Received: 25 May 2020 | Revised: 31 July 2020 | Accepted: 3 August 2020

DOI: 10.1111/cge.13826

ORIGINAL ARTICLE

CLINICAL GENETICS WILEY

Mutation spectrum and polygenic score in German patients with familial hypercholesterolemia

Lorenz Rieck[1] | Frieda Bardey[1] | Thomas Grenkowitz[2] | Lars Bertram[3,4] | Johannes Helmuth[5] | Claudia Mischung[5] | Joachim Spranger[1] | Elisabeth Steinhagen-Thiessen[1] | Thomas Bobbert[1] | Ursula Kassner[1] | Ilja Demuth[1,6]

[1]Department of Endocrinology and Metabolism, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[2]Department of Cardiology, Charité - University Medicine Berlin (Campus Benjamin Franklin), Berlin, Germany

[3]Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and Cardiogenetics, University of Lübeck, Lübeck, Germany

[4]Center for Lifespan Changes in Brain and Cognition (LCBC), Dept of Psychology, University of Oslo, Oslo, Norway

[5]Department Molecular Diagnostics, Labor Berlin – Charité Vivantes GmbH, Berlin, Germany

[6]Charité - Universitätsmedizin Berlin, BCRT - Berlin Institute of Health Center for Regenerative Therapies, Berlin, Germany

Correspondence
Ilja Demuth PhD, Charité - Universitätsmedizin Berlin, Department of Endocrinology and Metabolism, Biology of Aging Group, Augustenburger Platz 1, Berlin 13353, Germany.
Email: ilja.demuth@charite.de

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Numbers: 01GL1716A, 01GL1716B, 01UW0808, 16SV5536K, 16SV5537, 16SV5538, 16SV5837; Max-Planck-Gesellschaft; Charité Universitätsmedizin Berlin

Abstract

Autosomal-dominant familial hypercholesterolemia (FH) is characterized by increased plasma concentrations of low-density lipoprotein cholesterol (LDL-C) and a substantial risk to develop cardiovascular disease. Causative mutations in three major genes are known: the *LDL receptor* gene (*LDLR*), the *apolipoprotein B* gene (*APOB*) and the *proprotein convertase subtilisin/kexin 9* gene (*PCSK9*). We clinically characterized 336 patients suspected to have FH and screened them for disease causing mutations in *LDLR*, *APOB*, and *PCSK9*. We genotyped six single nucleotide polymorphisms (SNPs) to calculate a polygenic risk score for the patients and 1985 controls. The 117 patients had a causative variant in one of the analyzed genes. Most variants were found in the *LDLR* gene (84.9%) with 11 novel mutations. The mean polygenic risk score was significantly higher in FH mutation negative subjects than in FH mutation positive patients ($P < .05$) and healthy controls ($P < .001$), whereas the score of the two latter groups did not differ significantly. However, the score explained only about 3% of the baseline LDL-C variance. We verified the previously described clinical and genetic variability of FH for German hypercholesterolemic patients. Evaluation of a six-SNP polygenic score recently proposed for clinical use suggests that it is not a reliable tool to classify hypercholesterolemic patients.

Ursula Kassner and Ilja Demuth contributed equally to this study.

1 | INTRODUCTION

Autosomal-dominant familial hypercholesterolemia (FH, OMIM 143890) is a common genetic disorder with a prevalence of up to 1:250 in the European population and an even higher prevalence in some populations due to founder effects. In its typical form, FH is characterized by increased plasma concentrations of low-density lipoprotein cholesterol (LDL-C) and a high risk to develop cardiovascular disease (CVD).[1,2] A widely used tool in Europe for the clinical diagnosis of FH is the Dutch lipid clinic network (DLCN) score which is a point-based score system to assess the patients according to their clinical phenotypes considering the plasma LDL-C concentration, the family and/or patients history of premature CVD, clinical signs like tendon xanthoma and/or premature corneal arcus, and the results of molecular genetic analysis. According to the DLCN score in patients with a score below 3 and between 3 to 5 a diagnosis of FH is either *unlikely* or *possible*, whereas those with a score of 6 to 8 and above 8 have a *probable* or *definite* FH diagnosis, respectively.[2] Furthermore, mutations in three genes are known to cause FH: the *LDL receptor* gene (*LDLR*), the *apolipoprotein B* gene (*APOB*) and the *proprotein convertase subtilisin/kexin 9* gene (*PCSK9*).[3-5] Mutations in the *LDL receptor adapter protein 1* gene (*LDLRAP1*) cause a very rare autosomal-recessive (AR) form of the disease, that is, AR hypercholesterolemia with a FH-like clinical phenotype.[6] The majority of FH causing variants have been found in the *LDLR* gene with currently over 1900 reported sequence alterations (https://databases.lovd.nl/shared/genes/LDLR).[7] This corresponds with findings in the German population where the most common FH causing variants are found in *LDLR*, mostly of the missense type.[8] The independent CVD risk factor lipoprotein (a) is important to take into consideration when discussing treatments available in the near future and risk stratification for hypercholesterolemic patients.[9]

In about 60% to 80% of the definite FH cases, a causative mutation can be found in one of the three known genes, but only in about 20% to 30% of *possible* FH cases.[10,11] Since the unequivocal diagnosis of FH is primarily based on molecular testing, this leaves medical scientists with a conundrum because up to 60% of the clinically diagnosed patients turn out to be nonmutation carriers, that is, with no identified pathogenic sequence variant in one of the above mentioned genes.[12] This diagnostic gap might be explained by several factors: additional FH causing gene loci that have yet to be identified, a multifactorial polygenic genetic cause of the disease and the presence of epigenetic modifications. To this end, recent studies have evaluated a possible fourth FH causing gene, the *APOE* variant c.500_502delTCC (p.Leu167del), and found evidence for a potential disease causing role.[13] Furthermore *STAP1* was suggested to be associated with FH, however, we were not able to confirm this in a recent study on 75 hypercholesterolemic patients from Berlin, Germany, who were negative for mutations in canonical FH genes.[14,15] In a recent study on mouse models and samples from *STAP1* carriers the gene was ruled out as a FH causing candidate.[16] In 2010, a meta-analysis of genome-wide association studies identified multiple common single nucleotide polymorphisms (SNPs) to be strongly associated with elevated LDL-C concentrations.[17] Talmud et al hypothesized that the mutation negative FH patients might carry a greater-than-average number of these common LDL-C raising variants and assembled a 12-SNP polygenic score based on the variants with the strongest LDL-C associations. They found the score to be significantly higher when compared to the score of Whitehall II control cohort or to FH patients carrying a disease causing mutation ("mutation positive").[11] Futema et al successfully refined the score by selecting six SNPs from the original 12-SNP score, and were able to replicate the earlier findings, that is, the six-SNP polygenic score performed as well as the 12-SNP score suggesting that mutation negative FH patients have a significantly higher score when compared to mutation positive patients or controls. The authors concluded that hypercholesterolemia in almost 90% of the mutation negative patients has a polygenic basis.[18]

In the current study we analyzed the mutational spectrum and clinical phenotypes in 336 German patients with hypercholesterolemia. Furthermore, we determined the six-SNP polygenic score as proposed by Futema and colleagues[18] in these subjects and compared these to a control group of almost 2000 individuals from the Berlin Aging Study II (BASE-II),[19] assessed its impact on LDL-C levels, and discussed its putative clinical benefit.

2 | MATERIALS AND METHODS

2.1 | Subjects and recruitment

We included 336 patients ascertained between 2016 and 2019 in the specialized Lipid Clinic of the Department of Endocrinology and Metabolism, Charité - Universitätsmedizin Berlin, Germany diagnosed with hypercholesterolemia. All participants gave written informed consent prior to participation.

As a control group, we evaluated data from the BASE-II,[19] which is a multi-institutional and multidisciplinary study consisting of 2200 predominantly healthy individuals from the Berlin metropolitan area of which 1946 participants had information about both baseline LDL-C levels and genotypes necessary to calculate the individual SNP score. All participants gave written informed consent and the study was approved by the Ethics Committee of the Charité-Universitätsmedizin Berlin (approval number EA2/029/09).

2.2 | Clinical diagnostics

The hypercholesterolemic patients were assessed by a physician of the Charité Lipid Clinic. All patients underwent comprehensive patient history including family history for lipid disorders and premature CVD and physical examination. Standard laboratory parameters including LDL-C, total cholesterol (TC), HDL-cholesterol (HDL-C), triglycerides (TG) and lipoprotein (a) (Lp [a]) were determined. When native LDL-C was not available, we calculated it as described previously.[20-28] We assembled the clinical information to compute the individual DLCN score, which is a point-based tool to simplify the clinical diagnostic of

FH patients.[29] The score classifies each individual and proposes an FH diagnosis to be *unlikely* (score < 3), *possible* (score 3-5), *probable* (score 6-8), or *definite* (score > 8). Discriminatory power to distinguish between FH-mutation carriers and subjects without a FH causing variation was evaluated by calculating the area under the receiver operating characteristics (ROC) curve. LDL-C serum levels are provided in mg/dL (1 mmol/L = 38.66 mg/dL).

2.3 | Mutation screening

DNA from 336 Ethylenediaminetetraacetic acid blood samples was extracted from whole-blood samples using the QiaCube Kit (QIAgen, Hilden, Germany), following the manufacturer's protocol and molecular genetic analysis were based on next generation sequencing (NGS) technology. Further information on the mutation screening procedure is displayed in the supplements (supplementary methods).

We compared all detected sequence variants with the Human gene mutation database (HGMD),[30] Leiden open variation database,[7] Clinvar variant database[31] and their frequency in the gnomAD database (https://gnomad.broadinstitute.org). All sequence variants were evaluated with respect to their potential molecular function using the prediction tools PolyPhen2, Mutation Taster, and Human Splicing Finder (the latter for variants with a possible effect on splicing such as intronic variants).[32-34] We rated the mutations following the American college of medical genetics and genomics (ACMG) standards and guidelines[35] and all novel variants were rated according to the most recent refinement of the ACMG guidelines by Nykamp et al.[36] No allele specific analysis was performed in putative compound heterozygous patients.

2.4 | SNP Genotyping

The six SNPs used to calculate the SNP score are listed in Table S1 and were originally published by Futema et al (2015). The genotypes for rs1367117, rs429358 and rs7412 were available for the hypercholesterolemic patients from the NGS-based data.

The SNPs rs629301, rs6511720, and rs6544713 (proxy for rs4299376, $r2 = 1.0$ in European populations (CEU, ie, Utah Residents with Northern and Western European Ancestry) based on (https://ldlink.nci.nih.gov/)) were genotyped employing the ABI Prism SNaPshot Multiplex Kit (Applied Biosystems) and following the manufacturer's protocol with a few modifications. All polymerase chain reactions (PCRs) were consistently performed using either the GeneAmp PCR System 9700 (Applied Biosystems) or ProFlex PCR System (Applied Biosystems) depending on sample input amount. To measure the SNaPshot-Reaction a 3730 DNA Analyzer XL instrument (Applied Biosystems, HITACHI) was used. We designed three primer triplets for each SNP consisting of two PCR-Primers and a SNaPshot-Primer specific to detect the SNP in question using the tool available at https://www.eurofinsgenomics.eu/de. The PCR was performed separately for each SNP due to the differences in annealing temperatures of the PCR primers; the SNaPshot-Reaction was performed as a multiplex reaction. Before carrying out the genotyping with the patient samples, we optimized all reactions using control DNA. Comprehensive primer data are given in Table S2. Experimental PCR and SNaPshot conditions used for genotyping are available upon request.

Procedures for generating SNP data in the BASE-II dataset are available in the supplementary material methods section.

2.5 | Data analysis

For the statistical analysis we used the International Business Machines Corporation (IBM) statistical package for the Social Sciences version 24.0 [IBM Statistical Product and Service Solutions (SPSS) Statistics for Windows, Armonk, NY: IBM Corp.]. Graphs were designed using GraphPad Prism 7, MS Excel 2016 and SPSS version 24.0. Sequencing analysis was completed using GeneMapper 5 (ThermoFisher Scientific) software. Cutoff for peak height was defined 100 for homozygous and 150 for heterozygous peaks. Statistical significance was defined as $P < .05$.

If one or two genotypes for the SNPs were missing, we assumed a calculated risk of zero for that particular SNP. If more than two genotypes were missing in a subject, this individual was excluded. The repetition of calculations using the most common genotype or excluding the all samples with missing genotypes from the analysis did not lead to significant differences (data not shown).

3 | RESULTS

3.1 | Patient characteristics

In 44 patients of the 336 patients analyzed we detected sequence variants in the canonical FH genes, which were rated as *variants of uncertain significance* (VUS) when applying the ACMG guidelines (Table S3 showing all VUS, that is, also patients carrying pathogenic and VUS rated variants). Since there is currently no certainty to rate their effect on the corresponding protein, we excluded them from all statistical analysis (information on excluded subjects is provided in Figure S1). This resulted in a cohort of 292 hypercholesterolemic subjects of whom 273 were unrelated based on the available clinical information. The repetition of the calculations excluding related patients did not lead to significant differences in the results (data not shown). The mean age of this cohort was 56.1 ± 14.3 years with 191 (65.4%) females and 101 (34.6%) males. The mean baseline LDL-C level was 278.4 ±99.2 mg/dL. Patients with an identified FH causing mutation are henceforth referred as FH/M+ and patients without a pathogenic FH variant are referred as FH/M-.

Our control cohort from BASE-II consisted of 2171 healthy individuals with 1127 (51.9%) females and 1044 (48.1%) males. The mean LDL-C level was 122.72±36.67 mg/dL was available in 2116 cases with the median age of 67.2 years.

TABLE 1 Baseline characteristics of the 292 FH patients and 2171 BASE-II controls

Variable	FH-cohort	BASE-II
	N (%)	N (%)
Female	191(65.4)	1127(51.9)
Male	101(34.6)	1044(48.1)
	Mean (+ - SD)	Mean (+ - SD)
Age (years)[a]	56.1(14.3)	59.6(17.2)
SNP-Score	0.61364(0.181669)	0.57797(0.247789)
Baseline LDL-C (mg/dl)	278.4(99.2)	122.7(36.7)
DLCN-Score	5.2(2.9)	/
Maximum TC (mg/dl)[b]	211.3(71.4)	/
LDL-C (mg/dl)[b]	140.7(68.6)	/

[a]At the time of study recruitment.
[b]On maximum therapy.
Abbreviations: DLCN, Dutch lipid clinic network; FH, familial hypercholesterolemia; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol.

The baseline characteristics for both cohorts are shown in Table 1/S4.

3.2 | Mutation spectrum

One hundred and seventeen of all 336 patients (34.8%, 19 related) were found to carry an FH causing mutation. Overall, we identified 73 different sequence variants including the eleven novel variants all of which were rated *likely pathogenic* or *pathogenic* by ACMG Standard Guidelines (Table S5). In 44 patients, we discovered mutations that were rated VUS according to ACMG criteria (Table S3). As expected, most of the mutations were located in the *LDLR* gene (107 mutations, 84.9%). Nineteen (15.1%) of the disease causing variants were found in the *APOB* gene, and no disease causing variants were found in the *PCSK9* gene or the *LDLRAP1* gene. The highest proportion of variants (N = 56, 44.4%) were of the missense type with the most numerous sequence alteration being c.798 T > A (p.Asp266Glu) in the *LDLR* gene. The most frequent mutation, c.10580G > A (p.Arg3527Gln) was located in the *APOB* gene (18 cases, 14.3%) cases (Figure 1A). We identified 10 patients with seven different large deletions/duplications which amount to 8% of all detected variants and are shown in Table S5. The distribution of the variants in the *LDLR* gene and their relative location is presented in Figure S2.

We identified four putatively compound heterozygous, two double heterozygous and one both double and possibly compound heterozygous case in our study. With 13 out of 16 alterations, the *LDLR* mutations were the most prominent. The mean baseline LDL-C level in the double heterozygous patients was 300.53± 104.73 mg/dL and substantially (and significantly, see below) lower than in the compound heterozygous subjects (408.5±233.7 mg/dL), which resulted in less distinct phenotypes for the double heterozygous group displayed by lower mean DLCN scores (6.3 ±2.88 double heterozygous; 8.0±6.27 compound heterozygous). Interestingly, we found four apparently unrelated cases of compound heterozygous patients with identical mutations. In one case with an additional mutation c.10580G > A (p.Arg3527Gln) in the *APOB* gene. Also, of interest was the fact that the mean of baseline LDL-C for these double and/or compound heterozygote patients was only second highest with 362.23± 185.19 mg/dL after subjects with single mutations that led to a truncated protein (366.85± 106.87 mg/dL). That could be explained by the fact that there were only seven patients with such a condition. The variants c.1690A > C and c.2393_2401delTCCT CGTCT are present in four patients suggesting a possible linkage. However, no specific analysis such as a segregation analysis or other testing for allelic independence was conducted to explore these possibilities. In one younger patient carrying these two variants putatively in a compound heterozygous state we observed a very mild phenotype with baseline LDL-C levels around 150 mg/dL. Detailed characteristics in Table S6.

3.3 | Novel mutations

In the current study we found 11 sequence variants that were previously not described as FH causing mutations in 14 patients. All of these were located within the *LDLR* gene (Figure 1B, Table S7), and all were classified to be disease causing by the used in silico tools, except loss of function variants like nonsense and frameshift mutations which were directly considered as pathogenic variants according to the most recent refinement of ACMG guidelines.[35] Two of the novel disease causing variants, c.1988-50_2007del70bp (p.Gly663Aspfs*47) and c.940 + 1_940 + 4delGTGA (p.), were identified within intronic regions of the *LDLR* gene. Six variants were rated to be *likely pathogenic* and five to be *pathogenic*. All of them were absent from the gnomAD and HGMD databases. The mean baseline LDL-C level for the patients carrying the novel variants was 357.41±157.88 mg/dL, that is, similar to levels in patients with established FH causing mutations.

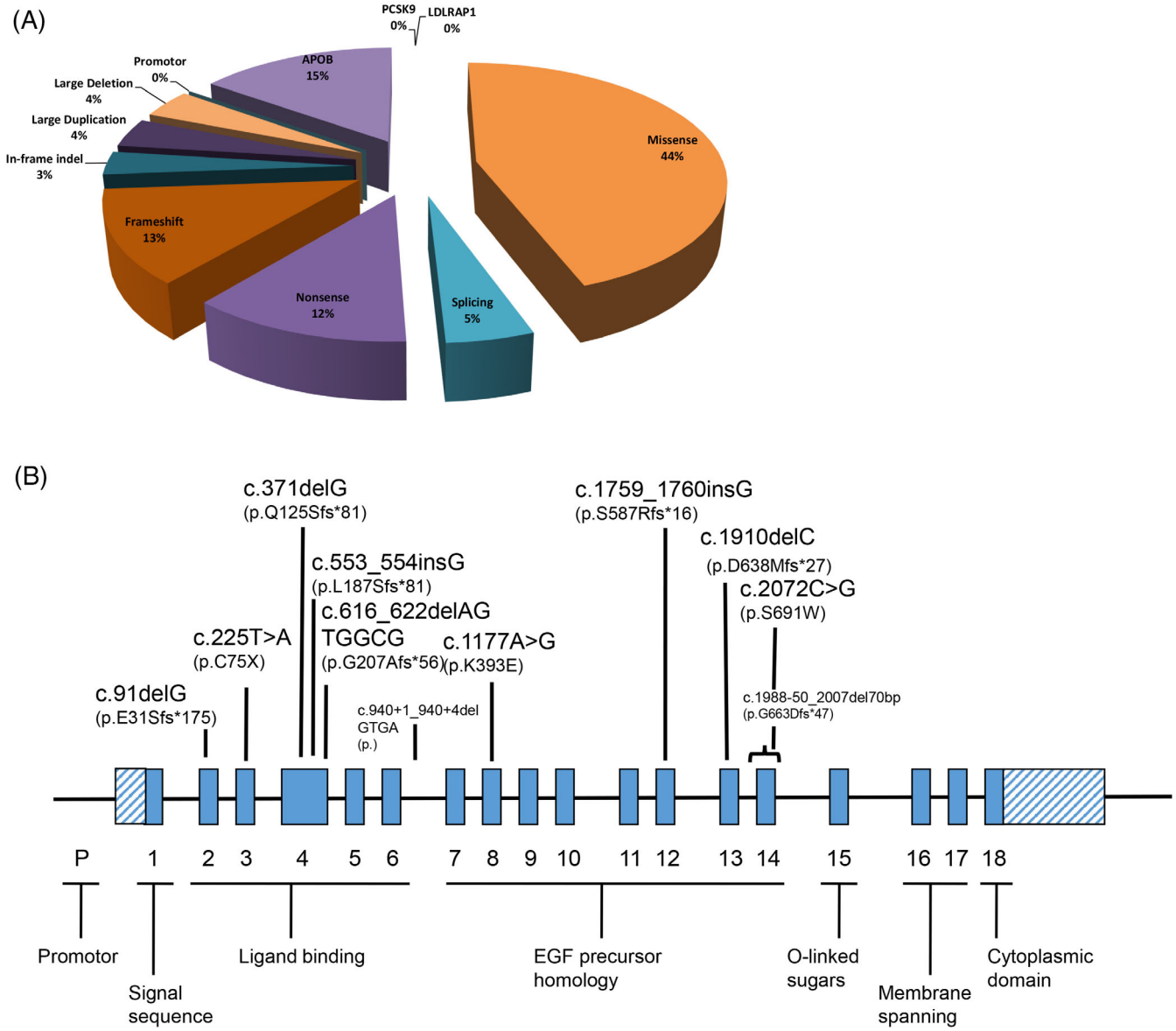

FIGURE 1 Mutation spectrum and novel sequence variants (A) Distribution of the different mutation types among the 117 FH patients. Variants c.13480_13482delCAGhet (p.Gln2294del) and c.10580G > A (p.Arg3527Gln) are located in the *APOB* gene. No pathogenic mutations were found in *PCSK9* and *LDLRAP1* gene. (B) Schematic presentation of novel sequence variants according to their relative positions in the *LDLR* gene. Blue boxes show the exonic regions and below the gene, the different functional protein domains are indicated. LDL-C, low-density lipoprotein cholesterol [Colour figure can be viewed at wileyonlinelibrary.com]

3.4 | Clinical vs molecular FH diagnosis

We were able to compute the DLCN score in 226 of 292 hypercholesterolemic patients where either a FH causing mutation was found or no mutation was identified. In patients with a DLCN score below 3 or between 3 to 5, (corresponding to an *unlikely* or *possible* FH diagnosis), 81.8% and 67.6%, respectively belonged to the group where no disease causing mutation was found. As expected, the proportion of mutation negative patients was considerably lower in the groups with higher DLCN score, with 39.1% and 17.9% mutation negatives in patients with a DLCN score between 6 and 8 and > 8, respectively (Figure 2A). When calculating the ROC curve for both the DLCN score AUC = 0.76 (95% CI = 0.69-0.82, $P < .001$) and baseline LDL-C AUC = 0.77 (95% CI = 0.71-0.83, $P < .001$) there was only a minor difference in discriminatory power. The corresponding ROC curves are shown in Figure S3. This demonstrates that both the DLCN score and the baseline LDL-C discriminate well between FH/M- and FH/M+ patients. As expected when analyzing the discriminatory power of baseline LDL-C to distinguish between BASE-II controls and FH/M+ or FH/M- patients the ROC curve was almost ideal with an AUC = 0.99 (95% CI = 0.98-0.99, $P < .001$) and AUC = 0.95 (95% CI = 0.93-0.97, $P < .001$), respectively. The prevalence of CVD in mutation positive patients was 78%. With respect to risk and positive family history of CVD, which is mainly portrayed by the DLCN score there is a

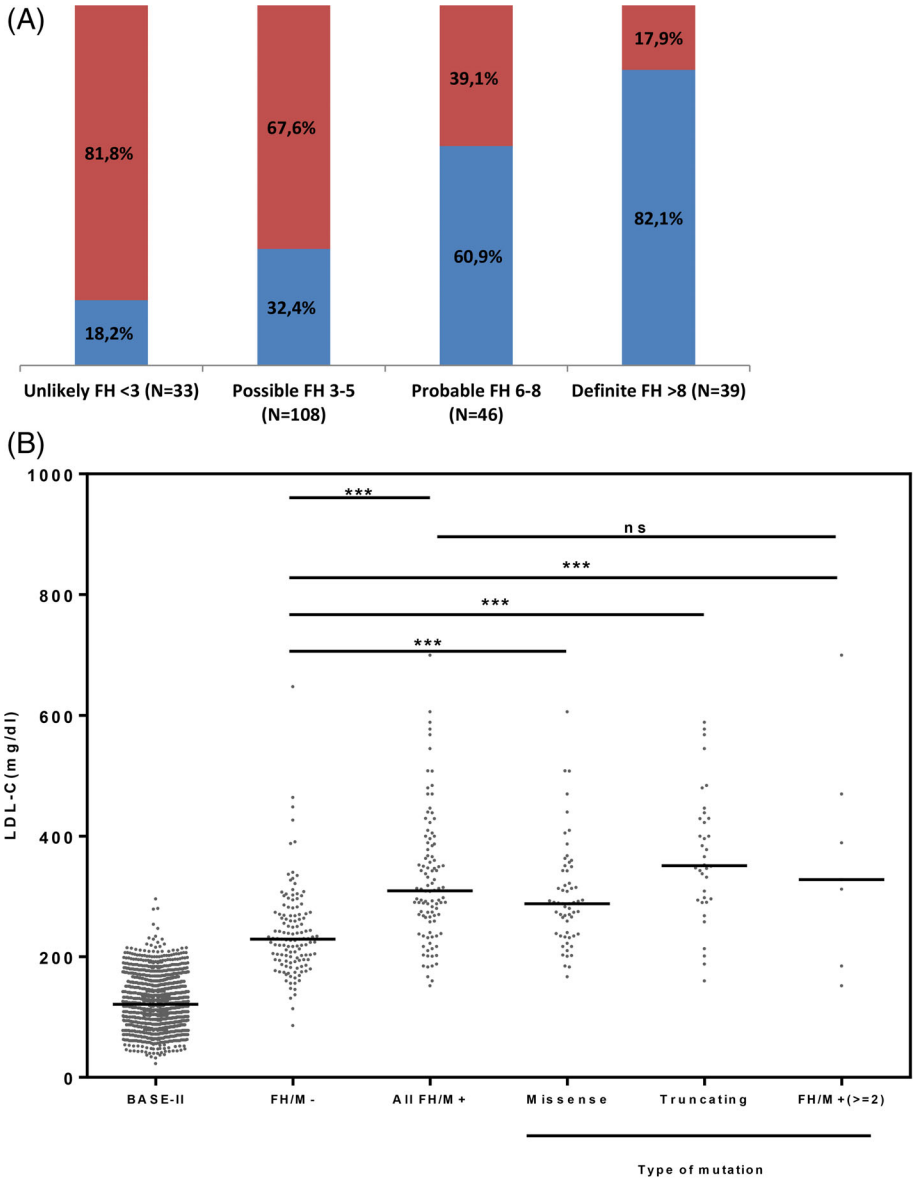

FIGURE 2 DLCN score and mutation status (A) Proportions of identified disease causing sequence variants according to DLCN score. The DLCN score was available for 226 out of 292 hypercholesterolemic patients. The graph displays the proportion of FH mutation negative patients (red column) compared to the proportion of patients with an identified FH causing mutation (blue column) in relation to the different DLCN score intervals. (B) LDL-C serum levels according to mutation status. LDL-C concentrations were compared between the BASE-II cohort (used here as controls) and different subgroups of the cohort of hypercholesterolemic patients: patients with no identified mutation (FH/M-), all different types of mutations (FH/M+), with missense variants, truncating mutations (including nonsense, and frameshift) and all patients who are compound and/or double heterozygous (or heterozygotes) with two or more sequence variations either in the same or a different gene (FH/M+[≥2]). Medians are indicated by black lines. Statistical analysis was performed by one way ANOVA and post hoc Tukey's test: ***$P < .001$, ns = not significant. DLCN, Dutch lipid clinic network; FH, familial hypercholesterolemia; LDL-C, low-density lipoprotein cholesterol [Colour figure can be viewed at wileyonlinelibrary.com]

significant difference between mutation positive and mutation negative patients (data not shown).

In addition, we compared the baseline LDL-C of our FH/M- patients with FH/M+ patients and the different types of mutations using the ANOVA and post hoc Tukey test to identify significant differences in mean LDL-C levels. This revealed a significant difference ($P < .001$) between FH/M- patients and all other subgroups. Furthermore, there was no significant difference ($P = 0.461$) between all mutation positive patients and the double and/or compound heterozygous group (Figure 2B).

Additionally, we compared the baseline LDL-C of FH/M- patients and subjects carrying variants with the different ACMG classifications. As expected, Tukey post hoc analysis revealed a significant difference ($P < .001$) between BASE-II controls and all other groups. The test revealed a significant difference ($P < .001$) between the FH/M- subgroup and group 5 which represented the patients with ACMG rated *pathogenic / likely pathogenic* mutations. Between FH/M- patients and the subjects with VUS (III) there was no significant difference ($P = .981$) (Figure S4). However, three patients carrying VUS (III) had a baseline LDL-C > 400 mg/dL (Table S3).

3.5 | Six-SNP LDL-C polygenic score

The LDL-C polygenic score was calculated from 336 hypercholesterolemic patients and from 1985 individuals of the BASE-II cohort. When comparing the mean SNP score of BASE-II controls with FH/M+ and FH/M- subjects, the highest score, 0.639(±0.164), was found in the FH/M- patients. Interestingly, the LDL-C polygenic score from the FH/M+ subjects, 0.573(±0.2), did not differ significantly from the "healthy" BASE-II controls 0.578(±0.248), t (137.7) = –0.282, $P = 0.778$. There was a statistically significant difference between the scores of the BASE-II group and the FH/M- group, with a mean

difference of 0.061 (95%-CI [0.035, 0.088]) lower for the BASE-II controls, t (250.177) = 4.520, $P < .001$. A scatterplot of these results can be seen in Figure 3.

Additionally, we assigned the BASE-II controls and both, FH/M- and FH/M+ patients, into three groups excluding individuals from BASE-II with a baseline LDL-C > 155 mg/dL and from the cohort of patients with baseline LDL-C < 155 mg/dL (the cutoff for the DLCN score). This allowed us to evaluate how well the score discriminates between "affected" (>155 mg/dL) and "nonaffected" (<155 mg/dL) individuals. The Area under the receiver operating characteristic statistic method indicated that the SNP score does not discriminate well between BASE-II controls and FH/M+ subjects (AUC = 0.50 [95% CI = 0.45-0.54, $P = 0.8971$]), BASE-II controls and FH/M- subjects (AUC = 0.60 [95% CI = 0.56-0.64, $P < .001$]) and between FH/M+ and FH/M- subjects (AUC = 0.62 [95% CI = 0.55-0.69, $P < .001$]). The corresponding ROC curves are shown in Figure S5. When computing the Youden index to the ROC curve of BASE-II controls and FH/M- patients, we were able to determine a cutoff at which the LDL-C SNP score discriminates best between the two groups. This allowed us to retrieve the basic quality criteria in a diagnostic test. After adjusting the positive predictive value for the disease prevalence of FH using Bayesian statistics, it was 0.57%. According to this finding the SNP score would allow 0.701 (less than one patient) correct positive diagnoses among the 123 mutation negative subjects with baseline LDL-C above 155 mg/dL. A "correct diagnosis" in this case signifies that these individuals were correctly assigned to the group in which the (elevated and likely disease causing baseline) LDL-C levels can be explained by a polygenic etiology derived from the SNP score. Information is shown in Table S8.

To analyze the association of the SNP score with baseline LDL-C in the control group and in our mutation negative (FH/M-) patients we computed a linear regression model. Despite being statistically significant, the score only explained a small fraction of the variance in baseline LDL-C (ß = 34.769, $P < .001$, $R^2 = .031$[$R_{adj}^2 = .030$]). In a next analysis step, we considered sex and age along with the SNP score as independent variables in the regression model. This model provided a better prediction of baseline LDL-C (F [3.20] = 76.26, $P < .001$, $R^2 = .099$[$R_{adj}^2 = .098$]). Two of the three variables except sex ($P = .542$) added statistically significant to the model ($P < .001$). Interestingly, this revealed that the most robust association of baseline LDL-C was observed with age followed by the SNP score and sex (Table 2). These findings indicate that when combining the SNP score with the covariate age up to 10% of the variance of LDL-C levels can be explained. No significant correlation was detected between the different groups of the DLCN score and the SNP score (data not shown).

4 | DISCUSSION

In this study, we analyzed the clinical FH phenotype and the mutational status of 336 patients from Berlin, Germany. In addition, we genotyped six SNPs and calculated a previously proposed six-SNP polygenic score.[18] The applied BASE-II dataset comprising N = 1985 individuals, served as a control group in the current study. To the best of our knowledge, this is the first report systematically evaluating the polygenic etiology in German hypercholesterolemic patients.

We found 73 different FH causing sequence variants including 11 novel mutations in 117 patients. Most of the identified variants were located in the *LDLR* gene and the majority of these were of the missense type. No disease causing variants were identified in the genes *PCSK9* and *LDLRAP1*. Similar to previous research[37] we found disease causing variants in the *LDLR* gene to be clustered in exons

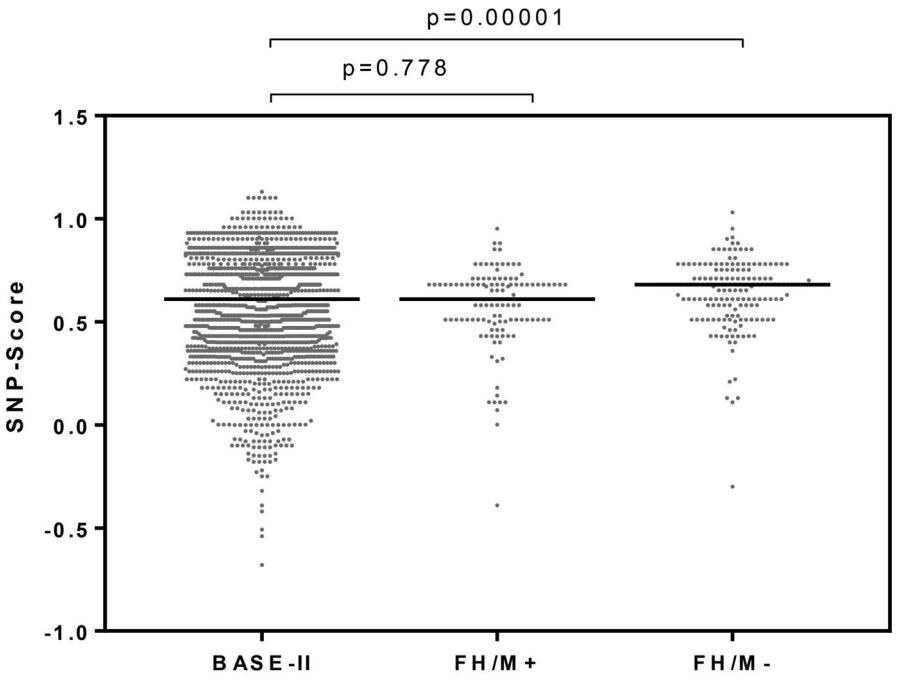

FIGURE 3 Six-SNP polygenic score in BASE-II controls and subgroups of FH patients. The six-SNP polygenic score was available for 1985 individuals in BASE-II controls, 117 patients in the FH/M+ (i.e., individuals carrying an FH causing mutation) group and 175 patients in the FH/M- (no FH causing mutation) group. Median LDL-C values are indicated by black lines. Unpaired *t* test revealed significant differences between BASE-II and FH/M- ($P < .001$). No statistically significant difference was found between BASE-II and FH/M+.FH, familial hypercholesterolemia; LDL-C, low-density lipoprotein cholesterol

TABLE 2 Results of linear regression analysis of LDL-C values on the six-SNP polygenic score adjusted for sex and age (years)

Covariates	ß	Standardized ß	SE	95% CI	*P*-Value
SNP score	34.529	.174	4.140	26.410-42.649	<.001
Sex	1.234	.013	2.024	−2.736-5.204	.542
Age	.767	.263	.061	.647–.886	<.001
		R^2 = .099, adjusted R^2 = .098			

Note: Dependent variable: Serum-LDL-C (mg/dl).
Abbreviations: LDL-C, low-density lipoprotein cholesterol; SNP, single nucleotide polymorphisms.

4 and 5, a region also containing the most frequent *LDLR* variant among German patients, that is, c.798 T > A (p.Asp266Glu). The most frequent variant was *APOB* c.10580G > A (p.Arg3527Gln) which corresponds to previous findings in German patients.[8] Our study further substantiates the mutation spectrum underlying FH in a dataset from Germany. Surprisingly, the mean LDL-C of the six patients carrying two or more disease causing variants was not higher when compared to patients with only one sequence variant. This is likely due to the combination of the variants detected and their putative impact on LDL receptor function, since if both encode a protein with residual activity instead of a null mutation, then LDL-C elevations can be expected to be less severe, as has been reported before.[38] In agreement with a consensus study from the European Atherosclerosis Society, mean LDL-C levels and corresponding clinical phenotypes according to the DLCN score were lower in heterozygous patients with variants in two different genes than in compound heterozygous patients. This observation, which could be explained by the notion that carrying more than one alteration in the same protein is more likely to decrease its function more drastically.[39] Especially in the putatively compound heterozygous patients further investigation is needed to test for allelic independency. A possible allelic linkage in these patients could explain the lack of difference in mean LDL-C values between all mutation positive patients and patients carrying two variants in Figure 2B.

As expected, the baseline LDL-C was significantly higher in patients with all types of detected variants, when compared to mutation negative patients. When comparing the LDL-C levels of the mutation negative patients with those of patients with alterations rated as VUS by ACMG guidelines, there was no significant difference suggesting that most of these variants are likely to be benign. It is a general problem that a significant proportion of the variants detected in standard molecular diagnostic procedures cannot be comprehensively interpreted due to a lack of information on their functional impact. This has some serious diagnostic implications since less than 10% of the variants detected in the three known FH causing genes have been validated as *pathogenic* based on *in vitro* or other functional assays.[40] A strategy to improve interpretability of noncoding variants was recently proposed by Kircher et al,[41] who used a combination of saturation mutagenesis and reporter assays to conduct functional measurements of over 30 000 single-nucleotide substitutions and deletions in the 20 disease-associated gene promoters and enhancers, including the *LDLR* gene. It would be desirable to have a similar catalog containing functional consequences of variants in the coding region of the FH genes, demonstrating for example, the impact of the large proportion of unclassified missense variants in the *LDLR* gene. Using the DLCN score we categorized our patients based on their clinical phenotypes and calculated its ability to discriminate between subjects being mutation negative and mutation positive in the three major FH genes in ROC curves. We detected a variant causative for the disease in over 70% of the patients with a diagnosis rated as *probable* and *definite* according to DLCN criteria which is lower than reported by Scicali et al with over 90.5%. This might be explained by the higher prevalence of the disease (FH/M+) in their study with 46.4% compared to 40.0% in our cohort.[42] The DLCN score discriminated between mutation positive and mutation negative patients with an AUC of 0.76 (95% CI = 0.69-0.82, P < .001) and the baseline LDL-C with an AUC of 0.77 (95% CI = 0.71-0.83 P < .001). In other words, using baseline LDL-C is just as good in discriminating mutation positive and mutation negative patients, as the score calculated based on the DLNC criteria in our dataset. This is in agreement with our earlier findings from Grenkowitz et al (2016), and results from other studies.[8,43,44] The discriminatory power of LDL-C, however, is lower in older patients (>60 years), with the DLCN criteria being useful for the clinical diagnosis.[45]

In our study, we were unable to detect disease causing variants in 65.2% of the patients clinically suspected to have FH, a proportion which is in the range found in a recently published review article.[46] Futema and colleagues proposed a six-SNP polygenic score to recognize patients in which a polygenic FH etiology is likely.[18] We followed this strategy and overall observed similar results: the polygenic score was highest in patients without a detected disease causing variant (=mutation negative). In contrast to Futema et al we found no difference in the six-SNP polygenic score between patients with a disease causing variant detected (=mutation positive) and our control group. Another difference in relation to the original description of the score was that we observed on average lower scores in our cohorts, which limits a direct comparison to the results reported by Futema and colleagues. This could be explained by the fact that we assigned a score of *zero* for a particular SNP, if we were unable to evaluate one or two of the genotypes in a particular patient, unlike Futema et al who in some cases assigned the SNP genotypes that are most common, if one was missing.

In order to further evaluate the six-SNP polygenic score we tested its power to discriminate between "nonaffected" individuals from our control group and mutation negative subjects from our patient cohort. These analyses revealed only a modest discrimination

power with an AUC = 0.60 (95% CI = 0.56-0.64, $P < .001$). This result was supported by both the very low positive predictive value of the score calculated for our mutation negative patients and our control group (0.57%) and our linear regression models showing that the overall variance explained by the baseline LDL-C levels is small (~3%) and that age is a better predictor of LDL-C levels than the six-SNP polygenic score. In our view it does not come as a surprise that patients with a clinical FH phenotype but without evidence for disease causing variants in one of the known FH genes have a higher SNP score (=frequency) of LDL-C raising variants than controls, since most of them have become noticed because of their high LDL-C. Because the six-SNP polygenic score explained only 3% of the variance in LDL-C, we do not think, however, that a high SNP score is the primary (polygenic) cause of the disease in these patients. A higher SNP score rather has increased the chance to be diagnosed with clinical FH in patients, which might have a disease causing variant in an unknown gene, or with a so far unknown epigenetic modification. Less severity of variants in these unknown disease genes or modifications, when compared to alterations of the known FH genes, might then explain the intermediate LDL-C levels in these patients.

Interestingly, other groups have recently reported polygenic risk scores predictive for clinical outcomes in the context of FH: Trinder et al reported that patients with a monogenic cause for FH and a high polygenic risk score based on 28 LDL-C associated SNPs (four of which were also included in our study) had a significantly increased risk for premature CVD. An increased CVD risk, however, was not detected for the fraction of mutation negative patients in that same study.[47] Paquette et al reported similar results using a genetic risk score comprised of 192 SNPs associated with coronary artery disease (CAD) which were found to be strongly associated with CVD events in 725 FH mutation positive patients.[48] In clinical practice a score like this could be used for screening for "general CVD risk" and perhaps as additional information alongside the established measures, such as the molecular diagnosis and the clinical phenotype, for high risk patients with a known monogenetic cause and superimposed polygenic risk.

Other scores based on LDL-C associated SNPs have been proposed by Wang et al (2016), Lamiquiz-Moneo et al (2017).[49,50] Considering their results and the reviews of both Hooper et al (2018) and Iacocca & Hegele et al (2017)[51,52] leads to analogous conclusions. There was no strong association found between LDL-C levels and a SNP score in other studies, for example, Lamiquiz-Moneo et al (2017) observed that only 3.1% of the LDL-C variance could be explained in their examined cohort, which supports our findings.[49,50] Finally, Wang et al (2016) and Iacocca & Hegele et al (2017) proposed that the aggregation of SNPs will cluster within families, a hypothesis we could not assess owing to the lack of sufficient family-based data in our datasets.

In summary, we did not find evidence that the polygenic risk score proposed by Futema and colleagues represents a reliable tool for FH in clinical practice. However, further research should be conducted to evaluate the potential of polygenic risk score(s) in clinical risk prediction of CVD. Therefore, we would argue that patients with a clinical FH phenotype lacking evidence for disease causing variants in the canonical FH genes should not be referred to as *hypercholesterolemic with a polygenic cause*. Instead, these patients should be the subject of future research to reveal the real underlying cause of the disease, be it genetic, epigenetic or of a different in nature.

ACKNOWLEDGEMENTS

We would like to acknowledge Gabriele Hildebrand and Susanne Rothe for their technical advice on the conduction of laboratory assays.

CONFLICT OF INTEREST

Nothing to declare.

AUTHOR CONTRIBUTIONS

Lorenz Rieck, Ursula Kassner, and Ilja Demuth conceived the study, interpreted and discussed data. Lorenz Rieck generated the genotype data for the patient cohort, analyzed the data and wrote the first manuscript draft. Frieda Bardey contributed to the statistical analysis and Thomas Grenkowitz helped conceptualizing the study design. Lars Bertram, Johannes Helmuth, Claudia Mischung, Joachim Spranger, Elisabeth Steinhagen-Thiessen, Thomas Bobbert, Ursula Kassner, and Ilja Demuth provided data. Ursula Kassner, and Ilja Demuth supervised the study. All authors reviewed the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/cge.13826.

DATA AVAILABILITY STATEMENT

Due to concerns for participant privacy, BASE-II data are available only upon request. External scientists may apply to the Steering Committee of BASE-II for data access. Please refer to the BASE-II website (https://www.base2.mpg.de/en/project-information/datadocumentation) for additional information. Please contact Ludmila Müller, scientific coordinator, at lmueller@mpib-berlin.mpg.de.

ORCID

Lorenz Rieck https://orcid.org/0000-0003-3493-2423

REFERENCES

1. Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European atherosclerosis society consensus panel. *Eur Heart J*. 2017;38:2459-2472.
2. Nordestgaard BG, Chapman MJ, Humphries SE, et al. Familial hypercholesterolaemia 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*. 2013;34:3478-3490a.
3. Kassner U, Wühle-Demuth M, Missala I, et al. Clinical utility gene card for: hyperlipoproteinemia, TYPE II. *Eur J Hum Genet*. 2014;22(7).

4. Mbikay M, Mayne J, Chrétien M. Proprotein convertases subtilisin/kexin type 9, an enzyme turned escort protein: hepatic and extra hepatic functions. *J Diabetes*. 2013;5:391-405.
5. Soutar AK, Naoumova RP. Mechanisms of disease: genetic causes of familial hypercholesterolemia. *Nat Clin Pract Cardiovasc Med*. 2007;4:214-225.
6. Fellin R, Arca M, Zuliani G, Calandra S, Bertolini S. The history of autosomal recessive hypercholesterolemia (ARH). From clinical observations to gene identification. *Gene*. 2015;555:23-32.
7. Leigh S, Futema M, Whittall R, et al. The UCL low-density lipoprotein receptor gene variant database: pathogenicity update. *J Med Genet*. 2017;54:217-223.
8. Grenkowitz T, Kassner U, Wühle-Demuth M, et al. Clinical characterization and mutation spectrum of German patients with familial hypercholesterolemia. *Atherosclerosis*. 2016;253:88-93.
9. Vuorio A, Watts GF, Schneider WJ, et al. Familial hypercholesterolemia and elevated lipoprotein(a): double heritable risk and new therapeutic opportunities. *J Intern Med*. 2020;287(1):2-18.
10. Graham CA, McIlhatton BP, Kirk CW, et al. Genetic screening protocol for familial hypercholesterolemia which includes splicing defects gives an improved mutation detection rate. *Atherosclerosis*. 2005;182:331-340.
11. Talmud PJ, Shah S, Whittall R, et al. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet*. 2013;381:1293-1301.
12. Umans-Eckenhausen MA, Defesche JC, Sijbrands EJ, et al. Review of first 5 years of screening for familial hypercholesterolaemia in The Netherlands. *Lancet*. 2001;357:165-168.
13. Marduel M, Ouguerram K, Serre V, et al. Description of a large family with autosomal dominant hypercholesterolemia associated with the APOE p.Leu167del mutation. *Hum Mutat*. 2013;34:83-87.
14. Fouchier SW, Dallinga-Thie GM, Meijers JC, et al. Mutations in STAP1 are associated with autosomal dominant hypercholesterolemia. *Circ Res*. 2014;115:552-555.
15. Danyel M, Ott CE, Grenkowitz T, et al. Evaluation of the role of STAP1 in familial hypercholesterolemia. *Sci Rep*. 2019;9:11995.
16. Loaiza N, Hartgers ML, Reeskamp LF, et al. Taking one step Back in familial hypercholesterolemia. *Arterioscler Thromb Vasc Biol*. 2020;40(4):ATVBAHA119313470.
17. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713.
18. Futema M, Shah S, Cooper JA, et al. Refinement of variant selection for the LDL cholesterol genetic risk score in the diagnosis of the polygenic form of clinical familial hypercholesterolemia and replication in samples from 6 countries. *Clin Chem*. 2015;61:231-238.
19. Bertram L, Böckenhoff A, Demuth I, et al. Cohort profile: the Berlin aging study II (BASE-II). *Int J Epidemiol*. 2014;43:703-712.
20. Adams SP, Tsang M, Wright JM. Lipid lowering efficacy of atorvastatin. *Cochrane Database Syst Rev*. 2012;12:CD008226.
21. Ara R, Tumur I, Pandor A, et al. Ezetimibe for the treatment of hypercholesterolaemia: a systematic review and economic evaluation. *Health Technol Assess*. 2008;12:1-212.
22. Ballantyne CM, Hoogeveen RC, Raya JL, et al. Efficacy, safety and effect on biomarkers related to cholesterol and lipoprotein metabolism of rosuvastatin 10 or 20 mg plus ezetimibe 10 mg vs. simvastatin 40 or 80 mg plus ezetimibe 10 mg in high-risk patients: results of the GRAVITY randomized study. *Atherosclerosis*. 2014;232:86-93.
23. Ballantyne CM, Houri J, Notarbartolo A, et al. Effect of ezetimibe coadministered with atorvastatin in 628 patients with primary hypercholesterolemia: a prospective, randomized, double-blind trial. *Circulation*. 2003;107:2409-2415.
24. Ballantyne CM, Pazzucconi F, Pintó X, et al. Efficacy and tolerability of fluvastatin extended-release delivery system: a pooled analysis. *Clin Ther*. 2001;23:177-192.
25. Ballantyne CM, Weiss R, Moccetti T, et al. Efficacy and safety of rosuvastatin 40 mg alone or in combination with ezetimibe in patients at high risk of cardiovascular disease (results from the EXPLORER study). *Am J Cardiol*. 2007;99:673-680.
26. Foody JM, Toth PP, Tomassini JE, et al. Changes in LDL-C levels and goal attainment associated with addition of ezetimibe to simvastatin, atorvastatin, or rosuvastatin compared with titrating statin monotherapy. *Vasc Health Risk Manag*. 2013;9:719-727.
27. Jones P, Kafonek S, Laurora I, et al. Comparative dose efficacy study of atorvastatin versus simvastatin, pravastatin, lovastatin, and fluvastatin in patients with hypercholesterolemia (the CURVES study). *Am J Cardiol*. 1998;81:582-587.
28. Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *BMJ*. 2003;326:1423-1420.
29. Organization, WH, Familial hypercholesterolemia—report of a second WHO consultation, Geneva: World Health Organization, 1999.
30. Stenson PD, Mort M, Ball EV, et al. The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*. 2017;136:665-677.
31. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46:D1062-D1067.
32. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248-249.
33. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11:361-362.
34. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37:e67.
35. 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*. 2015;17:405-424.
36. Nykamp K, Anderson M, Powers M, et al. Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet Med*. 2017;19:1105-1117.
37. Austin MA, Hutter CM, Zimmern RL, Humphries SE. Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. *Am J Epidemiol*. 2004;160:407-420.
38. Santos RD, Gidding SS, Hegele RA, et al. Defining severe familial hypercholesterolaemia and the implications for clinical management: a consensus statement from the international atherosclerosis society severe familial hypercholesterolemia panel. *Lancet Diabetes Endocrinol*. 2016;4:850-861.
39. Cuchel M, Bruckert E, Ginsberg HN, et al. Homozygous familial hypercholesterolaemia: new insights and guidance for clinicians to improve detection and clinical management. A position paper from the consensus panel on familial Hypercholesterolaemia of the European atherosclerosis society. *Eur Heart J*. 2014;35:2146-2157.
40. Chora JR, Medeiros AM, Alves AC, Bourbon M. Analysis of publicly available LDLR, APOB, and PCSK9 variants associated with familial hypercholesterolemia: application of ACMG guidelines and implications for familial hypercholesterolemia diagnosis. *Genet Med*. 2018;20:591-598.
41. Kircher M, Xiong C, Martin B, et al. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat Commun*. 2019;10:3583.

42. Scicali R, Di Pino A, Platania R, et al. Detecting familial hypercholesterolemia by serum lipid profile screening in a hospital setting: clinical, genetic and atherosclerotic burden profile. *Nutr Metab Cardiovasc Dis*. 2018;28(1):35-43.
43. Silva PRS, Jannes CE, Oliveira TGM, et al. Evaluation of clinical and laboratory parameters used in the identification of index cases for genetic screening of familial hypercholesterolemia in Brazil. *Atherosclerosis*. 2017;263:257-262.
44. Garg A, Garg V, Hegele RA, Lewis GF. Practical definitions of severe versus familial hypercholesterolaemia and hypertriglyceridaemia for adult clinical practice. *Lancet Diabetes Endocrinol*. 2019;7:880-886.
45. Wald DS, Bestwick JP, Wald NJ. Child-parent screening for familial hypercholesterolaemia: screening strategy based on a meta-analysis. *BMJ*. 2007;335:599.
46. Sharifi M, Futema M, Nair D, Humphries SE. Polygenic hypercholesterolemia and cardiovascular disease risk. *Curr Cardiol Rep*. 2019;21:43.
47. Trinder M, Li X, DeCastro ML, et al. Risk of premature atherosclerotic disease in patients with monogenic versus polygenic familial hypercholesterolemia. *J Am Coll Cardiol*. 2019;74:512-522.
48. Paquette M, Chong M, Thériault S, Dufour R, Paré G, Baass A. Polygenic risk score predicts prevalence of cardiovascular disease in patients with familial hypercholesterolemia. *J Clin Lipidol*. 2017;11:725-732. e725.
49. Wang J, Dron JS, Ban MR, et al. Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically. *Arterioscler Thromb Vasc Biol*. 2016;36:2439-2445.
50. Lamiquiz-Moneo I, Pérez-Ruiz MR, Jarauta E, et al. Single nucleotide variants associated with polygenic hypercholesterolemia in families diagnosed clinically with familial hypercholesterolemia. *Rev Esp Cardiol* (Engl Ed. 2018;71:351-356.
51. Hooper AJ, Burnett JR, Bell DA, Watts GF. The present and the future of genetic testing in familial hypercholesterolemia: opportunities and caveats. *Curr Atheroscler Rep*. 2018;20:31.
52. Iacocca MA, Hegele RA. Recent advances in genetic testing for familial hypercholesterolemia. *Expert Rev Mol Diagn*. 2017;17:641-651.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Rieck L, Bardey F, Grenkowitz T, et al. Mutation spectrum and polygenic score in German patients with familial hypercholesterolemia. *Clin Genet*. 2020;98:457–467. https://doi.org/10.1111/cge.13826