# **Supplementary Online Content**

Patel AP, Wang M, Fahed AC, et al. Association of rare pathogenic DNA variants for familial hypercholesterolemia, hereditary breast and ovarian cancer syndrome, and Lynch syndrome with disease risk in adults according to family history. *JAMA Netw Open*. 2020;3(4):e203959. doi:10.1001/jamanetworkopen.2020.3959

**eMethods.** Study Populations, Gene Sequencing, Variant Quality Control, and Variant **Classification** 

**eTable 1.** Adjustment of Lipid Levels in the UK Biobank Based on Self-report of Lipid-Lowering Medications

**eTable 2.** Variant Characteristics and Evidence in Support of Pathogenic or Likely Pathogenic Classifications

**eTable 3.** Baseline Characteristics of Familial Hypercholesterolemia Variant Carriers and Noncarriers

**eTable 4.** Baseline Characteristics of Women Hereditary Breast and Ovarian Cancer Syndrome Variant Carriers and Noncarriers

**eTable 5.** Baseline Characteristics of Lynch Syndrome Variant Carriers and Noncarriers **eTable 6.** Familial Hypercholesterolemia, Hazard Ratios for Each Component of the Composite End Point

**eTable 7.** Hereditary Breast and Ovarian Cancer Syndrome, Hazard Ratios for Primary **Cancers** 

**eTable 8.** Lynch Syndrome, Hazard Ratios for Primary Cancers

**eTable 9.** Prevalence and Clinical Importance of Pathogenic or Likely Pathogenic Variants for 3 Genomic Conditions Stratified by Gene

**eFigure 1.** Frequency of Carriers of Pathogenic or Likely Pathogenic Variants, for Each of 3 Genomic Conditions According to Gene

**eFigure 2.** Observed and Estimated Untreated Low-Density Lipoprotein Cholesterol Levels According to Familial Hypercholesterolemia Variant Status

**eFigure 3.** Age-Dependent Cumulative Probability of Disease According to Pathogenic or Likely Pathogenic Variant Carrier Status

**eFigure 4.** Predicted Risk of Disease at Ages 55, 65, and 75 According to Pathogenic or Likely Pathogenic Variant Carrier Status and Family History **eReferences.**

This supplementary material has been provided by the authors to give readers additional information about their work.

## **eMethods.**

### *Study populations*

The study population consisted of a subset of the UK biobank consisting of 49,738 participants who underwent exome sequencing at the Regeneron Genetics Center.1 Extensive clinical data including diagnosis of prevalent and incident atherosclerotic cardiovascular disease and cancers are available on all participants.2,3 Coronary artery disease was defined based on self-report of "heart attack/myocardial infarction", hospitalization records confirming a diagnosis of acute myocardial infarction or its acute complications, ischemic heart disease, coronary revascularization procedures (coronary artery bypass graft surgery or percutaneous angioplasty/stent placement), or death register indicating ischemic heart disease or myocardial infarction as a cause of death. Ischemic stroke was defined based on self-report of ischemic stroke, hospitalization records confirming a diagnosis of cerebral infarction due to thrombosis, cerebral atherosclerosis, cerebrovascular syndromes, and cerebrovascular stenoses, or death register indicating ischemic stroke as a cause of death. Breast cancer, ovarian cancer, colorectal cancer, uterine cancer, and cancers of the gastrointestinal tract, genitourinary tract, and skin were each defined based on self-report of the diagnosis, hospitalization records, cancer register data specifying type of cancer, and death register. For each of the three disease groups, the earliest date at which the diagnosis was ascertained was considered as the diagnosis date.

#### *Gene sequencing*

Whole-exome sequencing of 49,960 UK Biobank participants was performed at the Regeneron Genetics Center as previously described, $\frac{1}{2}$  and sequencing reads were aligned to the human reference genome build GRCh38 using the Burrows-Wheeler Aligner algorithm.4 Coverage exceeded 20X at 94.6% of sites on average. Variant calls through two separate pipelines, an "SPB pipeline" that used WeCall (GenomicsPLC) and GLnexus software and a functional equivalence (FE) pipeline that used GATK, were made available by the UK Biobank for 49,960 samples.<sup>1,5,6</sup> Variants from the FE pipeline that were also present in the SPB pipeline were included. 222 samples were excluded, for which there were no genotyping data available  $(n=51)$  or that failed additional sample quality control using genotyping data: heterozygous missingness outlier (n=112), putative sex chromosome aneuploidy  $(n=56)$  and discordance between reported and genetic sex  $(n=20)$ . No individuals within the group analyzed withdrew consent at the time of analysis. The remaining variants of 49,738 participants were carried forward for further analysis. PLINK formatted files were converted to VCF and a liftover was performed from GRCh38 to GRCh37.p13 by CrossMap (v0.3.3).<sup>7</sup>

### *Variant quality control*

Analysis was limited to the protein-coding regions and canonical splice sites of 9 genes for any of the three Tier 1 genomic conditions: familial hypercholesterolemia (*LDLR*, *APOB* and *PCSK9*), hereditary breast and ovarian cancer syndrome (*BRCA1* and *BRCA2*), and Lynch syndrome (*MLH1*, *MSH2*, *MSH6* and *PMS2*). Observed variants were filtered to a candidate list of variants that excludes synonymous variants or variants present at allele frequency of >=0.005 in any racial subpopulation of the gnomAD Genome Aggregation Database.<sup>8</sup> Additional variant quality control filters excluded variants that fall in low complexity regions, variants that fall in regions with segmental duplications, or variants that do not pass the threshold for the random forest algorithm of gnomAD.<sup>6,8</sup> The final variants were called both by the FE and SPB pipeline.

#### *Variant classification*

Candidate variants were filtered to select variants meeting clinical criteria of pathogenicity (pathogenic or likely pathogenic) based on American College of Medical Genetics and Genomics (ACMG)/Association of Molecular Pathology (AMP) criteria,<sup>9</sup> by an American Board of Genetics and Genomics (AMBGG)-certified clinical geneticist, blinded to the phenotype of the participants, at the Partners HealthCare Laboratory of Molecular Medicine (Boston, MA). In summary, the ACMG/AMP criteria for classifying pathogenic variants look at the effect of the variant on the gene, the previous reports of pathogenicity of the variant, functional studies supporting the damaging effect of the gene, and the prevalence of the variant in cohorts of cases with the disease and controls.<sup>9</sup> Findings of the ClinGen and ClinVar expert panels were all incorporated into interpretation of each variant. Variants with limited available data were independently reviewed by multiple geneticists.<sup>10,11</sup>



**eTable 1**. Adjustment of Lipid Levels in the UK Biobank Based on Self-report of Lipid-Lowering Medications

In participants who are on lipid-lowering medications, lipid levels were adjusted depending on the type of lipid-lowering medication intake based on prior reports of effect size for each medication type from the literature.<sup>12–17</sup> For example, in the case of statin intake, total cholesterol was divided by 0.8, lower density cholesterol by 0.7 and triglycerides by 0.85. \* When the lipid-lowering medication was not specified, it was assumed a statin.

# **eTable 2.** Variant Characteristics and Evidence in Support of Pathogenic or Likely Pathogenic Classifications











© 2020 Patel AP et al. *JAMA Network Open*.































© 2020 Patel AP et al. *JAMA Network Open*.

















© 2020 Patel AP et al. *JAMA Network Open*.











**eTable 3.** Baseline Characteristics of Familial Hypercholesterolemia Variant Carriers and **Noncarriers** 

Characteristics of carriers and noncarriers of pathogenic or likely pathogenic mutations in genes associated with familial hypercholesterolemia. Comparison of background variables between carriers of the variants of interest and noncarriers was performed with the analysis of variance (ANOVA) for continuous variables, chi-squared test for categorical variables, and the Kruskal-Wallis test for non-normally distributed measurement variables; values represent mean (standard deviation), n (% of individuals), or median [interquartile range], respectively. SI conversion factor: To convert cholesterol to mmol/L, multiply values by 0.0259. To convert triglyceride levels to mmol/l, multiple values by 0.01129. Family history is defined as known history of heart disease in a first degree relative. Severe hypercholesterolemia is defined as estimated untreated LDL-C greater than or equal to 190 mg/dL.



**eTable 4.** Baseline Characteristics of Women Hereditary Breast and Ovarian Cancer Syndrome Variant Carriers and Noncarriers

Characteristics of carriers and noncarriers of pathogenic or likely pathogenic mutations in genes associated with hereditary breast and ovarian cancer syndrome in women. Comparison of background variables between carriers of the variants of interest and noncarriers was performed with the analysis of variance (ANOVA) for continuous variables, chi-squared test for categorical variables, and the Kruskal-Wallis test for non-normally distributed measurement variables; values represent mean (standard deviation), n (% of individuals), or median [interquartile range], respectively. Family history is defined as known history of breast cancer in a first degree relative.



**eTable 5.** Baseline Characteristics of Lynch Syndrome Variant Carriers and Noncarriers

Characteristics of carriers and noncarriers of pathogenic or likely pathogenic mutations in genes associated with Lynch syndrome. Comparison of background variables between carriers of the variants of interest and noncarriers was performed with the analysis of variance (ANOVA) for continuous variables, chi-squared test for categorical variables, and the Kruskal-Wallis test for non-normally distributed measurement variables; values represent mean (standard deviation), n (% of individuals), or median [interquartile range], respectively. Family history is defined as known history of bowel cancer in a first degree relative.

## **eTable 6.** Familial Hypercholesterolemia, Hazard Ratios for Each Component of the Composite End Point



Hazard ratios with corresponding 95% confidence intervals and P values for primary disease endpoints of familial hypercholesterolemia, comparing carriers to noncarriers calculated using Cox proportional-hazards model with covariates of enrollment age, sex, and genetic ancestry as quantified by the first four genetic principal components.18



**eTable 7.** Hereditary Breast and Ovarian Cancer Syndrome, Hazard Ratios for Primary Cancers

Hazard ratios with corresponding 95% confidence intervals and P values for primary disease endpoints related to hereditary breast and ovarian cancer syndrome, comparing carriers to noncarriers calculated using Cox proportional-hazards model with covariates of enrollment age, sex, and genetic ancestry as quantified by the first four genetic principal components.





Hazard ratios with corresponding 95% confidence intervals and P values for primary disease endpoints related to Lynch syndrome, comparing carriers to noncarriers calculated using Cox proportional-hazards model with covariates of enrollment age, sex, and genetic ancestry as quantified by the first four genetic principal components.

**eTable 9.** Prevalence and Clinical Importance of Pathogenic or Likely Pathogenic Variants for 3 Genomic Conditions Stratified by Gene



Hazard ratios with corresponding 95% confidence intervals and P values for disease comparing carriers to noncarriers calculated using Cox proportional-hazards model with covariates of enrollment age, sex, and genetic ancestry as quantified by the first four genetic principal components, for all individuals with genomic condition and stratified by individual gene component. Heterogeneity in risk based on gene was observed among the Lynch syndrome variants ( $p=0.006$ ), but not in familial hypercholesterolemia ( $p=0.25$ ) or hereditary breast and ovarian cancer ( $p=0.99$ ).



**eFigure 1.** Frequency of Carriers of Pathogenic or Likely Pathogenic Variants, for Each of 3 Genomic Conditions According to Gene

Number of individuals carrying pathogenic or likely pathogenic mutations in *APOB, LDLR, PCSK9, BRCA1, BRCA2, MLH1, MSH2, MSH6,* and *PMS2*, with percentages out of 49,738 individuals in UK Biobank with exome sequencing data.

**eFigure 2.** Observed and Estimated Untreated Low-Density Lipoprotein Cholesterol Levels According to Familial Hypercholesterolemia Variant Status



A: Distribution of low-density lipoprotein (LDL) cholesterol among carriers (mean 161 mg/dL) and noncarriers (mean 137 mg/dL) of a familial hypercholesterolemia variant. B: Distribution of estimated untreated LDL cholesterol among carriers (mean 198 mg/dL) and noncarriers (mean 145 mg/dL) of a familial hypercholesterolemia variant, based on adjustment for average effect of any cholesterol-lowering therapies reported in each individual (eTable 1 in Supplement).

**eFigure 3.** Age-Dependent Cumulative Probability of Disease According to Pathogenic or Likely Pathogenic Variant Carrier Status



Age-dependent cumulative probability of disease according to pathogenic or likely pathogenic variant carrier status. Age-dependent cumulative probability estimated using Cox proportional-hazards model, standardized to the average of the first four genetic principal components and sex and stratified by respective variant carrier status for familial hypercholesterolemia (coronary artery disease, stroke, or peripheral artery disease, A, B), hereditary breast and ovarian cancer (C, D), and Lynch syndrome (colorectal and uterine cancer, E, F) in women and men, respectively.

**eFigure 4.** Predicted Risk of Disease at ages 55, 65, and 75 According to Pathogenic or Likely Pathogenic Variant Carrier Status and Family History



Cumulative probability of developing disease at ages 55, 65, and 75 years, estimated using Cox proportional-hazards model, standardized to the average of the first four genetic principal components, sex, family history, and family history-carrier status interaction (if significant) and stratified by variant carrier status and family history for familial hypercholesterolemia (coronary artery disease, stroke, or peripheral artery disease, A, B), hereditary breast and ovarian cancer (C, D), and Lynch syndrome (colorectal and uterine cancer, E, F) in women and men, respectively. The p-values for interaction between family history and carrier status were  $p=0.08$  for familial hypercholesterolemia, p=0.20 for hereditary breast and ovarian cancer syndrome, and p=0.003 for Lynch syndrome.

## **eReferences.**

- 1. Van Hout CV, Tachmazidou I, Backman JD, et al. *Whole Exome Sequencing and Characterization of Coding Variation in 49,960 Individuals in the UK Biobank*. Genomics; 2019. doi:10.1101/572347
- 2. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
- 3. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779
- 4. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
- 5. Regier AA, Farjoun Y, Larson DE, et al. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun*. 2018;9(1):4038. doi:10.1038/s41467-018-06159-4
- 6. Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinforma Oxf Engl*. 2014;30(20):2843-2851. doi:10.1093/bioinformatics/btu356
- 7. Zhao H, Sun Z, Wang J, Huang H, Kocher J-P, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinforma Oxf Engl*. 2014;30(7):1006-1007. doi:10.1093/bioinformatics/btt730
- 8. Karczewski KJ, Francioli LC, Tiao G, et al. *Variation across 141,456 Human Exomes and Genomes Reveals the Spectrum of Loss-of-Function Intolerance across Human Protein-Coding Genes*. Genomics; 2019. doi:10.1101/531210
- 9. Richards S, Aziz N, Bale S, 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 Off J Am Coll Med Genet*. 2015;17(5):405-424. doi:10.1038/gim.2015.30
- 10. Rehm HL, Berg JS, Brooks LD, et al. ClinGen--the Clinical Genome Resource. *N Engl J Med*. 2015;372(23):2235-2242. doi:10.1056/NEJMsr1406261
- 11. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862-868. doi:10.1093/nar/gkv1222
- 12. Cholesterol Treatment Trialists' (CTT) Collaboration. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised trials. *Lancet*. 2010;376(9753):1670-1681. doi:10.1016/S0140-6736(10)61350-5
- 13. Sudhop T, Lütjohann D, Kodal A, et al. Inhibition of intestinal cholesterol absorption by ezetimibe in humans. *Circulation*. 2002;106(15):1943-1948. doi:10.1161/01.cir.0000034044.95911.dc
- 14. Cannon CP, Blazing MA, Giugliano RP, et al. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. *N Engl J Med*. 2015;372(25):2387-2397. doi:10.1056/NEJMoa1410489
- 15. Zhao Z, Du S, Shen S, et al. Comparative efficacy and safety of lipid-lowering agents in patients with hypercholesterolemia: A frequentist network meta-analysis. *Medicine (Baltimore)*. 2019;98(6):e14400. doi:10.1097/MD.0000000000014400
- 16. Lloyd-Jones DM, Morris PB, Ballantyne CM, et al. 2017 Focused Update of the 2016 ACC Expert Consensus Decision Pathway on the Role of Non-Statin Therapies for LDL-Cholesterol Lowering in the Management of Atherosclerotic Cardiovascular Disease Risk: A Report of the American College of Cardiology Task Force on Expert Consensus Decision Pathways. *J Am Coll Cardiol*. 2017;70(14):1785-1822. doi:10.1016/j.jacc.2017.07.745
- 17. Birjmohun RS, Hutten BA, Kastelein JJP, Stroes ESG. Efficacy and safety of high-density lipoprotein cholesterol-increasing compounds: a meta-analysis of randomized controlled trials. *J Am Coll Cardiol*. 2005;45(2):185-197. doi:10.1016/j.jacc.2004.10.031
- 18. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909. doi:10.1038/ng1847