**Variants in *LPA* are associated with Familial Hypercholesterolaemia: whole genome sequencing analysis in the 100,000 Genomes Project.**

Martin Bird1, Antoine Rimbert2, Alan M Pittman3, Steve E Humphries4 and Marta Futema1,4

1Experimental Cardiology, Cardiovascular and Genomics Research Institute, School of Health & Medical Sciences, City St George's, University of London, 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

4Institute of Cardiovascular Science, Faculty of Population Health, University College London, London, UK

Corresponding author: [mfutema@sgul.ac.uk](mailto:mfutema@sgul.ac.uk)

**ABSTRACT**

**Background:**

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:**

WGS 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 (GWAS) 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.

**Results:**

An FH-causing variant was found in 17.4% of FH cases. GWAS identified the *LPA* gene locus being significantly associated (p<1x10-8). FH variant-negative participants had higher LDL and Lp(a) PRSs in comparison to the controls (p<1.0×10-16 and p<4.09×10-6, respectively). Similar associations were found in the monogenic FH with both LDL and Lp(a) PRSs being higher than in controls (p<4.03×10-4 and p<3.01x10-3, 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.

**Conclusions:**

This genome-wide analysis of monogenic and polygenic FH causes confirms a complex and heterogenous 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.

**Keywords:** Familial hypercholesterolaemia, whole genome sequencing, Lp(a), *LPA*, genome wide association study

**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’s called a "polygenic risk score." Polygenic score for LDL-C and Lp(a) are 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.

**INTRODUCTION**

Familial Hypercholesterolaemia (FH) is characterised by high concentrations of low-density lipoprotein 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). FH-causing variants in the aforementioned genes are found at a frequency of 1 in 250 individuals (95% CI, 1:345–1:192) in heterogenous 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 is 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 Netherlands (10) and was 17% in a UK survey of over 9,500 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 ≥80th percentile) has been shown to significantly increase the risk of premature atherosclerotic cardiovascular disease when compared to 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 to those with monogenic FH (19,20). Lp(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 CHD (22) also associated with the severity of coronary artery plaque (23). Lp(a) concentration threshold of 120 nmol/L has been shown to discriminate well patients at risk of CHD (24). Lp(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) PRS (26) has been recently shown to be significantly higher in FH variant-negative individuals in comparison to 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 United Kingdom 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 United Kingdom in 2013, generated whole genome sequencing (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.

**100KGP Familial Hypercholesterolaemia cohort**

A cohort of 467 probands and 69 affected relatives, identified as having Possible or Definite FH according to the clinical Simon Broome diagnostic criteria, as recommended by the National Institute for Health and Care Excellence (NICE) (1). Patients presenting with LDL-C >4.9mmol/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 HPO (Human Phenotype Ontology) terms at the time of recruitment to 100KGP. The most commonly assigned HPO terms to the FH cohort participants are summarised in **Supplementary 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** **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** **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 (ACMG) criteria (29). Adapted criteria as defined by the ClinGen consortium was used for variants in *LDLR* (30).

**Structural variants analysis**

Structural variant (SV) calls were generated using MANTA. Putative copy number variations (CNVs) 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 GWAS. The cohort selection is summarised in **Supplementary 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, hyperlipoproteinemia, increased LDL cholesterol concentration, myocardial infarction, premature coronary artery atherosclerosis). GWAS 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 × 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 section.

**Polygenic risk score analysis: LDL PRS**

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, 8.3% Hispanic or Latin American) (15). The cohort selection is summarised in **Supplementary** **Figure** **2A**. Genotype data for calculating LDL PRS was 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 was 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: Lp(a) PRS**

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. Lp(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, 8.3% South Asian). The cohort selection is summarised in **Supplementary** **Figure** **2B**. 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. Lp(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 versus control where data was normally distributed, data was analysed using an unpaired t-test. Comparison of Lp(a) PRS case groups versus control where data was 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 Figure S3**.

**Likely pathogenic and pathogenic variants in Tier1 FH 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 (**Supplementary** **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 (**Supplementary** **Table S4**). The *APOE* gene was analysed for the previously published (3) single FH variant p.Leu167del, which was found in 10 probands (**Supplementary** **Table S****4**). 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=8x10-5 [12 out of 77,275 individuals], Fisher’s exact test p<3.54x10-17). 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 (**Supplementary Table S5**). Four variants in the *APOB* gene were also classified as VUS (**Supplementary** **Table S5**). The ‘likely benign’ *APOB* variant p.Pro994Leu was significantly enriched in the FH cohort (8 out of 467 individuals) in comparison to 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=7.53x10-6), 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 (**Supplementary** **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. WGS data enabled the mapping of the variants’ break points, as illustrated in **Supplementary 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, loss-of-function variants. Several variants were identified (**Supplementary 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 sitosterolaemia (31), however we were unable to confirm levels of plant sterols in the patient.

**Genome Wide Association Study**

GWAS 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** **Figure S5,** with λ=1.14. The analysis identified one genome-wide significant signal at the *LPA* *locus* (p< 1x10-8) (**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 comparing to controls (MAF= 0.0163 vs. controls 0.0138, p=0.92) (see Locus Zoom plot in **Supplementary 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 SupplementaryResults.

**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 to controls (mean=0.592, SD=±0.272) both were significantly higher (p<2.2 x 10-16, p<4.52 x 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 (**Supplementary** **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 x10-15).

Lp(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. Lp(a) PRS was found to be non-normally distributed among 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 (**Supplementary Figure 7**), 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 to controls (median=17.6, IQR=56.3; p<6.82 ×10-7 and p<5.02 ×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** **Table S9**. Of the FH variant-negative cases 83 (15.5%) had Lp(a) PRS >120 nmol/L, while among the 100KGP controls, 8,206 (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 x 10-3).

In summary (**Supplementary 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 (**Fig.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 whole genome sequencing data we examined the genetic causes of FH in 100KGP participants. Our analysis identified 48 different likely pathogenic and pathogenic variants in known FH Tier1 and Tier2 genes, providing genetic diagnosis of monogenic FH to 93 (17.4%) participants of the 100KGP FH cohort. WGS 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 didn’t have access to a genetic diagnostic test or had negative diagnostic tests that did not include genome sequencing. The 100KGP genomic data are fully anonymised and unidentifiable, therefore we were unable 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 to 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. GWAS 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 -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 antisense 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) PRSs 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 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 > 4.9mmol/l to conform with the Simon Broome diagnostic criteria of Possible or Definite FH. 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 ~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 polygenic risk score (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.

**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.

**FUNDING**

MB was funded by St George’s University of London PhD studentship (internal funding 12729-58). SEH and MF received additional support from the National Institute for Health Research University College London Hospitals Biomedical Research Centre. SEH and MF were supported by a grant from the British Heart Foundation (BHF grant PG 08/008).

**CONFLICT OF INTEREST**

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

**AUTHORS’ CONTRIBUTIONS**

MB: formal analysis, investigation, visualisation, writing original draft. AP and AR: formal analysis, methodology, critical revision of the manuscript. SEH and MF: conceptualisation, design, critical revision of the manuscript. AP and MF: supervision.

**DATA AVAILABILITY STATEMENT**

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/join-a-gecip-domain>. We provided summary statistics of the GWAS in an online data repository with the link shown in the supplement.

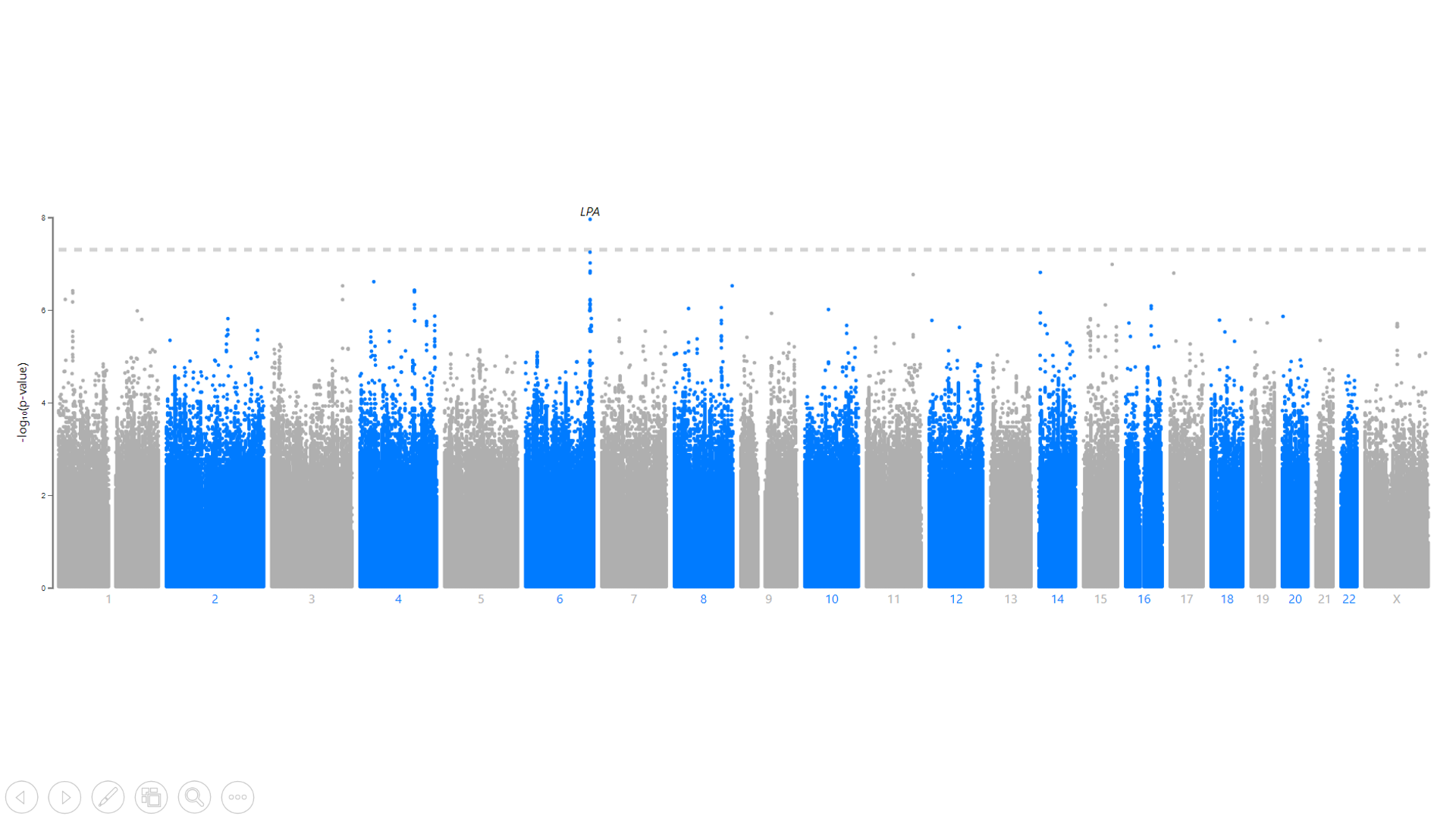
**FIGURE LEGENDS**

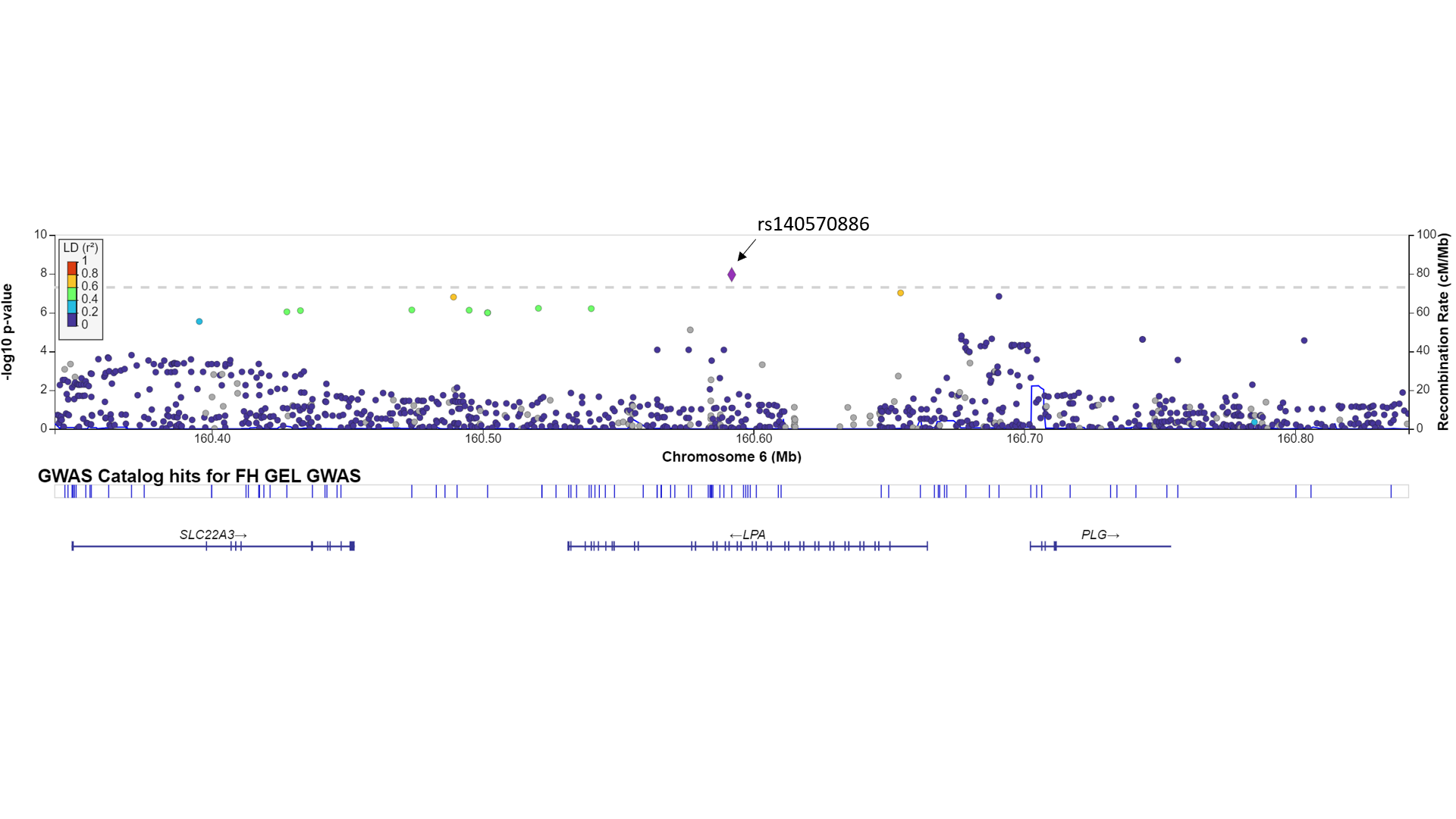
**Figure 1.** GWAS results of the FH variant-negative cases versus non-FH controls. **A.** Manhattan plot of the GWAS analysis. The dotted indicates the genome wide significance p value cut-off. **B.** Regional plot of the top GWAS signal, the *LPA* locus (generated by Locus Zoom http://locuszoom.org/).

**Figure 2.** Violin plot of LDL PRS in controls, monogenic FH and FH variant-negative individuals.

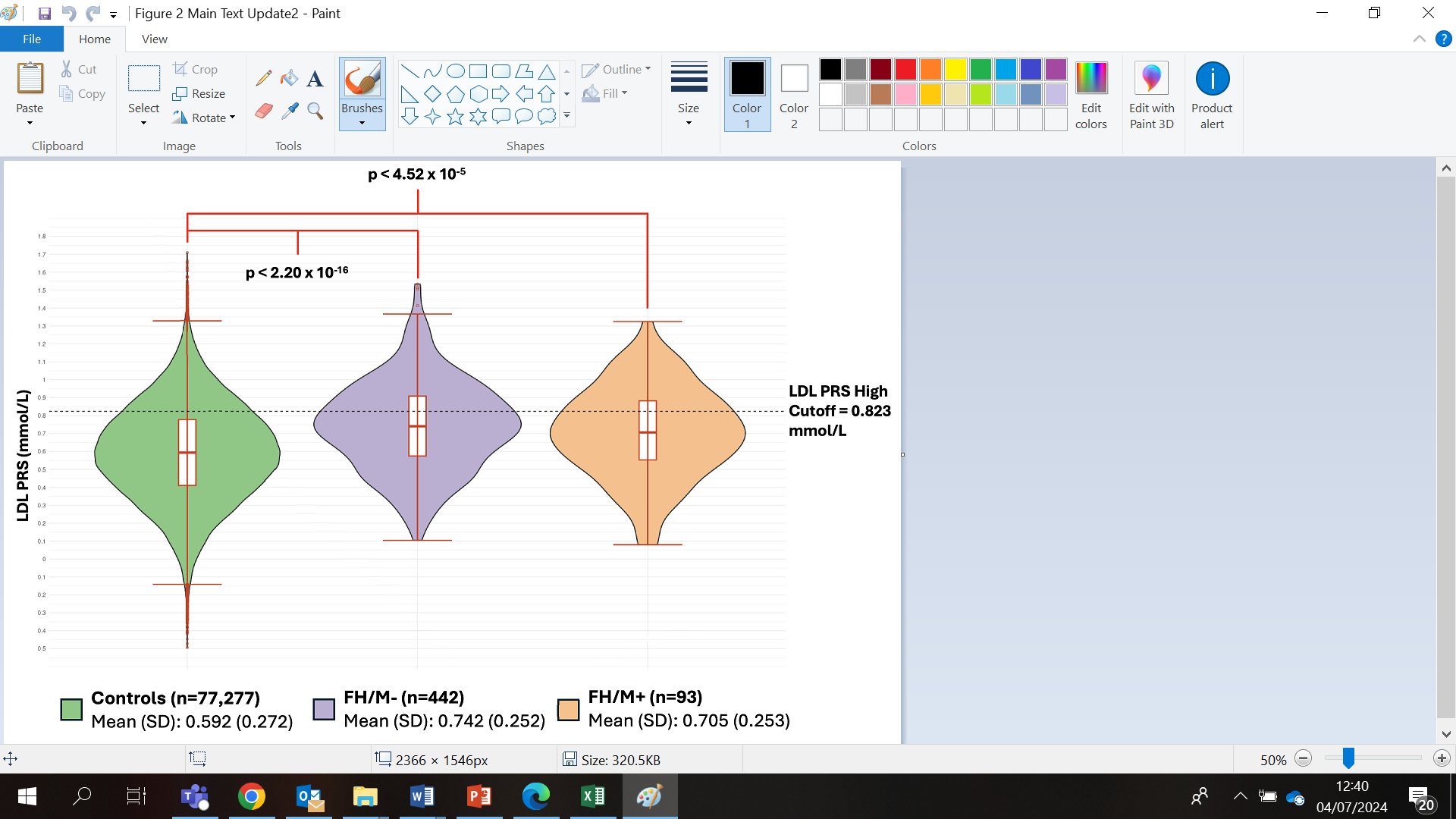
**Figure 3.** Violin plot of Lp(a) PRS in controls, monogenic FH and FH variant-negative individuals.

**Figure 1.**

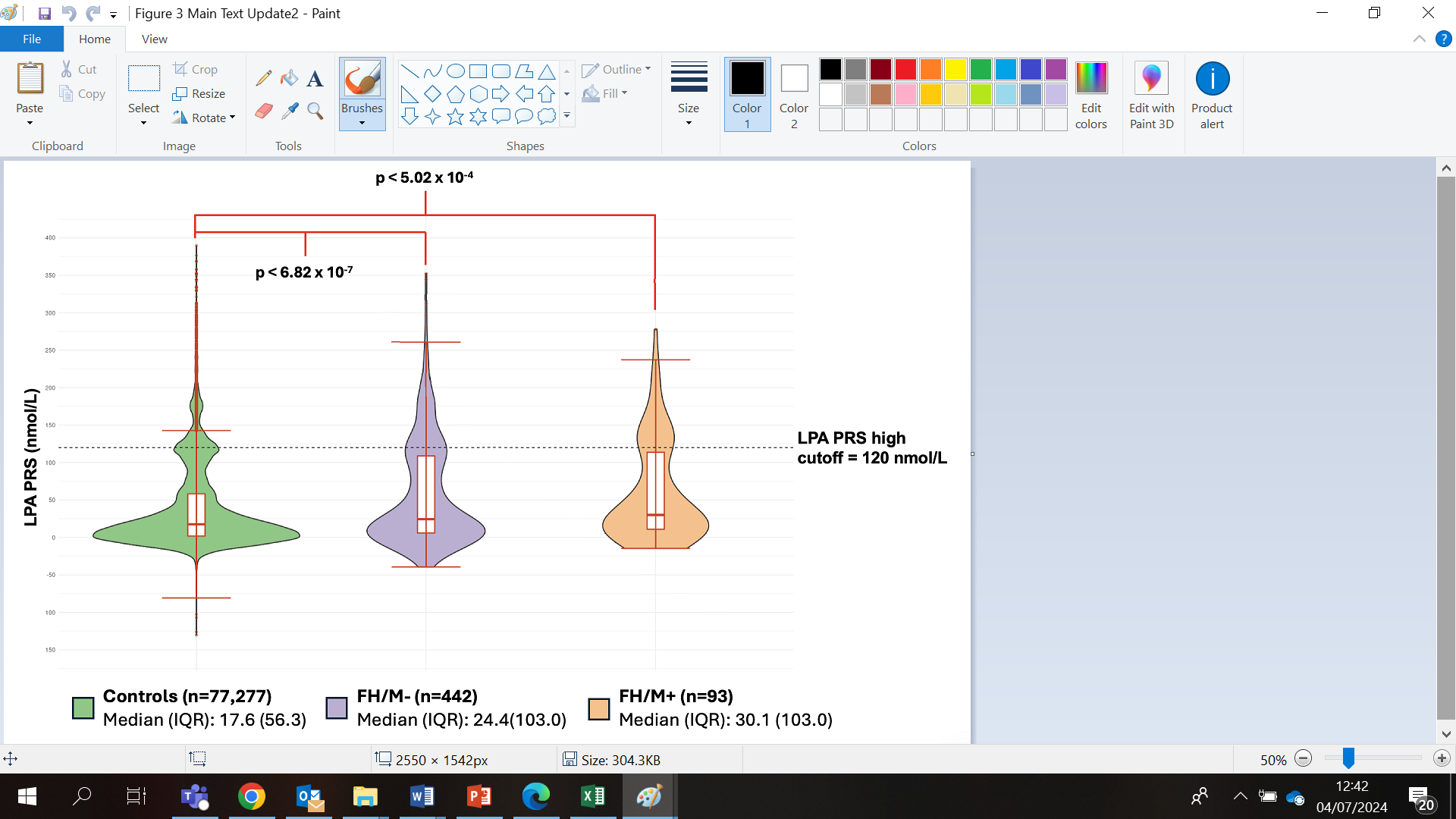
**A**. 

**B.** 

**Figure 2.**



**Figure 3.**



**REFERENCES**:

1. Marks D, Thorogood M, Neil HAW, Humphries SE. A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. Atherosclerosis. 2003 May;168(1):1–14.

2. Sharifi M, Rakhit RD, Humphries SE, Nair D. Cardiovascular risk stratification in familial hypercholesterolaemia. Heart. 2016 Jul 1;102(13):1003–8.

3. Akioyamen 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 Sep 1;7(9):e016461.

4. 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 Sep;43(9):1737–42.

5. Khera A V, 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 Jun 7;67(22):2578–89.

6. 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 Nov 11;337:a2423.

7. 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 Oct;1(1):e000021.

8. Pears R, Griffin M, Futema M, Humphries SE. Improving the cost-effectiveness equation of cascade testing for familial hypercholesterolaemia. Curr Opin Lipidol. 2015 Jun;26(3):162–8.

9. 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 Jun;77(6):572–80.

10. 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 Jul 23;28(8):875–83.

11. 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? Atherosclerosis Plus. 2023 Dec;54:S9.

12. 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 Aug;51(8):537–44.

13. Zanoni P, Panteloglou G, Othman A, Haas JT, Meier R, Rimbert A, et al. Posttranscriptional Regulation of the Human LDL Receptor by the U2-Spliceosome. Circ Res. 2022 Jan 7;130(1):80–95.

14. 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 Mar 16;29(3):679–88.

15. 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 Oct;13(5):515–23.

16. 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 Apr 13;381(9874):1293–301.

17. 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. Available from: https://doi.org/10.1101/2022.06.16.22276246

18. Khera A V, 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 Sep;50(9):1219–24.

19. Olmastroni E, Gazzotti M, Averna M, Arca M, Tarugi P, Calandra S, et al. Lipoprotein(a) Genotype Influences the Clinical Diagnosis of Familial Hypercholesterolemia. J Am Heart Assoc. 2023 May 16;12(10):e029223.

20. de Boer LM, Hutten BA, Zwinderman AH, Wiegman A. Lipoprotein(a) levels in children with suspected familial hypercholesterolaemia: a cross-sectional study. Eur Heart J. 2023 Apr 21;44(16):1421–8.

21. Tsimikas S. A Test in Context: Lipoprotein(a): Diagnosis, Prognosis, Controversies, and Emerging Therapies. J Am Coll Cardiol. 2017 Feb 14;69(6):692–711.

22. Nordestgaard BG, Langsted A. Lipoprotein (a) as a cause of cardiovascular disease: insights from epidemiology, genetics, and biology. J Lipid Res. 2016 Nov;57(11):1953–75.

23. 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 Aug 19;

24. Nordestgaard 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 Dec;31(23):2844–53.

25. 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. 330(6144):132–7.

26. 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 Mar 1;6(3):287.

27. 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 Sep;104(3):334–43.

28. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. New England Journal of Medicine. 2021 Nov 11;385(20):1868–80.

29. 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. Genetics in Medicine. 2015 May;17(5):405–24.

30. 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 Feb;24(2):293–306.

31. Berge KE, Tian H, Graf GA, Yu L, Grishin N V, Schultz J, et al. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science. 2000 Dec 1;290(5497):1771–5.

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 Mar 16;118(4):1088–102.

33. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015 Oct 7;47(10):1121–30.

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 Sep;58(9):1834–44.

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 Dec;600(7890):675–9.

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 Mar 19;139(12):1472–82.

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 Jan;34(1):83–7.

38. Langsted A, Kamstrup PR, Benn M, Tybjærg-Hansen A, Nordestgaard BG. High lipoprotein(a) as a possible cause of clinical familial hypercholesterolaemia: a prospective cohort study. Lancet Diabetes Endocrinol. 2016 Jul;4(7):577–87.

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 Sep 20;

40. Nicholls SJ. Therapeutic Potential of Lipoprotein(a) Inhibitors. Drugs. 2024 Jun;84(6):637–43.

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 Aug 20;11(1):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 Jul;44(7):1683–93.

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 Advances. 2023 Nov;2(9):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 Mar;71(3):788–92.