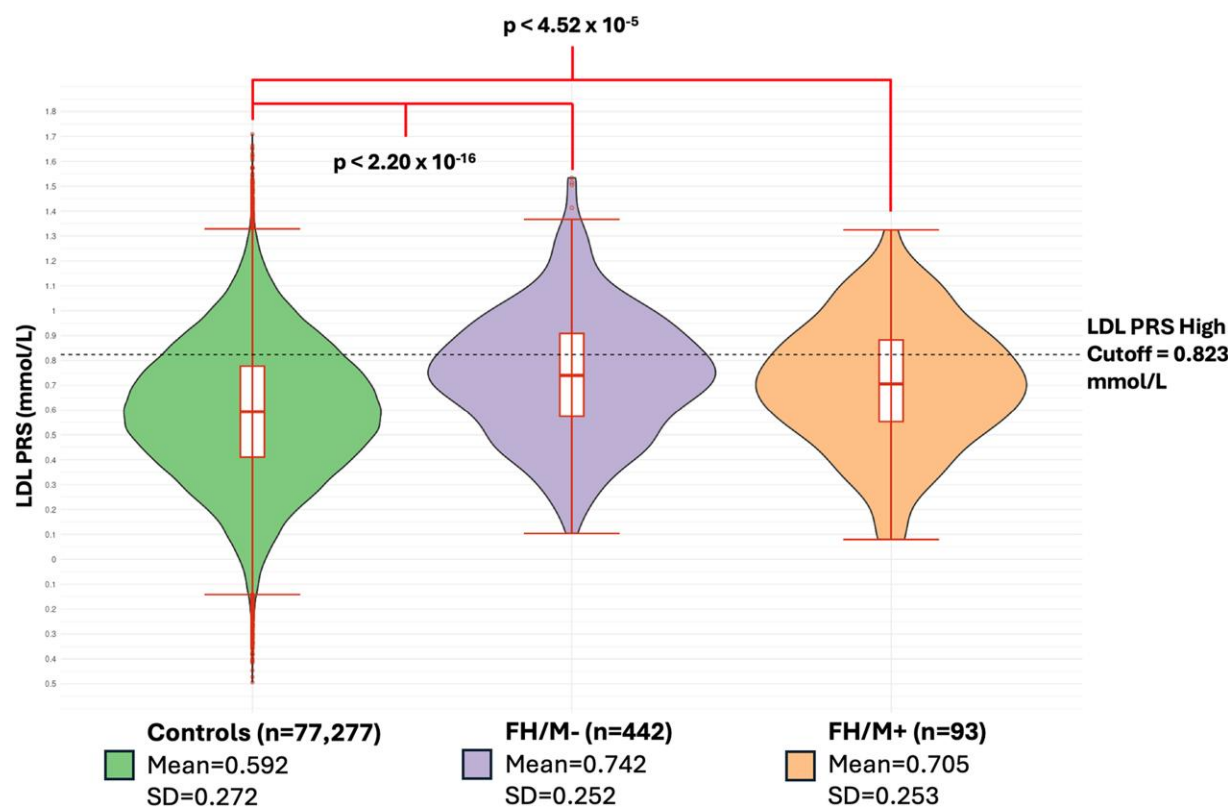


Figure 2 Violin plot of LDL polygenic risk score (PRS) in controls, familial hypercholesterolaemia variant-negative (FH/M-), and monogenic familial hypercholesterolaemia (FH/M+) individuals.

to define the reasons to explain why the 48 FH-causing variants were not initially detected, although this might reflect the patchy access to FH DNA testing between UK regions until recent times. We however, expected to find carriers of the *APOE* p.Leu167del,³⁷ which at the time of recruitment to 100KGP was not included in the standard FH genetic testing in the UK. The significantly higher prevalence of this variant in the clinical FH cohort compared with the non-FH control group confirms previous reports that carriers of this variant have an FH phenotype.³⁷

The availability of WGS data from non-FH participants of 100KGP (controls), which were generated and processed in the same manner as the FH cases, allowed us to perform case-control association study. Genome-wide association study between 443 FH variant-negative unrelated FH cases and 77 275 controls identified a significant signal at the *LPA* gene locus. Variants at the *LPA* locus have been shown to strongly determine concentration of Lp(a), which is an independent risk factor for CHD.²¹ Our study is the first to show a genetic association of *LPA* variants with variant-negative FH at a genome-wide significance level, after a comprehensive whole genome investigation and exclusion of monogenic causes of FH. Previous studies using a sum of up to two Lp(a)-raising SNPs similarly observed more frequent *LPA* risk alleles in the variant-negative FH when compared to monogenic FH.^{19,38} The much larger Lp(a) PRS applied in our study showed higher Lp(a) PRS in both variant-negative and variant-positive FH participants, highlighting that Lp(a) plays an important role in the FH phenotype. The enrichment of Lp(a)-increasing genotypes in cohorts of FH patients has been previously demonstrated as the likely result of ascertainment bias.

These findings underscore the importance of measuring both Lp(a) and LDL-C to more accurately define a patient's risk and to identify the responsible lipoprotein. This is particularly crucial for cardiovascular disease prevention, as recently highlighted,³⁹ especially given the numerous promising therapies specifically targeting Lp(a), including anti-sense oligonucleotides and RNA interference approaches, currently in clinical trials.⁴⁰ The *LPA* genotype information can aid the identification of individuals at risk of CHD caused by high Lp(a) concentrations and the future choice of treatment, therefore should be considered to be included in the genetic testing of individuals clinically diagnosed with FH.

We have previously observed a high LDL PRS in those who are clinically diagnosed as FH but do not have a pathogenic variant in one of the FH genes.¹⁶ Here we confirm the findings using a larger LDL PRS. Again, we observe that the mean LDL PRS is significantly higher also in the monogenic FH patient group. Additionally, although affecting a smaller number of individuals than high LDL PRS, a high Lp(a) PRS was observed at a similar proportion in both variant-positive and variant-negative FH cases. LDL and Lp(a) PRS analyses confirm the significant impact of polygenic background in FH, which is likely to explain the wide spectrum of LDL-C concentrations observed in FH variant carriers,⁷ and penetrance of likely pathogenic variants.⁴¹ It is also likely to contribute to the differences in age of onset of CHD in individuals with FH, even if they are carrying the identical FH-causing variant. Similar findings have been recently observed in the FH Canada National Registry, where a significantly higher mean score for both LDL and Lp(a) PRSs, as well as CAD PRS, was observed in FH

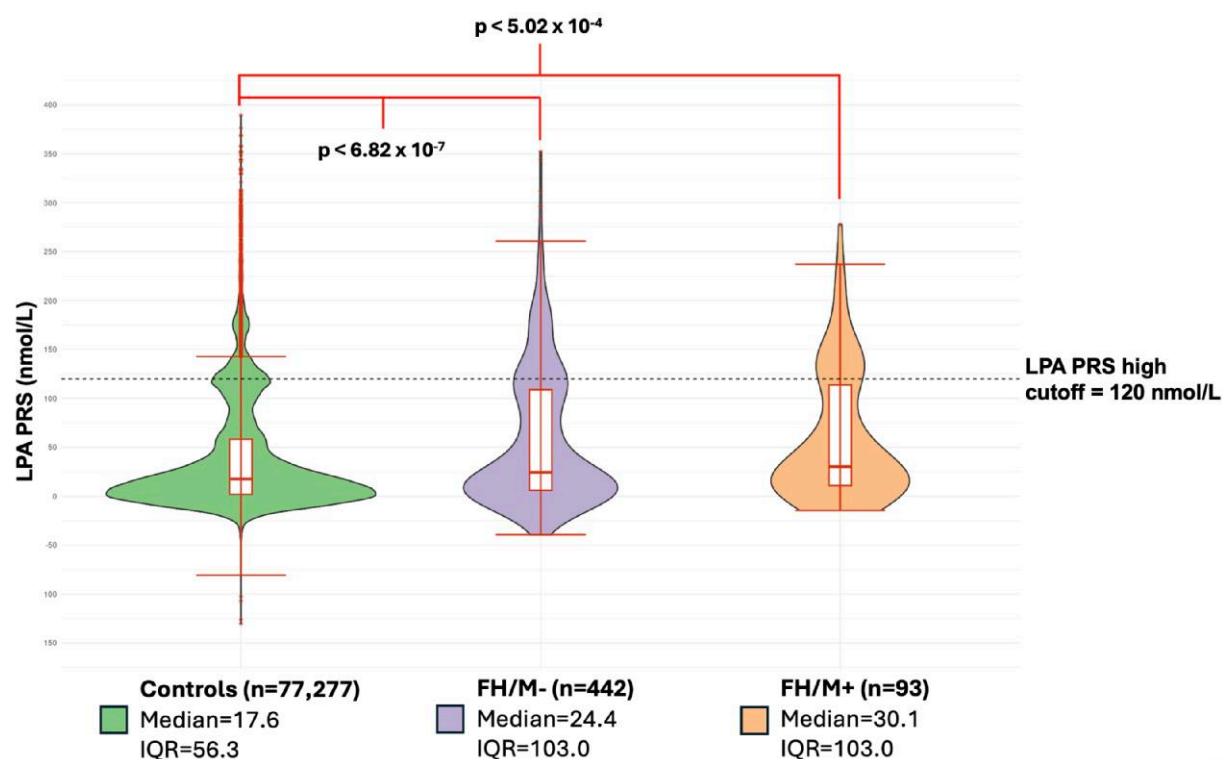


Figure 3 Violin plot of lipoprotein(a) polygenic risk score (PRS) in controls, familial hypercholesterolaemia variant-negative (FH/M-), and monogenic familial hypercholesterolaemia (FH/M+) individuals.

variant-negative individuals.⁴² The study, in addition to a previous report,⁴³ underlines the important role of polygenic background on the risk of atherosclerotic cardiovascular disease in monogenic FH, indicating the potential utility of PRSs in CAD risk prediction and the choice of intensity of therapy in FH patients.

Limitations

The main limitation of our study is the lack of the individual lipid profile data, apart from the HPO terms, on the non-FH comparison cohort or the FH participants, although all of the FH patients should have LDL-C concentrations > 4.9 mmol/L to conform with the Simon-Broome diagnostic criteria of PFH or DFH. All recruiting clinicians were experts in the field of lipidology with extensive experience in the identification and management of individuals with FH. While we cannot rule out the possible inclusion of a small proportion of individuals with hyperlipidaemia who do not fulfil the clinical diagnostic criteria for FH, their inclusion would only have a modest effect to reduce the power of the sample to detect a statistically significant difference between the FH and non-FH cohorts. Similarly, it is likely that a small proportion of the non-FH participants used as a comparison group will carry an FH-causing variant. Since the prevalence of this in the UK Biobank sample is $\sim 1/280$,⁴ this means that less than 0.4% of this cohort will have FH, which again would have only a modest impact on the power of the study. In terms of Lp(a) measurements, we predict that carriers of the significantly associated LPA SNPs and those with high Lp(a) PRS will have higher concentrations of Lp(a), since these are strongly genetically determined. This would be in line with recent findings.¹⁹

The genetic association analyses are constrained by the relatively small number of FH cases, which may mean that other variants influencing the phenotype of FH variant-negative participants have not been identified. While the majority of the FH cohort is of European genetic ancestry, 15% belong to other ancestries. Recent findings indicate that genetic ancestry is unlikely to affect monogenic FH detection.⁴ However, the PRS analyses, particularly the Lp(a) PRS, may not be optimal for participants of non-European ancestry.

Conclusions

In summary, our comprehensive analysis of both monogenic and genome wide polygenic causes of FH reveals a complex genetic architecture. The data presented here raise the issue of whether having high Lp(a) concentration should be included as a *cause* of FH. We believe that, in line with the original designation of Brown and Goldstein and others,⁴⁴ FH should be considered as a disorder of failure of clearance of LDL-C, and all genes currently designated as FH-causing encode for proteins directly or indirectly involved in this process. While having a high Lp(a) concentration is also associated with having high LDL-C and a high risk of early CHD, its cause is over production of the Lp(a) particle and this results in such an individual being incorrectly identified clinically as having FH, which by the above-mentioned definition should be reserved to FH mutation carriers. However, as our findings confirm, the clinical FH phenotype can be influenced by numerous genetic factors, and in some cases, clinical FH is mimicked by having inherited a high Lp(a) predisposition, which highlights the importance of measuring both Lp(a) and LDL-C for precision FH diagnosis. While

CHD risk will be reduced by LDL-C reduction in individuals with either high Lp(a) or high LDL-C, upcoming specific therapies to lower Lp(a) should be targeted to those who will benefit most, therefore having a precise genetic diagnosis will be clinically useful.

Supplementary material

Supplementary material is available at *European Journal of Preventive Cardiology*.

Acknowledgements

This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK, and the Medical Research Council have also funded research infrastructure.

Author contribution

M.B.: formal analysis, investigation, visualization, and writing original draft. A.M.P. and A.R.: formal analysis, methodology, and critical revision of the manuscript. S.E.H. and M.F.: conceptualization, design, and critical revision of the manuscript. A.M.P. and M.F.: supervision.

Funding

M.B. was funded by St George's University of London PhD studentship (internal funding 12729-58). S.E.H. and M.F. received additional support from the National Institute for Health Research University College London Hospitals Biomedical Research Centre scheme. S.E.H. and M.F. were supported by a grant from the British Heart Foundation (BHF grant PG 08/008).

Conflict of interest: S.E.H. is the medical director of a UCL Spin-off company (StoreGene) that offers genetic testing for cardiovascular risk including for FH. S.E.H. also reports payment for expert testimony from Verve Therapeutics. The remaining authors have no conflicts of interest related to this study.

Data availability

Genetic and phenotypic data for the 100KGP study participants are available through the Genomics England Research Environment via the application at <https://www.genomicsengland.co.uk/research/academic/join-research-network>. We provided summary statistics of the GWAS in an online data repository with the link shown in the supplement.

References

- Marks D, Thorogood M, Neil HAW, Humphries SE. A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. *Atherosclerosis* 2003;**168**:1–14.
- Sharifi M, Rakhit RD, Humphries SE, Nair D. Cardiovascular risk stratification in familial hypercholesterolaemia. *Heart* 2016;**102**:1003–1008.
- Akiyamen LE, Genest J, Shan SD, Reel RL, Albaum JM, Chu A, et al. Estimating the prevalence of heterozygous familial hypercholesterolaemia: a systematic review and meta-analysis. *BMJ Open* 2017;**7**:e016461.
- Gratton J, Humphries SE, Futema M. Prevalence of FH-causing variants and impact on LDL-C concentration in European, South Asian, and African Ancestry groups of the UK Biobank-brief report. *Arterioscler Thromb Vasc Biol* 2023;**43**:1737–1742.
- Khera AV, Won HH, Peloso GM, Lawson KS, Bartz TM, Deng X, et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J Am Coll Cardiol* 2016;**67**:2578–2589.
- Versmissen J, Oosterveer DM, Yazdanpanah M, Defesche JC, Basart DCG, Liem AH, et al. Efficacy of statins in familial hypercholesterolaemia: a long term cohort study. *BMJ* 2008;**337**:a2423.
- Gratton J, Humphries SE, Schmidt AF, Patel RS, Sofat R, Finan C, et al. Modelling a two-stage adult population screen for autosomal dominant familial hypercholesterolaemia: cross-sectional analysis within the UK Biobank. *BMJ Public Health* 2023;**1**:e000021.
- Pears R, Griffin M, Futema M, Humphries SE. Improving the cost-effectiveness equation of cascade testing for familial hypercholesterolaemia. *Curr Opin Lipidol* 2015;**26**:162–168.
- Taylor A, Wang D, Patel K, Whittall R, Wood G, Farrer M, et al. Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. *Clin Genet* 2010;**77**:572–580.
- Reeskamp LF, Tromp TR, Defesche JC, Grefhorst A, Stroes ESG, Hovingh GK, et al. Next-generation sequencing to confirm clinical familial hypercholesterolemia. *Eur J Prev Cardiol* 2021;**28**:875–883.
- Humphries SE, Challis R, Downes K, Howard E, Legerton T, Macanulty C, et al. How many FH genetic tests were performed by the UK genetic laboratory hubs in 2022? *Atheroscler Plus* 2023;**54**:S9.
- Futema M, Plagnol V, Li K, Whittall RA, Neil HAW, Seed M, et al. Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations. *J Med Genet* 2014;**51**:537–544.
- Zanoni P, Panteloglou G, Othman A, Haas JT, Meier R, Rimbart A, et al. Posttranscriptional regulation of the human LDL receptor by the U2-spliceosome. *Circ Res* 2022;**130**:80–95.
- Greene D, Pirri D, Frudd K, Sackey E, Al-Owain M, Giese APJ, et al. Genetic association analysis of 77,539 genomes reveals rare disease etiologies. *Nat Med* 2023;**29**:679–688.
- Trinder M, Paquette M, Cermakova L, Ban MR, Hegele RA, Baass A, et al. Polygenic contribution to low-density lipoprotein cholesterol levels and cardiovascular risk in monogenic familial hypercholesterolemia. *Circ Genom Precis Med* 2020;**13**:515–523.
- Talmud PJ, Shah S, Whittall R, Futema M, Howard P, Cooper JA, et al. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet* 2013;**381**:1293–1301.
- Thompson DJ, Wells D, Selzam S, Peneva I, Moore R, Sharp K, et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. *medRxiv* 22276246, <https://doi.org/10.1101/2022.06.16.22276246>, 16 June 2022, preprint: not peer reviewed.
- Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;**50**:1219–1224.
- Olmastoni E, Gazzotti M, Aversa M, Arca M, Tarugi P, Calandra S, et al. Lipoprotein(a) genotype influences the clinical diagnosis of familial hypercholesterolemia. *J Am Heart Assoc* 2023;**12**:e029223.
- de Boer LM, Hutten BA, Zwiderman AH, Wiegman A. Lipoprotein(a) levels in children with suspected familial hypercholesterolaemia: a cross-sectional study. *Eur Heart J* 2023;**44**:1421–1428.
- Tsimikas S. A test in context: lipoprotein(a): diagnosis, prognosis, controversies, and emerging therapies. *J Am Coll Cardiol* 2017;**69**:692–711.
- Nordstgaard BG, Langsted A. Lipoprotein (a) as a cause of cardiovascular disease: insights from epidemiology, genetics, and biology. *J Lipid Res* 2016;**57**:1953–1975.
- Clarke SL, Huang RDL, Hilliard AT, Levin MG, Sharma D, Thomson B, et al. Genetically predicted lipoprotein(a) associates with coronary artery plaque severity independent of low-density lipoprotein cholesterol. *Eur J Prev Cardiol* 2024.
- Nordstgaard BG, Chapman MJ, Ray K, Borén J, Andreotti F, Watts GF, et al. Lipoprotein(a) as a cardiovascular risk factor: current status. *Eur Heart J* 2010;**31**:2844–2853.
- McLean JW, Tomlinson JE, Kuang WJ, Eaton DL, Chen EY, Fless GM, et al. cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature* 1987;**330**:132–137.
- Trinder M, Uddin MM, Finneran P, Aragam KG, Natarajan P. Clinical utility of lipoprotein(a) and LPA genetic risk score in risk prediction of incident atherosclerotic cardiovascular disease. *JAMA Cardiol* 2021;**6**:287–295.
- Schwaninger G, Forer L, Ebenbichler C, Dieplinger H, Kronenberg F, Zschocke J, et al. Filling the gap: genetic risk assessment in hypercholesterolemia using LDL-C and LPA genetic scores. *Clin Genet* 2023;**104**:334–343.
- 100,000 Genomes Project Pilot Investigators. 100,000 genomes pilot on rare-disease diagnosis in health care—preliminary report. *N Engl J Med* 2021;**385**:1868–1880.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;**17**:405–424.
- Chora JR, Iacocca MA, Tichý L, Wand H, Kurtz CL, Zimmermann H, et al. The Clinical Genome Resource (ClinGen) familial hypercholesterolemia variant curation expert panel consensus guidelines for LDLR variant classification. *Genet Med* 2022;**24**:293–306.

31. Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J, et al. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science* 2000;**290**:1771–1775.
32. Zeng L, Moser S, Mirza-Schreiber N, Lamina C, Coassin S, Nelson CP, et al. Cis-epistasis at the LPA locus and risk of cardiovascular diseases. *Cardiovasc Res* 2022;**118**: 1088–1102.
33. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;**47**:1121–1130.
34. Mack S, Coassin S, Rueedi R, Yousri NA, Seppälä I, Gieger C, et al. A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. *J Lipid Res* 2017;**58**:1834–1844.
35. Graham SE, Clarke SL, Wu KHH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* 2021;**600**:675–679.
36. Paré G, Çaku A, McQueen M, Anand SS, Enas E, Clarke R, et al. Lipoprotein(a) levels and the risk of myocardial infarction among 7 ethnic groups. *Circulation* 2019;**139**: 1472–1482.
37. Marduel M, Ouguerram K, Serre V, Bonnefont-Rousselot D, Marques-Pinheiro A, Erik Berge K, et al. Description of a large family with autosomal dominant hypercholesterolemia associated with the APOE p.Leu167del mutation. *Hum Mutat* 2013;**34**:83–87.
38. Langsted A, Kamstrup PR, Benn M, Tybjaerg-Hansen A, Nordestgaard BG. High lipoprotein(a) as a possible cause of clinical familial hypercholesterolaemia: a prospective cohort study. *Lancet Diabetes Endocrinol* 2016;**4**:577–587.
39. Vuorio A, Raal FJ, Kovanen PT. Extension of heterozygous familial hypercholesterolemia treatment recommendations by including both low-density lipoprotein and lipoprotein(a) burden—a unique opportunity to improve patient prognosis. *Eur J Prev Cardiol* 2024.
40. Nicholls SJ. Therapeutic potential of lipoprotein(a) inhibitors. *Drugs* 2024;**84**:637–643.
41. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* 2020;**11**:3635.
42. Trinder M, Cermakova L, Ruel I, Baass A, Paquette M, Wang J, et al. Influence of polygenic background on the clinical presentation of familial hypercholesterolemia. *Arterioscler Thromb Vasc Biol* 2024;**44**:1683–1693.
43. Reeskamp LF, Shim I, Dron JS, Ibrahim S, Tromp TR, Fahed AC, et al. Polygenic background modifies risk of coronary artery disease among individuals with heterozygous familial hypercholesterolemia. *JACC Adv* 2023;**2**:100662.
44. Brown MS, Goldstein JL. Familial hypercholesterolemia: defective binding of lipoproteins to cultured fibroblasts associated with impaired regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity. *Proc Natl Acad Sci U S A* 1974;**71**: 788–792.