Genetics inMedicine | ORIGINAL RESEARCH ARTICLE

© American College of Medical Genetics and Genomics

Open

The use of next-generation sequencing in clinical diagnosis of familial hypercholesterolemia

Jana Vandrovcova, PhD[1], Ellen R.A. Thomas, MRCP[1], Santosh S Atanur, MSc[2], Penny J. Norsworthy, BSc[1], Clare Neuwirth, SRN, RN[1,4], Yvonne Tan, RN, BA[1,4], Dalia Kasperaviciute, PhD[3], Jennifer Biggs, BSc, MRes[1], Laurence Game, PhD[1], Michael Mueller, PhD[3], Anne K. Soutar, PhD[1] and Timothy J. Aitman, FRCP, DPhil[1,3]

Purpose: Familial hypercholesterolemia is a common Mendelian disorder associated with early-onset coronary heart disease that can be treated by cholesterol-lowering drugs. The majority of cases in the United Kingdom are currently without a molecular diagnosis, which is partly due to the cost and time associated with standard screening techniques. The main purpose of this study was to test the sensitivity and specificity of two next-generation sequencing protocols for genetic diagnosis of familial hypercholesterolemia.

Methods: Libraries were prepared for next-generation sequencing by two target enrichment protocols; one using the SureSelect Target Enrichment System and the other using the PCR-based Access Array platform.

Results: In the validation cohort, both protocols showed 100% specificity, whereas the sensitivity for short variant detection was 100% for the SureSelect Target Enrichment and 98% for the Access Array protocol. Large deletions/duplications were only detected using the SureSelect Target Enrichment protocol. In the prospective cohort, the mutation detection rate using the Access Array was highest in patients with clinically definite familial hypercholesterolemia (67%), followed by patients with possible familial hypercholesterolemia (26%).

Conclusion: We have shown the potential of target enrichment methods combined with next-generation sequencing for molecular diagnosis of familial hypercholesterolemia. Adopting these assays for patients with suspected familial hypercholesterolemia could improve cost-effectiveness and increase the overall number of patients with a molecular diagnosis.

Genet Med advance online publication 16 May 2013

Key Words: familial hypercholesterolemia; *LDLR*; molecular diagnostic testing; next-generation sequencing; target enrichment

Familial hypercholesterolemia (FH, OMIM no. 143890) is a common autosomal dominant condition with a prevalence of 1 in 500. Patients with FH have raised serum cholesterol levels and increased arterial deposition of low-density lipoprotein (LDL) cholesterol, leading to premature coronary heart disease. Despite the fact that early-onset coronary heart disease can be prevented by cholesterol-lowering drugs such as statins, less than a quarter of FH patients have currently been identified in the United Kingdom.[1] Several diagnostic criteria have been developed to identify individuals with FH, including the Dutch Lipid Clinic[2] criteria and the MedPed[3] criteria. In the United Kingdom, the clinical diagnosis of FH is based on the Simon Broome criteria of cholesterol levels, presence of tendon xanthomata, family history, and genetic testing.[4] The UK National Institute for Health and Clinical Excellence guidelines recommend initial mutation screening of index cases fulfilling Simon Broome criteria followed by cascade screening in at least first- and second-degree relatives.[5]

Most cases of FH are caused by mutations in the *LDLR* gene that encodes the LDL receptor protein, which binds LDL particles at the hepatic cell membrane and internalizes them for processing and excretion. FH-causing mutations in *LDLR* are found throughout the gene and include missense, truncating, and splice site mutations; small insertion/deletion mutations; and large insertions/deletions that can encompass multiple exons. Some mutations have been found in many unrelated individuals with FH, whereas others are found rarely.[6] Mutations in two other genes, *PCSK9* and *APOB*, can also cause the FH phenotype but in <20% of cases.[7] Rare autosomal-recessive hypercholesterolemia is caused by mutations in the *LDLRAP1* gene.[8]

Conventional DNA testing of FH disease–causing genes is mostly based on direct capillary sequencing, with multiplex ligation-dependent probe amplification (MLPA) used for the detection of large insertions or deletions.[9] These molecular techniques are sensitive and specific, but because of the cost and time involved, they are impractical for screening large numbers of patients. To overcome some of these limitations, assays such as the Amplification Refractory Mutation System (Elucigene FH20; Tepnel Molecular Diagnostics, Abingdon, UK) or array-based sequencing methods (LIPOchip; Progenika Biopharma, Derio, Spain) have been developed. These assays are tailored to

The first three authors contributed equally to this work.

[1]MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, London, UK; [2]National Heart and Lung Institute, Imperial College London, London, UK; [3]Institute of Clinical Sciences, Imperial College London, London, UK; [4]Imperial College Healthcare NHS Trust, London, UK. Correspondence: Timothy J. Aitman (t.aitman@imperial.ac.uk)

Submitted 7 December 2012; accepted 19 March 2013; advance online publication 16 May 2013. doi:10.1038/gim.2013.55

specific mutations and populations and therefore do not detect less common mutations or any novel mutations.[10,11]

During the past few years, high-throughput next-generation sequencing (NGS)–based methods have become available for DNA analysis. They have not only proven successful in new disease gene identification, but following the availability of economic "benchtop" sequencers, they have become more easily applicable to targeted diagnostic sequencing. The combination of high-throughput and relatively small DNA target selection allows for many genes and samples to be processed simultaneously, making it an attractive solution for the processing of large sample numbers in a diagnostic laboratory. Moreover, the flexible design of these assays enables molecular screening to be extended to other relevant genes and polymorphisms that are not covered by routine tests.

In this study, two target enrichment protocols, the hybridization-based SureSelect Target Enrichment System and the PCR-based Access Array System platform, followed by NGS sequencing, have been validated for the detection of FH-causing mutations. We show excellent performance for both approaches and discuss their potential for clinical screening programs and discovery of novel FH-causing genes.

MATERIALS AND METHODS

Patients

We studied two groups of patients, a validation group including patients with known molecular diagnosis and a prospective cohort of previously unscreened FH patients. In the validation group, DNA from 104 patients from the Hammersmith Hospital Lipid Clinic were analyzed. All 104 samples had previously undergone complete or partial screening of *LDLR*-coding regions and exon/intron boundaries, p.Asp374Tyr in *PCSK9*, and p.Arg3527Gln in *APOB* exon 26 using Sanger sequencing. In addition, MLPA was used to detect large deletions and duplications in a subset (>80%) of the patients in the validation cohort, as described in the study by Tosi et al.[9] Forty-five samples were known to carry heterozygous FH-causing point mutations or small insertions/deletions (<20 bp) in *LDLR*, *PCSK9*, or *APOB*, and six had large heterozygous insertions/deletions in *LDLR*. One sample was homozygous for *LDLR* p.Gln384Pro, and one sample was a compound heterozygote with one missense mutation and one large deletion in *LDLR*. The remaining 51 samples in the validation cohort had negative molecular diagnoses. Of the 104 samples, 29 were processed using the SureSelect Target Enrichment System alone, 42 using the Access Array System microfluidic platform alone, and the remaining 33 using both platforms. In the prospective cohort, 84 consecutive patients referred for molecular testing by the Hammersmith Hospital Lipid Clinic over a period of 1 year were studied by the PCR-based Access Array System microfluidic platform. Fifteen of the 84 samples were also analyzed using the SureSelect Target Enrichment System. One individual was referred with suspected homozygous FH. All the remaining 83 were suspected to be heterozygous. Six patients had a diagnosis of definite FH based on Simon Broome criteria. The median highest total cholesterol in this group was 9.7 mmol/l (minimum: 7.9 mmol/l; maximum: 15.9 mmol/l). Sixty-five individuals had a diagnosis of possible FH, with a median highest total cholesterol level of 8.7 mmol/l (minimum: 5.2 mmol/l; maximum: 13 mmol/l), and 13 patients did not fulfill Simon Broome criteria. The median highest total cholesterol in this group was 8.3 mmol/l (minimum: 6.3 mmol/l; maximum: 9.6 mmol/l). The mean age at measurement was 43 years (minimum: 2 years; maximum: 68 years).

The study was approved under ethics committee references REC2002/6451 and REC11/LO/0883. All patients provided informed consent.

DNA extraction

DNA was extracted from blood using a standard phenol–chloroform protocol or the Maxwell 16 system (Promega, Madison, WI) or from saliva samples using the Oragene (Genotek, Ottawa, Ontario, Canada) protocol. Both protocols followed the manufacturer's recommendations.

RNA extraction and reverse-transcription PCR

Blood for RNA extraction was collected in Tempus blood RNA tubes (Applied Biosystems, Foster City, CA) and extracted using the Paxgene blood RNA kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. Reverse transcription was performed using the iScript cDNA synthesis kit (BioRad, Hercules, CA). Primers (forward: TCGAGTTCCACTGCCTAAGTG and reverse: GTTGTTGTCCAAGCATTCGTT) were designed to amplify exons 4–7 of *LDLR*.

Custom SureSelect Target Enrichment System

Four genes with mutations known to cause FH (*LDLR*, *APOB*, *PCSK9*, and *LDLRAP1)*, a myopathy-associated variant in *SLCO1B1*, and 13 genes (*APOE*, *HMGCR*, *HNRNPD*, *INSIG1*, *KHSRP*, *NPC1L1*, *PTBP1*, *SREBF1*, *SREBF2*, *MESDC2*, *SCAP*, *INSIG2*, and *CYP7A1*) functioning within cholesterol-processing pathways were included in the SureSelect Target Enrichment System design (Agilent, Santa Clara, CA). The design contained 120-mer baits spanning the entire nonrepetitive sequence of the selected genes, including exons and introns, 2 kb of upstream sequence (10 kb for the four known FH genes), and 1 kb of downstream sequence.

The genomic coordinates of the 18 targeted genes were determined using the March 2006 build (NCBI36/hg18) of the human genome in the Ensembl genome browser.[12] The density of bait tiling was fivefold, and the baits were allowed to overlap into repeat regions by 30 bp. The total targeted DNA length was 399 kb. All libraries were generated from sheared DNA (Covaris, Woburn, MA) with an average insert size of 200 bp following the SureSelect Target Enrichment System XT protocol for Illumina multiplexed sequencing version 1.2 (Illumina, San Diego, CA). After dilution to 2 nmol/l, up to 30 libraries were pooled and sequenced on one lane on the HiSeq2000 platform (Illumina) to generate 2 × 100 bp paired-end reads.

PCR-based Access Array System

The design included 43 amplicons covering all exons of *LDLR*, with the majority of the coding sequence covered by more than one overlapping fragment. Amplicons were also designed to cover exons 2, 4, 7, and 9 of *PCSK9*, containing the most common gain-of-function pathogenic mutations; *APOB* (one amplicon covering the most common familial defective apolipoprotein B-100 mutation, p.Arg3527Gln); *APOE* (one amplicon covering the *APOE* E2 variant site, rs7412); and *SLCO1B1* (one amplicon covering rs4149056, the myopathy-associated variant). The average amplicon length was 184 bp, with 57% GC content. Primer sequences are shown in **Supplementary Table S1** online. Samples were processed using the Access Array System (Fluidigm, South San Francisco, CA) according to the manufacturer's 4-Primer amplicon tagging protocol generating paired-end libraries for Illumina sequencing. Purified pooled products of 47 barcoded samples and one negative control were sequenced in one run using the MiSeq sequencer (Illumina). One amplicon, a GC-rich (75%) amplicon in *APOE* targeting rs7412, failed to amplify and was excluded from subsequent data analysis.

Data analysis

Sequences were mapped to the GRCh37/hg19 human reference sequence using Burrows Wheeler Aligner v0.6.1.[13] PCR duplicate reads were removed from SureSelect Target Enrichment System data (Picard tools v1.35; http://picard.sourceforge.net). Sequence reads from both data sets were further processed, and variants were called using GATK v1.0.6[14] with hard filtering options. Variant annotation was carried out with Ensembl's Variant Effect Predictor tool[12] and was based on the transcripts ENST00000558518 (*LDLR*), ENST00000233242 (*APOB*), and ENST00000302118 (*PCSK9*). The annotation included Sorting Intolerant From Tolerant, Condel, and PolyPhen. Conservation scores (Genomic Evolutionary Rate Profiling scores) were obtained from Ensembl[12] version 68. To identify potentially pathogenic mutations, all nonsynonymous, splice site, frameshift, and truncating mutations were examined and compared with an FH locus–specific database[6,15] as a guide to interpretation of variant pathogenicity. Synonymous and intronic variants located outside exon/intron boundaries as well as single-nucleotide polymorphisms with minor allele frequency >1% in the International HapMap Project[16] or the 1000 Genomes Project[17] were excluded from further analysis. All single-nucleotide variants and short insertions/deletions that were potentially disease causing were verified by conventional Sanger sequencing. Novel variants were followed up by segregation analysis wherever possible.

Copy-number variant analysis

Copy-number variant (CNV) analysis from NGS data was performed for the samples sequenced using targeted capture. A read depth–based method,[18] as implemented in R package ExomeDepth, was used to identify deletions and duplications spanning at least one exon. Each sequencing batch of samples was processed separately to increase the quality of a reference set for each sample and therefore to maximize the power to detect CNVs. Read depth was assessed for each exon in the target region, and the ratio of expected and observed read count was obtained, as well as a Bayes factor for the CNV calls, as implemented in the ExomeDepth method.

As an independent method of CNV analysis, MLPA was performed using the kit *LDLR*-P062 (MRC-Holland, Amsterdam, The Netherlands) following the manufacturer's protocol. The novel exon 16 deletion was confirmed by PCR. The previously described large deletions and duplications are detailed in the study by Tosi et al.[9]

Statistical analysis

The sensitivity of an assay was defined as the percentage of pathogenic mutations correctly identified with respect to previous or new Sanger sequencing and MLPA. The specificity is defined as the percentage of mutation-negative samples correctly identified as negative with respect to previous or new Sanger sequencing and MLPA.

Myopathy-associated variant in *SLCO1B1*

Genotypes of the *SLCO1B1* myopathy–associated variant rs4149056 were scored from the Access Array System microfluidic platform and SureSelect Targeted Enrichment System data in all patients, and any history of adverse effects was obtained by review of medical records. Side effects were defined biochemically (transaminase or creatine kinase levels more than three times the upper limit of the normal range) or symptomatically for myalgia and other side effects that coincided temporally with statin treatment. Tests for deviation from Hardy–Weinberg equilibrium and association tests were performed using the DeFinetti software (http://ihg.gsf.de/cgi-bin/hw/hwa1.pl).

RESULTS

Validation study

Custom SureSelect Target Enrichment System. To validate the assay, DNA was analyzed from 62 previously screened individuals who either had a confirmed molecular diagnosis of FH (n = 28) or were mutation-negative (n = 34). An average of 330 Gb (191–381 Gb) of sequence was obtained per sample, with an average coverage of ×827. Overall, 64% of total mapped reads aligned within the target region, and 99.8 and 98.8% of nucleotides were covered at ×4 and ×25, respectively. The insufficiently covered regions are consistent among runs and were found mostly outside coding regions. Hybridization-based capture is known to target GC-rich regions poorly; however, all regions containing known FH-causing mutations were covered sufficiently (more than ×25) for confident variant calling. The initial analysis of the sequencing results was carried out by a researcher blinded to the gene and mutation details for each sample. All 20 heterozygous and 1 homozygous short pathogenic mutations, including point mutations and insertions/deletions of <15 bp, were detected (**Table 1**). In addition, one

Table 1 Mutations identified in the validation study

ID	Detected by SureSelect	Detected by Access Array	Gene	cDNA	Protein
990	ND	Yes	*APOB*	c.10580G>A	p.Arg3527Gln
747	Yes	Yes	*APOB*	c.10580G>A	p.Arg3527Gln
1086	ND	Yes	*APOB*	c.10580G>A	p.Arg3527Gln
667	ND	Yes	*PCSK9*	c.385G>A	p.Asp129Asn
735	ND	Yes	*LDLR*	c.1618G>A	p.Ala540Thr
604	Yes	Yes	*LDLR*	c.1880C>A	p.Ala627Asp
660	ND	Yes	*LDLR*	c.1048C>T	p.Arg350X
728	Yes	Yes	*LDLR*	c.1048C>T	p.Arg350X
631	Yes	Yes	*LDLR*	c.2231G>A	p.Arg744Gln
384	Yes	Yes	*LDLR*	c.1217G>A	p.ArgR406Gln
895	ND	Yes	*LDLR*	c.1691A>G	p.Asn564Ser
896	ND	Yes	*LDLR*	c.1691A>G	p.Asn564Ser
766	Yes	Yes	*LDLR*	c.2061dupC	p.Asn688Glnfs*29
430	Yes	Yes	*LDLR*	c.661G>A	p.Asp221Asn
570[a]	Yes	Yes	*LDLR*	c.681C>G	p.Asp227Glu
862	ND	Yes	*LDLR*	c.681C>G	p.Asp227Glu
1095	ND	Yes	*LDLR*	c.1444G>A	p.Asp482Asn
438	Yes	Yes	*LDLR*	c.1444G>C	p.Asp482His
1031	ND	Yes	*LDLR*	c.401G>A	p.Cys134Tyr
714	Yes	Yes	*LDLR*	c.939C>A	p.Cys313X
1067	ND	Yes	*LDLR*	c.1130G>A	p.Cys377Tyr
1068	ND	Yes	*LDLR*	c.1130G>A	p.Cys377Tyr
819	Yes	Yes	*LDLR*	c.2029T>C	p.Cys677Arg
818	ND	Yes	*LDLR*	c.2043C>A	p.Cys681X
450	Yes	Yes	*LDLR*	c.2043C>A	p.Cys681X
814	Yes	Yes	*LDLR*	c.2043C>A	p.Cys681X
415	Yes	Yes	*LDLR*	c.[1151A>C]; [1151A>C]	p.Gln384Pro
813	ND	Yes	*LDLR*	c.303delG	p.Glu101Aspfs*105
996	ND	Yes	*LDLR*	c.301G>A	p.Glu101Lys
771	ND	Yes	*LDLR*	c.301G>A	p.Glu101Lys
820	ND	Yes	*LDLR*	c.420G>C	p.Glu140Asp
770	Yes	Yes	*LDLR*	c.654_656delTGG	p.Glu219del
586	Yes	Yes	*LDLR*	c.682G>T	p.Glu228X
1071	ND	Yes	*LDLR*	c.1646G>A	p.Gly549Asp
853	Yes	Yes	*LDLR*	c.1436T>C	p.Leu479Pro
857	Yes	No	*LDLR*	c.2184_2194delGCTAAAGGTCA	p.Lys730Hisfs*48
743	ND	Yes	*LDLR*	c.1A>T	p.Met?
663	Yes	Yes	*LDLR*	c.1118_1121dupGTGG	p.Tyr375Trpfs*7
764	Yes	Yes	*LDLR*	c.1118_1121dupGTGG	p.Tyr375Trpfs*7
1102	ND	Yes	*LDLR*	c.1285G>A	p.Val429Met
1104	ND	Yes	*LDLR*	c.1285G>A	p.Val429Met
666	Yes	Yes	*LDLR*	c.1285G>A	p.Val429Met
852	ND	Yes	*LDLR*	c.1586+5G>C	
1011	ND	Yes	*LDLR*	c.1706-1G>T	
1094	ND	Yes	*LDLR*	c.1845+1G>A	
721	ND	Yes	*LDLR*	c.1845+11C>G	
566	Yes	Yes	*LDLR*	c.313+1G>A	
556	Yes; Yes	Yes; No	*LDLR*	c.[266G>A]; [2312-?_2547+?del]	p.Cys89Tyr; Exons 16–17 deletion
683	Yes	No	*LDLR*	c.-187-?_67+?del	Promoter and exon 1 deletion
836[a]	Yes	No	*LDLR*	c.-187-?_67+?del	Promoter and exon 1 deletion
547	Yes	No	*LDLR*	c.-187-?_67+?del	Promoter and exon 1 deletion
662[a]	**Yes**	**No**	***LDLR***	**c.2312-?_2389+?del**	**Exon 16 deletion**
436	Yes	No	*LDLR*	c.695-?_817+?del	Exon 5 deletion
564	Yes	No	*LDLR*	c.68-?_2583+?del	Exons 2–18 deletion
565	Yes	No	*LDLR*	c.68-?_940+?del	Exons 2–6 deletion
712	Yes	No	*LDLR*	c.1187-?_2140+?dup	Exons 9–14 duplication

Boldface indicates novel variants that were not previously identified.

cDNA, complementary DNA; ND, not done; ?, exact breakpoint boundaries not known.

[a]New diagnosis in samples that had previously undergone partial screening only.

compound heterozygote and six large insertions/deletions in the *LDLR* gene were detected, resulting in 100% sensitivity for this assay. No false positives were detected, and specificity for this assay was also 100%. A further three mutations were identified in *LDLR* in patients without a previous molecular diagnosis. One was a known single-nucleotide substitution, p.Asp227Glu, and two were large deletions (deletion of exon 16 and deletion of promoter/exon 1). In addition, a novel variant was identified in the *LDLR* promoter (c. −227G>T; Genomic Evolutionary Rate Profiling score: 3.29; located in the highly conserved footprint 1 site), and two rare variants were identified in *APOB* in patients without a previous molecular diagnosis. The two variants, p.Pro877Leu (rs12714097) and p.Asp2213del (rs72653087), have not been described previously in FH patients and are not reported in either the 1000 Genomes Project[17] or the International HapMap Project[16] databases. No other rare missense variants were detected in the known FH genes in individuals without a molecular diagnosis.

PCR-based Access Array System. DNA samples from 53 previously characterized patients including 40 with point mutations (39 heterozygotes and 1 homozygote), 6 with insertions/deletions (all heterozygotes), 6 with large deletions or duplications (all heterozygotes), and 1 compound heterozygote with 1 missense mutation and 1 large deletion in *LDLR* (**Table 1**) were amplified together with samples from 22 patients who had been previously screened as negative for mutations. The median coverage per sample was ×572 (minimum: 461; maximum: 625). All amplicons except *APOE* (see Materials and Methods) amplified with a mean coverage of ×506. The coverage for individual amplicons is listed in **Supplementary Table S2** online. Overall, 90% of bases were covered more than 25-fold. *LDLR* had a mean coverage of ×656 and >98% of bases were above ×25. All single-nucleotide changes in the coding sequence and intron/exon boundaries were correctly identified (**Table 1**), including one mutation in *PCSK9* exon 2 and one in *APOB*, despite the lower sequencing coverage of these amplicons with 13 and 15 reads, respectively (the percentage of mutated alleles was 54 and 53%). One variant was not detected. This 11 bp deletion (p.Lys730Hisfs*48) was not present in any aligned reads because the deletion overlapped with a forward primer, which prevented amplification of the mutated allele. The sensitivity for short variant detection was 98% (47 of 48), and specificity was 100%. Similar to the SureSelect Target Enrichment System results, one mutation, p.Asp227Glu, was identified in a sample that previously had no molecular diagnosis. No pathogenic mutations were found in the remaining samples that were negative on previous screening. Large deletions could not be detected with the PCR-based Access Array System because no reduction in coverage was observed within deleted regions. The overall sensitivity of this assay as compared with SureSelect Target Enrichment System was therefore 82% (47 of 57).

Prospective cohort

To test the feasibility of an NGS-based mutation screening in a clinical setting, a consecutive cohort of 84 unrelated patients referred for genetic screening over a single year was analyzed. Sixty-nine of these patients were screened using the Access Array System assay alone, and 15 were processed using both protocols. Sequencing and coverage metrics are provided in **Supplementary Table S2** online. In total, 23 variants that have previously been reported as pathogenic in FH patients were identified in 22 individuals (**Table 2**), including 13 single-nucleotide changes, 4 variants predicted to affect splicing, and 6 short insertions/deletions. In addition, two novel variants (p.Ala521Thr and p.Asn316Lysfs*54) were identified that had not been previously reported (**Table 2**). There was a good correlation between the results obtained from the Access Array System assay and SureSelect Target Enrichment System, with five of six mutations concordant between the data sets. The discordant variant, p.Lys730Hisfs*48, is the same one that was not detected using the Access Array System protocol in the validation cohort. Despite the presence of the same variant, these patients in the validation and prospective cohort were not known to be related. No additional mutations were identified using MLPA.

Of the two previously unreported variants (**Table 2**), the first was a missense change, p.Ala521Thr. This variant is conserved (Genomic Evolutionary Rate Profiling score: 2.66), is predicted to be damaging (PolyPhen score: 0.67, Sorting Intolerant From Tolerant score: 0.01, and Condel score: 0.643), and segregates with affected status in the family (**Supplementary Figure S1a** online). The second variant was a single-nucleotide deletion in exon 7 of *LDLR*, p.Asn316Lysfs*54. This frameshift mutation creates a premature stop codon and also segregates with hypercholesterolemia in this family (**Supplementary Figure S1b** online).

Two (c.817+1G>A and c.941-4G>A) of the four splice site variants identified (**Table 2**) were previously detected in FH patients,[6] but their pathogenicity had not been experimentally validated before. An RNA sample was available for one of them (patient ID: 35) and reverse-transcriptase PCR in this patient showed that the c.817+1G>A variant disrupted correct splicing (data not shown). Alternative splice products were generated by partial intron retention that led to truncation of the protein, by exon skipping that caused the loss of exon 5 sequence, and by a predicted frameshift of the remaining protein. Of the remaining variants previously identified in FH patients, one missense variant, p.Val827Ile, was conservative, without a published proof of pathogenicity and therefore was considered to be of unknown significance. Two children of this index case were available for segregation analysis, and the results showed that an unaffected daughter (total cholesterol: 4.6 mmol/l; age: 27 years) of the proband had inherited the p.Val827Ile variant.

Mutation detection rates

The prospective cohort included six patients with definite FH as defined by the Simon Broome criteria, 65 with possible FH, and

Table 2 Previously reported and novel variants identified in known disease-causing genes in the prospective cohort

DNA ID	Detected by SureSelect	Detected by Access Array	Gene	cDNA	Protein	Simon Broome criteria	Age at measurement	Highest total cholesterol (mmol/l)
1	Yes	Yes	*LDLR*	c.938G>A	p.Cys313Tyr	Definite	48	15.9
2	**ND**	**Yes**	***LDLR***	**c.948delC**	**p.Asn316Lysfs*54**	**Definite**	**12**	**11.2**
3	ND	Yes	*LDLR*	c.190+4A>T		Definite	44	9.4
5	ND	Yes	*LDLR*	c.2479G>A	p.Val827Ile	Definite	58	7.9
6	ND	Yes	*LDLR*	c.2054C>T/c.736delG	p.Pro685Leu/ p.Gly246Glufs*19	Definite	2	9.7
9	Yes	Yes	*LDLR*	c.1476_1477delCT	p.Ser493Cysfs*42	Possible	20	9.2
12	Yes	No	*LDLR*	c.2187_2197delAAAGGTCAGCT	p.Lys730Hisfs*48	Possible	25	13.0
14	ND	Yes	*LDLR*	c.204C>A	p.Cys68*	Possible	5	8.6
15	ND	Yes	*LDLR*	c.1681C>T	p.Gln561*	Possible	3	5.2
18	ND	Yes	*APOB*	c.10579C>T	p.Arg3527Trp	Possible	25	9.2
21	ND	Yes	*LDLR*	c.1476_1477delCT	p.Ser493Cysfs*42	Possible	29	8.1
22	Yes	Yes	*LDLR*	c.283T>C	p.Cys95Arg	Possible	38	9.0
33	**Yes**	**Yes**	***LDLR***	**c.1561G>A**	**p.Ala521Thr**	**Possible**	**33**	**10.0**
35	Yes	Yes	*LDLR*	c.817+1G>A		Possible	13	10.5
40	ND	Yes	*LDLR*	c.670_675delinsTTT	p.Asp224_ Lys225delinsPhe	Possible	41	11.6
45	ND	Yes	*LDLR*	c.1730G>A	p.Trp577*	Possible	56	10.3
50	ND	Yes	*LDLR*	c.906C>G	p.Cys302Trp	Possible	NA	NA
52	ND	Yes	*APOB*	c.10580G>A	p.Arg3527Gln	Possible	23	9.2
56	ND	Yes	*LDLR*	c.266G>A	p.Cys89Tyr	Possible	NA	NA
58	ND	Yes	*LDLR*	c.654_656delTGG	p.Gly219del	Possible	40	9.6
61	ND	Yes	*LDLR*	c.941-4G>A		Possible	43	NA
63	ND	Yes	*LDLR*	c.2054C>T	p.Pro685Leu	Possible	11	9.0
68	ND	Yes	*LDLR*	c.1130G>A	p.Cys377Tyr	Possible	12	7.0
72	ND	Yes	*LDLR*	c.313+1G>A		Not fulfilling criteria	47	9.6

Boldface indicates novel, not previously identified variants. These variants are expected to cause and were shown to segregate with the disease. Underlining indicates variants of unknown significance. These rare variants were previously reported in patients, but their pathogenicity has not been established.

cDNA, complementary DNA; NA, not available; ND, not done.

13 hypercholesterolemic patients not fulfilling Simon Broome criteria for FH. The highest detection rate of clearly pathogenic mutations was in the group of patients with a definite FH diagnosis (4 of 6, 67%) followed by the group with a diagnosis of possible FH (17 of 65, 26%). One mutation (1 of 13, 8%) was identified among the 13 hypercholesterolemic patients who did not fulfill FH diagnostic criteria.

Variants in genes involved in cholesterol metabolism

In addition to the three known FH-causing genes and *SLCO1B1*, the SureSelect Target Enrichment System assay included 13 genes that reside on cholesterol metabolism pathways. Rare variants identified in these genes, particularly in hypercholesterolemic individuals who had been screened negative for mutations in *LDLR*, *PCSK9*, and *APOB*, are potentially responsible for patients' raised cholesterol levels. Rare variants (minor allele frequency < 0.01) identified in such individuals are listed in **Supplementary Table S3** online. In patients with no previously known molecular diagnosis, seven rare nonsynonymous variants were found, of which six were not predicted to be functionally significant by Sorting Intolerant From Tolerant and PolyPhen. The single variant that was most likely to be of functional significance and therefore potentially pathogenic, was p.Val809Met in *SREBF1*. However, this was excluded from further analysis because it did not segregate with the phenotype in the family (data not shown).

SLCO1B1 myopathy-associated variant

Data on exposure to statins were available for 149 patients. Twenty-seven individuals were heterozygous for *SLCO1B1* rs4149056, and two were homozygous. One patient (a heterozygote) had suffered severe biopsy-proven hepatitis while taking

Table 3 The frequency of side effects of statin drugs in FH patients stratified by rs4149056

	Total	Heterozygous	Homozygous	Wild-type
Creatine kinase >3 times the upper limit of normal range	6	2	0	4
Creatine kinase ≤3 times the upper limit of normal range	5	1	1	3
Myalgia	12	3	1	8
Transaminase levels >3 times the upper limit of normal range	3	1	0	2
Transaminase levels ≤3 times the upper limit of normal range	9	2	0	7
Transaminase levels unknown	3	1	0	2
Other patient-reported side effects	11	4	0	7
Total	49	14	2	33

FH, familial hypercholesterolemia.

statin drugs. A further 48 patients had suffered less-severe side effects, either patient-reported clinical effects or asymptomatic biochemical disturbances (Table 3). On association testing for rs4149056 in all individuals who had suffered side effects (n = 49) versus those who had experienced no adverse effects on statins (n = 65), the odds ratio was 3.95 (confidence interval: 1.58–9.89; P = 0.002).

DISCUSSION

We investigated the sensitivity and specificity of two target enrichment protocols, combined with NGS, for the detection of disease-causing mutations in patients with proven or suspected FH. Using the SureSelect Target Enrichment System and the HiSeq 2000 system, 98.8% of targeted regions were covered at more than ×25 as compared with the PCR-based Access Array System followed by MiSeq sequencing, for which on average 10% of nucleotides failed to reach this coverage. Regions with low coverage were found outside the coding sequence of *LDLR* and mutation hotspots in *APOB* and *PCSK9* and therefore did not affect the overall success of mutation detection. Both techniques showed high sensitivity and specificity for the detection of single base substitutions and short insertions/deletions. In the validation part of the study, 100% of previously detected mutations were correctly identified using the SureSelect/HiSeq protocol. In comparison, the Access Array System/MiSeq approach led to correct identification of 98% of all variants detected previously by Sanger sequencing, although large insertions/deletions could not be detected (see below). The single variant that was not detected by the PCR-based Access Array System protocol was an 11 bp deletion (p.Lys730Hisfs*48) that overlapped a primer site. However, the recent availability of longer sequencing reads, currently up to 250 bp when using MiSeq, will be anticipated in most cases to eliminate the need to have primer sites within coding regions or allow design modification to include longer overlaps between amplicons.

Three mutations in the *LDLR* coding sequence (one single-nucleotide variant, p.Asp227Glu, and two large deletions) and one *LDLR* promoter variant (c.–227G>T) were identified in patients who were previously classified as negative for mutations. Inspection of the previous laboratory database indicated that the p.Asp227Glu variant and the two large deletions had not been detected because of incomplete capillary sequencing in the case of the p.Asp227Glu and promoter variants and absence of MLPA data in the case of the two deletions.

A small proportion of FH cases are caused by large deletions or duplications of *LDLR*.[19] The current standard screening is based on MLPA, a technique that is highly reliable but that is a costly and time-consuming addition to sequencing. The detection of large variants from NGS data has been shown previously, but its use in FH diagnostics has not yet been investigated.[20,21] Here, the combination of SureSelect Target Enrichment System/HiSeq and data analysis using ExomeDepth software[18] led to correct identification of all eight large deletions and one large duplication. This shows the potential of using hybrid capture for the detection of both short and large sequence variants in FH. The read depth from the PCR-based Access Array System assay was not used for this analysis method because read depth in this study did not correlate with exon deletions, although it has recently been shown that amplicon multiplex PCR can be optimized for CNV detection.[22]

When testing a new molecular workflow for routine diagnostics, in addition to sensitivity and specificity, it is important to consider the cost and time efficiency of the protocol. The SureSelect Target Enrichment System protocol allows for comprehensive coverage of targeted regions (current custom design up to 24 Mb), whereas the Access Array System protocol is limited to 480 amplicons with a maximum length of 400 bp when sequenced using the latest MiSeq system. Sequence capture also allows the detection of all types of variants, including large deletions and duplications, which was not possible in this study using the PCR-based Access Array System. On the other hand, the Access Array System is considerably cheaper to run,[20,23] with reagent costs approximately 10-fold lower than those for the SureSelect Target Enrichment System protocol,[23] and, in addition, the library preparation turnaround time is shorter. In our hands, 96 samples can be processed within a day using the Access Array System protocol as compared with at least 3 days needed for the SureSelect Target Enrichment System in-solution capture. Initial DNA requirements of the Access Array System/MiSeq protocol are also low (50 ng vs. 1 ug), and most of the process is automated, reducing the risks of contamination and human error to a minimum. Recently, Hollants et al.[24]

published a protocol for FH mutation detection that was also based on the Access Array System, but the amplicons were sequenced using the pyrosequencing-based GS-FLX (Roche, Branford, CT) platform. They successfully identified all short variants but, as with our study, could not detect large rearrangements. Our Access Array System protocol is here combined with the MiSeq sequencing platform (a benchtop personal NGS system specifically developed to suit the needs of the routine laboratory setting) and offers similar quality to GS-FLX but at a lower sequencing cost.[25] The use of the MiSeq system is also reported to limit errors of variant calls within homopolymer regions that are known to occur in pyrosequencing.[26]

Novel variants identified in the validation study

The *LDLR* promoter variant c.–227G>T has not previously been reported in FH patients, or in other populations, but was investigated previously as part of a study delineating the conserved footprint 1 site. A luciferase assay showed that the c. –227G>T variant had ~75% transcription levels as compared with the wild-type site.[27] Although a 25% reduction is not a definitive decrease in promoter activity, it cannot be excluded that this change is sufficient to cause raised cholesterol levels in this patient. Family members were unavailable for segregation analysis, and we therefore classify this variant as being of uncertain significance pending further functional and segregation data. The remaining two novel variants in *APOB*, rs12714097 and rs72653087, were located outside *LDLR*-binding sites, regions currently not associated with hypercholesterolemia, and therefore their pathogenicity remains to be elucidated.

Genotype–phenotype correlation in the prospective cohort

In the United Kingdom, FH diagnosis is made based on the Simon Broome criteria that identify patients with definite or possible FH. Following the National Institute for Health and Clinical Excellence guidelines,[6] genetic testing is recommended for all suspected index cases and should be followed by family cascade screening if a mutation is found. Our study focused on a group of consecutive patients referred for genetic screening by a local lipid clinic over the course of 1 year.

To assess clinical significance, all rare variants (minor allele frequency < 0.01) that were identified in *LDLR*, *APOB*, or *PCSK9* were compared against locus-specific FH databases.[6,15] All nonsense variants and insertion/deletion variants that have been found previously among FH patients were also considered pathogenic. Missense variants were considered to be pathogenic if they were predicted by Sorting Intolerant From Tolerant and PolyPhen to be deleterious and probably damaging, respectively, and were previously identified in FH patients. On the basis of these criteria, a missense change, p.Val827Ile, was included among the list of pathogenic variants, although the conservative valine to isoleucine substitution may not affect protein function. This residue is the third amino acid within the internalization signal for LDLR, NPVY, a position generally not necessary for correct internalization.[28] In addition, our segregation data suggest that this variant may not be disease causing, and we therefore classified this variant as being of unknown significance. Of the four splice site variants, three (c.313+1G>A, c.190+4A>T, and c.817+1G>A) were shown experimentally to disrupt splicing, two in published data[29,30] and one within this study, and are therefore listed as pathogenic. The remaining splice variant, c.941-4G>A, was described previously in FH patients, but its pathogenicity remains to be experimentally confirmed. An RNA sample from this individual was not available for confirmation within this study.

The mutation detection rates among patients with definite (66%) and possible (26%) FH are comparable to those of previously published studies.[7] One mutation was identified among 13 patients who did not fulfill Simon Broome criteria. This suggests that a number of patients with high cholesterol in the general population who do not fulfill formal clinical criteria for FH diagnosis may have FH-causing mutations. Such patients may therefore be diagnosed by the assays developed here, and, because these mutations are dominantly inherited, family cascade screening would probably identify the same mutations in 50% of their first-degree relatives.

The types of mutation identified here reflect the distribution of variants published in the *LDLR* locus–specific database,[15] with exonic substitutions (56%) being the most common, followed by short insertions/deletions (28%). Most of the mutations identified here were unique, with only 20% common among UK FH patients that would be identified using the Elucigene FH20 commercial ARMS kit. In our prospective study, we did not identify any mutations in *PCSK9* or large rearrangements of *LDLR*. These variants are generally rare among patients with FH,[7] and we therefore consider it unlikely that our results are biased in any way.

Recently, the importance of screening additional regions of *APOB* has been highlighted by Motazacker et al.,[31] who identified two patients with novel mutations located outside the commonly sequenced region of *APOB* exon 26. Screening of entire coding regions instead of focusing on mutation hot spots is therefore likely to enhance the discovery of FH-causing mutations. The format of next-generation assays is flexible and can be readily extended to include full coding regions of *APOB* and *PCSK9*, as well as coding regions of other medically relevant genes, such as *APOE* and *SLCO1B1*, which currently need to be genotyped separately.

Our current Access Array System covers the same regions that would be screened using conventional Sanger sequencing, and therefore the number of variants identified would be no larger than that identified after Sanger sequencing. The SureSelect Target Enrichment System design also included genes in cholesterol metabolism pathways and regions of *APOB* and *PCSK9* outside those known to be causative of FH. To establish the pathogenicity of such variants, further *in silico* and experimental analyses will be required. For detection of variants outside known FH-causing genes or gene regions, we focused on variants that are rare (minor allele frequency < 0.01), are probably pathogenic (i.e., located in coding regions, promoters, and exon/intron boundaries), and

are present only in individuals without an existing molecular diagnosis of FH. Using these filters, we were left with a manageable list of variants (<10) that we decided to follow up. When screening larger regions in more patients, the identification of large numbers of variants of unknown significance will necessitate further extensive family segregation and functional assays. However, the increase of data available in public databases and improvements of bioinformatics tools will allow more efficient filtering of variants than is available at present.

Statin-induced myopathy

Previous published data on larger cohorts show an odds ratio of 4.5 for association of rs4149056 with myopathy,[32] and on that basis, guidelines have recommended avoiding higher doses of simvastatin in heterozygotes or homozygotes for rs4149056.[33] Association testing restricted to more severe side effects (transaminases or creatine kinase greater than three times the upper limit of normal) was not informative in our cohort due to the small sample size. However, less-severe patient-reported adverse effects to statin treatment such as myalgia have an effect on patient adherence to treatment in this high-risk population. Therefore, the association identified here between *SLCO1B1* genotype and a wider range of milder adverse effects in this lipid clinic cohort gives an insight into the potential benefits of prospective genotyping before the initiation of statin treatment.

In conclusion, we have shown the potential utility of sequence target enrichment methods in combination with NGS in molecular diagnostics of FH. Owing to the comprehensive coverage, SureSelect Target Enrichment System protocols (either whole-exome or region-specific) may provide the most benefit for studies that aim to identify novel disease-causing genes and for diseases for which the number of genes needing to be screened is very large. In FH diagnostics for which <5 genes need to be analyzed, PCR-based enrichment techniques offer more streamlined protocols that provide high sensitivity for mutation detection and may offer an affordable option for clinical screening of large numbers of patients with suspected FH. If adopted, greater numbers of FH patients and their relatives may potentially benefit from early diagnosis and treatment.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

ACKNOWLEDGMENTS

We acknowledge funding from the MRC Clinical Sciences Centre, from the British Heart Foundation to S.A. and T.J.A.; from a Wellcome Trust Clinical Training Fellowship to E.R.A.T.; and from the National Institute for Health Research–funded Imperial Biomedical Research Centre to D.K., M.M., and T.J.A.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

1. Marks D, Thorogood M, Farrer JM, Humphries SE. Census of clinics providing specialist lipid services in the United Kingdom. *J Public Health (Oxf)* 2004;26:353–354.
2. Defesche JC. Familial hypercholesterolemia. In: Betteridge DJ (ed). *Lipids and Vascular Disease*. Martin Dunitz: London, 2000:65–76.
3. Williams RR, Hunt SC, Schumacher MC, et al. Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. *Am J Cardiol* 1993;72:171–176.
4. Scientific Steering Committee on behalf of the Simon Broome Register Group. Risk of fatal coronary heart disease in familial hypercholesterolaemia. *BMJ* 1991;303(6807):893–896.
5. Wierzbicki AS, Humphries SE, Minhas R; Guideline Development Group. Familial hypercholesterolaemia: summary of NICE guidance. *BMJ* 2008;337:a1095.
6. Leigh SE, Foster AH, Whittall RA, Hubbart CS, Humphries SE. Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. *Ann Hum Genet* 2008;72(Pt 4):485–498.
7. Taylor A, Wang D, Patel K, et al. Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. *Clin Genet* 2010;77:572–580.
8. Soutar AK, Naoumova RP. Mechanisms of disease: genetic causes of familial hypercholesterolemia. *Nat Clin Pract Cardiovasc Med* 2007;4:214–225.
9. Tosi I, Toledo-Leiva P, Neuwirth C, Naoumova RP, Soutar AK. Genetic defects causing familial hypercholesterolaemia: identification of deletions and duplications in the LDL-receptor gene and summary of all mutations found in patients attending the Hammersmith Hospital Lipid Clinic. *Atherosclerosis* 2007;194:102–111.
10. Taylor A, Tabrah S, Wang D, et al. Multiplex ARMS analysis to detect 13 common mutations in familial hypercholesterolaemia. *Clin Genet* 2007;71:561–568.
11. Alonso R, Defesche JC, Tejedor D, et al. Genetic diagnosis of familial hypercholesterolemia using a DNA-array based platform. *Clin Biochem* 2009;42:899–903.
12. Flicek P, Amode MR, Barrell D, et al. Ensembl 2012. *Nucleic Acids Res* 2012;40(database issue):D84–D90.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–1760.
14. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
15. Usifo E, Leigh SE, Whittall RA, et al. Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: update and pathological assessment. *Ann Hum Genet* 2012;76:387–401.
16. The International HapMap Project. *Nature* 2003;426(6968):789–796.
17. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467(7319):1061–1073.
18. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* 2012;28:2747–2754.
19. Taylor A, Martin B, Wang D, Patel K, Humphries SE, Norbury G. Multiplex ligation-dependent probe amplification analysis to screen for deletions and duplications of the LDLR gene in patients with familial hypercholesterolaemia. *Clin Genet* 2009;76:69–75.
20. Walsh T, Lee MK, Casadei S, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci USA* 2010;107:12629–12633.
21. Redin C, Le Gras S, Mhamdi O, et al. Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström syndromes. *J Med Genet* 2012;49:502–512.
22. Goossens D, Moens LN, Nelis E, et al. Simultaneous mutation and copy number variation (CNV) detection by multiplex PCR-based GS-FLX sequencing. *Hum Mutat* 2009;30:472–476.
23. Ware JS, John S, Roberts AM, et al. Next generation diagnostics in inherited arrhythmia syndromes: a comparison of two approaches. *J Cardiovasc Transl Res* 2013;6:94–103.
24. Hollants S, Redeker EJ, Matthijs G. Microfluidic amplification as a tool for massive parallel sequencing of the familial hypercholesterolemia genes. *Clin Chem* 2012;58:717–724.
25. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012;30:434–439.

26. Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin JF. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* 2011;12:245.
27. Dhawan P, Chang R, Mehta KD. Identification of essential nucleotides of the FP1 element responsible for enhancement of low density lipoprotein receptor gene transcription. *Nucleic Acids Res* 1997;25:4132–4138.
28. Chen WJ, Goldstein JL, Brown MS. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. *J Biol Chem* 1990;265:3116–3123.
29. Jensen HK, Jensen LG, Hansen PS, Faergeman O, Gregersen N. Two point mutations (313 + 1G –> A and 313 + 1G –> T) in the splice donor site of intron 3 of the low-density lipoprotein receptor gene are associated with familial hypercholesterolemia. *Hum Mutat* 1996;7:269–271.
30. Holla ØL, Nakken S, Mattingsdal M, et al. Effects of intronic mutations in the LDLR gene on pre-mRNA splicing: comparison of wet-lab and bioinformatics analyses. *Mol Genet Metab* 2009;96:245–252.
31. Motazacker MM, Pirruccello J, Huijgen R, et al. Advances in genetics show the need for extending screening strategies for autosomal dominant hypercholesterolaemia. *Eur Heart J* 2012;33:1360–1366.
32. Link E, Parish S, Armitage J, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med* 2008;359(8): 789–799.
33. Wilke RA, Ramsey LB, Johnson SG, et al.; Clinical Pharmacogenomics Implementation Consortium (CPIC). The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin Pharmacol Ther* 2012;92:112–117.

This work is licensed under a Creative Commons Attribution Unported 3.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/deed.en_US