

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
13 November 2014 (13.11.2014)



(10) International Publication Number
WO 2014/181107 A9

- (51) International Patent Classification:
C12Q 1/68 (2006.01)
- (21) International Application Number:
PCT/GB2014/051399
- (22) International Filing Date:
8 May 2014 (08.05.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
1308313.4 9 May 2013 (09.05.2013) GB
- (71) Applicant: **MEDICAL RESEARCH COUNCIL**
[GB/GB]; 2nd Floor, David Philips Building, Polaris
House, North Star Avenue, Swindon SN2 1FL (GB).
- (72) Inventors: **AITMANN, Timothy J.**; MRC Clinical Sci-
ences Centre, Imperial College, Hammersmith Hospital,
Du Cane Road, London W12 0NN (GB).
VANDROVKOVA, Jana; MRC Clinical Sciences Centre,
Imperial College, Hammersmith Hospital, Du Cane Road,
London W12 0NN (GB).

(74) Agent: **SCRIPT IP LIMITED**; 18 Bridge Street, Frome
Somerset BA11 1BB (GB).

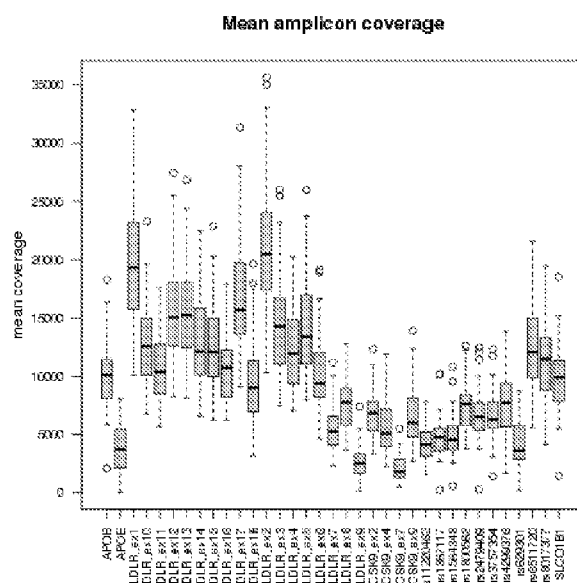
(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.

(84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

[Continued on next page]

(54) Title: GENETIC METHOD OF AIDING THE DIAGNOSIS AND TREATMENT OF FAMILIAL HYPERCHOLESTERO-
LAEMIA

Figure 1



(57) Abstract: The invention relates to a method for collect-
ing information useful in aiding the diagnosis and treatment
of familial hypercholesterolaemia (FH), the method com-
prising providing a sample of DNA from the subject, assay-
ing said DNA for (i) mutation in one or more genes which is
indicative of FH diagnosis (ii) presence of one or more
SNPs indicative of incremental risk of high cholesterol; and
(iii) mutation in a gene which is indicative of likelihood of
statin toxicity; wherein assaying said DNA comprises the
step of contacting said DNA sample with one or more
primers for amplification and/or sequencing of the relevant
segment(s) of said DNA characterised in that each of said
assays of steps (i), (ii) and (iii) is carried out in a single as-
say cycle. The invention also relates to a method of treat-
ment, and to primers.



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(48) Date of publication of this corrected version:

12 February 2015

Published:

- *with international search report (Art. 21(3))*
- *with information concerning authorization of rectification of an obvious mistake under Rule 91.3 (b) (Rule 48.2(i))*

(15) Information about Correction:

see Notice of 12 February 2015

GENETIC METHOD OF AIDING THE DIAGNOSIS AND TREATMENT OF FAMILIAL HYPERCHOLESTEROLAEMIA

FIELD OF THE INVENTION

- 5 The invention is in the field of hypercholesterolaemia. In particular, the invention relates to the provision of a comprehensive technique for patient management in the field of familial hypercholesterolaemia.

BACKGROUND TO THE INVENTION

- 10 Familial hypercholesterolaemia (FH, OMIM #143890) is a common Mendelian disorder that affects 1 in 500 individuals and it is estimated that 12 million of people are affected worldwide (Goldstein et al., 2001). Patients with FH have raised serum cholesterol levels from birth and increased arterial deposition of low density lipoprotein (LDL) cholesterol leading to premature coronary heart disease (CHD). If
15 untreated or inadequately treated half of men with FH develop heart disease before the age of 55 and one third of women before the age of 60. Despite the fact that early onset CHD can be prevented by cholesterol-lowering drugs such as statins, less than a quarter of FH patients have currently been identified in the UK. (Marks, Thorogood et al. 2004) One of the main reasons for FH being underdiagnosed are the cost and time associated
20 with current diagnostic techniques.

- In the UK the clinical diagnosis of FH is based on the Simon Broome criteria of cholesterol levels, presence of tendon xanthomata, family history and genetic testing. (Scientific Steering Committee on behalf of the Simon Broome Register Group 1991)
- 25 The UK National Institute for Health and Clinical Excellence (NICE) has recommended molecular diagnosis of FH probands and cascade screening of their families as the most effective way of identifying the estimated 75% of FH patients in the UK who are unaware of their condition and are therefore not on treatment which could reduce their high cardiovascular risks to near population levels (DeMott, 2008; Neil et al., 2000),
30 and a recent cost-benefit analysis indicated that this approach (used together with LDL cholesterol measurement) applied to patients with probable FH was more effective than screening using LDL measurement alone or limiting screening to patients with definite FH (Nherera et al., 2011). Monogenic FH is inherited in an autosomal dominant manner and therefore an affected individual has a 50% chance of
35 transmitting the disease causing mutation to a child.

Most cases of FH are caused by mutations in the *LDLR* gene that encodes the low density lipoprotein receptor protein, which binds LDL particles at the hepatic cell membrane and internalises 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 which can encompass multiple exons. Some mutations have been found in many unrelated individuals with FH, while others are found rarely. (Leigh, Foster et al. 2008) Mutations in two other genes, *PCSK9* and *APOB*, can also cause the FH phenotype, but in less than 10% of mutation positive cases. (Taylor, Wang et al. 2010)

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 (Tosi, Toledo-Leiva et al. 2007). These molecular techniques are sensitive and specific, but due to 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 specific mutations and populations, and therefore do not detect less common known mutations or any novel mutations (Taylor, Tabrah et al. 2007, Alonso, Defesche et al. 2009) .

Thus, familial hypercholesterolaemia (FH) is a known condition. Sufferers monogenic FH possess a mutation in one of three known FH diagnostic genes. These genes are *APOB*, *LDLR* and *PCSK9*. Standard molecular diagnostic techniques detect mutations in 40% of FH patients and it is believed that a proportion of remaining cases have raised cholesterol due to a combined, polygenic, effect of several LDL-C-raising common variants. This is discussed in more detail below. The current standard test for FH is a genetic test. This involves the analysis of the three diagnostic genes in an individual for mutations. This is currently accomplished by conventional sequencing techniques. This is most often accomplished using capillary sequencing. This is a very laborious process. This is a very time consuming process. For these reasons, many people lack a genetic diagnosis. These patients who lack a genetic diagnosis are typically handled solely on the basis of the symptoms which they present with. This is a further problem in the art.

The gold standard treatment for hypercholesterolaemia is a prescription of statins. Statins control (reduce) cholesterol levels when administered to patients. Statins are a long term treatment. Many patients will be prescribed statins for the rest of their natural lives when presenting with hypercholesterolaemia.

5

Cholesterol lowering therapy has been very successful in FH and has led to reduced mortality and morbidity. Statin drugs (HMGCoA reductase inhibitors) are the mainstay of treatment for FH, but are not well tolerated by all patients. Muscular side-effects ranging from muscle pain through to life-threatening rhabdomyolysis are among the principal reasons for discontinuation of treatment. The rs4149056 variant in *SLCO1B1* (solute carrier organic anion transporter family member 1B1) has been shown to be associated with statin-induced myopathy (The Search Collaborative Group, 2008). Thus, a proportion of the population is predisposed to statin toxicity. In other words, when statins are administered to a cohort of patients, a proportion of those patients will have an underlying statin toxicity issue. Statin toxicity causes a range of harrowing effects, including muscular pain, muscle weakness, death and/or erosion of muscle fibres, and in severe cases a resulting trauma to the kidneys which can result in kidney failure.

20 It is a problem in the art that patients suffering from statin toxicity are typically suffering a long term decline. As the toxicity problem runs on a patient gets weaker and weaker.

25 It is a problem in the prior art that the effects of statin toxicity are prone to end up as a separate diagnostic investigation unless the physician is alert to the risk and recognises the symptoms early in the treatment.

Casual cholesterol testing is becoming increasingly common outside mainstream medical service provision. As a result of raising of the public awareness of the dangers of hypercholesterolaemia, individuals in the population are becoming concerned about their own cholesterol levels. In response to this concern, third party providers are offering cheap cholesterol testing. These tests can be as simple as a skin prick conducted on a "while you wait" basis by operatives with little or no medical training. Although these tests might be useful as an initial indicator of a cholesterol level, the operatives providing these tests are usually not medically qualified and cannot offer any prescription or treatment recommendations. Therefore, individuals taking a casual cholesterol test and receiving an indication of high cholesterol are referred back to

mainstream healthcare providers. This is creating an influx of patients into doctors' surgeries presenting results of casual cholesterol tests and asking for treatment or further investigation. This is placing a large burden on mainstream health service provision. In addition, simple cholesterol levels obtained from such casual prick tests
5 provide no indication of any underlying cause or diagnosis. As indicated above, the current technique for analysing or diagnosing patients presenting with hypercholesterolaemia is laborious and expensive. The extra demand for these resource intensive tests is creating a further problem to medical healthcare providers.

10 The present invention seeks to overcome problems associated with the prior art.

SUMMARY OF THE INVENTION

In the prior art, the only genetic test for familial hypercholesterolaemia (FH) involves
15 conventional sequencing. There are three main genes which are diagnostic of FH. The prior art screen typically requires PCR amplification of the relevant fragments from these genes. This is followed by conventional sequencing. The approach used is usually capillary sequencing, which is laborious and costly. This results in many people lacking a genetic diagnosis of their condition. In addition, it means that the physician
20 attempting to treat the patient has a paucity of information about their condition.

The present inventors have studied FH in detail. The invention is based on a combined test which individually examines multiple factors which are involved in FH. In summary, the inventors teach the parallel analysis of three patient attributes – (i) assay
25 of the three main diagnostic genes for FH (APOB, LDLR and PCSK9); (ii) assay of polygenic FH SNPs (a collection of lower penetrance incremental risk indicators); and (iii) assay of the patient's SLCO1B1 genotype (the genetic indicator of statin toxicity). The inventors teach for the first time the combination of these three elements into a single diagnostic test. The advantage of this is that the physician is then appraised with
30 a complete picture of the patient's condition, including their incremental risk indicators, and also at the same time an indication of whether they are likely to have a negative response to the gold standard statin treatment. This combination has never been attempted in the prior art. This combination has advantages in targeting the treatment to the patient in the correct manner, and also avoiding the extremely
35 negative side effects for that proportion of the population who would typically be placed on statin treatment even though they are at risk of serious side effects from statin

toxicity. The present invention advantageously alleviates these problems associated with the prior art.

In addition to this main principle of the invention, the inventors teach that
5 advantageously the tests are conducted using a paired approach of micro-fluidics with next generation sequencing. This is also a departure from the prior art techniques which have relied on conventional PCR amplification coupled to ordinary sequencing techniques such as capillary sequencing. By embracing the micro-fluidic and next
10 generation sequencing (NGS) approaches in a combined method, the inventors are able to dramatically increase the efficiency of the test, and at the same time reduce the cost to a fraction of the cost of the current conventional testing.

These and other advantages are explained in more detail below.

15 Thus, in one aspect the invention provides a method for collecting information useful in aiding the diagnosis and treatment of familial hypercholesterolaemia (FH), the method comprising providing a sample of nucleic acid such as DNA from the subject, assaying said nucleic acid such as DNA for
(i) mutation in one or more genes which is indicative of FH diagnosis
20 (ii) presence of one or more SNPs indicative of incremental risk of high cholesterol; and
(iii) mutation in a gene which is indicative of likelihood of statin toxicity;
wherein assaying said nucleic acid such as DNA comprises the step of contacting said nucleic acid such as DNA sample with one or more primers for amplification and/or
25 sequencing of the relevant segment(s) of said nucleic acid such as DNA
characterised in that each of said assays of steps (i), (ii) and (iii) is carried out in a single assay cycle.

In another aspect, the invention relates to a method as described above wherein
30 assaying said nucleic acid such as DNA further comprises determining the nucleotide sequence of the relevant segment(s) of said nucleic acid such as DNA, and inferring from said nucleic acid such as DNA sequence whether said mutations of steps (i) and (iii), and said one or more SNPs of step (ii), are present.

35 The invention may be applied to any nucleic acid; suitably the nucleic acid is DNA. The invention is described with reference to DNA as an exemplary nucleic acid.

By 'single assay cycle' is meant that the assay steps are carried out on the same sample. Suitably this means in a multiplexed assay. Suitably this means at the same time (i.e. simultaneously). Suitably this means that the single assay cycle, which may consist of multiple tubes or reactions as necessary, is carried out as a single coherent test event.

5 Suitably this means that the results of the individual assays of steps (i), (ii) and (iii) are delivered (e.g. produced/determined) together.

In another aspect, the invention relates to a method as described above wherein said one or more genes indicative of FH diagnosis is selected from the group consisting of
10 APOB, LDLR and PCSK9. Each gene in this group of genes shares the common property of being indicative of monogenic FH.

In another aspect, the invention relates to a method as described above wherein said mutation in one or more genes which is indicative of FH diagnosis is selected from
15 those listed in Table 2.

In another aspect, the invention relates to a method as described above wherein said SNP is selected from Table A or Table B. Each SNP in this group of SNPs shares the common property of being indicative of incremental risk of high cholesterol. Suitably
20 each SNP in this group of SNPs shares the common property of being indicative of polygenic FH.

In another aspect, the invention relates to a method as described above wherein said gene indicative of risk of statin toxicity is SLCO1B1.

25 Suitably step (iii) comprises determining the presence of the rs4149056 or rs4149056 variant of SLCO1B1.

Suitably step (iii) comprises determining the presence of the rs4149056 variant of SLCO1B1.

30 In another aspect, the invention relates to a method as described above wherein said DNA sample is subjected to micro-fluidic PCR amplification of the DNA segment(s) of interest.

Suitably said PCR is carried out using at least one primer selected from Table 3.

Suitably said PCR is carried out using each of the primers in Table 3.

35

Enrichment is suitably accomplished by PCR.

In another aspect, the invention relates to a method as described above wherein sequence information is determined by next generation sequencing (NGS).

5 In another aspect, the invention relates to a method of treating a subject comprising performing the method according to any preceding claim wherein if an increased likelihood of FH is identified, then cholesterol lowering medicament is administered to said subject, wherein said medicament is selected according to the subject's SLCO1B1 genotype.

10 Suitably selection of the medicament comprises selection of one or more of the upper limit of the dose of said medicament for said subject, the dose of said medicament to be administered to said subject, or the particular medicament to be administered to said subject.

Suitably said medicament is a statin.

15 In another aspect, the invention relates to a primer selected from Table 3.

In another aspect, the invention relates to a set of primers comprising at least one pair of the primers in Table 3.

20 In another aspect, the invention relates to a set of primers according to claim 17 comprising each of the primers in Table 3.

DETAILED DESCRIPTION OF THE INVENTION

25 In one embodiment the invention may relate to a method for determining whether a subject has a predisposition for monogenic familial hypercholesterolaemia by virtue of a mutation in one of the genes selected from LDLR, PCSK9, APOB or polygenic familial hypercholesterolaemia by virtue of combination of genotypes in (rs6511720) LDLR, (rs2479409) PCSK9, (rs1367117) APOB, (rs629301) CELSR2, (rs4299376) ABCG8, 30 (rs1564348) SLC22A1, (rs1800562) HFE, (rs3757354) MYLIP, (rs11220462) ST3GAL4, (rs8017377) NYNRIN, (rs429358, rs7412) APOE and for determining individual's response to statin treatment by virtue of a variant (rs4149056) in SLCO1B1 gene comprising:

- 35 - carrying out a nucleic acid amplification step on a biological sample from the subject, wherein the nucleic acid amplification step uses all primer pairs, the primer pairs comprising a forward primer with a forward primer sequence

- selected from SEQ ID NO: 1 to SEQ ID NO:60; or a reverse primer with a reverse primer sequence selected from SEQ ID NO: 61 to SEQ ID NO:120;
- sequencing at least part of the amplification product, the amplification product being a sequence of at least a part of one of the genes LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLCO1B1 and
 - determining whether the LDLR, PCSK9 or APOB amplification product represents a sequence that has a mutation that is indicative of a risk for a familial hypercholesterolaemia
 - determining whether combination of LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE amplification products have SNPs that are indicative of a risk for a polygenic familial hypercholesterolaemia
 - determining whether the SLCO1B1 amplification product represents a sequence that has an SNP that is indicative of a risk for adverse statin effect.

In detail the invention may be operated as follows. The exemplary PCR primers described (or alternate primers if desired) may be used to assist in sequencing the full gene LDLR and selected genomic regions CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLCO1B1 of the subject.

The subject's gene sequence(s) is then compared with known variants of these genes associated with monogenic or polygenic FH. From the comparison, it is determined whether any particular DNA variant or variants exist in the subject. It is then possible to correlate the variants to risk of FH and statin related myopathy.

The comparison of the sequence information of the subject with known allele sequences can be carried out by making comparisons using established and freely available bio-informatic techniques (e.g. GATK or platform-specific software) with comparison to the reference human genome, 1000 genomes data and in high-quality controls, which we have established.

After comparing the subject's gene sequences to known variants, an informed diagnostic interpretation of the risk of FH can be made. Whenever a cause of monogenic FH, a mutation, is identified through genetic testing, family-specific genetic testing can also be used to identify relatives at-risk for the disease. Thereafter, a treatment decision can be made. Clinical decision algorithms for assigning pathogenicity are well established. In short, if a mutation is known to cause the disease

from previous publications that have shown segregation of the mutation in a family or enrichment of a mutation in cases versus controls then it is assigned as disease causing. New mutations, not present in the reference human genomes, that segregate in a family may also be assigned pathogenicity in combination with pathogenicity scoring tools (conservation, polyphen and sift). Variants of unknown significance are assigned as such and common benign variants are also identified.

Polygenic predisposition to FH is established by calculating a score, an LDL-c gene score, based on presence or absence of selected variants in LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE.

As described above, the invention provides a method of determining whether a subject has a predisposition for familial hypercholesterolaemia and adverse reaction to statin treatment.

DEFINITIONS

"Amplification" of polynucleotides may utilise methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. Also useful are strand displacement amplification (SDA), thermophilic SDA, and nucleic acid sequence based amplification (3SR or NASBA).

One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes. These methods are well known and widely practiced in the art. See, e.g., U.S. Patents 4,683,195 and 4,683,202 and Innis et al., 1990 (for PCR); Wu and Wallace, 1989 (for LCR); U.S. Patents 5,270,184 and 5,455,166 and Walker et al., 1992 (for SDA); Spargo et al., 1996 (for thermophilic SDA) and U.S. Patent 5,409,818, Fahy et al., 1991 and Compton, 1991 for 3SR and NASBA. Reagents and hardware for conducting PCR are commercially available.

Primers useful to amplify sequences from the LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLCO1B1 genes are suitably complementary to, and hybridise specifically to, sequences in the LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLCO1B1 genes or in regions that flank a region therein. Sequences in the LDLR, PCSK9, APOB,

CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLCO1B1 genes generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified
5 genomic segments has been described by Scharf et al., 1986. The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases.

10 The term “comprising” as used in this specification and claims means “consisting at least in part of”, that is to say when interpreting statements in this specification and claims which include the term, the features, prefaced by that term in each statement, all need to be present but other features can also be present. Related terms such as “comprise” and “comprised” are to be interpreted in a similar manner.

15 The term “fragment” of a polynucleotide sequence provided herein is a subsequence of contiguous nucleotides that is capable of specific hybridisation to a target of interest, e.g., a sequence that is at least 10 nucleotides in length. The fragments may comprise 10, preferably 15 nucleotides, preferably at least 20 nucleotides, more preferably at least 30 nucleotides, more preferably at least 40 nucleotides, more preferably at least
20 50 nucleotides and most preferably at least 60 nucleotides of contiguous nucleotides of a polynucleotide. A fragment of a polynucleotide sequence can be used as a primer, a probe, included in a microarray, or used in polynucleotide-based identification methods.

25 The term “hybridise under stringent conditions”, and grammatical equivalents thereof, refers to the ability of a polynucleotide molecule to hybridise to a target polynucleotide molecule (such as a target polynucleotide molecule immobilised on a DNA or RNA blot, such as a Southern blot or Northern blot) under defined conditions of temperature and salt concentration. The ability to hybridise under stringent hybridisation conditions can
30 be determined by initially hybridising under less stringent conditions then increasing the stringency to the desired stringency.

With respect to polynucleotide molecules greater than about 100 bases in length, typical stringent hybridisation conditions are no more than 25 to 30°C (for example,
35 10°C) below the melting temperature (T_m) of the native duplex (see generally, Sambrook et al, Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Ausubel et al, 1987, Current Protocols in Molecular Biology,

Greene Publishing). T_m for polynucleotide molecules greater than about 100 bases can be calculated by the formula $T_m = 81.5 + 0.41\% (G + C - \log(Na^+))$ (Sambrook et. al, Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Bolton and McCarthy, 1962, PNAS 84:1390). Typical stringent conditions for a polynucleotide of greater than 100 bases in length would be hybridisation conditions such as prewashing in a solution of 6X SSC, 0.2% SDS; hybridising at 65°C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65°C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65°C.

In one embodiment stringent conditions use 50% formamide, 5 x SSC, 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulphate at 42°C, with washes at 42°C in 0.2 x SSC and 50% formamide at 55°C, followed by a wash comprising of 0.1 x SSC containing EDTA at 55°C.

With respect to polynucleotide molecules having a length less than 100 bases, exemplary stringent hybridisation conditions are 5 to 10°C below T_m . On average, the T_m of a polynucleotide molecule of length less than 100 bp is reduced by approximately (500/oligonucleotide length) °C.

The term “nucleic acid” as used herein, means a single or double-stranded deoxyribonucleotide or ribonucleotide polymer of any length, and include as non-limiting examples, coding and non-coding sequences of a gene, sense and antisense sequences, exons, introns, genomic DNA, cDNA, pre-mRNA, mRNA, rRNA, siRNA, miRNA, tRNA, ribozymes, recombinant polynucleotides, isolated and purified naturally occurring DNA or RNA sequences, synthetic RNA and DNA sequences, nucleic acid probes, primers, and fragments thereof. Reference to a polynucleotide(s) is to be similarly understood. Most suitably nucleic acid is DNA.

The term “next generation sequencing” (NGS) is DNA sequencing using high-throughput sequencing technologies that parallelise the sequencing process, so that thousands or millions of sequences can be generated at once. This speed allows for low-cost sequencing of DNA, allowing, for example, for DNA sequencing to be viably used in clinical practice. There are generally three steps in NGS: template preparation, sequencing and imaging, and data analysis. Template preparation commonly requires amplification of the DNA, for example by using PCR. Three commonly used methods of sequencing are classified as cyclic reversible termination (CRT), single-nucleotide

addition (SNA) and real-time sequencing. The imaging coupled with these techniques often range from measuring bioluminescent signals to four-colour imaging of single molecular events. Examples of specific NGS technologies include, but are not limited to, Massively Parallel Signature Sequencing (MPSS), polony sequencing, 454
5 pyrosequencing, Illumina (Solexa) sequencing (sequencing based on reversible dye-terminators), Sequencing by Oligonucleotide Ligation and Detection (SOLiD) sequencing, ion semiconductor sequencing, DNA nanoball sequencing, Helioscope™ single scope sequencing, single molecule real time (SMRT) sequencing, single molecule real time (RNAP) sequencing and nanopore DNA sequencing.

10

The term “oligonucleotide(s)” are nucleic acids that are usually between 5 and 200 contiguous bases, and often between 5-10, 5-20, 10-20, 10-50, 15-50, 15-100, 20-50, 20-100, 20-200, 50-200, or 100-200 contiguous bases. An oligonucleotide that is longer than about 20 contiguous bases may be referred to as a polynucleotide. A
15 polymorphic site (polymorphism) can occur at any position within an oligonucleotide.

A “polymorphic site” refers the position in a nucleic acid sequence at which a polymorphism occurs. A polymorphic site may be as small as one base pair.

20 The term “polymorphism” refers to a genetic variation, or the occurrence of two or more genetically determined alternative sequences at a single genetic locus in a population. Each version of the sequence with respect to the polymorphic site is referred to as an “allele” of the polymorphic site. Preferred polymorphisms have two alleles, with the minor allele occurring at a frequency of greater than 1%, and more
25 preferably greater than 5% or 10% of a selected population. The allelic form occurring most frequently in a selected population is sometimes referenced as the “wildtype” form. The allelic form occurring less frequently in a selected population is sometimes referenced as the “mutant” form. Diploid organisms may be homozygous or heterozygous for allelic forms. A biallelic polymorphism has two forms. A triallelic
30 polymorphism has three forms. Examples of polymorphisms include restriction fragment length polymorphisms (RFLPs), variable number of tandem repeats (VNTRs), single nucleotide polymorphisms (SNPs), dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu.

35 The term “pre-disposition”, when used in relation to FH syndrome or any similar phrase such as “propensity” or “susceptibility”, means that certain alleles have been discovered to be associated with, or predictive of, FH. These “high-risk” alleles may be

the minor (or mutant) allele, or the major (or wild-type) allele. These alleles are thus over-represented in frequency or carriage rate in individuals who are at risk of developing FH compared to individuals who are not susceptible to FH. Hence, the term “an individual’s susceptibility to FH” refers to a statistically higher, or lower, frequency
5 of FH in an individual carrying a particular polymorphic allele, or genotype (i.e. allelic or polymorphism pattern) in comparison to the frequency in a member of the population that does not carry the particular polymorphic allele, or genotype.

An individual that carries one or both high-risk alleles at a polymorphic site is said to
10 have a heterozygous or homozygous “high-risk” genotype for that particular polymorphic site, respectively. An individual that does not carry a particular high-risk allele is said to have a homozygous “low-risk” genotype.

The term “primer” refers to a polynucleotide, usually having a free 3' OH group, that
15 may be hybridised to a template and used for priming polymerisation of a polynucleotide complementary to the target.

“Probing”: In general, it is expected that analysis of the amplification products will take place by sequencing. However, it may also be desirable to analyse the amplification
20 products by “wet” binding probing to detect the presence or absence of specific sequences. Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte. The region of the probes which is used to bind to the analyte can be made completely complementary to the targeted region of the gene of
25 interest such as LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE or SLCO1B1 genes. Therefore, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency are used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of
30 factors during hybridization and during the washing procedure, including temperature, ionic strength, base composition, probe length, and concentration of formamide. These factors are outlined in, for example, Maniatis et al., 1982 and Sambrook et al., 1989.

The term “SNP” or “single nucleotide polymorphism” is a polymorphism that occurs at
35 a polymorphic site occupied by a single nucleotide. The site of the SNP is usually preceded by and followed by highly conserved sequences (e. g., sequences that vary in less than 1/100 or 1/1000 members of a population). As used herein, “SNPs” is the

plural of SNP. SNPs are most frequently diallelic. A most common allele of a SNP is called a “major” or “wild-type” allele and an alternative allele of said SNP is called a “minor” or “mutant” allele. A SNP usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. SNPs can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

SNPs tend to be evolutionarily stable from generation to generation and, as such, can be used to study specific genetic abnormalities throughout a population. If SNPs occur in the protein coding region it can lead to the expression of a variant, sometimes defective, form of the protein that may lead to development of a genetic disease. Such SNPs can therefore serve as effective indicators of the genetic disease. Some SNPs may occur in non-coding regions, but nevertheless, may result in differential or defective splicing, or altered protein expression levels. SNPs can therefore be used as diagnostic tools for identifying individuals with a predisposition for certain diseases, genotyping the individual suffering from the disease in terms of the genetic causes underlying the condition, and facilitating drug development based on the insight revealed regarding the role of target proteins in the pathogenesis process.

A “SNP location” or “SNP locus” is a polymorphic site at which a SNP occurs.

SAMPLE

The term “biological sample” as used herein means a biological sample derived from a patient to be screened. The biological sample may be any suitable sample known in the art in which the expression of the selected markers can be detected. Included are individual cells or cell populations such as obtained from bodily tissues or fluids. The biological sample may comprise for example a blood, tissue, saliva or buccal smear sample.

Suitably the biological sample comprises blood.

Suitably the biological sample comprises saliva.

In one embodiment, a biological sample comprising nucleic acid is obtained from a subject.

Nucleic acid is suitably extracted from the biological sample for assay/analysis according to any suitable method known in the art. If further guidance is required, the reader is referred to the examples section.

- 5 Thus suitably the sample for analysis comprises nucleic acid. Suitably the sample for analysis consists essentially of nucleic acid. Suitably the sample for analysis consists of nucleic acid.

Suitably the sample is an *in vitro* sample.

- 10 Suitably the sample is an extracorporeal sample.

- In one embodiment suitably the method is an *in vitro* method. In one embodiment suitably the method is an extracorporeal method. In one embodiment suitably the actual sampling of the subject (collection of biological sample) is not part of the method of the invention. Suitably the method does not involve collection of the biological sample. Suitably the sample is a sample previously collected. Suitably the method does not require the presence of the subject whose nucleic acid is being assayed. Suitably the sample is an *in vitro* sample. Suitably the method does not involve the actual medical decision, *stricto sensu*; such a decision *stricto sensu* would typically be taken by the physician.
- 15
20

Suitably the method of the invention is conducted *in vitro*. Suitably the method of the invention is conducted extracorporeally.

25 ADVANTAGES

It is an advantage of the invention that extra information is extracted from a single patient sample compared to the prior art techniques which provide less information per sample.

30

It is an advantage of the invention that the diagnostic testing is greatly reduced in cost, and is significantly cheaper per patient than existing prior art techniques.

- It is an advantage of the invention that performing each of the three tests described in parallel enables the physician to design a more effective treatment regime for each patient.
- 35

It is an advantage of the invention that a new diagnostic pathway is provided.

It is an advantage of the invention that a “one stop” diagnostic suite is provided which enables the diagnosis and management of patients presenting with high cholesterol.

5

Performing individual tests for the different sections of the analysis described herein according to prior art techniques would require a myriad of separate evaluations to be carried out, mostly by PCR and conventional sequencing. This would be an enormously expensive exercise. Apart from the fact that this combination has never been taught
10 before in the art, the inventors have also made this much more efficient and much more economical by teaching the parallel/multiplexed analysis of each of these elements in a single analysis.

15

The inventors teach the application of emerging technologies such as micro-fluidic amplification and next generation sequencing (NGS) in the field of familial hypercholesterolaemia. It is an advantage of these techniques that the tests can be performed on a small amount of DNA, meaning a smaller sample needing to be collected from the patient. In addition the combined method is more efficient on both time and labour.

20

The inventors teach for the first time the comprehensive joining together of mutation detection for FH with the assay for *SLCO1B1* and the polygenic FH genes. This is a novel approach and has not been taught previously and it is this tripartite assay that provides numerous advantages as set out herein.

25

The methods taught herein are advantageously both very sensitive and highly specific.

A new level of effectiveness is shown for the combined tests according to the invention which has not been demonstrated in the prior art.

30

As demonstrated by the accompanying data, the combined method of the invention works very well, in fact it is surprising how well it is shown to work.

35

The combined method of the invention advantageously provides a comprehensive picture of the patient's need for treatment, coupled to their tolerance of the conventional statin based therapies. This type of comprehensive picture was not available by any prior art technique.

The inventors have broken new ground in combining three types of assays into a single procedure. Nowhere in the prior art has anybody attempted to make such a combination before. The provision of a single management method for handling
5 patients with hypercholesterolaemia is an advance in the field.

Although certain individual tests are known in the prior art, they each provide only a single answer. Carrying out any one of the individual tests known in the art does not allow the skilled worker to arrive at the invention. It is an advantage of the invention
10 that toxicity/tolerance of the intended therapy is assessed at the same point in time as the diagnosis is facilitated by analysis of the underlying genetic predisposition.

The invention advantageously provides a comprehensive management package for patients with high cholesterol.

15

MULTIPLEXING

It will be noted that the SLCO1B1 statin induced myopathy variant and polygenic FH SNPs are not covered by current diagnostic tests in the art.

20

The assay of the invention is advantageously carried out as a multiplex. In other words, the separate and individual test which make up the assay of the invention are advantageously carried out on a single DNA sample.

25 Advantageously, the assays of the invention are carried out in a single run of a sequence analyser.

Advantageously, the assays of the invention are enriched in a single PCR run, such a micro-fluidic PCR run.

30

It is a specific advantage of the invention that the assays are conducted in a multiplex format. This provides the advantage of presenting a complete set of information to the physician so that they may diagnose and design a suitable treatment regime on the basis of the output of the multiplexed assay.

35

A multiplex assay has its normal meaning in the art of molecular biology. In other words, the individual tests being carried out as taught by the invention are performed

by a single sample. The individual tests of the invention are performed simultaneously. Suitably the individual tests of the invention are performed in a single PCR assay. Suitably the assays of the invention are performed in a single sequencing step.

- 5 The combination of the various tests taught into a single simultaneous (multiplexed) analysis is a departure from the prior art because the phenomenon of familial hypercholesterolaemia (whether monogenic or polygenic) is a separate medical condition from the status of sensitivity to statin toxicity. It is an advantage of the invention that analysis of these two phenomena is brought together in a single step
10 procedure. This is more efficient. This is cheaper. This saves labour. This provides a greater armoury of information to the physician at a single time point, bringing together diagnosis and design of an appropriate treatment regime.

- In addition, the cost of multiple sample collection is eliminated since only a single cell
15 run is required.

In addition, the trauma or inconvenience of multiple sample collection on the patient is also eliminated by performing the invention on a single sample.

- 20 It is a further advantage of the invention that the patient is assessed for both monogenic and polygenic FH in a single assay. Of course, the patient will require treatment if they have either monogenic FH or polygenic FH. It is exceptionally rare for a patient to have both monogenic FH and polygenic FH. Even if they do possess both genetic predispositions to FH, in any case the treatment would typically be the
25 same (although clearly the investigation of related family members would be different if they were diagnosed as having monogenic as opposed to polygenic FH). It is a result of the insight of the present inventors that it is taught to carry out the tests for both monogenic FH and polygenic FH in the same multiplexed sample. Clearly, this leads to an inbuilt redundancy since individual patients will typically only have either have only
30 monogenic FH or polygenic FH as explained above. Without the insight of the inventors, this would appear to be a very wasteful process incurring avoidable cost. For example, if the patient has monogenic FH there would appear to be no clinical advantage to also testing them for polygenic FH, and vice versa. However, the inventors teach against this prejudice, with their assay requiring simultaneous
35 examination for both monogenic and polygenic FH. The benefit of this dual approach is the elimination of the labour involved in carrying out the tests separately, as is currently standard in the art. A further benefit is the avoidance of the need to

investigate family members if an index case is found not to have a cause of monogenic FH but instead has polygenic FH. Thus, by designing a deliberate element of redundancy into the multiplexed analysis, the invention actually reaps rewards in terms of reduced cost and increased efficiency by eliminating the need for subsequent or
5 alternative tests based on the results of a first test being carried out.

MICRO-FLUIDICS

Micro-fluidics offers a method for conducting highly parallel PCR. Instead of carrying
10 out conventional PCR which allows amplification of (for example) 48 or 96 samples per run, parallel PCR permits up to approximately 2500 amplifications to be carried out in the same machine. For example, up to approximately 500 primer pairs can be used in a single run of the machine. This technique provides a massively parallel PCR amplification procedure. The inventors teach use of this technique to prepare patient
15 samples for sequencing.

It should be noted that micro-fluidics/highly parallel PCR is an especially suitable approach to the enrichment of the sample for sequencing. Other techniques for enrichment may equally be used if desired. For example, there are hybridisation
20 techniques which may be used to enrich the nucleic acids of interest. One example of this is the "Sure Select" hybridisation technique, such as from Agilent. However, this is a lengthy and expensive approach to enrichment. It is an advantage to use micro-fluidic PCR based enrichment as described in more detail herein.

25 GENETIC TESTING

In many disorders with genetic links, mutations in the exons of genes (the coding regions) have been found to have the greatest effects on disease, and can also act as indicators of disease risk. Mutations in the introns of a gene (the non-coding regions)
30 can also affect disease and indicate disease risk but are less readily interpreted. In order to detect mutations of a gene, it is often desirable or necessary to sequence the genomic DNA for the gene in question. That, in turn, generally requires the entire genomic DNA for the gene in question to be amplified as a first step. Typically PCR is used for that amplification.

35 Complete or full coverage of all genes or parts of genes known to predispose to a disease or to modify treatment outcome is optimal for comprehensive genetic testing.

During the last few years high-throughput 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 are becoming 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.

We teach the use of “next generation sequencing” (NGS) of genes or gene segments after enrichment such as by amplification (e.g. PCR amplification).

In order for the genomic DNA to be amplified, it is necessary for appropriate primers to be provided. A skilled person working the invention may design their own primers, for example using commercially available software to aid the process. The invention should not be regarded as limited solely to the particular exemplary primer sequences provided herein. However, determining the sequences to be used for such primers is not always simple: long amplicons can incorporate erroneous bases; mutations in the gene can interfere with primer annealing, and tertiary structure of the DNA can make certain sequences inaccessible. Automated platforms for primer design can have variable performance and custom design is often important. In this regard, the inventors have also provided a set of especially useful exemplary primers according to the present invention. These are presented in Table 3.

Thus the present inventors have designed a new series of primers which together provides complete coverage of the LDLR gene and clinically relevant parts of PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLCO1B1 genes. This level of completeness to be achieved by a single genetic assay provided by the complete new series of primer pairs set out in Table 3 has not previously been described. Further, because of the streamlined nature and very low cost of the assay, the assay is applicable to screening large numbers of patients for genetic causes of hypercholesterolaemia where previous assays were unsuitable.

Table 3 - (sometimes referred to as Table S1): Exemplary Primers

Gene	No.	Sense primer (F-sp)	SEQ ID NO	Anti-sense primer (R-sp)	SEQ ID NO	Len
APOB	APOB_ex26_3	TGTGGCTTCTCCAGCAAAAT	1	ACTTTTCCATTGAGTCATCTACCA	61	200
LDLR	LDLR_t1_17r_2	GCTATTGGAGGATCTTGAAAGGC	3	GGCAGGCGACTTGATTTGTGTGA	63	175
LDLR	LDLR_t1_19	CTTCTGTGGACAACAACAGCAAA	4	GCTTTTAAACCCGTGAAGCTCTG	64	198
LDLR	LDLR_t1_21	GGCgACACTTTTGAAGGACTG	5	CCCACGTCATTACAGCATTTCA	65	179
LDLR	LDLR_t1_22	TCTTGCAGTGAGGTGAAGACATTT	6	cCATGCTCGCAGCCTCT	66	190
LDLR	LDLR_t1_22r_1	CAGCTAGGACACAGCAGGT	7	GGGGCTCCCTCTCAACCTATTC	67	171
LDLR	LDLR_t1o_3	ATCCACAGCAACATCTACFGAC	8	CcCTTCTCTGCTCCCTCCATTC	68	191
LDLR	LDLR_t1or_1	CTGATGCCCTTCTCTCCTCCT	9	TTCCTCTTCACGCCCTTGGTAT	69	197
LDLR	LDLR_t11_1	GGCTGTTTCTTCCAGAATTCGTT	10	GTTTTCAAGTCACCGAGTAGA	70	179
LDLR	LDLR_t11_2	ATGTACTGGACTGACTGGGGAA	11	GAGCCACCTGGCAACCC	71	190
LDLR	LDLR_t12_1	CtCTGGGACTGGCATCAGC	12	CAGCCTCTTTTCATCCTCCAAGA	72	162
LDLR	LDLR_t12_2	TACTGGGFTGACTCCAAACTTCA	13	GTCTGTGTCTATCCGCCACCTAA	73	171
LDLR	LDLR_t13_1	CCTGTGTCTCATCCCAGTGTITA	14	CTCTGGGGACAGTAGGTTTTCAG	74	162
LDLR	LDLR_t13_2	TGGACAGATATCATCAACGAAGCC	15	TITCCACAAGGAGGTTTCAAGGT	75	181
LDLR	LDLR_t14_1	CTTCTGGTATAGCTGATGATCTCGT	16	GCAGGCGCAGGTAAACTTGG	76	163
LDLR	LDLR_t14_2	GAGTGAAGTGGTgTGAGAGGAC	17	GACACAGGACGCAGAAACAAGG	77	187
LDLR	LDLR_t15_1	ATTAGGCGCACACCTATGAGAAG	18	GGAGGTGTCGGGAACAGG	78	180

LDLR	LDLR_t15_2	CGTCAGGCTAAAGGTCAGCTC	19	GACCCGCTCTCTGGGTGAAGA	79	181
LDLR	LDLR_t16_1	CCATTTCTTGgTGGCCTTCCTTT	20	GGAGgCTCTGACCTGGG	80	180
LDLR	LDLR_t17_1	GGAGCTGGGTCTCTGGTCT	21	GACCTCATCCTCTGTGGTCTTCT	81	177
LDLR	LDLR_t17_2	TACAGTGCTCgTCGTCTTCCTTT	22	CCATGGGCTCTGGCTTTCTA	82	198
LDLR	LDLR_t18_1	CgTGTTTCCTGAaTGCTGGACTG	23	CAGGCAATGCTTTGGTCTTCTCT	83	185
LDLR	LDLR_t2_1	GTCAGTTTCTGAFTCTGGCGTTG	24	CTGGGACTCATCAGAGCCATC	84	167
LDLR	LDLR_t2_2	CTCCTTTTCCCTCTCTCAGTGG	25	ataaaTGCATATCATGCCAAAGG	85	175
LDLR	LDLR_t3_1	gccTCAGTGGGTCTTTCCCTT	26	ACACFTACgACAGCCTTGCTC	86	183
LDLR	LDLR_t3_2	CTGTAGTGTCTGTACCTGCAAA	27	CgTAGAGACAAAGTCAGACCACTC	87	199
LDLR	LDLR_t4_1	GGTGtTGGGAGACTTCACACG	28	GCTGGCGGGACACAG	88	197
LDLR	LDLR_t4_2	GAAGTGCATCTCTCGGCAGTTC	29	CCACTCATCCGAGCCATCTTC	89	184
LDLR	LDLR_t4_3	TTCCAGTGCAACAGCTCC	30	GTGGATGCACTCGCCACTTA	90	177
LDLR	LDLR_t4_4	GTAGGGGTCTTTACGTGTTCCAA	31	GAGCCCAGGGACAGGTGATA	91	198
LDLR	LDLR_t5_1	CCCTGCTTCTTTTCTCTGGTTG	32	CATTAACGCAGCCAACFTCATCG	92	188
LDLR	LDLR_t5_2	ACACACTCTGTCTCTGTTTCCA	33	CAGCAAGGCACAGAGAATGGG	93	196
LDLR	LDLR_t6_1	TGAATGAGTGCCAAGCAAACTGA	34	cgCACTCTTTGATGGGTTCATCT	94	194
LDLR	LDLR_t6_2	GAGGGACCCAACAAGTTCAAGTG	35	CACTCATGTCTCAGTCCCTTTCC	95	178
LDLR	LDLR_t7_1	CGGCGAAGGGATGGGTAG	36	CCCACCcGGaAATCACCTTC	96	196
LDLR	LDLR_t7_1	CCAACGAATGCTTGGACAACAAC	37	gttGGTTGCCATGTCAGGAAG	97	179
LDLR	LDLR_t8_1	TTCGAAGGTGTgGTTTTGGC	38	CCTTCCTCACACTGGCACTT	98	178
LDLR	LDLR_t8_2	AGCCTCTTCTCTCtCTTCCAGAT	39	GCCTGCAAGGGGTGAGG	99	196
LDLR	LDLR_t9_1	GGCACTCTTGGTTCCATCGAC	40	TTCTATTGCTGGCCACCTCC	100	198

LDLR	LDLR_t9_2	CAGGAAGATGACGCTGGAC	41	gCCAGGAGCCCCTCATCT	101	188
PCSK9	PCSK9_t2_2	TACGTGGTGGTGTGAAGGAG	42	GAAGTGCCATTCCCAAAAAGGG	102	198
PCSK9	PCSK9_t4_2	GCAGCCTGGTGGAGGTGTAT	43	ACCTCCAGGATGGGGATATGG	103	185
PCSK9	PCSK9_t7_2	GTGACCCCTGGGACTTTGG	44	CCAAAGGGGCTGTTAGCATCA	104	197
PCSK9	PCSK9_t9_2	GGCCCTTTTTCAGAGTTGG	45	acacgAAGGAGGGGTACA	105	186
LDLR	LDLR_t1_22s_1	CCCCCTGCTAGAAACCTCACATT	47	CGGTCCAGCGCAATTTCACAG	107	171
LDLR	LDLR_t1_22s_3	GCTGGAAATTGGGCTGGAC	48	CgTGCCATTACCCACAAAGT	108	176
APOE	APOE_t3_4	GCCAGAGCACCGAGGaG	2	CGCGGCCCGTTCAC	62	189
SLCO1B1	SLCO1B1_ex6_11	AAAATGAAACACTCTCTATCTACATAGGTT	46	AGAATGTCTTCTTTAGCGAAATCATC	106	162
APOE	APOE_a6_2	GCGGACATGGAGGACGTG	49	GTCATCGGCATCGCGGAG	109	147
APOE	APOE_a6_1	GGTGGGGAGGAGACG	50	AGCTCTCGGTGCTCTGG	110	147
ST3GAL4	rs11220462	GCCTCTTCCACTCTGTTC	51	CAACTCCACACACCCCAACAC	111	200
APOB	rs1367117	TGTACAACATGACTTACCTGGACA	52	TCAATGCTCTGCTACCCCTGA	112	191
SLC22A1	rs1564348	ACCAATTCCATAAATAGTAGCCACA	53	TTTGGGAGGATTTCGGTGA	113	180
HFE	rs1800562	CAGGGCTGGATAACCTTGG	54	AGCAGATCCTCATCTCACTGC	114	196
PCSK9	rs2479409	CCTCCTGCCTGGTACACAAT	55	CCAGCCTACATGCATTTCAA	115	197
MYLIP	rs3757354	TTTACTCGGTCAGCTTTAGCAA	56	AAGTTCAAGTGCCTTCACTGCT	116	180
ABCG8	rs4299376	CAAGCCAAATGAGCCCTTCTC	57	ATCAGATGTTTACTATGTGCAAGACA	117	194
CELSR2	rs629301	GCAATTCTTGCAAAAGGGTTA	58	TGCTGTACCCCGCACAGAAC	118	206

LDLR	rs6511720	TTGCAAACATTCTGCCACTC	59	CTCCTTGACAACAGAGGAAACA	119	193
NYNRIN	rs8017377	CCCTTCCTCTTCCATCTGT	60	TCCFCATCCFGCCTCTCCT	120	192

The primers of SEQ ID NO 1 to SEQ ID NO 46 and SEQ ID NO 61 to SEQ ID NO 106 make up a set of primers for the sequencing of the coding regions of the currently
5 screened genes LDLR, PCSK9 and APOB. That set of primers however does not cover certain genomic regions that are important in determining polygenic FH and/or in outcome of statin treatment. Primers of sequences SEQ ID NO 47 to SEQ ID NO 60 and SEQ ID NO 107 to SEQ ID NO 120 provide coverage of these genomic regions, and are specifically taught to achieve improved and/or complete coverage of the clinically
10 important regions in FH.

Primers SEQ ID NO 1 to SEQ ID NO 120 have been specifically designed to allow simultaneous PCR amplification and subsequent sequencing to achieve a greater than 4-fold reduction in time required for FH genetic testing.

15

It will be apparent to the skilled reader that individual primer pairs from Table 3 (or subsets of pairs taken from Table 3) also find utility in examination of individual loci of interest (or several loci of interest) independently of the simultaneous (multiplexed) use of the complete set presented in Table 3

20

MONOGENIC FH

The genes prominent in monogenic FH are discussed in turn below.

All sequences are human sequences unless otherwise stated.

25 Unless otherwise stated, reference sequences described herein are as deposited in the Gene Sequences Genome Assembly, release number (assembly number): GRCh37.p10, Feb 2009 (database version 71.37 - Feb 2013)

http://www.ensembl.org/Homo_sapiens/Info/Annotation#assembly.

30 **LDLR**

The *LDLR* gene encodes low density lipoprotein receptor protein, which binds low density lipoprotein (LDL) particles at the cell membrane and internalises them for hepatic processing and excretion. *LDLR* spans nearly 45kb and is made up of 18 exons, with pathogenic mutations found throughout the gene, some of which have been found
35 in many unrelated individuals with FH, while many others are rare. The intronic regions of *LDLR* contain multiple Alu repeat sequences, which can recombine incorrectly leading to large FH-causing indels which encompass one or many exons.

Patients with homozygous or compound heterozygous mutations in *LDLR* have a severe phenotype with onset of cardiovascular disease in childhood requiring aggressive early treatment of lipid levels. The gene is located on chromosome 19 between base pairs 11,200,225 and 11,241,992 on forward strand.

5

PCSK9

Autosomal dominant gain-of function dominant mutations in *PCSK9* (proprotein convertase subtilin-kexin-type 9) can cause a severe form of FH; the PCSK9 protein is thought to bind LDLR and promote its degradation, although the details of its mechanism of action are still not fully understood (Soutar, 2011). The identification of this gene and the realisation that loss-of-function variants in *PCSK9* lower circulating lipid levels have caused great interest in this protein as a target for alternative lipid-lowering therapies. The gene is located on chromosome 1 between base pairs 55,505,221-55,530,525 on forward strand.

15

APOB

Two pathogenic mutations in *APOB* (apolipoprotein B) cause the autosomal dominant condition familial defective apolipoprotein-B100 (FDB). The defective APOB ligand has reduced affinity for LDLR, reducing LDL clearance and causing hypercholesterolaemia with a similar phenotype to *LDLR* mutations. *APOB* is a highly polymorphic gene, complicating identification of pathogenic variants. The gene is located on chromosome 2 between base pairs 21,224,301-21,266,945 on reverse strand.

Regarding monogenic FH, the mutations discussed herein are exemplary. For example there are more than 1000 different mutations reported in LDLR. Many of them are only found in one or two families. It is very likely that more mutations may be employed. See (https://grenada.lumc.nl/LOVD2/UCL-Heart/home.php?select_db=LDLR).

STATIN TOXICITY/SLCO1B1

The rs4149056 variant in *SLCO1B1* (solute carrier organic anion transporter family member 1B1) is associated with statin-induced myopathy, particularly in patients on high dose treatment (The Search Collaborative Group, 2008).

It is important to note that due to linkage disequilibrium between SNPs one can get the same answer by genotyping different SNPs. For example rs4363657 and rs4149056 give almost identical results. Thus it should be noted that rs4363657 may be used in place of rs4149056 if convenient - both SNPs give almost identical answers.

Most suitably the rs4149056 variant in *SLCO1B1* is assessed.

Statin drugs (HMGCoA reductase inhibitors) are the mainstay of treatment for FH, and muscular side-effects ranging from muscle pain through to life-threatening rhabdomyolysis are among the principle reasons for discontinuation of treatment.

SLCO1B1 encodes the organic anion--transporting polypeptide OATP1B1, which mediates hepatic uptake of many drugs including most statins. It has been suggested that patients being treated with statins should be routinely genotyped for this variant and that an upper limit on statin doses should be applied according to genotype (Niemi, 2009). The rs4149056 variant is located on chromosome 12 at position 21,368,722 on forward strand.

LDLR MUTATIONS

As will be apparent to the skilled worker, assessing the mutation status and/or sequences of the various genes or SNPs as taught herein is a matter of routine. The bioinformatics required for this analysis are modest, and are routine in the art.

Occasionally, in particular with the analysis of LDL receptor (LDLR) mutations, it may be necessary for the skilled worker operating the invention to pay attention to the mutation which they have detected. A first step in analysing an LDLR mutation would be to check if it has been reported in the literature or is a novel variant detected for the first time. Clearly if the mutation has been reported before in the literature, then it is a simple matter of routine for the skilled operator to consult the literature to know if the mutation is associated with FH or not. If the variant is new, then a judgement may need to be made as to whether or not it is likely to be a mutation predisposing to FH. In case further guidance is required, novel variants that cause premature stop codon, frameshift or overlap with splice donor or acceptor sites are expected to be mutations causing FH. Non-synonymous substitutions, in-frame deletions or insertions and variants that overlap splice regions may or may not be causal. These variants should be accessed for their pathogenicity using algorithms that predict possible impact of such variants on protein function. Most commonly used pathogenicity predictive web-sites include SIFT and PolyPhen (Kumar et al., 2009; Adzhubei et al., 2010). Where possible, variants with unclear pathogenicity should also be followed up by segregation analyses and in vitro functional studies. Functional studies include LDL binding and/or uptake assays (Soutar et. al., 1982; Goldstein et. al., 1972).

POLYGENIC FH

Currently 12 single nucleotide polymorphisms, SNPs, (Table A) located in 11 genes should be genotyped in order to establish the risk of polygenic FH. Identifying such individuals will improve the efficiency of cascade testing, since this type of test will only benefit patients with monogenic FH. Much less than the expected half of children of individuals with polygenic FH are expected to inherit the disease.

Table A: Single nucleotide polymorphisms identified as risk indicators for polygenic FH

SNP	Gene	Description
rs4299376	ABCG8	ATP-binding cassette, sub-family G (WHITE), member 8
rs1367117	APOB	apolipoprotein B (including Ag(x) antigen)
rs429358	APOE	apolipoprotein E
rs7412	APOE	apolipoprotein E
rs629301	CELSR2	cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)
rs1800562	HFE	hemochromatosis
rs6511720	LDLR	low density lipoprotein receptor
rs3757354	MYLIP	myosin regulatory light chain interacting protein
rs8017377	NYNRIN	NYN domain and retroviral integrase containing
rs2479409	PCSK9	proprotein convertase subtilisin/kexin type 9
rs1564348	SLC22A1	solute carrier family 22 (organic cation transporter), member 1
rs11220462	ST3GAL4	ST3 beta-galactoside alpha-2,3-sialyltransferase 4

The SNPs used to determine predisposition to polygenic FH and their genomic locations are listed in Table B.

Table B: The position of single nucleotide polymorphisms identified as risk indicators for polygenic FH

SNP	Gene	Chromosome	Position on chromosome
<u>rs11220462</u>	ST3GAL4	<u>11</u>	<u>126243952</u>
<u>rs1367117</u>	APOB	<u>2</u>	<u>21263900</u>
<u>rs1564348</u>	SLC22A1	<u>6</u>	<u>160578860</u>
<u>rs1800562</u>	HFE	<u>6</u>	<u>26093141</u>
<u>rs2479409</u>	PCSK9	<u>1</u>	<u>55504650</u>
<u>rs3757354</u>	MYLIP	<u>6</u>	<u>16127407</u>
<u>rs429358</u>	APOE	<u>19</u>	<u>45411941</u>

<u>rs4299376</u>	ABCG8	<u>2</u>	<u>44072576</u>
<u>rs629301</u>	CELSR2	<u>1</u>	<u>109818306</u>
<u>rs6511720</u>	LDLR	<u>19</u>	<u>11202306</u>
<u>rs7412</u>	APOE	<u>19</u>	<u>45412079</u>
<u>rs8017377</u>	NYNRIN	<u>14</u>	<u>24883887</u>

It will be appreciated that the SNPs in Table A and Table B are the same 12 SNPs, the tables present different information in connection with them. Suitably one or more of these SNPs are assayed; suitably two or more; suitably three or more; suitably four or
5 more; suitably five or more; suitably six or more; suitably eight or more; suitably ten or more; suitably all twelve SNPs of Table A or Table B are assayed.

STATINS

10 Approximately 99.9% of patients who receive a diagnosis of familial hypercholesterolaemia are prescribed statins. They are typically then sent away with their prescription. At most the patient may be advised to return if they experience side effects such as muscle pain etc. Patients are not automatically given a statin toxicity test. However, statin toxicity is a serious problem. A certain proportion of the
15 population carries a genetic predisposition to statin toxicity. Symptoms of statin toxicity include muscle weakness, pain such as muscle pain, death of muscle fibres, and kidney trauma which can lead to kidney failure from decomposing muscle fibre. Up to 5% of patients exhibit some form of statin toxicity. As many as 1 in 1000 patients can experience the most severe statin toxicity including loss of muscle fibre and kidney
20 trauma or kidney failure.

The inventors teach the parallel assay of the statin toxicity predisposition locus SLCO1B1 at the same time as collecting the information useful in aiding the diagnosis of FH together with the incremental risk factors. As many as 15-20% of the population
25 carry an SLCO1B1 allele which might predispose them to statin toxicity. A proportion of these people will actually go on to experience statin toxicity.

Currently in the art, typically no assessment of SLCO1B1 is made. The current thinking in the art is that it does not justify the cost of a separate laborious test for the SLCO1B1
30 locus.

The inventors teach incorporation of the SLCO1B1 assay in to the multiplexed PCR and sequencing approach as set out herein. By incorporating this test into the main information collection procedure, the results for the SLCO1B1 locus are obtained at a minimal or negligible extra cost (at most the cost of the extra primer pair included in the assay). By contrast, the advantages reaped in information for this miniscule extra investment are enormous. Firstly, the cost of a further separate genetic test for SLCO1B1 is avoided. More importantly, there is additional value to the physician who is better informed when designing the treatment for the patient under examination.

Patients presenting and undergoing the combined tests of the invention showing a predisposition to statin toxicity can be counselled in great detail about the side effects to be alert to. In addition, the patient can be prescribed a re-examination at early time points. The patients can be placed under a more frequent surveillance regime. The patient can be prescribed smaller doses of the statin. It is even possible to design a lower dose/better tolerated statin treatment regime based on their SLCO1B1 status. For example, subjects who are heterozygous for an SLCO1B1 allele predisposing to statin toxicity show approximately a three times greater risk of statin toxicity than a patient with no risk allele. Furthermore, a subject who is homozygous for an SLCO1B1 allele predisposing to statin toxicity has a ten times greater risk of statin toxicity.

Thus, in one embodiment, it is possible to design a statin dose or treatment regime according to the number of copies of an SLCO1B1 allele predisposing to statin toxicity which the patient is carrying. For example, an upper limit on statin doses may be set for subjects in whom a risk of statin toxicity is detected according to the present invention. For example, a particular statin dose may be set for subjects in whom a risk of statin toxicity is detected according to the present invention. For example, a particular statin may be selected for subjects in whom a particular SLCO1B1 mutation is detected. Guidance on statin choice, dose choice or dose limits may be found in (Niemi, 2009 - Clinical Pharmacology & Therapeutics (2010) 87 1, 130–133), which is incorporated herein by reference expressly for the teachings of alternate statin regimes according to SLCO1B1 genotype, as well as Wilke et al 2012, Clinical Pharmacology and Therapeutics 92:112-7.

Thus it can be appreciated that there are many important advantages provided by collating all of the information including SLCO1B1 together at the outset according to the methods of the invention.

It is a problem in the art that a large proportion of patients (millions of patients per annum) are currently treated in the primary care setting by their physician or general practitioner. In this setting, the cost of routinely conducting statin toxicity tests on each patient presenting with hypercholesterolaemia is a significant barrier. This problem is alleviated by the present invention, which provides a combined comprehensive information set at a fraction of the cost of the current standard analytical method.

FURTHER EMBODIMENTS AND APPLICATIONS

- In one embodiment the invention relates to a method for determining whether a subject has a predisposition for monogenic familial hypercholesterolaemia by virtue of a mutation in one of the genes selected from LDLR, PCSK9, APOB or polygenic familial hypercholesterolaemia by virtue of combination of genotypes in (rs6511720) LDLR, (rs2479409) PCSK9, (rs1367117) APOB, (rs629301) CELSR2, (rs4299376) ABCG8, (rs1564348) SLC22A1, (rs1800562) HFE, (rs3757354) MYLIP, (rs11220462) ST3GAL4, (rs8017377) NYNRIN, (rs429358, rs7412) APOE and for determining individual's response to statin treatment by virtue of a variant (rs4149056) in SLCO1B1 gene comprising:
- carrying out a nucleic acid amplification step on a biological sample from the subject, wherein the nucleic acid amplification step uses all primer pairs, the primer pairs comprising a forward primer with a forward primer sequence selected from SEQ ID NO: 1 to SEQ ID NO:60; or a reverse primer with a reverse primer sequence selected from SEQ ID NO: 61 to SEQ ID NO:120;
 - sequencing at least part of the amplification product, the amplification product being a sequence of at least a part of one of the genes LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLCO1B1;
 - and
 - determining whether the LDLR, PCSK9 or APOB amplification product represents a sequence that has a mutation that is indicative of a risk for a familial hypercholesterolaemia
 - determining whether combination of LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE amplification products have SNPs that are indicative of a risk for a polygenic familial hypercholesterolaemia
 - determining whether the SLCO1B1 amplification product represents a sequence that has a SNP that is indicative of a risk for adverse statin effect.

Suitably the amplification step uses at least one primer pair, the pair comprising a forward primer with a forward primer sequence selected from SEQ ID NO: 1 to SEQ ID NO: 60; and its accompanying reverse primer, with appropriate sequence selected from SEQ ID NO: 61 to SEQ ID NO: 120.

- 5 The invention may also relate to assessing a subject for the diagnosis and/or treatment of familial hypercholesterolaemia (FH).

The invention may also relate to designing a treatment for familial hypercholesterolaemia (FH).

The invention may also relate to managing a subject suspected of having

- 10 familial hypercholesterolaemia (FH).

The invention may also relate to determining an appropriate programme for the diagnosis and/or treatment of familial hypercholesterolaemia (FH).

- 15 A method for collecting information useful in aiding the diagnosis and treatment of familial hypercholesterolaemia (FH), the method comprising providing a sample of DNA from the subject, assaying said DNA for

(a) mutation in one or more genes which is indicative of FH diagnosis and

(b) mutation in a gene which is indicative of likelihood of statin toxicity;

- 20 wherein assaying said DNA comprises the step of contacting said DNA sample with one or more primers for amplification and/or sequencing of the relevant segment(s) of said DNA

characterised in that each of said assays of steps (a) and (b) is carried out in a single assay cycle.

- 25 A method for collecting information useful in aiding the diagnosis and treatment of familial hypercholesterolaemia (FH), the method comprising providing a sample of DNA from the subject, assaying said DNA for

(1) presence of one or more SNPs indicative of incremental risk of high cholesterol; and

- 30 (2) mutation in a gene which is indicative of likelihood of statin toxicity;

wherein assaying said DNA comprises the step of contacting said DNA sample with one or more primers for amplification and/or sequencing of the relevant segment(s) of said DNA

characterised in that each of said assays of steps (1) and (2) is carried out in a single assay cycle.

Further particular and preferred aspects are set out in the accompanying independent and dependent claims. Features of the dependent claims may be combined with features of the independent claims as appropriate, and in combinations other than those explicitly set out in the claims.

Where an apparatus feature is described as being operable to provide a function, it will be appreciated that this includes an apparatus feature which provides that function or which is adapted or configured to provide that function.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the present invention will now be described further, with reference to the accompanying drawings, in which:

Figure 1 shows the depth of coverage of primer pairs (SEQ ID NO: 1 – 120), as shown in Table 3, in an *in vitro* experiment using a sample of human DNA.

Figure 2 shows the quality of sequence data shown in Figure 1. 96.1% of bases passed standard quality filter ($\geq Q30$).

Figure 3 shows the time improvements of the current test compared to standard methods.

Figure 4 shows Figure S1. Family pedigree showing affected individuals as filled black boxes/circles and unaffected individuals as unfilled boxes/circles. The arrow indicates the proband. Total cholesterol (TC) levels (mmol/l) are shown for each individual as well as mutation status. WT, wild type; NA, not available;

Examples of the invention are disclosed in detail herein, with reference to the accompanying drawings. It is understood that the invention is not limited to the precise embodiments described in the examples and that various changes and modifications can be effected therein by one skilled in the art without departing from the scope of the invention as defined by the appended claims and their equivalents.

EXAMPLES**Example 1: Standard Procedure for Target Enrichment**

- 5 Target enrichment of the genes involved in FH is undertaken using the micro-fluidic PCR amplification 48.480 Access Array IFC (Fluidigm Inc).

Equipment:

- Micropipettes capable of dispensing volumes in the range of 1-1000µl, with
 10 compatible sterile filtered tips.
 Nitrile gloves and laboratory coat dedicated for use in each area
 Assorted sterile racks
 Sterile 1.5 ml micro-tubes
 IFC Controller AX, pre- and Post-PCR units (Fluidigm)
 15 IFC thermal cycler (Fluidigm)
 48.48 Access Array chip (Fluidigm)
 0.2ml thin walled micro-reaction PCR plates
 Large bench-top centrifuge with plate rotors, microcentrifuge and vortex mixer
 Yellow discard bins and fume cupboard
 20 Adhesive plate sealer
 Ice bath
 Class II biosafety air-flow hood

Reagents:

- 25 Sterile, nuclease-free water
 5x primer stock multiplex plates (Fluidigm)
 Control line fluid (Fluidigm)
 1x Access array harvest solution (Fluidigm)
 20x Access array loading reagent (Fluidigm)
 30 FastStart High fidelity PCR system with dNTPs kit (Roche)
 Access Array oligonucleotide multiplex primer plates (Fluidigm)
 Oligonucleotide primers, individual tubes (Sigma)

Procedures:

35

Primer preparation

- 1.1 Working in the Class II DNA free hood, appropriately label a fresh 0.2ml PCR plate.
- 1.2 Resuspend the primers in individual tubes to a concentration of 100mM, as per manufacturer's instructions.
- 1.3 For each primer pair, add 5µl from each tube to a single well of a fresh 0.2ml PCR plate. This will give 10µl of 100µM.
- 1.4 Add 10µl molecular biology grade water to wells containing the primer pairs. This will give 50µM stock of each primer pair.
- 1.5 To prepare the 20x primer solutions. In a fresh 0.2ml PCR plate, add 20µl of the primer stock to each well, along with 5µl 20x Access Array loading buffer and 75µl water.
- 1.6 Seal, vortex and briefly centrifuge. Label all plates with the date, assay name and concentration.
- 1.7 Primer solutions are valid for use for one year after resuspension, stored at -20°C

Access Array preparation:

- 1.8 Remove an Access Array chip from its packaging.
- 1.9 Inject control line fluid into each of the ports on either side of the chip. Tilt the chip to 45° whilst injecting the chip. Ensure the port is firmly pushed in.
- 1.10 Add 500 µl of 1x Access array harvest solution to wells H1-H4.
- 1.11 Place the chip into the Pre-PCR IFC AX controller. Run Prime (151x)

- 1.12 Into a fresh 1.5ml microtube, make the amplification mastermix of (per 60 reactions) 30 μ l FastStart 10x reaction buffer, 54 μ l 25mM MgCl₂, 15 μ l DMSO, 6 μ l 10mM dNTPs, 3 μ l FastStart enzyme blend, 15 μ l 20x Access array loading reagent and 57 μ l Nuclease free water, in the Class II air-hood.
- 1.13 Vortex and briefly centrifuge the tube.
- 1.14 In a fresh 0.2ml PCR plate, mix 3 μ l of master mix along with 1 μ l genomic DNA (50 ng/ μ l) and 1 μ l water into 48 separate wells.
- 1.15 Pipette 4 μ l 20x primer solution into the primer inlets.
- 1.16 Pipette 4 μ l sample/master mix into the sample inlets. One appropriate positive and one negative control (water) should be included in each run.
- 1.17 Load the chip into the Pre-PCR AX controller. Run Load mix (151x):

PCR protocol:

Temp	Time	Cycles
50°C	2 min	1
70°C	20 min	1
95°C	10 min	1
95°C	15 sec	10
60°C	30 sec	
72 °C	1 min	
95°C	15 sec	2
80°C	30 sec	
60°C	30 sec	
72°C	1 min	
95°C	15 sec	8
60°C	30 sec	
72°C	1 min	
95°C	15 sec	2
80°C	30 sec	
60°C	30 sec	
		36

72°C	1 min	
95°C	15 sec	} 8
60°C	30 sec	
72°C	1 min	
95°C	1 min	} 5
80°C	30 sec	
60°C	30 sec	
72°C	1 min	

- 1.18 After the PCR has completed, remove the plate and move to the post-PCR area.
- 1.19 Remove the remaining fluid in H1-4 and discard.
- 5 1.20 Pipette 600 µl of 1x Access Array Harvest Reagent into wells H1-4 and 2 µl into 1-48 sample inlets.
- 1.21 Place chip into the post-PCR IFC AX and run programme “Harvest (151x)”.
- 1.22 Once complete, remove chip from IFC AX and transfer
10 the harvest volume (approx. 10-15 µl) into columns 1-6 of a fresh 0.2ml PCR plate.
- 1.23 Perform QC by running the harvests on the Bioanalyzer.

Example 2: Target enrichment of sample DNAs

Overall the region of interest (ROI) in the study covered 8497 base pairs from the
15 LDLR, PCSK9, APOB, CELSR2, ABCG8, SLC22A1, HFE, MYLIP, ST3GAL4, NYNRIN, APOE and SLC01B1 genes. Genomic DNA templates were amplified using the Access Array IFC, according to the manufacturer's instructions (<http://www.fluidigram.com>).

Briefly, the technique employed a microfluidic chip that systematically combined 48
sample DNAs, along with 120 primer pairs (SEQ ID NOs 1 - 120). PCR reagents were
20 then drawn into the chip reaction chambers prior to PCR cycling. Common flanking sequences on each primer pair permit attachment of barcode indexes and sequencing adaptors. This allows for the attachment of platform-specific adaptors and barcodes pre target enrichment. The multiplex amplicon harvests were recovered for each DNA template, approximately 15µl was used to generate barcode and adapter products

specific to the MiSeq platform (Illumina, San Diego, CA). Amplicon harvest volume was adjusted to 20µl using PCR certified water. For each harvest, barcode-fusion PCR reactions were prepared using CS tagged primer pairs, for instance, a reaction with A_BC6_CS1 and CS2_P1. This strategy permitted amplicon sequencing in both

5 orientations. For each reaction, 10 µl of the Fluidigm IFC harvest was added to 86 µl of a Herculase II Fusion PCR mix, as per manufacturer's instructions (Agilent Technologies Inc, Santa Clara, CA) along with 20 pmol each primer. Cycling conditions were as follows: an initial incubation of 98°C for 30 sec, followed by two cycles of 98°C for 30 sec, 54°C for 30 sec and 72°C for 30 sec; after the last cycle reactions were held

10 at 72°C for 2 min, and then at 4°C.

Example 3: Illumina sequencing

Three microliters of 2nM Fluidigm library were denatured by adding 3µl of 0.1N NaOH and incubated for 5min at room temperature. Denatured library was then diluted to

15 20pM by adding 294µl of HT1 buffer (Illumina MiSeq sequencing kit). The library was further diluted to 6pM in a 1ml final volume using HT1 buffer prior to loading into the Miseq sequencing cartridge (sample position #17).

Fluidigm sequencing primers FL1 and FL2 (500µM stock) were diluted to 0.5µM final in a 200µl volume and spiked into the appropriate Miseq cartridge positions ie FL1 in

20 positions #12 for Read1 and #14 for Read2, FL2 in position #13 for Index Read. The MiSeq cartridge was loaded into Miseq and the run was set up following Illumina MiSeq sequencing standard procedures.

Example 4: The use of next generation sequencing in clinical diagnosis of

25 **familial hypercholesterolaemia**

Summary:

Familial hypercholesterolaemia (FH) is a common Mendelian disorder associated with early onset coronary heart disease that can be treated by cholesterol lowering drugs.

30 The majority of cases in the UK are currently without a molecular diagnosis partly due to the cost and time associated with standard screening techniques. This study tested the sensitivity and specificity of two next-generation sequencing (NGS) protocols for genetic diagnosis of FH.

Methods:

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

5 Results:

In the validation cohort, both protocols showed 100% specificity while 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
10 detection rate using the Access Array was highest in patients with clinically definite FH (67%) followed by patients with possible FH (26%).

Conclusions:

We have shown the potential of target enrichment methods combined with NGS for molecular diagnosis of FH. Adopting these assays for patients with suspected FH could
15 improve cost-effectiveness and increase the overall number of patients with a molecular diagnosis.

In this study, two target enrichment protocols, the hybridisation-based SureSelect Target Enrichment System (Agilent, Santa Clara, CA, USA) and the PCR-based Access
20 Array System platform (Fluidigm, South San Francisco, CA, USA), followed by NGS sequencing, have been validated for 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.

25 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 was analysed. 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
5 used to detect large deletions and duplications in a subset (>80%) of the patients in the validation cohort, as described in Tosi *et al.* ⁹ Forty-five samples were heterozygous and were known to carry FH-causing point mutations or small insertions/deletions (less than 20 base pairs) in *LDLR*, *PCSK9* or *APOB* and six had large insertions/deletions in *LDLR*. One sample was homozygous for *LDLR* p.Gln384Pro
10 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,
15 84 consecutive patients referred for molecular testing by the Hammersmith Hospital lipid clinic over a period of one year were studied by the PCR-based Access Array System microfluidic platform. Fifteen of the 84 samples were also analysed 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
20 had a diagnosis of definite FH based on Simon Broome criteria. The median highest total cholesterol in this group was 9.7 mmol/l (min 7.9 mmol/l, max 15.9 mmol/l). Sixty five individuals had a diagnosis of possible FH with a median highest total cholesterol levels 8.7 mmol/l (min 5.2 mmol/l, max 13 mmol/l) and thirteen patients did not fulfil Simon Broome criteria. The median highest total cholesterol in this group
25 was 8.3 mmol/l (min 6.3 mmol/l, max 9.6 mmol/l).
The study was approved under ethics committee references: REC2002/6451 and REC11/LO/o883. 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, USA) or from saliva samples using the Oragene (Genotek Inc, Ottawa, ON, Canada) protocol. Both protocols followed the
5 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, USA) and extracted using the Paxgene blood RNA kit
10 (Qiagen, Hilden, Germany) following the manufacturer's protocol. Reverse transcription was performed using the iScript cDNA synthesis kit (BioRad, Hercules, CA, USA). Primers (forward TCGAGTTCCACTGCCTAAGTG and reverse GTTGTGTTGTCCAAGCATTCGTT) were designed to amplify exons 4 to 7 of *LDLR*.

15 Custom SureSelect Target Enrichment System

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

The genomic co-ordinates of the 18 targeted genes were determined using the March
25 2006 build (NCBI36/hg18) of the human genome in the Ensembl genome browser¹². The density of bait tiling was 5-fold and the baits were allowed to overlap into repeat regions by 30bp. The total targeted DNA length was 399kb. All libraries were generated from sheared DNA (Covaris, Woburn, MA, USA) with an average insert size

of 200bp following the SureSelect Target Enrichment System XT (Agilent, Santa Clara, CA, USA) protocol for Illumina multiplexed sequencing version 1.2. After dilution to 2nM, up to 30 libraries were pooled and sequenced on one lane on the HiSeq2000 platform (Illumina, San Diego, CA, USA) to generate 2x100bp paired-end reads.

5

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 184bp with 57% GC content. Primer sequences are shown in Table S1. Samples were processed using the Access Array System (Fluidigm, South San Francisco, CA, USA) 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, San Diego, CA, USA). One amplicon, a GC rich (75%) amplicon in *APOE* targeting rs7412 failed to amplify and was excluded from subsequent data analysis.

20

Data Analysis

Sequences were mapped to the GRCh37/hg19 human reference sequence using Burrows Wheeler Aligner vo.6.1¹³. PCR duplicate reads were removed from SureSelect Target Enrichment System data (Picard tools v1.35). Sequence reads from both datasets were further processed and variants were called using GATK v1.0.6 with hard filtering options. Variant annotation was carried out with Ensembl's Variant Effect Predictor

25

tool ¹² and was based on transcripts ENST00000558518 (*LDLR*), ENST00000233242 (*APOB*), and ENST00000302118 (*PCSK9*). The annotation included Sorting Tolerant From Intolerant (SIFT), Condel and PolyPhen. Conservation scores (Genomic Evolutionary Rate Profiling, GERP) were obtained from Ensembl¹² version 68. To
5 identify potentially pathogenic mutations, all non-synonymous, splice site, frameshift and truncating mutations were examined and compared to an FH locus-specific databases ^{6,14} 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 (MAF) above 1% in the International
10 HapMap Project¹⁵ or the 1000 Genomes Project ¹⁶ 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 where possible.

15 **Copy Number Variant Analysis**

Copy number variant analysis (CNV) from NGS data was performed for the samples sequenced using targeted capture. A read-depth based method ¹⁷, 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 in order to
20 increase the quality of a reference set for each sample and therefore to maximise the power to detect copy number variants. 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 copy number variant calls, as implemented in the method. As an independent method of CNV analysis, MLPA was performed using the kit *LDLR*-
25 Po62 (MRC-Holland) following the manufacturer's protocol. The novel exon 16 deletion was confirmed by PCR. The previously-described large deletions and duplications are detailed in 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
5 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
10 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 a 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
15 tests were performed using the DeFinetti software (<http://ihg.gsf.de/cgi-bin/hw/hwa1.pl>).

RESULTS

Validation Study

20 Custom SureSelect Target Enrichment System

To validate the assay, DNA was analysed 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 827x. Overall 64% of total mapped reads aligned within the
25 target region and 99.8% and 98.8% of nucleotides were covered at 4x and 25x, respectively. The insufficiently covered regions are consistent among runs and were found mostly outside coding regions. Hybridisation-based capture is known to target GC-rich regions poorly; however, all regions containing known FH-causing mutations

were covered sufficiently (>25x) for confident variant calling. The initial analysis of the sequencing results was carried out blinded to the gene and mutation details for each sample. All twenty heterozygous and one homozygous short pathogenic mutations, including point mutations and insertions/deletions of less than 15 base pairs, were
5 detected (Table 1). In addition one 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
10 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, GERP score 3.29, located in the highly-conserved footprint 1 (FP1) site) and two rare variants were identified in *APOB* in patients without a previous molecular diagnosis. The two
15 variants, p.Pro877Leu (rs12714097) and p. Asp2213_Glu 2214delinsGlu (rs72653087), have not been described previously in FH patients and are not reported in either the 1000 Genomes Project¹⁶ or the International HapMap Project¹⁵ databases. No other rare missense variants were detected in the known FH genes in individuals without a molecular diagnosis.

PCR-based Access Array System

20 DNA samples from 53 previously characterized patients including 40 with point mutations (39 heterozygotes and one homozygote), 6 with insertions/deletions (all heterozygotes), 6 with large deletions or duplications (all heterozygotes), and one compound heterozygote with one missense mutation and one large deletion in *LDLR* (Table 1) were amplified together with 22 patients that had previously screened
25 negative. The median coverage per sample was 572x (min 461, max 625). All amplicons except *APOE* (see Methods) amplified with a mean coverage of 506x. The coverage for individual amplicons is listed in Table S2. Overall, 90% of bases were covered more than 25-fold. *LDLR* had a mean coverage 656x and more than 98% of bases were above

25x. 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
5 detected. This 11bp deletion (p.Lys730Hisfs*48) was not present in any aligned reads as the deletion overlapped with a forward primer, which prevented amplification of the mutated allele. The sensitivity for short variant detection was 98% (47/48) and specificity 100%. Similarly to the SureSelect Target Enrichment System results one mutation, the p.Asp227Glu, was identified in a sample that previously had no
10 molecular diagnosis. No pathogenic mutations were found in 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 compared to SureSelect Target Enrichment System was therefore 82% (47/57).

15 **Prospective Cohort**

To test the feasibility of an NGS-based mutation screen in a clinical setting, a consecutive cohort of 84 unrelated patients referred for genetic screening over a single year was analysed. Sixty-nine of these patients were screened using the Access Array System assay alone and fifteen were processed using both protocols. Sequencing and
20 coverage metrics are provided in Table S2. 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).
25 There was a good correlation between the results obtained from the Access Array System assay and SureSelect Target Enrichment System with 5 of 6 mutations concordant between the datasets. 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,
 5 p.Ala521Thr. This variant is conserved (GERP score 2.66), is predicted to be damaging (PolyPhen score 0.67, SIFT score 0.01, and Condel score 0.643) and segregates with affected status in the family (Figure S1a). 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 hypercholesterolaemia in this family
 10 (Figure S1b).

Two (c.817+1G>A and c.941-4G>A) of the four splice site variants identified (Table 2) were previously detected in FH patients⁶, but their pathogenicity had not been experimentally validated before. An RNA sample was available for one of them (patient ID 35) and RT-PCR in this patient showed that the c.817+1G>A variant disrupted
 15 correct splicing (data not shown). Alternative splice products were generated by partial intron retention that led to truncation of the protein, and by exon skipping that caused the loss of exon 5 sequence and 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
 20 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) of the proband had inherited the p.Val827Ile variant.

25 Mutation Detection Rates

The prospective cohort included six patients with definite FH as defined by the Simon Broome criteria, 65 with possible FH and 13 hypercholesterolaemic 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 (4/6, 67%) followed by the group with a diagnosis of possible FH (17/65, 26%). One mutation (1/13, 8%) was identified amongst the 13 hypercholesterolaemic patients that did not fulfill FH diagnostic criteria.

5

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 hypercholesterolaemic individuals who screened negative for mutations in *LDLR*, *PCSK9* and *APOB*, are potentially responsible for patients' raised cholesterol. Rare variants (MAF<0.01) identified in such individuals are listed in Table S3. In patients with no previously known molecular diagnosis, seven rare non-synonymous variants were found, of which six were not predicted to be functionally significant by SIFT 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 as it did not segregate with the phenotype in the family (data not shown).

10

15

***SLCO1B1* Myopathy-associated Variant**

Data on exposure to statins was 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 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 (p=0.002, 1.58-9.89).

20

25

DISCUSSION

We demonstrated the sensitivity and specificity of two target enrichment protocols, combined with NGS, for detection of disease-causing mutations in patients with proven or suspected FH. Using the SureSelect Target Enrichment System and the HiSeq 2000
5 system, 98.8% of targeted regions were covered at more than 25x compared to the PCR-based Access Array System followed by MiSeq sequencing where 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
10 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, though large
15 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 11bp deletion (p.Lys730Hisfs*48) that overlapped a primer site. However, the recent availability of longer sequencing reads, currently up to 250bp when using MiSeq, will be anticipated in most cases to eliminate the need to have primer sites within coding regions, or allow
20 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 mutation negative. Inspection of the previous laboratory database indicated that the p.Asp227Glu variant
25 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*.¹⁸ The current standard screening is based on MLPA, a technique that is highly reliable, but 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 been investigated.^{19, 20} Here, the combination of SureSelect Target Enrichment System/HiSeq and data analysis using ExomeDepth software¹⁷ 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 as read depth in this study did not correlate with exon deletions, although it has recently been shown that amplicon multiplex PCR can be optimised for copy number variant detection.²¹

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 24Mb) while the Access Array System protocol is limited to 480 amplicons with a maximum length of 400bp 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^{19,22} with reagent costs approximately 10-fold lower than for the SureSelect Target Enrichment System protocol²² 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 compared to at least three days needed for the SureSelect Target Enrichment System in-solution capture. Initial DNA requirements of the Access Array System/MiSeq protocol are also low (50ng vs 100) and most of the process is automated, reducing the risks of contamination and human error to a minimum.

Recently Hollants *et al.*²³ 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) platform. They successfully identified all short variants but, as with our study, could not detect large rearrangements. Our Access
5 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 (Roche), but at a lower sequencing cost.²⁴ 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.²⁵

10

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 FP1 site. A luciferase assay showed that the c.-227G>T
15 variant had around 75% transcription levels compared with the wild-type site²⁶. Whilst 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
20 data. The remaining two novel variants in *APOB*, rs12714097 and rs72653087 were located outside *LDLR* binding sites, regions currently not associated with hypercholesterolaemia, and their pathogenicity therefore also remains to be elucidated.

Genotype-phenotype Correlation in the Prospective Cohort

25 In the UK, FH diagnosis is made based on the Simon Broome criteria that identify patients with definite or possible FH. Following NICE guidelines⁵, 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 one year.

To assess clinical significance, all rare variants ($MAF < 0.01$) that were identified in *LDLR*, *APOB* or *PCSK9* were compared against a locus specific FH databases^{6,14}. All

5 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 SIFT and PolyPhen to be deleterious and probably damaging respectively, and were previously identified in FH patients. Based on these criteria a missense change p.Val827Ile was included among
10 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 internalisation signal for LDLR, NPVY, a position generally not necessary for correct internalisation.²⁷ In addition, our segregation data suggest that this variant may not be disease-causing and we therefore classified this variant as being of
15 unknown significance. Of the four splice site variants, three (c.313+1G>A, c.190+4A>T, c.817+1G>A) were shown experimentally to disrupt splicing, two in published data^{28,29}, 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
20 available for confirmation within this study.

The mutation detection rates among patients with definite (66%) and possible (26%) FH are comparable to previously published studies.⁷ One mutation was identified among 13 patients that did not fulfill Simon Broome criteria. This suggests that a number of patients with high cholesterol in the general population who do not fulfill
25 formal clinical criteria for FH diagnosis may have FH-causing mutations. Such patients may therefore be diagnosed by the assays developed here and, as these mutations are dominantly inherited, family cascade screening would likely 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 ¹⁴ 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 which would be
5 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
10 by Motazacker *et al.* ³⁰ 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
15 coding regions of other medically relevant genes, such as *APOE* and *SLCO1B1* that 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 after Sanger sequencing. The SureSelect Target Enrichment System
20 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 analysis will be required. For detection of variants outside known FH-causing genes or gene regions, we focused on variants that are rare (MAF < 0.01), likely pathogenic (i.e. located in coding regions,
25 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 (less than 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.

5

Statin Induced Myopathy

Previous published data on larger cohorts show an odds ratio of 4.5 for association of rs4149056 with myopathy³¹ and on that basis, guidelines have been recommended avoiding higher doses of simvastatin in heterozygotes or homozygotes for rs4149056.³²

10 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
15 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 prior to initiation of statin treatment.

In conclusion, we have shown the utility of sequence target enrichment methods in
20 combination with NGS in molecular diagnostics of FH. Due 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 where the number of genes needing to be screened is very large. In FH diagnostics where less than five genes need to be analysed,
25 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.

REFERENCES TO EXAMPLE 4

5

1. Marks D, Thorogood M, Farrer JM, Humphries SE. Census of clinics providing specialist lipid services in the United Kingdom. *J Public Health (Oxf)*. Dec 2004;26(4):353-354.
2. Defesche JC. Familial hypercholesterolemia. In: Betteridge DJ, ed. *Lipids and Vascular Disease*. London: Martin Dunitz Ltd; 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*. Jul 15 1993;72(2):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. 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*. Jul 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*. Jun 2010;77(6):572-580.
8. Soutar AK, Naoumova RP. Mechanisms of Disease: genetic causes of familial hypercholesterolemia. *Nat Clin Pract Cardiovasc Med*. 2007;4(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(1):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(6):561-568.
11. Alonso R, Defesche JC, Tejedor D, et al. Genetic diagnosis of familial hypercholesterolemia using a DNA-array based platform. *Clinical Biochemistry*. 2009;42(9):899-903.
12. Flicek P, Amode MR, Barrell D, et al. Ensembl 2012. *Nucleic Acids Res*. Jan 2012;40(Database issue):D84-90.
13. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. Jul 15 2009;25(14):1754-1760.
14. Usifo E, Leigh SE, Whittall RA, et al. Low-density lipoprotein receptor gene familial hypercholesterolemia variant database: update and pathological assessment. *Ann Hum Genet*. Sep 2012;76(5):387-401.
15. The International HapMap Project. *Nature*. Dec 18 2003;426(6968):789-796.
- 20 16. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature*. Oct 28 2010;467(7319):1061-1073.
17. 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*. Aug 31 2012.
- 25 18. 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*. Jul 2009;76(1):69-75.

19. 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 U S A*. Jul 13 2010;107(28):12629-12633.
20. Redin C, Le Gras S, Mhamdi O, et al. Targeted high-throughput sequencing for
5 diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alstrom Syndromes. *J Med Genet*. Aug 2012;49(8):502-512.
21. 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*. Mar 2009;30(3):472-476.
- 10 22. Ware JS, John S, Roberts AM, et al. Next Generation Diagnostics in Inherited Arrhythmia Syndromes : A Comparison of Two Approaches. *J Cardiovasc Transl Res*. Sep 7 2012.
23. Hollants S, Redeker EJ, Matthijs G. Microfluidic amplification as a tool for massive parallel sequencing of the familial hypercholesterolemia genes. *Clin*
15 *Chem*. Apr 2012;58(4):717-724.
24. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. May 2012;30(5):434-439.
25. Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, Martin JF. Accuracy and
20 quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*. 2011;12:245.
26. 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 Research*. October 1, 1997 1997;25(20):4132-
25 4138.
27. 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. *Journal of Biological Chemistry*. February 25, 1990 1990;265(6):3116-3123.
28. 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(3):269-271.
29. Holla OL, 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*. Apr 2009;96(4):245-252.
30. 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*. Jun 2012;33(11):1360-1366.
31. Link E, Parish S, Armitage J, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med*. Aug 21 2008;359(8):789-799.
32. Wilke RA, Ramsey LB, Johnson SG, et al. The clinical pharmacogenomics implementation consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin Pharmacol Ther*. Jul 2012;92(1):112-117.

Table 1. Mutations identified in the validation study

ID	Detected by SureSelect	Detected by Access Array	Gene	cDNA	Protein
990	ND	Yes	<i>APOB</i>	c.10580G>A	p.Arg3527Gln
747	Yes	Yes	<i>APOB</i>	c.10580G>A	p.Arg3527Gln
1086	ND	Yes	<i>APOB</i>	c.10580G>A	p.Arg3527Gln
667	ND	Yes	<i>PCSK9</i>	c.385G>A	p.Asp129Asn
735	ND	Yes	<i>LDLR</i>	c.1618G>A	p.Ala540Thr
604	Yes	Yes	<i>LDLR</i>	c.1880C>A	p.Ala627Asp
660	ND	Yes	<i>LDLR</i>	c.1048C>T	p.Arg350X
728	Yes	Yes	<i>LDLR</i>	c.1048C>T	p.Arg350X
631	Yes	Yes	<i>LDLR</i>	c.2231G>A	p.Arg744Gln
384	Yes	Yes	<i>LDLR</i>	c.1217G>A	p.ArgR406Gln
895	ND	Yes	<i>LDLR</i>	c.1691A>G	p.Asn564Ser
896	ND	Yes	<i>LDLR</i>	c.1691A>G	p.Asn564Ser

766	Yes	Yes	<i>LDLR</i>	c.2061dupC	p.Asn688Glnfs*29
430	Yes	Yes	<i>LDLR</i>	c.661G>A	p.Asp221Asn
570#	Yes	Yes	<i>LDLR</i>	c.681C>G	p.Asp227Glu
862	ND	Yes	<i>LDLR</i>	c.681C>G	p.Asp227Glu
1095	ND	Yes	<i>LDLR</i>	c.1444G>A	p.Asp482Asn
438	Yes	Yes	<i>LDLR</i>	c.1444G>C	p.Asp482His
1031	ND	Yes	<i>LDLR</i>	c.401G>A	p.Cys134Tyr
714	Yes	Yes	<i>LDLR</i>	c.939C>A	p.Cys313X
1067	ND	Yes	<i>LDLR</i>	c.1130G>A	p.Cys377Tyr
1068	ND	Yes	<i>LDLR</i>	c.1130G>A	p.Cys377Tyr
819	Yes	Yes	<i>LDLR</i>	c.2029T>C	p.Cys677Arg
818	ND	Yes	<i>LDLR</i>	c.2043C>A	p.Cys681X
450	Yes	Yes	<i>LDLR</i>	c.2043C>A	p.Cys681X
814	Yes	Yes	<i>LDLR</i>	c.2043C>A	p.Cys681X
415	Yes	Yes	<i>LDLR</i>	c.[1151A>C];[1151A>C]	p.Gln384Pro

813	ND	Yes	<i>LDLR</i>	c.303delG	p.Glu101Aspfs*105
996	ND	Yes	<i>LDLR</i>	c.301G>A	p.Glu101Lys
771	ND	Yes	<i>LDLR</i>	c.301G>A	p.Glu101Lys
820	ND	Yes	<i>LDLR</i>	c.420G>C	p.Glu140Asp
770	Yes	Yes	<i>LDLR</i>	c.654_656delTGG	p.Glu219del
586	Yes	Yes	<i>LDLR</i>	c.682G>T	p.Glu228X
1071	ND	Yes	<i>LDLR</i>	c.1646G>A	p.Gly549Asp
853	Yes	Yes	<i>LDLR</i>	c.1436T>C	p.Leu479Pro
857	Yes	No	<i>LDLR</i>	c.2184_2194delGCTAAAGGTCA	p.Lys730Hisfs*48
743	ND	Yes	<i>LDLR</i>	c.1A>T	p.Met?
663	Yes	Yes	<i>LDLR</i>	c.1118_1121dupGTGG	p.Tyr375Trpfs*7
764	Yes	Yes	<i>LDLR</i>	c.1118_1121dupGTGG	p.Tyr375Trpfs*7
1102	ND	Yes	<i>LDLR</i>	c.1285G>A	p.Val429Met
1104	ND	Yes	<i>LDLR</i>	c.1285G>A	p.Val429Met
666	Yes	Yes	<i>LDLR</i>	c.1285G>A	p.Val429Met

852	ND	Yes	<i>LDLR</i>	c.1586+5G>C	
1011	ND	Yes	<i>LDLR</i>	c.1706-1G>T	
1094	ND	Yes	<i>LDLR</i>	c.1845+1 G>A	
721	ND	Yes	<i>LDLR</i>	c.1845+11C>G	
566	Yes	Yes	<i>LDLR</i>	c.313+1 G>A	
556	Yes;Yes	Yes;No	<i>LDLR</i>	c.[266G>A];[2312-?_2547+?]	
683	Yes	No	<i>LDLR</i>	c.-187-?_67+?del	promoter and exon 1 deletion
836#	Yes	No	<i>LDLR</i>	c.-187-?_67+?del	promoter and exon 1 deletion
547	Yes	No	<i>LDLR</i>	c.-187-?_67+?del	promoter and exon 1 deletion
662#	Yes	No	<i>LDLR</i>	c.2312-?_2389+?del	exon 16 deletion
436	Yes	No	<i>LDLR</i>	c.695-?_817+?del	exon 5 deletion
564	Yes	No	<i>LDLR</i>	c.68-?_2583+?del	exons 2-18 deletion
565	Yes	No	<i>LDLR</i>	c.68-?_940+?del	exons 2-6 deletion
712	Yes	No	<i>LDLR</i>	c.1187-?_2140+?dup	exons 9-14 duplication

bold, novel variants that were not previously identified. ND, not done. #, new diagnosis in samples that had previously undergone partial screening only.

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

DNA	Detected	Detected	Gene	cDNA	Protein	Simon	Age at	Highest
ID	by	by				Broome	measurement	total
	SureSelect	Access				criteria		cholesterol
		Array						(mmol/l)
1	Yes	Yes	<i>LDLR</i>	c.938G>A	p.Cys313Tyr	definite	48	15.9
2	ND	Yes	<i>LDLR</i>	c.948delC	p.Asn316Lysfs*54	definite	12	11.2
3	ND	Yes	<i>LDLR</i>	c.190+4A>T		definite	44	9.4
5	ND	Yes	<i>LDLR</i>	c.2479G>A	p.Val827Ile	definite	58	7.9
6	ND	Yes	<i>LDLR</i>	c.2054C>T/ c.736delG	p.Pro685Leu/ p.Gly246Glufs*19	definite	2	9.7
9	Yes	Yes	<i>LDLR</i>	c.1476_1477delCT	p.Ser493Cysfs*42	possible	20	9.2
12	Yes	no	<i>LDLR</i>	c.2187_2197delAAAGGTCAGCT	p.Lys730Hisfs*48	possible	25	13.0
14	ND	Yes	<i>LDLR</i>	c.204C>A	p.Cys68*	possible	5	8.6
15	ND	Yes	<i>LDLR</i>	c.1681C>T	p.Gln561*	possible	3	5.2

18	ND	Yes	<i>APOB</i>	c.10579C>T	p.Arg3527Trp	possible	25	9.2
21	ND	Yes	<i>LDLR</i>	c.1476_1477delCT	p.Ser493Cysfs*42	possible	29	8.1
22	Yes	Yes	<i>LDLR</i>	c.283T>C	p.Cys95Arg	possible	38	9.0
33	Yes	Yes	<i>LDLR</i>	c.1561G>A	p.Ala521Thr	possible	33	10.0
35	Yes	Yes	<i>LDLR</i>	c.817+1G>A		possible	13	10.5
40	ND	Yes	<i>LDLR</i>	c.670_675delinsTTT	p.Asp224_Lys225delinsPhe	possible	41	11.6
45	ND	Yes	<i>LDLR</i>	c.1730G>A	p.Trp577*	possible	56	10.3
50	ND	Yes	<i>LDLR</i>	c.906C>G	p.Cys302Trp	possible	NA	NA
52	ND	Yes	<i>APOB</i>	c.10580G>A	p.Arg3527Gln	possible	23	9.2
56	ND	Yes	<i>LDLR</i>	c.266G>A	p.Cys89Tyr	possible	NA	NA
58	ND	Yes	<i>LDLR</i>	c.654_656delTGG	p.Gly219del	possible	40	9.6
61	ND	Yes	<i>LDLR</i>	c.941-4G>A		possible	43	NA
63	ND	Yes	<i>LDLR</i>	c.2054C>T	p.Pro685Leu	possible	11	9.0
68	ND	Yes	<i>LDLR</i>	c.1130G>A	p.Cys377Tyr	possible	12	7.0
72	ND	Yes	<i>LDLR</i>	c.313+1G>A		not	47	9.6

fulfilling
criteria

In **bold**, novel, not previously identified variants. These variants are expected to cause and were shown to segregate with the disease.

In *italic*, variants of unknown significance. These rare variants were previously reported in patients, but their pathogenicity has not been established. ND, not done; NA, not available

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

Table S2. Average amplicon coverage for PCR-based Fluidigm runs sequenced using the MiSeq system

Gene	Exon	Chromosome	Start	End	Average coverage
<i>PCSK9</i>	2	1	55509540	55509737	21.5
<i>PCSK9</i>	4	1	55517957	55518141	664.2
<i>PCSK9</i>	7	1	55523043	55523239	173.3
<i>PCSK9</i>	9	1	55524158	55524343	450.7
<i>APOB</i>		2	21229075	21229274	37.2
<i>SLCO1B1</i>		12	21331482	21331643	163.3
<i>LDLR</i>	1	19	11199777	11200413	794.1
<i>LDLR</i>	2	19	11210845	11211053	839.8
<i>LDLR</i>	3	19	11213288	11213532	807.0
<i>LDLR</i>	4	19	11215842	11216330	231.8
<i>LDLR</i>	5	19	11217176	11217413	885.9
<i>LDLR</i>	6	19	11217997	11218256	411.8
<i>LDLR</i>	7	19	11221268	11221509	533.5
<i>LDLR</i>	8	19	11222097	11222363	176.8
<i>LDLR</i>	9	19	11223891	11224181	446.5

<i>LDLR</i>	10	19	11224183	11224492	636.8
<i>LDLR</i>	11	19	11226682	11226963	279.8
<i>LDLR</i>	12	19	11227483	11227725	661.8
<i>LDLR</i>	13	19	11230711	11230960	654.1
<i>LDLR</i>	14	19	11230990	11231232	670.6
<i>LDLR</i>	15	19	11233780	11234067	426.8
<i>LDLR</i>	16	19	11238627	11238806	740.9
<i>LDLR</i>	17	19	11240131	11240381	751.7
<i>LDLR</i>	18	19	11241903	11242087	817.9

Table S3. List of rare variants (MAF < 0.01), identified in hypercholesterolaemic patients with no molecular diagnosis, in genes on cholesterol metabolism pathways

Gene	Variant	Number of	Consequence	Protein	Amino acid	PolyPhen
		heterozygotes		position	change	
<i>SCAP</i>	3_47467597_C/T	1	NON_SYNONYMOUS_CODING	267	R/Q	benign(0.006)
<i>NPC1L1</i>	7_44579062_C/G	1	NON_SYNONYMOUS_CODING	312	D/H	benign(0.003)
<i>INSIG1</i>	7_155090174_C/G	1	NON_SYNONYMOUS_CODING	60	P/R	benign(0)
<i>CYP7A1</i>	8_59404257_T/C	1	NON_SYNONYMOUS_CODING	431	K/R	benign(0.131)
<i>SREBF1</i>	17_17718592_C/T	1	NON_SYNONYMOUS_CODING	842	R/Q	benign(0.015)
<i>SREBF1</i>	17_17719222_C/T	1	NON_SYNONYMOUS_CODING	809	V/M	possibly_damaging(0.801)
<i>PTBP1</i>	19_806440_G/A	1	NON_SYNONYMOUS_CODING	335	A/T	benign(0)

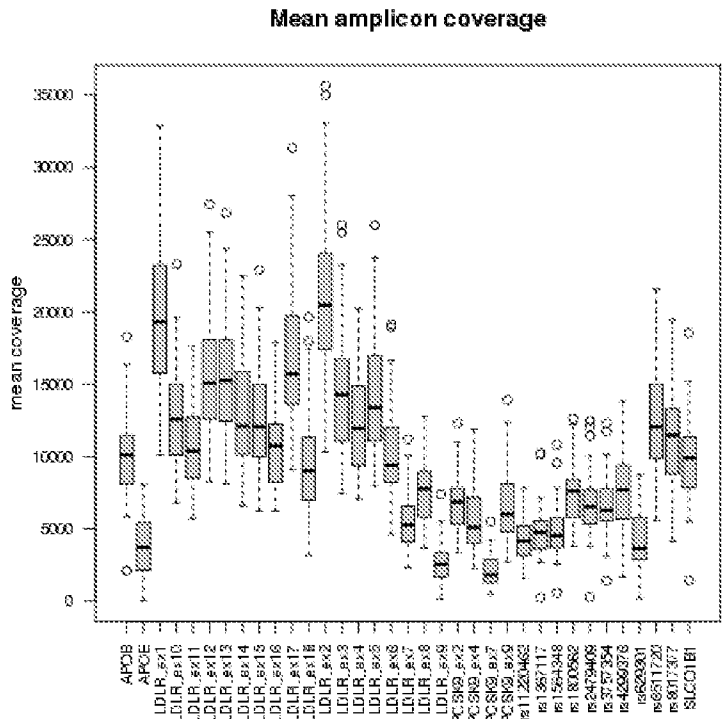
5 All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described aspects and embodiments of the present invention will be apparent to those skilled in the art without departing from the scope of the present invention. Although the present invention has been described in connection with specific preferred embodiments, it
10 should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are apparent to those skilled in the art are intended to be within the scope of the following claims.

CLAIMS

1. A method for collecting information useful in aiding the diagnosis and treatment of familial hypercholesterolaemia (FH), the method comprising providing a
5 sample of DNA from the subject, assaying said DNA for
- (i) mutation in one or more genes which is indicative of FH diagnosis
 - (ii) presence of one or more SNPs indicative of incremental risk of high cholesterol; and
 - (iii) mutation in a gene which is indicative of likelihood of statin toxicity;
- 10 wherein assaying said DNA comprises the step of contacting said DNA sample with one or more primers for amplification and/or sequencing of the relevant segment(s) of said DNA
characterised in that each of said assays of steps (i), (ii) and (iii) is carried out in a single assay cycle.
- 15
2. A method according to claim 1 wherein assaying said DNA further comprises determining the nucleotide sequence of the relevant segment(s) of said DNA, and inferring from said DNA sequence whether said mutations of steps (i) and (iii), and said one or more SNPs of step (ii), are present.
- 20
3. A method according to claim 1 or claim 2 wherein said one or more genes indicative of FH diagnosis is selected from the group consisting of APOB, LDLR and PCSK9.
- 25
4. A method according to any preceding claim wherein said mutation in one or more genes which is indicative of FH diagnosis is selected from those listed in Table 2.
5. A method according to any preceding claim wherein said SNP is selected from Table A or Table B.
- 30
6. A method according to any preceding claim wherein said gene indicative of risk of statin toxicity is SLCO1B1.
7. A method according to claim 6 wherein step (iii) comprises determining the
35 presence of the rs4149056 or rs4363657 variant of SLCO1B1.

8. A method according to claim 7 wherein step (iii) comprises determining the presence of the rs4149056 variant of SLCO1B1.
9. A method according to any preceding claim wherein said DNA sample is
5 subjected to micro-fluidic PCR amplification of the DNA segment(s) of interest.
10. A method according to claim 9 wherein said PCR is carried out using at least one primer selected from Table 3.
- 10 11. A method according to claim 10 using wherein said PCR is carried out using each of the primers in Table 3.
12. A method according to any preceding claim wherein sequence information is determined by next generation sequencing (NGS).
15
13. A method of treating a subject comprising performing the method according to any preceding claim wherein if an increased likelihood of FH is identified, then cholesterol lowering medicament is administered to said subject, wherein said medicament is selected according to the subject's SLCO1B1 genotype.
20
14. A method according to claim 13 wherein selection of the medicament comprises selection of one or more of the upper limit of the dose of said medicament for said subject, the dose of said medicament to be administered to said subject, or the particular medicament to be administered to said subject.
25
15. A method according to claim 14 wherein said medicament is a statin.
16. A primer selected from Table 3.
- 30 17. A set of primers comprising at least one pair of the primers in Table 3.
18. A set of primers according to claim 17 comprising each of the primers in Table 3.

Figure 1



5 Figure 2

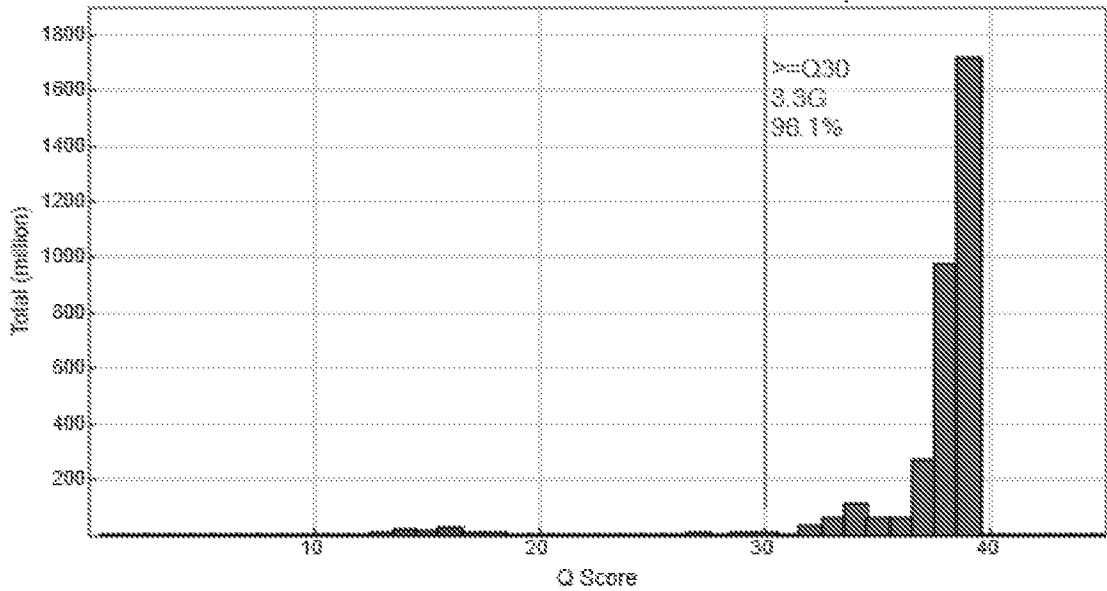


Figure 3

	Conventional Sanger sequencing	SureSelect Target Enrichment (Agilent)	TrueSeq Custom Amplicons (Illumina)	Access Array (Fluidigm)
Hands on time	15 hours	15 hours ^a	2-3h	0.5-1h
Total laboratory time	3 days	7 days ^b	3 days	1.5 days
Reagent and sequencing cost (per sample)	£140 ^c	£120	£110	£20
Usual number of samples per batch	10-20	10-20	48-96	48
DNA input	~1µg	~1µg	250ng	50ng
Advantages	Standard, proven accuracy	Comprehensive Detects large deletions and duplications	High sensitivity, fast	High sensitivity, little hands-on time, cheap
Disadvantages	Laborious, high cost	Laborious, high cost	Cannot detect large deletions or duplications	Cannot detect large deletions or duplications

Note: other target enrichment methods include HaloPlex (Agilent), Rainier's Targeted Sequencing (Rainier's Testrol), MASTR (Molgelabs), Ion Torrent PGM

^a Based on bidirectional sequencing of 20 amplicons and £2.00 per sequence

^b Time may be decreased with robotic automation

Figure 4 (Figure S1)

5

A

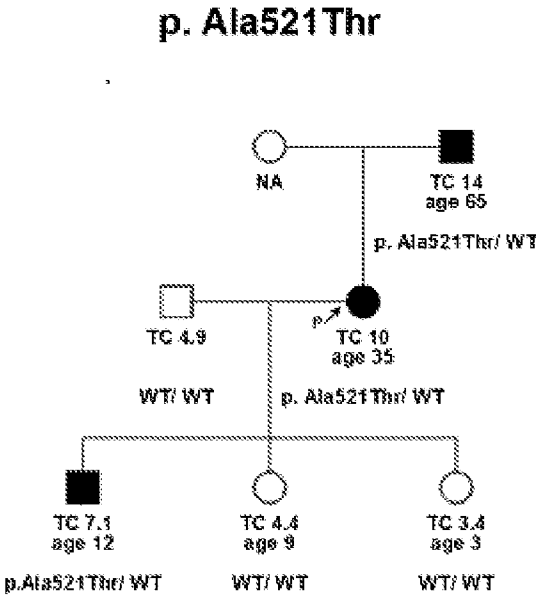
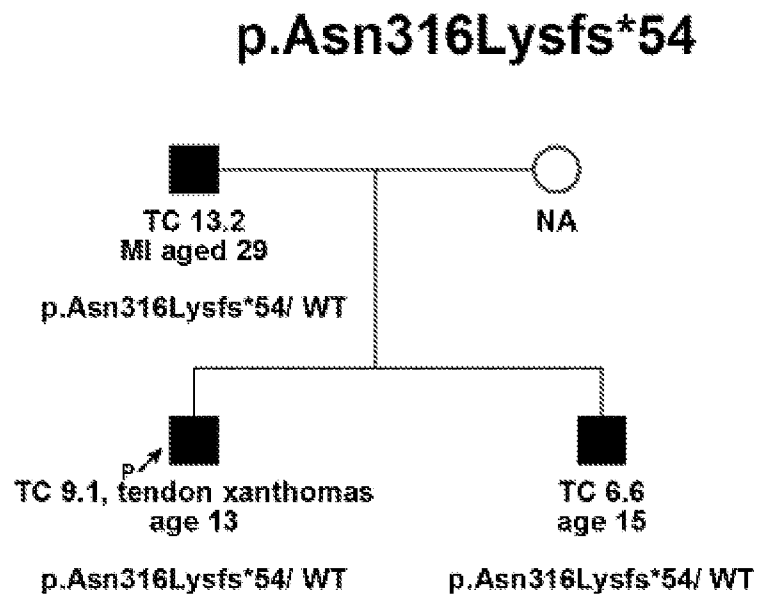


Figure 4 (Figure S1) cont'd



B

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2014/051399

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12Q1/68

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2011/027219 A1 (PROGENIKA BIOPHARMA SA [ES]; STEF MARIANNE [ES]; TEJEDOR DIEGO [ES]; M) 10 March 2011 (2011-03-10) table 1 sequence 719	16,17
X	WO 2005/072151 A2 (GENAISSANCE PHARMACEUTICALS [US]; AERSSSENS JEROEN [BE]; ATHANASIOU MAR) 11 August 2005 (2005-08-11) example 1 sequence 40	16,17

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

12 August 2014

Date of mailing of the international search report

08/09/2014

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Ulbrecht, Matthias

INTERNATIONAL SEARCH REPORT

International application No.

PCT/GB2014/051399

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

a. (means)

☒

on paper

☒

in electronic form

b. (time)

☐

in the international application as filed

☐

together with the international application in electronic form

☒

subsequently to this Authority for the purpose of search

2. ☒ In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2014/051399

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LITTLE DANIEL P ET AL: "Identification of apolipoprotein E polymorphisms using temperature cycled primer oligo base extension and mass spectrometry", EUROPEAN JOURNAL OF CLINICAL CHEMISTRY AND CLINICAL BIOCHEMISTRY, WALTER DE GRUYTER, BERLIN, DE, vol. 35, no. 7, 1 January 1997 (1997-01-01), pages 545-548, XP009107192, ISSN: 0939-4974 page 546, left-hand column, paragraph 1 -----	16,17
X	JP 2010 200701 A (FUKUOKA PREFECTURE) 16 September 2010 (2010-09-16) chemical formula 24 -----	16,17
Y	PHILIPPA J TALMUD ET AL: "Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study", THE LANCET, vol. 381, no. 9874, 22 February 2013 (2013-02-22), pages 1293-1301, XP055133974, ISSN: 0140-6736, DOI: 10.1016/S0140-6736(12)62127-8 abstract page 1294, right-hand column, last paragraph - page 1295, left-hand column, paragraph 1 supplementary data p. 1-2 supplementary Tab. 4 -----	1-18
Y	WO 2013/043041 A2 (INFOVALLEY GROUP OF COMPANIES [MY]) 28 March 2013 (2013-03-28) page 2, last paragraph table 1 claims 1-24 -----	1-15
Y	R A WILKE ET AL: "The Clinical Pharmacogenomics Implementation Consortium: CPIC Guideline for SLC01B1 and Simvastatin-Induced Myopathy", CLINICAL PHARMACOLOGY & THERAPEUTICS, vol. 92, no. 1, 23 May 2012 (2012-05-23), pages 112-117, XP55134179, ISSN: 0009-9236, DOI: 10.1038/clpt.2012.57 abstract figure 1 ----- -/--	1-18

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2014/051399

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	S. HOLLANTS ET AL: "Microfluidic Amplification as a Tool for Massive Parallel Sequencing of the Familial Hypercholesterolemia Genes", CLINICAL CHEMISTRY, vol. 58, no. 4, 1 April 2012 (2012-04-01), pages 717-724, XP55080574, ISSN: 0009-9147, DOI: 10.1373/clinchem.2011.173963 the whole document	1-18
A	----- LOURDES PALACIOS ET AL: "Molecular characterization of familial hypercholesterolemia in Spain", ATHEROSCLEROSIS, vol. 221, no. 1, 1 March 2012 (2012-03-01), pages 137-142, XP55133996, ISSN: 0021-9150, DOI: 10.1016/j.atherosclerosis.2011.12.021 abstract page 138, left-hand column, paragraph 5	1-18
A	----- G. K. HOVINGH ET AL: "Diagnosis and treatment of familial hypercholesterolaemia", EUROPEAN HEART JOURNAL, vol. 34, no. 13, 1 April 2013 (2013-04-01), pages 962-971, XP055133938, ISSN: 0195-668X, DOI: 10.1093/eurheartj/eh015 the whole document	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2014/051399

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2011027219	A1	10-03-2011	NONE

WO 2005072151	A2	11-08-2005	AU 2005208566 A1 11-08-2005
		CA 2555367 A1	11-08-2005
		EP 1743036 A2	17-01-2007
		JP 2007526764 A	20-09-2007
		US 2005277129 A1	15-12-2005
		WO 2005072151 A2	11-08-2005

JP 2010200701	A	16-09-2010	JP 5526337 B2 18-06-2014
		JP 2010200701 A	16-09-2010

WO 2013043041	A2	28-03-2013	NONE
