(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 June 2013 (06.06.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/080227 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/IS2012/050015

(22) International Filing Date:
27 November 2012 (27.11.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
050023 28 November 2011 (28.11.2011) IS

(71) Applicant: **DECODE GENETICS EHF** [IS/IS]; Sturlugata 8, IS-101 Reykjavik (IS).

(72) Inventor: **HELGADOTTIR, Anna;** Hamravik 86, IS-112 Reykjavik (IS).

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

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2013/080227 A1

(54) **Title:** GENETIC VARIANTS USEFUL FOR RISK ASSESSMENT OF ARTERIAL DISEASE

(57) **Abstract:** It has been discovered that certain genetic variants are predictive of risk of arterial disease in humans, such as large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease. The present invention discloses such variants and their use in methods of determining susceptibility to arterial diseases in humans. The variant is at least one allele of a polymorphism selected from rs3798220 and rs10455872. Also disclosed are kits and computer systems for use in such methods.

GENETIC VARIANTS USEFUL FOR RISK ASSESSMENT OF ARTERIAL DISEASE

BACKGROUND OF THE INVENTION

Genome-wide association studies and recent large scale meta-analyses of such studies have identified many genetic variants that confer risk of arterial diseases, including Coronary Artery Disease (CAD), Myocardial Infarction (MI), Abdominal Aortic Aneurysm (AAA) and Peripheral Arterial Disease (PAD). These variants have been discovered by comparing sequence data (genotypes) of cases with apparently healthy controls. For most of these newly discovered susceptibility variants the potential mechanisms and intermediate pathways by which they confer risk of these arterial diseases is not unknown, i.e. the identified loci may for example promote the initiation of coronary atheroma, the progression of atherosclerotic plaques, increase susceptibility of plaque rupture, or faciltate arterial thrombosis, thus leading to the observed phenotype. Thus, knowledge about genetic variation that confers risk of these diseases may provide valuable information about their etiology.

A knowledge of the underlying mechanism whereby particular genetic variants mediate their risk is also clinically important as it may modify personalized risk assessment and it can further be used in the developement of novel therapies. The present invention discloses the utility of certain genetic variants in lipoprotein (a) (Lp(a)) in the risk management of certain vascular conditions.

SUMMARY OF THE INVENTION

The present inventors have discovered that certain genetic variants in the human LPA gene are associated with risk of arterial conditions such as large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease. These variants are useful in diagnostic methods and related aspects, as described in detail herein.

In one aspect, the invention provides a method of determining a susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject, the method comprising analyzing data representative of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872 in the subject, wherein the presence of allele C in rs3798220 and/or allele G in rsl0455872 is associated with increased susceptibility to the condition in humans, and determining a susceptibility to the condition for the human subject from the data.

The invention also provides a method of determining whether a human subject is at increased risk of developing a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, the method comprising steps of (i) obtaining a biological sample containing nucleic acid from the subject; and (ii) determining, in the biological sample, the allelic identity of at least one polymorphic marker selected from rs3798220 and rsl0455872; wherein identification of at least one allele selected from allele C in rs3798220 and allele G in rsl0455872 in the biological sample is indicative that the subject is at increased risk of developing the condition.

Further provided is a method of determining a susceptibility to severe coronary artery disease in a human subject, the method comprising analyzing data representative of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872 in the subject, wherein the presence of allele C in rs3798220 and allele G in rsl0455872 is associated with increased number of obstructed coronary artery in humans, and determining a susceptibility to severe coronary artery disease for the human subject from the data.

The invention also provides an assay for determining a susceptibility to a condition selected from the group consisting of abdominal aortic aneurysm, peripheral arterial disease, large arterial atherosclerosis, early-onset coronary artery disease and severe coronary artery disease in a human subject, the assay comprising steps of (i) obtaining a biological sample comprising nucleic acid from the subject; (ii) determining, in the nucleic acid, the presence or absence of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872, wherein the presence of allele C in rs3798220 and allele G in rsl0455872 is associated with increased susceptibility to the condition in humans; and (iii) determining a susceptibility to the condition for the subject from the presence or absence of the at least one allele.

The invention also relates to prognosis. In one aspect, the invention provides a method of determining the prognosis of a human subject diagnosed with coronary artery disease, the method comprising steps of (i) obtaining a biological sample containing nucleic acid from the subject; and (ii) determining, in the biological sample, the allelic identity of at least one polymorphic marker selected from rs3798220 and rsl0455872; wherein identification of at least one allele selected from allele C in rs3798220 and allele G in rsl0455872 in the biological sample is indicative of a worse prognosis of the coronary artery disease for the subject.

Further provided are computer-implemented aspects, as described in more detail herein. One such aspect relates to a system for identifying susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, the system comprising (i) at least one processor; (ii) at least one computer-readable medium; (iii) a susceptibility database operatively coupled to a computer-readable medium of the system and containing

population information correlating the presence or absence of one or more alleles selected from allele C in rs3798220 and allele G in rsl0455872 to susceptibility of the condition in a population of humans; (iv) a measurement tool that receives an input about the human subject and generates information from the input about the presence or absence of the at least one allele in the human subject; and (v) an analysis tool that (a) is operatively coupled to the susceptibility database and the measurement tool; (b) is stored on a computer-readable medium of the system; and (c) is adapted to be executed on a processor of the system, to compare the information about the human subject with the population information in the susceptibility database and generate a conclusion with respect to susceptibility to the condition for the human subject.

Further details of these and other aspects of the inventions are described in the following detailed description of the invention.

It should be understood that all combinations of features described herein are contemplated, even if the combination of features is not specifically found in the same sentence or paragraph as set forth in the following detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention.

FIG 1 provides a diagram illustrating a system comprising computer implemented methods utilizing risk variants as described herein.

FIG 2 shows an exemplary system for determining risk of vascular conditions as described further herein.

FIG 3 shows a system for selecting a treatment protocol for a subject diagnosed with a vascular condition.

FIG 4 demonstrates association of *LPA* score (markers rs3798220 and rsl0455872) with vascular diseases. Forest plot shows the association of the *LPA* score with ischemic stroke, and TOAST subtypes (LAA= large artery atherosclerosis, CE=cardioembolic stroke, SVD= small vessel disease), peripheral arterial disease (PAD), abdominal aortic aneurysm (AAA), venous thromboembolism (VTE), coronary artery disease (CAD), and intracranial aneurysm (IA) in individuals of European origin. Combined odds ratios (OR) from multiple studies are indicated by square, with an area in proportion to the sample size, and horizontal lines representing the 95% confidence intervals (CI). The plot shows the effective sample size (n_e cases = n_e controls) and P-value (*P*) for the association.

FIG 5 demonstrates the association of *LPA* score with CAD, with and without atherosclerosis in other vascular territories, and with age-group at onset of CAD. A), shows a forest plot of the association of *LPA* score with CAD in Iceland, for cases with concomitant cardiovascular disease (CVD) (CVD = LAA, PAD, or AAA) and for CAD after excluding cases with known CVD.

The effective sample sizes (n_e cases = n_e controls), the odds ratios (OR), 95% confidence intervals (CI) and the P-value (P) are shown. The P-value contrasting the two groups of CAD cases is shown (P*). B). is a forest plot of the association of *LPA* score with early onset of CAD (diagnosed before the age of 50 in men and 60 in women), intermediate onset (diagnosed at the age 50-69 for men and 60-74 for women), and late onset (diagnosed >70 in men and ≥75 in women). Individuals of European origin from Iceland and Atlanta (USA) were combined using an inverse variance weighted meta-analysis approach (fixed-effect model).

FIG 6 shows the distribution of the number of coronary vessels with ≥50% stenosis according to *LPA* score is shown in the boxplots A - C. The boxes represent the interquartile range and the median is represented by the center line within the box. Embedded in the figure is the Spearman rank correlation coefficients (rho) for the correlation between the *LPA* score (coded as 0,1,2) and the different CAD severity scores, together with the P-value. N= number of subjects.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

Unless otherwise indicated, nucleic acid sequences are written left to right in a 5' to 3' orientation. Numeric ranges recited within the specification are inclusive of the numbers defining the range and include each integer or any non-integer fraction within the defined range. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by the ordinary person skilled in the art to which the invention pertains.

The following terms shall, in the present context, have the meaning as indicated:

A "polymorphic marker", sometime referred to as a "marker", as described herein, refers to a genomic polymorphic site. Each polymorphic marker has at least two sequence variations characteristic of particular alleles at the polymorphic site. Thus, genetic association to a polymorphic marker implies that there is association to at least one specific allele of that particular polymorphic marker. The marker can comprise any allele of any variant type found in the genome, including SNPs, mini- or microsatellites, translocations and copy number variations (insertions, deletions, duplications). Polymorphic markers can be of any measurable frequency in the population. For mapping of disease genes, polymorphic markers with population frequency higher than 5-10% are in general most useful. However, polymorphic markers may also have lower population frequencies, such as 1-5% frequency, or 0.1-1% frequency, or even lower frequency. The term shall, in the present context, be taken to include polymorphic markers with any population frequency.

An "allele" refers to the nucleotide sequence of a given locus (position) on a chromosome. A polymorphic marker allele thus refers to the composition (i.e., sequence) of the marker on a

chromosome. Genomic DNA from an individual contains two alleles (e.g., allele-specific sequences) for any given polymorphic marker, representative of each copy of the marker on each chromosome. Sequence codes for nucleotides used herein are: A = 1, C = 2, G = 3, T = 4. For microsatellite alleles, the CEPH sample (Centre d'Etudes du Polymorphisme Humain, genomics repository, CEPH sample 1347-02) is used as a reference, the shorter allele of each microsatellite in this sample is set as 0 and all other alleles in other samples are numbered in relation to this reference. Thus, e.g., allele 1 is 1 bp longer than the shorter allele in the CEPH sample, allele 2 is 2 bp longer than the shorter allele in the CEPH sample, allele 3 is 3 bp longer than the lower allele in the CEPH sample, etc., and allele -1 is 1 bp shorter than the shorter allele in the CEPH sample, allele -2 is 2 bp shorter than the shorter allele in the CEPH sample, etc.

Sequence conucleotide ambiguity as described herein is as proposed by IUPAC-IUB. These codes are compatible with the codes used by the EMBL, GenBank, and PIR databases.

IUB code	Meaning
A	Adenosine
C	Cytidine
G	Guanine
T	Thymidine
R	G or A
Y	T or C
K	G or T
M	A or C
S	G or C
W	A or T
B	C, G or T
D	A, G or T
H	A, C or T
V	A, C or G
N	A, C, G or T (Any base)

A nucleotide position at which more than one sequence is possible in a population (either a natural population or a synthetic population, e.g., a library of synthetic molecules) is referred to herein as a "polymorphic site".

A "Single Nucleotide Polymorphism" or "SIMP" is a DNA sequence variation occurring when a single nucleotide at a specific location in the genome differs between members of a species or between paired chromosomes in an individual. Most SNP polymorphisms have two alleles. Each individual is in this instance either homozygous for one allele of the polymorphism (i.e. both chromosomal copies of the individual have the same nucleotide at the SNP location), or the individual is heterozygous (i.e. the two sister chromosomes of the individual contain different nucleotides). SNP nomenclature sometimes refers to the official Reference SNP (rs) ID identification tag as assigned to each unique SNP by the National Center for

Biotechnological Information (NCBI). The SNPs disclosed herein are further identified by their genomic flanking sequences, as illustrated by the accompanying sequence listing.

A "variant", as described herein, refers to a segment of DNA that differs from the reference DNA. A "marker" or a "polymorphic marker", as defined herein, is a variant. Alleles that differ from the reference are referred to as "variant" alleles. A "variant" is sometimes also referred to as a "mutant".

A "microsatellite" is a polymorphic marker that has multiple small repeats of bases that are 2-8 nucleotides in length (such as CA repeats) at a particular site, in which the number of repeat lengths varies in the general population. An "indel" is a common form of polymorphism comprising a small insertion or deletion that is typically only a few nucleotides long.

A "haplotype," as described herein, refers to a segment of genomic DNA that is characterized by a specific combination of alleles arranged along the segment. For diploid organisms such as humans, a haplotype comprises one member of the pair of alleles for each polymorphic marker or locus along the segment. In a certain embodiment, the haplotype can comprise two or more alleles, three or more alleles, four or more alleles, or five or more alleles.

The term "susceptibility", or "risk", as described herein, refers to the proneness of an individual towards the development of a certain state (e.g., a certain trait, phenotype or disease), or towards being less able to resist a particular state than the average individual. The term encompasses both increased susceptibility and decreased susceptibility. Thus, particular alleles at polymorphic markers of the invention as described herein may be characteristic of increased susceptibility (i.e., increased risk) of a vascular condition as described herein, as characterized by a relative risk (RR) or odds ratio (OR) of greater than one for the particular allele. Alternatively, the markers of the invention are characteristic of decreased susceptibility (i.e., decreased risk) of the condition, as characterized by a relative risk of less than one.

The term "and/or" shall in the present context be understood to indicate that either or both of the items connected by it are involved. In other words, the term herein shall be taken to mean "one or the other or both".

The term "look-up table", as described herein, is a table that correlates one form of data to another form, or one or more forms of data to a predicted outcome to which the data is relevant, such as phenotype or trait. For example, a look-up table can comprise a correlation between allelic data for at least one polymorphic marker and a particular trait or phenotype, such as a particular disease diagnosis, that an individual who comprises the particular allelic data is likely to display, or is more likely to display than individuals who do not comprise the particular allelic data. Look-up tables can be multidimensional, *i.e.* they can contain information about multiple alleles for single markers simultaneously, or they can contain information about multiple markers, and they may also comprise other factors, such as

particulars about diseases diagnoses, racial information, biomarkers, biochemical measurements, therapeutic methods or drugs, etc.

A "computer-readable medium", is an information storage medium that can be accessed by a computer using a commercially available or custom-made interface. Exemplary computer-readable media include memory (e.g., RAM, ROM, flash memory, etc.), optical storage media (e.g., CD-ROM), magnetic storage media (e.g., computer hard drives, floppy disks, etc.), punch cards, or other commercially available media. Information may be transferred between a system of interest and a medium, between computers, or between computers and the computer-readable medium for storage or access of stored information. Such transmission can be electrical, or by other available methods, such as IR links, wireless connections, etc.

A "nucleic acid sample" as described herein, refers to a sample obtained from an individual that contains nucleic acid (DNA or RNA). In certain embodiments, i.e. the detection of specific polymorphic markers and/or haplotypes, the nucleic acid sample comprises genomic DNA. Such a nucleic acid sample can be obtained from any source that contains genomic DNA, including a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta, gastrointestinal tract or other organs.

The term "antisense agent" or "antisense oligonucleotide" refers, as described herein, to molecules, or compositions comprising molecules, which include a sequence of purine an pyrimidine heterocyclic bases, supported by a backbone, which are effective to hydrogen bond to a corresponding contiguous bases in a target nucleic acid sequence. The backbone is composed of subunit backbone moieties supporting the purine and pyrimidine hetercyclic bases at positions which allow such hydrogen bonding. These backbone moieties are cyclic moieties of 5 to 7 atoms in size, linked together by phosphorous-containing linkage units of one to three atoms in length. In certain preferred embodiments, the antisense agent comprises an oligonucleotide molecule.

The term "Lp(a)", as described herein, refers to lipoprotein(a). The Lp(a) protein is encoded by the lipoprotein(a) gene, which is referred to herein as *"LPA".*

The term "systemic atherosclerotic disease", as described herein, refers to atherosclerotic disease that affects multiple (at least two) regional circulations, including heart, brain, kidneys, limbs, and mesentery.

Association of genetic variants to vascular conditions and severity of cardiovascular disease

It has been discovered that variants in the human lipoprotein a gene *{LPA)* are predictive of risk of vascular conditions and severity of cardiovascular disease. Two variants in the human LPA gene, rsl0455872 and rs3798220, were studied. It was surprisingly found that the previously known association of these variants to lipoprotein(a) (Lp(a)) levels and coronary

artery disease (CAD) extends throughout the arterial tree. Moreoer, the variants were found to be predictive of the severity of coronary artery disease.

Lipoprotein structure is similar to plasminogen and tPA (tissue plasminogen activator) and it has been shown to affect (reduce) fibrinolysis and (increase) thrombogenesis. In addition, the LDL cholesterol content of Lp(a) contributes to atherosclerosis.

High LPA levels in blood has been shown to be a risk factor for coronary heart disease (CHD), cerebrovascular disease (CVD), atherosclerosis, thrombosis, and stroke. However, few reported studies have investigated the association between genetic variants that affect Lp(a) concentration and coronary artery disease. The frequency of small apo(a) isoforms has been shown to be higher in patients with myocardial infarction compared with healthy controls (Atheroslerosis **144** :323-333 (1999)) and another study has shown the isoforms to be higher in those with advanced atherosclerosis of the carotid arteries compared to those without advanced atherosclerosis (Circulation **100** :1154-1160 (1999)).

Previous studies have reported that the single nucleotide polymorphisms rs3798220 and rs10455872 in the coding sequence of the *LPA* gene are associated with coronary artery disease and Lp(a) levels (Arterioscler Thromb Vase Biol **27**:2030-2036 (2007); N Eng J Med **361**:2518-28 (2009)).

The association of these markers with age at onset of CAD, extent of CAD observed at coronary angiography or their association with other cardiovascular diseases have not been previously described. In fact, the skilled person would have predicted that the association with CAD was mediated primarily through the reported thrombotic properties of the Lp(a) in blood (as the variants increase serum Lp(a) levels). However, the present inventors have shown that *LPA* risk alleles confer increased susceptibility to atherosclerotic manifestations outside the coronary tree and that individuals carrying such risk alleles are more likely to be diagnosed earlier with CAD than individuals with CAD that are not carrying these variants. Further, the present inventors provide evidence for an association between *LPA* variants and the number of coronary arteries (out of 4 index vessels) with ≥50% stenosis, *i.e.* severity of CAD. Thus these *LPA* variants, and by implication Lp(a) levels, have an effect on the atherosclerotic burden of large vessels throughout the arterial tree.

Given the ability to predict risk of increased atherosclerotic burden among CAD cases, genotyping LPA variants may be clinically useful, not only in determining risk for developement of disease, but also for risk stratification among cases with documented CAD. The present inventors have found that LPA alleles can predict an important disease subtype, that carries an adverse prognosis. Evaluating such *LPA* risk alleles may therefore compliment traditional risk assessment in secondary CAD prevention as well as in primary prevention.

Evaluating *LPA* risk alleles in patients may further prove important as the risk alleles may interact with therapeutic agents aimed at lowering Lp(a) levels. The inventors have furthered the knowledge about the relationship between the LPA variants and CAD. Such insight into

the pathogenic mechanism is important when GWAS discoveries for CAD are translated into prognostic and therapeutic tools.

Methods of determining susceptibility to vascular conditions and severity of cardiovascular disease

Accordingly, the present invention in a first aspect provides a method of determining a susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject, the method comprising analyzing data representative of at least one allele of a polymorphism selected from the group consisting of rs3798220 (SEQ ID NO:20) and rsl0455872 (SEQ ID NO: 106) in the subject, wherein the presence of allele C in rs3798220 and/or allele G in rsl0455872 is associated with increased susceptibility to the condition in humans, and determining a susceptibility to the condition for the human subject from the data.

In one embodiment, the method comprises analyzing data for at least one allele of rs3798220 and rsl0455872, wherein a determination of the presence of allele C in rs3798220 and/or allele G in rsl0455872 is associated with increased susceptibility of the condition in the human subject. In one such embodiment, both alleles of markers rs3798220 and rsl0455872 are analyzed, wherein a determination of the presence of allele C in rs3798220 and/or allele G in rsl0455872 is associated with increased susceptibility of the condition in the human subject.

In one embodiment, the condition is severe coronary artery disease. In one embodiment, severe coronary artery disease is a disease that is characterized by at least one of (i) at least 50% stenosis of more than one coronary artery; (ii) systemic atherosclerotic disease in the human subject, and (iii) the concomitant diagnosis of coronary artery disease and at least one of (a) abdominal aortic aneurysm; (b) peripheral arterial disease; and (c) large arterial atherosclerosis.

It may be convenient to analyze data for a polymorphic marker in linkage disequilibrium with a marker of interest. As described further herein, due to linkage disequilibrium in the human genome, certain genetic markers are correlated due to being in linkage disequilibrium. Such correlated markers may be analyzed to obtain information about susceptibility. As a consequence, in one aspect, the invention provides a method of determining a susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject, the method comprising analyzing data representative of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872 in the subject, or a polymorphism in linkage disequilibrium therewith, and determining a susceptibility to the condition for the human subject from the data. In one embodiment, determination of the presence of an allele of the at least one polymorphism that

is associated with increased susceptibility to the condition in humans is indicative that the human subject is at increased susceptibility to the condition.

In certain embodiments, polymorphisms in linkage disequilibrium with rs3798220 and/or rsl0455872 are correlated with rs3798220 and/or rsl0455872 by values of the correlation coefficient $r^2 > 0.2$. In one preferred embodiment, polymorphisms in linkage disequilibrium with rs3798220 and/or rsl0455872 are correlated with rs3798220 and/or rsl0455872 by values of the correlation coefficient $r^2 > 0.5$. In one embodiment, polymorphisms correlated with rs3798220 are selected from the polymorphic markers set forth in the Surrogate Marker Table (a). In one embodiment, polymorphisms correlated with rsl0455872 are selected from the polymorphic markers set forth in the Surrogate Marker Table (b).

The analyzing data in certain embodiments comprises determining the allelic identity of at least one polymorphic marker. The analyzing may further comprise obtaining a biological sample from the human subject.

A second aspect therefore provides a method of determining whether a human subject is at increased risk of developing a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, the method comprising steps of (1) obtaining a biological sample containing nucleic acid from the subject; (2) determining, in the biological sample, the allelic identity of at least one polymorphic marker selected from rs3798220 and rsl0455872; (3) wherein identification of at least one allele selected from allele C in rs3798220 and allele G in rsl0455872 in the biological sample is indicative that the subject is at increased risk of developing the condition.

In certain embodiments, the data is nucleic acid sequence data. Nucleic acid sequence data identifying particular alleles of polymorphic markers is sometimes also referred to as genotype data. Allelic identity of allele(s) of a polymorphic marker is sometimes also referred to as a genotype. In one embodiment, nucleic acid sequence data is obtained from a biological sample from the individual. The nucleic acid sequence data can thus be sequence data obtained by analysis of a biological sample from an individual. The biological sample in one embodiment is a nucleic acid sample, i.e. the sample contains nucleic acid from the individual.

Nucleic acid sequence data can be obtained for example by analyzing sequence of the at least one polymorphic marker in a nucleic acid sample from the individual. Alternatively, nucleic acid sequence data can be obtained in a genotype dataset from the human individual and by analyzing sequence of the at least one polymorphic marker in the dataset. Such analysis in certain embodiments comprises determining the presence or absence of a particular allele of specific polymorphic markers.

In certain embodiments, the method of the invention comprises steps of (i) obtaining a nucleic acid sample from an individual; (ii) determining the nucleic acid sequence of at least

one polymorphic marker in the nucleic acid sample; and (iii) determining a susceptibility to the condition from the nucleic acid sequence of the at least one polymorphic marker.

A further aspect of the invention relates to method for determining susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject, the method comprising analyzing nucleic acid from the individual for evidence of the presence of allele C in rs3798220 and/or allele G in rsl0455872; and determining an increased susceptibility to the condition from evidence that allele C in rs3798220 and/or allele G in rsl0455872 is present in the subject, or determining a decreased susceptibility to the condition from evidence that allele C in rs3798220 and/or allele G in rsl0455872 is absent in the subject. The evidence of the presence of allele C in rs3798220 and/or allele G in rsl0455872 may conveniently be obtained by assaying or detecting allele C in rs3798220 and/or allele G in rsl0455872 directly. Alternatively, the skilled person will appreciate that it may also be convenient to obtain such evidence by assaying marker(s) that are surrogates of rs3798220 and/ rsl0455872. Such surrogates, which are markers that are in linkage disequilibrium and correlated with in rs3798220 and/or rsl0455872, provide information about rs3798220 and/or rsl0455872 through their genetic relationship. The surrogate markers are in essence useful proxies for rs3798220 and/or rsl0455872, and the evidence of the presence of allele C in rs3798220 and/or allele G in rsl0455872 may be obtained by assaying for specific alleles of such surrogate markers.

Surrogate markers can be selected based on certain values of the linkage disequilibrium measures D' and r^2, as described further herein. The skilled person will appreciate that since linkage disequilibrium is a continuous measure, certain values of the LD measure r^2 may be suitably chosen to define markers that are useful as surrogate markers in LD with the markers described herein. In one embodiment, suitable markers in linkage disequilibrium are correlated with the anchor marker *{i.e.,* rs3798220 and/or rsl0455872) by values of r^2 greater than 0.2. In another such embodiment, suitable markers in linkage disequilibrium are correlated with the anchor marker by values of r^2 greater than 0.5. In yet another such embodiment, suitable markers in linkage disequilibrium are correlated with the anchor marker by values of r^2 greater than 0.8. In one embodiment, suitable markers in linkage disequilibrium are correlated with the anchor marker by values of r^2 of 1.0. Such markers are perfect surrogates of the anchor marker, and will give identical association results, i.e. they provide identical genetic information.

To define markers that are useful in a diagnostic for determining a susceptibility to vascular conditions as described herein *{e.g.,* large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease), it may be useful to compare the frequency of markers alleles in individuals with the condition to their corresponding frequency in control individuals. In one embodiment, an increase in frequency of the at least one allele in the at least one polymorphism in individuals

diagnosed with the condition, as compared with the frequency of the at least one allele in the control group is indicative of the at least one allele being useful for assessing increased susceptibility to the condition.

In another embodiment, a decrease in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with the condition, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one allele being useful for assessing decreased susceptibility to, or protection against, the condition. In certain embodiments, the T allele of rs3798220 is indicative a decreased susceptibility to the condition. In certain embodiments, the G allele of rsl0455872 is indicative of a decreased susceptibility to the condition.

In general, sequence data can be obtained by analyzing a sample from an individual, or by analyzing information about specific markers in a genotype database. In certain embodiments, sequence data can be obtained through nucleic acid sequence information or amino acid sequence information from a preexisting record about a human individual. Such a preexisting record can be any documentation, database or other form of data storage containing such information.

Determination of a susceptibility or risk of a particular individual in general comprises comparison of the genotype information (sequence information about particular marker or a plurality of markers) to a record or database providing a correlation about particular polymorphic marker(s) and susceptibility to the vascular condition. Thus, in specific embodiments, determining a susceptibility comprises comparing the sequence data to a database containing correlation data between the at least one polymorphic marker and susceptibility to the condition. In certain embodiments, the database comprises at least one measure of susceptibility to the condition for the at least one polymorphic marker. In certain embodiments, the database comprises a look-up table comprising at least one measure of susceptibility to the condition for the at least one polymorphic marker. Determination of susceptibility is based on sequence information about particular markers identifying particular alleles at those markers. A calculation of susceptibility (risk) of the vascular condition is performed based on the information, using risk measures that have been determined for the particular alleles or combination of alleles. The measure of susceptibility may in the form of relative risk (RR), absolute risk (AR), percentage (%) or other convenient measure for describing genetic susceptibility of individuals.

Another aspect of the invention relates to a method for determining a susceptibility to a condition selected from the group consisting of abdominal aortic aneurysm, peripheral arterial disease, large arterial atherosclerosis, early-onset coronary artery disease and severe coronary artery disease in a human subject, comprising determining the presence or absence of at least one allele of at least one polymorphic marker in a nucleic acid sample obtained from the individual, or in a genotype dataset from the individual, wherein the at least one polymorphic marker is selected from the group consisting of rs3798220 and rsl0455872, and

wherein determination of the presence of the at least one allele is indicative of a susceptibility to the condition. Determination of the presence of an allele that correlates with the condition is indicative of an increased susceptibility (increased risk) to the condition. In certain embodiments, determination of the presence of an allele selected from the group consisting of allele C in rs3798220 and allele G in rsl0455872 is associated with increased susceptibility to the condition in humans. Individuals who are homozygous for these alleles are particularly susceptible to the condition. On the other hand, individuals who do not carry these at-risk alleles are at a decreased susceptibility of developing the condition. For SNPs, such individuals will be homozygous for the alternate (protective) allele of the polymorphism.

Determination of susceptibility is in some embodiments reported using non-carriers of the at-risk alleles of polymorphic markers as a reference. In certain embodiments, susceptibility is reported based on a comparison with the general population, *e.g.* compared with a random selection of individuals from the population. Such embodiments thus reflect the susceptibility (risk) of an individual compared with a randomly selected individual from the population.

Obtaining nucleic acid sequence data

In certain embodiments, polymorphic markers are detected by sequencing technologies. Obtaining sequence information about an individual identifies particular nucleotides in the context of a sequence. For SNPs, sequence information about a single unique sequence site is sufficient to identify alleles at that particular SNP. For markers comprising more than one nucleotide, sequence information about the genomic region of the individual that contains the polymorphic site identifies the alleles of the individual for the particular site. The sequence information can be obtained from a sample from the individual. In certain embodiments, the sample is a nucleic acid sample. In certain other embodiments, the sample is a protein sample.

Sequence data can be nucleic acid sequence data, which may be obtained by means known in the art. Sequence data is suitably obtained from a biological sample of genomic DNA, RNA, or cDNA (a "test sample") from an individual ("test subject). For example, nucleic acid sequence data may be obtained through direct analysis of the sequence of the polymorphic position (allele) of a polymorphic marker. Suitable methods, some of which are described herein, include, for instance, whole genome sequencing methods, whole genome analysis using SNP chips (e.g., Infinium HD BeadChip), cloning for polymorphisms, non-radioactive PCR-single strand conformation polymorphism analysis, denaturing high pressure liquid chromatography (DHPLC), DNA hybridization, computational analysis, single-stranded conformational polymorphism (SSCP), restriction fragment length polymorphism (RFLP), automated fluorescent sequencing; clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE), mobility shift analysis, restriction enzyme analysis; heteroduplex analysis, chemical mismatch cleavage (CMC), RNase protection assays, use of polypeptides that recognize nucleotide mismatches, such as E. coli mutS

protein, allele-specific PCR, and direct manual and automated sequencing. These and other methods are desribed in the art (see, for instance, Li et al., *Nucleic Acids Research,* 28(2): e1 (i-v) (2000); Liu et al., *Biochem Cell Bio* 80: 17-22 (2000); and Burczak et al., *Polymorphism Detection and Analysis,* Eaton Publishing, 2000; Sheffield et al., *Proc. Natl. Acad. Sci. USA,* 86:232-236 (1989); Orita et al., *Proc. Natl. Acad. Sci. USA,* 86:2766-2770 (1989); Flavell et al., *Cell,* 15:25-41 (1978); Geever et al., *Proc. Natl. Acad. Sci. USA,* 78:5081-5085 (1981); Cotton et al., *Proc. Natl. Acad. Sci. USA,* 85:4397-4401 (1985); Myers et al., *Science* 230: 1242-1246 (1985); Church and Gilbert, *Proc. Natl. Acad. Sci. USA,* 81: 1991-1995 (1988); Sanger et al., *Proc. Natl. Acad. Sci. USA,* 74:5463-5467 (1977); and Beavis et al., U.S. Patent No. 5,288,644).

Recent technological advances have resulted in technologies that allow massive parallel sequencing to be performed in relatively condensed format. These technologies share sequencing-by-synthesis principle for generating sequence information, with different technological solutions implemented for extending, tagging and detecting sequences. Exemplary technologies include 454 pyrosequencing technology (Nyren, P. *et al. Anal Biochem* 208: 171-75 (1993); http://www.454.com), Illumina Solexa sequencing technology (Bentley, D.R. *Curr Opin Genet Dev* 16:545-52 (2006); http://www.illumina.com), and the SOLiD technology developed by Applied Biosystems (ABI) (http://www.appliedbiosystems.com; see also Strausberg, R.L., *et al. Drug Disc Today* 13:569-77 (2008)). Other sequencing technologies include those developed by Pacific Biosciences (http://www.pacificbiosciences.com), Complete Genomics (http://www.completegenomics.com), Intelligen Bio-Systems (http://www.intelligentbiosystems.com), Genome Corp (http://www.genomecorp.com), ION Torrent Systems (http://www.iontorrent.com) and Helicos Biosciences (http://www.helicosbio.som). It is contemplated that sequence data useful for performing the present invention may be obtained by any such sequencing method, or other sequencing methods that are developed or made available. Thus, any sequence method that provides the allelic identity at particular polymorphic sites *{e.g.,* the absence or presence of particular alleles at particular polymorphic sites) is useful in the methods described and claimed herein.

Alternatively, hybridization methods may be used (see *Current Protocols in Molecular Biology,* Ausubel et al., eds., John Wiley & Sons, including all supplements). For example, a biological sample of genomic DNA, RNA, or cDNA (a "test sample") may be obtained from a test subject. The subject can be an adult, child, or fetus. The DNA, RNA, or cDNA sample is then examined. The presence of a specific marker allele can be indicated by sequence-specific hybridization of a nucleic acid probe specific for the particular allele. The presence of more than one specific marker allele or a specific haplotype can be indicated by using several sequence-specific nucleic acid probes, each being specific for a particular allele. A sequence-specific probe can be directed to hybridize to genomic DNA, RNA, or cDNA. A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe that hybridizes to a

complementary sequence. One of skill in the art would know how to design such a probe so that sequence specific hybridization will occur only if a particular allele is present in a genomic sequence from a test sample.

In certain embodiments, determination of a susceptibility to intracranial aneurysm and abdominal aortic aneurysm comprises forming a hybridization sample by contacting a test sample, such as a genomic DNA sample, with at least one nucleic acid probe. A non-limiting example of a probe for detecting mRNA or genomic DNA is a labeled nucleic acid probe that is capable of hybridizing to mRNA or genomic DNA sequences described herein. The nucleic acid probe can be, for example, a full-length nucleic acid molecule, or a portion thereof, such as an oligonucleotide of at least 10, 15, 30, 50, 100, 250 or 500 nucleotides in length that is sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. For example, the nucleic acid probe can comprise all or a portion of the nucleotide sequence of a sequence that comprises one of the markers described herein, or a sequence that is flanking one of the markers described herein, or the probe can be the complementary sequence of such a sequence. Hybridization can be performed by methods well known to the person skilled in the art (see, e.g., *Current Protocols in Molecular Biology,* Ausubel et al., eds., John Wiley & Sons, including all supplements). In one embodiment, hybridization refers to specific hybridization, i.e., hybridization with no mismatches (exact hybridization). In one embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is detected using standard methods. If specific hybridization occurs between the nucleic acid probe and the nucleic acid in the test sample, then the sample contains the allele that is complementary to the nucleotide that is present in the nucleic acid probe.

Allele-specific oligonucleotides can also be used to detect the presence of a particular allele in a nucleic acid. An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of any suitable size, for example an oligonucleotide of approximately 10-50 base pairs or approximately 15-30 base pairs, that specifically hybridizes to a nucleic acid which contains a specific allele at a polymorphic site (e.g., a polymorphic marker). An allele-specific oligonucleotide probe that is specific for one or more particular alleles at polymorphic markers can be prepared using standard methods (see, e.g., *Current Protocols in Molecular Biology, supra).* PCR can be used to amplify the desired region. Specific hybridization of an allele-specific oligonucleotide probe to DNA from a subject is indicative of the presence of a specific allele at a polymorphic site (see, e.g., Gibbs et al., *Nucleic Acids Res.* 17:2437-2448 (1989) and WO 93/22456).

In certain embodiments, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from a subject can be used to identify particular alleles in a nucleic acid. For example, an oligonucleotide array can be used. Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These arrays can generally be produced using

mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods, or by other methods known to the person skilled in the art

Suitable biological sample in the methods described herein can be any sample containing nucleic acid (e.g., genomic DNA) and/or protein from the human individual. For example, the biological sample can be a blood sample, a serum sample, a leukapheresis sample, an amniotic fluid sample, a cerbrospinal fluid sample, a hair sample, a tissue sample from skin, muscle, buccal, or conjuctival mucosa, placenta, gastrointestinal tract, or other organs, a semen sample, a urine sample, a saliva sample, a nail sample, a tooth sample, and the like. Preferably, the sample is a blood sample, a saliva sample or a buccal swab.

Indirect sequence analysis

It is possible to impute or predict genotypes for un-genotyped relatives of genotyped individuals. For every un-genotyped case, it is possible to calculate the probability of the genotypes of its relatives given its four possible phased genotypes. In practice it may be preferable to include only the genotypes of the case's parents, children, siblings, half-siblings (and the half-sibling's parents), grand-parents, grand-children (and the grand-children's parents) and spouses. It will be assumed that the individuals in the small sub-pedigrees created around each case are not related through any path not included in the pedigree. It is also assumed that alleles that are not transmitted to the case have the same frequency - the population allele frequency. The probability of the genotypes of the case's relatives can then be computed by:

$$\Pr(\text{genotypes of relatives};\Theta) = \sum_{he\{AA,\ AG,\ GA.GG\}} Pr(h;\Theta)\Pr(\text{genotypes of relatives I}h),$$

where Θ denotes the A allele's frequency in the cases. Assuming the genotypes of each set of relatives are independent, this allows us to write down a likelihood function for Θ:

$$L(\theta) = |^{}_{i}[\Pr(\text{genotypesof relativesof case}\, i;\Theta). \qquad (*)$$

This assumption of independence is usually not correct. Accounting for the dependence between individuals is a difficult and potentially prohibitively expensive computational task. The likelihood function in (*) may be thought of as a pseudolikelihood approximation of the full likelihood function for Θ which properly accounts for all dependencies. In general, the genotyped cases and controls in a case-control association study are not independent and applying the case-control method to related cases and controls is an analogous approximation. The method of genomic control (Devlin, B. et al., *Nat Genet 36,* 1129-30; author reply 1131 (2004)) has proven to be successful at adjusting case-control test statistics

for relatedness. We therefore apply the method of genomic control to account for the dependence between the terms in our pseudolikelihood and produce a valid test statistic.

Fisher's information can be used to estimate the effective sample size of the part of the pseudolikelihood due to un-genotyped cases. Breaking the total Fisher information, I, into the part due to genotyped cases, I_g, and the part due to ungenotyped cases, I_u, $I = I_g + I_u$, and denoting the number of genotyped cases with N, the effective sample size due to the un-genotyped cases is estimated by $\frac{I_u}{I_g} N$.

It is also possible to impute genotypes for markers with no genotype data. For example, using the IMPUTE software (Marchini, J. *et al*. *Nat Genet* 39:906-13 (2007)) and the HapMap CEU data (for example NCBI Build 36 (dbl26b)) as reference (Frazer, K.A., *et al*. *Nature* 449:851-61 (2007)) it is possible to impute genotypes for ungenotyped markers in a cohort, provided that the markers have been typed in the HapMap dataset. This can be useful for extending genotype coverage.

Protein analysis

Missense, nonsense and frameshift nucleic acid variations may lead to an altered amino acid sequence, as compared to the non-variant (e.g., wild-type) protein, due to amino acid substitutions, deletions, or insertions, or truncations. Variations at splice sites may also lead to splice variation. In such instances, detection of an amino acid substitution or a truncated amino acid sequence of the variant protein may be useful. Thus, nucleic acid sequence data may be obtained through indirect analysis of the nucleic acid sequence of the allele of the polymorphic marker, *i.e.* by detecting a protein variation.

The C allele of rs3798220 encodes an isoleucing to methionine substitution in the human Lp(a) protein, at position 4399 in Lp(a) (I4399M). Thus, this variant may suitably be detected through detection of the identity of the amino acid at position 4399 in an encoded Lp(a) protein.

Methods of detecting variant proteins are known in the art. For example, direct amino acid sequencing of the variant protein followed by comparison to a reference amino acid sequence can be used. Alternatively, SDS-PAGE followed by gel staining can be used to detect variant proteins of different molecular weights. Also, Immunoassays, e.g., antibody assays, e.g., immunofluorescent immunoassays, immunoprecipitations, radioimmunoasays, ELISA, and Western blotting, in which an antibody specific for an epitope comprising the variant sequence among the variant protein and non-variant or wild-type protein can be used. In certain embodiments, the amino acid sequence data is obtained or deduced from a preexisting record.

Number of Polymorphic Markers/Genes Analyzed

With regard to the methods of determining a susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease described herein, the methods can comprise obtaining sequence data about any number of polymorphic markers and/or about any number of genes. For example, the method can comprise obtaining sequence data for about at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 100, 500, 1000, 10,000 or more polymorphic markers. In certain embodiments, the sequence data is obtained from a microarray comprising probes for detecting a plurality of markers. In certain other embodiments, the sequence data is obtained through nucleic acid sequencing, for example by high-throughput nucleic acid sequencing. The sequence data may also be obtained by imputation of nucleic acid sequence using known methods, such as those described herein. The polymorphic markers can be the ones of the group specified herein or they can be different polymorphic markers that are not specified herein. In a specific embodiment, the method comprises obtaining sequence data about at least two polymorphic markers. In certain embodiments, each of the markers may be associated with a different gene. For example, in some instances, if the method comprises obtaining nucleic acid data about a human individual identifying at least one allele of a polymorphic marker, then the method comprises identifying at least one allele of at least one polymorphic marker. Also, for example, the method can comprise obtaining sequence data about a human individual identifying alleles of multiple, independent markers, which are not in linkage disequilibrium.

Linkage Disequilibrium

The natural phenomenon of recombination, which occurs on average once for each chromosomal pair during each meiotic event, represents one way in which nature provides variations in sequence (and biological function by consequence). It has been discovered that recombination does not occur randomly in the genome; rather, there are large variations in the frequency of recombination rates, resulting in small regions of high recombination frequency (also called recombination hotspots) and larger regions of low recombination frequency, which are commonly referred to as Linkage Disequilibrium (LD) blocks (Myers, S. *et al., Biochem Soc Trans* 34:526-530 (2006); Jeffreys, A.J., *et al., Nature Genet* 29:217-222 (2001); May, C.A., *et al., Nature Genet* 31:272-275(2002)).

Linkage Disequilibrium (LD) refers to a non-random assortment of two genetic elements. For example, if a particular genetic element (e.g., an allele of a polymorphic marker, or a haplotype) occurs in a population at a frequency of 0.50 (50%) and another element occurs at a frequency of 0.50 (50%), then the predicted occurrance of a person's having both elements is 0.25 (25%), assuming a random distribution of the elements. However, if it is discovered that the two elements occur together at a frequency higher than 0.25, then the

elements are said to be in linkage disequilibrium, since they tend to be inherited together at a higher rate than what their independent frequencies of occurrence *{e.g.,* allele or haplotype frequencies) would predict. Roughly speaking, LD is generally correlated with the frequency of recombination events between the two elements. Allele or haplotype frequencies can be determined in a population by genotyping individuals in a population and determining the frequency of the occurence of each allele or haplotype in the population. For populations of diploids, *e.g.,* human populations, individuals will typically have two alleles or allelic combinations for each genetic element (e.g., a marker, haplotype or gene).

Many different measures have been proposed for assessing the strength of linkage disequilibrium (LD; reviewed in Devlin, B. & Risch, N., *Genomics* 29:311-22 (1995))). Most capture the strength of association between pairs of biallelic sites. Two important pairwise measures of LD are r^2 (sometimes denoted Δ^2) and |D'| (Lewontin, R., *Genetics* 49:49-67 (1964); Hill, W.G. & Robertson, A. *Theor. Appl. Genet* 22:226-231 (1968)). Both measures range from 0 (no disequilibrium) to 1 ('complete' disequilibrium), but their interpretation is slightly different. |D'| is defined in such a way that it is equal to 1 if just two or three of the possible haplotypes are present, and it is < 1 if all four possible haplotypes are present. Therefore, a value of |D'| that is < 1 indicates that historical recombination may have occurred between two sites (recurrent mutation can also cause |D'| to be < 1, but for single nucleotide polymorphisms (SNPs) this is usually regarded as being less likely than recombination). The measure r^2 represents the statistical correlation between two sites, and takes the value of 1 if only two haplotypes are present.

The r^2 measure is arguably the most relevant measure for association mapping, because there is a simple inverse relationship between r^2 and the sample size required to detect association between susceptibility loci and SNPs. These measures are defined for pairs of sites, but for some applications a determination of how strong LD is across an entire region that contains many polymorphic sites might be desirable (e.g., testing whether the strength of LD differs significantly among loci or across populations, or whether there is more or less LD in a region than predicted under a particular model). Measuring LD across a region is not straightforward, but one approach is to use the measure r, which was developed in population genetics. Roughly speaking, r measures how much recombination would be required under a particular population model to generate the LD that is seen in the data. This type of method can potentially also provide a statistically rigorous approach to the problem of determining whether LD data provide evidence for the presence of recombination hotspots. For the methods described herein, a significant r^2 value can be at least 0.1 such as at least 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, or at lesat 0.99. In one preferred embodiment, the significant r^2 value can be at least 0.2. Alternatively, linkage disequilibrium as described herein, refers to linkage disequilibrium characterized by values of |D'| of at least 0.2, such as 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, or at least 0.99.

Thus, linkage disequilibrium represents a correlation between alleles of distinct markers. It is measured by correlation coefficient or |D'| (r^2 up to 1.0 and |D'| up to 1.0). In certain embodiments, linkage disequilibrium is defined in terms of values for both the r^2 and |D'| measures. In one such embodiment, a significant linkage disequilibrium is defined as $r^2 >$ 0.1 and |D'| >0.8. In another embodiment, a significant linkage disequilibrium is defined as $r^2 > 0.2$ and |D'| >0.9. Other combinations and permutations of values of r^2 and |D'| for determining linkage disequilibrium are also contemplated, and are also within the scope of the invention. Linkage disequilibrium can be determined in a single human population, as defined herein, or it can be determined in a collection of samples comprising individuals from more than one human population. In one embodiment of the invention, LD is determined in a sample from one or more of the HapMap populations (Caucasian, african, japanese, Chinese), as defined (http://www.hapmap.org). In one such embodiment, LD is determined in the CEU population of the HapMap samples. In another embodiment, LD is determined in the YRI population. In yet another embodiment, LD is determined in samples from the Icelandic population.

The following surrogate marker tables list markers that are correlated with rs3798220 and rs10455872 by values of $r^2 > 0.2$. These tables thus provide exemplary surrogate markers that are useful in the methods and other aspects of the invention as described in more detail herein.

Surrogate Marker Table. Surrogate markers on Chromosome 6 that are correlated by values of the correlation coefficient r^2>0.2 to anchor marker rs3798220 **(A)** and rs10455872 **(B)**. Correlation data was obtained from whole-genome sequencing data of over 700 Icelandic individuals. Shown is: marker identity, rs-name in dbSNP where available, position of that marker in NCBI Build 36 and Build 37, values of r^2, the predicted risk allele that is correlated with risk allele of the anchor-marker (allele C of rs3798220 and allele G of rs10455872), the other allele and finally sequence ID number of flanking sequence for the marker. Allelic codes are A = 1, C = 2, G = 3, T = 4.

(A)

Marker	dbSNP-134 name	Build36 position	Build37 position	r^2	Risk allele	Other allele	Seq ID NO:
chr6:159900330		159900330	159980339	0.265102	C	T	1
chr6:160210523		160210523	160290532	0.281774	T	C	2
chr6:160357007		160357007	160437016	0.294515	T	C	3
chr6:160535857		160535857	160615866	0.438387	G	A	4
chr6:160607402	rs10080815	160607402	160687411	0.961028	G	T	5
chr6:160671113	rs9295127	160671113	160751122	0.946698	A	C	6
chr6:160671521	rs9295128	160671521	160751530	0.501996	T	G	7
chr6:160679543	rs9457906	160679543	160759552	0.948117	G	C	8
chr6:160736399	rs10510226	160736399	160816408	0.999109	C	T	9
chr6:160762715	rs4708867	160762715	160842724	0.206531	G	A	10
chr6:160768733	rs9457925	160768733	160848742	1.000000	A	G	11

chr6:160773736	rs4709431	160773736	160853745	1.000000	C	A	12
chr6:160814868	rsl510224	160814868	160894877	1.000000	C	T	13
chr6:160830272	rs9457927	160830272	160910281	0.986791	G	A	14
chr6:160836048	rsl l7791490	160836048	160916057	1.000000	T	C	15
chr6:160842860	rsl l7733303	160842860	160922869	1.000000	G	A	16
chr6:160846072	rs4708870	160846072	160926081	0.202809	C	T	17
chr6:160849894	rs9457931	160849894	160929903	0.20 1164	G	A	18
chr6:160861631	rs7452960	160861631	160941640	0.578010	A	G	19
chr6:160881 127	**rs3798220**	160881 127	160961 136	1.000000	C	T	20
chr6:160887 186		160887 186	160967 195	0.463055	C	T	21
chr6:160907393		160907393	160987402	0.463434	G	A	22
chr6:160933003	rsl40570886	160933003	161013012	0.478270	C	T	23
chr6:161000289		161000289	161080298	0.461824	C	T	24
chr6:161073926		161073926	161 153935	0.385699	C	A	25
chr6:161288939	rsl50471561	161288939	161368948	0.202253	T	C	26
chr6:161297273	rs79784699	161297273	161377282	0.202981	A	G	27
chr6:161405938	rsl39068512	161405938	161485947	0.476171	T	C	28
chr6:161465216		161465216	161545225	0.385699	A	T	29
chr6:161470247	rs34749753	161470247	161550256	0.269796	C	T	30
chr6:161494549		161494549	161574558	0.375409	A	G	31
chr6:161606143		161606143	161686152	0.461824	C	T	32
chr6:161808714		161808714	161888723	0.213702	G	A	33

(B)

Marker	**dbSNP-134 name**	**Build36 position**	**Build37 position**	**r^2**	**Risk allele**	**Other allele**	**Seq ID NO:**
chr6:160623367	rs3127587	160623367	160703376	0.249508	A	G	34
chr6:160624902	rs3127586	160624902	16070491 1	0.248436	C	T	35
chr6:160630841	rsl41320342	160630841	160710850	0.247255	T	**c**	36
chr6:160631008	rs3125060	160631008	16071 1017	0.248334	A	G	37
chr6:160632801	rs56378 129	160632801	160712810	0.248520	A	G	38
chr6:160633289	rsl2193418	160633289	160713298	0.249486	G	C	39
chr6:160633514	rsl2206585	160633514	160713523	0.249603	A	G	40
chr6:160633686	rsl2203303	160633686	160713695	0.248008	C	T	41
chr6:160633694	rsl2194962	160633694	160713703	0.248234	T	**c**	42
chr6:1606341 17	rsl l755806	1606341 17	160714126	0.24701 1	G	T	43
chr6:160634593	rs78801493	160634593	160714602	0.248544	C	T	44
chr6:160634827	rs77410834	160634827	160714836	0.249651	T	**c**	45
chr6:160635083	rs2183470	160635083	160715092	0.249507	T	**c**	46
chr6:160636056	rs3103351	160636056	160716065	0.240761	G	A	47

chr6 :160636143	rs3127581	160636143	160716152	0.249684	C	T	48
chr6 :160637289	rs3106163	160637289	160717298	0.250722	C	T	49
chr6 :160637462	rs3120151	160637462	160717471	0.250221	T	**c**	50
chr6 :160638634	rs3127582	160638634	160718643	0.248397	A	G	51
chr6 :160638746	rs3106165	160638746	160718755	0.249547	C	T	52
chr6 :160639582	rs79390 162	160639582	160719591	0.249641	T	**c**	53
chr6 :160639731	rs932925	160639731	160719740	0.249641	A	G	54
chr6 :160642995	rs3103350	160642995	160723004	0.249143	C	T	55
chr6 :160643891	rsl l755965	160643891	160723900	0.249506	T	**c**	56
chr6 :160645759	rs3120149	160645759	160725768	0.249505	G	A	57
chr6 :160645885	rs3120148	160645885	160725894	0.249640	T	G	58
chr6 :160646334	rs3120147	160646334	160726343	0.248412	T	C	59
chr6 :160646357	rs3127585	160646357	160726366	0.249352	A	G	60
chr6 :160646422	rs3127584	160646422	160726431	0.249640	A	G	61
chr6 :160647047	rs3106175	160647047	160727056	0.249620	A	G	62
chr6 :160647 121	rs3127583	160647 121	160727 130	0.249620	A	G	63
chr6 :160647455	rs3120146	160647455	160727464	0.246640	T	C	64
chr6 :160647532	rs3106174	160647532	160727541	0.248759	A	G	65
chr6 :160647856	rs3120145	160647856	160727865	0.247376	G	A	66
chr6 :160647940	rs3106173	160647940	160727949	0.249141	G	A	67
chr6 :160647965	rs3106172	160647965	160727974	0.249141	G	A	68
chr6 :160648 131	rs3106171	160648 131	160728 140	0.248356	T	A	69
chr6 :160649288	rs76725059	160649288	160729297	0.250219	G	A	70
chr6 :160649914	rs3106170	160649914	160729923	0.249062	A	G	71
chr6 :160650795	rs3106169	160650795	160730804	0.249639	A	T	72
chr6 :160650930	rs3127598	160650930	160730939	0.249619	T	**c**	73
chr6 :160651094	rs3127597	160651094	160731 103	0.249586	A	T	74
chr6 :16065 1467	rs3120142	160651467	160731476	0.248025	G	A	75
chr6 :160651493	rs3106168	160651493	160731502	0.248543	G	A	76
chr6 :160651662	rs3125057	160651662	160731671	0.247815	A	T	77
chr6 :160651885	rsl2192401	160651885	160731894	0.252079	T	**c**	78
chr6 :160653245	rsl848830	160653245	160733254	0.247453	A	G	79
chr6 :160655271	rs3125056	160655271	160735280	0.246620	C	T	80
chr6 :160656777	rs3125055	160656777	160736786	0.249636	A	T	81
chr6 :160656787	rs3125054	160656787	160736796	0.249636	T	**c**	82
chr6 :160656930	rs3125053	160656930	160736939	0.248925	**c**	T	83
chr6 :160657380	rs3106167	160657380	160737389	0.248691	G	T	84
chr6 :160657792	rs3127591	160657792	160737801	0.249635	T	G	85
chr6 :160658363	rs3125052	160658363	160738372	0.249436	**c**	T	86
chr6 :160658720	rs3125051	160658720	160738729	0.248938	A	G	87
chr6 :160658821	rs3120140	160658821	160738830	0.249436	A	G	88

chr6:160660711	rs3103349	160660711	160740720	0.249822	T	C	89
chr6:160660826	rs3125050	160660826	160740835	0.233945	**c**	A	90
chr6:160661612	rs3120139	160661612	160741621	0.248341	A	G	91
chr6:160665228	rs3103348	160665228	160745237	0.248207	C	T	92
chr6:160666502	rs3103347	160666502	160746511	0.248341	T	**c**	93
chr6:160670459	rs6933264	160670459	160750468	0.241683	**c**	A	94
chr6:160670633	rs6911381	160670633	160750642	0.249962	T	C	95
chr6:160674702	rs3125049	160674702	160754711	0.234805	A	G	96
chr6:160676726	rsl2209517	160676726	160756735	0.229664	G	C	97
chr6:160691182	rs3120137	160691182	160771191	0.228464	A	G	98
chr6:160765025	rsl2209391	160765025	160845034	0.210616	A	G	99
chr6:160879821	rsl2194138	160879821	160959830	0.209031	T	A	100
chr6:160883220	rsl l751605	160883220	160963229	0.293830	**c**	T	101
chr6:160898987	rsl2210 186	160898987	160978996	0.206672	G	A	102
chr6:160905516	rsl l8039278	160905516	160985525	0.995528	A	G	103
chr6:160917108	rs74617384	160917108	160997117	0.989216	T	A	104
chr6:160925600	rs55730499	160925600	161005609	0.983647	T	C	105
chr6:160930108	**rsl0455872**	160930108	161010117	1.000000	G	A	106
chr6:160952214		160952214	161032223	0.214440	G	A	107
chr6:160952218		160952218	161032227	0.214847	C	T	108
chr6:160952233		160952233	161032242	0.201516	G	T	109
chr6:161006706	rs76735376	161006706	161086715	0.237739	A	G	110
chr6:161009297	rs56393506	161009297	161089306	0.262526	T	C	111
chr6:161023795		161023795	161103804	0.493952	A	T	112
chr6:161028134	rs2315065	161028134	161108143	0.604930	A	**c**	113
chr6:161043441	rs4252185	161043441	161123450	0.552536	C	T	114

Haplotype blocks (LD blocks) can be used to map associations between phenotype and haplotype status, using single markers or haplotypes comprising a plurality of markers. The main haplotypes can be identified in each haplotype block, and then a set of "tagging" SNPs or markers (the smallest set of SNPs or markers needed to distinguish among the haplotypes) can then be identified. These tagging SNPs or markers can then be used in assessment of samples from groups of individuals, in order to identify association between phenotype and haplotype. If desired, neighboring haplotype blocks can be assessed concurrently, as there may also exist linkage disequilibrium among the haplotype blocks.

Suitable surrogate markers may be selected using public information, such as from the International HapMap Consortium (http://www.hapmap.org) and the International 1000Genomes Consortium (http://www.1000genomes.org). The markers may also be suitably selected from results of whole-genome sequencing. The stronger the linkage disequilibrium (i.e., the higher the correlation) to the anchor marker, the better the

surrogate, and thus the mores similar the association detected by the surrogate is expected to be to the association detected by the anchor marker. Markers with values of r^2 equal to 1 are perfect surrogates for the at-risk variants, i.e. genotypes for one marker perfectly predicts genotypes for the other. In other words, the surrogate will, by necessity, give exactly the same association data to any particular disease as the anchor marker. Markers with smaller values of r^2 than 1 can also be surrogates for the at-risk anchor variant.

Association analysis

For single marker association to a disease, the Fisher exact test can be used to calculate two-sided p-values for each individual allele. Correcting for relatedness among patients can be done by extending a variance adjustment procedure previously described (Risch, N. & Teng, J. *Genome Res.,* 8:1273-1288 (1998)) for sibships so that it can be applied to general familial relationships. The method of genomic controls (Devlin, B. & Roeder, K. *Biometrics* 55:997 (1999)) can also be used to adjust for the relatedness of the individuals and possible stratification.

For both single-marker and haplotype analyses, relative risk (RR) and the population attributable risk (PAR) can be calculated assuming a multiplicative model (haplotype relative risk model) (Terwilliger, J.D. & Ott, J., *Hum. Hered.* 42:337-46 (1992) and Falk, C.T. & Rubinstein, P, *Ann. Hum. Genet. 51 (Pt* 3J:227-33 (1987)), i.e., that the risks of the two alleles/haplotypes a person carries multiply. For example, if RR is the risk of A relative to a, then the risk of a person homozygote AA will be RR times that of a heterozygote Aa and RR^2 times that of a homozygote aa. The multiplicative model has a nice property that simplifies analysis and computations — haplotypes are independent, *i.e.,* in Hardy-Weinberg equilibrium, within the affected population as well as within the control population. As a consequence, haplotype counts of the affecteds and controls each have multinomial distributions, but with different haplotype frequencies under the alternative hypothesis. Specifically, for two haplotypes, h_i and h_j, $\text{risk}(h_i)/\text{risk}(h_j) = (f_i/p_i)/(f_j/p_j)$, where f and p denote, respectively, frequencies in the affected population and in the control population. While there is some power loss if the true model is not multiplicative, the loss tends to be mild except for extreme cases. Most importantly, p-values are always valid since they are computed with respect to null hypothesis.

In the present context, an individual who is at an increased susceptibility (i.e., increased risk) for a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, is an individual in whom at least one specific allele at one or more polymorphic marker conferring increased susceptibility (increased risk) for the condition is identified. The at-risk marker allele is one that confers an increased risk (increased susceptibility) of the condition. In one embodiment, significance associated with a marker is measured by a relative risk (RR). In another embodiment, significance associated with a marker is

measured by an odds ratio (OR). In a further embodiment, the significance is measured by a percentage. In one embodiment, a significant increased risk is measured as a risk (relative risk and/or odds ratio) of at least 1.05, including but not limited to: at least 1.05, at least 1.06, at least 1.07, at least 1.08, at least 1.09, at least 1.10, at least 1.11, at least 1.12, at least 1.13, at least 1.14, at least 1.15, at least 1.16, at least 1.17, at least 1.18, at least 1.19, at least 1.20, at least 1.30, at least 1.40, at least 1.50, at least 1.60, at least 1.70, at least 1.80, at least 1.90, and at least 2.0. In a particular embodiment, a risk (relative risk and/or odds ratio) of at least 1.05 is significant. In another particular embodiment, a risk of at least 1.10 is significant. In yet another embodiment, a risk of at least 1.13 is significant. Other cutoffs are also contemplated, *e.g.,* at least 1.15, 1.25, 1.35, and so on, and such cutoffs are also within scope of the present invention. In other embodiments, a significant increase in risk is at least about 5%, including but not limited to at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 11%, at least about 12%, at least about 13%, at least about 14%, at least about 15%, at least about 16%, at least about 17%, at least about 18%, at least about 19%, at least about 20%, at least about 25%, at least about 30%, about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 60%, at least about 70%, and at least about 80%. Other cutoffs or ranges as deemed suitable by the person skilled in the art to characterize the invention are however also contemplated, and those are also within scope of the present invention. In certain embodiments, a significant increase in risk is characterized by a p-value, such as a p-value of less than 0.05, less than 0.01, less than 0.001, less than 0.0001, less than 0.00001, less than 0.000001, less than 0.0000001, less than 0.00000001, or less than 0.000000001.

An at-risk polymorphic markers (polymorphisms) as described herein is one where at least one allele of at least one polymorphic marker is more frequently present in an individual at risk for the condition (affected), or diagnosed with the condition, compared to the frequency of its presence in a comparison group (control), such that the presence of the marker or haplotype is indicative of susceptibility to the condition. The control group may in one embodiment be a population sample, i.e. a random sample from the general population. In another embodiment, the control group is represented by a group of individuals who are disease-free, *i.e.* have not been diagnosed with the condition or are do not have one or more specific disease-associated symptoms.

In other embodiments of the invention, an individual who is at a decreased susceptibility (i.e., at a decreased risk) for the condition is an individual in whom at least one specific marker allele conferring decreased susceptibility for the condition is identified. The marker alleles conferring decreased risk are also said to be protective. In one aspect, the protective marker allele is one that confers a significant decreased risk (or susceptibility) of the condition. In one embodiment, significant decreased risk is measured as a relative risk (or odds ratio) of less than 0.95, including but not limited to less than 0.9, less than 0.85, less than 0.8, less than 0.75, and less

than 0.7. In one particular embodiment, significant decreased risk is less than 0.9. In another embodiment, significant decreased risk is less than 0.85. In another embodiment, the decrease in risk (or susceptibility) is at least 5%, including but not limited to at least 10% at least 13%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, and at least 50%. In one particular embodiment, a significant decrease in risk is at least about 5%. In another embodiment, a significant decrease in risk is at least about 10%. In another embodiment, the decrease in risk is at least about 15%. Other cutoffs or ranges as deemed suitable by the person skilled in the art to characterize the invention are however also contemplated, and those are also within scope of the present invention.

The person skilled in the art will appreciate that for markers with two alleles present in the population being studied (such as SNPs), and wherein one allele is found in increased frequency in a group of individuals with a condition such as large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, compared with controls, the other allele of the marker will be found in decreased frequency in the group of individuals with the condition, compared with controls. In such a case, one allele of the marker (the one found in increased frequency in individuals with the condition) will be the at-risk allele, while the other allele will be a protective allele.

A genetic variant associated with a disease or condition can be used alone to predict the risk of the disease or condition for a given genotype. For a biallelic marker, such as a SNP, there are 3 possible genotypes: homozygote for the at risk variant, heterozygote, and non carrier of the at risk variant. Risk associated with variants at multiple loci can be used to estimate overall risk. For multiple SNP variants, there are k possible genotypes $k = 3^n \times 2^p$; where n is the number autosomal loci and p the number of gonosomal (sex chromosomal) loci. Overall risk assessment calculations for a plurality of risk variants usually assume that the relative risks of different genetic variants multiply, *i.e.* the overall risk (e.g., RR or OR) associated with a particular genotype combination is the product of the risk values for the genotype at each locus. If the risk presented is the relative risk for a person, or a specific genotype for a person, compared to a reference population with matched gender and ethnicity, then the combined risk - is the product of the locus specific risk values - and which also corresponds to an overall risk estimate compared with the population. If the risk for a person is based on a comparison to non-carriers of the at risk allele, then the combined risk corresponds to an estimate that compares the person with a given combination of genotypes at all loci to a group of individuals who do not carry risk variants at any of those loci. The group of non-carriers of any at risk variant has the lowest estimated risk and has a combined risk, compared with itself *{i.e.,* non-carriers) of 1.0, but has an overall risk, compare with the population, of less than 1.0. It should be noted that the group of non-carriers can potentially be very small, especially for large number of loci, and in that case, its relevance is correspondingly small.

The multiplicative model is a parsimonious model that usually fits the data of complex traits reasonably well. Deviations from multiplicity have been rarely described in the context of common variants for common diseases, and if reported are usually only suggestive since very large sample sizes are usually required to be able to demonstrate statistical interactions between loci.

Risk assessment and Diagnostics

Within any given population, there is an absolute risk of developing a disease or trait, defined as the chance of a person developing the specific disease or trait over a specified time-period. For example, a woman's lifetime absolute risk of breast cancer is one in nine. That is to say, one woman in every nine will develop breast cancer at some point in their lives. Risk is typically measured by looking at very large numbers of people, rather than at a particular individual. Risk is often presented in terms of Absolute Risk (AR) and Relative Risk (RR). Relative Risk is used to compare risks associating with two variants or the risks of two different groups of people. For example, it can be used to compare a group of people with a certain genotype with another group having a different genotype. For a disease, a relative risk of 2 means that one group has twice the chance of developing a disease as the other group. The risk presented is usually the relative risk for a person, or a specific genotype of a person, compared to the population with matched gender and ethnicity. Risks of two individuals of the same gender and ethnicity could be compared in a simple manner. For example, if, compared to the population, the first individual has relative risk 1.5 and the second has relative risk 0.5, then the risk of the first individual compared to the second individual is 1.5/0.5 = 3.

The creation of a model to calculate the overall genetic risk involves two steps: i) conversion of odds-ratios for a single genetic variant into relative risk and ii) combination of risk from multiple variants in different genetic loci into a single relative risk value.

Deriving risk from odds-ratios. Most gene discovery studies for complex diseases that have been published to date in authoritative journals have employed a case-control design because of their retrospective setup. These studies sample and genotype a selected set of cases (people who have the specified disease condition) and control individuals. The interest is in genetic variants (alleles) which frequency in cases and controls differ significantly.

The results are typically reported in odds-ratios, that is the ratio between the fraction (probability) with the risk variant (carriers) versus the non-risk variant (non-carriers) in the groups of affected versus the controls, i.e. expressed in terms of probabilities conditional on the affection status:

$$OR = (Pr(c|A)/Pr(nc|A)) \ / \ (Pr(c|C)/Pr(nc|C))$$

Sometimes it is however the absolute risk for the disease that we are interested in, i.e. the fraction of those individuals carrying the risk variant who get the disease or in other words the probability of getting the disease. This number cannot be directly measured in case-

control studies, in part, because the ratio of cases versus controls is typically not the same as that in the general population. However, under certain assumption, we can estimate the risk from the odds-ratio.

It is well known that under the rare disease assumption, the relative risk of a disease can be approximated by the odds-ratio. This assumption may however not hold for many common diseases. Still, it turns out that the risk of one genotype variant relative to another can be estimated from the odds-ratio expressed above. The calculation is particularly simple under the assumption of random population controls where the controls are random samples from the same population as the cases, including affected people rather than being strictly unaffected individuals. To increase sample size and power, many of the large genome-wide association and replication studies used controls that were neither age-matched with the cases, nor were they carefully scrutinized to ensure that they did not have the disease at the time of the study. Hence, while not exactly, they often approximate a random sample from the general population. It is noted that this assumption is rarely expected to be satisfied exactly, but the risk estimates are usually robust to moderate deviations from this assumption.

Calculations show that for the dominant and the recessive models, where we have a risk variant carrier, "c", and a non-carrier, "nc", the odds-ratio of individuals is the same as the risk-ratio between these variants:

$$OR = Pr(A|c)/Pr(A|nc) = r$$

And likewise for the multiplicative model, where the risk is the product of the risk associated with the two allele copies, the allelic odds-ratio equals the risk factor:

$$OR = Pr(A|aa)/Pr(A|ab) = Pr(A|ab)/Pr(A|bb) = r$$

Here "a" denotes the risk allele and "b" the non-risk allele. The factor "r" is therefore the relative risk between the allele types.

For many of the studies published in the last few years, reporting common variants associated with complex diseases, the multiplicative model has been found to summarize the effect adequately and most often provide a fit to the data superior to alternative models such as the dominant and recessive models.

The risk relative to the average population risk. It is most convenient to represent the risk of a genetic variant relative to the average population since it makes it easier to communicate the lifetime risk for developing the disease compared with the baseline population risk. For example, in the multiplicative model we can calculate the relative population risk for variant "aa" as:

$$RR(aa) = Pr(A|aa)/Pr(A) = (Pr(A|aa)/Pr(A|bb))/(Pr(A)/Pr(A|bb)) = r^2/(Pr(aa)\ r^2 + Pr(ab)\ r + Pr(bb)) = r^2/(p^2 r^2 + 2pq\ r + q^2) = r^2/R$$

Here "p" and "q" are the allele frequencies of "a" and "b" respectively. Likewise, we get that RR(ab) = r/R and RR(bb) = 1/R. The allele frequency estimates may be obtained from the publications that report the odds-ratios and from the HapMap database. Note that in the case where we do not know the genotypes of an individual, the relative genetic risk for that test or marker is simply equal to one.

Combining the risk from multiple markers. When genotypes of many SNP variants are used to estimate the risk for an individual, unless otherwise stated, a multiplicative model for risk can be assumed. This means that the combined genetic risk relative to the population is calculated as the product of the corresponding estimates for individual markers, e.g. for two markers gl and g2:

$$RR(g1,g2) = RR(g1)RR(g2)$$

The underlying assumption is that the risk factors occur and behave independently, i.e. that the joint conditional probabilities can be represented as products:

$$Pr(A|g1,g2) = Pr(A|g1)Pr(A|g2)/Pr(A) \quad \text{and} \quad Pr(g1,g2) = Pr(g1)Pr(g2)$$

Obvious violations to this assumption are markers that are closely spaced on the genome, i.e. in linkage disequilibrium such that the concurrence of two or more risk alleles is correlated. In such cases, we can use so called haplotype modeling where the odds-ratios are defined for all allele combinations of the correlated SNPs.

As is in most situations where a statistical model is utilized, the model applied is not expected to be exactly true since it is not based on an underlying bio-physical model. However, the multiplicative model has so far been found to fit the data adequately, i.e. no significant deviations are detected for many common diseases for which many risk variants have been discovered.

Adjusted life-time risk

The lifetime risk of an individual is derived by multiplying the overall genetic risk relative to the population with the average life-time risk of the disease in the general population of the same ethnicity and gender and in the region of the individual's geographical origin. As there are usually several epidemiologic studies to choose from when defining the general population risk, we will pick studies that are well-powered for the disease definition that has been used for the genetic variants.

For example, if the overall genetic risk relative to the population for a disease is 1.8 for a white male, and if the average life-time risk of the disease for individuals of his demographic is 20%, then the adjusted lifetime risk for him is 20% x 1.8 = 36%.

Note that since the average RR for a population is one, this multiplication model provides the same average adjusted life-time risk of the disease. Furthermore, since the actual life-time risk cannot exceed 100%, there must be an upper limit to the genetic RR.

Study population

In a general sense, the methods and kits of the invention can be utilized from samples containing nucleic acid material (DNA or RNA) from any source and from any individual, or from genotype data derived from such samples. In preferred embodiments, the individual is a human individual. The individual can be an adult, child, or fetus. The nucleic acid source may be any sample comprising nucleic acid material, including biological samples, or a sample comprising nucleic acid material derived therefrom. The present invention also provides for assessing markers and/or haplotypes in individuals who are members of a target population. Such a target population is in one embodiment a population or group of individuals at risk of developing the disease, based on other genetic factors, biomarkers *{e.g.,* leukotrienes), biophysical parameters, or general health and/or lifestyle parameters (e.g., history of a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease).

The invention provides for embodiments that include individuals from specific age subgroups, such as those over the age of 40, over age of 45, or over age of 50, 55, 60, 65, 70, 75, 80, or 85. Other embodiments of the invention pertain to other age groups, such as individuals aged less than 85, such as less than age 80, less than age 75, or less than age 70, 65, 60, 55, 50, 45, 40, 35, or age 30. Other embodiments relate to individuals with age at onset of a condition such as large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, coronary artery disease and severe coronary artery disease in any of the age ranges described in the above. In one embodiment of the invention, the polymorphic markers of the invention are predictive of increased susceptibility to coronary arter disease in individuals with onset of the disease before age 50 for males and before age 60 for females. It is also contemplated that a range of ages may be relevant in certain embodiments, such as age at onset at more than age 45 but less than age 60. Other age ranges are however also contemplated, including all age ranges bracketed by the age values listed in the above. The invention furthermore relates to individuals of either gender, males or females.

The Icelandic population is a Caucasian population of Northern European ancestry. A large number of studies reporting results of genetic linkage and association in the Icelandic population have been published in the last few years. Many of those studies show replication of variants, originally identified in the Icelandic population as being associating with a particular disease, in other populations (Styrkarsdottir, U., *et al*. *N Engl J Med* Apr 29 2008 (Epub ahead of print); Thorgeirsson, T., *et al*. *Nature* 452:638-42 (2008); Gudmundsson, J., *et al. Nat Genet.* 40:281-3 (2008); Stacey, S.N., *et al., Nat Genet.* 39:865-69 (2007); Helgadottir, A., *et al., Science* 316: 1491-93 (2007); Steinthorsdottir, V., *et al., Nat Genet.* 39:770-75 (2007); Gudmundsson, J., *et al., Nat Genet.* 39:631-37 (2007); Frayling, TM, *Nature Reviews Genet* 8:657-662 (2007); Amundadottir, L.T., *et al., Nat Genet.* 38:652-58

(2006); Grant, S.F., *et al., Nat Genet.* 38:320-23 (2006)). Thus, genetic findings in the Icelandic population have in general been replicated in other populations, including populations from Africa and Asia.

It is thus believed that the markers of the present invention will show similar association in other human populations. Particular embodiments comprising individual human populations are thus also contemplated and within the scope of the invention. Such embodiments relate to human subjects that are from one or more human population including, but not limited to, Caucasian populations, European populations, American populations, Eurasian populations, Asian populations, Central/South Asian populations, East Asian populations, Middle Eastern populations, African populations, Hispanic populations, and Oceanian populations. European populations include, but are not limited to, Swedish, Norwegian, Finnish, Russian, Danish, Icelandic, Irish, Kelt, English, Scottish, Dutch, Belgian, French, German, Spanish, Portuguese, Italian, Polish, Bulgarian, Slavic, Serbian, Bosnian, Czech, Greek and Turkish populations.

In certain embodiments, the invention relates to markers and/or haplotypes identified in specific populations, as described in the above. The person skilled in the art will appreciate that measures of linkage disequilibrium (LD) may give different results when applied to different populations. This is due to different population history of different human populations as well as differential selective pressures that may have led to differences in LD in specific genomic regions. It is also well known to the person skilled in the art that certain markers, *e.g.* SNP markers, have different population frequency in different populations, or are polymorphic in one population but not in another. The person skilled in the art will however apply the methods available and as thought herein to practice the present invention in any given human population. This may include assessment of polymorphic markers in the LD region of the present invention, so as to identify those markers that give strongest association within the specific population. Thus, the at-risk variants of the present invention may reside on different haplotype background and in different frequencies in various human populations. However, utilizing methods known in the art and the markers of the present invention, the invention can be practiced in any given human population.

Diagnostic Methods

Polymorphic markers associated with increased susceptibility of a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease are useful in diagnostic methods. While methods of diagnosing cardiovascular diseases are known in the art, the detection risk markers for the condition advantageously may be useful for detection of these condition at their early stages and may also reduce the occurrence of misdiagnosis. In this regard, the invention further provides methods of diagnosing a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, comprising obtaining sequence

data identifying at least one risk allele as described herein, in conjunction with carrying out one or more clinical diagnostic steps for the identification of the condition. The clinical diagnostics steps may include any such steps known to the skilled person and routinely used in the art.

In certain embodiments, a sample containing genomic DNA or protein from an individual is collected. Such sample can for example be a buccal swab, a saliva sample, a blood sample, or other suitable samples containing genomic DNA or protein, as described further herein. In certain embodiments, the sample is obtained by non-invasive means (e.g., for obtaining a buccal sample, saliva sample, hair sample or skin sample). In certain embodiments, the sample is obtained by non-surgical means, *i.e.* in the absence of a surgical intervention on the individual that puts the individual at substantial health risk. Such embodiments may, in addition to non-invasive means also include obtaining sample by extracting a blood sample (e.g., a venous blood sample). The genomic DNA or protein obtained from the individual is then analyzed using any common technique available to the skilled person, such as high-throughput technologies for genotyping and/or sequencing. Results from such methods are stored in a convenient data storage unit, such as a data carrier, including computer databases, data storage disks, or by other convenient data storage means. In certain embodiments, the computer database is an object database, a relational database or a post-relational database. The genotype data is subsequently analyzed for the presence of certain variants known to be susceptibility variants for a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease. Genotype and/or sequencing data can be retrieved from the data storage unit using any convenient data query method. Calculating risk conferred by a particular genotype for the individual can be based on comparing the genotype of the individual to previously determined risk (expressed as a relative risk (RR) or and odds ratio (OR), for example) for the genotype, for example for an heterozygous carrier of an at-risk variant. The calculated risk for the individual can be the relative risk for a person, or for a specific genotype of a person, compared to the average population with matched gender and ethnicity. The average population risk can be expressed as a weighted average of the risks of different genotypes, using results from a reference population, and the appropriate calculations to calculate the risk of a genotype group relative to the population can then be performed. Alternatively, the risk for an individual is based on a comparison of particular genotypes, for example heterozygous carriers of an at-risk allele of a marker compared with non-carriers of the at-risk allele. The calculated risk estimated can be made available to the customer via a website, preferably a secure website.

Kits

Kits useful in the methods of the invention comprise components useful in any of the methods described herein, including for example, primers for nucleic acid amplification,

hybridization probes, restriction enzymes (e.g., for RFLP analysis), allele-specific oligonucleotides, antibodies that bind to an altered polypeptide encoded by a nucleic acid of the invention as described herein (e.g., a genomic segment comprising at least one polymorphic marker of the present invention) or to a non-altered (native) polypeptide encoded by a nucleic acid of the invention as described herein, means for amplification of a nucleic acid associated with a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, means for analyzing the nucleic acid sequence of a nucleic acid associated with the condition, means for analyzing the amino acid sequence of a polypeptide encoded by a nucleic acid associated with the condition, etc. The kits can for example include necessary buffers, nucleic acid primers for amplifying nucleic acids of the invention (e.g., a nucleic acid segment comprising one or more of the polymorphic markers as described herein), and reagents for allele-specific detection of the fragments amplified using such primers and necessary enzymes (e.g., dna polymerase). Additionally, kits can provide reagents for assays to be used in combination with the methods of the present invention, e.g., reagents for use with other diagnostic assays for the condition.

In one embodiment, the invention pertains to a kit for assaying a sample from a subject to detect a susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a subject, wherein the kit comprises reagents necessary for selectively detecting at least one allele of at least one polymorphism of the present invention in the genome of the individual. In a particular embodiment, the reagents comprise at least one contiguous oligonucleotide that hybridizes to a fragment of the genome of the individual comprising at least one polymorphism of the present invention. In another embodiment, the reagents comprise at least one pair of oligonucleotides that hybridize to opposite strands of a genomic segment obtained from a subject, wherein each oligonucleotide primer pair is designed to selectively amplify a fragment of the genome of the individual that includes at least one polymorphism associated with risk of the condition. In one such embodiment, the polymorphism is selected from the group consisting of the markers described herein, and polymorphic markers in linkage disequilibrium therewith. In yet another embodiment the fragment is at least 20 base pairs in size. Such oligonucleotides or nucleic acids (e.g., oligonucleotide primers) can be designed using portions of the nucleic acid sequence flanking polymorphisms (e.g., SNPs or microsatellites) that are associated with risk of the condition. In another embodiment, the kit comprises one or more labeled nucleic acids capable of allele-specific detection of one or more specific polymorphic markers or haplotypes, and reagents for detection of the label. Suitable labels include, e.g., a radioisotope, a fluorescent label, an enzyme label, an enzyme co-factor label, a magnetic label, a spin label, an epitope label.

In particular embodiments, the polymorphic marker to be detected by the reagents of the kit comprises at least the marker rs3798220 and/or rsl0455872. The reagents of the kit may also be suitable for detecting a marker in linkage disequilibrium with rs3798220 and/or

rsl0455872. In another embodiment, the marker to be detected comprises at least one marker from the group of markers in strong linkage disequilibrium, as defined by values of r^2 greater than 0.2, to marker rs3798220 and/or rsl0455872. In another embodiment, the marker to be detected comprises at least one marker from the group of markers in strong linkage disequilibrium, as defined by values of r^2 greater than 0.5, to marker rs3798220 and/or rsl0455872.

In a preferred embodiment, the DNA template containing the SIMP polymorphism is amplified by Polymerase Chain Reaction (PCR) prior to detection, and primers for such amplification are included in the reagent kit. In such an embodiment, the amplified DNA serves as the template for the detection probe and the enhancer probe.

In one such embodiment, determination of the presence of a particular marker allele using the reagents of the kit is indicative of a susceptibility (increased susceptibility or decreased susceptibility) to the condition. In another embodiment, determination of the presence of a marker allele is indicative of response to a therapeutic agent for the condition. In another embodiment, the presence of the marker allele is indicative of prognosis of the condition. In yet another embodiment, the presence of the marker allele is indicative of progress of treatment of the condition. Such treatment may include intervention by surgery, medication or by other means (e.g., lifestyle changes).

In a further aspect of the present invention, a pharmaceutical pack (kit) is provided, the pack comprising a therapeutic agent and a set of instructions for administration of the therapeutic agent to humans diagnostically tested for one or more variants of the present invention, as disclosed herein. The therapeutic agent can be a small molecule drug, an antibody, a peptide, an antisense or rnai molecule, or other therapeutic molecules. In one embodiment, an individual identified as a carrier of at least one variant of the present invention is instructed to take a prescribed dose of the therapeutic agent. In one such embodiment, an individual identified as a homozygous carrier of at least one variant of the present invention (e.g., a homozygous carrier for allele C in rs3798220 and/or allele G in rsl0455872) is instructed to take a prescribed dose of the therapeutic agent. In another embodiment, an individual identified as a non-carrier of at least one variant of the present invention is instructed to take a prescribed dose of the therapeutic agent.

In certain embodiments, the kit further comprises a set of instructions for using the reagents comprising the kit. In certain embodiments, the kit further comprises a collection of data comprising correlation data between the polymorphic markers assessed by the kit and susceptibility to the condition.

Antisense agents

The nucleic acids and/or variants described herein, or nucleic acids comprising their complementary sequence, may be used as antisense constructs to control gene expression in cells, tissues or organs. The methodology associated with antisense techniques is well known

to the skilled artisan, and is for example described and reviewed in *AntisenseDrug Technology: Principles, Strategies, and Applications,* Crooke, ed., Marcel Dekker Inc., New York (2001). In general, antisense agents (antisense oligonucleotides) are comprised of single stranded oligonucleotides (RNA or DNA) that are capable of binding to a complimentary nucleotide segment. By binding the appropriate target sequence, an RNA-RNA, DNA-DNA or RNA-DNA duplex is formed. The antisense oligonucleotides are complementary to the sense or coding strand of a gene. It is also possible to form a triple helix, where the antisense oligonucleotide binds to duplex DNA.

Several classes of antisense oligonucleotide are known to those skilled in the art, including cleavers and blockers. The former bind to target RNA sites, activate intracellular nucleases *{e.g.,* RnaseH or Rnase L), that cleave the target RNA. Blockers bind to target RNA, inhibit protein translation by steric hindrance of the ribosomes. Examples of blockers include nucleic acids, morpholino compounds, locked nucleic acids and methylphosphonates (Thompson, *Drug Discovery Today,* 7:912-917 (2002)). Antisense oligonucleotides are useful directly as therapeutic agents, and are also useful for determining and validating gene function, for example by gene knock-out or gene knock-down experiments. Antisense technology is further described in Lavery *et* a/., *Curr. Opin. Drug Discov. Devel.* 6:561-569 (2003), Stephens *et al., Curr. Opin. Mol. Ther.* 5:118-122 (2003), Kurreck, *Eur. J. Biochem.* 270: 1628-44 (2003), Dias *et al., Mol. Cancer Ter.* 1:347-55 (2002), Chen, *Methods Mol. Med.* 75:621-636 (2003), Wang *et al., Curr. Cancer Drug Targets* 1:177-96 (2001), and Bennett, *Antisense Nucleic Acid Drug. Dev.* 12:215-24 (2002).

Antisense nucleotides can be from 5-500 nucleotides in length, including 5-200 nucleotides, 5-100 nucleotides, 10-50 nucleotides, and 10-30 nucleotides. In certain preferred embodiments, the antisense nucleotides is from 14-50 nucleotides in length, includign 14-40 nucleotides and 14-30 nucleotides. In certain such embodiments, the antisense nucleotide is capable of binding to a nucleotide segment with sequence as set forth in any one of SEQ ID N0: l-114.

The variants described herein can also be used for the selection and design of antisense reagents that are specific for particular variants. Using information about the variants described herein, antisense oligonucleotides or other antisense molecules that specifically target mRNA molecules that contain one or more variants of the invention can be designed. In this manner, expression of mRNA molecules that contain one or more variant of the present invention (markers and/or haplotypes) can be inhibited or blocked. In one embodiment, the antisense molecules are designed to specifically bind a particular allelic form (i.e., one or several variants (alleles and/or haplotypes)) of the target nucleic acid, thereby inhibiting translation of a product originating from this specific allele or haplotype, but which do not bind other or alternate variants at the specific polymorphic sites of the target nucleic acid molecule. As antisense molecules can be used to inactivate mRNA so as to inhibit gene expression, and thus protein expression, the molecules can be used for disease treatment. The methodology can involve cleavage by means of ribozymes containing nucleotide

sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Such mRNA regions include, for example, protein-coding regions, in particular protein-coding regions corresponding to catalytic activity, substrate and/or ligand binding sites, or other functional domains of a protein.

Clinical applications of RNAi include the incorporation of synthetic siRNA duplexes, which preferably are approximately 20-23 nucleotides in size, and preferably have 3' overlaps of 2 nucleotides. Knockdown of gene expression is established by sequence-specific design for the target mRNA. Several commercial sites for optimal design and synthesis of such molecules are known to those skilled in the art.

Other applications provide longer siRNA molecules (typically 25-30 nucleotides in length, preferably about 27 nucleotides), as well as small hairpin RNAs (shRNAs; typically about 29 nucleotides in length). The latter are naturally expressed, as described in Amarzguioui *et al.* *{FEBS Lett.* 579:5974-81 (2005)). Chemically synthetic siRNAs and shRNAs are substrates for *in vivo* processing, and in some cases provide more potent gene-silencing than shorter designs (Kim *et al., Nature Biotechnol.* 23:222-226 (2005); Siolas *et al., Nature Biotechnol.* 23:227-231 (2005)). In general siRNAs provide for transient silencing of gene expression, because their intracellular concentration is diluted by subsequent cell divisions. By contrast, expressed shRNAs mediate long-term, stable knockdown of target transcripts, for as long as transcription of the shRNA takes place (Marques *et al., Nature Biotechnol.* 23:559-565 (2006); Brummelkamp *et al., Science* 296: 550-553 (2002)).

Since RNAi molecules, including siRNA, miRNA and shRNA, act in a sequence-dependent manner, the variants presented herein can be used to design RNAi reagents that recognize specific nucleic acid molecules comprising specific alleles and/or haplotypes *{e.g.,* the alleles and/or haplotypes of the present invention), while not recognizing nucleic acid molecules comprising other alleles or haplotypes. These RNAi reagents can thus recognize and destroy the target nucleic acid molecules. As with antisense reagents, RNAi reagents can be useful as therapeutic agents (i.e., for turning off disease-associated genes or disease-associated gene variants), but may also be useful for characterizing and validating gene function *{e.g.,* by gene knock-out or gene knock-down experiments).

Delivery of RNAi may be performed by a range of methodologies known to those skilled in the art. Methods utilizing non-viral delivery include cholesterol, stable nucleic acid-lipid particle (SNALP), heavy-chain antibody fragment (Fab), aptamers and nanoparticles. Viral delivery methods include use of lentivirus, adenovirus and adeno-associated virus. The siRNA molecules are in some embodiments chemically modified to increase their stability. This can include modifications at the 2' position of the ribose, including 2'-0-methylpurines and 2'-fluoropyrimidines, which provide resistance to Rnase activity. Other chemical modifications are possible and known to those skilled in the art.

The following references provide a further summary of RNAi, and possibilities for targeting specific genes using RNAi: Kim & Rossi, *Nat. Rev. Genet.* 8:173-184 (2007), Chen &

Rajewsky, *Nat. Rev. Genet.* 8: 93-103 (2007), Reynolds, *et al., Nat. Biotechnol.* 22:326-330 (2004), Chi *et al., Proc. Natl. Acad. Sci. USA* 100:6343-6346 (2003), Vickers *et al., J. Biol. Chem.* 278:7108-7118 (2003), Agami, *Curr. Opin. Chem. Biol.* 6:829-834 (2002), Lavery, *et al., Curr. Opin. Drug Discov. Devel.* 6:561-569 (2003), Shi, *Trends Genet.* 19:9-12 (2003), Shuey *et al., Drug Discov. Today* 7:1040-46 (2002), McManus *et al., Nat. Rev. Genet.* 3:737-747 (2002), Xia *et al., Nat. Biotechnol.* 20: 1006-10 (2002), Plasterk *et al., curr. Opin. Genet. Dev.* 10:562-7 (2000), Bosher *et al., Nat. Cell Biol.* 2:E31-6 (2000), and Hunter, *Curr. Biol.* 9:R440-442 (1999).

Methods of assessing probability of response to therapeutic agents, methods of monitoring progress of treatment and methods of treatment

As is known in the art, individuals can have differential responses to a particular therapy (e.g., a therapeutic agent or therapeutic method). Pharmacogenomics addresses the issue of how genetic variations *{e.g.,* the variants (markers and/or haplotypes) of the present invention) affect drug response, due to altered drug disposition and/or abnormal or altered action of the drug. Thus, the basis of the differential response may be genetically determined in part. Clinical outcomes due to genetic variations affecting drug response may result in toxicity of the drug in certain individuals *{e.g.,* carriers or non-carriers of the genetic variants of the present invention), or therapeutic failure of the drug. Therefore, the variants of the present invention may determine the manner in which a therapeutic agent and/or therapeutic method acts on the body, or the way in which the body metabolizes the therapeutic agent.

Accordingly, in one embodiment, the presence of a particular at-risk allele for a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease is indicative of a different response, *e.g.* a different response rate, to a particular treatment modality for the condition. This means that a patient diagnosed with the condition, and carrying a certain allele at a polymorphic of the present invention *(e.g.,* the at-risk alleles allele C in rs3798220 and/or allele G in rsl0455872) would respond better to, or worse to, a specific therapeutic, drug and/or other therapy used to treat the condition. Therefore, the presence or absence of the marker allele could aid in deciding what treatment should be used for the patient. For example, for a newly diagnosed patient, the presence of a marker of the present invention may be assessed *(e.g.,* through testing DNA derived from a blood sample, as described herein). If the patient is positive for a marker allele, then the physician recommends one particular therapy, while if the patient is negative for the at least one allele of a marker, or a haplotype, then a different course of therapy may be recommended (which may include recommending that no immediate therapy, other than serial monitoring for progression of the condition, be performed). Thus, the patient's carrier status could be used to help determine whether a particular treatment modality should be administered. The value lies within the possibilities of being able to diagnose the disease at an early stage, to

select the most appropriate treatment, and provide information to the clinician about prognosis/aggressiveness of the disease in order to be able to apply the most appropriate treatment. The treatment modality may be any suitable treatment modality used for treating the condition. In certain embodiments, the treatment modality is surgical intervention.

The present invention also relates to methods of monitoring progress or effectiveness of a treatment for a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease. This can be done based on the genotype status (i.e., the presence or absence of particular alleles) of an individual for the at-risk alleles for the condition described herein, i.e., by assessing the absence or presence of at least one allele of at least one polymorphic marker as disclosed herein. Alternatively, or concomitantly, the genotype status of at least one risk variant for the condition as presented herein is determined before and during treatment to monitor its effectiveness.

In a further aspect, the markers and haplotypes of the present invention can be used for targeting the selection of pharmaceutical agents for specific individuals. Personalized selection of treatment modalities, lifestyle changes or combination of lifestyle changes and administration of particular treatment, can be realized by the utilization of the at-risk variants of the present invention. Thus, the knowledge of an individual's status for particular risk alleles for a cardiovascular condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, as described herein, can be useful for selection of appropriate treatment options.

Nucleic acids and polypeptides

The nucleic acids and polypeptides described herein can be used in methods and kits of the present invention. An "isolated" nucleic acid molecule, as used herein, is one that is separated from nucleic acids that normally flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (e.g., as in an RNA library). For example, an isolated nucleic acid of the invention can be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material can be purified to essential homogeneity, for example as determined by polyacrylamide gel electrophoresis (PAGE) or column chromatography (e.g., HPLC). An isolated nucleic acid molecule of the invention can comprise at least about 50%, at least about 80% or at least about 90% (on a molar basis) of all macromolecular species present. With regard to genomic DNA, the term "isolated" also can refer to nucleic acid molecules that are separated from the chromosome with which the

genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 250 kb, 200 kb, 150 kb, 100 kb, 75 kb, 50 kb, 25 kb, 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of the nucleotides that flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

The nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of "isolated" as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells or heterologous organisms, as well as partially or substantially purified DNA molecules in solution. "Isolated" nucleic acid molecules also encompass *in vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleotide sequence can include a nucleic acid molecule or nucleotide sequence that is synthesized chemically or by recombinant means. Such isolated nucleotide sequences are useful, for example, in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (e.g., from other mammalian species), for gene mapping (e.g., by *in situ* hybridization with chromosomes), or for detecting expression of the gene in tissue (e.g., human tissue), such as by Northern blot analysis or other hybridization techniques.

The invention also pertains to nucleic acid molecules that hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (e.g., nucleic acid molecules that specifically hybridize to a nucleotide sequence containing a polymorphic site associated with a marker or haplotype described herein). Such nucleic acid molecules can be detected and/or isolated by allele- or sequence-specific hybridization (e.g., under high stringency conditions). Stringency conditions and methods for nucleic acid hybridizations are well known to the skilled person (see, e.g., *Current Protocols in Molecular Biology,* Ausubel, F. *et al,* John Wiley & Sons, (1998), and Kraus, M. and Aaronson, S., *Methods Enzymol.,* 200:546-556 (1991), the entire teachings of which are incorporated by reference herein.

The percent identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first sequence). The nucleotides or amino acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions x 100). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%, of the length of the reference sequence. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A non-limiting example of such a mathematical algorithm is described in Karlin, S. and Altschul, S., *Proc. Natl. Acad. Sci. USA,* 90:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0), as described in Altschul, S. *et al., Nucleic Acids Res.,* 25:3389-3402

(1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. See the website on the world wide web at ncbi.nlm.nih.gov. In one embodiment, parameters for sequence comparison can be set at score= 100, word length = 12, or can be varied (e.g., W=5 or W=20). Another example of an algorithm is BLAT (Kent, W.J. *Genome Res.* **12**:656-64 (2002)).

The present invention also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleic acid that comprises, or consists of, the nucleotide sequence of at least one polymorphic allele contained in the markers and haplotypes described herein. The nucleic acid fragments of the invention are at least about 15, at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200, 500, 1000, 10,000 or more nucleotides in length. In certain embodiments, the nucleic acid fragments are from about 15 to about 1000 nucleotides in length. In certain other embodiments, the nucleic acid fragments are from about 18 to about 100 nucleotides in length, from about 12 to about 50 nucleotides in length, from about 12 to about 40 nucleotides in length, or from about 12 to about 30 nucleotides in length.

The present invention further provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleic acid that comprises, or consists of, the nucleotide sequence of any one of SEQ ID NO: 1 - 114, as described herein. The nucleic acid fragments can be from 10-600 nucleotides in length, such as from 10 - 500 nucleotides, 12 - 200 nucleotides, 12 - 100 nucleotides, 12 - 50 nucleotides and 12 - 30 nucleotides in length.

The nucleic acid fragments of the invention may be used as probes or primers in assays such as those described herein. "Probes" or "primers" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of a nucleic acid molecule. In addition to DNA and RNA, such probes and primers include polypeptide nucleic acids (PIMA), as described in Nielsen, P. *et al., Science 254:*1497- 1500 (1991). A probe or primer comprises a region of nucleotide sequence that hybridizes to at least about 15, typically about 20-25, and in certain embodiments about 40, 50 or 75, consecutive nucleotides of a nucleic acid molecule. In one embodiment, the probe or primer comprises at least one allele of at least one polymorphic marker or at least one haplotype described herein, or the complement thereof. In particular embodiments, a probe or primer can comprise 100 or fewer nucleotides; for example, in certain embodiments from 6 to 50 nucleotides, or, for example, from 12 to 30 nucleotides. In other embodiments, the probe or primer is at least 70% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence. In another embodiment, the probe or primer is capable of selectively hybridizing to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence. Often, the probe or primer further comprises a label, e.g., a radioisotope, a fluorescent label, an enzyme label, an enzyme co-factor label, a magnetic label, a spin label, an epitope label.

The nucleic acid molecules of the invention, such as those described above, can be identified and isolated using standard molecular biology techniques well known to the skilled person. The amplified DNA can be labeled (e.g., radiolabeled, fluorescently labeled) and used as a probe for screening a cDNA library derived from human cells. The cDNA can be derived from mRNA and contained in a suitable vector. Corresponding clones can be isolated, DNA obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art- recognized methods to identify the correct reading frame encoding a polypeptide of the appropriate molecular weight. Using these or similar methods, the polypeptide and the DNA encoding the polypeptide can be isolated, sequenced and further characterized.

Computer-implemented aspects

As understood by those of ordinary skill in the art, the methods and information described herein may be implemented, in all or in part, as computer executable instructions on known computer readable media. For example, the methods described herein may be implemented in hardware. Alternatively, the method may be implemented in software stored in, for example, one or more memories or other computer readable medium and implemented on one or more processors. As is known, the processors may be associated with one or more controllers, calculation units and/or other units of a computer system, or implanted in firmware as desired. If implemented in software, the routines may be stored in any computer readable memory such as in RAM, ROM, flash memory, a magnetic disk, a laser disk, or other storage medium, as is also known. Likewise, this software may be delivered to a computing device via any known delivery method including, for example, over a communication channel such as a telephone line, the Internet, a wireless connection, etc., or via a transportable medium, such as a computer readable disk, flash drive, etc.

More generally, and as understood by those of ordinary skill in the art, the various steps described above may be implemented as various blocks, operations, tools, modules and techniques which, in turn, may be implemented in hardware, firmware, software, or any combination of hardware, firmware, and/or software. When implemented in hardware, some or all of the blocks, operations, techniques, etc. may be implemented in, for example, a custom integrated circuit (IC), an application specific integrated circuit (ASIC), a field programmable logic array (FPGA), a programmable logic array (PLA), etc.

When implemented in software, the software may be stored in any known computer readable medium such as on a magnetic disk, an optical disk, or other storage medium, in a RAM or ROM or flash memory of a computer, processor, hard disk drive, optical disk drive, tape drive, etc. Likewise, the software may be delivered to a user or a computing system via any known delivery method including, for example, on a computer readable disk or other transportable computer storage mechanism.

Thus, another aspect of the invention is a system that is capable of carrying out a part or all of a method of the invention, or carrying out a variation of a method of the invention as described herein in greater detail. Exemplary systems include, as one or more components, computing systems, environments, and/or configurations that may be suitable for use with the methods and include, but are not limited to, personal computers, server computers, hand-held or laptop devices, multiprocessor systems, microprocessor-based systems, set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like. In some variations, a system of the invention includes one or more machines used for analysis of biological material (e.g., genetic material), as described herein. In some variations, this analysis of the biological material involves a chemical analysis and/or a nucleic acid amplification.

With reference to Fig. 1, an exemplary system of the invention, which may be used to implement one or more steps of methods of the invention, includes a computing device in the form of a computer 110. Components shown in dashed outline are not technically part of the computer 110, but are used to illustrate the exemplary embodiment of Fig. 1. Components of computer 110 may include, but are not limited to, a processor 120, a system memory 130, a memory/graphics interface 121, also known as a Northbridge chip, and an I/O interface 122, also known as a Southbridge chip. The system memory 130 and a graphics processor 190 may be coupled to the memory/graphics interface 121. A monitor 191 or other graphic output device may be coupled to the graphics processor 190.

A series of system busses may couple various system components including a high speed system bus 123 between the processor 120, the memory/graphics interface 121 and the I/O interface 122, a front-side bus 124 between the memory/graphics interface 121 and the system memory 130, and an advanced graphics processing (AGP) bus 125 between the memory/graphics interface 121 and the graphics processor 190. The system bus 123 may be any of several types of bus structures including, by way of example, and not limitation, such architectures include Industry Standard Architecture (ISA) bus, Micro Channel Architecture (MCA) bus and Enhanced ISA (EISA) bus. As system architectures evolve, other bus architectures and chip sets may be used but often generally follow this pattern. For example, companies such as Intel and AMD support the Intel Hub Architecture (IHA) and the Hypertransport™ architecture, respectively.

The computer 110 typically includes a variety of computer-readable media. Computer-readable media can be any available media that can be accessed by computer 110 and includes both volatile and nonvolatile media, removable and non-removable media. By way of example, and not limitation, computer readable media may comprise computer storage media. Computer storage media includes both volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or

other memory technology, CD-ROM, digital versatile disks (DVD) or other optical disk storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other physical medium which can be used to store the desired information and which can accessed by computer 110.

The system memory 130 includes computer storage media in the form of volatile and/or nonvolatile memory such as read only memory (ROM) 131 and random access memory (RAM) 132. The system ROM 131 may contain permanent system data 143, such as identifying and manufacturing information. In some embodiments, a basic input/output system (BIOS) may also be stored in system ROM 131. RAM 132 typically contains data and/or program modules that are immediately accessible to and/or presently being operated on by processor 120. By way of example, and not limitation, Fig. 1 illustrates operating system 134, application programs 135, other program modules 136, and program data 137.

The I/O interface 122 may couple the system bus 123 with a number of other busses 126, 127 and 128 that couple a variety of internal and external devices to the computer 110. A serial peripheral interface (SPI) bus 126 may connect to a basic input/output system (BIOS) memory 133 containing the basic routines that help to transfer information between elements within computer 110, such as during start-up.

A super input/output chip 160 may be used to connect to a number of 'legacy' peripherals, such as floppy disk 152, keyboard/mouse 162, and printer 196, as examples. The super I/O chip 160 may be connected to the I/O interface 122 with a bus 127, such as a low pin count (LPC) bus, in some embodiments. Various embodiments of the super I/O chip 160 are widely available in the commercial marketplace.

In one embodiment, bus 128 may be a Peripheral Component Interconnect (PCI) bus, or a variation thereof, may be used to connect higher speed peripherals to the I/O interface 122. A PCI bus may also be known as a Mezzanine bus. Variations of the PCI bus include the Peripheral Component Interconnect-Express (PCI-E) and the Peripheral Component Interconnect - Extended (PCI-X) busses, the former having a serial interface and the latter being a backward compatible parallel interface. In other embodiments, bus 128 may be an advanced technology attachment (ATA) bus, in the form of a serial ATA bus (SATA) or parallel ATA (PATA).

The computer 110 may also include other removable/non-removable, volatile/nonvolatile computer storage media. By way of example only, Fig. 1 illustrates a hard disk drive 140 that reads from or writes to non-removable, nonvolatile magnetic media. The hard disk drive 140 may be a conventional hard disk drive.

Removable media, such as a universal serial bus (USB) memory 153, firewire (IEEE 1394), or CD/DVD drive 156 may be connected to the PCI bus 128 directly or through an interface 150. A storage media 154 may be coupled through interface 150. Other removable/ non-removable, volatile/nonvolatile computer storage media that can be used in the exemplary

operating environment include, but are not limited to, magnetic tape cassettes, flash memory cards, digital versatile disks, digital video tape, solid state RAM, solid state ROM, and the like.

The drives and their associated computer storage media discussed above and illustrated in Fig. 1, provide storage of computer readable instructions, data structures, program modules and other data for the computer 110. In Fig. 1, for example, hard disk drive 140 is illustrated as storing operating system 144, application programs 145, other program modules 146, and program data 147. Note that these components can either be the same as or different from operating system 134, application programs 135, other program modules 136, and program data 137. Operating system 144, application programs 145, other program modules 146, and program data 147 are given different numbers here to illustrate that, at a minimum, they are different copies. A user may enter commands and information into the computer 20 through input devices such as a mouse/keyboard 162 or other input device combination. Other input devices (not shown) may include a microphone, joystick, game pad, satellite dish, scanner, or the like. These and other input devices are often connected to the processor 120 through one of the I/O interface busses, such as the SPI 126, the LPC 127, or the PCI 128, but other busses may be used. In some embodiments, other devices may be coupled to parallel ports, infrared interfaces, game ports, and the like (not depicted), via the super I/O chip 160.

The computer 110 may operate in a networked environment using logical connections to one or more remote computers, such as a remote computer 180 via a network interface controller (NIC) 170. The remote computer 180 may be a personal computer, a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements described above relative to the computer 110. The logical connection between the NIC 170 and the remote computer 180 depicted in Fig. 1 may include a local area network (LAN), a wide area network (WAN), or both, but may also include other networks. Such networking environments are commonplace in offices, enterprise-wide computer networks, intranets, and the Internet. The remote computer 180 may also represent a web server supporting interactive sessions with the computer 110, or in the specific case of location-based applications may be a location server or an application server.

In some embodiments, the network interface may use a modem (not depicted) when a broadband connection is not available or is not used. It will be appreciated that the network connection shown is exemplary and other means of establishing a communications link between the computers may be used.

In some variations, the invention is a system for identifying susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject. For example, in one variation, the system includes tools for performing at least one step, preferably two or more steps, and in some aspects all steps of a method of

the invention, where the tools are operably linked to each other. Operable linkage describes a linkage through which components can function with each other to perform their purpose.

In some variations, a system of the invention is a system for identifying susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject, and comprises:

(a) at least one processor;

(b) at least one computer-readable medium;

(c) a susceptibility database operatively coupled to a computer-readable medium of the system and containing population information correlating the presence or absence of one or more alleles of at least one polymorphic marker selected from the group consisting of allele C in rs3798220 and allele G in rsl0455872, and marker alleles in linkage disequilibirium therewith susceptibility to the condition in a population of humans;

(d) a measurement tool that receives an input about the human subject and generates information from the input about the presence or absence of the at least one allele in the human subject; and

(e) an analysis tool or routine that:

(i) is operatively coupled to the susceptibility database and the information generated by the measurement tool,

(ii) is stored on a computer-readable medium of the system,

(iii) is adapted to be executed on a processor of the system, to compare the information about the human subject with the population information in the susceptibility database and generate a conclusion with respect to susceptibility to the condition for the human subject.

Exemplary processors (processing units) include all variety of microprocessors and other processing units used in computing devices. Exemplary computer-readable media are described above. When two or more components of the system involve a processor or a computer-readable medium, the system generally can be created where a single processor and/or computer readable medium is dedicated to a single component of the system; or where two or more functions share a single processor and/or share a single computer readable medium, such that the system contains as few as one processor and/or one computer readable medium. In some variations, it is advantageous to use multiple processors or media, for example, where it is convenient to have components of the system at different locations. For instance, some components of a system may be located at a testing laboratory dedicated to laboratory or data analysis, whereas other components, including components (optional) for supplying input information or obtaining an output communication, may be located at a medical treatment or counseling facility (e.g., doctor's

office, health clinic, HMO, pharmacist, geneticist, hospital) and/or at the home or business of the human subject (patient) for whom the testing service is performed.

Referring to Fig. 2, an exemplary system includes a susceptibility database 208 that is operatively coupled to a computer-readable medium of the system and that contains population information correlating the presence or absence of at least one allele of at least one polymorphic marker selected from the group consisting of rs3798220 and marker rsl0455872, and correlated markers in linkage disequilibrium therewith and susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a population of humans.

In a simple variation, the susceptibility database contains 208 data relating to the frequency that a particular allele of the at least one marker has been observed in a population of humans with the condition and a population of humans free of the condition. Such data provides an indication as to the relative risk or odds ratio of developing the condition for a human subject that is identified as having the allele in question. In another variation, the susceptibility database includes similar data with respect to two or more alleles of at least one marker, thereby providing a useful reference if the human subject has any of the two or more alleles. In still another variation, the susceptibility database includes additional quantitative personal, medical, or genetic information about the individuals in the database diagnosed with the condition or free of the condition. Such information includes, but is not limited to, information about parameters such as age, sex, ethnicity, race, medical history, weight, diabetes status, blood pressure, blood lipids, family history of the condition, smoking history, and alcohol use in humans and impact of the at least one parameter on susceptibility to the condition. The information also can include information about other genetic risk factors for the condition besides these aforementioned markers. These more robust susceptibility databases can be used by an analysis routine 210 to calculate a combined score with respect to susceptibility or risk for developing the condition.

In addition to the susceptibility database 208, the system further includes a measurement tool 206 programmed to receive an input 204 from or about the human subject and generate an output that contains information about the presence or absence of the at least one allele of interest. (The input 204 is not part of the system per se but is illustrated in the schematic Figure 2.) Thus, the input 204 will contain a specimen or contain data from which the presence or absence of the at least one allele can be directly read, or analytically determined. In a simple variation, the input contains annotated information about genotypes or allele counts for the markers predictive of risk of the condition in the genome of the human subject, in which case no further processing by the measurement tool 206 is required, except possibly transformation of the relevant information about the presence/absence of the at least one allele into a format compatible for use by the analysis routine 210 of the system.

In another variation, the input 204 from the human subject contains data that is unannotated or insufficiently annotated with respect to the at least one marker, requiring analysis by the measurement tool 206. For example, the input can be genetic sequence of a chromosomal region or chromosome on which the marker resides, or whole genome sequence information, or unannotated information from a gene chip analysis of a variable loci in the human subject's genome. In such variations of the invention, the measurement tool 206 comprises a tool, preferably stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to receive a data input about a subject and determine information about the presence or absence of the at least one mutant allele in a human subject from the data. For example, the measurement tool 206 contains instructions, preferably executable on a processor of the system, for analyzing the unannotated input data and determining the presence or absence of the allele of interest in the human subject. In one variation, the input data is genomic sequence information, and the measurement tool optionally comprises a sequence analysis tool stored on a computer readable medium of the system and executable by a processor of the system with instructions for determining the presence or absence of the at least one mutant allele from the genomic sequence information.

In yet another variation, the input 204 from the human subject comprises a biological sample, such as a fluid (e.g., blood) or tissue sample, that contains genetic material that can be analyzed to determine the presence or absence of the allele of interest. In this variation, an exemplary measurement tool 206 includes laboratory equipment for processing and analyzing the sample to determine the presence or absence (or identity) of the allele(s) in the human subject. For instance, in one variation, the measurement tool includes: an oligonucleotide microarray (e.g., "gene chip") containing a plurality of oligonucleotide probes attached to a solid support; a detector for measuring interaction between nucleic acid obtained from or amplified from the biological sample and one or more oligonucleotides on the oligonucleotide microarray to generate detection data; and an analysis tool stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to determine the presence or absence of the at least one allele of interest based on the detection data.

To provide another example, in some variations the measurement tool 206 includes: a nucleotide sequencer (e.g., an automated DNA sequencer) that is capable of determining nucleotide sequence information from nucleic acid obtained from or amplified from the biological sample; and an analysis tool stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to determine the presence or absence of the at least one mutant allele based on the nucleotide sequence information.

In some variations, the measurement tool 206 further includes additional equipment and/or chemical reagents for processing the biological sample to purify and/or amplify nucleic acid of the human subject for further analysis using a sequencer, gene chip, or other analytical equipment.

The exemplary system further includes an analysis tool or routine 210 that: is operatively coupled to the susceptibility database 208 and operatively coupled to the measurement tool 206, is stored on a computer-readable medium of the system, is adapted to be executed on a processor of the system to compare the information about the human subject with the population information in the susceptibility database 208 and generate a conclusion with respect to susceptibility to the condition for the human subject. In simple terms, the analysis tool 210 looks at the alleles identified by the measurement tool 206 for the human subject, and compares this information to the susceptibility database 208, to determine a susceptibility to the condition for the subject. The susceptibility can be based on the single parameter (the identity of one or more alleles of one or more markers predictive of risk of the condition), or can involve a calculation based on other genetic and non-genetic data, as described above, that is collected and included as part of the input 204 from the human subject, and that also is stored in the susceptibility database 208 with respect to a population of other humans. Generally speaking, each parameter of interest is weighted to provide a conclusion with respect to susceptibility to the condition. Such a conclusion is expressed in the conclusion in any statistically useful form, for example, as an odds ratio, a relative risk, or a lifetime risk for subject developing the condition.

In some variations of the invention, the system as just described further includes a communication tool 212. For example, the communication tool is operatively connected to the analysis routine 210 and comprises a routine stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to: generate a communication containing the conclusion; and to transmit the communication to the human subject 200 or the medical practitioner 202, and/or enable the subject or medical practitioner to access the communication. (The subject and medical practitioner are depicted in the schematic Fig. 2, but are not part of the system per se, though they may be considered users of the system. The communication tool 212 provides an interface for communicating to the subject, or to a medical practitioner for the subject (e.g., doctor, nurse, genetic counselor), the conclusion generated by the analysis tool 210 with respect to susceptibility to the condition for the subject. Usually, if the communication is obtained by or delivered to the medical practitioner 202, the medical practitioner will share the communication with the human subject 200 and/or counsel the human subject about the medical significance of the communication. In some variations, the communication is provided in a tangible form, such as a printed report or report stored on a computer readable medium such as a flash drive or optical disk. In some variations, the communication is provided electronically with an output that is visible on a video display or audio output (e.g., speaker). In some variations, the communication is transmitted to the subject or the medical practitioner, e.g., electronically or through the mail. In some variations, the system is designed to permit the subject or medical practitioner to access the communication, e.g., by telephone or computer. For instance, the system may include software residing on a memory and executed by a processor of a computer used by the human subject or the medical practitioner, with which

the subject or practitioner can access the communication, preferably securely, over the internet or other network connection. In some variations of the system, this computer will be located remotely from other components of the system, e.g., at a location of the human subject's or medical practitioner's choosing.

In some variations of the invention, the system as described (including embodiments with or without the communication tool) further includes components that add a treatment or prophylaxis utility to the system. For instance, value is added to a determination of susceptibility to the condition when a medical practitioner can prescribe or administer a standard of care that can reduce susceptibility to the condition; and/or delay onset of the condition; and/or increase the likelihood of detecting the condition at an early stage. Exemplary lifestyle change protocols include loss of weight, increase in exercise, cessation of unhealthy behaviors such as smoking, and change of diet. Exemplary medicinal and surgical intervention protocols include administration of pharmaceutical agents for prophylaxis; and surgery. Exemplary diagnostic protocols include non-invasive and invasive imaging; monitoring metabolic biomarkers; and biopsy screening.

For example, in some variations, the system further includes a medical protocol database 214 operatively connected to a computer-readable medium of the system and containing information correlating the presence or absence of the at least one marker allele of interest and medical protocols for human subjects at risk for the condition. Such medical protocols include any variety of medicines, lifestyle changes, diagnostic tests, increased frequencies of diagnostic tests, and the like that are designed to achieve one of the aforementioned goals. The information correlating a marker allele with protocols could include, for example, information about the success with which the condition or its symptoms is avoided or delayed, or success with which the condition is detected early and treated, if a subject has a marker susceptibility allele and follows a protocol.

The system of this embodiment further includes a medical protocol tool or routine 216, operatively connected to the medical protocol database 214 and to the analysis tool or routine 210. The medical protocol tool or routine 216 preferably is stored on a computer-readable medium of the system, and adapted to be executed on a processor of the system, to: (i) compare (or correlate) the conclusion that is obtained from the analysis routine 210 (with respect to susceptibility to the condition for the subject) and the medical protocol database 214, and (ii) generate a protocol report with respect to the probability that one or more medical protocols in the medical protocol database will achieve one or more of the goals of reducing susceptibility to the condition; delaying onset of the condition; and increasing the likelihood of detecting the condition at an early stage to facilitate early treatment. The probability can be based on empirical evidence collected from a population of humans and expressed either in absolute terms (e.g., compared to making no intervention), or expressed in relative terms, to highlight the comparative or additive benefits of two or more protocols.

Some variations of the system just described include the communication tool 212. In some examples, the communication tool generates a communication that includes the protocol report in addition to, or instead of, the conclusion with respect to susceptibility.

Information about allele status of risk markers of the condition not only can provide useful information about identifying or quantifying susceptibility to the condition; it can also provide useful information about possible causative factors for a human subject identified with the condition, and useful information about therapies for the patient. In some variations, systems of the invention are useful for these purposes.

For instance, in some variations the invention is a system for assessing or selecting a treatment protocol for a subject diagnosed with a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease. An exemplary system, schematically depicted in Figure 3, comprises:

(a) at least one processor;

(b) at least one computer-readable medium;

(c) a medical treatment database 308 operatively connected to a computer-readable medium of the system and containing information correlating the presence or absence of at least one marker allele selected from the group consisting of allele C in rs3798220 and/or allele G in rsl0455872, and marker alleles correlated therewith, and efficacy of treatment regimens for the condition;

(d) a measurement tool 306 to receive an input (304, depicted in Fig. 3 but not part of the system per se) about the human subject and generate information from the input 304 about the presence or absence of the at least one marker allele in a human subject diagnosed with the condition; and

(e) a medical protocol routine or tool 310 operatively coupled to the medical treatment database 308 and the measurement tool 306, stored on a computer-readable medium of the system, and adapted to be executed on a processor of the system, to compare the information with respect to presence or absence of the at least one marker allele for the subject and the medical treatment database, and generate a conclusion with respect to at least one of:

(i) the probability that one or more medical treatments will be efficacious for treatment of the condition for the patient; and

(ii) which of two or more medical treatments for the condition will be more efficacious for the patient.

Preferably, such a system further includes a communication tool 312 operatively connected to the medical protocol tool or routine 310 for communicating the conclusion to the subject 300, or to a medical practitioner for the subject 302 (both depicted in the schematic of Fig. 3, but not part of the system per se). An exemplary communication tool comprises a routine stored

on a computer-readable medium of the system and adapted to be executed on a processor of the system, to generate a communication containing the conclusion; and transmit the communication to the subject or the medical practitioner, or enable the subject or medical practitioner to access the communication.

The present invention will now be exemplified by the following non-limiting examples.

EXAMPLE 1

The association of LPA SNPs with vascular diseases was analysed in samples from 35 case-control studies including ischemic stroke (effective sample size = Ne case = Ne control = 9,396), PAD (Ne = 5,222), AAA (Ne = 4,586), VTE (Ne = 4,607), IA (Ne = 1,328), CAD (Ne = 12,716), as well as in 3,724 subjects with carotid IMT measurements. The association with CAD severity was also tested in 5,588 subjects who had undergone coronary angiography and with age at onset of CAD in 9,608 cases.

Associations between each SNP and the vascular diseases are shown in Table 1 A-D. As the estimated effects of the two SNPs were not significantly different, and the mechanisms by which both SNPs affect Lp(a) have been suggested to be similar [8], for reasons of simplicity (and to increase power) the main analyses (Figure 4-5 and Tables 2-4) report the effect estimates for the two SNPs combined (LPA score), which assumes that their effects are equal. LPA score results for each study group are detailed in Table 5, A-C. There was little evidence for heterogeneity in ORs across the studies for each of the vascular diseases, apart from the VTE studies (P=0.047), and across the two CAD studies (P=0.0029).

Results from all studies combined showed significant association between the LPA score and ischemic stroke, with an estimated OR of 1.10 (95% CI 1.02-1.18, P=0.016) (Figure 4). Ischemic stroke subtype analysis revealed that the association was restricted to LAA (OR=1.27, 95% CI 1.11-1.46, $P = 6.7 \times 10^{-4}$), as the effects of LPA variants on the CE and SVD subtypes were insignificant (P>0.50). The LPA score associated both with PAD (OR = 1.47 (95% CI 1.33-1.63, $P = 2.2 \times 10^{-14}$)) and with AAA (OR = 1.23 (95% CI 1.11-1.36, P=5.5 xlO $^{-5}$). The previously reported association between CAD and LPA score8 in individuals of European origin, was confirmed (OR = 1.32 (95% CI 1.24-1.42), $P = 1.1 \times 10^{-15}$). In addition, the LPA score was associated with CAD in African Americans (OR = 2.49 (95% CI 1.08 - 5.72, P = 0.032) (Table 6). In contrast, the LPA score did not associate with VTE (P=0.63), IA (P=0. 15) (Figure 4), or mean CCA IMT (Table 2) (P=0.093).

Given the substantial overlap between CAD and the other atherosclerotic diseases, it is conceivable that the association of the LPA variants with LAA stroke, PAD, and AAA is mediated through (or inflated by) the association of these diseases with CAD. Therefore, a re-analysis was conducted on the association between the LPA variants and the atherosclerotic phenotypes in cases with and without history of CAD in studies where this information was available (Table 7, A and B). After exclusion of CAD, the effect estimates for LPA score became lower for PAD (OR=1. 18, P=0. 10) and AAA (OR= l.ll, P=0.14), suggesting a stronger association of the LPA score with atherosclerotic disease manifested in more than one vascular bed, although the effect estimate was not lowered for LAA (OR= 1.30, P=0.013) (Table 7A). To investigate this further the Icelandic CAD cases were divided into those with confirmed atherosclerosis

related disease in other vascular beds (PAD, AAA, or LAA), and those without known history of these diseases. While, for CAD with concomitant PAD, AAA, or LAA the OR was 1.54 ($P = 8.9 \times 10^{-9}$), the OR for CAD without known PAD, AAA, or LAA the OR was 1.22 ($P = 2.4 \times 10^{-7}$). Direct comparison of the two groups of CAD cases shows that each LPA risk allele increased the risk of atherosclerosis in other vascular beds by 26% ($P = 0.0010$) (Figure 5).

Given that the data support stronger association with a more widespread atherosclerosis among CAD cases an investigation was performed on whether the LPA variants also associated with the angiographic extent of CAD in 5,048 individuals of European origin and 540 African Americans from Iceland and Atlanta, USA, who had undergone coronary angiography. The extent of CAD was expressed as the number of obstructed vessels (>50% stenosis) out of the 4 coronary index vessels (the left main artery, left anterior descending, left circumflex artery, and right coronary artery). After adjusting for gender, age at angiography, study site and ethnicity, in a linear regression model, each LPA risk allele increased the number of diseased vessels by 0.267 ($P = 4.8 \times 10^{-12}$) (Table 3). Similarly, the LPA score correlated with Gensini and Duke severity scores in the Atlanta-USA study (Figure 6). Effect estimates for African Americans were similar to those of European origin (Table 3 and Figure 6).

Finally, it was investigated whether there was a correlation between LPA score and the age of diagnosis of CAD (Table 4 and Figure 5). After adjusting for gender, each LPA risk allele associated with 1.58 years earlier diagnosis of CAD ($P = 8.2 \times 10^{-8}$) amongst 9,276 cases of European origin from Iceland and Atlanta, USA (Table 6). Restricting to cases with MI diagnoses the corresponding reduction in age of diagnosis of first MI was 0.92 years ($P = 0.028$). Effect estimates for age at first CAD diagnosis amongst 330 African Americans were consistent with those observed for whites, but did not reach nominal significance levels.

Discussion

An association has been found between two Lp(a) related SNPs (combined and examined as *LPA* score), and PAD, AAA, and the LAA subtype of ischemic stroke, with each variant allele increasing the risk by 46%, 24%, and 27% for the three diseases, respectively. In contrast, analyses show a clear lack of association of the *LPA* variants with VTE or with CE and SVD stroke subtypes. It is shown that CAD cases carrying *LPA* risk alleles have increased susceptibility to atherosclerotic manifestations outside the coronary tree and are more likely to be diagnosed earlier with CAD, than CAD cases not carrying these variants. Further, the study provides evidence for an association between the *LPA* variants and the number of coronary arteries (out of 4 index vessels) with ≥50% stenosis. Thus the *LPA* variants, and by implication Lp(a) levels, have an

effect on the atherosclerotic burden of large vessels throug hout the arteria l tree. The relationship between genetic varia nts affecting Lp(a) levels, and either systemic distribution of atherosclerosis, or angiog raphic CAD severity, has not been reported before. However, few stud ies, which in general have been limited by sma ll sample sizes, have assessed the association between Lp(a) levels and severity of CAD and provided conflicting results [27-31].

In apparent contrast to the correlation between *LPA* score and earlier age at onset of CAD, *LPA* score did not associate with IMT, a validated marker of subclinica l early atherosclerosis [32, 33]. This is consistent with results from a recent popu lation-based cohort study, showing that Lp(a) levels (and SNPs correlated with Lp(a)), do not associate with IMT[34]. As increased IMT is believed to represent arteria l remodelling and very early atherosclerotic cha nges, our results suggests that the effects of the *LPA* varia nts, and thus Lp(a) levels, on the risk of atherosclerotic manifestations are med iated at later stages in the pathogenic pathway, and are not reflected by variation in ca rotid IMT.

The current study is limited by the fact that measurements of Lp(a) levels were not ava ila ble, rendering it impossible to show directly that the association of the *LPA* risk varia nts with atherosclerotic phenotypes is med iated through Lp(a) levels. Therefore, the possibility that the risk conferred by the *LPA* score is med iated through mecha nisms other tha n increased Lp(a) levels, ca nnot be disrega rded. However, previous stud ies have shown a strong and consistent association between the two *LPA* variants and Lp(a) levels[8, 35], and the association between the *LPA* risk varia nts and CAD has been shown to disappea r with Lp(a) level adj ustments [8], strong ly supporting the view that the effect on atherosclerotic phenotypes is med iated through Lp(a) levels.

Lp(a) has been implicated in both atherogenesis and thrombosis. The *LPA* risk alleles that correlate with increased plasma Lp(a) levels, associate with vascu lar diseases with a strong atherosclerotic component, such as CAD, PAD and LAA, and AAA, but fa il to demonstrate an association with diseases without an atherosclerotic aetiology, such as VTE and IA, or with diseases less related to atherosclerosis, such as the CE and SVD stroke subtypes. Of the vascu la r diseases exa mined in this study, CAD, PAD and LAA are almost solely due to atherosclerosis. In add ition, AAA has been considered to be a manifestation of adva nced atherosclerosis, as atherosclerosis is very common both in the aneurysma l wall and other circu latory beds of AAA patients [36]; althoug h it has been debated whether it is causa lly related to the aneurysms [37]. The ischemic stroke subtypes, CE and SVD are less linked to atherosclerosis as the majority of CE strokes are secondary to atria l fibrillation and the SVD subtype is believed to result from a non-atherothrombotic obstruction of small arteries deep with in the bra in[38]. VTE is a venous thrombotic disease in which atherosclerosis is not a causative factor althoug h some stud ies have shown a link between arteria l and venous disorders [39]. Given the assumption that the risk of *LPA* variants on vascula r diseases is

mediated through Lp(a) levels, the presented results suggest that the implicit thrombogenic properties of Lp(a) [6, 7], are insufficient to increase the risk of the thrombotic stroke subtypes CE and SVD, or promoting VTE.

Evidence for a role of circulating Lp(a) in various vascular diseases, apart from CAD, has been previously reported. A meta-analysis of prospective studies reported association of Lp(a) concentration with ischemic stroke [9], and a meta-analysis of case-control studies demonstrated a correlation between small apo(a) isoforms and increased risk of ischemic stroke [10]. In contrast, a recent study found no relationship between the *LPA* variants, examined here, and ischemic stroke [11] after excluding cases with CAD. The current study, that is based on larger sample sizes than used in these previous studies [9-11], demonstrates that the association between *LPA* variant alleles and ischemic stroke is confined to the LAA stroke subtype. This is consistent with results from the Atherosclerosis Risk In Communities (ARIC) study that indicated that levels of plasma Lp(a) associated with non-lacunar stroke, but not with lacunar stroke (i.e. SVD), or cardioembolic stroke [40], and with results from another small study [41] that reported association between Lp(a) levels and the LAA stroke subtype, but not with the SVD subtype. Previous evidence for the association between circulating Lp(a) and PAD, AAA, and VTE, are mostly derived from relatively small case control studies, that have reported association for PAD [12-14], AAA [14, 15], or VTE [20], but results from prospective studies with relatively few outcomes are inconsistent [16-19].

Table 1-A. Association of rsl0455872[G], and rs3798220[C], with ischemic stroke

		rsl0455872[G]				rs3798220[C]			
	control/case	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***
Ischemic stroke									
Iceland	25617/2476	0.057	0.062	1.090	0.207	0.017	0.015	0.913	0.480
Sweden	718/868	0.063	0.069	1.105	0.489	0.012	0.019	1.624	0.100
Germany-S	1193/1185	0.062	0.068	1.105	0.397	0.013	0.014	1.149	0.581
Germany-W	1057/1387	0.061	0.077	1.281	0.031	0.015	0.015	1.029	0.905
UK-SE	639/650	0.076	0.089	1.188	0.237	0.017	0.025	1.475	0.162
UK-OXVASC	2246/1095	0.077	0.077	0.991	0.930	0.016	0.016	0.976	0.908
Combined	**31470/7661**			**1.129 (1.031-1.235)**	**0.0087**			**1.053 (0.893-1.242)**	**0.540**
Heterogeneity					0.86				0.40

Tables 1-A to 1-D:

An association test was not done for rs3798220[C] in the VTE Spain-GAITI study.

LAA=large artery atherosclerosis, CE=cardioembolism, SVD=small vessel disease, PAD=peripheral arterial disease, AAA=abdominal aortic aneurysm, VTE=venous thromboembolism, CAD=coronary artery disease, IA= intracranial aneurysm.

frq_con :allelic frequency in controls, frq_case: allelic frequency in cases, OR: odds ratio, 95% CI: 95% confidence interval.

*In UK-OXVASC study none of the LAA cases carried rs3798220[C]. For the calculation of combined OR of rs3798220[C] for LAA an OR of 0.00001 was used for the UK-OXVASC study.

Table 1-B. Association of rsl0455872[G], and rs3798220[C], with ischemic stroke TOAST subtypes

	rsl0455872[G]					rs3798220[C]			
	control/case	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***
LAA									
Iceland	25617/245	0.057	0.06	1.062	0.768	0.017	0.018	1.061	0.876
Sweden	718/175	0.063	0.095	1.573	0.039	0.012	0.023	2.024	0.121
Germany-S	1193/374	0.061	0.067	1.104	0.565	0.012	0.015	1.268	0.520
Germany-W	1057/557	0.061	0.080	1.329	0.050	0.015	0.012	0.810	0.522
UK-SE	639/231	0.075	0.121	1.711	0.004	0.016	0.028	1.826	0.126
UK-OXVASC*	2246/111	0.077	0.093	1.193	0.473	0.016	0	0	0.009
Combined	**31470/1693**			**1.309 (1.127-1.520)**	**4.3×10^{-4}**			**1.218 (0.874-1.696)**	**0.240**
Heterogeneity					0.42				0.049
CE									
Iceland	25617/403	0.057	0.067	1.188	0.255	0.017	0.012	0.730	0.333
Sweden	718/152	0.063	0.046	0.719	0.245	0.012	0.016	1.357	0.562
Germany-S	1193/297	0.061	0.070	1.150	0.452	0.012	0.012	1.016	0.970
Germany-W	1057/553	0.061	0.077	1.293	0.079	0.015	0.015	1.006	0.984
UK-SE	639/76	0.075	0.079	1.073	0.827	0.016	0.034	2.125	0.170
UK-OXVASC	2246/220	0.077	0.066	0.847	0.396	0.016	0.019	1.143	0.726
Combined	**31470/1701**			**1.099 (0.945-1.279)**	**0.220**			**1.044 (0.763-1.429)**	**0.790**
Heterogeneity					0.35				0.67

SVD									
Iceland	25617/260	0.057	0.044	0.764	0.206	0.017	0.021	1.258	0.475
Sweden	718/74	0.063	0.047	0.742	0.442	0.012	0.021	1.707	0.427
Germany-S	1193/117	0.061	0.075	1.244	0.424	0.012	0.017	1.492	0.481
Germany-W	1057/275	0.061	0.067	1.124	0.554	0.015	0.020	1.369	0.406
UK-SE	639/114	0.075	0.079	1.064	0.818	0.016	0.027	1.654	0.311
UK-OXVASC	2246/183	0.077	0.072	0.93	0.733	0.016	0.029	1.764	0.121
Combined	**31470/1023**			**0.973 (0.803-1.181)**	**0.780**			**1.479 (1.054-2.076)**	**0.023**
Heterogeneity					0.66				0.99

Table 1-C. Association of rsl0455872[G], and rs3798220[C], with peripheral arterial disease and abdominal aortic aneurysm

		rsl0455872[G]				rs3798220[C]			
	control/case	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***
PAD									
Iceland	25617/1602	0.057	0.074	1.337	4.5×10^{-4}	0.017	0.024	1.482	0.006
Denmark	4506/464	0.071	0.107	1.590	0.0001	0.012	0.017	1.554	0.123
Sweden	143/206	0.115	0.116	0.997	0.990	0.012	0.005	0.400	0.312
Austria	433/487	0.033	0.081	2.586	9.6×10^{-6}	0.014	0.015	1.130	0.767
Italy	203/176	0.044	0.048	1.099	0.789	0.002	0.014	5.801	0.061
USA-PA	260/434	0.054	0.085	1.636	0.028	0.021	0.020	1.086	0.840
New Zealand	526/465	0.064	0.108	1.777	6.6×10^{-4}	0.016	0.021	1.348	0.391
Combined	**31688/3834**			**1.487 (1.332-1.659)**	**1.4×10^{-12}**			**1.418 (1.144-1.757)**	**0.0014**
Heterogeneity					0.037				0.52
AAA									
Iceland	25617/420	0.057	0.070	1.269	0.105	0.017	0.026	1.605	0.049
Denmark	4506/323	0.071	0.071	0.992	0.962	0.012	0.008	0.682	0.378
UK	639/477	0.076	0.085	1.143	0.403	0.017	0.026	1.558	0.134
Belgium	257/171	0.082	0.079	0.970	0.907	0.011	0.012	1.091	0.897
Holland	1368/837	0.073	0.092	1.270	0.035	0.017	0.010	0.602	0.075
Canada	151/201	0.063	0.100	1.665	0.074	0.017	0.023	1.435	0.527
USA-PA	260/842	0.054	0.096	1.858	0.0018	0.021	0.020	1.146	0.702
USA-CA	208/415	0.080	0.079	0.990	0.960	0.014	0.025	1.788	0.192
New Zealand	526/592	0.066	0.089	1.371	0.048	0.015	0.015	0.983	0.962
Combined	**33532/4277**			**1.242 (1.112-1.387)**	**1.2×10^{-4}**			**1.151 (0.914-1.449)**	**0.230**
Heterogeneity					0.37				0.21

Table 1-D. Association of rsl0455872[G], and rs3798220[C], with venous thromboembolism

		rsl0455872[G]				rs3798220[C]			
	control/case	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***	**frq_con**	**frq_case**	**OR (95% CI)**	***P-value***
VTE									
Iceland	25617/948	0.057	0.063	1.114	0.289	0.017	0.014	0.853	0.423
Spain	949/713	0.063	0.059	0.937	0.663	0.02	0.017	0.853	0.55
Spain-GAITl	328/39	0.068	0.150	2.964	0.024	0.009	0.01	-	-
Spain-GAIT2	788/1 13	0.035	0.050	1.451	0.418	0.022	0.04	2.346	0.128
France-MARTHA	794/1 113	0.068	0.058	0.833	0.187	0.02	0.014	0.702	0.1842
France-FARIVE	589/595	0.087	0.075	0.852	0.323	0.015	0.015	0.932	0.846
Canada-ped	197/28	0.038	0.074	2.003	0.263	0.023	0.018	0.804	0.833
Canada-ace	273/203	0.089	0.085	0.926	0.741	0.031	0.007	0.229	0.007
Combined	**29535/3752**			**0.995 (0.882-1 .123)**	**0.94**			**0.813 (0.642-1 .030)**	**0.086**
Heterogeneity					0.13				0.15
CAD									
Iceland	25617/9424	0.057	0.071	1.268	1.2 xlO $^{-8}$	0.017	0.021	1.291	0.00033
USA-Atlanta	987/2032	0.056	0.090	1.682	1.8 xlO $^{-6}$	0.01	0.02	2.181	0.0013
Combined	**26604/11456**			**1.315 (1.218-1 .419)**	**2.0 x 10^{-12}**			**1.343 (1.177-1 .539)**	**1.4 x 10^{-5}**
Heterogeneity					0.01 5				0.038
African Americans	309/339	0.005	0.016	3.371	0.041	0.009	0.012	1.828	0.314

IA									
Iceland	25617/103	0.057	0.062	1.105	0.743	0.017	0.019	1.172	0.760
Holland	1374/569	0.073	0.068	0.911	0.501	0.017	0.006	0.362	0.007
Finland	315/321	0.032	0.030	0.924	0.807	0.011	0.003	0.278	0.081
Combined	**27306/993**			**0.939 (0.746-1. 183)**	**0.600**			**0.49 (0.283-0.849)**	**0.011**
Heterogeneity					0.840				0.130

Table 2. The association of ***LPA*** score with carotid intima-media thickness (IMT)

		IMT_{log} (mm)			
Study group	**N**	**β**	**SE**	***P-value***	***Phet***
IMPROVE	2994	-0.0061	0.0036	0.090	-
Oxfordshire Family Blood Pressure Study	730	-0.0002	0.0224	0.99	-
Combined	3724	-0.0060	0.0036	0.093	0.81

Linear regression coefficients (β) with corresponding standard errors (SE) and P-values (P), assuming additive allelic effects. N=the number of subjects.

Table 3. Association of ***LPA*** score with angiographic CAD severity

		Number of affected coronary index vessels		
Study group	**N**	**β**	**SE**	***P-value P***
Iceland	2330	0.245	0.061	5.8×10^{-5}
Atlanta-European American	2718	0.281	0.052	6.3×10^{-8}
Atlanta-African American	540	0.396	0.219	0.071
Combined	5588	0.267	0.038	4.8×10^{-12}

Linear regression coefficients (β) with corresponding standard errors (SE) and P-values (P),

Table 4. Association between ***LPA*** score and the age at diagnosis of CAD.

		Age at diagnosis (years)		
	N	**β**	**SE**	***P-value***
European ancestry				
Age at first CAD diagnosis	9275	-1.581	0.295	8.2×10^{-8}
Age at first MI	5148	-0.918	0.416	0.028
Age at first CAD diagnosis for MI cases	5148	-1.334	0.400	0.00085
African Americans				
Age at first CAD diagnosis	333	-3.541	2.521	0.16
Age at first MI	185	-2.221	3.249	0.50
Age at first CAD diagnosis for MI cases	185	-6.009	3.159	0.059

Linear regression coefficients (β) with corresponding standard errors (SE) and P-values (P), assuming additive allelic effects. N=the number of subjects.

Table 5-A. The association of *LPA* score with ischemic stroke and TOAST subtypes

Phenotype Study_population	control/case	n_e	frq_con	frq_case	OR (95% _CI)	*P-value*
Ischemic stroke						
Icela nd	25617/2476	4105	0.074	0.077	1.052 (0.930- 1.190)	0.419
Sweden	718/868	786	0.075	0.088	1.187 (0.919- 1.534)	0.19
Germa ny-South	1193/1 185	1189	0.075	0.082	1.113 (0.900- 1.377)	0.324
Germa ny_West	1057/1387	1200	0.076	0.091	1.232 (1.003- 1.514)	0.047
UK-SE	639/650	644	0.094	0.114	1.241 (0.960- 1.604)	0.099
UK-OXVASC	2246/1095	1472	0.094	0.093	0.989 (0. 832- 1.175)	0.9
Combined	**31470/7661**	**9396**			**1.097 (1.017-1 .183)**	**0.016**
Heterogeneity						0.51
LAA						
Icela nd	25617/245	459	0.074	0.078	1.061 (0. 742- 1.518)	0.746
Sweden	718/175	281	0.075	0.118	1.645 (1.111-2.436)	0.013
Germa ny-South	1193/374	569	0.073	0.082	1.130 (0. 829- 1.540)	0.439
Germa ny_West	1057/557	730	0.076	0.091	1.228 (0.945- 1.596)	0.125
UK-SE	639/231	339	0.092	0.149	1.731 (1.229-2.439)	0.0017
UK-OXVASC	2246/1 11	212	0.094	0.093	0.984 (0. 612- 1.583)	0.947
Combined	**31470/ 1693**	**2590**			**1.272 (1.107-1 .461)**	**6.7 x 10^{-4}**
Heterogeneity						0.19
CE						
Icela nd	25617/403	767	0.074	0.08	1.090 (0. 831- 1.430)	0.534
Sweden	718/152	251	0.075	0.063	0.820 (0. 500- 1.345)	0.432
Germa ny-South	1193/297	476	0.073	0.082	1.129 (0. 805- 1.583)	0.482
Germa ny_West	1057/553	726	0.076	0.092	1.237 (0.952- 1.608)	0.112
UK-SE	639/76	136	0.092	0.112	1.254 (0. 720-2. 184)	0.424
UK-OXVASC	2246/220	401	0.094	0.085	0.897 (0. 634- 1.269)	0.539
Combined	**31470/ 1701**	**2756**			**1.033 (0.878-1 .216)**	**0.69**
Heterogeneity						0.67

SVD						
Iceland	25617/260	503	0.074	0.065	0.879 (0.617-1.251)	0.474
Sweden	718/74	134	0.075	0.068	0.894 (0.461-1.733)	0.74
Germany-South	1193/117	213	0.073	0.092	1.284 (0.788-2.093)	0.316
Germany_West	1057/275	436	0.076	0.087	1.169 (0.821-1.664)	0.386
UK-SE	639/114	193	0.092	0.105	1.167 (0.726-1.877)	0.524
UK-OXVASC	2246/183	338	0.094	0.100	1.075 (0.744-1.553)	0.7
Combined	**31470/1023**	**1818**			**1.055 (0.895-1.244)**	**0.52**
Heterogeneity						0.85

Table 5-B. The association of *LPA* score with PAD, AAA, and VTE

Phenotype Study_population	control/ case	n_e	frq_con	frq_case	OR (95%_CI)	*P-value*
PAD						
Iceland	25617/1602	2716	0.074	0.098	1.386 (1.184-1.580)	2.0×10^{-5}
Denmark	4506/464	841	0.082	0.125	1.585(1.274-1.973)	3.7×10^{-5}
Sweden	143/206	169	0.126	0.121	0.948 (0.594-1.514)	0.823
Austria	433/487	458	0.046	0.096	2.191 (1.499-3.202)	5.1×10^{-5}
Italy	203/176	189	0.047	0.062	1.354 (0.715-2.563)	0.352
USA-PA	260/434	325	0.074	0.106	1.496 (1.010-2.216)	0.044
New Zealand	526/465	494	0.082	0.128	1.657 (1.238-2.217)	6.9×10^{-4}
Combined	**31688/ 3834**	**5222**			**1.473 (1.333-1.626)**	**2.2×10^{-14}**
Heterogeneity						0.13
AAA						
Iceland	25617/420	789	0.074	0.097	1.347(1.049-1.730)	0.020
Denmark	4506/323	603	0.082	0.079	0.949 (0.705-1.277)	0.731
UK	639/477	564	0.094	0.112	1.221 (0.923-1.615)	0.162
Belgium	257/171	205	0.092	0.091	0.984 (0.617-1.569)	0.946
Holland	1368/837	1039	0.090	0.103	1.151 (0.935-1.417)	0.184
Canada	151/201	172	0.079	0.122	1.620 (0.970-2.704)	0.065
USA-PA	260/842	397	0.074	0.117	1.671 (1.182-2.361)	0.003
USA-CA	208/415	277	0.087	0.104	1.220 (0.815-1.827)	0.334
New Zealand	526/592	557	0.082	0.103	1.299 (0.973-1.734)	0.076
Combined	**33532/ 4278**	**4586**			**1.231 (1.113-1.362)**	**5.5×10^{-5}**
Heterogeneity						0.37
VTE						
Iceland	25617/948	1725	0.074	0.077	1.054 (0.879-1.264)	0.570
Spain	949/713	814	0.083	0.076	0.918 (0.709-1.189)	0.517
Spain-GAITI	328/39	70	0.077	0.16	2.834 (1.094-7.343)	0.032
Spain-GAIT2	788/113	198	0.057	0.09	1.958 (0.925-4.144)	0.079
France-MARTHA	794/1113	927	0.088	0.072	0.806 (0.630-1.030)	0.085
France-FARIVE	589/595	592	0.102	0.089	0.864 (0.641-1.163)	0.334
Canada-ped	197/28	49	0.061	0.092	1.544 (0.541-4.408)	0.417
Canada-ace	273/203	233	0.120	0.092	0.743 (0.487-1.133)	0.168
Combined	**29535/ 3752**	**4607**			**0.972 (0.867-1.089)**	**0.63**
Heterogeneity						0.047

Table 5-C. The association of *LPA* score with IA

Phenotype Study population	control/case	n_e	frq_con	frq_case	OR (95% CI)	*P-value*
IA						
Iceland	25617/103	205	0.074	0.082	1.121 (0.665- 1.889)	0.668
Holland	1374/569	805	0.091	0.075	0.816 (0.630- 1.057)	0.124
Finland	315/321	318	0.044	0.033	0.758 (0.424-1 .355)	0.35
Combined	**27306/993**	**1328**			**0.853 (0.688-1.058)**	**0.15**
Heterogeneity						0.52

Tables 5-A, 5-B and 5-C;

LAA= large artery atherosclerosis, CE=cardioembolism, SVD=small vessel disease, PAD=peripheral arterial disease, AAA=abdominal aortic aneurysm, VTE=venous thromboembolism, IA= intracranial aneurysm) .

n_e: effective sample size (n_e case), frq_con :allelic frequency in controls, frq_case : allelic frequency in cases, OR: odds ratio, 95% CI : 95% confidence interval.

Table 6. Association of *LPA* score with age at onset of CAD

	control/case	n_e	frq_con	frq_case	OR (95% CI)	*P-value*
Early onset CAD						
Iceland	25617/1 194	2025	0.074	0.114	1.618 (1.381-1.896)	2.8×10^{-9}
USA-Atlanta	987/619	761	0.066	0.127	2.055 (1.609-2.625)	8.0×10^{-9}
Combined	**26604/ 1813**	**2785**			**1.736 (1.520-1 .984)**	**4.5×10^{-16}**
Heterogeneity						0.11
Intermediate onset CAD						
Iceland	25617/4378	6480	0.074	0.095	1.325 (1.203-1.460)	1.3×10^{-8}
USA-Atlanta	987/1 117	1048	0.066	0.111	1.769 (1.423-2.199)	2.8×10^{-7}
Combined	**26604/5495**	**7528**			**1.390 (1.272-1 .519)**	**3.2×10^{-13}**
Heterogeneity						0.018
Late onset CAD						
Iceland	25617/2465	4176	0.074	0.081	1.106 (0.976-1.253)	0.113
USA-Atlanta	987/292	451	0.066	0.076	1.165 (0.809-1.678)	0.412
Combined	**26604/2757**	**4626**			**1.112 (0.988-1 .250)**	**0.078**
Heterogeneity						0.79
All CAD						
Iceland	25617/9424	11387	0.074	0.092	1.272 (1.183-1.368)	7.6×10^{-11}
USA-Atlanta	987/2032	1329	0.066	0.110	1.753(1.438-2.137)	2.8×10^{-8}
Combined	**26604/ 11456**	**12716**			**1.322 (1.235-1.415)**	**1.1×10^{-15}**
Heterogeneity						0.0029
African Americans	309/339	323	0.012	0.028	2.489 (1.083-5.719)	0.032

CAD=Coronary artery disease;
Early onset: diagnosed before the age of 50 in men and 60 in women);
Intermediate onset: diagnosed at the age 50-69 for men and 60-74 for women;
Late onset: diagnosed >70 in men and >75 in women.
n_e: effective sample size (n_e case), frq_con :allelic frequency in controls, frq_case: allelic frequency in cases, OR: odds ratio, 95% CI: 95% confidence interval.

Table 7-A. Association of *LPA* score with LAA, PAD and AAA, without concomittant CAD

Study population (% without CAD)	control/case	n_e	frq_con	frq_case	OR (95% CI)	*P-value*
Ischemic stroke without CAD						
Iceland (55%)	25617/1352	2425	0.074	0.07	0.952 (0.813-1.115)	0.543
Sweden (81%)	718/703	710	0.075	0.087	1.166 (0.890-1.528)	0.266
Germany-W (71%)	1057/987	1021	0.076	0.097	1.314 (1.054-1.637)	0.015
UK-OXVASC (88%)	2246/884	1269	0.094	0.095	1.011 (0.831-1.230)	0.913
Combined	**29638/3926**	**5425**			**1.062 (0.961-1. 174)**	**0.24**
Heterogeneity						**0.11**
LAA without CAD						
Iceland (42%)	25617/103	199	0.074	0.069	0.928 (0.537-1.604)	0.789
Sweden (70%)	718/94	166	0.075	0.117	1.634 (0.983-2.715)	0.058
Germany-W (74%)	1057/412	593	0.076	0.103	1.400 (1.057-1.855)	0.019
UK-OXVASC (86%)	2246/93	179	0.094	0.100	1.072 (0.647-1.775)	0.787
Combined	**29638/702**	**1137**			**1.297 (1.056-1 .592)**	**0.013**
Heterogeneity						**0.38**
PAD without CAD						
Iceland (21%)	25617/341	638	0.074	0.065	0.871 (0.632-1.201)	0.400
Italy (63%)	203/1 11	144	0.047	0.051	1.089 (0.507-2.339)	0.827
USA-PA (43%)	260/189	219	0.074	0.066	0.887 (0.521-1.510)	0.658
New Zealand (74%)	526/327	403	0.082	0.136	1.779 (1.297-2.440)	3.5×10^{-4}
Combined	**26606/968**	**1404**			**1.181 (0.967-1.443)**	**0.10**
Heterogeneity						**0.010**
AAA without CAD						

Iceland (29%)	25617/121	238	0.074	0.054	0.725 (0.424-1.240)	0.240
UK (74%)	639/351	453	0.094	0.109	1.180 (0.866-1.607)	0.294
Belgium (87%)	257/152	191	0.092	0.083	0.885 (0.531-1.474)	0.639
USA-PA (50%)	260/393	207	0.074	0.093	1.290 (0.861-1.934)	0.217
USA-CA (52%)	208/206	313	0.087	0.100	1.167 (0.729-1.868)	0.520
New Zealand (61%)	526/347	418	0.082	0.100	1.248 (0.893-1.745)	0.195
The Netherlands (79%)	1368/661	891	0.090	0.096	1.070 (0.852-1.345)	0.561
Combined	**28875/22311271**				**1.106 (0.968-1.263)**	**0.140**
Heterogeneity						**0.61**

LAA= large artery atherosclerosis, PAD=peripheral arterial disease, AAA=abdominal aortic aneurysm, CAD=coronary artery disease

n_e: effective sample size (n_e case), frq_con :allelic frequency in controls, frq_case: allelic frequency in cases, OR: odds ratio, 95% CI : 95% confidence interval .

Table 7-B. Association of *LPA* score with LAA, PAD and AAA, with concomitta nt CAD

Study population (% with CAD)	control/case	n_e	frq_con	frq._case	OR (95% CI)	*P-value*
Ischemic stroke with CAD						
Icela nd (45%)	25617/1 124	2020	0.074	0.086	1.180 (0.995- 1.399)	0.057
Sweden (19%)	718/165	268	0.075	0.094	1.277 (0. 832- 1.959)	0.263
Germany-W (29%)	1057/400	580	0.075	0.078	1.034 (0. 759- 1.408)	0.832
UK-OXVASC (12%)	2246/1 19	226	0.094	0.069	0.713 (0.435- 1.168)	0.179
					1.1 16	
Combined	**29638/ 1808**	**3408**			**(0.975-1 .278)**	**0.11**
Heterogeneity						**0.24**
LAA with CAD						
Icela nd (58%)	25617/142	276	0.074	0.083	1.142 (0. 724- 1.802)	0.568
Sweden (30%)	718/41	78	0.075	0.098	1.333 (0. 611-2.907)	0.47
Germany-W (26%)	1057/145	255	0.076	0.059	0.764 (0.463- 1.261)	0.292
UK-OXVASC (14%)	2246/1 5	30	0.094	0.067	0.689 (0. 176-2. 691)	0.592
					0.985	
Combined	**29638/343**	**639**			**(0.729-1.332)**	**0.92**
Heterogeneity						**0.52**
PAD with CAD						
Iceland (79%)	25617/1260	2210	0.074	0.108	1.530 (1.306- 1.793)	1.4×10^{-7}
Ita ly (37%)	203/65	98	0.047	0.084	1.857 (0. 818-4. 219)	0.139
USA- PA (57%)	260/246	191	0.074	0.138	2.013 (1.321-3. 068)	0.001 1
New Zealand (26%)	526/1 17	253	0.082	0.099	1.231 (0. 753-2. 013)	0.408
					1.558	
Combined	**26606/ 1688**	**2752**			**(1.355-1 .792)**	**5.0×10^{-10}**
Heterogeneity						**0.47**
AAA with CAD						
Iceland (71%)	25617/299	576	0.074	0.114	1.623 (1.229-2. 144)	6.5×10^{-4}
UK (26%)	639/126	210	0.092	0.119	1.340 (0. 856-2. 099)	0.201
Belg ium (13%)	257/22	41	0.092	0.136	1.549 (0. 596-4. 024)	0.369
USA- PA (50%)	260/395	199	0.074	0.145	2.127 (1.471-3. 075)	6.5×10^{-5}
USA-CA (48%)	208/190	315	0.087	0.103	1.207 (0. 750- 1.942)	0.438
New Zealand (39)	526/225	315	0.082	0.098	1.228 (0. 835- 1.806)	0.297
The Netherlands (21%)	1368/178	314	0.091	0.129	1.488 (1.051-2. 108)	0.025
					1.523	
Combined	**28875/ 1435**	**1970**			**(1.314-1 .765)**	**2.5×10^{-8}**

Heterogeneity	**0.45**

LAA=large artery atherosclerosis, PAD=peripheral arterial disease, AAA=abdominal aortic aneurysm, CAD=coronary artery disease

n_e: effective sample size (n_e case=n_e control), frq_con :allelic frequency in controls, frq_case: allelic frequency in cases, OR:odds ratio, 95% CI: 95% confidence interval.

EXAMPLE 2

Study populations

Ischemic stroke

Iceland. The Icelandic study has been described elsewhere[42]. Briefly, prevalent stroke cases have been recruited since 1998 through an ongoing cardiovascular disease (CVD) genetics program at deCODE genetics in Iceland. Cases, irrespective of age, were identified from a registry of individuals diagnosed with ischemic stroke or TIA at Landspitali University Hospital in Reykjavik, the only tertiary referral centre in Iceland, during the years 1993 to 2006.

The ischemic stroke or TIA diagnoses were based on standard WHO criteria[43] and imaging evidence (either CT or MRI), and were clinically confirmed by neurologists. Eligible patients who survived the stroke were invited to participate the genetic study, either by attending a recruitment centre for deCODE's genetic studies, or they were visited at their home by a study nurse.

Control subjects were participants from a large variety of genetic programs at deCODE. Individuals with confirmed stroke (identified by cross-matching with hospital lists), who had participated in genetic studies other than those of cardiovascular diseases (CVD) (but not participated in CVD studies) were excluded as controls. In addition, individuals with other atherosclerotic diseases or VTE were not included as controls.

United Kingdom (UK), OXVASC study. The ascertainment of cases and controls from the UK has been described elsewhere[44,45]. In brief, incident cases were ascertained through a population-based study (OXVASC)[44] of all vascular events (including acute stroke) irrespective of age, in a urban population of 91,000 registered with 63 family physicians in nine general practices in Oxfordshire. OXVASC started 2002, and is on-going. Cases with DNA samples used for the current study were recruited 2002-2009.

Ascertainment criteria of cases included standard WHO definitions of TIA and ischemic stroke[43], with diagnoses based on clinical findings and computed tomography (CT) or magnetic resonance imaging (MRI) brain imaging. For the purpose of the current analysis, cases of non-European descent were excluded. All diagnoses were confirmed by a neurologist.

Controls from various geographic locations within the UK (mostly from the SE England) were originally recruited for a study (PROCARDIS) aimed at identifying susceptibility genes for CAD[45]. Eligible respondents with no personal or sibling history of CAD before age 66 years and were of European origin were asked to participate and to provide a blood sample at their general practice. Although the PROCARDIS study recruited siblings, only one of each sibling-pair (randomly chosen) was included in the current analysis.

UK, South East England. The second case control study from the UK, herein referred to as UK-SE included ischemic stroke patients of European descent attending a cerebrovascular service in South East England and were recruited 1995-2002, as previously described[42]. Briefly, all cases were phenotyped by one experienced stroke neurologist with review of original brain imaging with CT or MRI.

Community controls free of symptomatic cerebrovascular disease were recruited by sampling family doctor lists from the same geographical region as the patients.

Sweden. The case control study from Sweden has been described elsewhere[42]. In brief, cases with ischemic stroke attending the stroke unit or the stroke outpatient clinic at Karolinska University Hospital, Huddinge unit in Stockholm were recruited from 1996 to 2002 as part of an on-going genetic epidemiology study. The ischemic stroke diagnosis was based on clinical findings and brain imaging (CT or MRI). The Swedish controls were population-based controls recruited from the same region in central Sweden as the patients.

South Germany. This study, herein referred to as Germany-S, has been described previously[42]. Briefly, cases were ascertained at the stroke unit of the Department of Neurology, Klinikum Grosshadern, University of Munich, during the period 2001-2006. Diagnoses were based on clinical findings and imaging evidence (either CT or MRI), and were clinically confirmed by neurologists. The control group was selected from the KORA S4 study, a community based epidemiological project, and included individuals without history of cardiovascular disease.

Westphalia region, Germany. Another German case-control study, referred to as Germany-W, recruited ischemic stroke patients through hospitals participating in the regional Westphalian Stroke Register, during the period 2000-2003, as described elsewhere[42]. Diagnoses were based on clinical findings and imaging evidence (either CT or MRI), and were clinically confirmed by neurologists. Population controls without a self-reported history of stroke were drawn from the prospective, population based Dortmund Health Study.

TOAST subtype classification

In addition to the clinically relevant diagnostic workup performed, including brain imaging with CT and/or MRI, all ischemic stroke cases from all studies had ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries, echocardiography, Holter monitoring, MRA, CT angiography, and blood tests for prothrombotic state, to provide

information about ischemic stroke causative subtypes. Patients were classified into etiologic subtypes according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST)[46]. Classification was performed independently for stroke populations but in a standardized manner.

The TOAST classification [46] includes six categories: (1) large-artery occlusive disease (large vessel disease), (2) cardioembolism (cardiogenic stroke), (3) small vessel disease (lacunar stroke), (4) other determined etiology, (5) etiology unknown despite diagnostic efforts, or (6) more than one etiology. Patients classified into the TOAST categories 4-6 were excluded from the stroke population from Germany-W. In Iceland, patients were classified as having large-artery occlusive disease if stenosis was ≥ 70% which is a stricter criterion than usually used i.e. ≥ 50%.

Abdominal aortic aneurysm

Iceland. Icelandic AAA cases were enrolled as part of the CVD genetics program at deCODE[47] from a registry of individuals who were admitted at Landspitali University Hospital, in Reykjavik, during 1980 - 2006. AAA was defined as diameter of infarenal aorta of ≥ 30 mm. AAA cases were either followed up or treated by intervention for emergency repair of symptomatic or ruptured AAA or for an elective repair by surgery or endovascular intervention. The Icelandic controls were the same subjects as used for the Icelandic stroke study, described above.

New Zealand. AAA cases were recruited from the Otago-Southland region, the vast majority (>97%) being of Anglo-European ancestry as reported previously [47]. Approximately 80% of cases had undergone surgical AAA repair (typically infrarenal aorta > 50 mm). The control group consisted of elderly individuals with no previous history of vascular disease from the same geographical region. An abdominal ultrasound scan excluded concurrent AAA from the control group and Anglo-European ancestry was required for inclusion. Controls were also asymptomatic for PAD and had ankle brachial indexes >1.

United Kingdom (UK). As previously described [47], AAA cases included those randomized to surveillance in the UK Small Aneurysm Trial. AAA diameter at baseline was 45 mm (32-55 mm). Controls were of European descent, recruited from England.

Denmark. Danish AAA cases were ascertained from population-based screening programmes for 65-74 year old men in 1994-1998 [48] and 2008-2009 (Clinical trials: NCT00662480). In both studies, an AAA was defined as an infrarenal aorta ≥ 30 mm (mean 40.5 mm). None were ruptured. Danish controls were from the randomised population-based intervention study (Inter99), described in details elsewhere [49].

Belgium and Canada. As reported elsewhere in detail [47], AAA cases included those admitted for emergency repair of ruptured AAA or for an elective surgery to the University Hospital of Liege (Liege, Belgium), or to Dalhousie University Hospital (Halifax, Canada). Controls included spouses of individuals with AAA or subjects admitted to the same hospitals for reasons other than AAA. All cases and controls were of European descent.

The Netherlands. As previously described, AAA cases from Utrecht were recruited in 2007-2009 from eight centres in The Netherlands, mainly when individuals visited their vascular surgeon in the polyclinic or, in rare cases, during hospital admission for elective or emergency AAA surgery. An AAA was defined as an infrarenal aorta ≥ 30 mm. Cases (89.9% males) had a mean AAA diameter of 58.4 mm, 61.7% had received surgery, of which 8.1 % was after rupture. The Dutch controls were recruited as part of the Nijmegen Biomedical Study and the Nijmegen Bladder Cancer Study (http://dceg.cancer.gov/icbc/membership.html).

Pennsylvania, USA (USA-PA). The majority (94%) of cases were enrolled through the Department of Vascular Surgery at Geisinger Medical Center, Danville, Pennsylvania[50]. AAA was defined as infrarenal aortic diameter ≥ 30 mm. Four percent of cases were admitted to the University Hospital of Pittsburgh, Pennsylvania for either elective or emergency surgery for AAA, during the years 1986 to 1991[47]. Controls included primary care patients that were ascertained through the Geisinger MyCode Project and were frequency matched for age and gender to the Geisinger Vascular Clinic AAA cases.

California, USA (USA-CA). The AAA-STOP (Abdominal Aortic Aneurysm: Simple Treatment or Prevention) study recruited cases and controls from the San Francisco Bay Area between December 2006 and July 2010. The study concentrated on recruiting small aneurysm cases, defined as having a maximal abdominal aortic diameter of >3.0cm and <5.0cm by ultrasound at entry. Controls had maximal aortic diameter <3.0cm as verified by ultrasound examination. Mean age and range as well as male/female ratio was similar between cases and controls. For the purpose of the current analysis only subjects of European ancestry were used and subjects with CAD were excluded from the control group.

Peripheral arterial disease

Iceland. PAD cases, identified from a registry of individuals diagnosed with PAD during the year 1998-2006, at the Landspitali University Hospital in Reykjavik, were enrolled over a nine year period as part of the CVD genetics program at deCODE[42,47]. The PAD diagnosis was based on vascular imaging or segmental pressure measurements and was clinically confirmed by a vascular surgeon. The Icelandic controls included the same subjects as used for the Icelandic stroke study, described above.

New Zealand. PAD cases were recruited from the Otago-Southland region, the vast majority (>97%) being of Anglo-European ancestry as reported previously[47]. PAD diagnosis was confirmed by an ankle brachial index<0.7, pulse volume recordings and angiography/ultrasound imaging. An abdominal ultrasound scan excluded cases with concurrent AAA. Controls were the same vascular disease-free individuals as described above for the New Zealand AAA cases.

Austria. Cases and controls were recruited through the Linz Peripheral Arterial Disease (LIPAD) study during 2000 to 2002, at the St. John of God Hospital, Department of Surgery (Linz, Austria) as previously described[51]. In brief, the PAD diagnosis was based on clinical findings,

angiography and non-invasive investigations. Diagnoses were confirmed by the attending vascular surgeons. Controls included patients referred to same hospital that had no clinical indication of PAD by history and physical examination and had ankle brachial index >1.0 [51].

Italy. Cases and controls were recruited among subjects consecutively admitted to the Department of Medicine of the A. Gemelli University Hospital of Rome, from 2000 to 2001. PAD diagnosis was performed in accordance with established criteria [52]. All cases had an ankle brachial pressure index lower than 0.8 and were at Fontaine's stage II, with intermittent claudication and no rest pain or trophic lesions [53].

Denmark. Cases were enrolled at Vascular Surgery Department, Viborg Hospital, Denmark, during 1999 to 2004 [54]. The diagnosis was based on clinical findings and all cases had an ankle brachial pressure index lower than 0.9. Danish controls were the same subjects as used for the Danish AAA study, described above.

Sweden. Cases and controls were recruited at the Department of Vascular Diseases at Malmo University Hospital (Malmo, Sweden), as described elsewhere [47]. Briefly, the diagnosis of critical limb ischemia caused by PAD was made in accordance with TransAtlantic Inter-Society Consensus criteria [55] and was confirmed by an experienced vascular surgery consultant. The control group consisted of healthy individuals included in a health screening programme for a preventive medicine project. None of those had symptomatic PAD.

Pennsylvania, USA (USA-PA). Cases were enrolled through the Department of Vascular Surgery at Geisinger Medical Center, Danville, Pennsylvania [50]. PAD subjects had an ankle brachial index (ABI) $\leq$ 0.85 in at least one leg and were confirmed to be AAA-free (infrarenal aortic diameter $\leq$ 2 cm) based on abdominal imaging carried out within the preceding 5 years. Controls were the same individuals from the Geisinger MyCode Project [50] as described above for the USA-PA AAA group.

Venous thromboembolism

Iceland. The ascertainment of Icelandic venous thromboembolism (VTE) cases has been described in detail elsewhere [56]. Briefly, cases identified from a registry of ICD diagnostic codes for VTE at the Landspitali University Hospital in Reykjavik, during the years 1987 to 2002, were enrolled as part of the CVD genetics program at deCODE. Diagnoses were confirmed with accepted imaging techniques (compression ultrasound, venogram, ventilation perfusion scan, pulmonary angiograms, computerized pulmonary angiograms) or confirmed at autopsy. The hospital charts of all cases were reviewed by experienced physicians in order to exclude erroneous diagnoses. Controls were the same subjects as described above for the Icelandic stroke cases.

Canada. The ACE and PEDS studies have been described in detail elsewhere [56]. In brief, cases were recruited from the Thrombosis Clinic at the Ottawa Hospital.

ACE study: Consecutive patients with objectively confirmed idiopathic deep vein thrombosis (DVT) or pulmonary embolism (PE), who had been treated for at least 3 months, were eligible for inclusion [57]. Cases with a malignant disorder were excluded.

PEDS study: Consecutive patients presenting with symptoms or signs suspected by a physician of being caused by acute PE were eligible for the study. Cases with co-morbid condition making life expectancy less than three months, needed long-term use of anticoagulants, were pregnant, or less than 18 years of age, were excluded.

Controls: For each study, friends of cases were recruited as controls and they were frequency matched to cases by sex, ethnicity, and age. Controls were excluded if they had prior VTE or recent malignant disease.

Spain. Cases identified from a registry of ICD diagnostic codes for VTE from anticoagulation clinics in 4 hospitals in Spain: Hospital General Universitario (Murcia), Hospital de la Santa Creu i Sant Pau (Barcelona), Hospital Clinico Universitario (Salamanca), and Clinica Universitaria de Navarra (Pamplona). The study included unrelated cases of European descent with a first, objectively confirmed episode of venous thromboembolism before the age of 75 years. Diagnosis was made by clinical probability, D-dimer levels, compression ultrasonography, ventilation perfusion lung scan, and, when necessary, phlebography or pulmonary angiography. Patients with known malignant disorders were excluded. The control group included unrelated individuals without a history of vascular or thromboembolic disease. The controls were randomly selected among blood donors, and traumatology and ophthalmology patients from the same hospitals matched by age, sex, race, and geographic distribution with the cases[58].

GAIT 1 and GAIT 2, Spain. The GAIT 1 (Genetic Analysis of Idiopathic Thrombophilia) study is composed of 21 extended Spanish families recruited in Barcelona. Twelve families were ascertained though a proband with idiopathic thrombophilia, whereas 9 families were randomly selected. Idiopathic thrombophilia was defined as multiple thrombotic events, a single spontaneous episode with at least 1 affected first-degree relative, or an onset of thrombosis before age 45; and absence of known risk factors such as antithrombin, protein C or protein S deficiencies, activated protein C resistance, factor V Leiden, and lupus anticoagulant. The GAIT 2 study includes an additional set of 15 extended Spanish families recruited in Barcelona. The criteria for inclusion in the study have been described before [59], and were the same as for the GAIT 1 study. For the purpose of the current analyses family members without history of venous and/or arterial thrombotic disease were used as controls.

MARTHA, France. The MARseille THrombosis Association study (MARTHA) has been previously described [60]. Briefly, cases were consecutively recruited whites with a documented history of VTE and without strong known risk factors, including antithrombin, protein C, or protein S deficiency, homozygosity for FV Leiden or FII 20210A, and lupus anticoagulant, who attended the

Thrombophilia Center of La Timone Hospital in France between 1994 and 2005. The thrombotic events were documented by venography, Doppler ultrasound, spiral computed tomographic scanning angiography, and/or ventilation/perfusion lung scan. Approximately 60% of controls included healthy French whites without a personal history of cardiovascular disease (including VTE) from the Marseilles area. The remaining ~40% consisted of healthy French white heterozygotes for the FV Leiden or FII 20210A mutations selected from the national health examination centers of French Social Security in collaboration with the Hemostasis and Thrombosis Study Group.

FARIVE, France. The multicenter case-control FARIVE study conducted in France has been described elsewhere [60]. In brief, the study involves consecutive whites treated for a first episode of proximal DVT and/or PE. Recruitment began in January 2003 and is still ongoing. DVT diagnosis was confirmed by venography and compression-ultrasonography, and for those with PE by spiral computed tomography, high probability ventilation-perfusion lung scan, pulmonary angiography, and with physical findings compatible with DVT. Cases younger than 18 years, with previous VTE, a history of malignancy within 5 years, or have a short life expectancy because of other causes, were not eligible for the study. The control group was recruited during the same period and consisted of inpatients and outpatients, frequency matched to cases by age- and sex, and were free of venous and arterial thrombotic disease. Potential controls with cancer, liver or kidney failure, or a history of venous and/or arterial thrombotic disease were excluded.

Intracranial aneurysm

Iceland. As previously described [56], intracranial aneurysm (IA) cases were identified from a registry of those seen in the years 1994-2006 at Landspitali University Hospital with IA diagnoses (ICD9 code 430, or ICD10 code 160.0-7, 167. 1 and 169.0). The Icelandic controls included the same subjects as used for the Icelandic stroke study, described above.

The Netherlands. The previously described [56] Dutch IA study included cases with ruptured (91.5%) or unruptured (8.5%) IA admitted to the University Medical Center Utrecht. Diagnosis was based on clinical symptoms, conventional angiography, computed tomography or magnetic resonance angiogram. The controls were healthy Dutch blood bank donors of European origin.

Finland. The Finnish cases and controls have been described previously [61]. Cases were admitted for treatment of intracranial aneurysm at either the University Hospital of Kuopio or the University Hospital of Helsinki in Finland.

Intima media thickness (IMT)

IMPROVE. The IMPROVE study includes subjects, (aged 54-79 years), from 7 centers in Italy, Finland, Sweden, the Netherlands and France, who presented at least three vascular risk factors and who had no history of cardiovascular disease or other medical conditions that might limit longevity, were ascertained between 2004-2005 though a longitudinal cohort study. Blood sampling, ultrasound examinations for IMT, and medical history was performed/recorded at

baseline and after 30 months. The ultrasonic variable considered for the purpose of the current study was the mean IMT of the common carotid arteries (CCAs) excluding the first centimeter closest to the bifurcation (left and right). The far walls of the left and right CCA were visualized in anterior, lateral and posterior projections and mean IMT measured. IMT measurements were performed in a centralized laboratory using dedicated software.

Oxfordshire Family Blood Pressure Study. The ascertainment strategy of this study has been described previously [63]. In brief, white British families were ascertained between 1993 and 1997, from hypertensive probands for a genetic study of cardiovascular risk factors. Family members were later (1999-2001) invited to attend a visit for additional phenotyping including ultrasound measurement of CCA IMT. Ultrasonography was performed using a 7.5-MHz linear array transducer, and all of the measurements were made by 1 physician. The scanning protocol involved studying the right and left CCAs in all subjects according to a standard method [64]. End-diastolic frames (of anterior, lateral, and posterior projections) of the distal 1 cm of the CCA were analyzed for mean and maximal IMT, and the average reading from these frames was calculated.

Coronary artery disease (CAD)

Iceland. Icelandic CAD cases were enrolled as part of the CVD genetics program at deCODE. Cases were identified from a registry of over 10,000 individuals who suffered myocardial infarction (MI) before the age of 75 in Iceland between 1981 and 2002 and satisfied the MONICA criteria [65]. Additional subjects with CAD, were identified from a list of those who have undergone percutaneous coronary intervention (PCI) in the years 1993 to 2003, and from a registry of >23.000 individuals with CAD discharge diagnoses (ICD 9 codes 410.*, 411.*, 412.*, 414.*, or ICD 10 codes 120.0, 121.*, 122.*. 123.*, 124.*, 125.*), during the years 1983-2010, from the Landspitali University Hospital in Reykjavik. Age at onset of CAD was assigned as the age at first recorded CAD diagnosis identified from the abovementioned lists. The Icelandic controls included the same subjects as used for the Icelandic stroke study, described above.

Information on CAD extent was gathered from a registry of those undergoing cardiac catheterization at Landspitali University Hospital in Reykjavik during the years 2007-2009. Coronary angiograms of all subjects were analyzed by an experienced cardiologist, with the extent of CAD expressed as the number of vessel territories out of the 4 index vessels (left main stem, left anterior descending, left circumflex artery, and right coronary artery) with stenoses >50% of the vessel normal reference diameter, using visual assessment of lesion severity [66]. Those with <50% stenosis were assigned a normal/atheromatous status (0 vessel disease).

Atlanta, Georgia, USA. The study participants were enrolled at the Emory University Hospital, the Emory Clinic and Grady Memorial Hospitals through its Emory Genebank study and Clinical Registry in Neurology (CRIN). The Emory Genebank studies the association of biochemical and genetic factors to CAD in subjects undergoing cardiac catheterization. CAD cases included those that had at least one significant stenosis (>50%) in any of the major coronary arteries on angiography, or those without significant stenosis but had history of MI, coronary artery bypass

graft, or PCI. Controls included individuals with no or minimal CAD (<20% stenosis) on cardiac catheterization and had no prior history of MI or CAD. Additional controls were recruited from CRIN and included individuals with non-vascular neurological diseases (mainly Parkinson's and Alzheimer's diseases), their spouses, unrelated friends and community volunteers; excluding those with a known history of CAD. Angiography CAD extent was expressed as the number of coronary index vessels with ≥50% stenosis[66], and as Gensini[67], and Duke[68] severity scores. Information on ethnicity was self-reported.

Ascertainment of CAD information amongst cases with ischemic stroke, PAD, and AAA

For the majority of studies the information on CAD was ascertained at the time of recruitment of cases. For the Icelandic studies, however, the overlap with CAD was examined by cross-matching lists of cases with a comprehensive list of subjects with CAD discharge diagnoses hospitalized at the only tertiary care hospital in Iceland during a 27 years period (1983-2010). Hence, the Icelandic cases may have had their CAD diagnosis after the LAA, PAD, or AAA outcome. It should be noted that the proportion of CAD amongst cases with LAA, PAD, and AAA, was strikingly dissimilar between study groups. The overlap with CAD was by far the highest amongst the Icelandic study groups. This diversity is likely due to the difference in ascertainment of CAD information between the studies. This may have contributed to the significant difference in effect sizes between the Icelandic, Italian and New Zealand PAD without CAD studies.

Genotyping

DNA extraction and single SNP genotyping was carried out with various high-throughput methods with laboratory-specific quality control procedures. As the *LPA* SNPs were not represented on available genome-wide chips, genotypes for the Icelandic case-control samples were imputed. First, single SNP genotyping of the two SNPs was performed, using Centaurus (Nanogen) assays[21], on a subset of >900 (rs10455872) and >12,000 (rs3978220) Icelanders who also had genome-wide SNP data (either 317K or 370K Illumina chips). Haplotype information from the observed genetic data, along with the linkage disequilibrium structure in the Icelandic samples with genome-wide SNP data were used for imputation in cases and controls, as previously described[22].

Statistical analysis

For the main analyses we assessed the association between the outcome variables and the total number of minor alleles of either SNP (rs3798220(C) or rs10455872(G)), assuming that both minor alleles have equal effects on the outcome. The effect of each SNP individually (rs3798220(C) and rs10455872(G)) on the binary outcomes, as reported in **Table 1,** was also assessed by comparing with the major allele of both SNPs.

Logistic and linear regression models were used for binary and quantitative outcomes, respectively, to test for linear trend with allele score {*LPA* score=the sum of minor alleles of

rs3798220 and rs10455872) to derive effect estimates, standard errors, and two sided *P*-values, assuming additive allelic effects (on the linear or log scale as appropriate). These tests were performed as implemented in NEMO[23], the R-software, and STATA™ software, version 10.

The IMT trait was logarithmically transformed prior to analysis, and covariate adjusted for age, sex and study center in the IMPROVE study and for age, sex, body mass index, and physical exercise in the Oxfordshire Family Blood Pressure Study. Age at angiography and sex were included as covariates in the regression model when fitting the number of affected coronary vessels to the *LPA* score.

To allow for the relatedness between subjects in the GAIT studies the method of generalized estimating equations (GEE) was used as implemented in the GEE package[24] for the R-software. For the Icelandic studies, P-values, and 95% CI of ORs, are given after adjustment for the relatedness of the subjects and other possible population stratification using the method of genomic control. As the association tests for IMT in the Oxfordshire Family Blood Pressure studies were not significant, no adjustments for relatedness were done.

A fixed effects meta-analysis approach was used to estimate an inverse-variance weighted average of the ORs from all studies of each outcome category. Tests of whether summary ORs were equal to one, and whether the true effects in all studies were the same (likelihood ratio χ^2 heterogeneity test) were performed.

The power of the study to detect significant association (at $P \leq 0.05$ and assuming an OR = 1.3 which is lower than reported for CAD) between *LPA* score and ischemic stroke, PAD, AAA, or VTE, was >99%. The power was slightly lower for the ischemic stroke subtypes LAA, CE, or SVD, or 96%, 97%, and 88% respectively, and for IA the power was 75%. The power to detect an association between *LPA* score and IMT with effect size >0.1mm was >99%, at *P* even as low as 0.0005. Prospective studies suggest that a 0.1 mm increase in IMT is associated with a 20-30% increase in risk of developing CAD [25, 26], and each *LPA* minor allele increases the risk of CAD by 20-30%. An effect size of 0.1 mm per variant allele could therefore be expected if the effect of *LPA* variants on CAD (or other atherosclerotic diseases) was mediated entirely through early atherogenesis and reflected in changes in the carotid artery IMT. As the case-control ratio was very different between studies the effective sample size (n_e), representing the sample size for cases where case-control ratio is 1, for each study was estimated as $n_e = 2N_cN_a/(N_c + N_a)$ (where N_c = number of controls and N_a = number of cases) and reported in Figures 1 and 2. For the Icelandic studies the effect sample size was further adjusted for relatedness by dividing it by the genomic control inflation factor. The combined effective sample sizes were used in the power calculations.

REFERENCES

1. Hackam DG, Anand SS. Emerging risk factors for atherosclerotic vascular disease: a critical review of the evidence. Jama 2003; 290(7) :932-40.
2. Gaw A, Boerwinkle E, Cohen JC, Hobbs HH. Comparative analysis of the apo(a) gene, apo(a) glycoprotein, and plasma concentrations of Lp(a) in three ethnic groups. Evidence for no common "null" allele at the apo(a) locus. J Clin Invest 1994;93(6) :2526-34.
3. Brunner C, Lobentanz EM, Petho-Schramm A, et al. The number of identical kringle IV repeats in apolipoprotein(a) affects its processing and secretion by HepG2 cells. J Biol Chem 1996; 271 (50) :32403- 10.
4. Schreiner PJ, Morrisett JD, Sharrett AR, et al. Lipoprotein[a] as a risk factor for preclinical atherosclerosis. Arterioscler Thromb 1993; 13(6) :826-33.
5. Sotiriou SN, Orlova VV, Al-Fakhri N, et al. Lipoprotein(a) in atherosclerotic plaques recruits inflammatory cells through interaction with Mac-1 integrin. Faseb J 2006; 20(3) :559-61 .
6. Miles LA, Fless GM, Levin EG, Scanu AM, Plow EF. A potential basis for the thrombotic risks associated with lipoprotein(a) . Nature 1989; 339(6222) :301-3.
7. Caplice NM, Panetta C, Peterson TE, et al. Lipoprotein (a) binds and inactivates tissue factor pathway inhibitor: a novel link between lipoproteins and thrombosis. Blood 2001 ;98(10) :2980-7.
8. Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med 2009; 361 (26) :2518-28.
9. Erqou S, Kaptoge S, Perry PL, et al. Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvascular mortality. Jama 2009; 302(4) :412-23.
10. Erqou S, Thompson A, Di Angelantonio E, et al. Apolipoprotein(a) isoforms and the risk of vascular disease: systematic review of 40 studies involving 58,000 participants. J Am Coll Cardiol ;55(19) :2160-7.
11. Hopewell JC, Clarke R, Parish S, et al. Lipoprotein(a) Genetic Variants Associated with Coronary and Peripheral Vascular Disease, But Not with Stroke Risk in the Heart Protection Study. Circ Cardiovasc Genet.
12. Cheng SW, Ting AC, Wong J. Lipoprotein (a) and its relationship to risk factors and severity of atherosclerotic peripheral vascular disease. Eur J Vase Endovasc Surg 1997; 14(l) :17-23.
13. Volpato S, Vigna GB, McDermott MM, et al. Lipoprotein(a), inflammation, and peripheral arterial disease in a community-based sample of older men and women (the InCHIANTI study) . Am J Cardiol ;105(12) :1825-30.
14. Jones GT, van Rij AM, Cole J, et al. Plasma lipoprotein(a) indicates risk for 4 distinct forms of vascular disease. Clin Chem 2007; 53(4) :679-85.
15. Takagi H, Manabe H, Kawai N, Goto SN, Umemoto T. Circulating lipoprotein(a) concentrations and abdominal aortic aneurysm presence. Interact Cardiovasc Thorac Surg 2009;9(3) :467-70.
16. Pradhan AD, Shrivastava S, Cook NR, Rifai N, Creager MA, Ridker PM. Symptomatic peripheral arterial disease in women : nontraditional biomarkers of elevated risk. Circulation 2008; 117(6) :823-31 .
17. Ridker PM, Stampfer MJ, Rifai N. Novel risk factors for systemic atherosclerosis: a comparison of C-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. Jama 2001 ;285(19) :2481-5.
18. Aboyans V, Criqui MH, Denenberg JO, Knoke JD, Ridker PM, Fronek A. Risk factors for progression of peripheral arterial disease in large and small vessels. Circulation 2006; 113(22) :2623-9.
19. Sofi F, Marcucci R, Abbate R, Gensini GF, Prisco D. Lipoprotein (a) and venous thromboembolism in adults: a meta-analysis. Am J Med 2007; 120(8) :728-33.
20. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Polak JF, Folsom AR. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. Arch Intern Med 2002; 162(10) :1182-9.
21. Kutyavin IV, Milesi D, Belousov Y, et al. A novel endonuclease IV post-PCR genotyping system. Nucleic Acids Res 2006;34(19) :el28.
22. Kong A, Masson G, Frigge ML, et al. Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 2008.
23. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet 2003; 35(2) :131-8.
24. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data . Bioinformatics; 26(4) :580- 1.

25. Bots ML, Hoes AW, Koudstaal PJ, Hofman A, Grobbee DE. Common carotid intima-media thickness and risk of stroke and myocardial infarction: the Rotterdam Study. Circulation 1997;96(5) :1432-7.
26. Hodis HN, Mack WJ, LaBree L, et al. The role of carotid arterial intima-media thickness in predicting clinical coronary events. Ann Intern Med 1998; 128(4) :262-9.
27. Dahlen GH, Guyton JR, Attar M, Farmer JA, Kautz JA, Gotto AM, Jr. Association of levels of lipoprotein Lp(a), plasma lipids, and other lipoproteins with coronary artery disease documented by angiography. Circulation 1986; 74(4) :758-65.
28. Korhonen T, Savolainen MJ, Koistinen MJ, et al. Association of lipoprotein cholesterol and triglycerides with the severity of coronary artery disease in men and women. Atherosclerosis 1996; 127(2) :213-20.
29. Zampoulakis JD, Kyriakousi AA, Poralis KA, et al. Lipoprotein(a) is related to the extent of lesions in the coronary vasculature and to unstable coronary syndromes. Clin Cardiol 2000; 23(12) :895-900.
30. Wang XL, Tan C, McCredie RM, Wilcken DE. Determinants of severity of coronary artery disease in Australian men and women. Circulation 1994; 89(5) :1974-81 .
31. Gazzaruso C, Geroldi D, Garzaniti A, et al. Apolipoprotein(a) phenotypes as genetic markers of coronary atherosclerosis severity. Int J Cardiol 1998; 64(3) :277-84.
32. Rohani M, Jogestrand T, Ekberg M, et al. Interrelation between the extent of atherosclerosis in the thoracic aorta, carotid intima-media thickness and the extent of coronary artery disease. Atherosclerosis 2005; 179(2) :311-6.
33. O'Leary DH, Polak JF, Kronmal RA, Manolio TA, Burke GL, Wolfson SK, Jr. Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults. Cardiovascular Health Study Collaborative Research Group. N Engl J Med 1999; 340(1) :14-22.
34. Kivimaki M, Magnussen CG, Juonala M, et al. Conventional and Mendelian randomization analyses suggest no association between lipoprotein(a) and early atherosclerosis : the Young Finns Study. Int J Epidemiol .
35. Ronald J, Rajagopalan R, Cerrato F, et al. Genetic variation in LPAL2, LPA, and PLG predicts plasma lipoprotein(a) level and carotid artery disease risk. Stroke;42(l):2-9.
36. Johnsen SH, Forsdahl SH, Singh K, Jacobsen BK. Atherosclerosis in abdominal aortic aneurysms : a causal event or a process running in parallel? The Tromso study. Arterioscler Thromb Vase Biol ;30(6) :1263-8.
37. Nordon IM, Hinchliffe RJ, Loftus IM, Thompson MM. Pathophysiology and epidemiology of abdominal aortic aneurysms. Nat Rev Cardiol ;8(2) :92- 102.
38. Lodder J, Bamford JM, Sandercock PA, Jones LN, Warlow CP. Are hypertension or cardiac embolism likely causes of lacunar infarction? Stroke 1990; 21(3) :375-81 .
39. Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism : a meta-analysis. Circulation 2008; 117(1) :93- 102.
40. Ohira T, Shahar E, Chambless LE, Rosamond WD, Mosley TH, Jr., Folsom AR. Risk factors for ischemic stroke subtypes: the Atherosclerosis Risk in Communities study. Stroke 2006; 37(10) :2493-8.
41. Jones GT, Deng M, Hammond-Tooke GD, McCormick SP, van Rij AM. Increased plasma lipoprotein(a) found in large-artery atherosclerotic, but not small-artery occlusive, stroke. Clin Chem 2009; 55(10) :1888-90.
42. Gretarsdottir, S. et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. Ann Neurol 64, 402-9 (2008) .
43. Hatano, S. Experience from a multicentre stroke register: a preliminary report. Bull World Health Organ 54, 541-53 (1976) .
44. Rothwell, P.M. et al. Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study) . Lancet 366, 1773-83 (2005) .
45. Clarke, R. et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med 361, 2518-28 (2009) .
46. Adams, H.P., Jr. et al. Design of the Trial of Org 10172 in Acute Stroke Treatment (TOAST). Control Clin Trials 18, 358-77 (1997) .
47. Helgadottir, A. et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. Nat Genet 40, 217-24 (2008) .
48. Lindholt, J.S., Juul, S., Fasting, H. & Henneberg, E.W. Screening for abdominal aortic aneurysms : single centre randomised controlled trial . Bmj 330, 750 (2005) .
49. Jorgensen, T. et al. A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. Eur J Cardiovasc Prev Rehabil 10, 377-86 (2003) .

50. Elmore, J.R. et al. Identification of a genetic variant associated with abdominal aortic aneurysms on chromosome 3pl2. 3 by genome wide association. J Vase Surg 49, 1525-31 (2009).
51. Mueller, T. et al. Factor V Leiden, prothrombin G20210A, and methylenetetrahydrofolate reductase C677T mutations are not associated with chronic limb ischemia: the Linz Peripheral Arterial Disease (LIPAD) study. J Vase Surg 41, 808-15 (2005).
52. Suggested standards for reports dealing with lower extremity ischemia. Prepared by the Ad Hoc Committee on Reporting Standards, Society for Vascular Surgery/North American Chapter, International Society for Cardiovascular Surgery. J Vase Surg 4, 80-94 (1986).
53. Flex, A. et al. The -174 G/C polymorphism of the interleukin-6 gene promoter is associated with peripheral artery occlusive disease. Eur J Vase Endovasc Surg 24, 264-8 (2002).
54. Joensen, J.B. et al. Can long-term antibiotic treatment prevent progression of peripheral arterial occlusive disease? A large, randomized, double-blinded, placebo-controlled trial. Atherosclerosis 196, 937-42 (2008).
55. Dormandy, J.A. & Rutherford, R.B. Management of peripheral arterial disease (PAD). TASC Working Group. TransAtlantic Inter-Society Consensus (TASC). J Vase Surg 31, S1-S296 (2000).
56. Gretarsdottir, S. et al. Genome-wide association study identifies a sequence variant within the DAB2IP gene conferring susceptibility to abdominal aortic aneurysm. Nat Genet 42, 692-7.
57. Wells, P.S. et al. The ACE D/D genotype is protective against the development of idiopathic deep vein thrombosis and pulmonary embolism. Thromb Haemost 90, 829-34 (2003).
58. Corral, J. et al. A nonsense polymorphism in the protein Z-dependent protease inhibitor increases the risk for venous thrombosis. Blood 108, 177-83 (2006).
59. Vila, L. et al. Heritability of thromboxane A2 and prostaglandin E2 biosynthetic machinery in a Spanish population. Arterioscler Thromb Vase Biol 30, 128-34.
60. Tregouet, D.A. et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. Blood 113, 5298-303 (2009).
61. Weinsheimer, S. et al. Association of kallikrein gene polymorphisms with intracranial aneurysms. Stroke 38, 2670-6 (2007).
62. Baldassarre, D. et al. Cross-sectional analysis of baseline data to identify the major determinants of carotid intima-media thickness in a European population: the IMPROVE study. Eur Heart J 31, 614-22.
63. Mayosi, B.M. et al. Genotype at the -174G/C polymorphism of the interleukin-6 gene is associated with common carotid artery intimal-medial thickness: family study and meta-analysis. Stroke 36, 2215-9 (2005).
64. Adams, M.R. et al. Carotid intima-media thickness is only weakly correlated with the extent and severity of coronary artery disease. Circulation 92, 2127-34 (1995).
65. Nomenclature and criteria for diagnosis of ischemic heart disease. Report of the Joint International Society and Federation of Cardiology/World Health Organization task force on standardization of clinical nomenclature. Circulation 59, 607-9 (1979).
66. Ringqvist, I. et al. Prognostic value of angiographic indices of coronary artery disease from the Coronary Artery Surgery Study (CASS). J Clin Invest 71, 1854-66 (1983).
67. Gensini, G.G. A more meaningful scoring system for determining the severity of coronary heart disease. Am J Cardiol 51, 606 (1983).

CLAIMS

1. A method of determining a susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject, the method comprising:

analyzing data representative of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872 in the subject, wherein the presence of allele C in rs3798220 and/or allele G in rsl0455872 is associated with increased susceptibility to the condition in humans, and

determining a susceptibility to the condition for the human subject from the data.

2. The method of claim 1, wherein severe coronary artery disease is characterized by at least one of

(i) at least 50% stenosis of more than one coronary artery;

(ii) systemic atherosclerotic disease in the human subject, and

(iii) the concomitant diagnosis of coronary artery disease and at least one of

a. abdominal aortic aneurysm;

b. peripheral arterial disease; and

c. large arterial atherosclerosis

in the subject.

3. The method of claim 1, wherein early-onset coronary artery disease is characterized by age at onset of less than 50 years for males and less than 60 years for females.

4. The method of any one of the preceding claims, comprising analyzing the data for the presence or absence of allele C in rs3798220 and/or allele G in rsl0455872, wherein determination of the presence of the at least one allele is indicative of increased susceptibility to the condition.

5. The method of claim 4, wherein the analyzing data comprises analyzing a biological sample from the human subject to obtain nucleic acid sequence information about the subject, wherein the nucleic acid sequence information comprises sequence sufficient to identify the presence or absence of allele C in rs3798220 and/or allele G in rsl0455872 in the subject.

6. The method of claim 5, comprising analyzing the biological sample to obtain the nucleic acid sequence information.

7. The method of claim 4, wherein the analyzing data comprises analyzing data from a preexisting record about the human subject.

8. A method of determining whether a human subject is at increased risk of developing a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, the method comprising steps of

obtaining a biological sample containing nucleic acid from the subject;

determining, in the biological sample, the allelic identity of at least one polymorphic marker selected from rs3798220 and rsl0455872;

wherein identification of at least one allele selected from allele C in rs3798220 and allele G in rsl0455872 in the biological sample is indicative that the subject is at increased risk of developing the condition.

9. The method of claim 8, wherein severe coronary artery disease is characterized by at least one of

(i) at least 50% stenosis of more than one coronary artery;

(ii) systemic atherosclerotic disease [affecting multiple regional circulations, including heart, brain, kidneys, limbs, mesentery] in the human subject, and

(iii) the concomitant diagnosis of coronary artery disease and at least one of

a. abdominal aortic aneurysm;

b. peripheral arterial disease; and

c. large arterial atherosclerosis

in the subject.

10. The method of claim 8, wherein early-onset coronary artery disease is characterized by age at onset of less than 50 years for males and less than 60 years for females.

11. The method of claim 8, wherein the determining comprises analyzing the biological sample to obtain the allelic identity.

12. The method of claim 8, wherein the determining comprises obtaining nucleic acid sequence information about marker rs3798220 and/or marker rsl0455872 and determine the allelic identity of the marker from the nucleic acid sequence information.

13. A method of determining a susceptibility to severe coronary artery disease in a human subject, the method comprising:

analyzing data representative of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872 in the subject, wherein the presence of allele C in rs3798220 and allele G in rsl0455872 is associated with increased number of obstructed coronary artery in humans, and

determining a susceptibility to severe coronary artery disease for the human subject from the data.

14. The method of claim 13, wherein severe coronary artery disease is characterized by at least one of

(i) at least 50% stenosis of more than one coronary artery;

(ii) systemic atherosclerotic disease [affecting multiple regional circulations, including heart, brain, kidneys, limbs, mesentery] in the human subject, and

(iii) the concomitant diagnosis of coronary artery disease and at least one of

a. abdominal aortic aneurysm;

b. peripheral arterial disease; and

c. large arterial atherosclerosis

in the subject.

15. The method of claim 13, wherein the data is nucleic acid sequence data.

16. The method of claim 13, wherein the nucleic acid sequence data is obtained from a biological sample containing nucleic acid from the subject.

17. A method of determining a susceptibility to a condition selected from the group consisting of abdominal aortic aneurysm, peripheral arterial disease and large arterial atherosclerosis in a human subject previously diagnosed with coronary artery disease, the method comprising:

analyzing data representative of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872 in the subject, wherein the presence of allele C in

rs3798220 and allele G in rsl0455872 is associated with increased susceptibility to the condition in humans, and

determining a susceptibility to the condition for the human subject from the data.

18. The method of claim 17, wherein the coronary artery disease is early-onset coronary artery disease characterized by age at onset of less than 50 years for males and less than 60 years for females.

19. The method of claim 17 or claim 18, comprising analyzing the data for the presence or absence of allele C in rs3798220 and/or allele G in rsl0455872, wherein determination of the presence of the at least one allele is indicative of increased susceptibility to the condition.

20. The method of any one of the claims 17 to 19, wherein the analyzing data comprises analyzing a biological sample from the human subject to obtain nucleic acid sequence information about the subject, wherein the nucleic acid sequence information comprises sequence information sufficient to identify the presence or absence of allele C in rs3798220 and/or allele G in rsl0455872 in the subject.

21. The method of any one of the claims 17 to 20, wherein the analyzing data comprises analyzing data from a preexisting record about the human subject.

22. An assay for determining a susceptibility to a condition selected from the group consisting of abdominal aortic aneurysm, peripheral arterial disease, large arterial atherosclerosis, early-onset coronary artery disease and severe coronary artery disease in a human subject, the assay comprising:

(i) obtaining a biological sample comprising nucleic acid from the subject;

(ii) determining, in the nucleic acid, the presence or absence of at least one allele of a polymorphism selected from the group consisting of rs3798220 and rsl0455872, wherein the presence of allele C in rs3798220 and allele G in rsl0455872 is associated with increased susceptibility to the condition in humans; and

(iii) determining a susceptibility to the condition for the subject from the presence or absence of the at least one allele.

23. The assay of claim 22, wherein severe coronary artery disease is characterized by at least one of

(0 at least 50% stenosis of more than one coronary artery;

(ii) systemic atherosclerotic disease [affecting multiple regional circulations, including heart, brain, kidneys, limbs, mesentery] in the human subject, and

(iii) the concomitant diagnosis of coronary artery disease and at least one of

a. abdominal aortic aneurysm;

b. peripheral arterial disease; and

c. large arterial atherosclerosis

in the subject.

24. The assay of claim 22, wherein early-onset coronary artery disease is characterized by age at onset of less than 50 years for males and less than 60 years for females.

25. The assay of claim 22, wherein the determining comprises analyzing the biological sample to obtain the allelic identity.

26. A method of determining the prognosis of a human subject diagnosed with coronary artery disease, the method comprising steps of

obtaining a biological sample containing nucleic acid from the subject;

determining, in the biological sample, the allelic identity of at least one polymorphic marker selected from rs3798220 and rsl0455872;

wherein identification of at least one allele selected from allele C in rs3798220 and allele G in rsl0455872 in the biological sample is indicative of a worse prognosis of the coronary artery disease for the subject.

27. The method of claim 26, wherein the worse prognosis is characterized by an increased susceptibility of systemic atherosclerotic disease in the subject.

28. The method of claim 26, wherein the worse prognosis is characterized by an increased susceptibility of development of an atherosclerotic condition selected from abdominal arterial disease, peripheral arterial disease and large arterial disease in the subject.

29. The method of any one of the claims 26 to claim 28, wherein the determining comprises analyzing the biological sample to obtain the allelic identity.

30. The method of claim 29, wherein the determining comprises obtaining nucleic acid sequence information about marker rs3798220 and/or marker rsl0455872 and determine the allelic identity of the marker from the nucleic acid sequence information.

31. The method of claim 29 or claim 30, wherein obtaining allelic identity comprises identifying at least one allele of the at least one polymorphic marker in the biological sample.

32. A system for identifying susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, the system comprising:

at least one processor;

at least one computer-readable medium;

a susceptibility database operatively coupled to a computer-readable medium of the system and containing population information correlating the presence or absence of one or more alleles selected from allele C in rs3798220 and allele G in rsl0455872 to susceptibility of the condition in a population of humans;

a measurement tool that receives an input about the human subject and generates information from the input about the presence or absence of the at least one allele in the human subject; and

an analysis tool that:

is operatively coupled to the susceptibility database and the the measurement tool,

is stored on a computer-readable medium of the system,

is adapted to be executed on a processor of the system, to compare the information about the human subject with the population information in the susceptibility database and generate a conclusion with respect to susceptibility to the condition for the human subject.

33. The system according to claims 32, further including:

a communication tool operatively coupled to the analysis tool, stored on a computer-readable medium of the system and adapted to be executed on a processor of the system to communicate to the subject, or to a medical practitioner for the subject, the conclusion with respect to susceptibility to the condition for the subject.

34. The system of claim 32 or claim 33, wherein severe coronary artery disease is characterized by at least one of

(0 at least 50% stenosis of more than one coronary artery;

(ii) systemic atherosclerotic disease [affecting multiple regional circulations, including heart, brain, kidneys, limbs, mesentery] in the human subject, and

(iii) the concomitant diagnosis of coronary artery disease and at least one of

a. abdominal aortic aneurysm;

b. peripheral arterial disease; and

c. large arterial atherosclerosis

in the subject.

35. The system according to any one of claims 32 to 34, wherein the measurement tool comprises a tool stored on a computer-readable medium of the system and adapted to be executed by a processor of the system to receive a data input about a subject and determine information about the presence or absence of allele C in rs3798220 and/or allele G in rsl0455872 in a human subject from the data.

36. The system according to claim 35, wherein the data is genomic sequence information, and the measurement tool comprises a sequence analysis tool stored on a computer readable medium of the system and adapted to be executed by a processor of the system to determine the presence or absence of allele C in rs3798220 and/or allele G in rsl0455872 from the genomic sequence information.

37. The system according to any one of claims 32 to 36, wherein the input about the human subject is a biological sample from the human subject, and wherein the measurement tool comprises a tool to identify the presence or absence of allele C in rs3798220 and/or allele G in rsl0455872 in the biological sample, thereby generating information about the presence or absence of the at least one allele in a human subject.

38. The system according to claim 37, wherein the measurement tool includes:

an oligonucleotide microarray containing a plurality of oligonucleotide probes attached to a solid support;

a detector for measuring interaction between nucleic acid obtained from or amplified from the biological sample and one or more oligonucleotides on the oligonucleotide microarray to generate detection data; and

an analysis tool stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to determine the presence or absence of the at least one allele based on the detection data.

39. The system according to claim 37, wherein the measurement tool includes:

a nucleotide sequencer capable of determining nucleotide sequence information from nucleic acid obtained from or amplified from the biological sample; and

an analysis tool stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to determine the presence or absence of the at least one allele based on the nucleotide sequence information.

40. The system according to any one of claims 32 to 39, further comprising:

a medical protocol database operatively connected to a computer-readable medium of the system and containing information correlating the presence or absence of the at least one allele and medical protocols for human subjects at risk for the condition; and

a medical protocol routine, operatively connected to the medical protocol database and the analysis routine, stored on a computer-readable medium of the system, and adapted to be executed on a processor of the system, to compare the conclusion from the analysis routine with respect to susceptibility to the condition for the subject and the medical protocol database, and generate a protocol report with respect to the probability that one or more medical protocols in the database will:

reduce susceptibility to the condition; or

delay onset of the condition; or

increase the likelihood of detecting the condition at an early stage to facilitate early treatment.

41. The system according to any one of claims 33 to 40, wherein the communication tool is operatively connected to the analysis routine and comprises a routine stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to:

generate a communication containing the conclusion; and

transmit the communication to the subject or the medical practitioner, or enable the subject or medical practitioner to access the communication.

42. The system according to claim 41, wherein the communication expresses the susceptibility to the condition in terms of odds ratio or relative risk or lifetime risk.

43. The system according to claim 41 or 42, wherein the communication further includes the protocol report.

44. The system according to any one of claims 32 to 43, wherein the susceptibility database further includes information about at least one parameter selected from the group consisting of age, sex, ethnicity, race, medical history, weight, lipid measurements, blood pressure, family history of the condition, and smoking history in humans and impact of the at least one parameter on susceptibility to the condition.

45. A system for assessing or selecting a treatment protocol for a subject diagnosed with a condition selected from the group consisting of large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease, comprising:

at least one processor;

at least one computer-readable medium;

a medical treatment database operatively connected to a computer-readable medium of the system and containing information correlating the presence or absence of at least one allele selected from allele C in rs3798220 and allele G in rsl0455872 and efficacy of treatment regimens for the condition;

a measurement tool to receive an input about the human subject and generate information from the input about the presence or absence of the at least one allele in a human subject diagnosed with the condition; and

a medical protocol tool operatively coupled to the medical treatment database and the measurement tool, stored on a computer-readable medium of the system, and adapted to be executed on a processor of the system, to compare the information with respect to presence or absence of the at least one allele for the subject and the medical treatment database, and generate a conclusion with respect to at least one of:

the probability that one or more medical treatments will be efficacious for treatment of the condition for the patient; and

which of two or more medical treatments for the condition will be more efficacious for the patient.

46. The system according to claim 45, wherein the measurement tool comprises a tool stored on a computer-readable medium of the system and adapted to be executed by a processor of the system to receive a data input about a subject and determine information about the presence or absence of the at least one allele in a human subject from the data.

47. The system according to claim 46, wherein the data is genomic sequence information, and the measurement tool comprises a sequence analysis tool stored on a computer readable

medium of the system and adapted to be executed by a processor of the system to determine the presence or absence of the at least one allele from the genomic sequence information.

48. The system according to claim 45, wherein the input about the human subject is a biological sample from the human subject, and wherein the measurement tool comprises a tool to identify the presence or absence of the at least one allele in the biological sample, thereby generating information about the presence or absence of the at least one allele in a human subject.

49. The system according to any one of claims 45 to 48, further comprising a communication tool operatively connected to the medical protocol routine for communicating the conclusion to the subject, or to a medical practitioner for the subject.

50. The system according to claim 49, wherein the communication tool comprises a routine stored on a computer-readable medium of the system and adapted to be executed on a processor of the system, to:

generate a communication containing the conclusion; and

transmit the communication to the subject or the medical practitioner, or enable the subject or medical practitioner to access the communication.

51. A method for determining susceptibility to a condition selected from large artery atherosclerosis, peripheral arterial disease, abdominal aortic aneurysm, early-onset coronary artery disease and severe coronary artery disease in a human subject, the method comprising:

analyzing nucleic acid from the individual for evidence of the presence of allele C in rs3798220 and/or allele G in rsl0455872; and

determining an increased susceptibility to the condition from evidence that allele C in rs3798220 and/or allele G in rsl0455872 is present in the subject, or determining a decreased susceptibility to the condition from evidence that allele C in rs3798220 and/or allele G in rsl0455872 is absent in the subject.

52. The method of claim 51, further comprising a step of obtaining a biological sample comprising nucleic acid from the subject.

53. The method of claim 51 or claim 52, further comprising analyzing the nucleic acid from the subject to obtain information about the presence or absence of allele C in rs3798220 and/or

allele G in rsl0455872, wherein the analyzing of the nucleic acid from the subject comprises at least one nucleic acid analysis technique selected from the group consisting of: polymerase chain reaction, allele-specific hybridization, allele-specific primer extension, allele-specific amplification, nucleic acid sequencing, 5-exonuclease digestion, molecular beacon assay, oligonucleotide ligation assay, size analysis, and singe-stranded conformational analysis.

54. The method of any one of the claims 51 to 53, wherein evidence of the presence of allele C in rs3798220 and/or allele G in rsl0455872 is obtained by analyzing at least one polymorphic marker in linkage disequilibrium with rs3798220 and/or at least one polymorphic marker in linkage disequilibrium with rsl0455872, wherein the linkage disequilibrium is characterized by a correlation coefficient r^2 of greater than 0.2.

55. The method of claim 54, wherein the at least one polymorphic marker in linkage disequilibrium with rs3798220 is selected from the group consisting of the markers listed in Surrogate Marker Table A.

56. The method of claim 54, wherein the at least one polymorphic marker in linkage disequilibrium with rsl0455872 is selected from the group consisting of the markers listed in Surrogate Marker Table B.

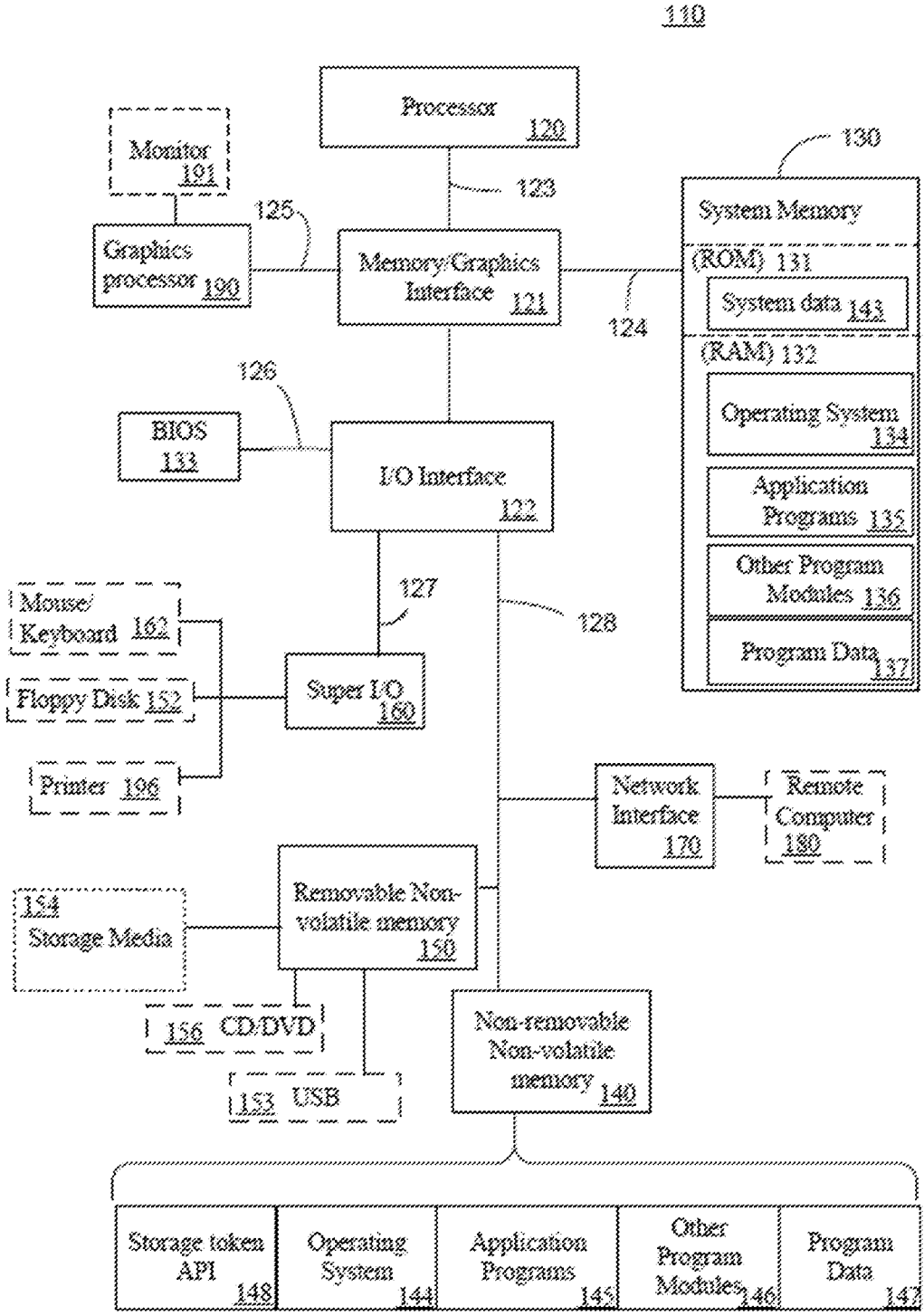

110
Processor 120
Monitor 191
123
130
System Memory
125
Graphics processor 190
Memory/Graphics Interface 121
(ROM) 131
System data 143
124
(RAM) 132
126
BIOS 133
I/O Interface 122
Operating System 134
Application Programs 135
Other Program Modules 136
Program Data 137
127
128
Mouse/ Keyboard 162
Floppy Disk 152
Super I/O 160
Printer 196
Network Interface 170
Remote Computer 180
Removable Non-volatile memory 150
154 Storage Media
156 CD/DVD
153 USB
Non-removable Non-volatile memory 140
Storage token API 148
Operating System 144
Application Programs 145
Other Program Modules 146
Program Data 147

FIG. 1

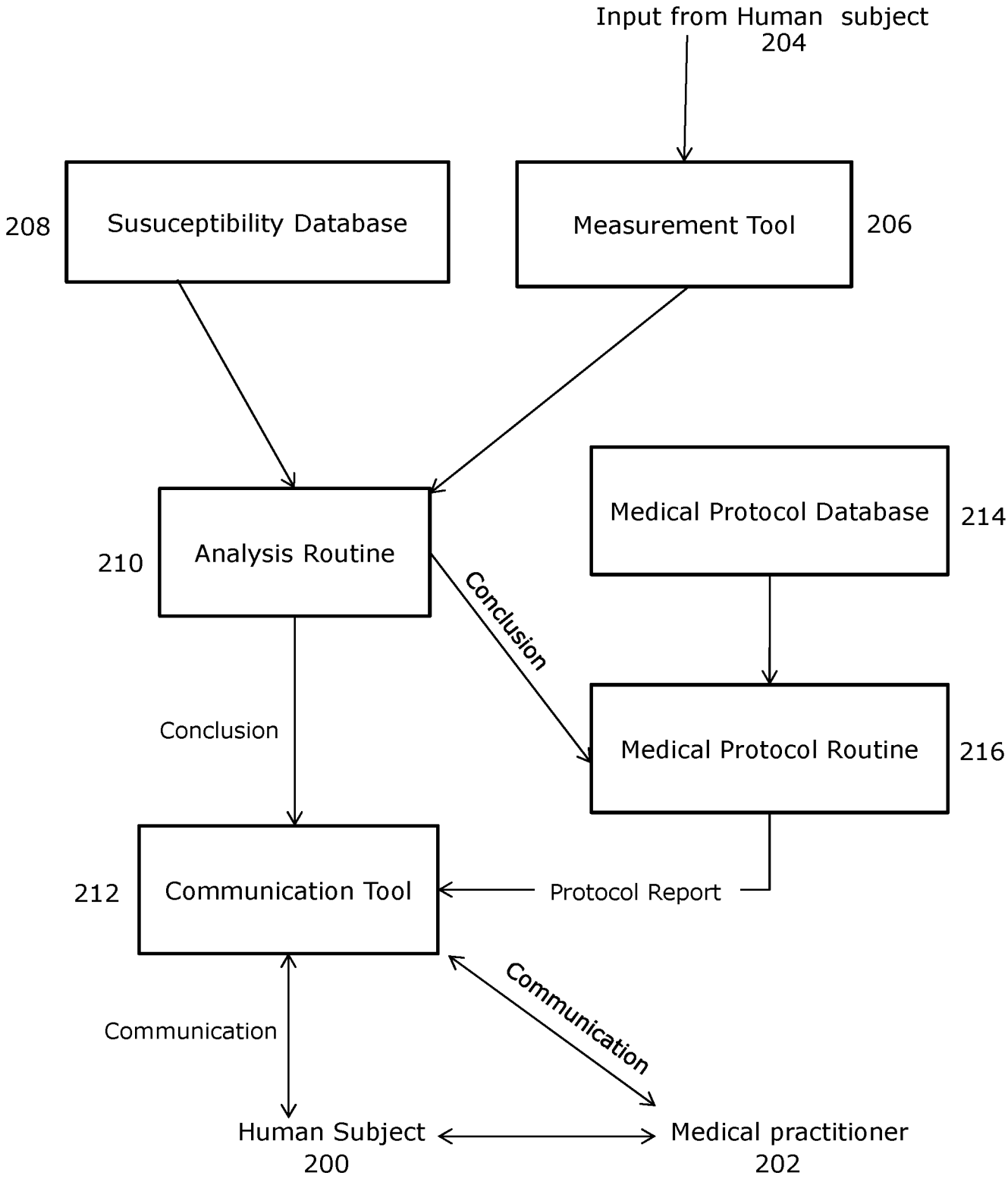

Input from Human subject
204
208
Susuceptibility Database
Measurement Tool
206
Medical Protocol Database
214
210
Analysis Routine
Conclusion
Conclusion
Medical Protocol Routine
216
212
Communication Tool
Protocol Report
Communication
Communication
Human Subject
200
Medical practitioner
202

FIG. 2

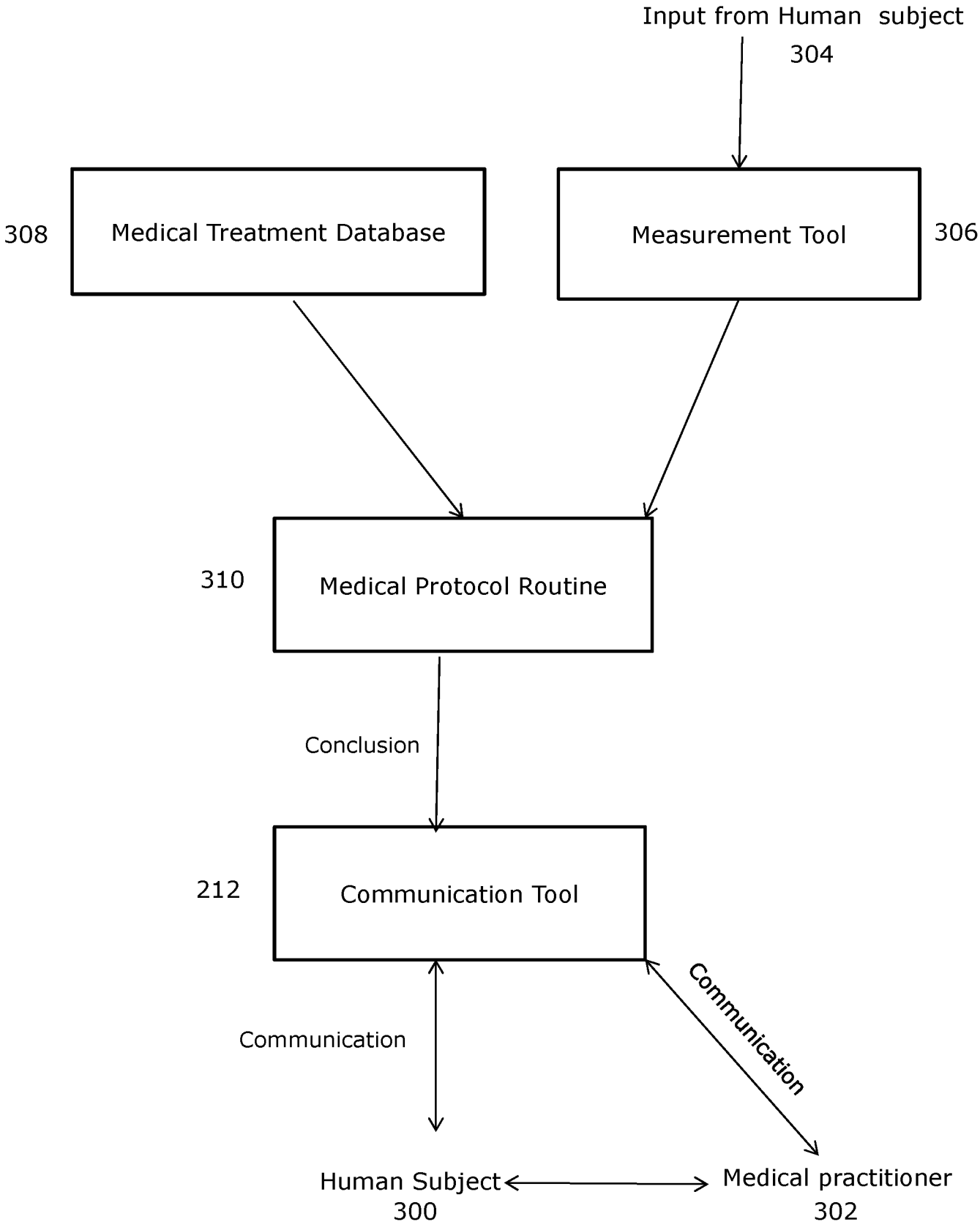

Input from Human subject
304
308
Medical Treatment Database
Measurement Tool
306
310
Medical Protocol Routine
Conclusion
212
Communication Tool
Communication
Communication
Human Subject
300
Medical practitioner
302

FIG. 3

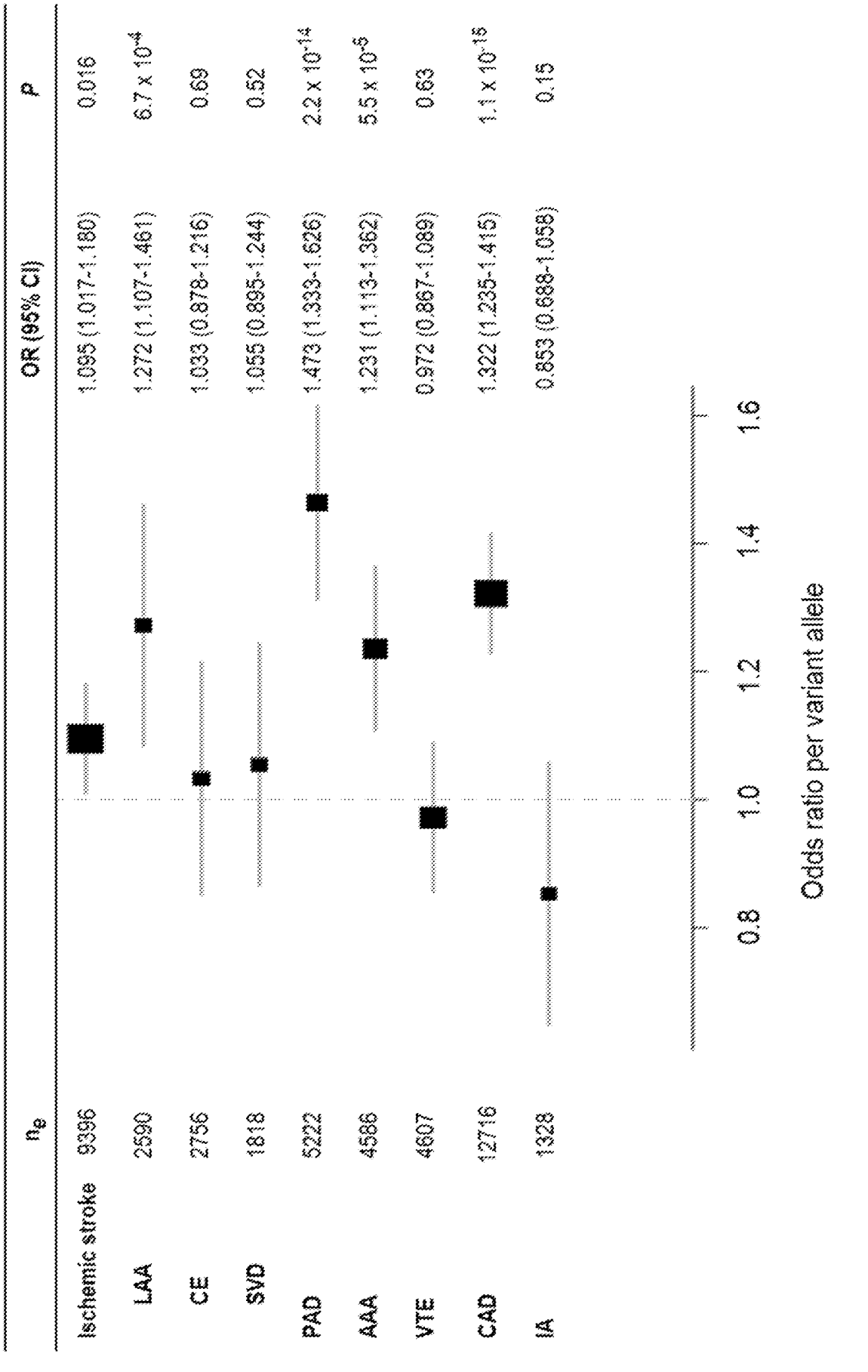

n_e
OR (95% CI)
P
Ischemic stroke 9396 1.095 (1.017-1.180) 0.016
LAA 2590 1.272 (1.107-1.461) 6.7×10^{-4}
CE 2756 1.033 (0.878-1.216) 0.69
SVD 1818 1.055 (0.895-1.244) 0.52
PAD 5222 1.473 (1.333-1.626) 2.2×10^{-14}
AAA 4586 1.231 (1.113-1.362) 5.5×10^{-5}
VTE 4607 0.972 (0.867-1.089) 0.63
CAD 12716 1.322 (1.235-1.415) 1.1×10^{-15}
IA 1328 0.853 (0.688-1.058) 0.15
0.8
1.0
1.2
1.4
1.6
Odds ratio per variant allele

FIG. 4

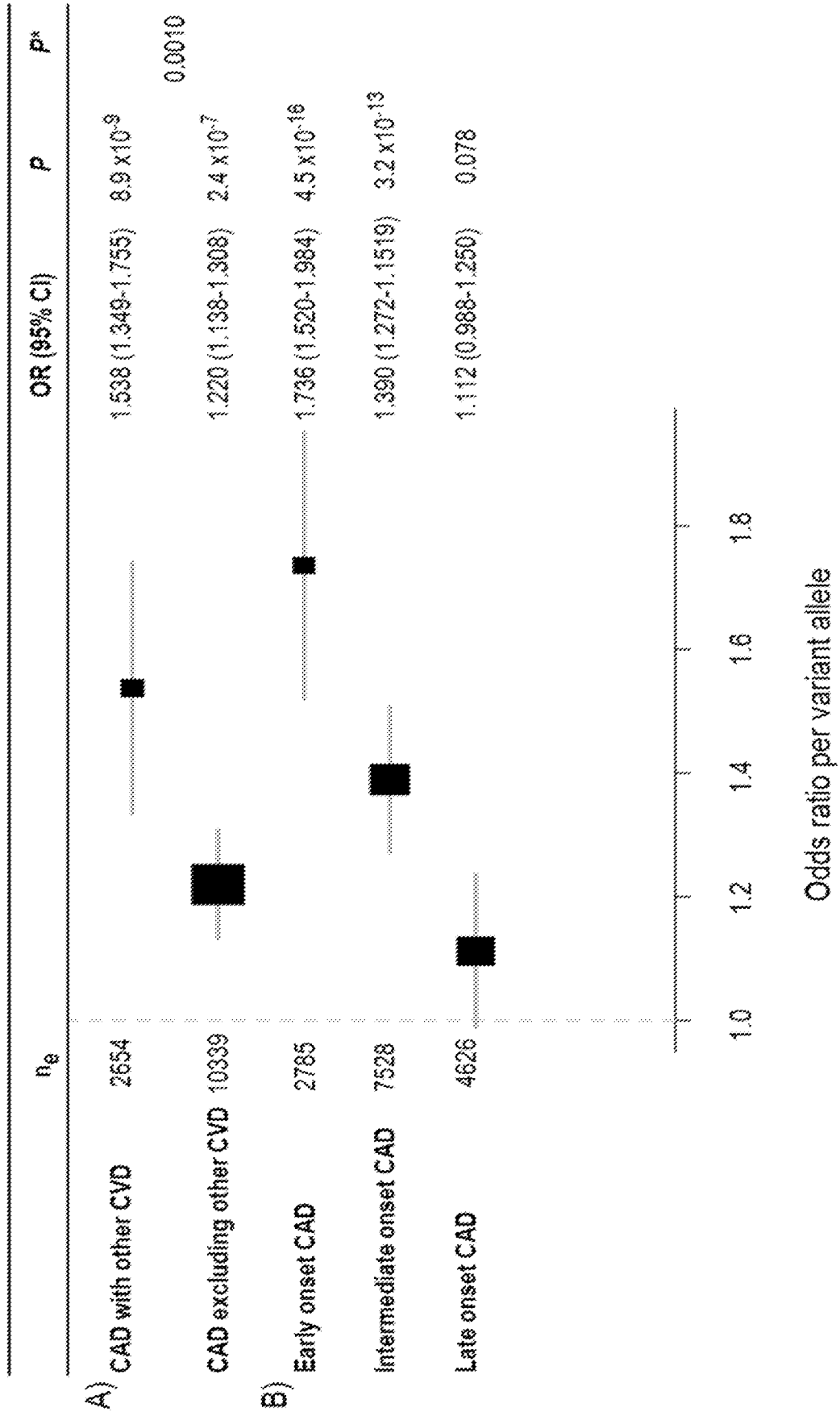

n_e
OR (95% CI)
P
P*
A)
CAD with other CVD
2654
1.538 (1.349-1.755)
8.9 x10-9
0.0010
CAD excluding other CVD
10339
1.220 (1.138-1.308)
2.4 x10-7
B)
Early onset CAD
2785
1.736 (1.520-1.984)
4.5 x10-16
Intermediate onset CAD
7528
1.390 (1.272-1.519)
3.2 x10-13
Late onset CAD
4626
1.112 (0.988-1.250)
0.078
1.0
1.2
1.4
1.6
1.8
Odds ratio per variant allele

FIG. 5

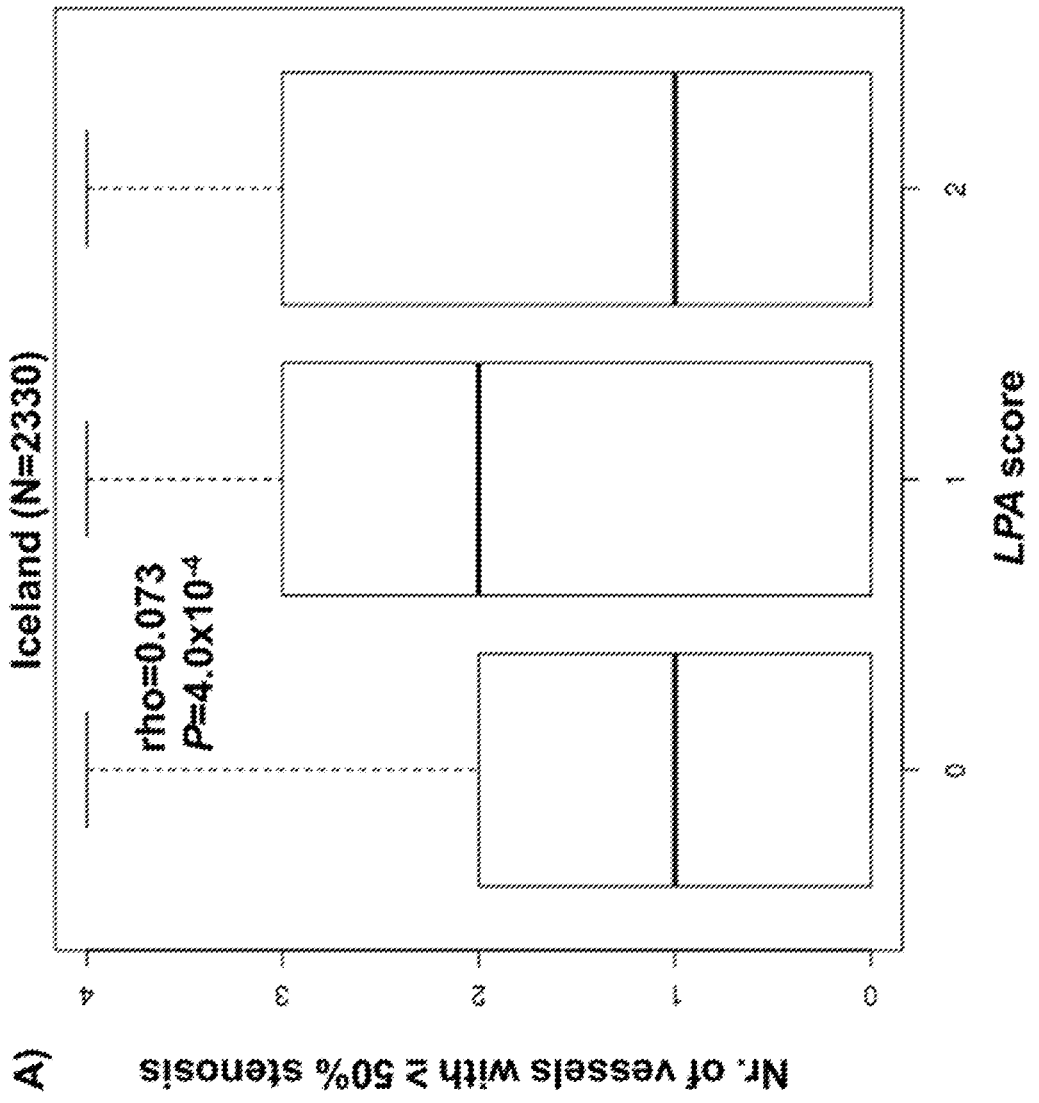

A)
Nr. of vessels with ≥ 50% stenosis
0
1
2
3
4
Iceland (N=2330)
rho=0.073
P=4.0x10-4
LPA score
0
1
2

FIG. 6 A)

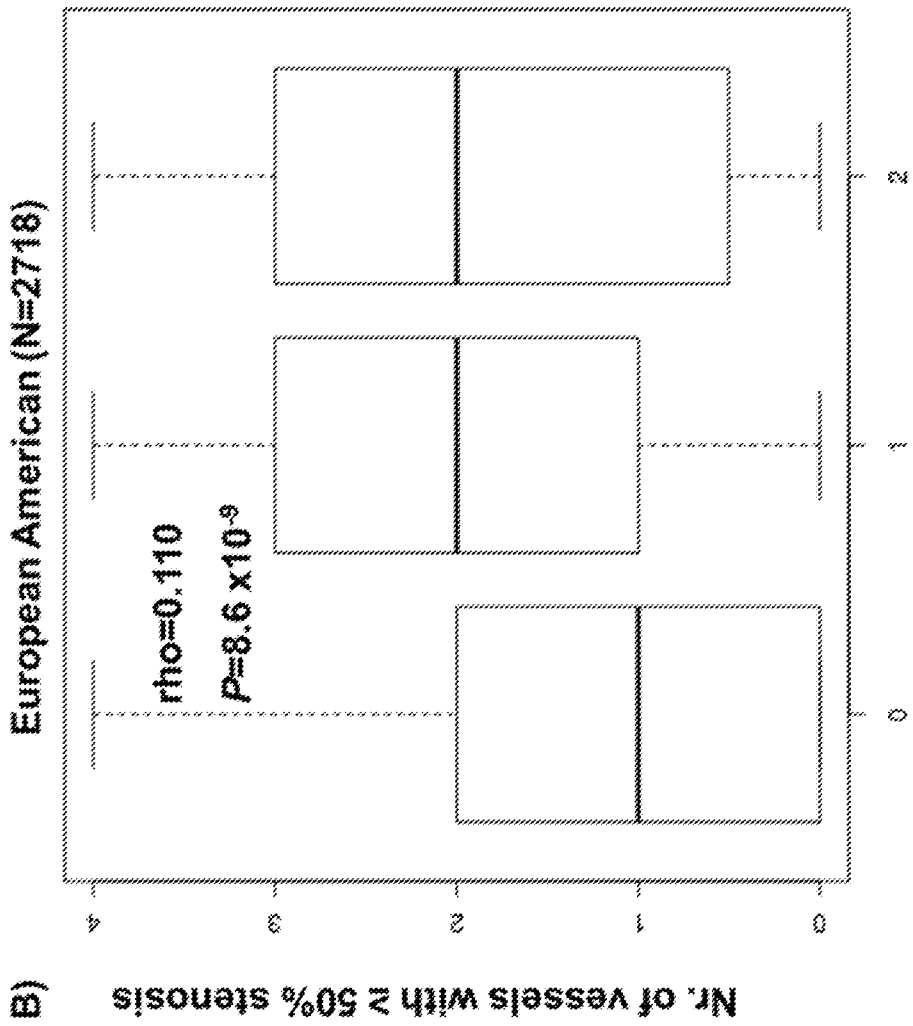

B)
Nr. of vessels with ≥ 50% stenosis
0
1
2
3
4
European American (N=2718)
rho=0.110
P=8.6 x10-9
0
1
2
LPA score

FIG. 6 B)

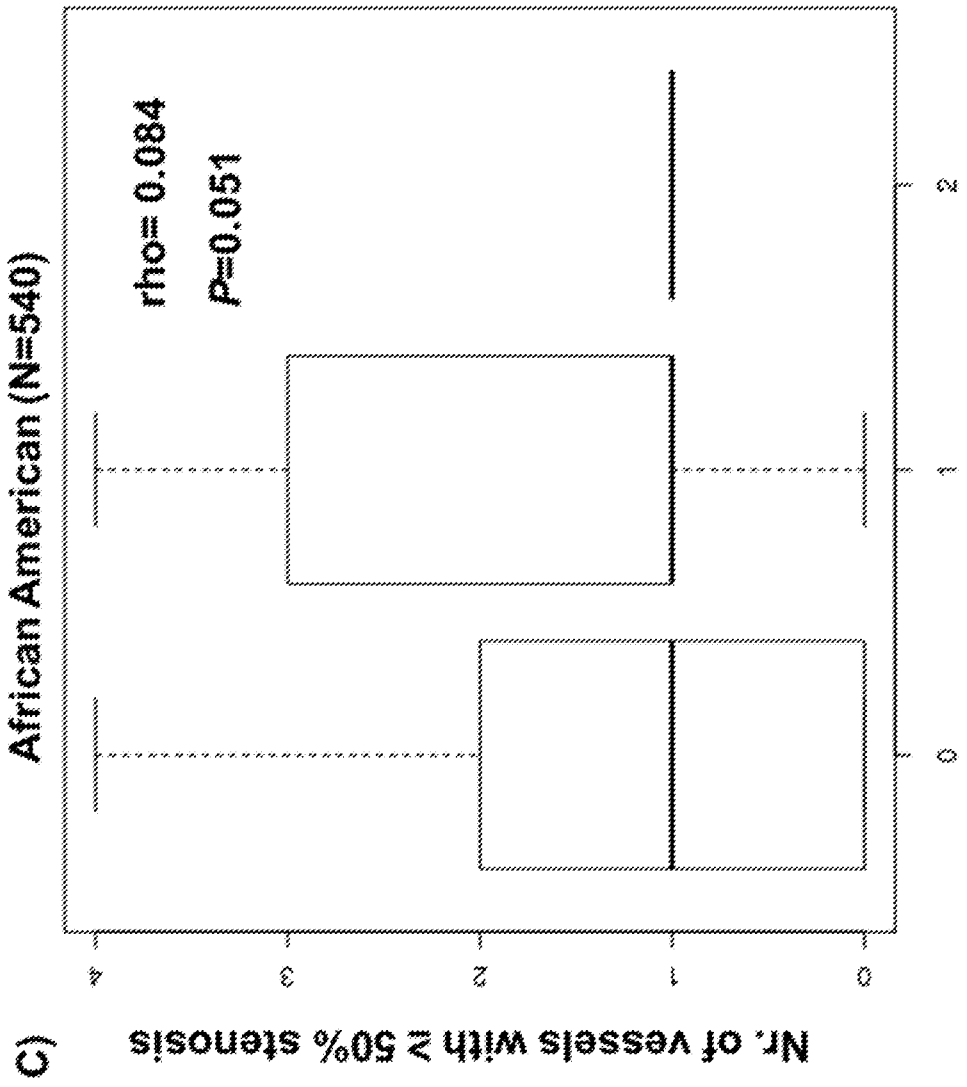

C)
Nr. of vessels with ≥ 50% stenosis
0
1
2
3
4
African American (N=540)
rho= 0.084
P=0.051
0
1
2
LPA score

FIG. 6 C)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IS20 12/0500 15

A. CLASSIFICATION OF SUBJECT MATTER
CI2Q 1/68 (2006.01)
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)
IPC/EC: C12Q

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

DK, NO, SE, FI: Classes as above

Electronic data base consulted during Ihe international search (name of data base and, where practicable, search terms used)
EPODOC, WPI, FULL TEXT: ENGLISH, GERMAN. FRENCH, XPESP, XPESP2. XPIOP, XPIPCOM, XPI3E, XPOAC, XPRD, XPTK. COMPDX. INSPEC. MEDLINE, PUBCOMP, PUBSUBS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Leibniz-Institut fur Arteriosklerosеforschung - LIFA, Research Report 2010, pages 1-40, http://www.lifa-muenster.de/fileadmin/redaktion/dokumente/LIFA-Forschunasbericht-20IO.pdf Page 22 SNP: rs3798220, rsl0455872	1, 4-8, 11-12, 17-22, 25, 32-33, 35-56
X	LUKE M. M. et al.: "A Polymorphism in the Protein-Like Domain of Apolipoprotein(a) la Associated With Severe Coronary Artery Disease", ARTERIOSCLEROSIS, THROMBOSIS, AND VASCULAR BIOLOGY, Vol. 27 (2007), pages 2030-2035 and Online Supplement, pages 1-17 Whole document SNP: rs3798220	1, 4-8, 11-12, 17-22, 25, 32-33, 35-56
X	CLARK R. et al.: "Genetic Variants Associated with Lp(a) Lipoprotein Level and Coronary Disease", THE NEW ENGLAND JOURNAL OF MEDICINE, Vol. 361, No. 26 (24/12/2009), pages 25 18-2528 Whole document; especially page 2518; page 2520, figure 1; page 2521, right hand column - page 2523 SNP: rs3798220, rs 10455872	1, 4-8, 11-12, 17-22, 25, 32-33, 35-56

☒ 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

Dale of the actual completion of the international search	Date of mailing of the international search report
20/02/2013	26/02/2013
Name and mailing address of the ISA Nordic Patent Institute Helgesh0j Alle 81 DK - 2630 Taastrup. Denmark. Facsimile No. + 45 43 50 80 08	Authorized officer Anita Thyrsted Telephone No. +45 43 50 8 1 32

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IS20 12/0500 15

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of ihe relevant passages	Relevant to claim No.
X	LI Y. et al.: "Genetic Variants in the Apolipoprotein(a) Gene and Coronary Heart Disease", CIRCULATION CARDIOVASCULAR GENETICS; JOURNAL OF THE AMERICAN HEART ASSOCIATION, Vol. 4 (October 201 1), pages 565-573 Whole document; especially page 476, right hand column; page 479, right hand column SNP: rs3798220, rsl0455872	1, 4-8, 11- 12, 17-22, 25, 32-33, 35-56
X	DAMANI. S. B. et al.: "Emerging Genomic Applications in Coronary Artery Disease", JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY, Vol. 4, No. 5 (May 2011), pages 473-482 Whole document SNP: rs3798220, rsl0455872	1, 4-8, 11- 12, 17-22, 25, 32-33, 35-56
P. X	HELGADOTTIR A. et el.: "Apolipoprotein(a) Genetic Sequence Variants Associated With Systemic Atherosclerosis and Coronary Atherosclerotic Burden But Not With Venous Thromboembolism", JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY, Vol. 60, No. 8 (21/08/2012), pages 722-729 Whole document SNP: rs3798220, rsl0455872	1, 4-8, 11-12, 17-22, 25, 32-33, 35-56
X	US 2010/0120045 A1 (HELGADOTTIR A. et al.) 13/05/2010 Whole document; especially claims; paragraph [0093] SNP: rsl 175 1605. rs3798220	1, 4-8, 11-12, 17-22, 25, 32-33, 35-56
X	US 2009/00041 87 A1 (RIDKER P. et al.) 01/01/2009 Whole document, especially example 2; paragraphs [0048], [0125], [0178] SNP: rs3798220, rs9457931, rs9457927	1, 4-8, 11- 12, 17-22, 25, 32-33, 35-56

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IS20 12/0500 15

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. S Claims Nos.: l, 4. 8, 11-12. 17-20, 22, 25. 5 1-56
because they relate to subject matter not required to be searched by this Authority, namely:

Claims 1. 4, 8, 11- 12, 17-20, 22, 25 and 5 1-56 relate to subject matter considered by this Authority to be covered by the provision of Rule 39. l(iv)/67. 1(iv) PCT. The methods comprise diagnostic methods of the human or animal body due to the phrasing "obtaining ... from the subject", "in a human subject", "in the subject", "from the human subject", "from the subject" or "from the individual". The search of claims 1, 4, 8, 11- 12, 17-20, 22, 25 and 5 1-56 has been carried out for these claims when relating to in vitro diagnostic methods only .

2. ☒ Claims Nos.: 51-54
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out. specifically:

The subject matter of claims 5 1-54 is unclear due to the expressions "from evidence" or the like and "markers in linkage disequilibrium with". It is not obvious for a person skilled in the art which markers fulfil this criterion. Thus, these expressions render the scope of the claims unclear. Accordingly, the search of claims 5 1-54 has been restricted to rs3798220 and rs 10455872 and the surrogate markers listed in table A and B.

3. □ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See extra sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention First mentioned in the claims; it is covered by claims Nos.:

1, 4-8, 11-12, 17-22, 25, 32-33, 35-56 (all partly)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/IS20 12/050015

Patent document cited in search report / Publication date	Patent family member(s) / Publication date
US201 0 I 20045 A I 201005 13	VVO2008 132763 A2 2008 I 106 WO2008132763 A3 20081218 EP2 155907 A2 20100224
US2009004 I87 A l 20090101	US201 1206667 A l 201 10825 WO2008 140776 A 1 2008 1120 US79433 17 B2 201 10517 JP2010528262 A 20100819 EP2150627 A l 20100210 EP2150627 A4 201 10223 CN101743325 A 20100616 CA2686874 A I 20081 I20

Continuation of Box No. III:

The International Searching Authority found multiple inventions in this international application as follows:

Invention 1 (claims 1, 4-8, 11-12, 17-22, 25, 32-33, 35-56 (all partly)): A method of determining a susceptibility to **large artery atherosclerosis,** comprising analyzing data representative of at least one allele of a polymorphism selected from rs3798220 and rsl0455872; a method of determining whether a human subject is at increased risk of developing **large artery atherosclerosis,** comprising determining in a biological sample the allelic identity of at least one polymorphic marker selected from rs3798220 and rsl0455872; a method of determining a susceptibility to **large artery atherosclerosis** in a human subject previously diagnosed with coronary disease, comprising analyzing data representative of at least one allele of a polymorphism selected from rs3798220 and rs 10455872; an assay for determining a susceptibility to **large artery atherosclerosis,** comprising determining, in a sample comprising nucleic acid from a subject, the presence or absence of at least one allele of a polymorphism selected from rs3798220 and rs 10455872; a system for identifying susceptibility to **large artery atherosclerosis,** comprising at least one processor, at least one computer-readable medium, a susceptibility database operatively coupled to a computer-readable medium and containing population information correlating the presence or absence of one or more alleles selected from allele **C** in rs3798220 and allele **G** in rs 10455872 to susceptibility of **large artery atherosclerosis** in a population, a measurement tool and an analyzis tool; a system for assessing or selecting a treatment protocol for a subject diagnosed with **large artery atherosclerosis,** comprising at least one processor, at least one computer-readable medium, a medical treatment database operatively connected to a computer-readable medium and containing information correlating the presence or absence of at least one allele selected from allele **C** in rs3798220 and allele **G** in rs 10455872 and efficacy of treatment regimens for **large artery atherosclerosis, a** measurement tool and an analyzis tool.

Invention 2 (claims 1, 4-8, 11-12, 17-22, 25, 32-33, 35-56 (all partly)): The same as for invention 1 but wherein the condition is **peripheral arterial disease.**

Invention 3 (claims 1, 4-8, 11-12, 17-22, 25, 32-33, 35-56 (all partly)): The same as for invention 1 but wherein the condition is **abdominal aortic aneurysm.**

Invention 4 (claims 1, 4-8, 11-12, 22, 25, 32-33, 35-56 (all partly); claims 3, 10, 24 (all completely)): A method of determining a susceptibility to **early-onset coronary artery disease,** comprising analyzing data representative of at least one allele of a polymorphism selected from rs3798220 and rsl0455872; a method of determining whether a human subject is at increased risk of developing **early-onset coronary artery disease,** comprising determining in a biological sample the allelic identity of at least one polymorphic marker selected from rs3798220 and rsl0455872; an assay for determining a susceptibility to **early-onset coronary artery disease,** comprising determining, in a sample comprising nucleic acid from a subject, the presence or absence of at least one allele of a polymorphism selected from rs3798220 and rsl0455872; a system for identifying susceptibility to **early-onset coronary artery disease,** comprising at least one processor, at least one computer-readable medium, a susceptibility database operatively coupled to a computer-readable medium and containing population information correlating the presence or absence of one or more alleles selected from allele **C** in rs3798220 and allele **G** in rsl0455872 to susceptibility of **early-onset coronary artery disease** in a population, a measurement tool and an analyzis tool; a system for assessing or selecting a treatment protocol for a subject diagnosed with **early-onset coronary artery disease,** comprising at least one processor, at least one computer-readable medium, a medical treatment database operatively connected to a computer-readable medium and containing information correlating the presence or absence of at least one allele selected from allele **C** in rs3798220 and allele **G** in rs 10455872 and efficacy of treatment regimens for **early-onset coronary artery disease,** a measurement tool and an analyzis tool.

Invention 5 (claims 1, 4-8, 11-12, 22, 25, 32-33, 35-56 (all partly); claims 2, 9, 13-16, 23, 34 (all completely)): A method of determining a susceptibility to **severe coronary artery disease,** comprising analyzing data representative of at least one allele of a polymorphism selected from rs3798220 and rsl0455872; a method of determining whether a human subject is at increased risk of developing **severe coronary artery disease,** comprising determining in a biological sample the allelic identity of at least one polymorphic marker selected from rs3798220 and rs 10455872; an assay for determining a susceptibility to **severe coronary artery disease,** comprising determining, in a sample comprising nucleic acid from a subject, the presence or absence of at least one allele of a polymorphism selected from rs3798220 and rs 10455872; a system for identifying susceptibility to **severe coronary artery disease,** comprising at least one processor, at least one computer-readable medium, a susceptibility database operatively coupled to a computer-readable medium and containing population information correlating the presence or absence of one or more alleles selected from allele **C** in rs3798220 and allele **G** in rs 10455872 to susceptibility of **severe coronary artery disease** in a population, a measurement tool and an analyzis tool; a system for assessing or selecting a treatment protocol for a subject diagnosed with **severe coronary artery disease,** comprising at least one processor, at least one computer-readable medium, a medical treatment database operatively connected to a computer-readable medium and containing information correlating the presence or absence of at least one allele selected from allele **C** in rs3798220 and allele **G** in rsI0455872 and efficacy of treatment regimens for **severe coronary artery disease,** a measurement tool and an analyzis tool.

Invention 6 (claims 26-31 (all completely)): A method of determining the prognosis of a human subject diagnosed with **coronary artery disease,** comprising the steps of determining, in a biological sample, the allelic identity of at least one polymorphic marker selected from rs3798220 and rs 10455872.

The common technical feature between inventions 1-6 is the use of a polymorphic marker selected from rs3798220 and rsl0455872 for determining the susceptibility to, the risk of developing, or the prognosis of a an arterial condition. However, this feature is already known in the art.

As stated in the present application page 8, lines 14-17, the single nucleotide polymorphism rs3798220 and rsl0455872 in the coding sequence of the LPA gene are associated with coronary artery disease. In Luke M. M. et al.: "A Polymorphism in the Protein-Like Domain of Apolipoprotein(a) Ia Associated With Severe Coronary Artery Disease", ARTERIOSCLEROSIS, THROMBOSIS, AND VASCULAR BIOLOGY, Vol. 27 (2007), pages 2030-2035 and Online Supplement, pages 1-17, an association of rs3798220 (I4399M) with severe coronary artery disease was shown. Likewise, both the C allele of rs3798220 and the G allele of rsl0455872 have been associated with coronary disease, e.g. coronary artery disease, see Clark R. et al.: "Genetic Variants Associated with Lp(a) Lipoprotein Level and Coronary Disease", THE NEW ENGLAND JOURNAL OF MEDICINE, Vol. 361, No. 26 (24/12/2009), pages 25 18-2528.

As rs3798220 and rs 10455872 are known to be associated with one of the arterial conditions in the claims, lack of unity arise between the six arterial conditions in the claims. Thus, the application does not meet the requirement of unity of inventions as defined in Rules 13.1 and 13.2 PCT.