ORIGINAL ARTICLE

Whole exome sequencing of familial hypercholesterolaemia patients negative for *LDLR*/*APOB*/*PCSK9* mutations

Marta Futema,[1] Vincent Plagnol,[2] KaWah Li,[1] Ros A Whittall,[1] H Andrew W Neil,[3] Mary Seed,[4] on behalf of the Simon Broome Consortium, Stefano Bertolini,[5] Sebastiano Calandra,[6] Olivier S Descamps,[7] Colin A Graham,[8] Robert A Hegele,[9] Fredrik Karpe,[10] Ronen Durst,[11,12] Eran Leitersdorf,[12] Nicholas Lench,[13] Devaki R Nair,[14] Handrean Soran,[15] Frank M Van Bockxmeer,[16] UK10K Consortium,[17] Steve E Humphries[1]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2014-102405).

For numbered affiliations see end of article.

Correspondence to
Professor Steve E Humphries, Centre for Cardiovascular Genetics, British Heart Foundation Laboratories, Institute of Cardiovascular Science, The Rayne Building University College London, London WC1E 6JF, UK; steve.humphries@ucl.ac.uk, rmhaseh@ucl.ac.uk

Received 10 March 2014
Revised 2 June 2014
Accepted 15 June 2014

To cite: Futema M, Plagnol V, Li KW, *et al*. *J Med Genet* 2014;**51**: 537–544.

ABSTRACT

Background Familial hypercholesterolaemia (FH) is an autosomal dominant disease of lipid metabolism, which leads to early coronary heart disease. Mutations in *LDLR*, *APOB* and *PCSK9* can be detected in 80% of definite FH (DFH) patients. This study aimed to identify novel FH-causing genetic variants in patients with no detectable mutation.

Methods and results Exomes of 125 unrelated DFH patients were sequenced, as part of the UK10K project. First, analysis of known FH genes identified 23 *LDLR* and two *APOB* mutations, and patients with explained causes of FH were excluded from further analysis. Second, common and rare variants in genes associated with low-density lipoprotein cholesterol (LDL-C) levels in genome-wide association study (GWAS) meta-analysis were examined. There was no clear rare variant association in LDL-C GWAS hits; however, there were 29 patients with a high LDL-C SNP score suggestive of polygenic hypercholesterolaemia. Finally, a gene-based burden test for an excess of rare (frequency <0.005) or novel variants in cases versus 1926 controls was performed, with variants with an unlikely functional effect (intronic, synonymous) filtered out.

Conclusions No major novel locus for FH was detected, with no gene having a functional variant in more than three patients; however, an excess of novel variants was found in 18 genes, of which the strongest candidates included *CH25H* and *INSIG2* ($p<4.3\times10^{-4}$ and $p<3.7\times10^{-3}$, respectively). This suggests that the genetic cause of FH in these unexplained cases is likely to be very heterogeneous, which complicates the diagnostic and novel gene discovery process.

INTRODUCTION

Familial hypercholesterolaemia (FH (OMIM #143890)) is a genetic disorder, inherited in an autosomal dominant fashion, characterised by the defective plasma clearance of low-density lipoprotein cholesterol (LDL-C) and caused by mutations in three genes: *LDLR*, *APOB* and *PCSK9*.[1] A recessive form of FH due to mutations in *LDLRAP1* is also known.[2] FH is estimated to affect one in 500 individuals[3] and if untreated leads to premature coronary heart disease (CHD).[4] In the UK, the FH Simon Broome criteria are used for the diagnosis, which classify patients into possible FH, when adults present with total cholesterol >7.5 mmol/L or LDL-C >4.9 mmol/L, and family history of high cholesterol or premature CHD, or the more severe form—definite FH (DFH), when in addition to the above, tendon xanthomas are present in the patient or first or second degree relative.[5] The FH mutation detection rate for DFH patients varies between 63% and 87%,[6–8] suggesting that there are other genetic causes, located outside of the currently screened regions, which are yet to be identified. The importance of identifying an FH-causing variant, which has clinical utility in providing an unequivocal diagnosis,[9] has been emphasised by the National Institute of Health and Care Excellence, which in 2008 recommended cascade testing using DNA information for finding the affected relatives of a patient.[10] The risk of early CHD can be significantly reduced by statin treatment,[11] and genetic information has been demonstrated to complement the management of treated patients.[12]

Of FH patients where a mutation can be found, ~93% occur in the *LDLR* gene.[13] The *APOB* variant (c.10580G>A, p.(Arg3527Gln)) accounts for ~5% of UK FH cases,[7 8 14] whereas a gain-of-function mutation in *PCSK9* (c.1120G>T, p.(Asp374Tyr)) can be found in roughly 1.7% of FH patients.[14] In the past few years, several loci have been reported to cosegregate with FH in family linkage studies; however, to date, this has not led to the identification of a specific causal gene.[15–17] It is likely that there are novel FH mutations located in unknown genes influencing lipid metabolism and that their discovery may contribute to the identification of novel treatment targets. In order to find novel causes of FH it was agreed that, as part of the UK10K project (http://www.uk10k.org/studies/rarediseases.html), the whole exomes of 125 unrelated DFH patients were sequenced at a high depth. We expected that an FH-causing mutation in a novel gene would be very rare accounting for fewer FH cases than the gain-of-function mutation in *PCSK9* (1.7%), since a higher frequency would have made likely its identification in previous studies. We also suspected that a proportion of

patients would have polygenic hypercholesterolaemia, due to the combined impact of common LDL-C-raising SNPs.[18]

MATERIALS AND METHODS

Patients

A total of 125 unrelated patients, diagnosed as DFH using the UK Simon Broome criteria on the basis of the presence or family history of tendon xanthomas, were initially screened and shown to be negative for mutations in known FH genes (*LDLR*, *APOB*, *PCSK9* and *LDLRAP1*). All consents and local review board approvals were in accordance with the UK10K project ethical framework. The initial mutation screening methods varied and are summarised in online supplementary table S1.

Controls

The association with FH was tested against consented 1926 UK10K samples with no lipid abnormalities (listed in online supplementary methods), sequenced in parallel, using the same sequence capture and variant calling methods (http://www.uk10k.org/studies/).

Exome sequencing and variant calling

The whole exome sequencing was performed and processed at the Wellcome Trust Sanger Institute (Cambridge, UK) as part of the UK10K project (see online supplementary methods). CNVs were called using the ExomeDepth package for R (freely available at the Comprehensive R Archive Network).[19]

Filtering of the variants

Variants were flagged as *rare* (frequency<0.5%) and *novel* (frequency=0) according to their frequency in publicly available databases including 1000 Genomes[20] and National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP6500) (http://evs.gs.washington.edu/EVS/). In addition to the frequency filters, a *functional* flag was added, which prioritised variants that are most likely to affect a protein's function, that is, non-synonymous, stop gain, stop loss, frameshift deletions and insertions, and splice site variants.

Burden test for association

Rare or *novel* variants were combined in a single gene manner and counted in cases versus 1926 controls (ie, gene by gene). A binomial test was used to assess the excess of *functional rare* and *novel* variants in cases in comparison with the controls. p Values lower than 4×10^{-3} were taken as evidence sufficient to be flagged for follow-up.

Analysis of the variants

Variants within Tier 1 genes (*LDLR*, *APOB*, *PCSK9*, *LDLRAP1*) were assessed on the basis of their frequency, and manually by looking at their annotations in the UCL FH mutation database.[21] Sanger sequencing was used to confirm all called mutations. Samples with known FH mutations and therefore an explained cause were removed from further analysis.

The Tier 2 list (see online supplementary table S2) consisted of genes associated with LDL-C as a lead trait in the largest (at the time) available Global Lipid Genetic Consortium (GLGC) meta-analysis of genome-wide association studies (GWASs).[22] *Functional rare* and *novel* variants in the Tier 2 genes were compared by the burden test against non-FH controls, as one group (counts in all genes combined) and by each single gene.

LDL-C gene score analysis

The possibility of polygenic hypercholesterolaemia in this cohort was assessed using the LDL-C gene score analysis, recently described.[18] Most of the 12 LDL-raising GWAS SNPs are located outside of the coding regions, and thus to obtain these genotypes, methods as in the original publication were used.[18] Gene scores were calculated by summing the weights of LDL-raising alleles provided by the GLGC (see online supplementary table S3) and the *APOE* haplotype was scored as follows: $\varepsilon2\varepsilon2=-0.9$, $\varepsilon2\varepsilon3=-0.4$, $\varepsilon2\varepsilon4=-0.2$, $\varepsilon3/\varepsilon3=0$, $\varepsilon3\varepsilon4=0.1$ and $\varepsilon4\varepsilon4=0.2$.[22] Gene scores of a randomly selected subjects from the UK Whitehall II (WHII) study (n=3020) were used as a healthy control comparison group.[23] Individuals with a gene score above 1.16, which was the top decile cut-off for the WHII subjects, were considered to have polygenic hypercholesterolaemia. The Welch two sample t test was used to test for an overall difference between the groups.

RESULTS

We first analysed variants in known FH genes (figure 1A) (for gene coverage see online supplementary results). For *LDLR*, 10 individuals were carrying a missense mutation, five a nonsense mutation, three had small deletions and two individuals had intronic changes known to affect splicing (see online supplementary table S4). Analysis with ExomeDepth for CNVs identified two large duplications and one deletion within the *LDLR* region (see online supplementary figure S1). For *APOB*, two individuals carried the known FH mutation, c.10580G>A (p.R3527Q), and several *novel* and cases-unique *APOB* variants, distributed across different gene exons, were identified (see online supplementary table S5). These included a recently identified mutation, p.R50W, which cosegregated with the disease.[24] Because *APOB* is highly polymorphic, the overall number of rare variants was not significantly different in comparison with controls. *PCSK9* had the lowest mean read depth (18×), with four exons (1, 5, 9 and 10) covered less than 10× due to a high guanine and/or cytosine (GC) content (see online supplementary figure S2). There were no FH-causing variants identified in this gene. There were no homozygous or compound heterozygous calls in the *LDLRAP1* gene in any of the samples. One patient was found to be heterozygous for a previously identified frameshift mutation (c.432_433insA (p.(Ala145LysfsX26))).[25]

LDL-C gene score analysis

Out of 109 FH samples (21 mutation positive, 88 mutation negative) with sufficient DNA for genotyping for all 12 SNPs, 31 had a gene score above the 1.16 cut-off (figure 1A), within which two samples, in addition to the high gene score, had an *LDLR* mutation, one in exon 11 (c.1690A>C (p.N564H)) found on the same allele as a 9bp deletion in exon 17 (c.2393_2401del9 (p.L799_V801del)), which has been demonstrated as not fully-penetrant.[26] The other was a deletion of a consensus splice site at the 5′ of exon 5, c.695-6_698del, which has not been examined in vivo to confirm its likely effect on splicing.

The mean LDL-C gene score for the FH mutation negative group was 1.08, which was significantly higher than 0.90 for the WHII study ($p<2.2\times10^{-16}$), and 0.96 for the FH mutation positive group (p=0.006) (figure 1B) (for the distribution of scores see online supplementary figure S3). The overall difference between the groups was significant (analysis of variance (ANOVA), $p=1.33\times10^{-12}$). Individuals with a gene score above the top decile cut-off for the WHII subjects (>1.16), were

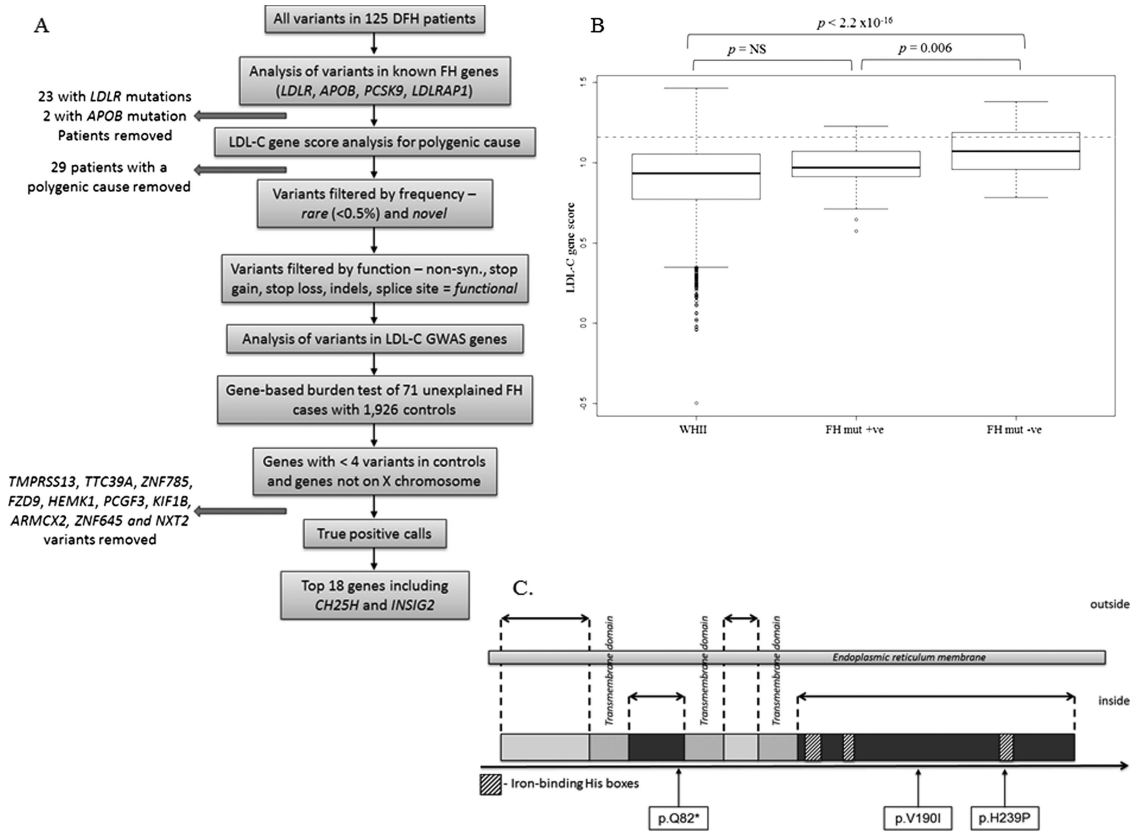


Figure 1 Novel familial hypercholesterolaemia (FH) gene discovery pipeline. (A) To increase the chance of detecting true FH-causing variants with a strong effect and reduce the noise, samples with a mutation in *LDLR* or *APOB* (apart from novel *APOB* variants of unknown effect) or those with a high low-density lipoprotein cholesterol (LDL-C) gene score were removed from the analysis. The remaining variants were filtered by their frequency and *functional* effect and compared against controls. Genes with more than four *novel functional* variants in controls or genes located on the X chromosome were filtered out to enhance the power of the test. The remaining variants were manually assessed and false positive calls were removed. (B) Comparison of the LDL-C SNPs score among the WHII control population (n=3020), FH mutation positive individuals (n=21) and FH mutation negative individuals (n=83) in a standard boxplot (the minimum, lower quartile, median, upper quartile and maximum). The overall difference between the groups was highly significant (ANOVA, $p<2.2\times10^{-16}$). Dashed line indicates the top decile cut-off for the WHII cohort (=1.16). A gene score was not attainable for 16 samples due to a poor DNA quality and insufficient concentration, which resulted in incomplete genotype data. (C) Schematic representation of the intronless *CH25H* gene and the localisation of novel variants identified in the FH cohort (in boxes). *CH25H* encodes an enzyme, cholesterol 25-hydroxylase, known to be spanning the endoplasmic reticulum membrane, with two domains (including the N-terminal) located outside of the membrane (in light grey), three 20 amino acid long transmembrane regions and two domains positioned inside the membrane, which contain three His boxes, essential for the catalytic activity of the enzyme.[30]

considered to have polygenic hypercholesterolaemia and excluded from further analysis as they were unlikely to carry a single mutation of a strong effect.

GWAS LDL-C genes

We next examined any gene identified through GWAS as being involved in determining levels of LDL-C in healthy individuals.[22] A burden test on all *functional rare* and *novel* variants in any gene singly or in all Tier 2 genes combined showed no obvious candidate for a novel FH locus (see online supplementary table S6). In addition, there were no loss-of-function variants (ie, premature stop codon formation, loss of a stop codon, frameshift indels, CNVs) observed in these genes in any sample (n=125), or in the 71 with no identified mutation and a low gene score. There was no association of *novel functional* variants in any gene located within the several loci identified by published family linkage studies (see online supplementary table S7).

Whole exome analysis

In all, 25 samples carrying a mutation in Tier 1 genes and 29 with the LDL-C gene score above 1.16 were removed from further analysis. To interrogate the whole exome, a burden test was performed between 71 cases and 1926 controls. There were 4407 genes with one or more *novel functional* variant in cases. In order to remove calls less likely to influence the FH phenotype and increase the power of the test, we limited further analysis only to genes where a maximum of four *novel functional* variants were seen in the controls, based on the expected prevalence of FH of 1 in 500, and therefore any gene with >4 *novel functional* variants in the controls were excluded (the original gene list is shown in online supplementary table S8). Variants in genes located on the X chromosome were removed from the final list (X chromosome genes shown in online supplementary table S8). The next step involved a visual validation of the quality of calls performed using the Human Genome 19 on the Integrative Genomic Viewer (IGV).[27] In order to avoid false negatives, calls that were filtered out due to inadequate quality were reanalysed in genes showing excess of *novel* variants. An additional loss-of-function variant, a premature stop codon at the position c.244C>T (p.Q81*), was found in the *CH25H* gene in an FH patient sample with a low LDL-C SNP score. After adjusting for the false positives and false negatives,

Table 1 Summary of genes and their variants which show an excess of *novel functional* variants in FH cases (n=71) in comparison with controls (n=1926)

Gene		Ch	Number of variants in cases (n=71)	Number of variants in controls (n=1926)	p Value
CH25H		**10**	**3**	**2**	**4.3×10^{-4}**
	Cases		ENST00000371852:exon1:c.G568A:p.V190I; exon1:c.A716C:p.H239P; exon1:c.C244T:p.Q82X		
	Controls		ENST00000371852:exon1:c.T742G:p.C248G; exon1:c.C590A:p.P197Q		
HSPB7		**1**	**2**	**0**	**1.3×10^{-3}**
	Cases		2X ENST00000311890:exon2:c.199+7G>A		
	Controls		None		
KLRC1		**12**	**2**	**0**	**1.3×10^{-3}**
	Cases		ENST00000544822:exon5:c.G333C:p.Q111H; exon3:c.C178T:p.H60Y		
	Controls		None		
MOAP1		**14**	**3**	**4**	**1.4×10^{-3}**
	Cases		ENST00000556883:exon2:c.C707T:p.A236V; exon2:c.G476C:p.C159S; exon2:c.A182G:p.N61S		
	Controls		ENST00000556883:exon2:c.C655G:p.R219G; exon2:c.C627A:p.S209R; exon2:c.C264G:p.I88M; exon2:c.A919G:p.I307V		
RBM25		**14**	**3**	**4**	**1.4×10^{-3}**
	Cases		ENST00000261973:exon6:c.A454T:p.I152F; exon2:c.T50C:p.L17P; exon11:c.C1364A:p.A455D		
	Controls		ENST00000261973:exon7:c.C671T:p.A224V; exon11:c.A1273G:p.R425G; exon18:c.G2392A:p.V798I; exon2:c.T7C:p.F3L		
ANP32E		**1**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000436748:exon3:c.G227C:p.S76T; ENST00000533654:exon4:c.A434G:p.K145R		
	Controls		ENST00000436748:exon6:c.G629T:p.R210L		
CABP5		**19**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000293255:exon4:c.C281A:p.T94N; exon3:c.G201A:p.M67I		
	Controls		ENST00000293255:exon3:c.A169C:p.M57L		
CELA2B		**1**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000375910:exon6:c.G576A:p.W192X; ENST00000422901:exon3:c.G271A:p.G91R		
	Controls		ENST00000375910:exon7:c.T739C:p.Y247H		
INSIG2		**2**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000245787:exon2:c.T89C:p.I30T; exon2:c.C236T:p.T79M		
	Controls		ENST00000245787:exon4:c.G376A:p.D126N		
KCTD7		**7**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000275532:exon4:c.G814A:p.V272M; exon4:c.C758T:p.S253L		
	Controls		ENST00000275532:exon4:c.G506A:p.R169Q		
MRO		**18**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000436348:exon5:c.G578A:p.R193Q; exon5:c.G565A:p.V189I		
	Controls		ENST00000436348:exon3:c.A223G:p.S75G		
NR2E1		**6**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000368983:exon1:c.G136A:p.G46S; exon5:c.A634G:p.M212V		
	Controls		ENST00000368983:exon7:c.G1000A:p.V334I		
PABPC1		**8**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000318607:exon9:c.A1250C:p.Q417P;exon10:c.G1364A:p.R455H		
	Controls		ENST00000523555:exon3:c.226+3A>G		
PODXL		**7**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000537928:exon3:c.G821A:p.R274K; exon5:c.A992G:p.H331R		
	Controls		ENST00000537928:exon8:c.C1246G:p.Q416E		
PUS3		**11**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000530811:exon1:c.T74C:p.V25A; exon2:c.T824C:p.L275P		
	Controls		ENST00000530811:exon4:c.945-8T>C		
TXNDC15		**5**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000511070:exon2:c.C130T:p.R44W; ENST00000507024:exon2:c.G91A:p.A31T		
	Controls		ENST00000358387:exon2:c.G534C:p.E178D		
WDR89		**14**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000394942:exon2:c.T821C:p.L274S; exon2:c.A553G:p.M185V		
	Controls		ENST00000394942:exon2:c.A860G:p.D287G		
ZNF720		**16**	**2**	**1**	**3.7×10^{-3}**
	Cases		ENST00000398696:exon2:c.T508G:p.L170V; exon2:c.A29G:p.H10R		
	Controls		ENST00000399681:exon6:c.A893G:p.H298R		

Ch, chromosome; FH, familial hypercholesterolaemia.

CH25H remained the top gene ($p<4.3\times10^{-4}$) with three variants in the cases and two in the controls (table 1). To examine the prevalence of nonsense variants in *CH25H* in public data sources, we analysed the NHLBI ESP database and found one nonsense allele (c.638delT) in 6503 individuals (Minor Allele Frequency (MAF)=0.00008), which was significantly lower than in the FH group (MAF=0.0003, $p<1.5\times10^{-3}$).

CH25H and *INSIG2* variants

CH25H codes for cholesterol 25-hydroxylase, known to catalyse the formation of the oxysterol—25-hydroxycholesterol (25-HC) (9). The *INSIG2* gene, which also exhibited an excess of *novel functional* variants in the FH cohort in comparison with the controls ($p=3.7\times10^{-3}$) (table 1), has been demonstrated to regulate the activity of Sterol Regulatory Element-Binding Protein (SREBPs), a family of major lipid metabolism transcription factors, via direct biding of 25-HC.[28] Thus, both genes, *CH25H* and *INSIG2*, are involved in the same pathway of cholesterol metabolism. There were three heterozygous variants found in *CH25H*, all confirmed by Sanger sequencing (see online supplementary figure S4), of which one leads to a formation of a premature stop codon at residue 81, predicted to have a damaging effect on the protein; the second affects a well-conserved residue across species, c.568G>A (p.V190I); and the third, c.716A>C (p.H239P), alters one of the crucial residues of the His Box 3 domain, known to play a crucial role, together with His Boxes 1 and 2, in the catalytic activity of *CH25H*[29] (figure 1C). Two *novel functional* variants were found in the control cohort, both being non-synonymous (p.P197Q and p.C248G). The p.P197Q is located in a conserved region of the protein; however, it is predicted as tolerated/benign/neutral by SIFT/PolyPhen/Mutation Taster. The p.C248G variant affects a residue that is not conserved.[30]

Sanger sequencing also confirmed two *novel functional* variants in the *INSIG2* gene called in the cases, both non-synonymous changes (see online supplementary figure S5). A mutation prediction report generated by Project HOPE[31] highlighted that the c.89T>C (p.I30T) variant will cause an empty space in the core of INSIG2 because of the size differences between the wild type Isoleucine and the smaller mutant —Threonine. The other variant, c.236C>T (p.T79M), located in the transmembrane domain of INSIG2, is predicted to have an effect on the hydrophobic interactions within the core of the protein or with the membrane lipids, because the mutant Methionine is more hydrophobic than the wild type Threonine. One rare missense variant was found in *INSIG2* in the controls (p.D126N), which was predicted as tolerated/probably damaging/disease causing (by SIFT/PolyPhen/Mutation Taster).

DISCUSSION

In this study, we have identified 25 mutations in known FH genes (23 in *LDLR* and two in *APOB*), which were missed by the current screening protocol. Because the sequencing coverage of the *PCSK9* gene was lower than for *LDLR* and *APOB*, we cannot rule out that there may have been undetected mutations in this gene also. This finding confirmed that *LDLR* locus is highly heterogeneous and mutations within this gene account for the majority of FH causes. The issue of genetic misdiagnosis and the need for an update of current screening methods have been previously discussed.[32] In addition to the known FH mutations, we identified six novel *APOB* variants, distributed across different exons, in five patients, which included the recently examined p.R50W variant.[33] The pathogenicity of these variants remains to be tested. Most of the current mutation screening strategies for FH are focused on a selected region of exon 26 of *APOB*, because of its established function;[34] however, the whole exome sequencing enabled us to analyse the entire coding sequence of the gene, by which we found novel variants unique to the FH cohort.

Polygenic hypercholesterolaemia

The cumulative effect of common LDL-raising alleles in genes identified by GWAS was shown to be the likely cause of high LDL-C in a significant proportion (27%) of the examined patients. A gene score above the top decile for a healthy population cut-off (1.16) was also observed in two patients with considerably mild *LDLR* mutations, which demonstrates that common polymorphisms can contribute to the presentation of an individual carrying a mild effect FH mutation with LDL-C levels above the diagnostic threshold.

GWAS LDL-C genes

Since common variants in the LDL-C-associated GWAS genes were found to be important in the FH pathogenesis, we looked for evidence that rare variants in these genes were causing FH. *Rare* and *novel functional* variants in genes associated with LDL-C levels in the GWAS meta-analysis were not significantly over-represented in the FH cohort, when compared with controls. This suggests that rare variants that have a major effect on function in these genes known to have common LDL-C variants of modest effect are unlikely to be a common cause of FH.

CH25H and *INSIG2* variants

Genes *CH25H* and *INSIG2* are the strongest candidates for novel FH loci among the final 18 genes, showing an excess of *novel functional* variants, based on the available reports on functions of the proteins for which they code. *CH25H* encodes 25-cholesterol hydroxylase, which catalyses the formation of 25-HC from cholesterol. The gene is located in close proximity to the *LIPA* gene in which mutations were recently found in patients with autosomal recessive FH phenotype.[35] It has been demonstrated that both cholesterol and 25-HC can regulate the function of SREBP, a transcription factor known to regulate the expression of several key players in the lipid metabolism.[36 37] It is known that the regulation of SREBP activity depends on binding of 25-HC to INSIG2, encoded by the *INSIG2* gene.[28] The recently updated GLGC GWAS study with >180 000 individuals has identified an association at the genome-wide level of LDL-C with an *INSIG2* gene variant (rs10490626, MAF=0.08).[38]

The *CH25H* variants identified in this study have not been observed in 1000 Genomes, 6500ESP and 69CG or the 1926 control exomes. We therefore decided to sequence the gene in an additional cohort of 150 mutation negative FH patients with a low gene score, but no additional amino acid changes were identified.

A detailed literature search and gene ontology analysis of the remaining 16 most significant genes did not reveal any clear association with lipid metabolism. We suspect that the majority of these associations are false positives, and that increasing the number of DFH cases would help to reduce the number of chance signals. It is also possible that some of the top genes are indeed affecting the plasma clearance of LDL-C; however, their biology is yet to be understood.

There are a number of limitations to our study. An alternative study design would be to use Next Generation Sequencing (NGS) of relatives (or trios) from selected families with clear autosomal dominant hypercholesterolaemia. The UK10K study

only allowed for 125 subjects with FH to be included, and we calculated that, if we selected 125 singleton no-mutation patients with a clinical diagnosis of DFH, we would expect four to carry a shared mutated locus leading to the defective plasma clearance of LDL cholesterol. The power of the study is clearly dependent on the number of singletons included, with the idea that any identified candidate locus would be sequenced in the family members of the affected proband. While a group of singletons may be genetically heterogeneous, the use of the 'burden' analysis and not a single-variant test means that heterogeneity should not reduce power to detect a novel FH-causing gene. Another limitation is that we did not have lipid profile information for individuals in the control comparison cohort, only their rare disease phenotype status, which did not overlap with FH pathogenesis. The possibility that the control cohort includes FH-affected individuals was considered. Assuming that the prevalence of FH is 1/500, we would expect by chance to find ~4 individuals in this cohort carrying an FH-causing mutation. We have analysed variants in *LDLR*, *APOB* and *PCSK9* in the control cohort and identified three *LDLR* and two *APOB* mutations as incidental findings, which was similar to the expected FH frequency of one in 500. We have also allowed for this prevalence in the control comparison cohort by using a frequency cut-off of four novel gene functional variants in controls, in case any of the novel variants identified in FH cases were also present in the controls. A final limitation is that it is possible that some of the identified variants in the 18 genes in table 1 may be technical false positives, since only for the *CH25H* and *INSIG2* genes were all variants confirmed by Sanger Sequencing, However, to be as certain as possible using bioinformatics that the variants we observed are not false positives, for all these variants we included a visual validation of the quality of calls performed using the Human Genome 19 on the IGV.[27]

In summary, in 125 DFH unrelated patients without an identified mutation by conventional screening methods, analyses identified 25 disease-causing variants in already known FH loci, as well as six previously unreported *APOB* variants in five patients. LDL-C gene score analysis found that 31 (29 mutation negative) patients had an SNP score in the top decile of the general population and therefore had a definite polygenic aetiology, and an additional five had a potential functional variant in *CH25H* or *INSIG2*. This means that the explanation for the FH phenotype is still lacking in 50% of the patients, suggesting that some causal variants may have been missed at different stages of the data processing or analysis. The variant calling pipeline used for this study was carefully optimised for the majority of the exome regions, though some calls in poorly covered regions could be missed. There is a possibility that there are genetic causes located outside of the protein coding region, affecting protein expression, posttranscriptional stability or altering gene splicing. Also, it is possible that the LDL-C gene score cut-off of 1.16 for polygenic hypercholesterolaemia is too stringent. Thus, using the 9th decile cut-off of 1.08, in which a 41% of WHII individuals had LDL-C above the 4.9 mmol/L (mean LDL-C=4.68±1.05 mmol/L) FH diagnostic level, could be more appropriate. By doing so, the phenotype would be explained in an additional nine mutation-negative patients. A polygenic explanation in additional subjects might also be achieved if SNPs in recently identified LDL-C-raising loci[38] were included in the score. Finally, because the burden test results are dependent on the number of associated variants and variants diluting the signal, it is possible that novel FH mutations are located in a highly polymorphic gene, in which it is difficult to pick up the true mutation.

Thus, overall, no major novel locus for FH was detected, with no gene having a functional variant in more than three patients. This suggests that the genetic cause of FH in these unexplained cases is likely to be very heterogeneous, which complicates the novel gene discovery and diagnostic process.

Author affiliations
[1]British Heart Foundation Laboratories, Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, the Rayne Building University College London, London, UK
[2]Department of Genetics, Environment and Evolution, UCL Genetics Institute, University College London, London, UK
[3]Department of Primary Care Health Sciences, NIHR School of Primary Care Research, University of Oxford, Oxford, UK
[4]Department of Cardiology, Imperial College Health Services, Charing Cross Hospital, London, UK
[5]Department of Internal Medicine, University of Genoa, Genoa, Italy
[6]Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
[7]Centre de Recherche Médicale de Jolimont, Haine St-Paul, Belgium
[8]Queens University Belfast & Regional Genetics Centre, Belfast Health and Social Care Trust/City Hospital Belfast BT9 7AB Northern Ireland UK
[9]Robarts Research Institute, London, Ontario, Canada
[10]OCDEM, Radcliffe Department of Medicine, University of Oxford, Churchill Hospital, Oxford, UK
[11]Cardiology Department, Hadassah Hebrew University Medical Center, Jerusalem, Israel
[12]Department of Medicine, Center for Research, Prevention and Treatment of Atherosclerosis, Hadassah Hebrew University Medical Centre, Jerusalem, Israel
[13]North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children, London, UK
[14]Consultant Lipidologist and Chemical Pathologist Director SAS Laboratory for Cardiac Biomarkers, Royal Free Hospital, London, UK
[15]Cardiovascular Trials Unit, University Department of Medicine, Central Manchester University Hospital NHS Foundation Trust, Manchester, UK
[16]Division of Laboratory Medicine, Department of Biochemistry, Royal Perth Hospital, Perth, Australia
[17]The University of Western Australia, Perth, Australia

Acknowledgements This study makes use of data generated by the UK10K Consortium. A full list of the investigators who contributed to the generation of the data is available online (http://www.UK10K.org). Funding for UK10K was provided by the Wellcome Trust under award WT091310.

Collaborators UK10K Consortium (http://www.uk10k.org).

Contributors All authors listed on the manuscript fulfil the criteria for authorship. Specific contributions are: MF: writing of the manuscript, data analysis. VP: data analysis, CNVs calling. KWL: LDL-C gene score genotyping. RAW: sample preparation. HAWN, MS, SB, SC, OSD, CAG, RAH, FK, RD, EL, NL, DRN, HS, FMVB: patient clinical diagnosis, selection and study samples providers. UK10K: exome sequencing and data production. SEH: data analysis, project supervision, guarantor of the publication.

Funding SEH holds a Chair funded by the British Heart Foundation, and SEH and RAW are supported by the BHF (PG08/008) and by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. MF is funded by an MRC CASE award with Gen-Probe Life Sciences, and VP is partially funded by a MRC research grant (G1001158). HAWN is a NIHR Senior Investigator.

Competing interests None.

Ethics approval Ethical Governance Framework UK10K. National Research Ethics Service, Cambridgeshire 2 Research Committee (http://www.uk10k.org/ethics.html).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Marks D, Thorogood M, Neil HA, Humphries SE. A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. *Atherosclerosis* 2003;168:1–14.
2 Ciccarese M, Pacifico A, Tonolo G, Pintus P, Nikoshkov A, Zuliani G, Fellin R, Luthman H, Maioli M. A new locus for autosomal recessive hypercholesterolemia maps to human chromosome 15q25-q26. *Am J Hum Genet* 2000;66:453–60.

3 Neil HA, Hammond T, Huxley R, Matthews DR, Humphries SE. Extent of underdiagnosis of familial hypercholesterolaemia in routine practice: prospective registry study. *BMJ* 2000;321:148.
4 Slack J. Risks of ischaemic heart-disease in familial hyperlipoproteinaemic states. *Lancet* 1969;2:1380–2.
5 [No authors listed]. Mortality in treated heterozygous familial hypercholesterolaemia: implications for clinical management. Scientific Steering Committee on behalf of the Simon Broome Register Group. *Atherosclerosis* 1999;142:105–12.
6 Taylor A, Patel K, Tsedeke J, Humphries SE, Norbury G. Mutation screening in patients for familial hypercholesterolaemia (ADH). *Clini Genet* 2010;77:97–9.
7 Futema M, Whittall RA, Kiley A, Steel LK, Cooper JA, Badmus E, Leigh SE, Karpe F, Neil HA. Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic. *Atherosclerosis* 2013;229:161–8.
8 Graham CA, McIlhatton BP, Kirk CW, Beattie ED, Lyttle K, Hart P, Neely RD, Young IS, Nicholls DP. Genetic screening protocol for familial hypercholesterolemia which includes splicing defects gives an improved mutation detection rate. *Atherosclerosis* 2005;182:331–40.
9 Humphries SE, Norbury G, Leigh S, Hadfield SG, Nair D. What is the clinical utility of DNA testing in patients with familial hypercholesterolaemia? *Curr Opin Lipidol* 2008;19:362–8.
10 Wierzbicki AS, Humphries SE, Minhas R. Familial hypercholesterolaemia: summary of NICE guidance. *BMJ* 2008;337:a1095.
11 Neil A, Cooper J, Betteridge J, Capps N, McDowell I, Durrington P, Seed M, Humphries SE. Reductions in all-cause, cancer, and coronary mortality in statin-treated patients with heterozygous familial hypercholesterolaemia: a prospective registry study. *Eur Heart J* 2008;29:2625–33.
12 Humphries SE, Whittall RA, Hubbart CS, Maplebeck S, Cooper JA, Soutar AK, Naoumova R, Thompson GR, Seed M, Durrington PN, Miller JP, Betteridge DJ, Neil HA. Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk. *J Med Genet* 2006;43:943–9.
13 Usifo E, Leigh SE, Whittall RA, Lench N, Taylor A, Yeats C, Orengo CA, Martin AC, Celli J, Humphries SE. Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: update and pathological assessment. *Ann Hum Genet* 2012;76:387–401.
14 Taylor A, Wang D, Patel K, Whittall R, Wood G, Farrer M, Neely RD, Fairgrieve S, Nair D, Barbir M, Jones JL, Egan S, Everdale R, Lolin Y, Hughes E, Cooper JA, Hadfield SG, Norbury G, Humphries SE. Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. *Clini Genet* 2010;77:572–80.
15 Cenarro A, Garcia-Otin AL, Tejedor MT, Solanas M, Jarauta E, Junquera C, Ros E, Mozas P, Puzo J, Pocovi M, Civeira F. A presumptive new locus for autosomal dominant hypercholesterolemia mapping to 8q24.22. *Clini Genet* 2011;79:475–81.
16 Marques-Pinheiro A, Marduel M, Rabes JP, Devillers M, Villeger L, Allard D, Weissenbach J, Guerin M, Zair Y, Erlich D, Junien C, Munnich A, Krempf M, Abifadel M, Jais JP, Boileau C, Varret M. A fourth locus for autosomal dominant hypercholesterolemia maps at 16q22.1. *Eur J Hum Genet* 2010;18:1236–42.
17 Wang X, Li X, Zhang YB, Zhang F, Sun L, Lin J, Wang DM, Wang LY. Genome-wide linkage scan of a pedigree with familial hypercholesterolemia suggests susceptibility loci on chromosomes 3q25-26 and 21q22. *PLoS ONE* 2011;6:e24838.
18 Talmud PJ, Shah S, Whittall R, Futema M, Howard P, Cooper JA, Harrison SC, Li K, Drenos F, Karpe F, Neil HA, Descamps OS, Langenberg C, Lench N, Kivimaki M, Whittaker J, Hingorani AD, Kumari M, Humphries SE. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet* 2013;381:1293–301.
19 Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, Wood NW, Hambleton S, Burns SO, Thrasher AJ, Kumararatne D, Doffinger R, Nejentsev S. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 2012;28:2747–54.
20 Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
21 Usifo E, Leigh SE, Whittall RA, Lench N, Taylor A, Yeats C, Orengo CA, Martin AC, Celli J, Humphries SE. Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: update and pathological assessment. *Ann Hum Genet* 2012;76:387–401.
22 Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, Konig IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA Jr, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466:707–13.
23 Marmot MG, Smith GD, Stansfeld S, Patel C, North F, Head J, White I, Brunner E, Feeney A. Health inequalities among British civil servants: the Whitehall II study. *Lancet* 1991;337:1387–93.
24 Thomas ER, Atanur SS, Norsworthy PJ, Encheva V, Snijders AP, Game L, Vandrovcova J, Siddiq A, Seed M, Soutar AK, Aitman TJ. Identification and biochemical analysis of a novel APOB mutation that causes autosomal dominant hypercholesterolemia. *Mol Genet Genomic Med* 2013;1:155–61.
25 Garcia CK, Wilund K, Arca M, Zuliani G, Fellin R, Maioli M, Calandra S, Bertolini S, Cossu F, Grishin N, Barnes R, Cohen JC, Hobbs HH. Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* 2001;292:1394–8.
26 Mozas P, Cenarro A, Civeira F, Castillo S, Ros E, Pocovi M. Mutation analysis in 36 unrelated Spanish subjects with familial hypercholesterolemia: identification of 3 novel mutations in the LDL receptor gene. *Hum Mutat* 2000;15:483–4.
27 Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–92.
28 Radhakrishnan A, Ikeda Y, Kwon HJ, Brown MS, Goldstein JL. Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: oxysterols block transport by binding to Insig. *Proc Natl Acad Sci USA* 2007;104:6511–18.
29 Fox BG, Shanklin J, Ai J, Loehr TM, Sanders-Loehr J. Resonance Raman evidence for an Fe-O-Fe center in stearoyl-ACP desaturase. Primary sequence identity with other diiron-oxo proteins. *Biochemistry* 1994;33:12776–86.
30 Holmes RS, Vandeberg JL, Cox LA. Genomics and proteomics of vertebrate cholesterol ester lipase (LIPA) and cholesterol 25-hydroxylase (CH25H). *3 Biotech* 2011;1:99–109.
31 Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC bioinformatics* 2010;11:548.
32 Futema M, Plagnol V, Whittall RA, Neil HA, Humphries SE. Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia. *J Med Genet* 2012;49:644–9.
33 Thomas ERA, Atanur SS, Norsworthy PJ, Encheva V, Snijders AP, Game L, Vandrovcova J, Siddiq A, Seed M, Soutar AK, Aitman TJ. Identification and biochemical analysis of a novel APOB mutation that causes autosomal dominant hypercholesterolemia. *Mol Genet Genomic Med* 2013;1:155–61.
34 Boren J, Lee I, Zhu W, Arnold K, Taylor S, Innerarity TL. Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *J Clin Invest* 1998;101:1084–93.
35 Stitziel NO, Fouchier SW, Sjouke B, Peloso GM, Moscoso AM, Auer PL, Goel A, Gigante B, Barnes TA, Melander O, Orho-Melander M, Duga S, Sivapalaratnam S, Nikpay M, Martinelli N, Girelli D, Jackson RD, Kooperberg C, Lange LA, Ardissino D, McPherson R, Farrall M, Watkins H, Reilly MP, Rader DJ, de Faire U, Schunkert H, Erdmann J, Samani NJ, Charnas L, Altshuler D, Gabriel S, Kastelein JJ, Defesche JC, Nederveen AJ, Kathiresan S, Hovingh GK. Exome sequencing and directed clinical phenotyping diagnose cholesterol ester storage disease presenting as autosomal recessive hypercholesterolemia. *Arterioscler Thromb Vasc Biol* 2013;33:2909–14.
36 Adams CM, Reitz J, De Brabander JK, Feramisco JD, Li L, Brown MS, Goldstein JL. Cholesterol and 25-hydroxycholesterol inhibit activation of SREBPs by different mechanisms, both involving SCAP and Insigs. *J Biol Chem* 2004;279:52772–80.
37 Shao W, Espenshade PJ. Expanding roles for SREBP in metabolism. *Cell Metab* 2012;16:414–19.
38 Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog HM, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF,

Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkila K, Hypponen E, Isaacs A, Jackson AU, Johansson A, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J, Lyytikainen LP, Magnusson PK, Mangino M, Mihailov E, Montasser ME, Muller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney AS, Doring A, Elliott P, Epstein SE, Eyjolfsson GI, Gigante B, Goodarzi MO, Grallert H, Gravito ML, Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T, Jones MR, Kaleebu P, Kastelein JJ, Khaw KT, Kim E, Klopp N, Komulainen P, Kumari M, Langenberg C, Lehtimaki T, Lin SY, Lindstrom J, Loos RJ, Mach F, McArdle WL, Meisinger C, Mitchell BD, Muller G, Nagaraja R, Narisu N, Nieminen TV, Nsubuga RN, Olafsson I, Ong KK, Palotie A, Papamarkou T, Pomilla C, Pouta A, Rader DJ, Reilly MP, Ridker PM, Rivadeneira F, Rudan I, Ruokonen A, Samani N, Scharnagl H, Seeley J, Silander K, Stancakova A, Stirrups K, Swift AJ, Tiret L, Uitterlinden AG, van Pelt LJ, Vedantam S, Wainwright N, Wijmenga C, Wild SH, Willemsen G, Wilsgaard T, Wilson JF, Young EH, Zhao JH, Adair LS, Arveiler D, Assimes TL, Bandinelli S, Bennett F, Bochud M, Boehm BO, Boomsma DI, Borecki IB, Bornstein SR, Bovet P, Burnier M, Campbell H, Chakravarti A, Chambers JC, Chen YD, Collins FS, Cooper RS, Danesh J, Dedoussis G, de Faire U, Feranil AB, Ferrieres J, Ferrucci L, Freimer NB, Gieger C, Groop LC, Gudnason V, Gyllensten U, Hamsten A, Harris TB, Hingorani A, Hirschhorn JN, Hofman A, Hovingh GK, Hsiung CA, Humphries SE, Hunt SC, Hveem K, Iribarren C, Jarvelin MR, Jula A, Kahonen M, Kaprio J, Kesaniemi A, Kivimaki M, Kooner JS, Koudstaal PJ, Krauss RM, Kuh D, Kuusisto J, Kyvik KO, Laakso M, Lakka TA, Lind L, Lindgren CM, Martin NG, Marz W, McCarthy MI, McKenzie CA, Meneton P, Metspalu A, Moilanen L, Morris AD, Munroe PB, Njolstad I, Pedersen NL, Power C, Pramstaller PP, Price JF, Psaty BM, Quertermous T, Rauramaa R, Saleheen D, Salomaa V, Sanghera DK, Saramies J, Schwarz PE, Sheu WH, Shuldiner AR, Siegbahn A, Spector TD, Stefansson K, Strachan DP, Tayo BO, Tremoli E, Tuomilehto J, Uusitupa M, van Duijn CM, Vollenweider P, Wallentin L, Wareham NJ, Whitfield JB, Wolffenbuttel BH, Ordovas JM, Boerwinkle E, Palmer CN, Thorsteinsdottir U, Chasman DI, Rotter JI, Franks PW, Ripatti S, Cupples LA, Sandhu MS, Rich SS, Boehnke M, Deloukas P, Kathiresan S, Mohlke KL, Ingelsson E, Abecasis GR. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45:1274–83.

SUPPLEMENTARY METHODS

Control samples cohorts:

UK10K_NEURO_ABERDEEN (n=387)

UK10K_NEURO_ASD_GALLAGHER (n=75)

UK10K_NEURO_EDINBURGH (n=233)

UK10K_NEURO_GURLING (n=48)

UK10K_NEURO_IOP_COLLIER (n=172)

UK10K_NEURO_MUIR (n=166)

UK10K_OBESITY_GS (n=421)

UK10K_OBESITY_TWINSUK (n=67)

UK10K_RARE_CILIOPATHIES (n=121)

UK10K_RARE_NEUROMUSCULAR (n=114)

UK10K_RARE_THYROID (n=122)

For more details see http://www.uk10k.org/studies/ .

Whole exome sequencing

Genomic DNA (1–3 μg), extracted from blood (*1*), was sheared to 100–400 bp using a Covaris E210 or LE220 (Covaris, Woburn, Massachusetts, USA). Sheared DNA was subjected to Illumina paired-end DNA library preparation and enriched for target sequences (Agilent Technologies, Santa Clara, CA, USA; Human All Exon 50 Mb - ELID S02972011) according to the manufacturer's recommendations (Agilent Technologies, Santa Clara, CA, USA; SureSelectXT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing). Enriched libraries were sequenced (eight samples over two lines) using the HiSeq 2000 platform (Illumina) as paired-end 75 base reads according to the manufacturer's protocol.

Variant calling

Calls were made using samtools/bcftools version 0.1.19-3-g4b70907 from all UK10K per-sample exome BAMs split by chromosome. A BCF file was created with samtools mpileup, calculating genotype likelihoods for every site in the bait (+/-100bp) regions file then variants (SNPs and Indels) were called by bcftools.

SUPPLEMENTARY RESULTS

Gene coverage

The overall mean coverage of *LDLR*, *APOB*, *PCSK9* and *LDLRAP1* ranged from 42x (*LDLR*), to 18x (*PCSK9*), with the first and the last exons of a gene having the lowest coverage. The read depth was highly dependent on the GC content of an exon (regression $p = 4.9 \times 10^{-14}$) (Figure S2). Exons of the *APOB* had the highest average read depth among Tier 1 genes (58x).

LDL-C SNPs score

The distribution of LDL-C SNPs scores in FH mutation negative patients and in the healthy WHII population was as shown in Figure S3.

The *APOE* ε2ε2 genotype was not observed among the genotyped patients. There were two individuals with the ε2ε3 genotype, both having an *LDLR* mutation. Five patients had the ε4ε4 isoform.

Figure S1.
Copy Number Variants (CNVs) in *LDLR* gene. A: Heterozygous duplication of exons 3 to 8. B: Heterozygous deletion of exons 11 and 12. C: Heterozygous duplication of exons 13 to 15. All identified by ExomeDepth in the exome sequencing data. The crosses show the ratio of observed/expected number of reads for the test sample. The grey shaded region shows the estimated 99% confidence interval for this observed ratio in the absence of CNV call. The presence of contiguous exons with read count ratio located outside of the confidence interval is indicative of a heterozygous deletion or duplication in a sample. Exons 1 and 18 were excluded from the analysis (not shown on the graph) as they did not reach the threshold of 100 for the total number of reads. All CNVs were confirmed by MLPA experiment. The deletion of exons 11-12 and duplication of exons 13-15 both lead to a frame shift. The duplication of exons 3-8 leads to elongated peptide and it has been previously found in FH patients (*2*).

A B C

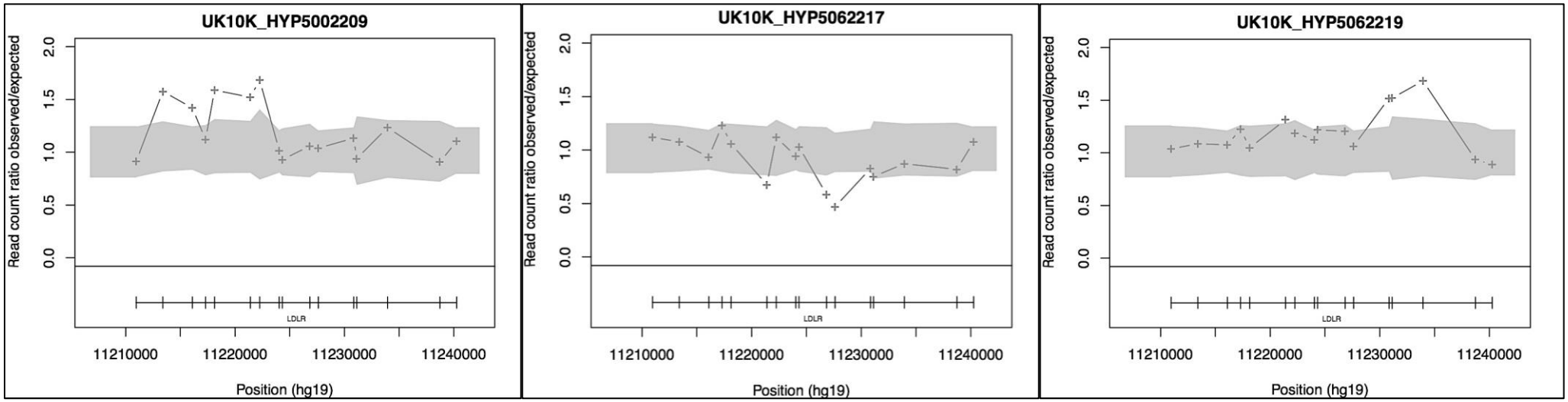


Figure S2.
The negative correlation of the median read depth and the GC content for each targeted exon of the four FH genes (*LDLR*, *APOB*, *PCSK9* and *LDLRAP1*).

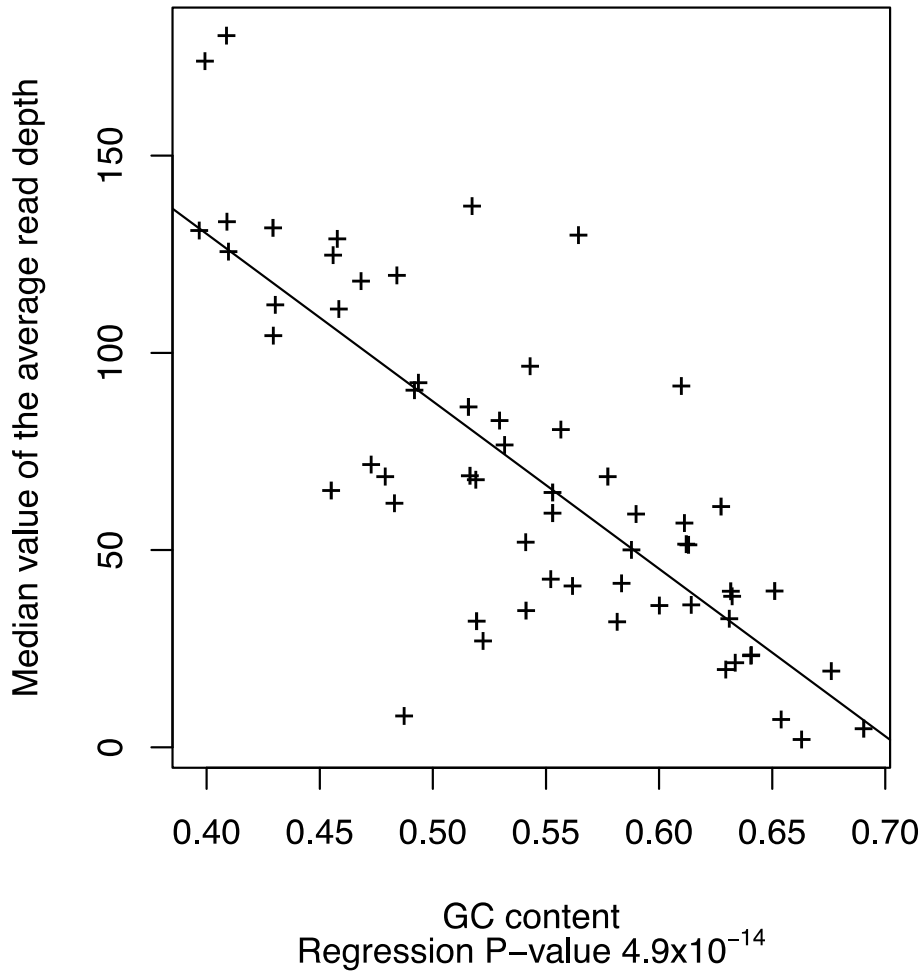


Figure S3.
Distribution of the LDL-C SNPs score in mutation negative DFH patients (in purple) and in the healthy WHII cohort (in grey). Red line indicates the LDL-C score top decile cutoff for WHII (=1.16).

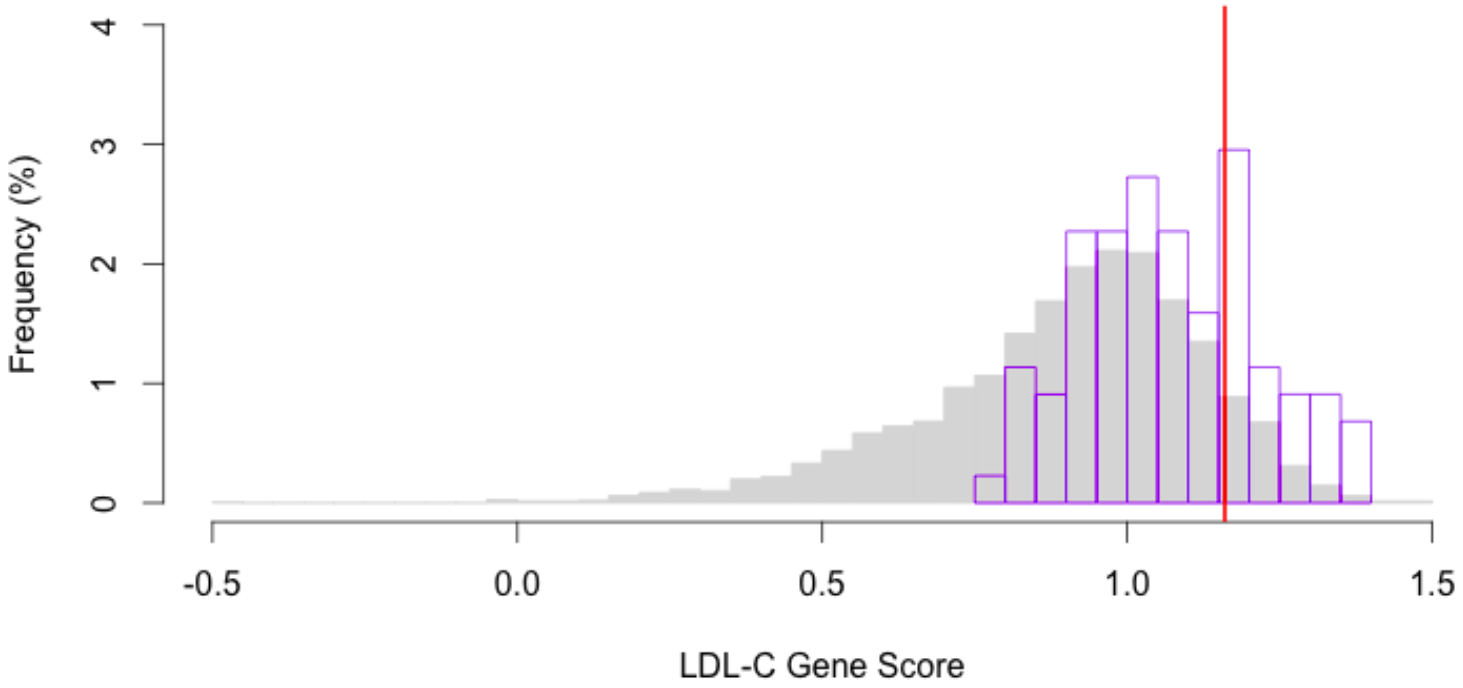


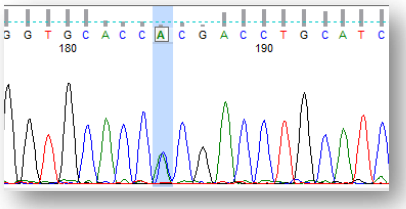
Figure S4.
Sanger sequencing confirmation of novel *CH25H* variants. Primers used for the amplification of the region are highlighted in blue and in purple

CH25H sequencing (order #410822401)

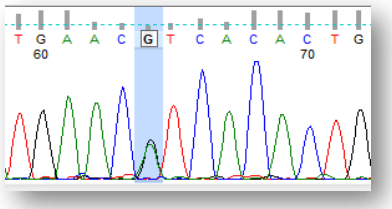
```
90966631 TYGTGTGGCACCTGCTGCACCACAAGGTGCCCTGGYTGTACCGCACCTTYCACAAGGTGC 90966572
90966571 ACCACCRGAACTCGTCCTCGTTCGCRCTGGCAACGCAGTATATGAGCGTCTGGGAACTGT 90966512
90966511 TTTCTTTGGGYTTCTTCGACATGATGAACGTCACACTGCTYGGGTGCCACCCGCTCACCA 90966452
90966451 CCCTGACCTTCCACGTGGTCAACATCTGKCTTTCCGTGGAGGWCCACTCCGGCTACAAYT 90966392
90966391 TCCCTTGGTCCACTCACAGACTGGTGCCCTTCGGGTGGTACGGGGGTGTGGTRCACCACG 90966332
90966331 ACCTGCATCACTCTCACTTTAACTGCAACTTCGCTCCRTACTTTACACACTGGGACAAAA 90966272
90966271 YACTGGGAACRCTGCGRACTGCATCTGTCCCAGCGCRRTGATGTGGCTGCGGTGGGTGCC 90966212
90966211 CCTAAGAMTCGGGACTGCTRTGCCTTTCACACTTGAATGAAGAGAAACACCTGAGCTATA 90966152
90966151 TATTTTTTTAAAGCAACTAACTTATTRCTTTATGTTTATCTATGAAAACCATAGATAAAA 90966092
90966091 TCTGATGCATTTTTGTAATCTGACAAAGTAATTTACATACTGTTTGTGTATCAATACAAT 90966032
```

*Primers:
CH25H_01F / CH25H_02R

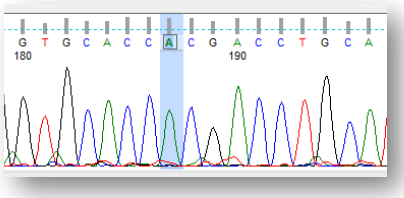
p.H239P (c.716A>C) (UK10K_HYP5231652)



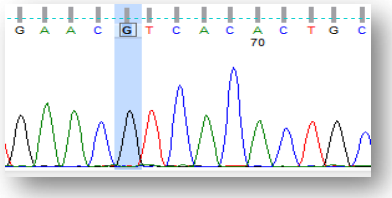
p.V190I (c.568G>A) (UK10K_HYP5231677)



WT (UK10K_HYP5231677)



WT (UK10K_HYP5231652)

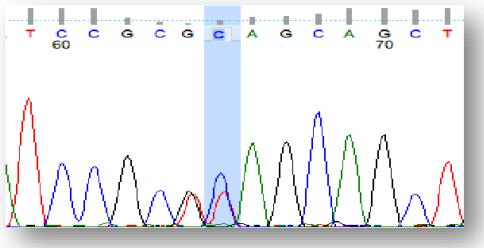


CH25H sequencing (order #4108796)

```
90966991 TGCAGCCCCTCTGGGACCACCTGAGGAGCTGGGAGGCCCTCCTACAGTCGCCCTTCTTCC 90966932
90966931 CGGTCATCTTCTCCATCACCACATACRTRGGCTTTTGCCTGCCCTTCGTGGTCCTGGATA 90966872
90966871 TCCTGTGCTCCTGGGTGCCCGCCCTGCGGCGCTACAAGATCCAYCCTGACTTCTCGCCAT 90966812
90966811 CCGCGCAGCAGCTGCTACCTTGCCTGVGRCAGACCCTCTACCAGCAKGTGATGTTTGTGT 90966752
90966751 TCCCCGTRACGCTGMTGCAYTGGGCCYGCAGCCCGGCCCTCCTGCCCCACGAAGCTCCCG 90966692
90966691 AGCTGCTCCTGCTGCTGCACCACATCCTGTTCTGCCTGCYACTCTTCGACATGGAGTTCT 90966632
90966631 TYGTGTGGCACCTGCTGCACCACAAGGTGCCCTGGYTGTACCGCACCTTYCACAAGGTGC 90966572
90966571 ACCACCRGAACTCGTCCTCGTTCGCRCTGGCAACGCAGTATATGAGCGTCTGGGAACTGT 90966512
90966511 TTTCTTTGGGYTTCTTCGACATGATGAACGTCACACTGCTYGGGTGCCACCCGCTCACCA 90966452
90966451 CCCTGACCTTCCACGTGGTCAACATCTGKCTTTCCGTGGAGGWCCACTCCGGCTACAAYT 90966392
90966391 TCCCTTGGTCCACTCACAGACTGGTGCCCTTCGGGTGGTACGGGGGTGTGGTRCACCACG 90966332
90966331 ACCTGCATCACTCTCACTTTAACTGCAACTTCGCTCCRTACTTTACACACTGGGACAAAA 90966272
90966271 YACTGGGAACRCTGCGRACTGCATCTGTCCCAGCGCRRTGATGTGGCTGCGGTGGGTGCC 90966212
```

*Primers:
CH25H _03F / CH25H_04R

p.A80A (c.243G>T) and p.Q81* (c.244C>T) (UK10K_HYP5159267)



WT (UK10K_HYP5002218)

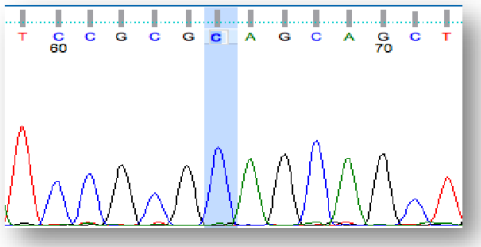


Figure S5.
Sanger sequencing confirmation of novel INSIG2 variants. Primers used for the amplification of the region are highlighted in yellow.

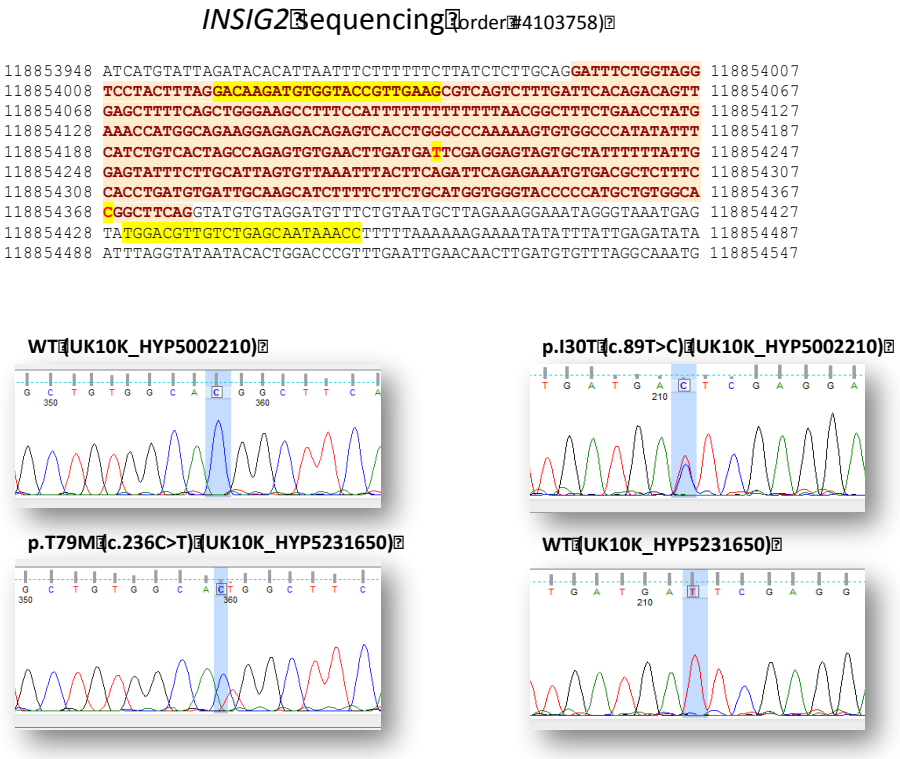


Table S1.
Summary of methods used for the initial FH mutation screening.

BATCH	UK10K ID	Original study cohort	Ref	*LDLR*	MLPA of *LDLR*	*APOB*	*PCSK9*
4	UK10K_HYP5231659	Australian FH	(*3*)	all exons and promoter by Big Dye Terminator chemistry (Applied Biosystems) sequencing	yes	fragment of exon 26, and exon 29 by Big Dye Terminator chemistry (Applied Biosystems) sequencing	exon 7 by Big Dye Terminator chemistry (Applied Biosystems) sequencing
4	UK10K_HYP5231660	Australian FH					
4	UK10K_HYP5231661	Australian FH					
4	UK10K_HYP5231662	Australian FH					
4	UK10K_HYP5231663	Australian FH					
4	UK10K_HYP5231664	Australian FH					
4	UK10K_HYP5231665	Australian FH					
4	UK10K_HYP5231666	Australian FH					
4	UK10K_HYP5231667	Australian FH					
4	UK10K_HYP5231668	Australian FH					
4	UK10K_HYP5231669	Australian FH					
4	UK10K_HYP5231670	Australian FH					
4	UK10K_HYP5231671	Australian FH					
4	UK10K_HYP5231672	Australian FH					
4	UK10K_HYP5231673	Australian FH					
4	UK10K_HYP5231674	Australian FH					
4	UK10K_HYP5231675	Australian FH					
5	UK10K_HYP5269604	Israeli FH	-	all exons and promoterby SSCP	no	APOB fragment of exon 26 by SSCP	no
5	UK10K_HYP5269605	Israeli FH					
5	UK10K_HYP5269606	Israeli FH					
5	UK10K_HYP5269607	Israeli FH					
5	UK10K_HYP5269608	Israeli FH					
5	UK10K_HYP5269609	Israeli FH					
7	UK10K_HYP5358903	Israeli FH					
7	UK10K_HYP5358904	Israeli FH					
7	UK10K_HYP5358905	Israeli FH					
7	UK10K_HYP5358906	Israeli FH					
4	UK10K_HYP5231679	Italy FH	(*4*)	all exons and	yes	c.9216 to	all exons by

4	UK10K_HYP5231676	Italy FH		promoter by Sanger sequencing		c.11788 +152 nt of intron 26 by Sanger sequencing	Sanger sequencing
4	UK10K_HYP5231677	Italy FH					
4	UK10K_HYP5231678	Italy FH					
5	UK10K_HYP5269570	Northern Ireland FH	(*5*)	all exons and promoter by TTGE/DDGE	yes	RFLP for p.R3527Q	Exon 7 by Sanger sequencing
5	UK10K_HYP5269571	Northern Ireland FH					
5	UK10K_HYP5269572	Northern Ireland FH					
5	UK10K_HYP5269573	Northern Ireland FH					
5	UK10K_HYP5269574	Northern Ireland FH					
5	UK10K_HYP5269575	Northern Ireland FH					
5	UK10K_HYP5269576	Northern Ireland FH					
5	UK10K_HYP5269577	Northern Ireland FH					
5	UK10K_HYP5269578	Northern Ireland FH					
5	UK10K_HYP5269581	Northern Ireland FH					
3	UK10K_HYP5159271	Oxford FH	(*6*)	all exons and promoter by HRM	yes	fragment of exon 26 by HRM and ARMS	ARMS for p.D374Y
3	UK10K_HYP5159272	Oxford FH					
3	UK10K_HYP5159273	Oxford FH					
3	UK10K_HYP5159274	Oxford FH					
3	UK10K_HYP5159275	Oxford FH					
4	UK10K_HYP5231650	Oxford FH					
4	UK10K_HYP5231651	Oxford FH					
4	UK10K_HYP5231652	Oxford FH					
4	UK10K_HYP5231653	Oxford FH					
4	UK10K_HYP5231654	Oxford FH					
4	UK10K_HYP5231655	Oxford FH					
4	UK10K_HYP5231656	Oxford FH					
4	UK10K_HYP5231657	Oxford FH					
4	UK10K_HYP5231658	Oxford FH					
3	UK10K_HYP5159266	RFH	(*7*)	all exons and promoter by SSCP or Sanger sequencing	yes	ARMS for p.R3527Q	ARMS for p.D374Y
3	UK10K_HYP5159267	RFH					
3	UK10K_HYP5159268	RFH					
3	UK10K_HYP5159269	RFH					
3	UK10K_HYP5159270	RFH					
1	UK10K_HYP5002209	SBBHF	(*8*-	all exons and	yes	RFLP for	all exons by HRM

1	UK10K_HYP5002210	SBBHF	*11*)	promoter by HRM		p.R3527Q	
1	UK10K_HYP5002211	SBBHF					
1	UK10K_HYP5002212	SBBHF					
1	UK10K_HYP5002213	SBBHF					
1	UK10K_HYP5002214	SBBHF					
1	UK10K_HYP5002215	SBBHF					
1	UK10K_HYP5002216	SBBHF					
1	UK10K_HYP5002217	SBBHF					
1	UK10K_HYP5002218	SBBHF					
1	UK10K_HYP5002219	SBBHF					
1	UK10K_HYP5002220	SBBHF					
1	UK10K_HYP5002221	SBBHF					
1	UK10K_HYP5002222	SBBHF					
1	UK10K_HYP5002223	SBBHF					
1	UK10K_HYP5002224	SBBHF					
1	UK10K_HYP5002225	SBBHF					
1	UK10K_HYP5002226	SBBHF					
1	UK10K_HYP5002227	SBBHF					
1	UK10K_HYP5002228	SBBHF					
1	UK10K_HYP5002229	SBBHF					
1	UK10K_HYP5002230	SBBHF					
1	UK10K_HYP5002231	SBBHF					
1	UK10K_HYP5002232	SBBHF					
2	UK10K_HYP5062209	SBBHF					
2	UK10K_HYP5062210	SBBHF					
2	UK10K_HYP5062211	SBBHF					
2	UK10K_HYP5062212	SBBHF					
2	UK10K_HYP5062213	SBBHF					
2	UK10K_HYP5062214	SBBHF					
2	UK10K_HYP5062215	SBBHF					
2	UK10K_HYP5062216	SBBHF					
2	UK10K_HYP5062217	SBBHF					
2	UK10K_HYP5062218	SBBHF					

2	UK10K_HYP5062219	SBBHF					
2	UK10K_HYP5062220	SBBHF					
2	UK10K_HYP5062221	SBBHF					
2	UK10K_HYP5062222	SBBHF					
2	UK10K_HYP5062223	SBBHF					
2	UK10K_HYP5062224	SBBHF					
2	UK10K_HYP5062225	SBBHF					
2	UK10K_HYP5062226	SBBHF					
2	UK10K_HYP5062227	SBBHF					
2	UK10K_HYP5062228	SBBHF					
2	UK10K_HYP5062229	SBBHF					
2	UK10K_HYP5062230	SBBHF					
2	UK10K_HYP5062231	SBBHF					
2	UK10K_HYP5062232	SBBHF					
5	UK10K_HYP5269585	SBBHF					
5	UK10K_HYP5269589	SBBHF					
5	UK10K_HYP5269595	SBBHF					
5	UK10K_HYP5269597	SBBHF					
5	UK10K_HYP5269598	SBBHF					
5	UK10K_HYP5269601	SBBHF					
5	UK10K_HYP5269602	SBBHF					
6	UK10K_HYP5315266	SBBHF					
6	UK10K_HYP5315268	SBBHF					
6	UK10K_HYP5315271	SBBHF					
6	UK10K_HYP5315273	SBBHF					
6	UK10K_HYP5315275	SBBHF					
7	UK10K_HYP5358898	SBBHF					
7	UK10K_HYP5358899	SBBHF					
7	UK10K_HYP5358900	SBBHF					
7	UK10K_HYP5358901	SBBHF					
7	UK10K_HYP5358902	SBBHF					

Table S2.

Tier 2 candidate genes – LDL-C (lead trait) associated loci from Teslovich et al. GWAS meta-analysis (either a plausible biological candidate gene in the locus or the nearest annotated gene to the lead SNP) (*12*). Where associated SNP was located in a gene cluster, other genes in the region were included.

Gene ID
ABCG5
ABCG8
ABO
APOE
APOC1(APOE locus)
TOMM40 (APOE locus)
PVRL2 (APOE locus)
HFE
LPA
MYLIP
NYNRIN
OSBPL7
SORT1
CELSR2 (SORT1 locus)
PSRC1 (SORT1 locus)
ST3GAL4
DCPS (ST3GAL4 locus)
KIRREL3
TOP1
PLCG1 (TOP1 locus)
ZHX3 (TOP1 locus)
LPIN3(TOP1 locus)
PLEC1
PARP10 (PLEC locus)
GRINA (PLEC locus)
SPATC1 (PLEC locus)
OPLAH (PLEC locus)
EXOSC4 (PLEC locus)
GPAA1 (PLEC locus)
KIAA1875 (PLEC locus)
CYC1 (PLEC locus)
SHARPIN (PLEC locus)
MAF1 (PLEC locus)

Table S3.
The top LDL-rising SNPs and their effects (as reported by the GLGC) used for the LDL-C gene score genotyping and calculation.

SNP ID	Nearest gene	Risk Allele	Beta coefficient (mmol/l)
rs2479409	*PCSK9*	G	0.051978278
rs629301	*CELSR2*	T	0.146108094
rs1367117	*APOB*	A	0.104732351
rs4299376 (rs6544713)	*ABCG8*	T	0.071114559
rs3757354	*MYLIP*	C	0.036979571
rs1800562	*HFE*	G	0.057408844
rs1564348	*SLC22A1*	T	0.01448151
rs4055111 (rs11220462)	*ST3GAL4*	G	0.050426687
rs8017377	*NYNRIN*	A	0.029480217
rs6511720	*LDLR*	G	0.180760279

Table S4.
Summary of the identified *LDLR* mutations and their *in silico* predicted effect, including calculated LDL-C gene scores for the mutations carriers (presented in bold are the gene scores that are above the top decile cutoff for the control population).

Mutation type/Exon	Mutation	Gene Score	PolyPhen	SIFT	Mutation Taster
Missense					
4	c.326G>A (p.C109Y)	1.03	Probably damaging	Not tolerated	Disease Causing
4	2X c.502G>C (p.D168H)	0.91, N/A	Probably damaging	Not tolerated	Disease Causing
4	c.681C>G (p.D227E)	N/A	Probably damaging	Not tolerated	Disease Causing
9	2X c.1196C>A (p.A399D)	1.03 and 1.07	Possibly damaging	Not tolerated	Disease Causing
11	c.1690A>C (p.N564H)[1]	1.17	Probably damaging	Tolerated	Disease Causing
12	c.1823C>T (p.P608L)	1.09	Probably damaging	Not tolerated	Disease Causing
14	c.2054C>T (p.P685L)	0.97	Probably damaging	Not tolerated	Disease Causing
17	c.2479G>A (p.V827I)	0.92	Probably damaging	Not tolerated	Disease Causing
Nonsense					
4	2X c.682G>T (p.E228*)	0.78 and 0.84	NA	NA	Disease Causing
7	c.1048C>T (p.R350*)	1.11	NA	NA	Disease Causing
8	c.1150C>T (p.Q384*)	0.65	NA	NA	Disease Causing
11	c.1685G>A (p.W562*)	0.95	NA	NA	Disease Causing
Indels					
5	c.695-6_698del	1.23	NA	NA	Disease Causing
12	2X c.1776_1778del p.G592del	N/A	NA	NA	Disease Causing
Intronic					
intron14	c.2140+1G>A	0.58	NA	NA	Disease Causing
intron9	c.1359-31_1359-23 delinsCGGCT	0.92	NA	NA	NA
Large rearrangements					
3_8	c.191-?_1186+?dup	1.03	10kb in frame duplication, peptide elongation		
11_12	c.1587-?_1845+?del	N/A	4kb out of frame deletion, truncated protein		
13_15	c.1846-?_2311+?dup	0.92	7kb out of frame duplication, truncated peptide		

N/A - not available

NA - not applicable

1 – carrier of this variant also has a deletion in exon 17 of *LDLR* c.2393_2401del9 (p.L799_V801del))

Table S5.

All *novel functional APOB* variants identified in the FH cases, including *in silico* predictions of their effect and LDL-C gene scores for the corresponding variant carriers. Using *in silico* mutation prediction tools (PolyPhen2, SIFT, Mutation Taster) the variant located in exon 3 of *APOB* (c.148C>T (p.R50W) has been predicted to be pathogenic by all three algorithms. The mutant Tryptophan is bigger than the wild type Arginine and it is predicted to cause a loss of hydrogen bonds in the core of the protein, which may result in an incorrect folding. The variant has been recently shown to co-segregate with the disease (Thomas et al., *Molecular Genetics & Genomic Medicine2013; 1(3) 155–161*). Other variants include c.598G>A (p.A200T), c.1199G>A (p.R400H), and c.G2700G>T (p.Q900H) in both cases the mutant differs in size and hydrophobicity from the wild type residue, which may affect the folding of the protein as well as the hydrophobic interactions within the protein's core. The novel c.10277C>T (p.A3426V) variant is located near to the LDL-receptor-binding site (*13*), and although it has been predicted as benign/tolerated/polymorphism by the *in silico* tools, it may affect the LDL-R/ApoB interaction. The known FH-causing mutation (p.R3527Q), which was found in two patients, is also listed.

Exon	Variant	Gene Score	PolyPhen	SIFT	MutationTaster	ID
3	c.148C>T(p.R50W)	0.83	Probably Damaging	Not Tolerated	Disease Causing	HYP5062228
6	c.598G>A (p.A200T)	0.98	Possibly Damaging	Not Tolerated	Polymorphism	HYP5269576
10	c.1199G>A(p.R400H)	N/A	Benign	Not Tolerated	Polymorphism	HYP5159267
18	c.G2700G>T (p.Q900H)	1.19	Probably Damaging	Not Tolerated	Polymorphism	HYP5358899
26	c.10277C>T (p.A3426V) and	1.17	Benign	Tolerated	Polymorphism	HYP5002222
	c.6639_6641delTGA (p.2213_2214delD)		NA	NA	Disease Causing	
	2 X c.G10580G>A (p.R3527Q)	0.71 and 1.01	Probably Damaging	Not Tolerated	Disease Causing	HYP5062226 and HYP5062216

NA- not applicable.

Table S6.
Top *p* values of the *novel functional* variant association between cases and controls in the Tier 2 candidate genes.

Gene	Variants in cases (n=71)	Variants in controls (n=1,926)	*p* value
KIAA1875	3	13	0.02
NYNRIN	3	18	0.04
CYC1	1	4	0.17
HFE	1	4	0.17
TOP1	1	4	0.17
ZHX3	2	20	0.18
PVRL2	1	7	0.25
ABCG8	1	18	0.50
OSBPL7	1	19	0.52
CELSR2	1	35	0.73
ABCG5	0	6	1
APOC1	0	2	1
APOE	0	2	1
DCPS	0	8	1
EXOSC4	0	8	1
GPAA1	0	21	1
GRINA	0	9	1
KIRREL3	0	9	1
LPA	0	34	1
LPIN3	0	17	1
MAF1	0	5	1
MYLIP	0	8	1
PARP10	0	11	1
PLCG1	0	25	1
PSRC1	0	4	1
SHARPIN	0	4	1
SORT1	0	9	1
SPATC1	0	12	1
ST3GAL4	0	13	1
TOMM40	0	3	1

Table S7.

Gene burden test of *novel functional* variants for genes in loci associated with FH in family linkage studies.

Chromosomal region	Gene name	Number of rare functional variants		p value
		cases (n=71)	controls (n=1926)	
21q22 (*14*)	*KRTAP10-11*	2	2	0.02
	PFKL	2	4	0.03
	DSCR8	1	0	0.05
	KRTAP11-1	1	0	0.05
	ERG	2	8	0.09
	KRTAP19-8	1	1	0.10
	LRRC3	2	10	0.13
	RCAN1	1	2	0.15
	SIM2	1	2	0.15
	SYNJ1	3	24	0.16
	CBR1	2	15	0.22
	COL18A1	3	30	0.24
	ZNF295	2	17	0.26
	C21orf59	1	5	0.27
	C21orf90	1	5	0.27
	ETS2	1	5	0.27
	C21orf56	1	6	0.31
	KRTAP12-4	1	6	0.31
	PKNOX1	1	6	0.31
	PCNT	5	68	0.33
	BACH1	1	7	0.35
	MX1	1	7	0.35
	BRWD1	2	24	0.39
	PRDM15	2	25	0.41
	HLCS	1	9	0.41
	PRMT2	1	9	0.41
	DOPEY2	3	43	0.43
	SUMO3	1	10	0.44
	MX2	1	11	0.47
	TTC3	2	29	0.48
	AIRE	1	13	0.53
	TRPM2	2	35	0.58
	ABCG1	1	16	0.60
	FTCD	1	17	0.62
	ITSN1	1	20	0.67
	LSS	1	20	0.67
	DSCAM	1	21	0.69
	COL6A2	1	23	0.72
	C21orf2	1	24	0.74
	TRAPPC10	1	24	0.74
	TSPEAR	1	24	0.74
	UMODL1	1	30	0.81
	ITGB2	1	33	0.84
	MCM3AP	1	36	0.86
	URB1	1	57	0.95

16q22 (*15*)	*CMTM2*	2	3	0.02
	HSF4	2	6	0.06
	KCTD19	2	7	0.08
	CES4A	2	8	0.09
	CMTM4	1	2	0.15
	TMEM208	1	3	0.19
	CMTM3	1	5	0.27
	DPEP3	1	5	0.27
	TMCO7	1	5	0.27
	PLEKHG4	2	19	0.30
	C16orf48	1	6	0.31
	CDH16	2	20	0.32
	CES3	1	11	0.47
	CDH3	1	13	0.53
	COG4	1	14	0.55
	GFOD2	1	16	0.60
	TSNAXIP1	1	17	0.62
	FHOD1	1	19	0.65
	SLC12A4	1	19	0.65
	FUK	1	23	0.72
8q24 (*16*)	*WISP1*	3	25	0.17
	ST3GAL1	1	8	0.38
	ZFAT	2	28	0.46
3q25 (*14*)	*HPS3*	3	15	0.06
	ZBBX	2	8	0.09
	NMD3	1	1	0.10
	TRIM59	1	1	0.10
	MLF1	2	11	0.14
	ANKUB1	1	2	0.15
	IL12A	1	2	0.15
	OTOL1	2	13	0.18
	C3orf80	1	3	0.19
	WWTR1	1	4	0.23
	SLITRK3	2	17	0.26
	SMC4	1	6	0.31
	B3GALNT1	1	9	0.41
	MFSD1	1	9	0.41
	SI	2	35	0.58
	MED12L	2	36	0.59
	MECOM	1	24	0.74
	IGSF10	1	34	0.84

Table S8.
All top genes showing a significant excess of *novel functional* variants in cases vs. controls before adjusting for false positive calls. The list includes genes located on chromosome X.

Gene	Number of variants in cases (n=71)	Number of variants in controls (n=1,926)	p value
TMPRSS13	6	7	2.79×10^{-6}
CH25H	3	2	4.3×10^{-4}
ARMCX2	3	3	8.29×10^{-4}
HSPB7	2	0	1.26×10^{-3}
KLRC1	2	0	1.26×10^{-3}
ZNF645	2	0	1.26×10^{-3}
MOAP1	3	4	1.41×10^{-3}
RBM25	3	4	1.41×10^{-3}
TTC39A	4	13	2.62×10^{-3}
ZNF785	4	13	2.62×10^{-3}
FZD9	3	6	3.21×10^{-3}
HEMK1	3	6	3.21×10^{-3}
PCGF3	4	14	3.28×10^{-3}
KIF1B	5	24	3.3×10^{-3}
ANP32E	2	1	3.7×10^{-3}
CABP5	2	1	3.7×10^{-3}
CELA2B	2	1	3.7×10^{-3}
INSIG2	2	1	3.7×10^{-3}
KCTD7	2	1	3.7×10^{-3}
MRO	2	1	3.7×10^{-3}
NR2E1	2	1	3.7×10^{-3}
NXT2	2	1	3.7×10^{-3}
PABPC1	2	1	3.7×10^{-3}
PODXL	2	1	3.7×10^{-3}
PUS3	2	1	3.7×10^{-3}
TXNDC15	2	1	3.7×10^{-3}
WDR89	2	1	3.7×10^{-3}
ZNF720	2	1	3.7×10^{-3}

REFERENCE

1. S. A. Miller, D. D. Dykes, H. F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215 (Feb 11, 1988).
2. S. W. Fouchier, J. C. Defesche, M. W. Umans-Eckenhausen, J. P. Kastelein, The molecular basis of familial hypercholesterolemia in The Netherlands. *Human genetics* **109**, 602 (Dec, 2001).
3. A. J. Hooper *et al.*, Genetic analysis of familial hypercholesterolaemia in Western Australia. *Atherosclerosis* **224**, 430 (Oct, 2012).
4. S. Bertolini *et al.*, Spectrum of mutations and phenotypic expression in patients with autosomal dominant hypercholesterolemia identified in Italy. *Atherosclerosis* **227**, 342 (Apr, 2013).
5. C. A. Graham *et al.*, Mutation screening and genotype:phenotype correlation in familial hypercholesterolaemia. *Atherosclerosis* **147**, 309 (Dec, 1999).
6. M. Futema *et al.*, Analysis of the frequency and spectrum of mutations recognised to cause familial hypercholesterolaemia in routine clinical practice in a UK specialist hospital lipid clinic. *Atherosclerosis*, (Apr, 2013).
7. A. Taylor *et al.*, Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. *Clinical genetics* **77**, 572 (Jun, 2010).
8. R. A. Whittall *et al.*, Development of a high-resolution melting method for mutation detection in familial hypercholesterolaemia patients. *Ann Clin Biochem* **47**, 44 (Jan, 2010).
9. C. D. Mamotte, F. M. van Bockxmeer, A robust strategy for screening and confirmation of familial defective apolipoprotein B-100. *Clinical chemistry* **39**, 118 (Jan, 1993).
10. S. E. Humphries *et al.*, Genetic causes of familial hypercholesterolaemia in patients in the UK: relation to plasma lipid levels and coronary heart disease risk. *J Med Genet* **43**, 943 (Dec, 2006).
11. A. Taylor *et al.*, Multiplex ligation-dependent probe amplification analysis to screen for deletions and duplications of the LDLR gene in patients with familial hypercholesterolaemia. *Clin Genet* **76**, 69 (Jul, 2009).
12. T. M. Teslovich *et al.*, Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707 (Aug 5, 2010).
13. J. Boren *et al.*, Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *The Journal of clinical investigation* **101**, 1084 (Mar 1, 1998).
14. X. Wang *et al.*, Genome-wide linkage scan of a pedigree with familial hypercholesterolemia suggests susceptibility loci on chromosomes 3q25-26 and 21q22. *PloS one* **6**, e24838 (2011).
15. A. Marques-Pinheiro *et al.*, A fourth locus for autosomal dominant hypercholesterolemia maps at 16q22.1. *European journal of human genetics : EJHG* **18**, 1236 (Nov, 2010).
16. A. Cenarro *et al.*, A presumptive new locus for autosomal dominant hypercholesterolemia mapping to 8q24.22. *Clin Genet* **79**, 475 (May, 2011).