

## RESEARCH ARTICLE SUMMARY

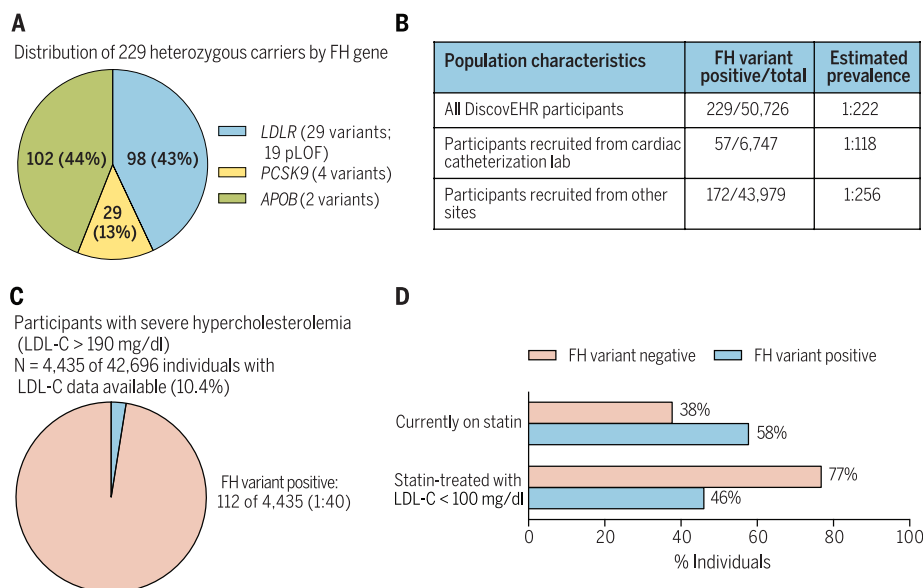
## HUMAN GENETICS

# Genetic identification of familial hypercholesterolemia within a single U.S. health care system

Noura S. Abul-Husn, Kandanurug Manickam, Laney K. Jones, Eric A. Wright, Dustin N. Hartzel, Claudia Gonzaga-Jauregui, Colm O'Dushlaine, Joseph B. Leader, H. Lester Kirchner, D'Andra M. Lindbuchler, Marci L. Barr, Monica A. Giovanni, Marylyn D. Ritchie, John D. Overton, Jeffrey G. Reid, Raghu P. R. Metpally, Amr H. Wardeh, Ingrid B. Borecki, George D. Yancopoulos, Aris Baras, Alan R. Shuldiner, Omri Gottesman, David H. Ledbetter, David J. Carey, Frederick E. Dewey, Michael F. Murray\*

**INTRODUCTION:** Familial hypercholesterolemia (FH) is a public health genomics priority but remains underdiagnosed and undertreated despite widespread cholesterol screening. This represents a missed opportunity to prevent FH-associated cardiovascular morbidity and mortality. Pathogenic variants in three genes (*LDLR*, *APOB*, and *PCSK9*) account for the majority of FH cases. We assessed the prevalence and clinical impact of FH-associated genomic variants in 50,726 individuals from the MyCode Community Health Initiative at Geisinger Health System who underwent exome sequencing as part of the DiscovEHR human genetics collaboration with the Regeneron Genetics Center.

**RATIONALE:** Genetic testing for FH is uncommon in clinical practice in the United States, and the prevalence of FH variants in U.S. populations has not been well established. We sought to evaluate FH prevalence in a large integrated U.S. health care system using genomic sequencing and electronic health record (EHR) data. We determined the impact of FH variants on low-density lipoprotein cholesterol (LDL-C) levels and coronary artery disease (CAD) risk. We assessed the likelihood of FH variant carriers achieving a presequencing EHR-based FH diagnosis according to established clinical diagnostic criteria. Finally, we examined the rates of statin medication use and outcomes in FH variant carriers.



**Prevalence and clinical impact of FH variants in a large U.S. clinical care cohort.** (A) Distribution of 229 heterozygous carriers of an FH variant in the DiscovEHR cohort by FH gene. (B) Prevalence of an FH variant in the DiscovEHR cohort and according to recruitment site. (C) Prevalence of an FH variant among individuals with severe hypercholesterolemia (LDL-C  $\geq$  190 mg/dl). (D) Statin treatment rates and outcomes in FH variant carriers and noncarriers.

**RESULTS:** Thirty-five known and predicted pathogenic variants in *LDLR*, *APOB*, and *PCSK9* were identified in 229 individuals. The estimated FH prevalence was 1:256 in unselected participants and 1:118 in participants ascertained via the cardiac catheterization laboratory. FH variants were found in only 2.5% of individuals with severe hypercholesterolemia (maximum EHR-documented LDL-C  $\geq$  190 mg/dl) in the cohort, and a maximum LDL-C of  $\geq$  190 mg/dl was absent in 45% of FH variant carriers. Overall, FH variant carriers had  $69 \pm 3$  mg/dl greater maximum LDL-C than sequenced noncarriers ( $P = 1.8 \times 10^{-20}$ ) and had significantly increased odds of general and premature CAD

## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aaf7000>

[odds ratio (OR), 2.6 ( $P = 4.3 \times 10^{-11}$ ) and 3.7 ( $P = 5.5 \times 10^{-14}$ ), respectively]. The increased odds of general and premature CAD were most pronounced in carriers of *LDLR* predicted loss-of-function variants [OR, 5.5 ( $P = 7.7 \times 10^{-13}$ ) and 10.3 ( $P = 9.8 \times 10^{-19}$ ), respectively]. Fourteen FH variant carriers were deceased; chart review revealed that none of these individuals had a clinical diagnosis of FH. Before genetic testing, only 15% of FH variant carriers had an ICD-10 (International Classification of Diseases, 10th revision) diagnosis code for pure hypercholesterolemia or had been seen in a lipid clinic, suggesting that few had been previously diagnosed with FH. Retrospectively applying Dutch Lipid Clinic Network diagnostic criteria to EHR data, we found presequencing criteria supporting a probable or definite clinical diagnosis of FH in 24% of FH variant carriers, highlighting the limitations of using existing clinical criteria for EHR-based screening in the absence of genetic testing. Active statin use was identified in 58% and high-intensity statin use in 37% of FH variant carriers. Only 46% of carriers currently on statin therapy had a most recent LDL-C level below 100 mg/dl compared to 77% of noncarriers.

**CONCLUSION:** In summary, we show that large-scale genomic screening in patients with longitudinal EHR data has the ability to detect FH, uncover and characterize novel pathogenic variants, determine disease prevalence, and enhance overall knowledge of clinical impact and outcomes. The 1:256 prevalence of FH variants in this predominantly European-American cohort is in line with prevalence estimates from recent work in European cohorts. Our findings highlight the undertreatment of FH variant carriers and demonstrate a potential clinical benefit for large-scale sequencing initiatives in service of precision medicine. ■

The list of author affiliations is available in the full article online.

\*Corresponding author. Email: [mfmurray1@geisinger.edu](mailto:mfmurray1@geisinger.edu)  
Cite this article as N. S. Abul-Husn et al., *Science* 354, aaf7000 (2016). DOI: 10.1126/science.aaf7000

## RESEARCH ARTICLE

## HUMAN GENETICS

# Genetic identification of familial hypercholesterolemia within a single U.S. health care system

Noura S. Abul-Husn,<sup>1</sup> Kandamurugu Manickam,<sup>2</sup> Laney K. Jones,<sup>2</sup> Eric A. Wright,<sup>2</sup> Dustin N. Hartzel,<sup>2</sup> Claudia Gonzaga-Jauregui,<sup>1</sup> Colm O'Dushlaine,<sup>1</sup> Joseph B. Leader,<sup>2</sup> H. Lester Kirchner,<sup>2</sup> D'Andra M. Lindbuchler,<sup>2</sup> Marci L. Barr,<sup>2</sup> Monica A. Giovanni,<sup>2</sup> Marylyn D. Ritchie,<sup>2</sup> John D. Overton,<sup>1</sup> Jeffrey G. Reid,<sup>1</sup> Raghu P. R. Metpally,<sup>2</sup> Amr H. Wardeh,<sup>2</sup> Ingrid B. Borecki,<sup>1</sup> George D. Yancopoulos,<sup>1</sup> Aris Baras,<sup>1</sup> Alan R. Shuldiner,<sup>1</sup> Omri Gottesman,<sup>1</sup> David H. Ledbetter,<sup>2</sup> David J. Carey,<sup>2</sup> Frederick E. Dewey,<sup>1</sup> Michael F. Murray<sup>2\*</sup>

Familial hypercholesterolemia (FH) remains underdiagnosed despite widespread cholesterol screening. Exome sequencing and electronic health record (EHR) data of 50,726 individuals were used to assess the prevalence and clinical impact of FH-associated genomic variants in the Geisinger Health System. The estimated FH prevalence was 1:256 in unselected participants and 1:118 in participants ascertained via the cardiac catheterization laboratory. FH variant carriers had significantly increased risk of coronary artery disease. Only 24% of carriers met EHR-based presequencing criteria for probable or definite FH diagnosis. Active statin use was identified in 58% of carriers; 46% of statin-treated carriers had a low-density lipoprotein cholesterol level below 100 mg/dl. Thus, we find that genomic screening can prompt the diagnosis of FH patients, most of whom are receiving inadequate lipid-lowering therapy.

Familial hypercholesterolemia (FH) is one of three genomic conditions designated by the Centers for Disease Control and Prevention as having potential for a significant positive impact on public health through improved diagnosis and treatment (1, 2). FH is characterized by substantial, lifelong elevation of low-density lipoprotein cholesterol (LDL-C) and a markedly increased risk of premature cardiovascular disease (3, 4). Known genetic causes of FH include inactivating variants in the gene encoding the LDL receptor (*LDLR*), protein-disrupting variants in apolipoprotein B (*APOB*), and activating variants in proprotein convertase subtilisin/kexin type 9 (*PCSK9*). Although estimated at a worldwide prevalence of 1:500 (5–7), recent studies in some European countries have revealed that FH could affect ~1:250 individuals (8, 9), with higher prevalence observed in certain founder populations (10, 11). The prevalence of FH in U.S. populations is not well established. Notably, despite widespread cholesterol screening, only a small fraction of FH cases are appropriately diagnosed and treated (4, 12, 13). This represents a missed opportunity to prevent FH-associated cardiovascular morbidity and mortality.

A diagnosis of FH can be made with a validated set of criteria, such as those established by

the Dutch Lipid Clinic Network (DLCN), Simon Broome, or Make Early Diagnosis to Prevent Early Death (MEDPED) (14–17). These diagnostic tools estimate the likelihood of FH on the basis of clinical features and, in the case of DLCN and Simon Broome criteria, also include identification of functional variants in the *LDLR*, *APOB*, or *PCSK9* genes. However, genetic testing for these variants is uncommon in clinical practice in the United States. We thus sought to understand the prevalence and clinical impact of FH variants in a clinical cohort by analyzing genomic sequence and electronic health record (EHR) data from 50,726 individuals from the Geisinger Health System, an integrated health care system with provider services in Pennsylvania and New Jersey.

## Exome sequencing of 50,726 individuals reveals a high genotypic prevalence of FH

This study included 50,726 consented adult participants from the MyCode Community Health Initiative of Geisinger Health System (18) who underwent exome sequencing as part of the DiscovEHR human genetics study (see table S1 for demographics and clinical characteristics of the study population) (19). Participants were 59.2% female, with a median age of 61 years, and predominantly Caucasian (98.4%). LDL-C values were available in the EHR for 42,696 (84.2%) of sequenced participants; of these, 4435 (10.4%) had severe hypercholesterolemia, defined as maximum EHR-documented LDL-C  $\geq$  190 mg/dl (20).

Statin use was documented at any point in the EHR in 27,402 (54.0%) of sequenced participants.

The exome sequence data were analyzed for known pathogenic variants in *LDLR*, *PCSK9*, and *APOB* and for predicted protein-truncating, loss-of-function (LoF) variants in *LDLR* (table S2), including exonic copy number variants identified via CLAMMS (Copy number estimation using Lattice-Aligned Mixture Models) (21). By positional intersection with the clinical genetics database ClinVar (22), we found 19 missense variants in *LDLR*, *PCSK9*, and *APOB* designated as “pathogenic” for FH. We identified 21 additional predicted LoF variants in *LDLR*, including 9 frameshift, 8 splice donor or acceptor, and 4 nonsense variants. We also identified a predicted pathogenic exon 13 to 17 tandem duplication in *LDLR*; this was the only copy number variant identified in *LDLR*. Upon manual review, six variants were removed to produce a more stringent set of variants (tables S1 and S2); these included two missense variants with conflicting evidence in ClinVar, two missense variants that were located at the same amino acid residue as a pathogenic variant but were not in ClinVar themselves, and two that were annotated as splice variants in an alternative transcript of *LDLR*, for which there is no evidence of protein expression. Thus, 35 known (that is, ClinVar-documented) and predicted (that is, protein-truncating LoF) pathogenic FH variants (29 *LDLR*, 4 *PCSK9*, and 2 *APOB* variants) were used for our analyses.

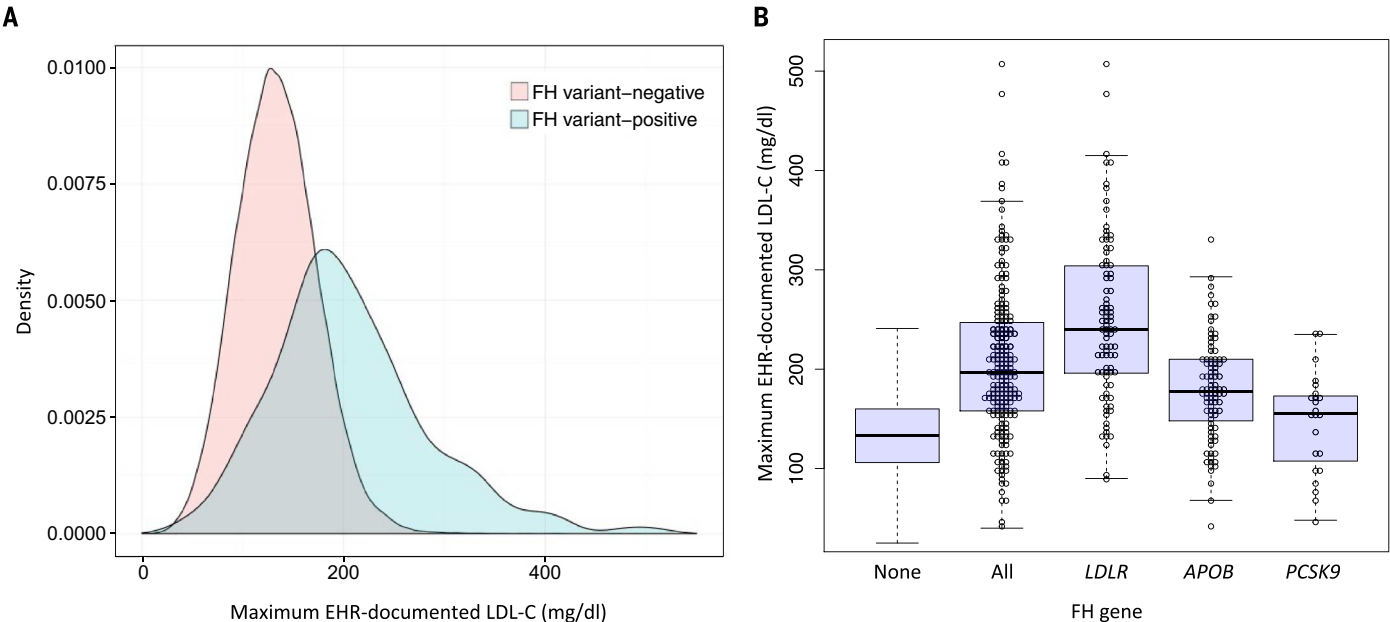
Among the 50,726 sequenced participants, we identified 229 heterozygous carriers of 1 of the 35 FH variants, corresponding to a total carrier frequency of 1:222 participants (Table 1). There were no cases of homozygous or compound heterozygous FH. Given that this prevalence estimate is based on a sampling of individuals within a single health care delivery system, it may be an overestimation of population frequency due to ascertainment bias. The MyCode cohort included 6747 participants recruited from the cardiac catheterization laboratory; the estimated prevalence of an FH variant among these participants was 1:118, and the prevalence in other participants was 1:256 (Table 1). Overall, 98 (42.8%) individuals were found with *LDLR* variants, 102 (44.5%) with *APOB* variants, and 29 (12.7%) with *PCSK9* variants (table S2). A recent study by Khera *et al.* (23) identified more individuals with variants in *LDLR* (86%) and relatively fewer *APOB* (13%) and *PCSK9* (0.6%); this discrepancy is likely due to the inclusion of different populations in each study, with theirs being 46% South Asian and 7% black.

We identified 30 first- or second-degree relationships among FH variant carriers using their exome data (24), including two pedigrees containing five sequenced noncarriers and eight carriers of the *APOB* p.Arg3527Gln variant, where segregation of high LDL-C levels with carrier status can be observed (fig. S1). The three variants with the largest number of related carriers accounted for more than half of FH cases in the study: *APOB* p.Arg3558Cys (46 carriers, 8 related), *APOB* p.Arg3527Gln (56 carriers, 10 related), and

<sup>1</sup>Regeneron Genetics Center, Tarrytown, NY 10591, USA.

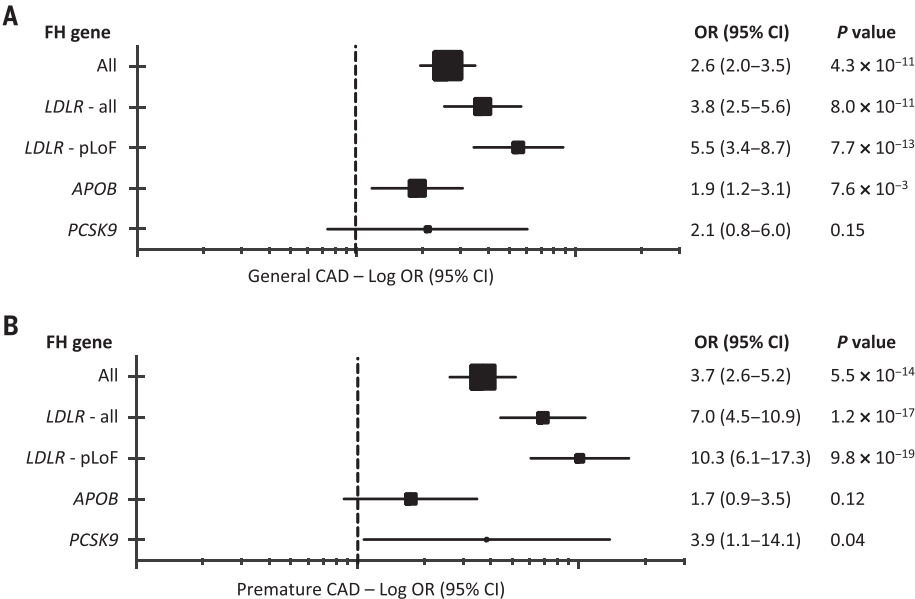
<sup>2</sup>Geisinger Health System, Danville, PA 17822, USA.

\*Corresponding author. Email: mfmurray1@geisinger.edu



**Fig. 1. FH variants are associated with increased LDL-C levels.** (A) Density of maximum EHR-documented LDL-C levels in heterozygous carriers of any FH variant (FH variant-positive,  $n = 204$ ) and in sequenced noncarriers (FH variant-negative,  $n = 42,442$ ) with LDL-C data available in the EHR. LDL-C levels approximated a normal distribution (Anderson-Darling normality test;  $P < 2.2 \times 10^{-16}$  in FH variant-negative and  $P = 6.1 \times 10^{-5}$  in FH variant-positive). FH variant carriers had  $69 \pm 3$  mg/dl greater LDL-C than noncarriers in a mixed linear model analysis adjusting for age, age<sup>2</sup>, sex, and the first five principal components of ancestry ( $P < 1.8 \times 10^{-20}$ ). (B) Maximum EHR-documented LDL-C levels in sequenced noncarriers ( $n = 42,442$ ) and

in carriers according to FH gene ( $n = 88, 92$ , and  $24$  for *LDLR*, *APOB*, and *PCSK9*, respectively). Open circles indicate individual LDL-C values for FH variant carriers. Median LDL-C level (in mg/dl) and IQR are shown in the box plots; these were 133.0 (106.0 to 160.0) in noncarriers, 240.3 (196.5 to 303.5) in *LDLR*, 178 (148.0 to 210.0) in *APOB*, and 155.0 (107.5 to 173.0) in *PCSK9* variant carriers. Maximum LDL-C was higher in carriers of variants in *LDLR* compared to *APOB* or *PCSK9* (one-way ANOVA with post hoc Tukey test;  $P = 1.5 \times 10^{-10}$  and  $P = 1.3 \times 10^{-9}$ , respectively). There was no significant difference in maximum LDL-C between carriers of *APOB* and *PCSK9* variants ( $P = 0.09$ ).



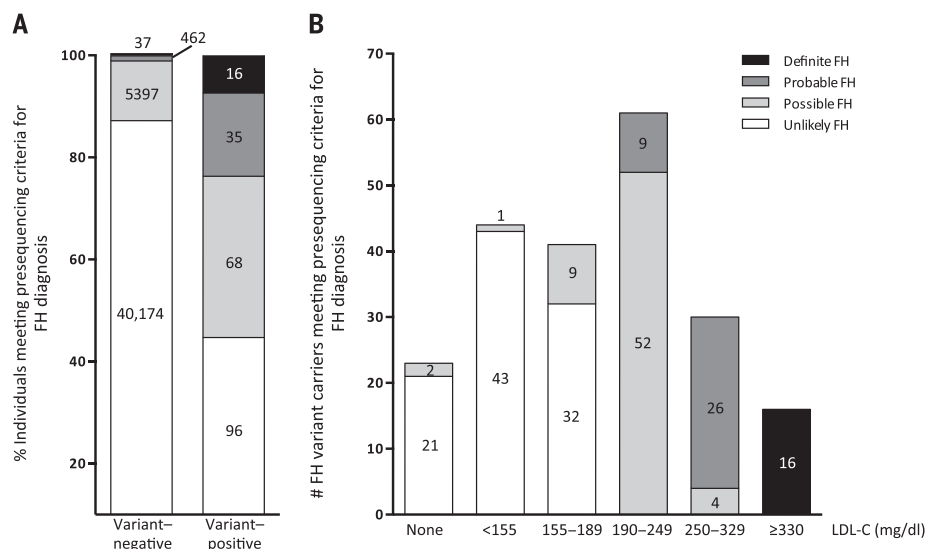
**Fig. 2. FH variants are associated with increased risk of CAD.** CAD cases and controls were defined with ICD-9 (International Classification of Diseases, Ninth Revision) diagnosis codes and cardiac catheterization report data extracted from EHRs. ORs for general CAD (A) and premature CAD (B) (defined as males  $\leq 55$  years and females  $\leq 65$  years) were calculated by logistic regression with adjustment for age, sex, and principal components of ancestry (see table S3). pLoF, predicted LoF (defined as variants leading to a premature stop codon, or loss of a start or stop codon; disrupting canonical splice acceptor or donor dinucleotides; or frameshifting leading to the formation of a premature stop codon); *LDLR* - all, all known and predicted pathogenic variants identified in *LDLR*; *LDLR* - pLoF, predicted LoF variants identified in *LDLR*.

the *LDLR* exon 13 to 17 duplication (29 carriers, 10 related). This underscores the likelihood of encountering close family members with FH even in an unselected clinical population within a regional health care system and the opportunity for family-based screening and clinical management. In a subset of the sequenced cohort in which only one individual in every first- and second-degree relationship was retained, the overall prevalence of an FH variant remained unchanged at 1:224 (Table 1).

**FH variant carriers have higher LDL-C levels than noncarriers and are at increased cardiovascular risk**

LDL-C levels were available in the EHR for 204 of 229 FH variant-positive and for 42,442 FH variant-negative individuals. We examined maximum EHR-documented LDL-C levels in the sequenced cohort to approximate the untreated or pretreated state. Among FH variant-positive and FH variant-negative individuals, maximum LDL-C levels approximated a normal distribution (Fig. 1A). Carriers of any FH variant had  $69 \pm 3$  mg/dl greater maximum LDL-C than sequenced noncarriers in a mixed linear model analysis ( $P = 1.8 \times 10^{-20}$ ). Maximum LDL-C values were significantly higher in carriers of *LDLR* variants [median, 240.3 mg/dl; interquartile range (IQR), 196.5 to 303.5] compared to carriers of *APOB* [median, 178.0 mg/dl; IQR, 148.0 to 210.0; one-way





**Fig. 3. Presequencing likelihood of FH diagnosis with DLCN criteria.** Criteria for diagnosis of FH (unlikely, possible, probable, or definite) were based exclusively on extracted EHR data. **(A)** Percentage of participants meeting clinical criteria for FH diagnosis among living noncarriers (variant-negative;  $n = 46,070$ ) and carriers (variant-positive;  $n = 215$ ) of an FH variant. **(B)** Number of living variant carriers ( $n = 215$ ) that would meet presequencing criteria for FH diagnosis per stratum of maximum EHR-documented LDL-C range (in mg/dl).

analysis of variance (ANOVA) with post hoc Tukey test,  $P = 1.5 \times 10^{-10}$ ] or *PCSK9* (median, 155.0 mg/dl; IQR, 107.5 to 173.0;  $P = 1.3 \times 10^{-9}$ ) variants (Fig. 1B). Despite overall increased LDL-C in FH carriers, maximum LDL-C levels for each identified FH variant were widely variable, ranging from a median of 90 to 479 mg/dl (table S2).

The frequency of known and predicted pathogenic FH variants was evaluated across categories of maximum EHR-documented LDL-C levels (Table 1 and table S1). A significantly higher proportion of FH variant carriers (112 of 204 or 54.9%) had severe hypercholesterolemia (maximum LDL-C  $\geq 190$  mg/dl) compared to noncarriers (4309 of 42,442 or 10.2%;  $\chi^2$  test,  $P < 0.0001$ ; table S1). The prevalence of an FH variant increased significantly across increasing thresholds of maximum LDL-C levels (Table 1; Cochran-Armitage test for trend,  $P < 0.0001$ ). Of the 4435 sequenced participants with LDL-C  $\geq 190$  mg/dl, only 112 harbored an FH variant (2.5% or 1:40). Of the 53 with maximum LDL-C  $\geq 330$  mg/dl, 17 had an FH variant (32.1% or 1:3).

We evaluated the association of known and predicted pathogenic FH variants with coronary artery disease (CAD) as defined by electronic phenotyping of CAD cases and controls through the EHR (25), using linear mixed models to account for population structure due to ancestry and relatedness (Fig. 2 and table S3). Individuals with an FH variant had increased odds of CAD [odds ratio (OR), 2.6; 95% confidence interval (CI), 2.0 to 3.5;  $P = 4.3 \times 10^{-11}$ ] compared to noncarriers. The increased risk of CAD was greatest in carriers of *LDLR* predicted LoF variants (OR, 5.5; 95% CI, 3.4 to 8.7;  $P = 7.7 \times 10^{-13}$ ) and smallest in carriers of *APOB* variants (OR, 1.9;

95% CI, 1.2 to 3.1;  $P = 7.6 \times 10^{-3}$ ). We identified 4150 individuals with premature CAD (defined as having CAD  $\leq 55$  years in males and  $\leq 65$  years in females) in the cohort. Of these, 53 individuals harbored an FH variant, indicating a prevalence of genotypically defined FH among individuals with premature CAD of 1:78 (1.3%). The odds of premature CAD were significantly increased among FH variant carriers (OR, 3.7; 95% CI, 2.6 to 5.2;  $P = 5.5 \times 10^{-14}$ ). As was observed with general CAD, the increased risk of premature CAD was greatest in carriers of *LDLR* predicted LoF variants (OR, 10.3; 95% CI, 6.1 to 17.3;  $P = 9.8 \times 10^{-19}$ ) and smallest in carriers of *APOB* variants (OR, 1.7; 95% CI, 0.9 to 3.5;  $P = 0.12$ ). Notably, 14 of the 229 FH variant carriers were deceased (mean age of death, 76.1 years; range, 58 to 91 years). Chart review revealed that the cause of death was cardiovascular-related in half of these individuals and that 13 had evidence of potential FH-related comorbid disease (table S4). None of the deceased FH variant carriers carried a clinical diagnosis of FH.

### EHR-based FH diagnosis is challenging without knowledge of an FH variant

We evaluated whether any of the 229 FH variant carriers might have been previously diagnosed with FH by mining the EHRs for a diagnosis code of “pure hypercholesterolemia” or for any encounter at Geisinger’s specialized Lipid Clinic (table S1); these were present in the EHRs of only 35 (15.3%) carriers.

We retrospectively applied the DLCN criteria for FH diagnosis to EHR data available from living individuals in the sequenced MyCode cohort (table S5). These criteria allow for a diagnosis of FH on the basis of clinical and family history of

elevated LDL-C, premature cardiovascular disease, physical stigmata of hypercholesterolemia, and, when known, molecular identification of FH variants (16). Among living variant-negative individuals ( $n = 46,070$ ), there were 37 (0.1%) “definite,” 462 (1.0%) “probable,” and 5397 (11.7%) “possible” FH diagnoses solely on the basis of EHR data (Fig. 3A). Among living variant-positive individuals ( $n = 215$ ), there were 16 (7.4%) definite, 35 (16.3%) probable, and 68 (31.6%) possible FH diagnoses (Fig. 3A and table S6). Individuals meeting probable or definite criteria for FH diagnosis (23.7% of carriers) were clustered among those with a maximum LDL-C level of  $\geq 190$  mg/dl (Fig. 3B). There were 96 (44.7%) variant carriers who would have been judged unlikely to have a diagnosis of FH; there were no significant differences in age or other characteristics between these individuals and those in the combined possible/probable/definite FH categories (table S6). When the same DLCN criteria were applied with the inclusion of identified FH variants, 188 (87.4%) individuals met criteria for definite and 27 (12.6%) met criteria for probable FH diagnosis. Application of MEDPED criteria (17) to EHRs produced similar results, with 53 (24.7%) living FH variant carriers meeting criteria for clinical FH diagnosis (table S7). Together, these data highlight the limitations of using clinical criteria for EHR-based screening and the opportunity for genomic data to augment the detection of individuals with FH.

Individuals deemed unlikely to have FH (by DLCN criteria) before identification of a pathogenic variant, despite having LDL-C data available in the EHR ( $n = 75$ ), were considered to have nonpenetrant disease. On the basis of these results, we predicted 18 FH variants to have reduced penetrance, which included 13 of 29 *LDLR*, 3 of 4 *PCSK9*, and 2 of 2 *APOB* variants (table S2). This estimate is limited by the incompleteness of EHR data (table S5), the inability to account fully for lipid-lowering treatment effects, the variable number of heterozygous carriers per variant, and familial relationships between some of the carriers within the cohort. For example, closely related carriers may have shared environmental and genetic factors influencing their phenotypes, which could affect estimates of penetrance.

### FH variant carriers are undertreated and have LDL-C levels above goal

Overall, 173 (80.5%) of the 215 living FH variant carriers had been prescribed lipid-lowering therapy (mostly statin medications; table S8). We examined the medication reconciliation records from 2015 to 2016 available for 189 of the FH variant carriers. An active prescription for statin medication was identified in 109 (57.7%) FH variant carriers, of whom 70 (37.0%) were prescribed high-intensity statin therapy. Among those not receiving high-intensity statin therapy ( $n = 119$ ), 10 individuals (8.4%) had EHR-documented evidence of statin intolerance. The average lowest LDL-C level among 192 FH variants carriers with these data available was 104.9 mg/dl. Among 106

**Table 1. Prevalence of an FH variant in the MyCode cohort.** We assessed the prevalence of an FH variant in all sequenced participants, in a subset in which only one individual in each first- and second-degree relationship was retained, according to recruitment site (from the cardiac catheterization laboratory or elsewhere), and across increasing LDL-C thresholds.

Population characteristics	n	FH variant–positive n (%)*	Estimated prevalence
All sequenced participants	50,726	229 (0.5)	1:222
Including only one individual in every first- and second-degree relationship	38,339	171 (0.5)	1:224
Recruitment site†			
Cardiac catheterization laboratory	6,747	57 (0.8)	1:118
Other sites	43,979	172 (0.4)	1:256
Maximum EHR-documented LDL-C (mg/dl)‡			
LDL-C < 155	28,512	48 (0.2)	1:607
LDL-C ≥ 155	14,184	156 (1.1)	1:91
LDL-C ≥ 190	4,435	112 (2.5)	1:40
LDL-C ≥ 250	390	50 (12.8)	1:8
LDL-C ≥ 330	53	17 (32.1)	1:3

\*Heterozygous carriers of 1 of the 35 identified FH variants (n = 229 total; n = 204 individuals with LDL-C data available). †n = 50,726 sequenced participants. ‡n = 42,696 sequenced participants with LDL-C data available.

variant carriers with LDL-C levels available within 12 months of the study, the average most recent LDL-C level was 130.3 mg/dl, and only 41 (38.7%) carriers had a most recent LDL-C level below 100 mg/dl, which is the LDL-C target recommended by the National Lipid Association's Expert Panel for adults with FH (26). Subsetting to carriers currently on statin therapy and with recent LDL-C levels available (n = 63), only 29 (46.0%) had a most recent LDL-C level below 100 mg/dl. In comparison, 76.8% of noncarriers currently on statin had a most recent LDL-C below 100 mg/dl ( $\chi^2$  test,  $P < 0.0001$ ; table S8).

Conclusion

Underdiagnosis and undertreatment of FH continue to be a concern, and are associated with premature CAD and stroke, indicating that genomic approaches to case identification may be needed as a front-end intervention to improve outcomes (4). Here, exome sequence data linked to longitudinal EHRs identified FH variants at a higher than previously estimated prevalence (6, 7, 10); excluding participants recruited from a cardiac catheterization laboratory, the prevalence was 1:256, in line with more recent estimates of 1:217 in Denmark (8) and 1:319 in the Netherlands (9). Our assessment of FH prevalence in a U.S. health care system using this methodology supports the claim that there is significant underdiagnosis of this condition (4, 12). Whether our estimated prevalence of FH variants in our patient population, largely a stable regional health care population in central Pennsylvania, is generalizable to other U.S. patient populations remains to be determined. We caution that extant familial (and cryptic) relatedness in our study population could result in overestimation of a given FH-causing allele that is rare in the general population but segregating in family members. For example, our study population was enriched for *APOB* p.Arg3527Gln, which is known to be common in individuals of Amish descent in Lancaster, Pennsylvania (11).

Surprisingly, FH variants explain only 2.5% of severe hypercholesterolemia in the cohort, which challenges the notion that patients with LDL-C levels of more than 190 mg/dl are likely to have genotypically defined FH (27); this is consistent with a recent observation that 1.7% of patients with severe hypercholesterolemia had an FH variant (23). There are several potential explanations for this. Polygenic hypercholesterolemia (28, 29) could account for a proportion of FH variant–negative cases of severe LDL-C elevation, as could common, secondary causes of elevated LDL-C, such as obesity, hypothyroidism, and nephrotic syndrome (20, 30). It is also possible that, by including only established pathogenic variants in *LDLR*, *APOB*, and *PCSK9*, and predicted pathogenic LoF variants in *LDLR* in our analysis, we excluded additional truly pathogenic FH variants that would explain some cases of severe hypercholesterolemia. Finally, there may be yet-to-be-identified genes that are causative of FH.

We applied a validated set of clinical FH diagnostic criteria to EHR data to estimate the likelihood of achieving a presequencing diagnosis in living FH variant carriers. Only 24% of carriers would have met criteria for probable or definite FH diagnosis before variant identification. This suggests that such diagnostic algorithms may be of limited utility when applied to EHRs in the absence of genetic testing, as others have observed (31), which has direct implications for current efforts seeking to use similar methodologies. Future large-scale EHR screening programs may use natural language processing and machine learning to improve reliability. For example, the FH foundation recently launched the Find FH program, which is developing a machine-learning algorithm to identify individuals with probable FH within EHR data, laboratory results, and claims databases (32).

By surveying EHR records from last year, we found that 58% of FH variant carriers were currently prescribed a statin medication, and only 46% of those on statin treatment had a most

recent LDL-C level under 100 mg/dl, the recommended LDL-C target for adult FH patients (20, 27, 33). These findings are consistent with previous reports and highlight the undertreatment of FH (4, 34). The risk for general and premature CAD was significantly higher in FH variant carriers, with ORs of 2.6 and 3.7, respectively, and this was most pronounced in those with predicted LoF variants in *LDLR*. A similarly increased risk of CAD was observed by Khera *et al.* (23), who reported ORs of 3.8 for carriers of any FH variant and 9.5 for predicted LoF variants in *LDLR*. This underscores the importance of early diagnosis and appropriate treatment of these patients. At present, there are no genotype-based guidelines for the treatment of hyperlipidemia; future studies evaluating the effects of new therapies on cardiovascular disease outcomes may lead to clarification of FH treatment guidelines.

Despite its strengths, this study has limitations. The study population does not necessarily reflect the global community, but rather individuals presenting for clinical care. The average age of sequenced MyCode participants was 61 years; therefore, there may be survival bias in this cohort, with underrepresentation of individuals having severe or homozygous FH (who often do not survive past the second decade if left untreated) (35). The cohort also had a higher percentage of ever-prescribed statin compared to the total ~1.4 million patients in the Geisinger Health System with EHR data (54% versus 13%), although surprisingly few FH variant carriers had an active statin order in the sequenced cohort. As with any real-world study, we were limited by the data collected, which is subject to error in entry and incompleteness, specifically missing data in the EHR (such as physical exam findings and family history). Because major efforts, including the national Precision Medicine Initiative, will also rely heavily on EHR data, this study helps to validate this approach despite the limitations cited. We assumed that the most recent outpatient LDL-C value was reflective

of current prescribed lipid-lowering therapy and that medications were being taken upon physician order and not by independent verification. We did not adjust LDL-C values to account for treatment effect; instead, we elected to analyze maximum EHR-documented LDL-C, because this was most likely to approximate the untreated state in this real-world cohort and because of some degree of uncertainty in the timing of statin initiation relative to EHR-documented LDL-C levels. Other studies have implemented a standard 30% reduction in LDL-C in individuals on statin treatment (23), but note that this does not account for different statin types or doses, or potential variability in response across FH carriers and noncarriers. Finally, we elected to use a conservative definition of pathogenic variants, which did not include missense variants designated as “likely pathogenic” or having in silico predictions of pathogenicity in our analysis; this could have resulted in an underestimate of the true frequency of FH.

This study helps build support for the emerging concept that, in conditions such as FH, genomic sequencing might soon be applied as part of a population screen (36). Given that genomic sequencing has not been routinized in U.S. health care except in research settings such as this, other health care systems would currently require programmatic funding beyond clinical reimbursements to apply a similar approach to their patient population (37). The costs of sequencing, interpretation, and follow-up of secondary and false-positive findings from genomic sequencing would need to be considered. Proof of clinical utility and cost-effectiveness of genomic screening for FH can be built on the Dutch model experience, where screening, diagnosis, and treatment led to more than 3 years of life gained in FH (38). A recent study modeling genetic screening for FH suggested that it would not currently be cost-effective in the United States, although the authors acknowledged several knowledge gaps regarding FH diagnosis, cascade screening, management, and clinical outcomes in the United States, as well as outdated genetic testing cost input data (39). Further analyses taking into account a changed DNA sequencing landscape with rapidly declining costs, and new and more effective treatment options, will be critical in considering implementation of genomic screening for FH and in determining the clinical settings and patient populations that would benefit most from this approach. Genetic identification of FH is just one clinical application of the many possible uses of genomic sequencing, and in and of itself may not justify the cost. Cost-effectiveness and clinical utility models in which genomic sequencing is used as a single test to inform many actionable genetic diseases require further evaluation.

This study targets a particular use of sequencing information, and at this point, we cannot fully model how genomic sequencing as a front-end intervention might alter clinical care by improving detection of FH-conferred risk. Nonetheless, our findings demonstrate a potential clinical benefit for the large-scale sequencing

planned by the national Precision Medicine Initiative (40). An important goal of precision medicine is to accurately identify and treat individuals at increased risk of medically actionable conditions (41); as a highly modifiable genetic condition, FH is an ideal starting point for implementation of a return of results program (42). The clinical impact of systematic genomic screening for FH may be further magnified by cascade screening of family members, identification of novel genetic causes of FH, and protective genetic or other factors to explain nonpenetrance, ultimately leading to more refined cardiovascular risk stratification for patients with elevated LDL-C levels.

## Materials and methods

### Setting and study participants

Human genetics studies were conducted as part of the DiscovEHR project of the Geisinger Health System (GHS) and the Regeneron Genetics Center (RGC). The study was approved by the GHS Institutional Review Board. The study population consisted of 50,726 consented participants  $\geq 18$  years from the MyCode® Community Health initiative of GHS, an integrated health services organization in Pennsylvania and New Jersey. A detailed description of the DiscovEHR study population can be found in a companion publication (19).

### Sequencing of LDLR, APOB, and PCSK9

Sample preparation and whole exome sequencing were performed at the RGC as previously described (25). In brief, exome capture was performed using NimbleGen probes according to the manufacturer's recommended protocol (Roche NimbleGen). The captured DNA was PCR amplified and quantified by qRT-PCR (Kapa Biosystems). The multiplexed samples were sequenced using 75 bp paired-end sequencing on an Illumina v4 HiSeq 2500 to a coverage depth sufficient to provide greater than 20x haploid read depth of over 85% of targeted bases in 96% of samples (approximately 80x mean haploid read depth of targeted bases). Raw sequence data from each Illumina HiSeq 2500 run were uploaded to the DNAnexus platform (43) for sequence read alignment and variant identification. In brief, raw sequence data were converted from BCL files to sample-specific FASTQ-files, which were aligned to the human reference build GRCh37.p13 with BWA-mem (44). Single nucleotide variants (SNV) and insertion/deletion (indel) sequence variants were identified using the Genome Analysis Toolkit (45, 46). Copy number variants were detected using the CLAMMS algorithm (21).

Sequence data for *LDLR*, *APOB*, and *PCSK9* were extracted from exome sequences generated at the RGC with the use of protocols described in detail in a companion publication (19). Sequence variants were annotated using SnpEff (47) and predicted LoF variants were defined as any of the following: SNVs leading to a premature stop codon, loss of a start codon, or loss of a stop codon; SNVs or indels disrupting canonical splice acceptor or donor dinucleotides; open reading frame shifting indels leading to the formation of a premature stop codon. Sequence variants

annotated by the clinical genetics database ClinVar (22) as “pathogenic” were also identified by positional intersection with exome sequence variant calls. The ClinVar database was accessed in March 2016. We considered LoF variants (previously known and novel) in *LDLR* as “predicted pathogenic” variants, and considered missense variants classified in ClinVar as “pathogenic” in *LDLR*, *APOB*, and *PCSK9* as “known pathogenic” variants. The union of “predicted pathogenic” and “known pathogenic” variants was the set of sequence variations used for all subsequent analyses.

### FH diagnostic criteria and clinical characteristics

A clinical diagnosis of FH was established using the DLCN criteria (15, 16). Relevant data were extracted from the EHR and a numerical score was assigned (see table S5). EHR data were unavailable for some of the phenotypic criteria, in particular physical exam findings (tendon xanthomata and arcus cornealis) in patients or relatives. Total scores were calculated without and with the knowledge of a functional variant in *LDLR*, *APOB*, or *PCSK9* genes. A diagnosis of FH was considered definite if the total score was greater than 8, probable if the score was 6–8, possible if the score was 3–5, and unlikely if the score was below 3 points.

We applied MEDPED criteria to EHRs in a similar manner. As MEDPED criteria rely on the knowledge of 1st, 2nd, or 3rd degree relatives of FH (17), which is not captured in the EHR, we expanded the criteria to include knowledge of any family history of hyperlipidemia, and used corresponding MEDPED age-based LDL-C thresholds for individuals with 1st degree relatives with FH (if positive family history of hyperlipidemia) or for general population (if negative family history of hyperlipidemia).

Maximum EHR-documented LDL-C values and lowest EHR-documented LDL-C values were extracted from EHRs. We extracted most recent EHR-documented LDL-C values from levels recorded in the EHR within 12 months of the study. Only outpatient LDL-C measurements are reported. Evidence of ever-prescribed lipid-lowering therapy was also extracted from the EHR. Current lipid-lowering therapy was defined as the last prescribed lipid-lowering agent documented in the EHR in patients with a medical reconciliation performed in 2015 or 2016. “High-intensity statins” were defined as high doses of specific statin medications: simvastatin 80 mg, atorvastatin 40 or 80 mg, and rosuvastatin 20 or 40 mg daily (20). Moderate-intensity statins were defined as moderate doses of specific statin medications: atorvastatin 10 or 20 mg, rosuvastatin 5 or 10 mg, simvastatin 20 or 40 mg, pravastatin 40 or 80 mg, lovastatin 40 mg, fluvastatin XL 80 mg, and pitavastatin 2 or 4 mg daily; or fluvastatin 40 mg twice daily.

### Coronary artery disease definitions

Participants were assigned a CAD case or control status using an electronic phenotyping algorithm



that is described in a previous publication (25). In brief, individuals were considered to have CAD if they had a history of coronary revascularization in the EHR, or history of acute coronary syndrome, ischemic heart disease, or exertional angina (ICD-9 codes 410\*, 411\*, 412\*, 413\*, 414\*) with angiographic evidence of obstructive coronary atherosclerosis (>50% stenosis in at least one major epicardial vessel from catheterization report). CAD controls were defined as individuals without any case criteria or any single encounter or problem list diagnosis code indicating CAD. Two case-control definitions were considered: “general CAD,” in which all individuals, regardless of age, were analyzed, and “premature CAD,” defined as cases and controls younger than 55 (for males) or 65 (for females) years of age.

### Statistical analysis

We used a mixed linear model, coding genotypes under an additive model (0 for sequenced non-carriers, 1 for heterozygotes), adjusting for age, age<sup>2</sup>, sex, and the first 5 principal components of ancestry to test for associations between predicted and known pathogenic variants in *LDLR*, *APOB*, and *PCKS9* individually and in aggregate, and maximum EHR-documented LDL-C values. ANOVA with post hoc Tukey honest significant difference tests was used to examine differences in maximum LDL-C values across *LDLR*, *APOB*, and *PCKS9* variant carriers.

We tested for associations between *LDLR*, *APOB*, and *PCKS9* variants in aggregate (coded as above) and “general” and “premature” CAD liability adjusting for age, age<sup>2</sup>, and sex using mixed linear model association analysis. Odds ratios for CAD were estimated using Firth’s penalized likelihood logistic regression (48) adjusting for age, age<sup>2</sup>, sex, and the first five principal components of ancestry. Wald 95% confidence intervals were estimated for odds ratios using standard error estimates back-calculated from p-values from the mixed linear models of association. GCTA v2.1.4 (49) and R version 3.2.1 (R Project for Statistical Computing) were used for all statistical analyses.

### REFERENCES AND NOTES

- M. S. Bowen, K. Kolor, W. D. Dotson, R. M. Ned, M. J. Khoury, Public health action in genomics is now needed beyond newborn screening. *Public Health Genomics* **15**, 327–334 (2012). doi: [10.1159/000341889](https://doi.org/10.1159/000341889); pmid: [22986915](https://pubmed.ncbi.nlm.nih.gov/22986915/)
- S. S. Gidding *et al.*, The agenda for familial hypercholesterolemia: A scientific statement from the American Heart Association. *Circulation* **132**, 2167–2192 (2015). doi: [10.1161/CIR.0000000000000297](https://doi.org/10.1161/CIR.0000000000000297); pmid: [26510694](https://pubmed.ncbi.nlm.nih.gov/26510694/)
- A. C. Goldberg *et al.*, Familial hypercholesterolemia: Screening, diagnosis and management of pediatric and adult patients: Clinical guidance from the National Lipid Association Expert Panel on Familial Hypercholesterolemia. *J. Clin. Lipidol.* **5**, S1–S8 (2011). doi: [10.1016/j.jacl.2011.04.003](https://doi.org/10.1016/j.jacl.2011.04.003); pmid: [21600525](https://pubmed.ncbi.nlm.nih.gov/21600525/)
- B. G. Nordestgaard *et al.*, Familial hypercholesterolemia is underdiagnosed and undertreated in the general population: Guidance for clinicians to prevent coronary heart disease: Consensus statement of the European Atherosclerosis Society. *Eur. Heart J.* **34**, 3478–3490 (2013). doi: [10.1093/eurheartj/ehz273](https://doi.org/10.1093/eurheartj/ehz273); pmid: [23956253](https://pubmed.ncbi.nlm.nih.gov/23956253/)
- J. L. Goldstein, H. G. Schrott, W. R. Hazzard, E. L. Bierman, A. G. Motulsky, Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J. Clin. Invest.* **52**, 1544–1568 (1973). doi: [10.1172/JCI107332](https://doi.org/10.1172/JCI107332); pmid: [4718953](https://pubmed.ncbi.nlm.nih.gov/4718953/)
- P. N. Hopkins, P. P. Toth, C. M. Ballantyne, D. J. Rader, Familial hypercholesterolemias: Prevalence, genetics, diagnosis and screening recommendations from the National Lipid Association Expert Panel on Familial Hypercholesterolemia. *J. Clin. Lipidol.* **5**, S9–S17 (2011). doi: [10.1016/j.jacl.2011.03.452](https://doi.org/10.1016/j.jacl.2011.03.452); pmid: [21600530](https://pubmed.ncbi.nlm.nih.gov/21600530/)
- D. Marks, M. Thorogood, H. A. Neil, S. E. Humphries, A review on the diagnosis, natural history, and treatment of familial hypercholesterolemia. *Atherosclerosis* **168**, 1–14 (2003). doi: [10.1016/S0021-9150\(02\)00330-1](https://doi.org/10.1016/S0021-9150(02)00330-1); pmid: [12732381](https://pubmed.ncbi.nlm.nih.gov/12732381/)
- M. Benn, G. F. Watts, A. Tybjaerg-Hansen, B. G. Nordestgaard, Mutations causative of familial hypercholesterolemia: Screening of 98 098 individuals from the Copenhagen General Population Study estimated a prevalence of 1 in 217. *Eur. Heart J.* **37**, 1384–1394 (2016). doi: [10.1093/eurheartj/ehw028](https://doi.org/10.1093/eurheartj/ehw028); pmid: [26908947](https://pubmed.ncbi.nlm.nih.gov/26908947/)
- B. Sjouke *et al.*, Homozygous autosomal dominant hypercholesterolemia in the Netherlands: Prevalence, genotype–phenotype relationship, and clinical outcome. *Eur. Heart J.* **36**, 560–565 (2015). doi: [10.1093/eurheartj/ehu058](https://doi.org/10.1093/eurheartj/ehu058); pmid: [24585268](https://pubmed.ncbi.nlm.nih.gov/24585268/)
- M. A. Austin, C. M. Hutter, R. L. Zimmern, S. E. Humphries, Genetic causes of monogenic heterozygous familial hypercholesterolemia: A HuGE prevalence review. *Am. J. Epidemiol.* **160**, 407–420 (2004). doi: [10.1093/aje/kwh236](https://doi.org/10.1093/aje/kwh236); pmid: [15321837](https://pubmed.ncbi.nlm.nih.gov/15321837/)
- H. Shen *et al.*, Familial defective apolipoprotein B-100 and increased low-density lipoprotein cholesterol and coronary artery calcification in the Old Order Amish. *Arch. Intern. Med.* **170**, 1850–1855 (2010). doi: [10.1001/archinternmed.2010.384](https://doi.org/10.1001/archinternmed.2010.384); pmid: [21059979](https://pubmed.ncbi.nlm.nih.gov/21059979/)
- H. A. Neil, T. Hammond, R. Huxley, D. R. Matthews, S. E. Humphries, Extent of underdiagnosis of familial hypercholesterolemia in routine practice: Prospective registry study. *BMJ* **321**, 148 (2000). doi: [10.1136/bmj.321.7254.148](https://doi.org/10.1136/bmj.321.7254.148); pmid: [10894692](https://pubmed.ncbi.nlm.nih.gov/10894692/)
- A. H. Pijlman *et al.*, Evaluation of cholesterol lowering treatment of patients with familial hypercholesterolemia: A large cross-sectional study in The Netherlands. *Atherosclerosis* **209**, 189–194 (2010). doi: [10.1016/j.atherosclerosis.2009.09.014](https://doi.org/10.1016/j.atherosclerosis.2009.09.014); pmid: [19818960](https://pubmed.ncbi.nlm.nih.gov/19818960/)
- Risk of fatal coronary heart disease in familial hypercholesterolemia. Scientific Steering Committee on behalf of the Simon Broome Register Group. *BMJ* **303**, 893–896 (1991). doi: [10.1136/bmj.303.6807.893](https://doi.org/10.1136/bmj.303.6807.893); pmid: [1933004](https://pubmed.ncbi.nlm.nih.gov/1933004/)
- F. Civeira; International Panel on Management of Familial Hypercholesterolemia, Guidelines for the diagnosis and management of heterozygous familial hypercholesterolemia. *Atherosclerosis* **173**, 55–68 (2004). doi: [10.1016/j.atherosclerosis.2003.11.010](https://doi.org/10.1016/j.atherosclerosis.2003.11.010); pmid: [15177124](https://pubmed.ncbi.nlm.nih.gov/15177124/)
- World Health Organization, *Familial Hypercholesterolemia: Report of a Second WHO Consultation* (World Health Organization, 1998).
- R. R. Williams *et al.*, Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. *Am. J. Cardiol.* **72**, 171–176 (1993). doi: [10.1016/0002-9149\(93\)90155-6](https://doi.org/10.1016/0002-9149(93)90155-6); pmid: [8328379](https://pubmed.ncbi.nlm.nih.gov/8328379/)
- D. J. Carey *et al.*, The Geisinger MyCode community health initiative: An electronic health record–linked biobank for precision medicine research. *Genet. Med.* **18**, 906–913 (2016). doi: [10.1038/gim.2015.187](https://doi.org/10.1038/gim.2015.187); pmid: [26866580](https://pubmed.ncbi.nlm.nih.gov/26866580/)
- F. E. Dewey *et al.*, Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, aaf6814 (2016).
- N. J. Stone *et al.*, 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* **63**, 2889–2934 (2014). doi: [10.1016/j.jacc.2013.11.002](https://doi.org/10.1016/j.jacc.2013.11.002); pmid: [24239923](https://pubmed.ncbi.nlm.nih.gov/24239923/)
- J. S. Packer *et al.*, CLAMMS: A scalable algorithm for calling common and rare copy number variants from exome sequencing data. *Bioinformatics* **32**, 133–135 (2016). doi: [10.1093/bioinformatics/btv547](https://doi.org/10.1093/bioinformatics/btv547); pmid: [26382196](https://pubmed.ncbi.nlm.nih.gov/26382196/)
- M. J. Landrum *et al.*, ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014). doi: [10.1093/nar/gkt1113](https://doi.org/10.1093/nar/gkt1113); pmid: [24234437](https://pubmed.ncbi.nlm.nih.gov/24234437/)
- A. V. Khara *et al.*, Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J. Am. Coll. Cardiol.* **67**, 2578–2589 (2016). doi: [10.1016/j.jacc.2016.03.520](https://doi.org/10.1016/j.jacc.2016.03.520); pmid: [27050191](https://pubmed.ncbi.nlm.nih.gov/27050191/)
- J. Staples, D. Qiao, M. H. Cho, E. K. Silverman, D. A. Nickerson, PRIMUS: Rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am. J. Hum. Genet.* **95**, 553–564 (2014). doi: [10.1016/j.ajhg.2014.10.005](https://doi.org/10.1016/j.ajhg.2014.10.005); pmid: [25439724](https://pubmed.ncbi.nlm.nih.gov/25439724/)
- F. E. Dewey *et al.*, Inactivating variants in *ANGPTL4* and risk of coronary artery disease. *N. Engl. J. Med.* **374**, 1123–1133 (2016). doi: [10.1056/NEJMoa1510926](https://doi.org/10.1056/NEJMoa1510926); pmid: [26933753](https://pubmed.ncbi.nlm.nih.gov/26933753/)
- J. G. Robinson, A. C. Goldberg, Treatment of adults with familial hypercholesterolemia and evidence for treatment: Recommendations from the National Lipid Association Expert Panel on Familial Hypercholesterolemia. *J. Clin. Lipidol.* **5**, S18–S29 (2011). doi: [10.1016/j.jacl.2011.03.451](https://doi.org/10.1016/j.jacl.2011.03.451); pmid: [21600526](https://pubmed.ncbi.nlm.nih.gov/21600526/)
- T. A. Jacobson *et al.*, National Lipid Association recommendations for patient-centered management of dyslipidemia: Part 2. *J. Clin. Lipidol.* **9**, S1–S122.e1 (2015). doi: [10.1016/j.jacl.2015.09.002](https://doi.org/10.1016/j.jacl.2015.09.002); pmid: [26699442](https://pubmed.ncbi.nlm.nih.gov/26699442/)
- H. Tada *et al.*, Multiple associated variants increase the heritability explained for plasma lipids and coronary artery disease. *Circ. Cardiovasc. Genet.* **7**, 583–587 (2014). doi: [10.1161/CIRCGENETICS.113.000420](https://doi.org/10.1161/CIRCGENETICS.113.000420); pmid: [25170055](https://pubmed.ncbi.nlm.nih.gov/25170055/)
- T. M. Teslovich *et al.*, Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010). doi: [10.1038/nature09270](https://doi.org/10.1038/nature09270); pmid: [20686565](https://pubmed.ncbi.nlm.nih.gov/20686565/)
- D. Vodnala, M. Rubenfire, R. D. Brook, Secondary causes of dyslipidemia. *Am. J. Cardiol.* **110**, 823–825 (2012). doi: [10.1016/j.amjcard.2012.04.062](https://doi.org/10.1016/j.amjcard.2012.04.062); pmid: [22658245](https://pubmed.ncbi.nlm.nih.gov/22658245/)
- L. Troeung *et al.*, A new electronic screening tool for identifying risk of familial hypercholesterolemia in general practice. *Heart* **102**, 855–861 (2016). doi: [10.1136/heartjnl-2015-308824](https://doi.org/10.1136/heartjnl-2015-308824); pmid: [26864669](https://pubmed.ncbi.nlm.nih.gov/26864669/)
- The FH Foundation, FIND FH Initiative, <https://theffhfoundation.org/find-fh/> [accessed 21 June 2016].
- Z. Reiner *et al.*, ESC/EAS guidelines for the management of dyslipidaemias: The Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS). *Atherosclerosis* **32**, 1769–1818 (2011). doi: [10.1016/j.atherosclerosis.2011.06.028](https://doi.org/10.1016/j.atherosclerosis.2011.06.028); pmid: [21712404](https://pubmed.ncbi.nlm.nih.gov/21712404/)
- M. Benn, G. F. Watts, A. Tybjaerg-Hansen, B. G. Nordestgaard, Familial hypercholesterolemia in the Danish general population: Prevalence, coronary artery disease, and cholesterol-lowering medication. *J. Clin. Endocrinol. Metab.* **97**, 3956–3964 (2012). doi: [10.1210/jc.2012-1563](https://doi.org/10.1210/jc.2012-1563); pmid: [22893714](https://pubmed.ncbi.nlm.nih.gov/22893714/)
- F. J. Raal *et al.*, Reduction in mortality in subjects with homozygous familial hypercholesterolemia associated with advances in lipid-lowering therapy. *Circulation* **124**, 2202–2207 (2011). doi: [10.1161/CIRCULATIONAHA.111.042523](https://doi.org/10.1161/CIRCULATIONAHA.111.042523); pmid: [21986285](https://pubmed.ncbi.nlm.nih.gov/21986285/)
- A. Andermann, I. Blancquaert, S. Beauchamp, V. Déry, Revisiting Wilson and Jungner in the genomic age: A review of screening criteria over the past 40 years. *Bull. World Health Organ.* **86**, 317–319 (2008). doi: [10.2471/BLT.00.000000](https://doi.org/10.2471/BLT.00.000000); pmid: [18435522](https://pubmed.ncbi.nlm.nih.gov/18435522/)
- L. G. Biesecker, R. C. Green, Diagnostic clinical genome and exome sequencing. *N. Engl. J. Med.* **370**, 2418–2425 (2014). doi: [10.1056/NEJMr1312543](https://doi.org/10.1056/NEJMr1312543); pmid: [24941179](https://pubmed.ncbi.nlm.nih.gov/24941179/)
- D. Wonderling *et al.*, Cost-effectiveness analysis of the genetic screening program for familial hypercholesterolemia in The Netherlands. *Semin. Vasc. Med.* **4**, 97–104 (2004). doi: [10.1055/s-2004-822992](https://doi.org/10.1055/s-2004-822992); pmid: [15199439](https://pubmed.ncbi.nlm.nih.gov/15199439/)
- C. X. Chen, J. W. Hay, Cost-effectiveness analysis of alternative screening and treatment strategies for heterozygous familial hypercholesterolemia in the United States. *Int. J. Cardiol.* **181**, 417–424 (2015). doi: [10.1016/j.ijcard.2014.12.070](https://doi.org/10.1016/j.ijcard.2014.12.070); pmid: [25569270](https://pubmed.ncbi.nlm.nih.gov/25569270/)
- E. A. Ashley, The precision medicine initiative: A new national effort. *JAMA* **313**, 2119–2120 (2015). doi: [10.1001/jama.2015.3595](https://doi.org/10.1001/jama.2015.3595); pmid: [25928209](https://pubmed.ncbi.nlm.nih.gov/25928209/)
- M. J. Khoury, J. P. Evans, A public health perspective on a national precision medicine cohort: Balancing long-term knowledge generation with early health benefit. *JAMA* **313**, 2117–2118 (2015). doi: [10.1001/jama.2015.3382](https://doi.org/10.1001/jama.2015.3382); pmid: [26034952](https://pubmed.ncbi.nlm.nih.gov/26034952/)
- R. C. Green *et al.*, ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013). doi: [10.1038/gim.2013.73](https://doi.org/10.1038/gim.2013.73); pmid: [23788249](https://pubmed.ncbi.nlm.nih.gov/23788249/)

43. J. G. Reid *et al.*, Launching genomics into the cloud: Deployment of Mercury, a next generation sequence analysis pipeline. *BMC Bioinformatics* **15**, 30 (2014). doi: [10.1186/1471-2105-15-30](https://doi.org/10.1186/1471-2105-15-30); pmid: [24475911](https://pubmed.ncbi.nlm.nih.gov/24475911/)
44. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324); pmid: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
45. A. McKenna *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110); pmid: [20644199](https://pubmed.ncbi.nlm.nih.gov/20644199/)
46. G. A. Van der Auwera *et al.*, From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013). doi: [10.1002/0471250953.bi1110s43](https://doi.org/10.1002/0471250953.bi1110s43); pmid: [25431634](https://pubmed.ncbi.nlm.nih.gov/25431634/)
47. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w<sup>1118</sup>*; iso-2; iso-3. *Fly* **6**, 80–92 (2012). doi: [10.4161/fly.19695](https://doi.org/10.4161/fly.19695); pmid: [22728672](https://pubmed.ncbi.nlm.nih.gov/22728672/)
48. G. Heinze, M. Schemper, A solution to the problem of separation in logistic regression. *Stat. Med.* **21**, 2409–2419 (2002). doi: [10.1002/\(ISSN\)1097-0258](https://doi.org/10.1002/(ISSN)1097-0258); pmid: [12210625](https://pubmed.ncbi.nlm.nih.gov/12210625/)
49. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). doi: [10.1016/j.ajhg.2010.11.011](https://doi.org/10.1016/j.ajhg.2010.11.011); pmid: [12210625](https://pubmed.ncbi.nlm.nih.gov/12210625/)

#### ACKNOWLEDGMENTS

We would like to thank the MyCode Community Health Initiative participants for their permission to use their health and genomics information. We would also like to thank K. Wilemon and K. Myers from the Familial Hypercholesterolemia Foundation for input on the manuscript. The study was funded by Regeneron Pharmaceuticals. In addition to employment by Regeneron Pharmaceuticals, G.D.Y. is a cofounder, member

of the board of directors, and stockholder in Regeneron Pharmaceuticals. Pathogenic and predicted pathogenic variants reported in this paper are tabulated in table S2. All variants will be clinically confirmed by the Laboratory of Molecular Medicine (Partners Healthcare) and deposited into ClinVar ([www.ncbi.nlm.nih.gov/clinvar/](http://www.ncbi.nlm.nih.gov/clinvar/)). At the time of publication, 11 of the 35 reported variants had been deposited into ClinVar. Additional information for reproducing and extending the results described in the article is available under a data use agreement with Regeneron Genetics Center LLC and Geisinger Clinic.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/354/6319/aaf7000/suppl/DC1](http://www.sciencemag.org/content/354/6319/aaf7000/suppl/DC1)  
Fig. S1  
Tables S1 to S8

17 March 2016; accepted 16 November 2016  
10.1126/science.aaf7000





EXTENDED PDF FORMAT  
SPONSORED BY



## Genetic identification of familial hypercholesterolemia within a single U.S. health care system

Noura S. Abul-Husn, Kandamurugu Manickam, Laney K. Jones, Eric A. Wright, Dustin N. Hartzel, Claudia Gonzaga-Jauregui, Colm O'Dushlaine, Joseph B. Leader, H. Lester Kirchner, D'Andra M. Lindbuchler, Marci L. Barr, Monica A. Giovanni, Marylyn D. Ritchie, John D. Overton, Jeffrey G. Reid, Raghu P. R. Metpally, Amr H. Wardeh, Ingrid B. Borecki, George D. Yancopoulos, Aris Baras, Alan R. Shuldiner, Omri Gottesman, David H. Ledbetter, David J. Carey, Frederick E. Dewey and Michael F. Murray  
(December 22, 2016)  
*Science* **354** (6319), . [doi: 10.1126/science.aaf7000]

Editor's Summary

### Unleashing the power of precision medicine

Precision medicine promises the ability to identify risks and treat patients on the basis of pathogenic genetic variation. Two studies combined exome sequencing results for over 50,000 people with their electronic health records. Dewey *et al.* found that ~3.5% of individuals in their cohort had clinically actionable genetic variants. Many of these variants affected blood lipid levels that could influence cardiovascular health. Abul-Husn *et al.* extended these findings to investigate the genetics and treatment of familial hypercholesterolemia, a risk factor for cardiovascular disease, within their patient pool. Genetic screening helped identify at-risk patients who could benefit from increased treatment.

*Science*, this issue p. 10.1126/science.aaf6814, p. 10.1126/science.aaf7000

---

This copy is for your personal, non-commercial use only.

---

#### Article Tools

Visit the online version of this article to access the personalization and article tools:  
<http://science.sciencemag.org/content/354/6319/aaf7000>

#### Permissions

Obtain information about reproducing this article:  
<http://www.sciencemag.org/about/permissions.dtl>

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.