GENETIC SUSCEPTIBILITY VARIANTS ASSOCIATED WITH CARDIOVASCULAR DISEASE

BACKGROUND OF THE INVENTION

Coronary artery disease and myocardial infarction

The major complications of Coronary Artery Disease, i.e. Myocardial infarction (MI) and Acute Coronary Syndrome (ACS), are the leading causes of hospital admissions in industrialized countries. Cardiovascular disease continues to be the principle cause of death in the United States, Europe and Japan. The costs of the disease are high both in terms of morbidity and mortality, as well as in terms of the financial burden on health care systems.

Myocardial infarction generally occurs when there is an abrupt decrease in coronary blood flow following a thrombotic occlusion of a coronary artery previously damaged by atherosclerosis (i.e. in subjects with coronary artery disease). In most cases, infarction occurs when an atherosclerotic plaque fissures, ruptures or ulcerates and when conditions favor thrombogenesis. In rare cases, infarction may be due to coronary artery occlusion caused by coronary emboli, congenital abnormalities, coronary spasm, and a wide variety of systemic, particularly inflammatory diseases. Medical risk factors for MI include cigarette smoking, diabetes, hypertension and serum total cholesterol levels > 200 mg/dL, elevated serum LDL cholesterol, and low serum HDL cholesterol. Event rates in individuals without a prior history of cardiovascular disease are about 1%. In individuals who have had a first MI or ACS, the risk of a repeat MI within the next year is 10-14%, despite maximal medical management including angioplasty and stent placement.

Atherosclerosis can affect vascular beds in many large and medium arteries. Myocardial infarction and unstable angina (acute coronary syndrome (ACS)) stem from coronary artery atherosclerosis (Coronary Artery Disease), while ischemic stroke most frequently is a consequence of carotid or cerebral artery atherosclerosis. Limb ischemia caused by peripheral arterial occlusive disease (PAOD) may occur as a consequence of iliac, femoral and popliteal artery atherosclerosis. The atherosclerotic diseases remain common despite the wide-spread use of medications that inhibit thrombosis (aspirin) or treat medical risk factors such as elevated cholesterol levels in blood (statins), diabetes, or hypertension (diuretics and anti-hypertensives).

Atherosclerotic disease is initiated by the accumulation of lipids within the artery wall, and in particular, the accumulation of low-density lipoprotein (LDL) cholesterol. The trapped LDL becomes oxidized and internalized by macrophages. This causes the formation of atherosclerotic lesions containing accumulations of cholesterol-engorged macrophages, referred to as "foam

cells". As disease progresses, smooth muscle cells proliferate and grow into the artery wall forming a "fibrous cap" of extracellular matrix enclosing a lipid-rich, necrotic core. Present in the arterial walls of most people throughout their lifetimes, fibrous atherosclerotic plaques are relatively stable. Such fibrous lesions cause extensive remodeling of the arterial wall, outwardly displacing the external, elastic membrane, without reduction in luminal diameter or serious impact on delivery of oxygen to the heart. Accordingly, patients can develop large, fibrous atherosclerotic lesions without luminal narrowing until late in the disease process. However, the coronary arterial lumen can become gradually narrowed over time and in some cases compromise blood flow to the heart, especially under high demand states such as exercise. This can result in reversible ischemia causing chest pain relieved by rest called stable angina.

In contrast to the relative stability of fibrous atherosclerotic lesions, the culprit lesions associated with myocardial infarction and unstable angina (each of which are part of the acute coronary syndrome) are characterized by a thin fibrous cap, a large lipid core, and infiltration of inflammatory cells such as T-lymphocytes and monocyte/macrophages. Non-invasive imaging techniques have shown that most MI´s occur at sites with low- or intermediate- grade stenoses, indicating that coronary artery occlusion is due most frequently to rupture of culprit lesions with consequent formation of a thrombus or blood clot and not solely due to luminal narrowing by stenosis. Plaque rupture may be due to erosion or uneven thinning of the fibrous cap, usually at the margins of the lesion where macrophages enter, accumulate, and become activated by a local inflammatory process. Thinning of the fibrous cap may result from degradation of the extracellular matrix by proteases released from activated macrophages. These changes producing plaque instability and risk of MI may be augmented by production of tissue-factor procoagulant and other factors increasing the likelihood of thrombosis.

In acute coronary syndrome, the culprit lesion showing rupture or erosion with local thrombosis typically is treated by angioplasty or by balloon dilation and placement of a stent to maintain luminal patency. Patients experiencing ACS are at high risk for a second coronary event due to the multi-vessel nature of coronary artery disease with event rates approaching 10-14% within 12 months after the first incident.

The emerging view of MI is as an inflammatory disease of the arterial vessel wall on preexisting chronic atherosclerotic lesions, sometimes triggering rupture of culprit lesions and leading to local thrombosis and subsequent myocardial infarction. The process that triggers and sustains arterial wall inflammation leading to plaque instability is unknown, however, it results in the release into the circulation of tumor necrosis factor alpha and interleukin-6. These and other cytokines or biological mediators released from the damaged vessel wall stimulate an inflammatory response in the liver causing elevation in several non-specific general inflammatory markers including C-reactive protein. Although not specific to atherosclerosis, elevated C-reactive protein (CRP) and serum amyloid A appear to predict risk for MI, perhaps as surrogates for vessel wall inflammation. Many general inflammatory markers predict risk of coronary heart disease, although these markers are not specific to atherosclerosis. For example, Stein (Stein,

S., *Am J Cardiol , 87* (suppl):21A-26A (2001)) discusses the use of any one of the following serum inflammatory markers as surrogates for predicting risk of coronary heart disease including C-reactive protein (CRP), serum amyloid A, fibrinogen, interleukin-6, tissue necrosis factor-alpha, soluble vascular cell adhesion molecules (sVCAM), soluble intervascular adhesion molecules (sICAM), E-selectin, matrix metalloprotease type-1, matrix metalloprotease type-2, matrix metalloprotease type-3, and matrix metalloprotease type-9. Elevation in one more of these serum inflammatory markers is not specific to coronary heart disease but also occurs with age or in association with cerebrovascular disease, peripheral vascular disease, non-insulin dependent diabetes, osteoarthritis, bacterial infection, and sepsis.

Elevated CRP or other serum inflammatory markers is also prognostic for increased risk of a second myocardial infarct in patients with a previous myocardial infarct (Retterstol, L. *et al., Atheroscler., 160*: 433-440 (2002)).

Although classical risk factors such as smoking, hyperlipidemia, hypertension, and diabetes are associated with many cases of coronary heart disease (CHD) and MI, many patients do not have involvement of these risk factors. In fact, many patients who exhibit one or more of these risk factors do not develop MI. Family history has long been recognized as one of the major risk factors. Although some of the familial clustering of MI reflects the genetic contribution to the other conventional risk factors, a large number of studies have suggested that there are significant genetic susceptibility factors, beyond those of the known risk factors (Friedlander Y, *et al., Br. Heart J.* 1985; 53:382-7, Shea S. *et al., J. Am. Coll. Cardiol.* 1984; 4:793-801, and Hopkins P.N., *et al., Am. J. Cardiol.* 1988; 62:703-7). Major genetic susceptibility factors have only been identified for the rare Mendelian forms of hyperlipidemia such as a familial hypercholesterolemia.

Genetic risk is conferred by subtle differences in genes among individuals in a population. Genes differ between individuals most frequently due to single nucleotide polymorphisms (SNP), although other variations are also important. SNP are located on average every 1000 base pairs in the human genome. Accordingly, a typical human gene containing 250,000 base pairs may contain 250 different SNP. Only a minor number of SNPs are located in exons and alter the amino acid sequence of the protein encoded by the gene. Most SNPs may have little or no effect on gene function, while others may alter transcription, splicing, translation, or stability of the mRNA encoded by the gene. Additional genetic polymorphism in the human genome is caused by insertion, deletion, translocation, or inversion of either short or long stretches of DNA. Genetic polymorphisms conferring disease risk may therefore directly alter the amino acid sequence of proteins, may increase the amount of protein produced from the gene, or may decrease the amount of protein produced by the gene.

As genetic polymorphisms conferring risk of disease are uncovered, genetic testing for such risk factors is becoming important for clinical medicine. Examples are apolipoprotein E testing to identify genetic carriers of the apoE4 polymorphism in dementia patients for the

differential diagnosis of Alzheimer's disease, and of Factor V Leiden testing for predisposition to deep venous thrombosis. More importantly, in the treatment of cancer, diagnosis of genetic variants in tumor cells is used for the selection of the most appropriate treatment regime for the individual patient. In breast cancer, genetic variation in estrogen receptor expression or heregulin type 2 (Her2) receptor tyrosine kinase expression determine if anti-estrogenic drugs (tamoxifen) or anti-Her2 antibody (Herceptin) will be incorporated into the treatment plan. In chronic myeloid leukemia (CML) diagnosis of the Philadelphia chromosome genetic translocation fusing the genes encoding the Bcr and Abl receptor tyrosine kinases indicates that Gleevec (STI571), a specific inhibitor of the Bcr-Abl kinase should be used for treatment of the cancer. For CML patients with such a genetic alteration, inhibition of the Bcr-Abl kinase leads to rapid elimination of the tumor cells and remission from leukemia.

Restenosis

Coronary balloon angioplasty was introduced in the late 1970s as a less invasive method for revascularization of coronary artery disease patients than the coronary artery bypass graft (CABG) surgeries. Since then there has been a quick progress in the development of new percutaneous devices to revascularize areas with limited blood flow. However, the expanded use of angioplasty has shown that the arteries react to angioplasty by a proliferative process that limits the success of this treatment. This process is known as restenosis.

Restenosis is defined as a re-narrowing of the treated segment, which equals or exceeds 50% of the lumen in the adjacent normal segment of the artery. Depending on the patient population studied, the restenosis rates range from 30% to 44% of lesions treated by balloon dilation. This problem prompted a search for interventional techniques that minimizes the risk of restenosis. Several clinical trials have shown a significant reduction in the restenosis rates with endovascular stenting. The purpose of stenting is to maintain the arterial lumen by a scaffolding process that provides radial support. Stents, usually made of stainless steel, are placed in the artery either by a self-expanding mechanism or, using balloon expansion. However, in-stent restenosis still remains a major problem in the field of percutaneous, transluminal coronary angioplasty (PTCA), requiring patients to undergo repeated procedures and surgery. Restenosis is the result of the formation of neointima, a composition of smooth muscle-like cells in a collagen matrix. The current treatment modalities for in-stent restenosis include repeat balloon angioplasty, repeat stenting, cutting balloon angioplasty, directional coronary atherectomy, rotational coronary atherectomy, brachytherapy, and drug-eluting stents (DES). The restenosis problem can be minimised by local intravascular irradiation (intracoronary brachytherapy) and by the introduction of DES and these treatments have been shown to successfully preventing cell proliferation after stent implantation or angioplasty.

Intracoronary brachytherapy is a treatment in which sealed sources of radioactive material are used to deliver radiation at a very short distance by placing them in the artery lumen at the site of the atherosclerotic lesion. The physical benefit of brachytherapy is that

doses of radiation can be delivered almost directly to the target with a very rapid falloff of dose to the surrounding normal tissue. The rationale underlining this modality is based on the ability of ionizing radiation to inhibit cell proliferation, in this case, the proliferation of smooth muscle cells that tend to form a neointima. In the near future, it would be important to be able to classify patients with respect to the risk of having in-stent restenosis. This classification can potentially be made on the basis of genetic risk factors. The outcome of the classification may determine which therapy is most appropriate and also where coronary bypass surgery has to be considered.

Aneurysms

Degenerative changes of the arterial wall may cause localized dilatation, or aneurysm, of the artery, including abdominal aorta aneurysm (AAA) and intracranial aneurysm (IA). Atherosclerotic changes of the vessel wall are found in the majority of AAA that are characterized histopathologically by chronic inflammation, destructive remodelling of elastic media and depletion of medial smooth muscle cells resulting in marked weakening of the aortic wall. In contrast, berry aneurysms of intracranial arteries are not associated with atherosclerosis. Furthermore, the histopathological features of IA are different. The typical berry aneurysms of intracranial arteries, located at arterial bifurcations, have a thin, or no, media and the internal elastic lamina is either absent or severely fragmented.

Both AAA and IA represent a degenerative process of the arteries leading to their enlargement that is usually asymptomatic with natural history culminating in either a therapeutic intervention or rupture. Rupture of IA leads to subarachnoid haemorrhage, and rupture of both IA and AAA have high morbidity and mortality. In the case of AAA the rupture risk increases with the growth rate as well as the size of the aneurysm.

Intracranial aneurysm (IA), also called cerebral aneurysm or brain aneurysm is a cerebrovascular disorder in which weakness in the wall of a cerebral artery or vein causes a localized dilation or ballooning of the blood vessel.

A common location of cerebral aneurysms is on the arteries at the base of the brain, known as the Circle of Willis. Approximately 85% of cerebral aneurysms develop in the anterior part of the Circle of Willis, and involve the internal carotid arteries and their major branches that supply the anterior and middle sections of the brain. It is believed that aneurysms may result from congenital defects, preexisting conditions such as high blood pressure and atherorsclerosis, or head trauma. Cerebral aneurysms occur more commonly in adults than in children but they may occur at any age.

Cerebral aneurysms are classified both by size and shape. Small aneurysms have a diameter of less than 15mm. Larger aneurysms include those classified as large (15 to 25mm),

giant (25 to 50mm), and super giant (over 50mm). Saccular aneurysms are those with a saccular outpouching and are the most common form of cerebral aneurysm. Berry aneurysms are saccular aneurysms with necks or stems resembling a berry. Fusiform aneurysms are aneurysms without stems.

A small, unchanging aneurysm will produce no symptoms. Before a larger aneurysm ruptures, the individual may experience such symptoms as a sudden and unusually severe headache, nausea, vision impairment, vomiting, and loss of consciousness, or the individual may be asymptomatic, experiencing no symptoms at all. Onset is usually sudden and without warning. Rupture of a cerebral aneurysm is dangerous and usually results in bleeding into the meninges or the brain itself, leading to a subarachnoid hemorrhage (SAH) or intracranial hematoma (ICH), either of which constitutes a stroke. Rebleeding, hydrocephalus (the excessive accumulation of cerebrospinal fluid), vasospasm (spasm, or narrowing, of the blood vessels), or multiple aneurysms may also occur. The risk of rupture from an unruptured cerebral aneurysm varies according to the size of an aneurysm, with the risk rising as the aneurysm size increases. The overall rate of aneurysm rupture is estimated at 1.3% per year. The risk of short term re-rupture increases dramatically after an aneurysm has bled, though after approximately 6 weeks the risk returns to baseline.

Emergency treatment for individuals with a ruptured cerebral aneurysm generally includes restoring deteriorating respiration and reducing intracranial pressure. Currently there are two treatment options for brain aneurysms: surgical clipping or endovascular coiling. Either surgical clipping or endovascular coiling is usually performed within the first three days to occlude the ruptured aneurysm and reduce the risk of rebleeding.

The prognosis for a patient with a ruptured cerebral aneurysm depends on the extent and location of the aneurysm, the person's age, general health, and neurological condition. Some individuals with a ruptured cerebral aneurysm die from the initial bleeding. Other individuals with cerebral aneurysm recover with little or no neurological deficit. The most significant factors in determining outcome are severity of the aneurysm and age.

Abdominal aortic aneurysm (AAA) is a localized dilatation of the abdominal aorta, that exceeds the normal diameter by more than 50%. The normal diameter of the infrarenal aorta is 2 cm. It is caused by a degenerative process of the aortic wall. The aneurysm is most commonly located infrarenally (90%), other possible locations are suprarenal and pararenal. The aneurysm can extend to include one or both of the iliac arteries. An aortic aneurysm may also occur in the thorax.

AAA is uncommon in individuals of African, African American, Asian, and Hispanic heritage. The frequency varies strongly between males and females. The peak incidence is among males around 70 years of age, the prevalence among males over 60 years totals 2-6%. The frequency is much higher in smokers than in non-smokers (8:1). Other risk factors include

hypertension and male sex. In the US, the incidence of AAA is 2-4% in the adult population. Rupture of the AAA occurs in 1-3% of men aged 65 or more, the mortality being 70-95%.

The exact causes of the degenerative process remain unclear. Known risk factors include genetic factors, hemodynamic influences, atherosclerosis, and various other factors such as infection, trauma, connective tissue disorders, arterities, etc. AAAs are commonly divided according to their size and symptomatology. An aneurysm is usually considered to be present if the measured outer aortic diameter is over 3 cm (normal diameter of aorta is around 2 cm). The natural history is of increasing diameter over time, followed eventually by the development of symptoms (usually rupture). If the outer diameter exceeds 5 cm, the aneurysm is considered to be large. For aneurysms under 5 cm, the risk of rupture is low, so that the risks of surgery usually outweigh the risk of rupture. Aneurysms less than 5cm are therefore usually kept under surveillance until such time as they become large enough to warrant repair, or develop symptoms. The vast majority of aneurysms are asymptomatic. The risk of rupture is high in a symptomatic aneurysm, which is therefore considered an indication for surgery. Possible symptoms include low back pain, flank pain, abdominal pain, groin pain or pulsating abdominal mass. The complications include rupture, peripheral embolisation, acute aortic occlusion, aortocaval or aortoduodenal fistulae. On physical examination, a palpable abdominal mass can be noted. Bruits can be present in case of renal or visceral arterial stenosis.

The main treatment options for asymptomatic AAA are immediate repair and surveillance with a view to eventual repair. Surveillance is indicated in small aneurysms, where the risk of repair exceeds the risk of rupture. As an AAA grows in diameter the risk of rupture increases. Although some controversy exists around the world, most vascular surgeons would not consider repair until the aneurysm reached a diameter of 5cm. The threshold for repair varies slightly from individual to individual, depending on the balance of risks and benefits when considering repair versus ongoing surveillance. The size of an individual's native aorta may influence this, along with the presence of comorbitities that increase operative risk or decrease life expectancy. Currently, the main modes of repair available for an AAA are open aneurysm repair (OR), and endovascular aneurysm repair (EVAR). Open repair is indicated in young patients as an elective procedure, or in growing or large, symptomatic or ruptured aneurysms. Open repair has been the mainstay of intervention from the 1950's until recently. Endovascular repair first became practical in the 1990's and although it is now an established alternative to open repair, its role is yet to be clearly defined. It is generally indicated in older, high-risk patients or patients unfit for open repair. However, endovascular repair is feasible for only a proportion of AAA's, depending on the morphology of the aneurysm. The main advantage over open repair is that the peri-operative period has less impact on the patient.

Stroke

Stroke is a group of diverse disorders encompassing several pathophysiological mechanisms. The clinical phenotype of stroke is complex but is broadly divided into: ischemic

and hemorrhagic stroke. The majority of stroke events, appr 80%, is due to ischemia (cerebral infarction), that occurs when a cerebral artery becomes completely occluded and the blood supply to a part of the brain is totally or partially blocked (due to thrombosis or an embolism). Ischemic stroke is further subdivided into large artery disease (LAA) (also called large vessel disease, LVD), cardioembolic stroke and small vessel disease. Approximately 25% of ischemic stroke events are due to large-artery disease of the carotid and vertebral arteries, the two pairs of large arteries that supply the brain with blood. The most common cause of large-artery disease is atherosclerosis. Cardioembolic strokes are caused by an embolism that originates inside the heart. Embolism of cardiac origin accounts for about 1/4 of ischemic strokes. Strokes due to cardioembolism are in general severe and prone to early and long-term recurrence. Ischemic heart disease, rheumatic mitral stenosis, and prosthetic cardiac valves are major sources of cardioembolic stroke but atrial fibrillation remains the commonest cause.

There is a continued and great need to understand the genetic variants conferring risk (increased and decreased) of the cardiovascular diseases. The present invention provides genetic variants that have been shown to be associated with susceptibility to cardiovascular disease, including MI, Coronary Artery Disease (CAD), Intracranial aneurysm (IA), Abdominal Aorta Aneurysm (AAA), Peripheral Arterial Disease (PAD) and Restenosis. These variants are useful in risk management and methods for therapeutic intervention of cardiovascular diseases.

SUMMARY OF THE INVENTION

The present invention relates to methods of determining a susceptibility to cardiovascular diseases, including Coronary Artery Disease, Myocardial Infarction, Peripheral Artery Disease, Stroke, Restenosis, Intracranial Aneurysm and Abdominal Aorta Aneurysm. The invention also relates to various uses, kits, procedures and apparati useful in the determination of a susceptibility to cardiovascular disease based on evaluation of certain polymorphic markers and/or haplotypes that have been found to be associated with susceptibility to cardiovascular disease.

In one aspect, the invention relates to a method for determining a susceptibility to cardiovascular disease in a human individual, 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 derived from the individual, wherein the at least one polymorphic marker is selected from the polymorphic markers set forth in Table 10, and markers in linkage disequilibrium therewith, and wherein the presence of the at least one allele is indicative of a susceptibility to cardiovascular disease. The method may in one embodiment relate to determination of the presence or absence of at least one allele of at least one polymorphic marker in a nucleic acid sample obtained from the individual. In another embodiment, the method relates to determination of the presence or absence of at least one

allele of at least one polymorhpic marker in a genotype dataset derived from the individual. The genotype dataset is derived from the individual in the sense that the information that is relates to a particular nucleic acid sample as a template relates to a single individual, for whom genetic information is derived.

In another aspect, the present invention relates to a method of determining a susceptibility to cardiovascular disease in a human individual, 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, wherein the at least one polymorphic marker is selected from markers associated with LD Block C09 (SEQ ID NO:94), wherein determination of the presence or absence of the at least one allele is indicative of a susceptibility to cardiovascular disease. In one embodiment, the at least one polymorphic marker is selected from the markers set forth in Table 3, and markers in linkage disequilibrium therewith.

In an alternative aspect, the invention relates to a method of diagnosing a susceptibility to Cardiovascular Disease in a human individual, the method 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 derived from the individual, wherein the at least one polymorphic marker is selected from the group of markers associated with LD block C09, wherein the presence of the at least one allele is indicative of a susceptibility to Cardiovascular Disease. In one embodiment, linkage disequilibrium is used as a quantitative measure of the degree to which specific markers are associated with LD Block C09.

In another aspect, the invention relates to a method of determining a susceptibility to cardiovascular disease in a human individual, comprising determining whether at least one at-risk allele in at least one polymorphic marker is present in a genotype dataset derived from the individual, wherein the at least one polymorphic marker is selected from the markers within the LD Block C09 (SEQ ID NO:94), and markers in linkage disequilibrium therewith, and wherein determination of the presence of the at least one at-risk allele is indicative of increased susceptibility to cardiovascular disease in the individual.

The genotype dataset comprises in one embodiment information about marker identity, and the allelic status of the individual, i.e. information about the identity of the two alleles carried by the individual for the marker. The genotype dataset may comprise allelic information about one or more marker, including two or more markers, three or more markers, five or more markers, one hundred or more markers, etc. In some embodiments, the genotype dataset comprises genotype information from a whole-genome assessment of the individual including hundreds of thousands of markers, or even one million or more markers.

In one embodiment, the at least one polymorphic marker is present within the genomic segment LD Block C09, with the nucleotide sequence as set forth in SEQ ID NO:94. In another embodiment, the at least one polymorphic marker comprises at least one marker selected from rs7041637, rs2811712, rs3218018, rs3217992, rs2069426, rs2069422, rs1333034, rs1011970,

rs10116277, rs1333040, rs2383207, rs1333050, D9S1814, rs10757278, rs10757274, rs1333049, D9S1870, and markers in linkage disequilibrium therewith. In another embodiment, the at least one polymorphic marker is selected from rs10757278, rs10757274, and rs10333049, and markers in linkage disequilibrium therewith. In another embodiment, the at least one polymorphic marker comprises at least one marker in strong linkage disequilibrium, as defined by numeric values for |D'| of greater than 0.8 and/or r^2 of greater than 0.2, with one or more markers selected from the group consisting of the markers set forth in Table 3.

In one embodiment, the method of determining a susceptibility, or diagnosing a susceptibility of, cardiovascular disease, further comprises assessing the frequency of at least one haplotype in the individual. In one such embodiment, the at least one haplotype is selected from the haplotypes that comprise at least one polymorphic marker within the genomic segment LD Block C09 (SEQ ID NO:94). In another embodiment, the at least one haplotype is selected from haplotypes that are in linkage disequilibrium with at least one marker as set forth in Table 3. In another embodiment, the at least one haplotype is selected from the haplotypes that comprise at least one polymorphic marker selected from at least one marker selected from rs7041637, rs2811712, rs3218018, rs3217992, rs2069426, rs2069422, rs1333034, rs1011970, rs10116277, rs1333040, rs2383207, rs1333050, D9S1814, rs10757278, rs10757274, rs1333049, D9S1870, and markers in linkage disequilibrium therewith.

In another aspect, the invention relates to a method of determining a susceptibility to Cardiovascular disease in a human individual, comprising determining whether at least one at-risk allele in at least one polymorphic marker is present in a genotype dataset derived from the individual, wherein the at least one polymorphic marker is selected from the markers set forth in Table 3, and markers in linkage disequilibrium therewith, and wherein determination of the presence of the at least one at-risk allele is indicative of increased susceptibility to Cardiovascular disease in the individual. The genotype dataset comprises in one embodiment information about marker identity, and the allelic status of the individual, i.e. information about the identity of the two alleles carried by the individual for the marker. The genotype dataset may comprise allelic information about one or more marker, including two or more markers, three or more markers, five or more markers, one hundred or more markers, etc. In some embodiments, the genotype dataset comprises genotype information from a whole-genome assessment of the individual including hundreds of thousands of markers, or even one million or more markers.

In one embodiment, the at least one polymorphic marker is present within SEQ ID NO:94, as set forth herein. In another embodiment, the at least one polymorphic marker comprises at least one marker selected from rs7041637, rs2811712, rs3218018, rs3217992, rs2069426, rs2069422, rs1333034, rs1011970, rs10116277, rs1333040, rs2383207, rs1333050, D9S1814, rs10757278, rs10757274, rs1333049, D9S1870, and markers in linkage disequilibrium therewith. In another embodiment, the at least one polymorphic marker comprises at least one marker in strong linkage disequilibrium, as defined by numeric values for |D'| of greater than 0.8

and/or r^2 of greater than 0.2, with one or more markers selected from the group consisting of the markers set forth in Table 3. In one preferred embodiment, the at least one polymorphic marker is selected from markers rs10757278, rs10757274, and rs1333049, and markers in linkage disequilibrium therewith. In another preferred embodiment, the at least one polymorphic marker is selected from markers rs10757278, rs10757274, and rs1333049. In yet another embodiment, the at least one polymorphic marker is selected from markers associated with LD Block C09 (SEQ ID NO:94). In one such embodiment, the at least one polymorphic marker is in linkage disequilibrium with at least one polymorphic marker within LD Block C09 (SEQ ID NO:94).

In one embodiment, the method of determining a susceptibility, or diagnosing a susceptibility of, Cardiovascular disease, further comprises assessing the frequency of at least one haplotype in the individual. In one such embodiment, the at least one haplotype is selected from the haplotypes that comprise at least one polymorphic marker as set forth in Table 10, and polymorphic markers in linkage disequilibrium therewith. In another embodiment, the at least one haplotype is selected from the haplotypes that comprise at least one polymorphic marker as set forth in Table 3, and polymorphic markers in linkage disequilibrium therewith. In another embodiment, the at least one haplotype is selected from the haplotypes that comprise at least one polymorphic marker selected from rs7041637, rs2811712, rs3218018, rs3217992, rs2069426, rs2069422, rs1333034, rs1011970, rs10116277, rs1333040, rs2383207, rs1333050, D9S1814, rs10757278, rs10757274, rs10333049, D9S1870, and markers in linkage disequilibrium therewith.

In certain embodiments of the invention, determination of the presence of at least one at-risk allele of at least one polymorphic marker in a nucleic acid sample from the individual is indicative of an increased susceptibility to the Cardiovascular disease. In one embodiment, the increased susceptibility is characterized by a relative risk (RR) or odds ratio (OR) of at least 1.15. In another embodiment, the increased susceptibility is characterized by a relative risk (RR) or odds ratio (OR) of at least 1.20. In another embodiment, the increased susceptibility is characterized by a relative risk (RR) or odds ratio (OR) of at least 1.30.

In some embodiments, the presence of rs7041637 allele A, rs2811712 allele A, rs3218018 allele A, rs3217992 allele A, rs2069426 allele C, rs2069422 allele A, rs1333034 allele A, rs1011970 allele G, rs10116277 allele T, rs1333040 allele T, rs2383207 allele G, rs1333050 allele T, D9S1814 allele 0, rs10757278 allele G, rs1333049 allele C, rs10757274 allele G, and/or D9S1870 allele X (composite allele of all alleles smaller than 2) is indicative of increased susceptibility of the Cardiovascular disease.

In particular embodiments, the presence of at least one protective allele in a nucleic acid sample from the individual is indicative of a decreased susceptibility of Cardiovascular disease. In another embodiment, the absence of at least one at-risk allele in a nucleic acid sample from the individual is indicative of a decreased susceptibility of Cardiovascular disease.

Another aspect of the invention relates to a method of assessing a susceptibility to Cardiovascular disease in a human individual, comprising screening a nucleic acid from the individual for at least one polymorphic marker or haplotype in the genomic segment with the sequence as set forth in SEQ ID NO:94, that correlates with increased occurrence of Cardiovascular disease in a human population, wherein the presence of an at-risk marker allele in the at least one polymorphism or an at-risk haplotype in the nucleic acid identifies the individual as having elevated susceptibility to the Cardiovascular disease, and wherein the absence of the at least one at-risk marker allele or at-risk haplotype in the nucleic acid identifies the individual as not having the elevated susceptibility.

In one such embodiment, the at least one polymorphic marker or haplotype comprises at least one polymorphic marker selected from the markers set forth in Table 10, and polymorphic markers in linkage disequilibrium therewith. In another embodiment, the at least one marker or haplotype comprises at least one polymorphic marker selected from the markers set forth in Table 3, and polymorphic markers in linkage disequilibrium therewith. In another embodiment, the at least one polymorphic marker or haplotype comprises at least one polymorphic marker selected from the markers rs7041637, rs2811712, rs3218018, rs3217992, rs2069426, rs2069422, rs1333034, rs1011970, rs10116277, rs1333040, rs2383207, rs1333050, D9S1814, rs10757278, rs10757274, rs10333049, D9S1870, and markers in linkage disequilibrium therewith. In certain embodiments, linkage disequilibrium is characterized by numeric values for $|D'|$ of greater than 0.8 and/or r^2 of greater than 0.2.

Certain embodiments of the invention further comprise a step of screening the nucleic acid for the presence of at least one at-risk genetic variant for a Cardiovascular disease not associated with LD Block C09 (SEQ ID NO:94). Such additional genetic variants can in specific embodiments include any variant that has been identified as a susceptibility or risk variant for Cardiovascular disease.

In another aspect of the present invention, the presence of the marker or haplotype found to be associated with Cardiovascular disease, and as such useful for determining a susceptibility to Cardiovascular disease, is indicative of a different response rate of the subject to a particular treatment modality for Cardiovascular disease.

In another aspect, the invention relates to a method of identification of a marker for use in assessing susceptibility to a Cardiovascular disease in human individuals, the method comprising:

identifying at least one polymorphic marker in linkage disequilibrium with at least one of the markers within LD Block C09 (SEQ ID NO:94);

determining the genotype status of a sample of individuals diagnosed with, or having a susceptibility to, Cardiovascular disease; and

determining the genotype status of a sample of control individuals;

wherein a significant difference in frequency of at least one allele in at least one polymorphism in individuals diagnosed with, or having a susceptibility to, the Cardiovascular disease, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing susceptibility to the Cardiovascular disease.

In one embodiment, "significant" is determined by statistical means, *e.g.* the difference is statistically significant. In one such embodiment, statistical significance is characterized by a P-value of less than 0.05. In other embodiments, the statistical significance is characterized a P-value of 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 0.0000000001, or less than 0.00000001.

In one embodiment, the at least one polymorphic marker is in linkage disequilibrium, as characterized by numerical values of r^2 of greater than 0.2 and/or |D'| of greater than 0.8 with at least one marker selected from markers set forth in Table 21. In another embodiment, the at least one polymorphic marker is in linkage disequilibrium, as characterized by numerical values of r^2 of greater than 0.2 and/or |D'| of greater than 0.8 with at least one marker selected from markers rs10757278, rs10757274, and rs1333049.

In one embodiment, an increase in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or having a susceptibility to, a Cardiovascular disease, as compared with the frequency of the at least one allele in the control sample, is indicative of the at least one polymorphism being useful for assessing increased susceptibility to the Cardiovascular disease. In another embodiment, a decrease in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or having a susceptibility to, a Cardiovascular disease, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing decreased susceptibility to, or protection against, the Cardiovascular disease.

Another aspect of the invention relates to a method of genotyping a nucleic acid sample obtained from a human individual, comprising determining the presence or absence of at least one allele of at least one polymorphic marker in the sample, wherein the at least one marker is selected from the markers set forth in Table 3 and Table 21, and markers in linkage disequilibrium therewith, and wherein determination of the presence or absence of the at least one allele of the at least one polymorphic marker is predictive of a susceptibility of a Cardiovascular disease.

In one embodiment, genotyping comprises amplifying a segment of a nucleic acid that comprises the at least one polymorphic marker by Polymerase Chain Reaction (PCR), using a nucleotide primer pair flanking the at least one polymorphic marker. In another embodiment, genotyping is performed using a process selected from allele-specific probe hybridization, allele-specific primer extension, allele-specific amplification, nucleic acid sequencing, 5'-exonuclease

digestion, molecular beacon assay, oligonucleotide ligation assay, size analysis, and single-stranded conformation analysis. In one particular embodiment, the process comprises allele-specific probe hybridization. In another embodiment, the process comprises DNA sequencing. In a preferred embodiment, the method comprises:

1) contacting copies of the nucleic acid with a detection oligonucleotide probe and an enhancer oligonucleotide probe under conditions for specific hybridization of the oligonucleotide probe with the nucleic acid;

wherein

a) the detection oligonucleotide probe is from 5-100 nucleotides in length and specifically hybridizes to a first segment of the nucleic acid whose nucleotide sequence is given by SEQ ID NO:94 that comprises at least one polymorphic site;

b) the detection oligonucleotide probe comprises a detectable label at its 3' terminus and a quenching moiety at its 5' terminus;

c) the enhancer oligonucleotide is from 5-100 nucleotides in length and is complementary to a second segment of the nucleotide sequence that is 5' relative to the oligonucleotide probe, such that the enhancer oligonucleotide is located 3' relative to the detection oligonucleotide probe when both oligonucleotides are hybridized to the nucleic acid; and

d) a single base gap exists between the first segment and the second segment, such that when the oligonucleotide probe and the enhancer oligonucleotide probe are both hybridized to the nucleic acid, a single base gap exists between the oligonucleotides;

2) treating the nucleic acid with an endonuclease that will cleave the detectable label from the 3' terminus of the detection probe to release free detectable label when the detection probe is hybridized to the nucleic acid; and

3) measuring free detectable label, wherein the presence of the free detectable label indicates that the detection probe specifically hybridizes to the first segment of the nucleic acid, and indicates the sequence of the polymorphic site as the complement of the detection probe.

In a particular embodiment, the copies of the nucleic acid are provided by amplification by Polymerase Chain Reaction (PCR). In another embodiment, the susceptibility determined is increased susceptibility. In another embodiment, the susceptibility determined is decreased susceptibility.

Another aspect of the invention relates to a method of assessing an individual for probability of response to a therapeutic agent for preventing and/or ameliorating symptoms associated with a Cardiovascular disease, 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, wherein the at least one polymorphic marker is selected from the markers set forth in Table 21, and markers in linkage disequilibrium therewith, wherein determination of the presence of the at least one allele of the at least one marker is indicative of a probability of a positive response to the Cardiovascular disease therapeutic agent. In one embodiment, the at least one polymorphic marker is selected from marker rs1333040, rs10116277, rs2383207 and rs10757278, and markers in linkage disequilibrium therewith. In one embodiment, the therapeutic agent is selected from beta blockers, anticoagulation agents, including heparin and/or low molecular weigth heparin, antiplatelet agents, such as clopidogrel, aspirin, beta blockers, including metoprolol and carvedilol, ACE inhibitors, Statins, Aldosterone antagonists, including eplerenone, leukotriene synthesis inhibitors, the agents set forth in Agent Table I, Agent Table II, (R)-(+)-alpha-cyclopentyl-4-(2-quinolinylmethoxy)-Benzeneacetic acid, atreleuton, and 4-{(S)-2-[4-(4-Chloro-phenoxy)-phenoxymethyl]-pyrrolidin-1-yl}-butyramide, also known as DG-051. Other embodiments may include any one or a combination of the therapeutic agents described herein to be useful for therapeutic intervention of Cardiovascular disease.

Yet another aspect of the invention relates to a method of predicting prognosis of an individual diagnosed with, a Cardiovascular disease, the method 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, wherein the at least one polymorphic marker is selected from the group consisting of rs1333040, rs10116277, rs2383207 and rs10757278, and markers in linkage disequilibrium therewith, wherein determination of the presence of the at least one allele is indicative of a worse prognosis of the Cardiovascular disease in the individual. The prognosis may in certain embodiment relate to susceptibility of recurrent MI events, recurrent stroke events, or susceptibility to other complications relating to a Cardiovascular disease.

A further aspect of the invention relates to a method of monitoring progress of a treatment of an individual undergoing treatment for a Cardiovascular disease, the method 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, wherein the at least one polymorphic marker is selected from the group consisting of rs1333040, rs10116277, rs2383207 and rs10757278, and markers in linkage disequilibrium therewith, wherein determination of the presence of the at least one allele is indicative of the treatment outcome of the individual. The treatment may in certain embodiments be surgical treatment. In other embodiments, the treatment is by administration of a therapeutic agent, optionally including lifestyle changes or alterations in environmental exposure to risk factors for cardiovascular disease, as described further herein.

In one embodiment, the method further comprises assessing at least one biomarker in a sample from the individual. The biomarker is in certain embodiments a cardiac marker or an inflammatory marker. In one embodiment, the at least one biomarker is selected from creatin kinase, troponin, glycogen phosphorylase, C-reactive protein (CRP), serum amyloid A, fibrinogen, interleukin-6, tissue necrosis factor-alpha, soluble vascular cell adhesion molecules (sVCAM), soluble intervascular adhesion molecules (sICAM), E-selectin, matrix metalloprotease type-1, matrix metalloprotease type-2, matrix metalloprotease type-3, matrix metalloprotease type-9, serum sCD40L, leukotrienes, leukotriene metabolites, interleukin-6, tissue necrosis factor-alpha, myeloperoxidase (MPO), and N-tyrosine. In one embodiment, the leukotriene is selected from LTB4, LTC4, LTD4 and LTE4. In another embodiment, the method further comprises analyzing non-genetic information to make risk assessment, diagnosis, or prognosis of the individual. The non-genetic information is in one embodiment selected from age, gender, ethnicity, socioeconomic status, previous disease diagnosis, medical history of subject, family history of Cardiovascular disease, biochemical measurements, and clinical measurements. In a particular preferred embodiment, a further step comprising calculating overall risk is employed.

Another aspect of the invention relates to analyzing a sample comprising genomic DNA from a human individual or a genotype dataset derived from a human individual for the presence or absence of at least one at-risk allele of at least one at-risk variant for cardiovascular disease not in linkage disequilibrium with any one of the markers set forth in Table 10. Thus, the variants described herein to be associated with Cardiovascular disease may be combined with other genetic variants for Cardiovascular disease, that are not genetically related (*i.e.*, not in linkage disequilibrium with) the markers described herein Such analysis may be undertaken in combination with any of the methods described herein. Furthermore any two markers herein, or any other combination of markers and/or haplotypes described herein to be associated with cardiovascular disease may be combined to assess an increased susceptibility to cardiovascular disease.

In some embodiments of the methods of the invention, non-genetic information is analyzed, to make risk assessment, diagnosis, or prognosis of the individual. The non-genetic information is in certain embodiments selected from age, gender, ethnicity, socioeconomic status, previous disease diagnosis, medical history of subject, family history of cardiovascular disease, biochemical measurements, and clinical measurements. Combined genetic factors and/or combinations of genetic and non-genetic factors may be analyzed by known methods, to generate a combined risk.

The invention also relates to a kit for assessing susceptibility to a Cardiovascular disease in a human individual, the kit comprising reagents for selectively detecting the presence or absence of at least one allele of at least one polymorphic marker in the genome of the individual, wherein the polymorphic marker is selected from the markers set forth in Tables 10,, and markers in linkage disequilibrium therewith, and wherein the presence of the at least one allele is indicative of a susceptibility to a Cardiovascular diseases.

In one embodiment, the at least one polymorphic marker is present within the genomic segment with the sequence as set forth in SEQ ID NO:94. In another embodiment, the at least one polymorphic marker is selected from the group of markers set forth in Table 21, and markers in linkage disequilibrium therewith. In another embodiment, the at least one polymorphic markers is selected from rs1333040, rs10116277, rs2383207 and rs10757278, and markers in linkage disequilibrium therewith.

In one embodiment, the reagents comprise at least one contiguous oligonucleotide that hybridizes to a fragment of the genome of the individual comprising the at least one polymorphic marker, a buffer and a detectable label. In one embodiment, the reagents comprise at least one pair of oligonucleotides that hybridize to opposite strands of a genomic nucleic acid segment obtained from the subject, wherein each oligonucleotide primer pair is designed to selectively amplify a fragment of the genome of the individual that includes one polymorphic marker, and wherein the fragment is at least 30 base pairs in size. In a particular embodiment the at least one oligonucleotide is completely complementary to the genome of the individual. In another embodiment, the at least one oligonucleotide can comprise at least one mismatch to the genome of the individual. In one embodiment, the oligonucleotide is about 18 to about 50 nucleotides in length. In another embodiment, the oligonucleotide is 20-30 nucleotides in length.

In one preferred embodiment, the kit comprises:

a detection oligonucleotide probe that is from 5-100 nucleotides in length; an enhancer oligonucleotide probe that is from 5-100 nucleotides in length; and an endonuclease enzyme;

wherein the detection oligonucleotide probe specifically hybridizes to a first segment of the nucleic acid whose nucleotide sequence is given by SEQ ID NO:94 that comprises at least one polymorphic site; and wherein the detection oligonucleotide probe comprises a detectable label at its 3' terminus and a quenching moiety at its 5' terminus; wherein the enhancer oligonucleotide is from 5-100 nucleotides in length and is complementary to a second segment of the nucleotide sequence that is 5' relative to the oligonucleotide probe, such that the enhancer oligonucleotide is located 3' relative to the detection oligonucleotide probe when both oligonucleotides are hybridized to the nucleic acid; wherein a single base gap exists between the first segment and the second segment, such that when the oligonucleotide probe and the enhancer oligonucleotide probe are both hybridized to the nucleic acid, a single base gap exists between the oligonucleotides; and wherein treating the nucleic acid with the endonuclease will cleave the detectable label from the 3' terminus of the detection probe to release free detectable label when the detection probe is hybridized to the nucleic acid.

A further aspect of the invention relates to the use of an oligonucleotide probe in the manufacture of a diagnostic reagent for diagnosing and/or assessing susceptibility to Cardiovascular disease in a human individual, wherein the probe hybridizes to a segment of a nucleic acid whose nucleotide sequence is given by SEQ ID NO: 94 that comprises at least one polymorphic site, wherein the fragment is 15-500 nucleotides in length. In one embodiment, the

polymorphic site is selected from the polymorphic markers rs1333040, rs10116277, rs2383207 and rs10757278, and markers in linkage disequilibrium therewith

Yet another aspect of the invention relates to a computer-readable medium on which is stored: an identifier for at least one polymorphic marker; an indicator of the frequency of at least one allele of said at least one polymorphic marker in a plurality of individuals diagnosed with a Cardiovascular disease; and an indicator of the frequency of the least one allele of said at least one polymorphic markers in a plurality of reference individuals; wherein the at least one polymorphic marker is selected from the polymorphic markers set forth in Table 10, and markers in linkage disequilibrium therewith. In one embodiment, the at least one polymorphic marker is selected from rs1333040, rs10116277, rs2383207 and rs10757278, and markers in linkage disequilibrium therewith.

Another aspect relates to an apparatus for determining a genetic indicator for Type 2 diabetes in a human individual, comprising: a computer readable memory; and a routine stored on the computer readable memory; wherein the routine is adapted to be executed on a processor to analyze marker and/or haplotype information for at least one human individual with respect to at least one polymorphic marker selected from the markers set forth in Table 10, and markers in linkage disequilibrium therewith, and generate an output based on the marker or haplotype information, wherein the output comprises a risk measure of the at least one marker or haplotype as a genetic indicator of a Cardiovascular disease for the human individual.

In one embodiment, the routine further comprises an indicator of the frequency of at least one allele of at least one polymorphic marker or at least one haplotype in a plurality of individuals diagnosed with a Cardiovascular disease, and an indicator of the frequency of at the least one allele of at least one polymorphic marker or at least one haplotype in a plurality of reference individuals, and wherein a risk measure is based on a comparison of the at least one marker and/or haplotype status for the human individual to the indicator of the frequency of the at least one marker and/or haplotype information for the plurality of individuals diagnosed with the Cardiovascular disease.

The present invention, as described herein, may be reduced to practice using any one, or a combination of, the polymorphic markers described herein as being useful for the determination of a susceptibility to cardiovascular disease. This includes markers that are shown herein to be associated with cardiovascular disease, but also includes markers that are in linkage disequilibrium with such variants. In one embodiment, the at least one marker is selected from the markers set forth in any of the Tables 3, 10, 21 and 26. In another embodiment, the at least one marker is selected from the markers set forth in Table 10. In another embodiment, the at least one marker is selected from the markers set forth in Table 3 and Table 21. In another embodiment, the at least one marker is selected from the markers set forth in Table 3. In another embodiment, the at least one marker is selected from the markers set forth in Table 21. In another embodiment, the at least one marker is selected from markers in linkage

disequilibrium with the CDKN2A and/or CDKN2B genes. In another embodiment, the at least one marker is selected from the markers rs10811650, rs10116277, rs1333040, rs10738607, rs4977574, rs6475608, D9S1870, rs2383207, rs1333045, rs1333046, rs10757278 and rs1333048. In another embodiment, the at least one marker is selected from the markers rs1333040, rs10116277, rs2383207 and rs10757278. In another embodiment the at least one marker is rs1333040 (SEQ ID NO:59). In another embodiment, the at least one marker is rs10116277 (SEQ ID NO:56). In another embodiment, the at least one marker is rs2383207 (SEQ ID NO:82). In another embodiment, the at least one marker is rs10757278 (SEQ ID NO:88). In some embodiments, the at least one marker is further optionally selected from markers in linkage disequilibrium with any on or a combination of more than one of the above mentioned markers.

The Cardiovascular disease in the various aspects of the invention relating to methods, uses, apparatus or kits is in some embodiments an arterial disease. In one such embodiment, the arterial disease phenotype is selected from Myocardial Infarction, Acute Coronary Syndrome (ACS), Coronary Artery Disease, Stroke, Peripheral Artery Disease, Restenosis, Intracranial Aneurysm and Aorta Abdominal Aneurysm, transluminal coronary angioplasty (PTCA), and coronary artery bypass surgery (CABG). In one embodiment, the Cardiovascular disease is Myocardial Infarction. In another embodiment, the Cardiovascular disease is Myocardial Infarction or Coronary Artery Disease. In yet another embodiment, the Cardiovascular disease is Myocardial Infarction, Coronary Artery Disease, Aorta Abdominal Aneurysm or Intracranial Aneurysm. In another embodiment, the Cardiovascular Disease is Myocardial Infarction, Coronary Artery Disease, Restenosis, Aorta Abdominal Aneurysm or Intracranial Aneurysm. In one embodiment, the Stroke phenotype is Large Artery Atherosclerotic Stroke and/or Cardiogenic Stroke. The Restenosis phenotype is in one embodiment Coronary In-stent Restenosis. In certain embodiments, the In-stent Restenosis is either Restenosis following Bare Metal Stent (BMS) placement, or it is Restenosis following placement of a Drug Eluting Stent (DES).

Variants (markers and/or haplotypes comprising polymorphic markers) in linkage disequilibrium with the markers and haplotypes of the present invention are also useful for the methods and kits of the invention. The invention therefore also pertains to markers in linkage disequilibrium with the markers and haplotypes of the invention. In certain embodiments of the methods, uses, apparatus or kits of the invention, linkage disequilibrium is characterized by specific cutoff values for a quantitative measure of linkage disequilibrium. In one such embodiment, linkage disequilibrium is characterized by specific cutoff values for r^2. In another such embodiment, linkage disequilibrium is characterized by specific cutoff values for |D'|. In yet another embodiment, linkage disequilibrium is characterized by specific cutoff values for r^2 and |D'|. In one preferred embodiment, linkage disequilibrium is characterized by values for r^2 of greater than 0.1. In another preferred embodiment, linkage disequilibrium is characterized by values for r^2 of greater than 0.2. Other cutoff values for r^2 are also possible, including, but not limited to, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99. In another preferred

embodiment, linkage disequilibrium is characterized by values for |D'| of greater than 0.5. In another preferred embodiment, linkage disequilibrium is characterized by values for |D'| of greater than 0.8. Other cutoff values for |D'| are also possible, including, but not limited to, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9, 0.95, 0.96, 0.97, 0.98 and 0.99. In certain embodiments, linkage disequilibrium is characterized by numeric cutoff values for either |D'| and r^2. In one such embodiment linkage disequilibrium is characterized by numeric cutoff values for either |D'| of greater than 0.8 and r^2 of greater than 0.2, or both.

In certain other embodiments of the methods, uses, apparatus or kits of the invention, the individual is of a specific human ancestry. In one embodiment, the ancestry is selected from black African ancestry, Caucasian ancestry and Chinese ancestry. In another embodiment, the ancestry is black African ancestry. In another embodiment, the ancestry is African American ancestry. In another embodiment, the ancestry is European ancestry. In another embodiment, the ancestry is Caucasian ancestry. The ancestry is in certain embodiment self-reported by the individual who undergoes genetic analysis or genotyping. In other embodiments, the ancestry is determined by genetic determination comprising detecting at least one allele of at least one polymorphic marker in a nucleic acid sample from the individual, wherein the presence or absence of the allele is indicative of the ancestry of the individual.

In other particular other embodiments of the methods, uses, apparatus or kits of the invention, the presence of at least one at-risk variant, i.e. an at-risk allele in at least one polymorphic marker or an at-risk haplotype, is indicative of an earlyonset of the Cardiovascular disease. Early onset is in some embodiments categorized as onset before age 75. In other embodiments, early onset is categorized as onset before age 70, before age 65, before age 60, before age 55, before age 50, before age 45, or before age 40. Other values for categorization of age at onset are also contemplated, including, but not limited to, all integer values of age, and such age categories are also within scope of the invention. In certain embodiments, the Cardiovascular disease is Myocardial Infarction, and the age at onset is below 50 for males and/or below 60 for females.

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 shows a) Association results for 127 SNPs located in a 1 Mb interval (21.6 - 22.6 Mb, Build 34) on chromosome 9. Plotted is - log *P*, where *P* is the *P*-value adjusted for relatedness of the individual against the chromosomal location of the SNPs. *b*) The corresponding pair-wise correlation r^2 between 1004 common SNPs in the same region from the HapMap release 19 for

the CEU population. *c*) Location of two recombination hot-spots based on the HapMap dataset (*Nature* 437, 1299-1320 (27 October 2005)) that define the LD-block (position 21,920,147 to 21,149,982 in NCBI Build 36; SEQ ID NO:94 that includes the strongest association results. *d*) The pair-wise correlation structure in the region measured by D' for the same set of SNPs as used in panel *b*. All four panel use the same horizontal Mb scale indicated in panel *a*.

DETAILED DESCRIPTION OF THE INVENTION

A description of preferred embodiments of the invention follows.

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, microsatellites, insertions, deletions, duplications and translocations.

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 "SNP" 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). The SNP nomenclature as reported herein refers to the official Reference SNP (rs) ID identification tag as assigned to each unique SNP by the National Center for Biotechnological Information (NCBI).

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 "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 . 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. Haplotypes are described herein in the context of the marker name and the allele of the marker in that haplotype, *e.g.*, "G rs10757278" refers to the 3 allele of marker rs7758851 being in the haplotype, and is equivalent to "rs10757278 allele G". Furthermore, allelic codes in haplotypes are as for individual markers, i.e. 1 = A, 2 = C, 3 = G and 4 = T.

The term "susceptibility", as described herein, refers to an individual (or group of individuals) being prone to developing a certain state (*e.g.*, a certain trait, phenotype or disease), or 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 and/or haplotypes of the invention as described herein may be characteristic of increased susceptibility (i.e., increased risk) of cardiovascular disease, as characterized by a relative risk (RR) or odds ratio (OR) of greater than one for the particular allele or haplotype. Alternatively, the markers and/or haplotypes of the invention are characteristic of decreased susceptibility (i.e., decreased risk) of cardiovascular disease, 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 compute-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 acess of stored information. Such transmission can be electrical, or by other available methods, such as IR links, wireless connections, etc.

A "nucleic acid sample" is a sample obtained from an individuals that contains nucleic acid. 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 as 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 "cardiovascular disease therapeutic agent", as described herein, refers to an agent that can be used to ameliorate or prevent symptoms associated with a cardiovascular disease.

The term "coronary artery disease therapeutic agent", as described herein, refers to an agent that can be used to ameliorate or prevent symptoms associated with coronary artery disease. Such agents can for example be statins, beta blockers, calcium channel blockers, cardiac glycosides, antihypertensive agents, diuretics, agents acting on the renin-angiotensin system, and aspirin.

The term "coronary stenosis" or "coronary stenosis therapeutic method", as described herein, refers to methods that can be used to ameliorate or prevent symptoms associated with coronary artery disease. Such methods can be balloon angioplasty, stenting, cutting balloon angioplasty, percutaneous transluminal coronary angioplasty (PTCA), directional coronary atherectomy, rotational coronary atherectomy, brachytherapy, drug-eluting stent (DES) insertion, metal stent insertion, or coronary artery surgeries, such as Coronary Artery Bypass Surgery (CABG).

The term "cardiovascular disease-associated nucleic acid", as described herein, refers to a nucleic acid that has been found to be associated to cardiovascular disease. This includes, but is not limited to, the markers and haplotypes described herein and markers and haplotypes in strong linkage disequilibrium (LD) therewith.

The term "LD Block C09", as described herein, refers to the Linkage Disequilibrium (LD) block on Chromosome 9 between positions 21,920,147 and 22,149,982 base pairs on Chromosome 9 of NCBI (National Center for Biotechnology Information) Build 34, Build 35 and Build 36. The nucleotide sequence of the LD Block region from these Builds is set forth in SEQ ID NO:94.

The term "cardiovascular disease", as described herein, refers to the class of diseases that involve the heart or blood vessels (arteries and veins). In one embodiment, the invention pertains to arterial disease, which relate to atherosclerotic events, which are believed to have

similar causes and mechansims. Cardiovascular diseases have certain common risk factors (age, smoking, Diabetes mellitus, hypercholesterolemia, obesity, high blood pressure, stress, depression, elevated heart rate, sleep deprivation, environmental exposure).

The abbreviations "PCTA", "CABG", "MI", "PAD", "CAD","LAA", "IA", and "AAA", as described herein, refer to the following: "PCTA" refers to Transluminal Coronary Angiopathy, "CABG" refers to Coronary Artery Bypass Surgery, "MI" refers to Myocardial Infarction, "PAD" refers to Peripheral Artery Disease, "CAD" refers to Coronoary Artery Disease, "LAA Stroke" refers to Large Artery Atherosclerotic Stroke, "IA" refers to Intrachranial Aneurysm and "AAA" refers to Abdominal Aortic Aneurysm.

The term "early onset", as described herein, refers to onset of a disease that is lower than is typically observed. In the present exemplary context, the term, as applied to the MI phenotype, is defined as a MI event before the age of 50 for males and before the age of 60 for females. The term can, in alternative embodiments of the invention, be defined in alternative manner as known to the skilled person and described in further detail herein.

Association of genetic variants to Coronary Artery Disease

Through an association study between SNP markers on a chip containing approximately 317,000 such SNPs, the present invention has identified association of certain markers on chromosome 9 with cardiovascular diseases. The original discovery was made when an analysis of SNP data from patients diagnosed with Myocardial Infarction was made, as illustrated in Table 1 and Table 12. Several markers in a region described herein as LD block C09 were found to be strongly associated with MI, with RR values as high as 1.2. Two microsatellites within the region, D9S1814 and D9S1870, were found also to be correlated with the MI phenotype. The composite allele X of D9S1870 (a composite of all alleles shorter than 2 (alleles, -6, -4, -2 and 0, respectively), was found to associate strongly with MI (Table 1). Further investigations identified close to 90 additional markers that are strongly correlated with the five markers giving strongest association to MI (rs10116277, rs1333040, rs2383207, D9S1814 and D9S1870; see Table 3). These markers could thus serve as surrogate markers for any of these five markers and therefore be used in the methods of the present invention.

The D9S1870 marker was subsequently genotyped in a very large sample of individuals (over 70,000), including additional MI cases as well as other cardiovascular diseases, plus tenths of thousands of additional population controls. A replication study in three cohorts from the US was also performed, all containing individuals of Caucasian origin. Results of these studies for the phenotype MI revealed replication of the original finding (Table 4 and 12-14), with a combined p-value of approximately 10^{-12}. The corresponding population attributable risk is about 17% for this variant.

Further studies of this variant revealed a significant correlation to age at onset of MI. Arbitrarily defining early-onset MI as MI before age 50 for males and below age 60 for females revealed an increase in this early-onset group of 1.33, compared to 1.21 for all MI cases. Multiple regression of the number of copies of the composite X allele of D9S1870 and the age of onset of MI revealed a very significant decrease in age at onset for each copy of X carried by the MI individuals (Table 8). This shows that other definitions of age at onset of MI than the cutoff of 50 for males and 60 for females could also be used for detecting this trend in age at onset with carrier status for the X allele.

The present invention has also identified association between other cardiovascular diseases and variants within the LD block C09, using the X allele of D9S1870 as surrogate marker. Thus significant association was found to Peripheral Artery Disease (PAD), even after removing individuals diagnosed with MI from the PAD cohort. We also observed increased risk of Stroke as broad phenotype, as well as the Stroke subphenotype Large Vessel disease (LVD) (Table 5 and Table 29). We have also investigated association of the at-risk variants to the related disorders peripheral artery disease (PAD) and abdominal aorta aneurysm (AAA) As can be seen in Table 29, these markers are associated with these related disorders. The association is particularly compelling for AAA, wherein significant association is observed for a large number of markers in addition to these three, as shown in Table 30. These results illustrate that the markers and haplotypes of the invention are indeed reflective of disorders related to coronary artery disease, MI and in-stent restenosis, such as abdominal aorta aneurysm.

A further analysis of individuals with diagnosis of in-stent restenosis was performed (Table 9). Significant association was detected for both mild restenosis (<50%) and severe restenosis (>50%). This indicates that the present invention can be used to indicate which individuals are at increased risk of in-stent restenosis after undergoing transluminal coronary angioplasty (PTCA).

The known genes located within the LD block C09, are called CDKN2A and CDKN2B. These genes encode three proteins that are known as ARF (also known as $p19^{ARF}$ and $p14^{ARF}$), $p15^{INK4b}$ and $p16^{INK4a}$, all of which encode tumor suppressor proteins. $p15^{INK4b}$ has its own open reading frame, but $p16^{INK4a}$ and ARF have different first exons that are spliced to a common second and third exons. Despite the sharing of exons between $p16^{INK4a}$ and ARF the proteins are encoded in different reading frames. Therefore, $p16^{INK4a}$ and ARF are proteins that do not share homology. The products of these genes have been extensively studied and are known to play a widespread role in tumor suppression. Recent data has suggested that the ARF, $p15^{INK4b}$ and $p16^{INK4a}$ locus also has a role in aging of cells, i.e. the decline of replicative potential of self-renewing cells. Several groups have shown that the expression of $p16^{INK4a}$ increases with aging in many tissues of rodents and humans. It has even been proposed that the expression of $p16^{INK4a}$ could be used as a biomarker of physiologic, as opposed to chronologic age.

Human cancers frequently have homozygous deletions of the ARF, p15 ^{INK4b} and p16^{INK4a} locus with reduced expression of all three proteins, and decreased tumor suppressor activity. Knock-out studies of mice deficient for ARF, p15 ^{INK4b} or p16^{INK4a} have revealed that these strains are more prone to cancers than wild-type mice. Furthermore, mice with overexpression of the ARF, p15 ^{INK4b} or p16^{INK4a} locus show reduction in incidence of spontaneous cancers. Since cancer is the principal cause of death in mice on this background one may argue that the tumor resistance of the mice overexpressing the ARF, p15 ^{INK4b} or p16^{INK4a} locus would also lead to longer lifespan of these mice. However, this is not the case since these mice demonstrate a normal lifespan. This may suggest that the increased ARF, p15 ^{INK4b} or p16^{INK4a} locus function and diminished tumor incidence may come at the expense of excess mortality from non-malignant causes related to aging (Cell, 127, Oct. 20, 2006), such as atherosclerotic disease.

Sequencing of the exons of CDKN2A and CDKN2B regions, including exons, exon-intron junctions and potential regulatory regions was performed using the primers as indicated in Table 12, resulting in the identification of a number of SNPs, as shown in Table 13. Three of those SNPs were not found in public databases, and the flanking sequences of those SNPs are indicated in Table 14. As it is possible that SNP markers or other polymorphisms in LD with the markers found to be associating to MI in this region of chromosome 9 show association with a higher risk, we genotyped these additional markers by sequencing, as indicated in Table 13. Several of the markers show association to MI with RR values as high as 1.7-1.8, in particular markers SG09S291 and rs2069416. These markers, and/or other markers within the CDKN2A and CDKN2B genes that are in LD with the markers of the present invention as described herein, are thus also within the scope of the invention, as those markers may represent either true disease-causing variants, or variants in strong LD with an underlying causative variant(s).

Investigation of association of the rs10757278 variant to Intrachranial Aneurysm shows significant association to this phenotype in the original Icelandic cohort, as well as replication in independent cohorts from the Netherlands and Finland (*e.g.*, Table 32). Furthermore, the original finding of association to AAA was replicated in several cohorts from Belgium, Canada, USA, Netherlands, UK and New Zealand (Table 32). These results show that the rs10757278 marker, and markers in LD therewith, are indeed significantly associated with cardiovascular disorders, including the arterial diseases than myocardial infarction, peripheral artery disease, stroke, intracranial aneurysm and abdominal aortic aneurysm.

The original discovery of the association between variants on chromosome 9p21 and cardiovascular diseases described herein (see also Helgadottir, A., *et. al., Science* **316:**1491-3 2007) has been replicated in several independent studies, including studies of subjects with CAD and controls. The association with CAD/MI in Caucasians has been replicated in 4,251 cases and 4,443 controls of the PROCARDIS Consortium (*Hum Mol Genet.* Epub 2007 Nov 29.), in an Italian population including 416 MI cases and 308 controls (*J Hum Genet.* 2008;**53(2):**144-50. Epub 2007 Dec 8.), in participants of the Framingham Heart Study (*BMC Med Genet.* 2007 Sep 19;8 Suppl 1:S5.), and in the Northwick Park Heart Study II (*Clin Chem.* Epub 2008 Feb 4).

The association of rs10757278, as well as other correlated SNPs, with CAD has also been confirmed in Asian populations from Japan and Korea with comparable odds ratios as published for Caucasians (see *Arterioscler Thromb Vasc Biol.* 2008 Feb;**28(2)**:360-5. Epub 2007 Nov 29, and *J Hum Genet*. Epub 2008 Feb 9). These studies, together with the data shown herein, clearly indicate that variants within the LD Block C09 region on Chromosome 9p21 are associated with cardiovascular disease in all populations.

Assessment for markers and haplotypes

The genomic sequence within populations is not identical when individuals are compared. Rather, the genome exhibits sequence variability between individuals at many locations in the genome. Such variations in sequence are commonly referred to as polymorphisms, and there are many such sites within each genome For example, the human genome exhibits sequence variations which occur on average every 500 base pairs. The most common sequence variant consists of base variations at a single base position in the genome, and such sequence variants, or polymorphisms, are commonly called Single Nucleotide Polymorphisms ("SNPs"). These SNPs are believed to have occurred in a single mutational event, and therefore there are usually two possible alleles possible at each SNPsite; the original allele and the mutated allele. Due to natural genetic drift and possibly also selective pressure, the original mutation has resulted in a polymorphism characterized by a particular frequency of its alleles in any given population. Many other types of sequence variants are found in the human genome, including microsatellites, insertions, deletions, inversions and copy number variations. A polymorphic microsatellite has multiple small repeats of bases (such as CA repeats, TG on the complimentary strand) at a particular site in which the number of repeat lengths varies in the general population. In general terms, each version of the sequence with respect to the polymorphic site represents a specific allele of the polymorphic site. These sequence variants can all be referred to as polymorphisms, occurring at specific polymorphic sites characteristic of the sequence variant in question. In general terms, polymorphisms can comprise any number of specific alleles. Thus in one embodiment of the invention, the polymorphism is characterized by the presence of two or more alleles in any given population. In another embodiment, the polymorphism is characterized by the presence of three or more alleles. In other embodiments, the polymorphism is characterized by four or more alleles, five or more alleles, six or more alleles, seven or more alleles, nine or more alleles, or ten or more alleles. All such polymorphisms can be utilized in the methods and kits of the present invention, and are thus within the scope of the invention.

In some instances, reference is made to different alleles at a polymorphic site without choosing a reference allele. Alternatively, a reference sequence can be referred to for a particular polymorphic site. The reference allele is sometimes referred to as the "wild-type"

allele and it usually is chosen as either the first sequenced allele or as the allele from a "non-affected" individual (*e.g.*, an individual that does not display a trait or disease phenotype).

Alleles for SNP markers as referred to herein refer to the bases A, C, G or T as they occur at the polymorphic site in the SNP assay employed. The allele codes for SNPs used herein are as follows: 1= A, 2=C, 3=G, 4=T. The person skilled in the art will however realise that by assaying or reading the opposite DNA strand, the complementary allele can in each case be measured. Thus, for a polymorphic site (polymorphic marker) containing an A/G polymorphism, the assay employed may either measure the percentage or ratio of the two bases possible, i.e. A and G. Alternatively, by designing an assay that determines the opposite strand on the DNA template, the percentage or ratio of the complementary bases T/C can be measured. Quantitatively (for example, in terms of relative risk), identical results would be obtained from measurement of either DNA strand (+ strand or – strand). Polymorphic sites (polymorphic markers) can allow for differences in sequences based on substitutions, insertions or deletions. For example, a polymorphic microsatellite has multiple small repeats of bases (such as CA repeats) at a particular site in which the number of repeat lengths varies in the general population. Each version of the sequence with respect to the polymorphic site represents a specific allele of the polymorphic site.

Typically, a reference sequence is referred to for a particular sequence. Alleles that differ from the reference are referred to as "variant" alleles. For example, the genomic DNA sequence from position 21,920,147 to position 22,149,982 base pairs on Chromosome 9 of NCBI Build 34 ("LD block C09"; SEQ ID NO:94) represents a reference sequence. A variant sequence, as used herein, refers to a sequence that differs from the reference sequence but is otherwise substantially similar. Alleles at the polymorphic genetic markers that make up the haplotypes described herein are variants. Additional variants can include changes that affect a polypeptide. Sequence differences, when compared to a reference nucleotide sequence, can include the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of a reading frame; duplication of all or a part of a sequence; transposition; or a rearrangement of a nucleotide sequence, as described in detail herein. Such sequence changes alter the polypeptide encoded by the nucleic acid. For example, if the change in the nucleic acid sequence causes a frame shift, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated polypeptide. Alternatively, a polymorphism associated with coronary artery disease and in-stent restenosis or a susceptibility to coronary artery disease and in-stent restenosis can be a synonymous change in one or more nucleotides (*i.e.*, a change that does not result in a change in the amino acid sequence). Such a polymorphism can, for example, alter splice sites, affect the stability or

transport of mRNA, or otherwise affect the transcription or translation of an encoded polypeptide. It can also alter DNA to increase the possibility that structural changes, such as amplifications or deletions, occur at the somatic level. The polypeptide encoded by the reference nucleotide sequence is the "reference" polypeptide with a particular reference amino acid sequence, and polypeptides encoded by variant alleles are referred to as "variant" polypeptides with variant amino acid sequences.

A haplotype refers to a segment of 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 . 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, each allele corresponding to a specific polymorphic marker along the segment. Haplotypes can comprise a combination of various polymorphic markers, *e.g.*, SNPs and microsatellites, having particular alleles at the polymorphic sites. The haplotypes thus comprise a combination of alleles at various genetic markers.

Detecting specific polymorphic markers and/or haplotypes can be accomplished by methods known in the art for detecting sequences at polymorphic sites. For example, standard techniques for genotyping for the presence of SNPs and/or microsatellite markers can be used, such as fluorescence-based techniques (*e.g.*, Chen, X. *et al.*, *Genome Res. 9(5)*: 492-98 (1999); Kutyavin *et al.*, *Nucleic Acid Res.* **34:**e128 (2006)), utilizing PCR, LCR, Nested PCR and other techniques for nucleic acid amplification. Specific methodologies available for SNP genotyping include, but are not limited to, TaqMan genotyping assays and SNPlex platforms (Applied Biosystems), mass spectrometry (*e.g.,* MassARRAY system from Sequenom), minisequencing methods, real-time PCR, Bio-Plex system (BioRad), CEQ and SNPstream systems (Beckman), Molecular Inversion Probe array technology (*e.g.,* Affymetrix GeneChip), and BeadArray Technologies (*e.g.,* Illumina GoldenGate and Infinium assays). By these or other methods available to the person skilled in the art, one or more alleles at polymorphic markers, including microsatellites, SNPs or other types of polymorphic markers, can be identified.

In certain methods described herein, an individual who is at an increased susceptibility (i.e., at risk) for cardiovascular disease is an individual in whom at least one specific allele at one or more polymorphic marker or haplotype conferring increased susceptibility for cardiovascular disease is identified (i.e., at-risk marker alleles or haplotypes). In one aspect, the at-risk marker or haplotype is one that confers a significant increased risk (or susceptibility) of cardiovascular disease. In one embodiment, significance associated with a marker or haplotype is measured by a relative risk. In one embodiment, significance associated with a marker or haplotype is measured by a relative risk (RR). In another embodiment, significance associated with a marker or haplotye 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 1.2, including but not limited to: at least 1.2, at least 1.3, at least 1.4, at least 1.5, at least 1.6, at least 1.7, 1.8, at least 1.9, at least 2.0, at

least 2.5, at least 3.0, at least 4.0, and at least 5.0. In a particular embodiment, a risk (relative risk and/or odds ratio) of at least 1.2 is significant. In another particular embodiment, a risk of at least 1.3 is significant. In yet another embodiment, a risk of at least 1.4 is significant. In a further embodiment, a relative risk of at least 1.5 is significant. In another further embodiment, a significant increase in risk is at least 1.7 is significant. However, 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 20%, including but not limited to about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 150%, 200%, 300%, and 500%. In one particular embodiment, a significant increase in risk is at least 20%. In other embodiments, a significant increase in risk is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% and at least 100%. 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 marker or haplotype of the present invention is one where at least one allele of at least one marker or haplotype is more frequently present in an individual at risk for cardiovascular disease (affected), compared to the frequency of its presence in a healthy individual (control), and wherein the presence of the marker or haplotype is indicative of susceptibility to cardiovascular disease. 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. Such disease-free control may in one embodiment be characterized by the absence of one or more specific disease-associated symptoms. In another embodiment, the disease-free control group is characterized by the absence of one or more disease-specific risk factors. Such risk factors are in one embodiment at least one environmental risk factor. Representative environmental factors are natural products, minerals or other chemicals which are known to affect, or contemplated to affect, the risk of developing the specific disease or trait. Other environmental risk factors are risk factors related to lifestyle, including but not limited to food and drink habits, geographical location of main habitat, and occupational risk factors. In another embodiment, the risk factors comprise at least one additional genetic risk factor.

As an example of a simple test for correlation would be a Fisher-exact test on a two by two table. Given a cohort of chromosomes, the two by two table is constructed out of the number of chromosomes that include both of the markers or haplotypes, one of the markers or haplotypes but not the other and neither of the markers or haplotypes. Other statistical tests of association known to the skilled person are also contemplated and are also within scope of the invention.

In other embodiments of the invention, an individual who is at a decreased susceptibility (i.e., at a decreased risk) for a disease or trait is an individual in whom at least one specific allele at one or more polymorphic marker or haplotype conferring decreased susceptibility for the disease or trait is identified. The marker alleles and/or haplotypes conferring decreased risk are also said to be protective. In one aspect, the protective marker or haplotype is one that confers a significant decreased risk (or susceptibility) of the disease or trait. In one embodiment, significant decreased risk is measured as a relative risk (or odds ratio) of less than 0.9, including but not limited to less than 0.9, less than 0.8, less than 0.7, less than 0.6, less than 0.5, less than 0.4, less than 0.3, less than 0.2 and less than 0.1. In one particular embodiment, significant decreased risk is less than 0.7. In another embodiment, significant decreased risk is less than 0.5. In yet another embodiment, significant decreased risk is less than 0.3. In another embodiment, the decrease in risk (or susceptibility) is at least 20%, including but not limited to at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% and at least 98%. In one particular embodiment, a significant decrease in risk is at least about 30%. In another embodiment, a significant decrease in risk is at least about 50%. In another embodiment, the decrease in risk is at least about 70%. 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 trait or disease in the population, compared with controls, the other allele of the marker will be found in decreased frequency in the group of individuals with the trait or disease, compared with controls. In such a case, one allele of the marker (the one found in increased frequency in individuals with the trait or disease) will be the at-risk allele, while the other allele will be a protective allele.

A genetic variant associated with a disease or a trait (*e.g.* cardiovascular disease) can be used alone to predict the risk of the disease 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 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.

By way of an example, let us consider a total of eight variants that have been described to associate with prostate cancer (Gudmundsson, J., *et al., Nat Genet* **39:**631-7 (2007), Gudmundsson, J., *et al., Nat Genet* **39:**977-83 (2007); Yeager, M., *et al, Nat Genet* **39:**645-49 (2007), Amundadottir, L., *el al., Nat Genet* **38:**652-8 (2006); Haiman, C.A., *et al., Nat Genet* **39:**638-44 (2007)). Seven of these loci are on autosomes, and the remaining locus is on chromosome X. The total number of theoretical genotypic combinations is then $3^7 \times 2^1 = 4374$. Some of those genotypic classes are very rare, but are still possible, and should be considered for overall risk assessment. It is likely that the multiplicative model applied in the case of multiple genetic variant will also be valid in conjugation with non-genetic risk variants assuming that the genetic variant does not clearly correlate with the "environmental" factor. In other words, genetic and non-genetic at-risk variants can be assessed under the multiplicative model to estimate combined risk, assuming that the non-genetic and genetic risk factors do not interact.

Using the same quantitative approach, the combined or overall risk associated with a plurality of variants associated with any cardiovascular disease, as described herein, may be assessed.

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.*, "alleles" of a polymorphic marker) occurs in a population at a frequency of 0.50 (50%) and another 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 allele 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). 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'|. 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.

If all polymorphisms in the genome were independent at the population level (*i.e.,* no LD), then every single one of them would need to be investigated in association studies, to assess all the different polymorphic states. However, due to linkage disequilibrium between polymorphisms, tightly linked polymorphisms are strongly correlated, which reduces the number of polymorphisms that need to be investigated in an association study to observe a significant association. Another consequence of LD is that many polymorphisms may give an association signal due to the fact that these polymorphisms are strongly correlated.

Genomic LD maps have been generated across the genome, and such LD maps have been proposed to serve as framework for mapping disease-genes (Risch, N. & Merkiangas, K, *Science* 273:1516-1517 (1996); Maniatis, N., *et al., Proc Natl Acad Sci USA* 99:2228-2233 (2002); Reich, DE *et al, Nature* 411:199-204 (2001)). It is also now established that many portions of the human genome can be broken into series of discrete haplotype blocks containing a few common haplotypes; for these blocks, linkage disequilibrium data provides little evidence indicating recombination (see, *e.g.,* Wall., J.D. and Pritchard, J.K., *Nature Reviews Genetics 4*:587-597 (2003); Daly, M. *et al., Nature Genet. 29*:229-232 (2001); Gabriel, S.B. *et al., Science 296*:2225-2229 (2002); Patil, N. *et al., Science 294*:1719-1723 (2001); Dawson, E. *et al., Nature 418*:544-548 (2002); Phillips, M.S. *et al., Nature Genet. 33*:382-387 (2003)). There are two main methods for defining these haplotype blocks: blocks can be defined as regions of DNA that have limited haplotype diversity (see, *e.g.,* Daly, M. *et al., Nature Genet. 29*:229-232 (2001); Patil, N. *et al., Science 294*:1719-1723 (2001); Dawson, E. *et al., Nature 418*:544-548 (2002); Zhang, K. *et al., Proc. Natl. Acad. Sci. USA 99*:7335-7339 (2002)), or as regions

between transition zones having extensive historical recombination, identified using linkage disequilibrium (see, *e.g.,* Gabriel, S.B. *et al., Science 296*:2225-2229 (2002); Phillips, M.S. *et al., Nature Genet. 33*:382-387 (2003); Wang, N. *et al., Am. J. Hum. Genet. 71*:1227-1234 (2002); Stumpf, M.P., and Goldstein, D.B., *Curr. Biol. 13*:1-8 (2003)). More recently, a fine-scale map of recombination rates and corresponding hotspots across the human genome has been generated (Myers, S., *et al., Science* 310:321-32324 (2005); Myers, S. *et al., Biochem Soc Trans* 34:526530 (2006)). The map reveals the enormous variation in recombination across the genome, with recombination rates as high as 10-60 cM/Mb in hotspots, while closer to 0 in intervening regions, which thus represent regions of limited haplotype diversity and high LD. The map can therefore be used to define haplotype blocks/LD blocks as regions flanked by recombination hotspots. As used herein, the terms "haplotype block" or "LD block" includes blocks defined by any of the above described characteristics, or other alternative methods used by the person skilled in the art to define such regions. 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.

It has thus become apparent that for any given observed association to a polymorphic marker in the genome, it is likely that additional markers in the genome also show association. This is a natural consequence of the uneven distribution of LD across the genome, as observed by the large variation in recombination rates. The markers used to detect association thus in a sense represent "tags" for a genomic region (i.e., a haplotype block or LD block; *e.g.,* the C09 LD block) that is associating with a given disease or trait, and as such are useful for use in the methods and kits of the present invention. One or more causative (functional) variants or mutations may reside within the region found to be associating to the disease or trait. Such variants may confer a higher relative risk (RR) or odds ratio (OR) than observed for the tagging markers used to detect the association. The present invention thus refers to the markers used for detecting association to the disease, as described herein, as well as markers in linkage disequilibrium with the markers. Thus, in certain embodiments of the invention, markers that are in LD with the markers and/or haplotypes of the invention, as described herein, may be used as surrogate markers. The surrogate markers have in one embodiment relative risk (RR) and/or odds ratio (OR) values smaller than for the markers or haplotypes initially found to be associating with the disease, as described herein. In other embodiments, the surrogate markers have RR or OR values greater than those initially determined for the markers initially found to be associating with the disease, as described herein. An example of such an embodiment would be a rare, or relatively rare (such as < 10% allelic population frequency) variant in LD with a more

common variant (> 10% population frequency) initially found to be associating with the disease, such as the variants described herein. Identifying and using such markers for detecting the association discovered by the inventors as described herein can be performed by routine methods well known to the person skilled in the art, and are therefore within the scope of the present invention.

Determination of haplotype frequency

The frequencies of haplotypes in patient and control groups can be estimated using an expectation-maximization algorithm (Dempster A. *et al.*, *J. R. Stat. Soc. B*, 39:1-38 (1977)). An implementation of this algorithm that can handle missing genotypes and uncertainty with the phase can be used. Under the null hypothesis, the patients and the controls are assumed to have identical frequencies. Using a likelihood approach, an alternative hypothesis is tested, where a candidate at-risk-haplotype, which can include the markers described herein, is allowed to have a higher frequency in patients than controls, while the ratios of the frequencies of other haplotypes are assumed to be the same in both groups. Likelihoods are maximized separately under both hypotheses and a corresponding 1-df likelihood ratio statistic is used to evaluate the statistical significance.

To look for at-risk and protective markers and haplotypes within a linkage region, for example, association of all possible combinations of genotyped markers is studied, provided those markers span a practical region. The combined patient and control groups can be randomly divided into two sets, equal in size to the original group of patients and controls. The marker and haplotype analysis is then repeated and the most significant p-value registered is determined. This randomization scheme can be repeated, for example, over 100 times to construct an empirical distribution of p-values. In a preferred embodiment, a p-value of <0.05 is indicative of an significant marker and/or haplotype association.

Haplotype Analysis

One general approach to haplotype analysis involves using likelihood-based inference applied to NEsted MOdels (Gretarsdottir S., *et al.*, *Nat. Genet. 35*:131-38 (2003)). The method is implemented in the program NEMO, which allows for many polymorphic markers, SNPs and microsatellites. The method and software are specifically designed for case-control studies where the purpose is to identify haplotype groups that confer different risks. It is also a tool for studying LD structures. In NEMO, maximum likelihood estimates, likelihood ratios and p-values are calculated directly, with the aid of the EM algorithm, for the observed data treating it as a missing-data problem.

Even though likelihood ratio tests based on likelihoods computed directly for the observed data, which have captured the information loss due to uncertainty in phase and missing genotypes, can be relied on to give valid p-values, it would still be of interest to know how much information had been lost due to the information being incomplete. The information measure for haplotype analysis is described in Nicolae and Kong (Technical Report 537, Department of Statistics, University of Statistics, University of Chicago; *Biometrics*, 60(2):368-75 (2004)) as a natural extension of information measures defined for linkage analysis, and is implemented in NEMO.

For single marker association to a disease, the Fisher exact test can be used to calculate two-sided p-values for each individual allele. Usually, all p-values are presented unadjusted for multiple comparisons unless specifically indicated. The presented frequencies (for microsatellites, SNPs and haplotypes) are allelic frequencies as opposed to carrier frequencies. To minimize any bias due the relatedness of the patients who were recruited as families for the linkage analysis, first and second-degree relatives can be eliminated from the patient list. Furthermore, the test can be repeated for association correcting for any remaining relatedness among the patients, by extending a variance adjustment procedure described in Risch, N. & Teng, J. (*Genome Res.*, 8:1273-1288 (1998)), DNA pooling (*ibid*) for sibships so that it can be applied to general familial relationships, and present both adjusted and unadjusted p-values for comparison. The differences are in general very small as expected. To assess the significance of single-marker association corrected for multiple testing we can carry out a randomization test using the same genotype data. Cohorts of patients and controls can be randomized and the association analysis redone multiple times (*e.g.*, up to 500,000 times) and the p-value is the fraction of replications that produced a p-value for some marker allele that is lower than or equal to the p-value we observed using the original patient and control cohorts.

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 3)*: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.

Linkage Disequilibrium Using NEMO

LD between pairs of markers can be calculated using the standard definition of D' and r^2 (Lewontin, R., *Genetics 49*:49-67 (1964); Hill, W.G. & Robertson, A. *Theor. Appl. Genet. 22*:226-231 (1968)). Using NEMO, frequencies of the two marker allele combinations are estimated by maximum likelihood and deviation from linkage equilibrium is evaluated by a likelihood ratio test. The definitions of D' and r^2 are extended to include microsatellites by averaging over the values for all possible allele combination of the two markers weighted by the marginal allele probabilities.

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.

As described herein, certain polymorphic markers and haplotypes comprising such markers are found to be useful for risk assessment of cardiovascular disease, *e.g.*, arterial diseases, *e.g.* myocardial infarction, coronary artery disease, restenosis, peripheral artery disease, stroke, intracranial aneurysm and abdominal aortic aneurysm. Risk assessment can involve the use of the markers for diagnosing a susceptibility to the cardiovascular disease. Particular alleles of polymorphic markers are found more frequently in individuals with cardiovascular disease, than in individuals without diagnosis of cardiovascular disease. Therefore, these marker alleles have predictive value for detecting cardiovascular disease, or a susceptibility to cardiovascular disease, in an individual. Tagging markers within haplotype blocks or LD blocks comprising at-risk markers, such as the markers of the present invention, can be used as surrogates for other markers and/or haplotypes within the haplotype block or LD

block. 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. Markers with smaller values of r^2 than 1 can also be surrogates for the at-risk variant, or alternatively represent variants with relative risk values as high as or possibly even higher than the at-risk variant. The at-risk variant identified may not be the functional variant itself, but is in this instance in linkage disequilibrium with the true functional variant. The present invention encompasses the assessment of such surrogate markers for the markers as disclosed herein. Such markers are annotated, mapped and listed in public databases, as well known to the skilled person, or can alternatively be readily identified by sequencing the region or a part of the region identified by the markers of the present invention in a group of individuals, and identify polymorphisms in the resulting group of sequences. As a consequence, the person skilled in the art can readily and without undue experimentation genotype surrogate markers in linkage disequilibrium with the markers and/or haplotypes as described herein. The tagging or surrogate markers in LD with the at-risk variants detected, also have predictive value for detecting association to the cardiovascular disease, or a susceptibility to the cardiovascular disease, in an individual. These tagging or surrogate markers that are in LD with the markers of the present invention can also include other markers that distinguish among haplotypes, as these similarly have predictive value for detecting susceptibility to cardiovascular disease.

The present invention can in certain embodiments be practiced by assessing a sample comprising genomic DNA from an individual for the presence of variants described herein to be associated with cardiovascular disease. Such assessment includes steps of detecting the presence or absence of at least one allele of at least one polymorphic marker, using methods well known to the skilled person and further described herein, and based on the outcome of such assessment, determine whether the individual from whom the sample is derived is at increased or decreased risk (increased or decreased susceptibility) of cardiovascular disease. Alternatively, the invention can be practiced utilizing a dataset comprising information about the genotype status of at least one polymorphic marker described herein to be associated with cardiovascular disease (or markers in linkage disequilibrium with at least one marker shown herein to be associated with cardiovascular disease). In other words, a dataset containing information about such genetic status, for example in the form of genotype counts at a certain polymorphic marker, or a plurality of markers (*e.g.*, an indication of the presence or absence of certain at-risk alleles), or actual genotypes for one or more markers, can be queried for the presence or absence of certain at-risk alleles at certain polymorphic markers shown by the present inventors to be associated with cardiovascular disease. A positive result for a variant (*e.g.*, marker allele) associated with cardiovascular disease, as shown herein, is indicative of the individual from which the dataset is derived is at increased susceptibility (increased risk) of at least one cardiovascular disease (*e.g.*, arterial diseases, *e.g.* myocardial infarction, coronary artery disease, restenosis, peripheral artery disease, stroke, intracranial aneurysm and abdominal aortic aneurysm).

In certain embodiments of the invention, a polymorphic marker is correlated to a cardiovascular disease by referencing genotype data for the polymorphic marker to a look-up table that comprises correlations between at least one allele of the polymorphism and the disease. In some embodiments, the table comprises a correlation for one polymorhpism. In other embodiments, the table comprises a correlation for a plurality of polymorhpisms. In both scenarios, by referencing to a look-up table that gives an indication of a correlation between a marker and cardiovascular disease, a risk for cardiovascular disesa, or a susceptibility to cardiovascular disesae, can be identified in the individual from whom the sample is derived. In some embodiments, the correlation is reported as a statistical measure. The statistical measure may be reported as a risk measure, such as a relative risk (RR), an absolute risk (AR) or an odds ratio (OR).

The markers and haplotypes of the invention, *e.g.*, the markers presented in Tables 1-36 herein, *e.g.* the markers in Table 3, 10 and 21, may be useful for risk assessment and diagnostic purposes for cardiovascular disease (*e.g.*, arterial diseases, *e.g.* myocardial infarction, coronary artery disease, restenosis, peripheral artery disease, stroke, intracranial aneurysm and abdominal aortic aneurysm), either alone or in combination. Thus, even in the cases where the increase in risk by individual markers is relatively modest, i.e. on the order of 10-30%, the association may have significant implications. Thus, relatively common variants may have significant contribution to the overall risk (Population Attributable Risk is high), or combination of markers can be used to define groups of individual who, based on the combined risk of the markers, is at signficant combined risk of developing a cardiovascular disease.

Biomarkers

The cardiovascular diseases are known to have several common biomarkers, which are belived to relate to increased risk of developing cardiovascular disease. These include elevated fibrinogen, PAI-1, homocysteine, asymmetric dimethylarginine, C-reactive protein and B-type natriuretic peptide (BNP). These common biomarkers underscore the common etiology for the cardiovascular diseases. Recently, urinary peptides have been shown to be promising biomarkers for Cardiovascular disease, in particular Coronary Artery Disease (CAD) (Zimmerli, L.U., *et al.,Mol Cell Proteomics* **7**:290-8 (2008)). These have the advantage of being non-invasive, only requiring a urine sample from the individual to be assessed. In one application, a pattern of polypeptides in the urine sample is characteristic of increased risk of CAD.

Many general inflammatory markers are predictive of risk of coronary heart disease, including CAD and MI, although these markers are not specific to atherosclerosis. For example, Stein (Stein, S., *Am J Cardiol , 87* (suppl):21A-26A (2001)) discusses the use of any one of the following serum inflammatory markers as surrogates for predicting risk of coronary heart disease including C-reactive protein (CRP), serum amyloid A, fibrinogen, interleukin-6, tissue necrosis factor-alpha, soluble vascular cell adhesion molecules (sVCAM), soluble intervascular adhesion

molecules (sICAM), E-selectin, matrix metalloprotease type-1, matrix metalloprotease type-2, matrix metalloprotease type-3, and matrix metalloprotease type-9.

A significant association between CRP levels in serum and increased risk for coronary heart disease was found in the Women's Health Study, with the highest relative risk of 4.5 seen for those women in the highest quintile of serum CRP (Ridker, P.M. *et al., New England. J. Med., 347*: 1557-1565 (2001)). A similar correlation between increased serum CRP and increased risk for coronary heart disease in women has been reported (Ridker, P.M *et al., New Engld. J. Med., 342*:836-843 (2000); Bermudez, E.A. *et .al., Arterioscler. Thromb. Vasc. Biol., 22*: 1668-1673 (2002)). A similar correlation between increased serum inflammatory markers such as CRP and increased risk for coronary heart disease has been reported for men (Doggen, C.J.M. *et al.*, J.. *Internal Med., 248*:406-414 (2000) and Ridker, P.M. *et al., New England. J. Med., 336*: 973-979 (1997)). Elevated CRP or other serum inflammatory markers is also prognostic for increased risk of a second myocardial infarct in patients with a previous myocardial infarct (Retterstol, L. *et al., Atheroscler., 160*: 433-440 (2002)). Emerging evidence also suggests that elevated CRP is an independent risk factor for adverse clinical outcomes. See, *e.g.*, Ridker *et al.*, N. Engl. J. Med. 352: 1 (January 6, 2005).

The end products of the leukotriene pathway are potent inflammatory lipid mediators derived from arachidonic acid. They can potentially contribute to development of atherosclerosis and destabilization of atherosclerotic plaques through lipid oxidation and/or proinflammatory effects, and LTC4, LTD4, and LTE4, are known to induce vasoconstriction. On the other hand, LTB4 is a strong proinflammatory agent. Increased production of these end products of the leukotriene pathway, could therefore serve as a risk factor for MI and atherosclerosis, whereas both inflammation and vasoconstriction/vasospasm have a well established role in the pathogenesis of MI and atherosclerosis.

In certain embodiments of the invention, the genetic risk variants for cardiovascular disease, such as MI, CAD, AAA, IA, stroke and/or PAD are assessed in combination with at least one biomarker. For example, levels of an inflammatory marker in an appropriate test sample (*e.g.*, serum, plasma or urine) can be measured and the determination of the biomarker level in the sample, relative to a control (either a normal, disease-free control, or a random sample from the population) is made. The result of the analysis can be analyzed in combination with genetic risk conferred by the variants described herein, to determine overall risk. Representative inflammatory markers include: C-reactive protein (CRP), serum amyloid A, fibrinogen, serum sCD40L, a leukotriene (*e.g.*, LTB4, LTC4, LTD4, LTE4), a leukotriene metabolite, interleukin-6, tissue necrosis factor-alpha, soluble vascular cell adhesion molecules (sVCAM), soluble intervascular adhesion molecules (sICAM), E-selectin, matrix metalloprotease type-1, matrix metalloprotease type-2, matrix metalloprotease type-3, matrix metalloprotease type-9, myeloperoxidase (MPO), and N-tyrosine. In a preferred embodiment, the marker is CRP, sCD40L or MPO. The determination of biomarkers can be made by standard methods known to the skilled person. For example, in one embodiment, production of a leukotriene metabolite is

stimulated in a first test sample from the individual, using a calcium ionophore. The level of production is compared with a control level. The control level is a level that is typically found in control individual(s), such as individual who are not at risk for MI, CAD, AAA, IA, stroke or PAD; alternatively, a control level is the level that is found by comparison of disease risk in a population associated with the lowest band of measurement (*e.g.*, below the mean or median, the lowest quartile or the lowest quintile) compared to higher bands of measurement (*e.g.*, above the mean or median, the second, third or fourth quartile; the second, third, fourth or fifth quintile).

As described in the above, the haplotype block structure of the human genome has the effect that a large number of variants (markers and/or haplotypes) in linkage disequilibrium with the variant originally associated with a disease or trait may be used as surrogate markers for assessing association to the disease or trait. The number of such surrogate markers will depend on factors such as the historical recombination rate in the region, the mutational frequency in the region (i.e., the number of polymorphic sites or markers in the region), and the extent of LD (size of the LD block) in the region. These markers are usually located within the physical boundaries of the LD block or haplotype block in question as defined using the methods described herein, or by other methods known to the person skilled in the art. However, sometimes marker and haplotype association is found to extend beyond the physical boundaries of the haplotype block as defined. Such markers and/or haplotypes may in those cases be also used as surrogate markers and/or haplotypes for the markers and/or haplotypes physically residing within the haplotype block as defined. As a consequence, markers and haplotypes in LD (typically characterized by r^2 greater than 0.1, such as r^2 greater than 0.2, including r^2 greater than 0.3, also including r^2 greater than 0.4) with the markers and haplotypes of the present invention are also within the scope of the invention, even if they are physically located beyond the boundaries of the haplotype block as defined. This includes markers that are described herein (*e.g.*, Tables 1-36; *e.g.*, Tables 3, 10, and 21), but may also include other markers that are in strong LD (*e.g.*, characterized by r^2 greater than 0.1 or 0.2 and/or $|D'| > 0.8$) with one or more of the markers listed in Tables 1-35, including the markers set forth in Tables 3, 10 and 21.

For the SNP markers described herein, the opposite allele to the allele found to be in excess in patients (at-risk allele) is found in decreased frequency in cardiovascular disease. These markers and haplotypes are thus protective for cardiovascular disease, i.e. they confer a decreased risk or susceptibility of individuals carrying these markers and/or haplotypes developing cardiovascular disease.

Certain variants of the present invention, including certain haplotypes comprise, in some cases, a combination of various genetic markers, *e.g.*, SNPs and microsatellites. Detecting haplotypes can be accomplished by methods known in the art and/or described herein for

detecting sequences at polymorphic sites. Furthermore, correlation between certain haplotypes or sets of markers and disease phenotype can be verified using standard techniques. A representative example of a simple test for correlation would be a Fisher-exact test on a two by two table.

In specific embodiments, a marker allele or haplotype associated with cardiovascular disease (*e.g.*, marker alleles as listed in Tables 3, 10 and 21) is one in which the marker allele or haplotype is more frequently present in an individual at risk for cardiovascular disease, (affected), compared to the frequency of its presence in a healthy individual (control), wherein the presence of the marker allele or haplotype is indicative of cardiovascular disease or a susceptibility to cardiovascular disease. In other embodiments, at-risk markers in linkage disequilibrium with one or more markers found to be associated with cardiovascular disease, including coronary artery disease and in-stent restenosis (*e.g.*, markers as listed in Tables 3, 10 and 21) are tagging markers that are more frequently present in an individual at risk for cardiovascular disease (affected), compared to the frequency of their presence in a healthy individual (control), wherein the presence of the tagging markers is indicative of increased susceptibility to cardiovascular disease. In a further embodiment, at-risk markers alleles (i.e. conferring increased susceptibility) in linkage disequilibrium with one or more markers found to be associated with cardiovascular disease (*e.g.*, marker alleles as listed in Tables 3, 10 and 21, and markers in linkage disequilibrium therewith), are markers comprising one or more allele that is more frequently present in an individual at risk for cardiovascular disease, compared to the frequency of their presence in a healthy individual (control), wherein the presence of the markers is indicative of increased susceptibility to the cardiovascular disease.

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. 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, biophysical parameters (*e.g.*, weight, BMD, blood pressure), or general health and/or lifestyle parameters (*e.g.*, history of disease or related diseases, previous diagnosis of disease, family history of 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 the disease in any of the age ranges described in the above. 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 (Stacey, S.N., *et al., Nat Genet.* May 27 2007 (Epub ahead of print; 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); 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.

The markers of the present invention found to be associated with cardiovascular disease are believed to 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, Portugues, Italian, Polish, Bulgarian, Slavic, Serbian, Bosnian, Chech, Greek and Turkish populations. The invention furthermore in other embodiments can be practiced in specific human populations that include Bantu, Mandenk, Yoruba, San, Mbuti Pygmy, Orcadian, Adygel, Russian, Sardinian, Tuscan, Mozabite, Bedouin, Druze, Palestinian, Balochi, Brahui, Makrani, Sindhi, Pathan, Burusho, Hazara, Uygur, Kalash, Han, Dai, Daur, Hezhen, Lahu, Miao, Oroqen, She, Tujia, Tu, Xibo, Yi, Mongolan, Naxi, Cambodian, Japanese, Yakut, Melanesian, Papuan, Karitianan, Surui, Colmbian, Maya and Pima.

In one preferred embodiment, the invention relates to populations that include black African ancestry such as populations comprising persons of African descent or lineage. Black African ancestry may be determined by self reporting as African-Americans, Afro-Americans, Black Americans, being a member of the black race or being a member of the negro race. For example, African Americans or Black Americans are those persons living in North America and having origins in any of the black racial groups of Africa. In another example, self-reported

persons of black African ancestry may have at least one parent of black African ancestry or at least one grandparent of black African ancestry. In another embodiment, the invention relates to individuals of Caucasian origin.

The racial contribution in individual subjects may also be determined by genetic analysis. Genetic analysis of ancestry may be carried out using unlinked microsatellite markers such as those set out in Smith *et al.* (*Am J Hum Genet* **74**, 1001-13 (2004)).

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 frequncy 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.

Utility of Genetic Testing

The person skilled in the art will appreciate and understand that the variants described herein in general do not, by themselves, provide an absolute identification of individuals who will develop a particular cardiovascular disease. The variants described herein do however indicate increased and/or decreased likelihood that individuals carrying the at-risk or protective variants of the invention will develop symptoms associated with at least one cardiovascular disease (*e.g.*, MI, CAD, IA, AAA, restenosis, stroke PAD). This information is however extremely valuable in itself, as outlined in more detail in the below, as it can be used to, for example, initiate preventive measures at an early stage, perform regular physical and/or mental exams to monitor the progress and/or appearance of symptoms, or to schedule exams at a regular interval to identify early symptoms, so as to be able to apply treatment at an early stage.

The knowledge about a genetic variant that confers a risk of developing cardiovascular disease offers the opportunity to apply a genetic test to distinguish between individuals with increased risk of developing the disease (i.e. carriers of the at-risk variant) and those with decreased risk of developing the disease (i.e. carriers of the protective variant). The core values

of genetic testing, for individuals belonging to both of the above mentioned groups, are the possibilities of being able to diagnose disease, or a predisposition to disease, at an early stage and provide information to the clinician about prognosis/aggressiveness of disease in order to be able to apply the most appropriate treatment.

Individuals with a family history of cardiovascular diseases and carriers of at-risk variants may benefit from genetic testing since the knowledge of the presence of a genetic risk factor, or evidence for increased risk of being a carrier of one or more risk factors, may provide increased incentive for implementing a healthier lifestyle (*e.g.,* lose weight, increase exercise, give um smoking, reduce stress, etc.), by avoiding or minimizing known environmental risk factors for cardiovascular diseases. Genetic testing of patients may furthermore give valuable information about the primary cause of the cardiovascular disease and can aid the clinician in selecting the best treatment options and medication for each individual.

The present invention can be thus be used for risk assessment for cardiovascular disease, including diagnosing whether an individual is at risk for developing a cardiovascular disease, such as Myocardial Infarction, Coronary Artery Disease, PAD, AAA, IA, stroke or restenosis. The polymorphic markers of the present invention can be used alone or in combination, as well as in combination with other factors, including known biomarkers, for risk assessment of an individual for cardiovascular disease. Many factors known to affect the predisposition of indiviudal towards developing risk of developing Cardiovascular disease are known to the person skilled in the art and can be utilized in such assessment. These include, but are not limited to, age, gender, smoking status, physical activiy, waist-to-hip circumference ratio, family history of Cardiovascular Disease, previously diagnosed cardiovascular disease, obesity, diagnosis of Diabetes mellitus, stress, depression, elevated heart rate, hypertriglyceridemia, low HDL cholesterol, hypertension, elevated blood pressure, cholesterol levels, HDL cholesterol, LDL cholesterol, triglycerides, apolipoprotein AI and B levels, fibrinogen, ferritin, C-reactive protein and leukotriene levels. Methods known in the art can be used for such assessment, including multivariate analyses or logistic regression, as described further herein.

METHODS

Methods for risk assessment cardiovascular disease are described herein and are encompassed by the invention. The invention also encompasses methods of assessing an individual for probability of response to a therapeutic agent for a cardiovascular disease, methods for predicting the effectiveness of a therapeutic agent for cardiovascular diseasee, nucleic acids, polypeptides and antibodies and computer-implemented functions. Kits for assaying a sample from a subject to detect susceptibility to cardiovascular disease are also encompassed by the invention.

Diagnostic methods

In certain embodiments, the present invention pertains to methods of diagnosing, or aiding in the diagnosis of, cardiovascular disease (*e.g.*, MI, CAD, IA, AAA, stroke, PAD, restenosis) or a susceptibility to cardiovascular disease, by detecting particular alleles at genetic markers that appear more frequently in subjects with at least one cardiovascular disease or subjects who are susceptible to cardiovascular disease. In a particular embodiment, the invention is a method of diagnosing a susceptibility to cardiovascular disease by detecting at least one allele of at least one polymorphic marker (*e.g.*, the markers described herein). The present invention describes methods whereby detection of particular alleles of particular markers or haplotypes is indicative of a susceptibility to cardiovascular disease. Such prognostic or predictive assays can also be used to determine prophylactic treatment of a subject prior to the onset of symptoms of the cardiovascular disease. The present invention pertains in some embodiments to methods of clinical applications of diagnosis, *e.g.*, diagnosis performed by a medical professional. In other embodiments, the invention pertains to methods of diagnosis or determination of a susceptibility performed by a layman. Recent technological advances in genotyping technologies, including high-throughput genotyping of SNP markers, such as Molecular Inversion Probe array technology (*e.g.*, Affymetrix GeneChip), and BeadArray Technologies (*e.g.*, Illumina GoldenGate and Infinium assays) have made it possible for individuals to have their own genome assessed for up to one million SNPs simultaneously, at relatively little cost. The resulting genotype information, made available to the individual can be compared to information from the public litterature about disease or trait risk associated with various SNPs. The diagnostic application of disease-associated alleles as described herein, can thus be performed either by the individual, through analysis of his/her genotype data, or by a health professional based on results of a clinical test. In other words, the diagnosis or assessment of a susceptibility based on genetic risk can be made by health professionals, genetic counselors or by the layman, based on information about his/her genotype and publications on various risk factors. In the present context, the term "diagnosing", "diagnose a susceptibility" and "determine a susceptibility" is meant to refer to any available diagnostic method, including those mentioned above.

In addition, in certain other embodiments, the present invention pertains to methods of diagnosing, or aiding in the diagnosis of, a decreased susceptibility to cardiovascular disease, by detecting particular genetic marker alleles or haplotypes that appear less frequently in patients diagnosed with cardiovascular disease than in individual not diagnosed with cardiovascular disease or in the general population.

As described and exemplified herein, particular marker alleles or haplotypes (*e.g.* the markers and haplotypes as listed in Tables 3, 10 and 21, and markers in linkage disequilibrium therewith) are associated with cardiovascular disease. In one embodiment, the marker allele or haplotype is one that confers a significant risk or susceptibility to cardiovascular disease. In another embodiment, the invention relates to a method of diagnosing a susceptibility to

cardiovascular disease in a human individual, the method 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 derived from the individual, wherein the at least one polymorphic marker is selected from the group consisting of the polymorphic markers listed in Table 3, 10 or 21, and markers in linkage disequilibrium therewith. In another embodiment, the invention pertains to methods of diagnosing a susceptibility to cardiovascular disease in a human individual, by screening for at least one marker allele or haplotype as listed in Table 21 or markers in linkage disequilibrium therewith. In another embodiment, the marker allele or haplotype is more frequently present in a subject having, or who is susceptible to, cardiovascular disease (affected), as compared to the frequency of its presence in a healthy subject (control, such as population controls). In certain embodiments, the significance of association of the at least one marker allele or haplotype is characterized by a p value < 0.05. In other embodiments, the significance of association is characterized by smaller p-values, such as < 0.01, <0.001, <0.0001, <0.00001, <0.000001, <0.0000001, <0.00000001 or <0.000000001.

In these embodiments, the presence of the at least one marker allele or haplotype is indicative of a susceptibility to cardiovascular disease (*e.g.,* MI, CAD, IA, Stroke, AAA, restenosis, PAD). These diagnostic methods involve detecting the presence or absence of at least one marker allele or haplotype that is associated with cardiovascular disease. The haplotypes described herein include combinations of alleles at various genetic markers (*e.g.,* SNPs, microsatellites). The detection of the particular genetic marker alleles that make up the particular haplotypes can be performed by a variety of methods described herein and/or known in the art. For example, genetic markers can be detected at the nucleic acid level (*e.g.,* by direct nucleotide sequencing or by other means known to the skilled in the art) or at the amino acid level if the genetic marker affects the coding sequence of a protein encoded by a cardiovascular disease, including coronary artery disease and in-stent restenosis -associated nucleic acid (*e.g.,* by protein sequencing or by immunoassays using antibodies that recognize such a protein). The marker alleles or haplotypes of the present invention correspond to fragments of a genomic DNA sequence associated with cardiovascular disease. Such fragments encompass the DNA sequence of the polymorphic marker or haplotype in question, but may also include DNA segments in strong LD (linkage disequilibrium) with the marker or haplotype (*e.g.,* as determined by particular values of r^2 and/or $|D'|$).

In one embodiment, diagnosis of a susceptibility to cardiovascular disease can be accomplished using hybridization methods, including, but not limited to, Southern analysis, Northern analysis, and/or *in situ* hybridizations (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements). 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. In one embodiment, a haplotype can be indicated by a

single nucleic acid probe that is specific for the specific haplotype (i.e., hybridizes specifically to a DNA strand comprising the specific marker alleles characteristic of the haplotype). 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.

To determine or diagnose a susceptibility to cardiovascular disease, a hybridization sample is formed by contacting the test sample containing an cardiovascular disease -associated nucleic acid, 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 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 LD Block C09 (SEQ ID NO:94), as described herein, optionally comprising at least one allele of a marker described herein, or at least one haplotype described herein, or the probe can be the complementary sequence of such a sequence. In a particular embodiment, the nucleic acid probe is a portion of the nucleotide sequence of LD Block C09 (SEQ ID NO:94), as described herein, optionally comprising at least one allele of a marker described herein, or at least one allele of one polymorphic marker or haplotype comprising at least one polymorphic marker described herein, or the probe can be the complementary sequence of such a sequence. Other suitable probes for use in the diagnostic assays of the invention are described herein. Hybridization can be performed by methods well known to the person skilled in the art (see, *e.g.*, Current Protocols in Molecular Biology, Ausubel, F. *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 then detected using standard methods. If specific hybridization occurs between the nucleic acid probe and the coronary artery disease and in-stent restenosis -associated 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. The process can be repeated for other markers of the present invention, or markers that make up a haplotype of the present invention, or multiple probes can be used concurrently to detect more than one marker alleles at a time. It is also possible to design a single probe containing more than one marker alleles of a particular haplotype (*e.g.*, a probe containing alleles complementary to 2, 3, 4, 5 or all of the markers that make up a particular haplotype). Detection of the particular markers of the haplotype in the sample is indicative that the source of the sample has the particular

haplotype (*e.g.*, a haplotype) and therefore is susceptible to cardiovascular disease (*e.g.*, MI, CAD, IA, Stroke, AAA, restenosis, PAD).

In one preferred embodiment, a method utilizing a detection oligonucleotide probe comprising a fluorescent moiety or group at its 3' terminus and a quencher at its 5' terminus, and an enhancer oligonucleotide, is employed, as described by Kutyavin *et al.* (*Nucleic Acid Res.* **34:**e128 (2006)). The fluorescent moiety can be Gig Harbor Green or Yakima Yellow, or other suitable fluorescent moieties. The detection probe is designed to hybridize to a short nucleotide sequence that includes the SNP polymorphism to be detected. Preferably, the SNP is anywhere from the terminal residue to -6 residues from the 3' end of the detection probe. The enhancer is a short oligonucleotide probe which hybridizes to the DNA template 3' relative to the detection probe. The probes are designed such that a single nucleotide gap exists between the detection probe and the enhancer nucleotide probe when both are bound to the template. The gap creates a synthetic abasic site that is recognized by an endonuclease, such as Endonuclease IV. The enzyme cleaves the dye off the fully complementary detection probe, but cannot cleave a detection probe containing a mismatch. Thus, by measuring the fluorescence of the released fluorescent moiety, assessment of the presence of a particular allele defined by nucleotide sequence of the detection probe can be performed.

The detection probe can be of any suitable size, although preferably the probe is relatively short. In one embodiment, the probe is from 5-100 nucleotides in length. In another embodiment, the probe is from 10-50 nucleotides in length, and in another embodiment, the probe is from 12-30 nucleotides in length. Other lengths of the probe are possible and within scope of the skill of the average person skilled in the art.

In a preferred embodiment, the DNA template containing the SNP polymorphism is amplified by Polymerase Chain Reaction (PCR) prior to detection. In such an embodiment, the amplified DNA serves as the template for the detection probe and the enhancer probe.

Certain embodiments of the detection probe, the enhancer probe, and/or the primers used for amplification of the template by PCR include the use of modified bases, including modified A and modified G. The use of modified bases can be useful for adjusting the melting temperature of the nucleotide molecule (probe and/or primer) to the template DNA, for example for increasing the melting temperature in regions containing a low percentage of G or C bases, in which modified A with the capability of forming three hydrogen bonds to its complementary T can be used, or for decreasing the melting temperature in regions containing a high percentage of G or C bases, for example by using modified G bases that form only two hydrogen bonds to their complementary C base in a double stranded DNA molecule. In a preferred embodiment, modified bases are used in the design of the detection nucleotide probe. Any modified base known to the skilled person can be selected in these methods, and the selection of suitable bases is well within the scope of the skilled person based on the teachings herein and known bases available from commercial sources as known to the skilled person.

In another hybridization method, Northern analysis (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, *supra*) is used to identify the presence of a polymorphism associated with cardiovascular disease. For Northern analysis, a test sample of RNA is obtained from the subject by appropriate means. As described herein, specific hybridization of a nucleic acid probe to RNA from the subject is indicative of a particular allele complementary to the probe. For representative examples of use of nucleic acid probes, see, for example, U.S. Patent Nos. 5,288,611 and 4,851,330.

Additionally, or alternatively, a peptide nucleic acid (PNA) probe can be used in addition to, or instead of, a nucleic acid probe in the hybridization methods described herein. A PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P., *et al.*, *Bioconjug. Chem. 5*:3-7 (1994)). The PNA probe can be designed to specifically hybridize to a molecule in a sample suspected of containing one or more of the marker alleles or haplotypes that are associated with cardiovascular disease. Hybridization of the PNA probe is thus diagnostic for cardiovascular disease.

In one embodiment of the invention, a test sample containing genomic DNA obtained from the subject is collected and the polymerase chain reaction (PCR) is used to amplify a fragment comprising one ore more markers or haplotypes of the present invention. As described herein, identification of a particular marker allele or haplotype associated with cardiovascular disease, can be accomplished using a variety of methods (*e.g.*, sequence analysis, analysis by restriction digestion, specific hybridization, single stranded conformation polymorphism assays (SSCP), electrophoretic analysis, etc.). In another embodiment, diagnosis is accomplished by expression analysis using quantitative PCR (kinetic thermal cycling). This technique can, for example, utilize commercially available technologies, such as TaqMan® (Applied Biosystems, Foster City, CA) . The technique can assess the presence of an alteration in the expression or composition of a polypeptide or splicing variant(s) that is encoded by a nucleic acid associated with a cardiovascular disease. Further, the expression of the variant(s) can be quantified as physically or functionally different.

In another embodiment of the methods of the invention, analysis by restriction digestion can be used to detect a particular allele if the allele results in the creation or elimination of a restriction site relative to a reference sequence. A test sample containing genomic DNA is obtained from the subject. PCR can be used to amplify particular regions that are associated with cardiovascular disease (*e.g.* the polymorphic markers and haplotypes of Table 3, 10 or 21, and markers in linkage disequilibrium therewith) nucleic acid in the test sample from the test subject. Restriction fragment length polymorphism (RFLP) analysis can be conducted, *e.g.*, as described in Current Protocols in Molecular Biology, *supra*. The digestion pattern of the relevant DNA fragment indicates the presence or absence of the particular allele in the sample.

Sequence analysis can also be used to detect specific alleles at polymorphic sites associated with cardiovascular disease, including coronary artery disease and in-stent restenosis (*e.g.* the polymorphic markers and haplotypes of Table 3, 10 or 21, and markers in linkage disequilibrium therewith). Therefore, in one embodiment, determination of the presence or absence of a particular marker alleles or haplotypes comprises sequence analysis. For example, a test sample of DNA or RNA can be obtained from the test subject. PCR or other appropriate methods can be used to amplify a portion of a nucleic acid associated with cardiovascular disease, and the presence of a specific allele can then be detected directly by sequencing the polymorphic site (or multiple polymorphic sites) of the genomic DNA in the sample.

In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from a subject, can be used to identify polymorphisms in a nucleic acid associated with cardiovascular disease (*e.g.* the polymorphic markers of Table 3, 10, and 21, and markers in linkage disequilibrium therewith). 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 (see, *e.g.*, Fodor, S. *et al.*, *Science, 251*:767-773 (1991); Pirrung *et al.*, U.S. Patent No. 5,143,854 (see also published PCT Application No. WO 90/15070); and Fodor. S. *et al.*, published PCT Application No. WO 92/10092 and U.S. Patent No. 5,424,186, the entire teachings of each of which are incorporated by reference herein). Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, e.g., U.S. Patent No. 5,384,261; the entire teachings of which are incorporated by reference herein. In another example, linear arrays can be utilized. Additional descriptions of use of oligonucleotide arrays for detection of polymorphisms can be found, for example, in U.S. Patent Nos. 5,858,659 and 5,837,832, the entire teachings of both of which are incorporated by reference herein.

Other methods of nucleic acid analysis that are available to those skilled in the art can be used to detect a particular allele at a polymorphic site associated with cardiovascular disease. Representative methods include, for example, direct manual sequencing (Church and Gilbert, *Proc. Natl. Acad. Sci. USA, 81*: 1991-1995 (1988); Sanger, F., *et al.*, *Proc. Natl. Acad. Sci. USA, 74*:5463-5467 (1977); Beavis, *et al.*, U.S. Patent No. 5,288,644); automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield, V., *et al.*, *Proc. Natl. Acad. Sci. USA, 86*:232-236 (1989)), mobility shift analysis (Orita, M., *et al.*, *Proc. Natl. Acad. Sci. USA, 86*:2766-2770 (1989)), restriction enzyme analysis (Flavell, R., *et al.*, *Cell*, 15:25-41 (1978); Geever, R., *et al.*, *Proc. Natl. Acad. Sci. USA, 78*:5081-5085 (1981)); heteroduplex analysis; chemical mismatch cleavage (CMC) (Cotton, R., *et al.*, *Proc. Natl. Acad. Sci. USA, 85*:4397-4401 (1985)); RNase protection assays (Myers, R., *et al.*, *Science, 230*:1242-

1246 (1985); use of polypeptides that recognize nucleotide mismatches, such as *E. coli* mutS protein; and allele-specific PCR.

In another embodiment of the invention, diagnosis of cardiovascular disease or a susceptibility to cardiovascular disease can be made by examining expression and/or composition of a polypeptide encoded by a nucleic acid associated with cardiovascular disease in those instances where the genetic marker(s) or haplotype(s) of the present invention result in a change in the composition or expression of the polypeptide. Thus, diagnosis of a susceptibility to cardiovascular disease can be made by examining expression and/or composition of one of these polypeptides, or another polypeptide encoded by a nucleic acid associated with cardiovascular disease, in those instances where the genetic marker or haplotype of the present invention results in a change in the composition or expression of the polypeptide. The haplotypes and markers of the present invention that show association to cardiovascular disease may play a role through their effect on one or more of these nearby genes (*e.g,* the CDKN2A and CDKN2B genes). Possible mechanisms affecting these genes include, e.g., effects on transcription, effects on RNA splicing, alterations in relative amounts of alternative splice forms of mRNA, effects on RNA stability, effects on transport from the nucleus to cytoplasm, and effects on the efficiency and accuracy of translation.

Thus, in another embodiment, the variants (markers or haplotypes) of the invention showing association to cardiovascular disease affect the expression of a nearby gene (*e.g.,* CDKN2A and/or CDKN2B). It is well known that regulatory element affecting gene expression may be located far away, even as far as tenths or hundreds of kilobases away, from the promoter region of a gene. By assaying for the presence or absence of at least one allele of at least one polymorphic marker of the present invention, it is thus possible to assess the expression level of such nearby genes. It is thus contemplated that the detection of the markers or haplotypes of the present invention can be used for assessing expression for one or more of the CDKN2A and/or CDKN2B genes.

A variety of methods can be used for detecting protein expression levels, including enzyme linked immunosorbent assays (ELISA), Western blots, immunoprecipitations and immunofluorescence. A test sample from a subject is assessed for the presence of an alteration in the expression and/or an alteration in composition of the polypeptide encoded by a nucleic acid associated with cardiovascular disease. An alteration in expression of a polypeptide encoded by a nucleic acid associated with cardiovascular disease can be, for example, an alteration in the quantitative polypeptide expression (i.e., the amount of polypeptide produced). An alteration in the composition of a polypeptide encoded by a nucleic acid associated with a cardiovascular disease is an alteration in the qualitative polypeptide expression (e.g., expression of a mutant polypeptide or of a different splicing variant). In one embodiment, diagnosis of a susceptibility to cardiovascular disease is made by detecting a particular splicing variant encoded by a nucleic acid associated with cardiovascular disease, or a particular pattern of splicing variants.

Both such alterations (quantitative and qualitative) can also be present. An "alteration" in the polypeptide expression or composition (*e.g.*, the CDKN2A and/or CDKN2B polypeptides), as used herein, refers to an alteration in expression or composition in a test sample, as compared to the expression or composition of the polypeptide in a control sample. A control sample is a sample that corresponds to the test sample (*e.g.*, is from the same type of cells), and is from a subject who is not affected by, and/or who does not have a susceptibility to, cardiovascular disease. In one embodiment, the control sample is from a subject that does not possess a marker allele or haplotype associated with cardiovascular disease, as described herein. Similarly, the presence of one or more different splicing variants in the test sample, or the presence of significantly different amounts of different splicing variants in the test sample, as compared with the control sample, can be indicative of a susceptibility to cardiovascular disease. An alteration in the expression or composition of the polypeptide in the test sample, as compared with the control sample, can be indicative of a specific allele in the instance where the allele alters a splice site relative to the reference in the control sample. Various means of examining expression or composition of a polypeptide encoded by a nucleic acid are known to the person skilled in the art and can be used, including spectroscopy, colorimetry, electrophoresis, isoelectric focusing, and immunoassays (e.g., David *et al.*, U.S. Pat. No. 4,376,110) such as immunoblotting (see, e.g., Current Protocols in Molecular Biology, particularly chapter 10, *supra*).

For example, in one embodiment, an antibody (e.g., an antibody with a detectable label) that is capable of binding to a polypeptide encoded by a nucleic acid associated with cardiovascular disease (*e.g.*, CDKN2A and/or CDKN2B polypeptides) can be used. Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g., Fv, Fab, Fab', $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a labeled secondary antibody (e.g., a fluorescently-labeled secondary antibody) and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin.

In one embodiment of this method, the level or amount of polypeptide encoded by a nucleic acid associated with cardiovascular disease in a test sample is compared with the level or amount of the polypeptide in a control sample. A level or amount of the polypeptide in the test sample that is higher or lower than the level or amount of the polypeptide in the control sample, such that the difference is statistically significant, is indicative of an alteration in the expression of the polypeptide encoded by the nucleic acid, and is diagnostic for a particular allele or haplotype responsible for causing the difference in expression. Alternatively, the composition of the polypeptide in a test sample is compared with the composition of the polypeptide in a control sample. In another embodiment, both the level or amount and the composition of the polypeptide can be assessed in the test sample and in the control sample.

In another embodiment, the diagnosis of a susceptibility to cardiovascular disease is made by detecting at least one marker or haplotypes of the present invention (e.g., associated alleles of the markers listed in Table 3, 10 and/or 21, and markers in linkage disequilibrium therewith), in combination with an additional protein-based, RNA-based or DNA-based assay.

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 associated with cardiovascular disease (*e.g.*, MI, CAD, IA, Stroke, AAA, restenosis, PAD) (*e.g.*, antibodies that bind to a polypeptide encoded by LD Block C09 (SEQ ID NO:94), or the CDKN2A and/or CDKN2B genes or fragments thereof, *e.g.*, a genomic segment comprising at least one polymorphic marker and/or haplotype of the present invention) or to a non-altered (native) polypeptide encoded by a nucleic acid associated with a cardiovascular disease, means for amplification of a nucleic acid associated with a cardiovascular disease, means for analyzing the nucleic acid sequence of a nucleic acid associated with a cardiovascular disease, means for analyzing the amino acid sequence of a polypeptide encoded by a nucleic acid associated with a cardiovascular disease, 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 as described herein.

In one embodiment, the invention pertains to a kit for assaying a sample from a subject to detect cardiovascular disease or a susceptibility to cardiovascular disease, 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 one polymorphism, wherein the polymorphism is selected from the group consisting of the polymorphisms as listed in any of the Tables 3, 10 and 21, 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 acids flanking polymorphisms (*e.g.,* SNPs or microsatellites) that are indicative of a cardiovascular disease. In another embodiment, the kit comprises one or more labeled nucleic acids capable of detecting one or more specific polymorphic markers or haplotypes associated with cardiovascular disease, 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 or haplotype to be detected by the reagents of the kit comprises one or more markers, two or more markers, three or more markers, four or more markers or five or more markers selected from the group consisting of the markers in Table 3, 10 and 21, and markers in linkage disequilbrium therewith. In another embodiment, the marker or haplotype to be detected comprises the markers listed in Table 3, 10 or 21. In another embodiment, the marker or haplotype 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 at least one of the group of markers consisting of rs10757278, rs10116277, rs1333040, D9S1814, D9S1870 and rs2383207. In yet another embodiment, the marker or haplotype to be detected comprises at least one marker selected from the group of markers consisting of rs10757278, rs7041637, rs2811712, rs3218018, rs3217992, rs2069426, rs2069422, rs1333034, rs1011970, rs10116277, rs1333040, rs2383207, rs1333050, D9S1814, and D9S1870.

In one preferred embodiment, the kit for detecting the markers of the invention comprises a detection oligonucleotide probe, that hybridizes to a segment of template DNA containing a SNP polymorphisms to be detected, an enhancer oligonucleotide probe and an endonuclease. As explained in the above, the detection oligonucleotide probe comprises a fluorescent moiety or group at its 3' terminus and a quencher at its 5' terminus, and an enhancer oligonucleotide, is employed, as described by Kutyavin *et al.* (*Nucleic Acid Res.* **34:**e128 (2006)). The fluorescent moiety can be Gig Harbor Green or Yakima Yellow, or other suitable fluorescent moieties. The detection probe is designed to hybridize to a short nucleotide sequence that includes the SNP polymorphism to be detected. Preferably, the SNP is anywhere from the terminal residue to -6 residues from the 3' end of the detection probe. The enhancer is a short oligonucleotide probe which hybridizes to the DNA template 3' relative to the detection probe. The probes are designed such that a single nucleotide gap exists between the detection probe and the enhancer nucleotide probe when both are bound to the template. The gap creates a synthetic abasic site that is recognized by an endonuclease, such as Endonuclease IV. The enzyme cleaves the dye off the fully complementary detection probe, but cannot cleave a detection probe containing a mismatch. Thus, by measuring the fluorescence of the released fluorescent moiety, assessment of the presence of a particular allele defined by nucleotide sequence of the detection probe can be performed.

The detection probe can be of any suitable size, although preferably the probe is relatively short. In one embodiment, the probe is from 5-100 nucleotides in length. In another

embodiment, the probe is from 10-50 nucleotides in length, and in another embodiment, the probe is from 12-30 nucleotides in length. Other lengths of the probe are possible and within scope of the skill of the average person skilled in the art.

In a preferred embodiment, the DNA template containing the SNP 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.

Certain embodiments of the detection probe, the enhancer probe, and/or the primers used for amplification of the template by PCR include the use of modified bases, including modified A and modified G. The use of modified bases can be useful for adjusting the melting temperature of the nucleotide molecule (probe and/or primer) to the template DNA, for example for increasing the melting temperature in regions containing a low percentage of G or C bases, in which modified A with the capability of forming three hydrogen bonds to its complementary T can be used, or for decreasing the melting temperature in regions containing a high percentage of G or C bases, for example by using modified G bases that form only two hydrogen bonds to their complementary C base in a double stranded DNA molecule. In a preferred embodiment, modified bases are used in the design of the detection nucleotide probe. Any modified base known to the skilled person can be selected in these methods, and the selection of suitable bases is well within the scope of the skilled person based on the teachings herein and known bases available from commercial sources as known to the skilled person.

In one such embodiment, the presence of the marker or haplotype is indicative of a susceptibility (increased susceptibility or decreased susceptibility) to Cardiovascular disease. In another embodiment, the presence of the marker or haplotype is indicative of response to a therapeutic agent for a Cardiovascular disease. In another embodiment, the presence of the marker or haplotype is indicative of prognosis of a Cardiovascular disease. In yet another embodiment, the presence of the marker or haplotype is indicative of progress of treatment of a cardiovascular disease. 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 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.

Therapeutic agents

Variants of the present invention (*e.g.*, the markers and/or haplotypes of the invention as described herein, *e.g.*, the markers listed in Table 3, 10 and 21) can be used to identify novel therapeutic targets for cardiovascular disease. For example, genes containing, or in linkage disequilibrium with, variants (markers and/or haplotypes) associated with cardiovascular disease, or their products (*e.g.*, the CDKN2A and CDKN2B genes and their gene products), as well as genes or their products that are directly or indirectly regulated by or interact with these genes or their products, can be targeted for the development of therapeutic agents to treat cardiovascular disease, or prevent or delay onset of symptoms associated with cardiovascular disease. Therapeutic agents may comprise one or more of, for example, small non-protein and non-nucleic acid molecules, proteins, peptides, protein fragments, nucleic acids (DNA, RNA), PNA (peptide nucleic acids), or their derivatives or mimetics which can modulate the function and/or levels of the target genes or their gene products.

The nucleic acids and/or variants of the invention, 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 described and reviewed in *AntisenseDrug Technology: Principles, Strategies, and Applications,* Crooke, ed., Marcel Dekker Inc., New York (2001). In general, antisense nucleic acid molecules are designed to be complementary to a region of mRNA expressed by a gene, so that the antisense molecule hybridizes to the mRNA, thus blocking translation of the mRNA into protein. 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 al.*, *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)

The variants described herein can 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 to treat a disease or disorder, such as a cardiovascular disease. 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.

The phenomenon of RNA interference (RNAi) has been actively studied for the last decade, since its original discovery in *C. elegans* (Fire *et al.,Nature* 391:806-11 (1998)), and in recent years its potential use in treatment of human disease has been actively pursued (reviewed in Kim & Rossi, *Nature Rev. Genet.* 8:173-204 (2007)). RNA interference (RNAi), also called gene silencing, is based on using double-stranded RNA molecules (dsRNA) to turn off specific genes. In the cell, cytoplasmic double-stranded RNA molecules (dsRNA) are processed by cellular complexes into small interfering RNA (siRNA). The siRNA guide the targeting of a protein-RNA complex to specific sites on a target mRNA, leading to cleavage of the mRNA (Thompson, *Drug Discovery Today,* 7:912-917 (2002)). The siRNA molecules are typically about 20, 21, 22 or 23 nucleotides in length. Thus, one aspect of the invention relates to isolated nucleic acid molecules, and the use of those molecules for RNA interference, i.e. as small interfering RNA molecules (siRNA). In one embodiment, the isolated nucleic acid molecules are 18-26 nucleotides in length, preferably 19-25 nucleotides in length, more preferably 20-24 nucleotides in length, and more preferably 21, 22 or 23 nucleotides in length.

Another pathway for RNAi-mediated gene silencing originates in endogenously encoded primary microRNA (pri-miRNA) transcripts, which are processed in the cell to generate precursor miRNA (pre-miRNA). These miRNA molecules are exported from the nucleus to the cytoplasm, where they undergo processing to generate mature miRNA molecules (miRNA), which direct translational inhibition by recognizing target sites in the 3' untranslated regions of mRNAs, and subsequent mRNA degradation by processing P-bodies (reviewed in Kim & Rossi, *Nature Rev. Genet.* 8:173-204 (2007)).

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 of the present invention (*e.g.*, the markers and haplotypes set forth in Tables 3, 10 and 21) 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'-O-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).

A genetic defect leading to increased predisposition or risk for development of a cardiovascular disease, or a defect causing the disease, may be corrected permanently by administering to a subject carrying the defect a nucleic acid fragment that incorporates a repair sequence that supplies the normal/wild-type nucleotide(s) at the site of the genetic defect. Such site-specific repair sequence may concompass an RNA/DNA oligonucleotide that operates to promote endogenous repair of a subject's genomic DNA. The administration of the repair sequence may be performed by an appropriate vehicle, such as a complex with polyethelenimine, encapsulated in anionic liposomes, a viral vector such as an adenovirus vector, or other pharmaceutical compositions suitable for promoting intracellular uptake of the adminstered nucleic acid. The genetic defect may then be overcome, since the chimeric oligonucleotides induce the incorporation of the normal sequence into the genome of the subject, leading to expression of the normal/wild-type gene product. The replacement is propagated, thus rendering a permanent repair and alleviation of the symptoms associated with the disease or condition.

The present invention provides methods for identifying compounds or agents that can be used to treat cardiovascular disease. Thus, the variants of the invention are useful as targets for the identification and/or development of therapeutic agents. In certain embodiments, such methods include assaying the ability of an agent or compound to modulate the activity and/or expression of a nucleic acid that includes at least one of the variants (markers and/or haplotypes) of the present invention, or the encoded product of the nucleic acid (*e.g,* one or both of the CDKN2A and CDKN2B genes). This in turn can be used to identify agents or compounds that inhibit or alter the undesired activity or expression of the encoded nucleic acid product. Assays for performing such experiments can be performed in cell-based systems or in cell-free systems, as known to the skilled person. Cell-based systems include cells naturally expressing the nucleic acid molecules of interest, or recombinant cells that have been genetically modified so as to express a certain desired nucleic acid molecule.

Variant gene expression in a patient can be assessed by expression of a variant-containing nucleic acid sequence (for example, a gene containing at least one variant of the present invention, which can be transcribed into RNA containing the at least one variant, and in turn translated into protein), or by altered expression of a normal/wild-type nucleic acid sequence due to variants affecting the level or pattern of expression of the normal transcripts, for example variants in the regulatory or control region of the gene. Assays for gene expression include direct nucleic acid assays (mRNA), assays for expressed protein levels, or assays of collateral compounds involved in a pathway, for example a signal pathway. Furthermore, the

expression of genes that are up- or down-regulated in response to the signal pathway can also be assayed. One embodiment includes operably linking a reporter gene, such as luciferase, to the regulatory region of the gene(s) of interest.

Modulators of gene expression can in one embodiment be identified when a cell is contacted with a candidate compound or agent, and the expression of mRNA is determined. The expression level of mRNA in the presence of the candidate compound or agent is compared to the expression level in the absence of the compound or agent. Based on this comparison, candidate compounds or agents for treating cardiovascular disease can be identified as those modulating the gene expression of the variant gene. When expression of mRNA or the encoded protein is statistically significantly greater in the presence of the candidate compound or agent than in its absence, then the candidate compound or agent is identified as a stimulator or up-regulator of expression of the nucleic acid. When nucleic acid expression or protein level is statistically significantly less in the presence of the candidate compound or agent than in its absence, then the candidate compound is identified as an inhibitor or down-regulator of the nucleic acid expression.

The invention further provides methods of treatment using a compound identified through drug (compound and/or agent) screening as a gene modulator (i.e. stimulator and/or inhibitor of gene expression).

Methods of assessing probability of response to therapeutic agents and methods, methods of monitoring treatment progress and methods for treating cardiovascular disease

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 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 allele of a polymorhpic marker, or the presence of a haplotype as described herein is indicative of a different response rate to a particular treatment modality for a cardiovascular disease. This means that a patient diagnosed with cardiovascular disease, or at risk for a cardiovascular disease, and carrying a certain allele at a polymorphic or haplotype of the present invention (*e.g.*, the at-risk alleles and/or haplotypes of the invention) would respond better to, or worse to, a specific therapeutic,

drug and/or other therapy used to treat the cardiovascular disease. Therefore, the presence or absence of the marker allele or haplotype could aid in deciding what treatment should be used for a the patient. For example, for a newly diagnosed patient, the presence of a marker or haplotype of the present invention may be assessed (*e.g.*, through testing DNA derived from a blood sample or other sample containing genomic DNA, as described herein). If the patient is positive for a marker allele or haplotype at (that is, at least one specific allele of the marker, or haplotype, is present), then the physician recommends one particular therapy (*e.g.*, one particular therapeutic agent or a combination of therapeutic agents), 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 disease, 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 and provide information to the clinician about prognosis/aggressiveness of the disease in order to be able to apply the most appropriate treatment.

As one example, the application of a genetic test for restenosis can identify subjects who are at high risk of developing restenosis after coronary stent procedure. While it is know that some treatment methods for coronary artery disease, such as introducing drug-eluting stents and brachytherapy, are associated with decreased risk of in-stent restenosis, the use of these methods are restricted because of number of reasons, including economical reasons. Identification of individuals within the group of those undergoing coronary stent procedure who are carriers of genetic risk variants for in-stent restenosis will allow targeting of those individuals that would benefit most from therapy associated with decreased risk of in-stent restenosis.

The present invention also relates to methods of monitoring effectiveness of a treatment for a cardiovascular disease, including coronary artery disease, MI, stroke, PAD, IA, AAA and restenosis. This can be done based on the genotype and/or haplotype status of the markers and haplotypes of the present invention, or by monitoring expression of genes that are associated with the variants (markers and haplotypes) of the present invention (*e.g.*, CDKN2A and CDKN2B). The risk gene mRNA or the encoded polypeptide can be measured in a tissue sample (*e.g.*, a peripheral blood sample, or a biopsy sample). Expression levels and/or mRNA levels can thus be determined before and during treatment to monitor its effectiveness. Alternatively, or concomitantly, the genotype and/or haplotype status of at least one risk variant for cardiovascular disease as presented herein is determined before and during treatement to monitor its effectivenes.

The treatment modules of a cardiovascular disease to which the invention pertains includes, but is not limited to, methods of treatment for myocardial infarction or susceptibility to myocardial infarction; methods of phophylaxis therapy for myocardial infarction; methods of treatment for transient ischemic attack or stroke, or susceptibility to stroke; methods of treatment for claudication, PAD or susceptibility to PAD; methods of treatment for acute

coronary syndrome (*e.g.*, unstable angina, non-ST-elevation myocardial infarction (NSTEMI) or ST-elevation myocardial infarction (STEMI)); methods for reducing risk of MI, stroke or PAD; methods for decreasing risk of a second myocardial infarction or stroke; methods of treatment for atherosclerosis, such as for patients requiring treatment (*e.g.*, angioplasty, stents, revascularization procedure) to restore blood flow in arteries (*e.g.*, coronary, carotid, and/or femoral arteries); methods of treatment for asymptomatic ankle/brachial index of less than 0.9; and/or methods for decreasing leukotriene synthesis (*e.g.*, for treatment of myocardial infarction, stroke or PAD), methods for treatment of abdominal aorta aneurysm, methods for treatment of intracranial aneurysm.

Treatment of coronary artery disease and MI may be categorized as (i) preventive treatment and (ii) disease management. The main goal of the latter is to minimize damage to the heart and prevent further complications. The first line of disease management typically includes one or more of adminstration of oxygen, aspirin, glyceryl nitrate (nitroglycerin) and analgesia, such as morphine or related drugs. Once diagnosis of MI is made, additional therapy may include beta blockers, anticoagulation agents, including heparin and/or low molecular weigth heparin, and possibly also antiplatelet agents, such as clopidogrel. Secondary prevention, i.e. the management of risk of a recurrent MI, typically includes one or more of the following: Antiplatelet drug therapy, including aspirin and/or clopidogrel, beta blocker therapy, including metoprolol and carvedilol, ACE inhibitor therapy, Statin therapy, Aldosterone antagonist therapy, including eplerenone. Further, non-therapeutic administration of food supplements such as omega-3 fatty acids may be benefitional.

New preventive therapy for cardiovascular disese, including CAD, MI and stroke, includes agents that act on the formation and/or rupture of plaques, and also includes phosphodiesterase inhibitors. Such therapeutic agents are useful in the methods of the invention, as described herein. This includes, but is not limited to, agents that target the leukotriene synthesis pathway. The leukotriene synthesis inhibitor can be any agent that inhibits or antagonizes a member of the leukotriene synthesis pathway (*e.g.*, FLAP, 5-LO, LTC4S, LTA4H, and LTB4DH). For example, the leukotriene synthesis inhibitor can be an agent that inhibits or antagonizes FLAP polypeptide activity (*e.g.*, a FLAP inhibitor) and/or FLAP nucleic acid expression (*e.g.*, a FLAP nucleic acid antagonist). In another embodiment, the leukotriene synthesis inhibitor is an agent that inhibits or antagonizes polypeptide activity and/or nucleic acid expression of another member of the leukotriene biosynthetic pathway (*e.g.*, LTC4S, LTA4H) or that increases breakdown of leukotrienes (*e.g.*, LTB4DH). In preferred embodiments, the agent alters activity and/or nucleic acid expression of FLAP, LTA4H or of 5-LO. Preferred agents include those set forth in the Agent Table I herein. In another embodiment, preferred agents can be: 1-((4-chlorophenyl)methyl)-3-((1,1-dimethylethyl)thio)-alpha,alpha-dimethyl-5-(2-quinolinylmethoxy)- 1H-Indole-2-propanoic acid otherwise known as MK-0591, (R)-(+)-alpha-cyclopentyl-4-(2-quinolinylmethoxy)-Benzeneacetic acid, otherwise known as BAY-x-1005, 3-(3-(1,1-dimethylethylthio-5-(quinoline-2-ylmethoxy)-1-(4-chloromethylphenyl)indole-2-yl)-2,2-dimethylpropionaldehyde oxime-0-2-acetic acid otherwise known as A-81834; or can be zileuton, atreleuton, 6-((3-fluoro-5-

(tetrahydro-4-methoxy-2H-pyran-4yl)phenoxy)methyl)-1-methyl-2(1H)-quinlolinone otherwise known as ZD-2138, 1-((4-chlorophenyl)methyl)-3-((1,1dimethylethyl)thio)-alpha,alpha-dimethyl-5-(2-quinolinylmethoxy)-1H-Indole-2-propanoic acid otherwise known as MK-886, 4-(3-(4-(2-Methyl-imidazol-1-yl)-phenylsulfanyl)-phenyl)-tetrahydro-pyran-4-carboxylic acid amide otherwise known as CJ-13610. Additional agents include those described in Penning *et al., Med Chem.* 2002 45(16):3482-90, Penning, *Curr Pharm Des.* 2001, 7(3):163-79 and Penning *et al., J Med Chem.* 2000 43(4):721-35.In another embodiment, the agent alters metabolism or activity of a leukotriene (*e.g.*, LTA4, LTB4, LTC4, LTD4, LTE4, Cys LT1, Cys LT2), such as leukotriene antagonists or antibodies to leukotrienes, as well as agents which alter activity of a leukotriene receptor (*e.g.*, BLT1, BLT2, CysLTR1, and CysLTR2).

In other preferred embodiments, the agent alters activity and/or nucleic acid expression of LTA4H. Preferred agents include those set forth in the Agent Table II; but also include the following agents: 1-[2-[4-(phenylmethyl)phenoxy]ethyl]-2-methyl-4-tetrazolylpieridine; 1-[2-[4-(4-oxazolyl)phenoxy)phenoxy]ethyl]pyrrolidine; 3-[methyl[3-[4-(2-thienylmethyl)phenoxy]propyl]amino]propionic acid; methyl 3-[methyl[3-[4-(2-thienylmethyl)phenoxy]propyl]amino]propionate; 3-[methyl[3-[4-(3-thienylmethyl)phenoxy]propyl]amino]propionic acid; methyl-3-[methyl[3-4-(3-theinylmethyl)phenoxy]propyl]amino]propionate; 3-[methyl[3-[4-(4-fluorophenoxy)phenoxy]propyl]amino]propionic acid; 3-[methyl[3-[4-(4-biphenyloxy)phenoxy]propyl]amino]propionic acid; N-[3-[[3-[4-(phenylmethyl)phenoxy] propyl]methylamino]propionyl]benzenesulfonamide; 1-[2-[4-(phenylmethyl)phenoxy]ethyl]-2-methyl-4-(1H-tetrazol-5-yl)piperidine; 1-[2-[4-(phenylmethyl)phenoxy]ethyl]-4-(1H-tetrazol-5-yl)piperidine. In another embodiment, preferred agents can be: ethyl-1-[2-[4-(phenylmethyl)phenoxy]ethyl]-4-piperidine-carboxylate, otherwise known as SC-56938; [4-[5-(3-Phenyl-propyl)thiophen-2-yl]butoxy]acetic acid, otherwise known as RP64966; (R)-S-[[4-(dimethylamino)phenyl]methyl]-N-(3-mercapto-2methyl-1-oxopropyl-L-cycteine, otherwise known as SA6541. In one preferred embodiment, the therapeutic agent is 4-{(S)-2-[4-(4-Chloro-phenoxy)-phenoxymethyl]-pyrrolidin-1-yl}-butyramide, also known as DG-051.

The agents for treating or preventing a cardiovascular disease can be adminstered alone, or in combination with a statin. Statins include, but are not limited to, the agents rovuvastatin, fluvastatin, atorvastatin, lovastatin (also known as mevolin), simvastatin, pravastatin, pitavastatin, mevastatin, crevastatin, ML-236A, ML-236B, MBV-530A and MB-530B.

All agents listed in the above and in Agent Table I and Agent Table II also include their optically pure enantiomers, salts, chemical derivatives, and analogues.

Agent Table I

	Product_Name (Code)	Structure	Chemical Name	Patent Ref	Date Patent Issued/Application Published	MOA
Abbott	atreleuton (ABT-761)		(R)-(+)-N-[3[5-[(4-fluorophenyl)methyl]-2thienyl]-1methyl-2-propynyl]--N-hydroxurea	US 5288751, US 5288743, US 5616596	2/22/94 04/01/97	5-LPO inhibitor
Abbott	A-81834		3-(3-(1,1-dimethylethylthio-5-(quinoline-2-ylmethoxy)-1-(4-chloromethylphenyl)indole-2-yl)-2,2-dimethylpropionaldehyde oxime-0-2-acetic acid	WO9203132, US 5459150	3/5/1992, 10/17/95	FLAP inhibitor
Abbott	A-86886		3-(3-(1,1-dimethylethylthio-5-(pyridin-2-ylmethoxy)-1-(4-chloromethylphenyl)indole-2-yl)-2,2-dimethylpropionaldehyde oxime-0-2-acetic acid	WO9203132, US 5459150	3/5/1992, 10/17/95	5-LPO inhibitor
Abbott	A-93178					FLAP inhibitor
AstraZeneca				EP 623614	09/11/94	5-LPO inhibitor

Agent Table I

AstraZeneca	ZD-2138		6-((3-fluoro-5-(tetrahydro-4-methoxy-2H-pyran-4yl)phenoxy)methyl)1-methyl-2(1H)-quinolinone (alternatively NH can be N-methyl)	EP 466452	5-LPO inhibitor
Bayer	BAY-X-1005		(R)-(+)-alpha-cyclopentyl 4-(2-quinolinylmethoxy)-Benzeneacetic acid	US 4970215 EP 344519, DE 19880531	FLAP inhibitor
Merck	MK-0591		1-((4-chlorophenyl)methyl)-3-((1,1-dimethylethyl)thio)alpha,alpha-dimethyl-5-(2-quinolinylmethoxy)-1H-Indole-2-propanoic acid	EP 419049, US 19890822	FLAP inhibitor
Merck	MK-866		(3(3-(4-chlorobenzyl)-3-t-butyl-thio-5-isopropylindol-2yl]2,2-dimethyl-proanoic acid		5-LPO inhibitor
Merck	MK-886		1-((4-chlorophenyl)methyl)-3-((1,1dimethylethyl)thio)-alpha,alpha-dimethyl-5-(2-quinolinylmethoxy)-1H-Indole-2-propanoic acid	EP 419049, US 19890822	5-LPO inhibitor
Pfizer	CJ-13610		4-(3-(4-(2-Methyl-imidazol-1-yl)-phenylsulfanyl)-phenyl)-tetrahydro-pyran-4-carboxylic acid amide		5-LPO inhibitor

Agent Table II

Target	Compound ID	Chemical Name	Patent / Reference
LTA4H Inhibitor	SC-57461A	3-[methyl[3-[4-(phenylmethyl)phenoxy]-propyl]amino]propionic acid	Penning, T.D. et.al. Bioorg Med. Chem. Letters (2003), 13, 1137-1139. ibid, (2002), 12, 3383-3386
LTA4H Inhibitor	SC-56938	Ethyl-1-[2-[4-(phenylmethyl)phenoxy]ethyl]-4-piperidine-carboxylate	Penning, T.D. et.al. Bioorg Med. Chem. Letters (2003), 13, 1137-1139. ibid, (2002), 12, 3383-3386. US 6,506,87 6A1
LTA4H Inhibitor	RP 64966	[4-[5-(3-Phenyl-propyl)thiophen-2-yl]butoxy]acetic acid	WO9627585
LTA4H Inhibitor	SA 6541	(R)-S-[[4-(dimethylamino)phenyl]methyl]-N-(3-mercapto-2methyl-1-oxopropyl-L-cycteine	WO9809943
LTA4H Inhibitor	SA-9499/ SA-6541	(R)-3-(4-Dimethylamino-benzylsulfanyl)-2-((R)-3-mercapto-2-methyl-propionylamino)-propionic acid	
LTB4 Receptor Antagonist	Amelubant/ BIIL-284	Carbamic acid,((4-((3-((4-(1-(4-hydroxyphenyl)-1-methylethyl)phenoxy)methyl)phenyl)methoxy)phenyl)iminomethyl)- ethyl ester	US 6,576,669
LTB4 Receptor Antagonist	BIRZ-227	5-Chloro-2-[3-(4-methoxy-phenyl)-2-pyridin-2-yl-pyrrolidin-1-yl]-benzooxazole	Journal of Organic Chemistry 1998,63:2(326-330).
LTB4 Receptor Antagonist	CP 195543	2-[(3S,4R)-3,4-dihydro-4-hydroxy-3-(phenylmethyl)-2H-1-benzopyran-7 -yl]-4-(trifluoromethyl)benzoic acid	Process: WO 98/11085 1998, priority US 60/26372 1996; J. Pharamacology and Expert. Therapy, 1998, 285: 946-54
LTB4 Receptor Antagonist	Ebselen	2-Phenyl-benzo[d]isoselenazol-3-one	Journal of Cerebral Blood Flow and Metabolism 1995, July 2-6 (S162); Drugs of the Future 1995, 20:10 (1057)
LTB4 Receptor Antagonist	LTB 019; CGS-25019C	4-[5-(4-Carbamimidoyl-phenoxy)-pentyloxy]-N,N-diisopropyl-3-methoxy-benzamide maleate	ACS Meeting 1994, 207th:San Diego (MEDI 003); International Congress of the Inflammation Research Association 1994, 7th:White Haven (Abs W23)
LTB4 Receptor Antagonist	LY 210073	5-(2-Carboxy-ethyl)-6-[6-(4-methoxy-phenyl)-hex-5-enyloxy]-9-oxo-9H-xanthene-2-carboxylic acid	J Med Chem 1993 36 (12) 1726-1734
LTB4 Receptor Antagonist	LY 213024	5-(3-carboxybenzoyl)-2-(decyloxy)benzenepropanoic acid	J Med Chem 1993 36 (12) 1726-1734

Target	Compound ID	Chemical Name	Patent / Reference
LTB4 Receptor Antagonist	LY 255283	1-[5-ethyl-2-hydroxy-4-[[6-methyl-6-(1H-tetrazol-5-yl)heptyl]oxy]phenyl]ethanone	EP 276064 B 1990, priority US 2479 1987
LTB4 Receptor Antagonist	LY 264086	7-carboxy-3-(decyloxy)-9-oxo-9H-xanthene-4-propanoic acid	US 4996230 1991, priority US 481413 1990
LTB4 Receptor Antagonist	LY 292728	7-carboxy-3-[3-[(5-ethyl-4'-fluoro-2-hydroxy[1,1'-biphenyl]-4-yl)oxy]propoxy]-9-oxo-9H-xanthene-4-propanoic acid disodium salt	EP 743064 A 1996, priority US 443179 1995
LTB4 Receptor Antagonist	LY-293111 (VML-295)	Benzoic acid,2-(3-(3-((5-ethyl-4'-fluoro-2-hydroxy(1,1'-biphenyl)-4-yl)oxy)propoxy)-2-propylphenoxy)-	Proceedings of the American Society for Clinical Oncology 2002, 21:1 (Abs 343) [LY-293111 for Cancer] SCRIP World Pharmaceutical News 1997, 2272 (13) [for VML-295]
LTB4 Receptor Antagonist	ONO 4057; LB 457	(E)-2-(4-carboxybutoxy)-6-[[6-(4-methoxyphenyl)-5-hexenyl]oxy]benzenepropanoic acid	EP 405116 A 1991
LTB4 Receptor Antagonist	PF 10042	1-[5-hydroxy-5-[8-(1-hydroxy-2-phenylethyl)-2-dibenzofuranyl]-1-oxopentyl]pyrrolidine	EP 422329 B 1995, priority US 409630 1989
LTB4 Receptor Antagonist	RG-14893	8-Benzyloxy-4-[(methyl-phenethyl-carbamoyl)-methyl]-naphthalene-2-carboxylic acid	SCRIP World Pharmaceutical News 1996, 2168 (20)
LTB4 Receptor Antagonist	SB-201993	3-{6-(2-Carboxy-vinyl)-5-[8-(4-methoxy-phenyl)-octyloxy]-pyridin-2-ylmethylsulfanylmethyl}-benzoic acid	WO-09500487
LTB4 Receptor Antagonist	SC-52798	7-[3-(2-Cyclopropylmethyl-3-methoxy-4-thiazol-4-yl-phenoxy)-propoxy]-8-propyl-chroman-2-carboxylic acid	Bioorganic and Medicinal Chemistry Letters 1994, 4:6 (811-816); Journal of Medicinal Chemistry 1995, 38:6 (858-868)
LTB4 Receptor Antagonist	SC-53228	3-{7-[3-(2-Cyclopropylmethyl-3-methoxy-4-methylcarbamoyl-phenoxy)-propoxy]-8-propyl-chroman-2-yl}-propionic acid	International Congress of the Inflammation Research Association 1994, 7th:White Haven (Abs W5)
LTB4 Receptor Antagonist	WAY 121006	3-fluoro-4'-(2-quinolinylmethoxy)-[1,1'-biphenyl]-4-acetic acid	Drugs under Experimental and Clinical research 1991, 17:8 (381-387)
LTB4 Receptor Antagonist	ZD-2138	3-Amino-3-(4-methoxy-tetrahydro-pyran-4-yl)-acrylic acid 1-methyl-2-oxo-1,2-dihydro-quinolin-6-ylmethyl ester	International Symposium on Medicinal Chemistry 1994, 13th:Paris (P 197)

Alternatively, biological networks or metabolic pathways related to the markers and haplotypes of the present invention can be monitored by determining mRNA and/or polypeptide levels. This can be done for example, by monitoring expression levels or polypeptides for several genes belonging to the network and/or pathway, in samples taken before and during treatment. Alternatively, metabolites belonging to the biological network or metabolic pathway can be determined before and during treatment. Effectiveness of the treatment is determined by comparing observed changes in expression levels/metabolite levels during treatment to corresponding data from healthy subjects.

In a further aspect, the markers of the present invention can be used to increase power and effectiveness of clinical trials. Thus, individuals who are carriers of the at-risk variants of the present invention may be more likely to respond to a particular treatment modality. In one embodiment, individuals who carry at-risk variants for gene(s) in a pathway and/or metabolic network for which a particular treatment (*e.g.,* small molecule drug, *e.g.* the small molecule drugs as listed in the above, *e.g.*, the drugs listed in Agent Table I and Agent Table II) is targeting, are more likely to be responders to the treatment. In another embodiment, individuals who carry at-risk variants for a gene, which expression and/or function is altered by the at-risk variant, are more likely to be responders to a treatment modality targeting that gene, its expression or its gene product.

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 (*e.g.,* change in diet, exercise, weight loss program, smoking abstinence, less stressful lifestyle, etc.) or combination of the two, 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 markers of the present invention, can be useful for selection of treatment options that target genes or gene products affected by the at-risk variants of the invention. Certain combinations of variants may be suitable for one selection of treatment options, while other gene variant combinations may target other treatment options. Such combination of variant may include one variant, two variants, three variants, or four or more variants, as needed to determine with clinically reliable accuracy the selection of treatment module.

In addition to the diagnostic and therapeutic uses of the variants of the present invention, the variants (markers and haplotypes) can also be useful markers for human identification, and as such be useful in forensics, paternity testing and in biometrics. The specific use of SNPs for forensic purposes is reviewed by Gill (*Int. J. Legal Med.* 114:204-10 (2001)). Genetic variations in genomic DNA between individuals can be used as genetic markers to identify individuals and to associate a biological sample with an individual. Genetic markers, including SNPs and microsatellites, can be useful to distinguish individuals. The more markers

that are analyzed, the lower the probability that the allelic combination of the markers in any given individual is the same as in an unrelated individual (assuming that the markers are unrelated, i.e. that the markers are in perfect linkage equilibrium). Thus, the variants used for these purposes are preferably unrelated, i.e. they are inherited independently. Thus, preferred markers can be selected from available markers, such as the markers of the present invention, and the selected markers may comprise markers from different regions in the human genome, including markers on different chromosomes.

In certain applications, the SNPs useful for forensic testing are from degenerate codon positions (i.e., the third position in certain codons such that the variation of the SNP does not affect the amino acid encoded by the codon). In other applications, such for applications for predicting phenotypic characteristics including race, ancestry or physical characteristics, it may be more useful and desirable to utilize SNPs that affect the amino acid sequence of the encoded protein. In other such embodiments, the variant (SNP or other polymorphic marker) affects the expression level of a nearby gene, thus leading to altered protein expression.

Computer-implemented aspects

The present invention also relates to computer-implemented applications of the polymorphic markers and haplotypes described herein to be associated with cardiovascular disease. Such applications can be useful for storing, manipulating or otherwise analyzing genotype data that is useful in the methods of the invention. One example pertains to storing genotype information derived from an individual on readable media, so as to be able to provide the genotype information to a third party (*e.g.*, the individual), or for deriving information from the genotype data, *e.g.*, by comparing the genotype data to information about genetic risk factors contributing to increased susceptibility to cardiovascular disease, and reporting results based on such comparison.

One such aspect relates to computer-readable media. In general terms, such medium has capabilities of storing (i) identifer information for at least one polymorphic marker or a haplotye; (ii) an indicator of the frequency of at least one allele of said at least one marker, or the frequency of a haplotype, in individuals with cardiovascular disease (*e.g.,* MI; CAD, IA, AAA, stroke, restenosis, PAD); and an indicator of the frequency of at least one allele of said at least one marker, or the frequency of a haplotype, in a reference population. The reference population can be a disease-free population of individuals. Alternatively, the reference population is a random sample from the general population, and is thus representativ of the population at large. The frequency indicator may be a calculated frequency, a count of alleles and/or haplotype copies, or normalized or otherwise manipulated values of the actual frequencies that are suitable for the particular medium.

Additional information about the individual can be stored on the medium, such as ancestry information, information about sex, physical attributes or characteristics (including height and weight), biochemical measurements (such as blood pressure, blood lipid levels, lipid levels, such as cholesterol levels), biomarkers relevant for cardiovascular disease, as described further herein, or other useful information that is desirable to store or manipulate in the context of the genotype status of a particular individual.

The invention furthermore relates to an apparatus that is suitable for determination or manipulation of genetic data useful for determining a susceptibility to cardiovascular disease in a human individual. Such an apparatus can include a computer-readable memory, a routine for manipulating data stored on the computer-readable memory, and a routine for generating an output that includes a measure of the genetic data. Such measure can include values such as allelic or haplotype frequencies, genotype counts, sex, age, phenotype information, values for odds ratio (OR) or relative risk (RR), population attributable risk (PAR), or other useful information that is either a direct statistic of the original genotype data or based on calculations based on the genetic data.

The markers and haplotypes shown herein to be associated with increased susceptibility (*e.g.*, increased risk) of cardiovascular disease, are in certain embodiments useful for interpretation and/or analysis of genotype data. Thus in certain embodiments, an identification of an at-risk allele for a cardiovascular disease, as shown herein, or an allele at a polymorphic marker in LD with any one of the markers shown herein to be associated with cardiovascular disease, is indicative of the individual from whom the genotype data originates is at increased risk of cardiovascular disease. In one such embodiment, genotype data is generated for at least one polymorphic marker shown herein to be associated with cardiovascular disease, or a marker in linkage disequilibrium therewith. The genotype data is subsequently made available to the individual from whom the data originates, for example via a user interface accessable over the internet, together with an interpretation of the genotype data, *e.g.*, in the form of a risk measure (such as an absolute risk (AR), risk ratio (RR) or odds ration (OR)) for the cardiovascular disease. In another embodiment, at-risk markers identified in a genotype dataset derived from an individual are assessed and results from the assessment of the risk conferred by the presence of such at-risk varians in the dataset are made available to the individual, for example via a secure web interface, or by other communication means. The results of such risk assessment can be reported in numeric form (*e.g.*, by risk values, such as absolute risk, relative risk, and/or an odds ratio, or by a percentage increase in risk compared with a reference), by graphical means, or by other means suitable to illustrate the risk to the individual from whom the genotype data is derived. In particular embodiments, the results of risk assessment is made available to a third party, *e.g.*, a physician, other healthcare worker or genetic counselor.

Markers useful in various aspects of the invention

The above-described applications can all be practiced with the markers and haplotypes of the invention that have in more detail been described with respect to methods of assessing susceptibility to cardiovascular disease and described in detail herein. Thus, these applications can in general be reduced to practice using any of the markers listed in Tables 1-35, and markers in linkage disequilibrium therewith. In some embodiments, the marker is selected from the markers set forth in Tables 3, 10 or 21, and markers in linkage disequilibrium therewith. In one embodiment, the markers or haplotypes are present within the genomic segment whose sequence is set forth in SEQ ID NO:94. In another embodiment, the markers and haplotypes comprise at least one marker selected from rs7041637, rs2811712, rs3218018, rs3217992, rs2069426, rs2069422, rs1333034, rs1011970, rs10116277, rs1333040, rs2383207, rs1333050, D9S1814, rs10757278, rs10757274, rs10333049, D9S1870, optionally including markers in linkage disequilibrium therewith. In one specific embodiment, linkage disequilibrium is defined by numerical values for r^2 of greater than 0.2. In another embodiment, the marker or haplotype comprises at least one marker selected from rs7041637 allele A, rs2811712 allele A, rs3218018 allele A, rs3217992 allele A, rs2069426 allele C, rs2069422 allele A, rs1333034 allele A, rs1011970 allele G, rs10116277 allele T, rs1333040 allele T, rs2383207 allele G, rs1333050 allele T, D9S1814 allele 0, rs10757278 allele G, rs1333049 allele C, rs10757274 allele G, and/or D9S1870 allele X (composite allele of all alleles smaller than 2), wherein the indicated allele is indicative of increased susceptibility of the Cardiovascular disease.

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 haplotype described herein). In one embodiment, the invention includes variants that hybridize under high stringency hybridization and wash conditions (*e.g.,* for selective hybridization) to a nucleotide sequence that comprises the nucleotide sequence of LD Block C09 (SEQ ID NO:94) or a fragment thereof (or a nucleotide sequence comprising the complement of the nucleotide sequence of LD Block C09 as set forth in SEQ ID NO:94), wherein the nucleotide sequence comprises at least one at-risk allele of at least one polymorphic marker, or at least one haplotype, as described 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, wordlength=12, or can be varied (*e.g.*, W=5 or W=20). Other examples include the algorithm of Myers and Miller, CABIOS (1989), ADVANCE and ADAM as described in Torellis, A. and Robotti, C., *Comput. Appl. Biosci. 10*:3-5 (1994); and FASTA described in Pearson, W. and Lipman, D., *Proc. Natl. Acad. Sci. USA, 85*:2444-48 (1988). In another embodiment, the percent identity between two amino acid sequences can be accomplished using the GAP program in the GCG software package (Accelrys, Cambridge, UK).

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 LD Block C09 (SEQ ID NO:94), or a nucleotide sequence comprising, or consisting of, the complement of the nucleotide sequence of LD Block C09 (SEQ ID NO:94) , wherein the nucleotide sequence comprises 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.

The nucleic acid fragments of the invention are 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 (PNA), 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 comprising a contiguous nucleotide sequence from LD Block C09 and comprising 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 and the sequence information provided by the nucleotide sequence of LD Block C09 (SEQ ID NO:94). See generally *PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich,

Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (Eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila, P. *et al.*, *Nucleic Acids Res.*, 19:4967-4973 (1991); Eckert, K. and Kunkel, T., *PCR Methods and Applications*, 1:17-24 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202, the entire teachings of each of which are incorporated herein by reference.

In general, the isolated nucleic acid sequences of the invention can be used as molecular weight markers on Southern gels, and as chromosome markers that are labeled to map related gene positions. The nucleic acid sequences can also be used to compare with endogenous DNA sequences in patients to identify a susceptibility to a cardiovascular disease, and as probes, such as to hybridize and discover related DNA sequences or to subtract out known sequences from a sample (*e.g.*, subtractive hybridization). The nucleic acid sequences can further be used to derive primers for genetic fingerprinting, to raise anti-polypeptide antibodies using immunization techniques, and/or as an antigen to raise anti-DNA antibodies or elicit immune responses.

Two polypeptides, as described herein (or a region of the polypeptides) are substantially homologous or identical when the amino acid sequences are at least about 45-55%. In other embodiments, two polypeptides (or a region of the polypeptides) are substantially homologous or identical when they are at least about 70-75%, at least about 80-85%, at least about 90%, at least about 95% homologous or identical, or are identical. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid molecule comprising the nucleotide sequence of LD Block C09 (SEQ ID NO:94) or a portion thereof, and further comprising at least one polymorphism as shown in Table 3, 10 or 21, wherein the encoding nucleic acid will hybridize to the nucleotide sequence of LD Block C09 (SEQ ID NO:94), under stringent conditions as more particularly described herein. In on embodiment, the polypeptide comprises all or a portion of the amino acid sequence of CDKN2A and/or CDKN2B.

Antibodies

Polyclonal antibodies and/or monoclonal antibodies that specifically bind one form of the gene product but not to the other form of the gene product are also provided. Antibodies are also provided which bind a portion of either the variant or the reference gene product that contains the polymorphic site or sites. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain antigen-binding sites that specifically bind an antigen. A molecule that specifically binds to a polypeptide of the invention is a molecule that binds to that polypeptide or a fragment thereof, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind to a polypeptide of the invention. The term

"monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a polypeptide of the invention. A monoclonal antibody composition thus typically displays a single binding affinity for a particular polypeptide of the invention with which it immunoreacts.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a desired immunogen, *e.g.*, polypeptide of the invention or afragment thereof. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules directed against the polypeptide can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, *Nature* 256:495-497 (1975), the human B cell hybridoma technique (Kozbor *et al.*, *Immunol. Today* 4: 72 (1983)), the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss,1985, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (see generally *Current Protocols in Immunology* (1994) Coligan *et al.*, (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds a polypeptide of the invention.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating a monoclonal antibody to a polypeptide of the invention (see, *e.g.*, *Current Protocols in Immunology, supra*; Galfre *et al.*, *Nature* 266:55052 (1977); R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner, *Yale J. Biol. Med.* 54:387-402 (1981)). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods that also would be useful.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody to a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide to thereby isolate immunoglobulin library members that bind the polypeptide. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *Surf*ZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT

Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.*, *Bio/Technology* 9: 1370-1372 (1991); Hay *et al.*, *Hum. Antibod. Hybridomas* 3:81-85 (1992); Huse *et al.*, *Science* 246: 1275-1281 (1989); and Griffiths *et al.*, *EMBO J.* 12:725-734 (1993).

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art.

In general, antibodies of the invention (*e.g.*, a monoclonal antibody) can be used to isolate a polypeptide of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. A polypeptide-specific antibody can facilitate the purification of natural polypeptide from cells and of recombinantly produced polypeptide expressed in host cells. Moreover, an antibody specific for a polypeptide of the invention can be used to detect the polypeptide (*e.g.*, in a cellular lysate, cell supernatant, or tissue sample) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. The antibody can be coupled to a detectable substance to facilitate its detection. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibodies may also be useful in pharmacogenomic analysis. In such embodiments, antibodies against variant proteins encoded by nucleic acids according to the invention, such as variant proteins that are encoded by nucleic acids that contain at least one polymorpic marker of the invention, can be used to identify individuals that require modified treatment modalities.

Antibodies can furthermore be useful for assessing expression of variant proteins in disease states, such as in active stages of a cardiovascular disease, or in an individual with a predisposition to a disease related to the function of the protein, in particular a cardiovascular disease. Examples are provided by biomarker (*e.g.*, cardiac markers), as described further herein.. Antibodies specific for a variant protein of the present invention that is encoded by a

nucleic acid that comprises at least one polymorphic marker or haplotype as described herein (*e.g.*, CDKN2A and/or CDKN2B)can be used to screen for the presence of the variant protein, for example to screen for a predisposition to cardiovascular disease as indicated by the presence of the variant protein.

Antibodies can be used in other methods. Thus, antibodies are useful as diagnostic tools for evaluating proteins, such as variant proteins of the invention, in conjunction with analysis by electrophoretic mobility, isoelectric point, tryptic or other protease digest, or for use in other physical assays known to those skilled in the art. Antibodies may also be used in tissue typing. In one such embodiment, a specific variant protein has been correlated with expression in a specific tissue type, and antibodies specific for the variant protein can then be used to identify the specific tissue type.

Subcellular localization of proteins, including variant proteins, can also be determined using antibodies, and can be applied to assess aberrant subcellular localization of the protein in cells in various tissues. Such use can be applied in genetic testing, but also in monitoring a particular treatment modality. In the case where treatment is aimed at correcting the expression level or presence of the variant protein or aberrant tissue distribution or developmental expression of the variant protein, antibodies specific for the variant protein or fragments thereof can be used to monitor therapeutic efficacy.

Antibodies are further useful for inhibiting variant protein function (*e.g.*, CDKN2A and/or CDKN2B), for example by blocking the binding of a variant protein to a binding molecule or partner. Such uses can also be applied in a therapeutic context in which treatment involves inhibiting a variant protein's function. An antibody can be for example be used to block or competitively inhibit binding, thereby modulating (i.e., agonizing or antagonizing) the activity of the protein. Antibodies can be prepared against specific protein fragments containing sites required for specific function or against an intact protein that is associated with a cell or cell membrane. For administration *in vivo*, an antibody may be linked with an additional therapeutic payload, such as radionuclide, an enzyme, an immunogenic epitope, or a cytotoxic agent, including bacterial toxins (diphtheria or plant toxins, such as ricin). The *in vivo* half-life of an antibody or a fragment thereof may be increased by pegylation through conjugation to polyethylene glycol.

The present invention further relates to kits for using antibodies in the methods described herein. This includes, but is not limited to, kits for detecting the presence of a variant protein in a test sample. One preferred embodiment comprises antibodies such as a labelled or labelable antibody and a compound or agent for detecting variant proteins in a biological sample, means for determining the amount or the presence and/or absence of variant protein in the sample, and means for comparing the amount of variant protein in the sample with a standard, as well as instructions for use of the kit.

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

EXEMPLIFICATION

The following contains description of the identifiction of susceptibility factors found to be associated with coronary artery disease and in-stent restenosis through single-point analysis of SNP markers and microsatellite markers.

METHODS

The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee.

Icelandic coronary artery disease and in-stent restenosis cohort

The association between markers in LD block C09 to coronary artery disease was originally discovered as an association between the markers and myocardial infarction, which is the most feared complication of coronary artery disease (subphenotype of coronary artery disease).

Over the last eight years individuals who have suffered an MI we have been recruited through cardiovascular disease (CVD) genetic programs at deCODE. Currently blood samples have been collected from 2525 MI patients. The individuals who had suffered an MI were identified from a registry of over 10,000 individuals who: a) had an MI before the age of 75 in Iceland in the years 1981 to 2002 and satisfy the MONICA criteria (*J Clin Epidemiol* **41**, 105-14 (1988)); b) participated in a large prospective epidemiology study (*1*) done by the Icelandic Heart Association (IHA) over the past 30 years and had MI prior to 1981; c) had MI discharge diagnosis from the major hospitals in Reykjavik in the years 2003 and 2004. MI diagnoses of all individuals in the registry follow strict diagnostic rules based on signs, symptoms, electrocardiograms, cardiac enzymes and necropsy findings (*2*). The patients were contacted through collaborating physicians in the CVD genetic programs at deCODE. Most of the participants in the study visited the IHA and had their blood drawn, although participants who lived outside the Reykjavik area visited their local health care center.

Additional subjects with coronary artery disease, but are without known history of myocardial infarction, are identified from a list of those who have undergone coronary stent procedure in the major hospitals in Reykjavik in the years 1993 and 2003.

For over 700 subjects on this list, information on in-stent restenosis is available, including subjects with variable degree of restenosis (0-100% in-stent restenosis). A confirmed proband

with restenosis is a subject who has 50% in-stent stenosis or more as determined by coronary angiography read by an intervention cardio/radiologist.

The controls used for the study were recruited as a part of various genetic programs at deCODE. The medical history for the controls were unknown unless if the control subjects also had participated in any of the CVD genetic programs (i.e. MI, stroke, peripheral vascular disease, type II diabetes, obesity, familial combined hyperlipidemia, coronary restenosis, and hypertension genetic programs). Individuals with known MI, stroke, peripheral vascular or coronary artery disease were excluded as controls.

Subjects from the United States

Cohort from Philadelphia

The study participants from Philadelphia were enrolled at the University of Pennsylvania Medical Center through the PENN CATH study program which studies the association of biochemical and genetic factors to coronary artery disease (CAD) in subjects undergoing cardiac catheterization. A total of 3850 subjects have participated. For the purpose of the current study we selected from the PENN CATH study individuals diagnosed with one of the following coronary artery disease: MI based on criteria for acute MI in terms of elevations of cardiac enzymes and electrocardiographic changes, or a self-reported history of MI, history of coronary artery bypass surgery (CABG) or percutaneous, transluminal coronary angioplasty (PTCA). To use as controls we selected individuals who were without significant luminal stenosis on coronary angiography (luminal stenosis less than 50%). Ethnicity information was self-reported.

The University of Pennsylvania Institutional Review Board approved the study and all subjects provided written informed consent.

Cohort from Cleveland

The study participants were enrolled at the Cleveland Clinic Heart Center through the Genebank program, which is a registry of data in conjunction with biological samples for individuals undergoing coronary catheterization. The diagnostic criteria for MI were based on at least two of the following: prolonged chest pain, ECG patterns consistent with acute MI or significant elevation of cardiac enzymes. Subjects from the Genebank registry who were without significant luminal stenosis (<50% stenosis), as assessed with coronary angiography, and were without previous history of CAD, were selected as controls for the current study.

This study was approved by the Cleveland Clinic Foundation Institutional Review Board on Human Subjects and all subjects gave written informed consent.

Cohort from Atlanta

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 with CAD in subjects undergoing cardiac catheterization. For the purpose of the current study those subjects who had a self-reported history of MI, CABG, or PTCA, were selected and used as a patient group. Control subjects were selected from a group of individuals with non-vascular neurological diseases (mainly Parkinson's and Alzheimer's diseases) recruited from CRIN, their spouses, unrelated friends and community volunteers. These subjects were matched for age, and ethnicity to the patient population. Controls were excluded if they had a known history of MI or coronary artery disease. All subjects provided written informed consent. Information on ethnicity was self-reported.

Genotyping

A genome-wide scan of 1570 Icelandic individuals diagnosed with myocardial infarction (MI) and 7088 population controls was performed using Infinium HumanHap300 SNP chips from Illumina for assaying approximately 317,000 single nucleotide polymorphisms (SNPs) on a single chip (Illumina, San Diego, CA, USA). SNP genotyping for replication in other case-control cohorts was carried using the Centaurus platform (Nanogen).

Statistical Methods for Association Analysis

To test individual markers for association to disease phenotypes such as coronary artery disease or myocardial infarction, we use a likelihood ratio test to calculate a two-sided P-value for each allele of the markes. We calculate relative risk (RR) and population attributable risk (PAR) assuming a multiplicative model (C. T. Falk, P. Rubinstein, *Ann Hum Genet* 51 (Pt 3), 227 (1987); J. D. Terwilliger, J. Ott, *Hum Hered* 42, 337 (1992)). To elucidate the linkage disequilibrium between markers in the region we used the CEPH Caucasian HapMap data. We calculated LD between pairs of SNPs using the standard definition of D' (R. C. Lewontin, *Genetics* 50, 757 (1964)) and for the correlation coefficient r^2 (W. G. Hill, A. Robertson, *Genetics* 60, 615 (Nov, 1968). For the Icelandic cohort, to take into account that some of the individuals are related to each other, we obtained the null statistic of the test statistic either by simulating genotypes through the Icelandic genealogy or from the test statistic for all the 300,000 tested

for association in the initial genome-wide association scan (citation). Model-free estimates of the genotype relative risk are generated as follows: RR of genotype G_1 compared to genotype G_0 was estimated by $[n(G_1)/n(G_0)]/ [m(G_1)/m(G_0)]$ where n and m denote genotype counts in patients and controls respectively. Results from different cohorts were combined using a Mantel-Hanezel model (citation) where cohorts are allowed to have different population frequencies for the alleles/genotypes but assume to have common relative risks.

We use multiple regression to test for association between markers and quantitative traits, such as ago of onset of MI in the cases, where the number of copies of the at-risk variant carried by an individual is taken as explanatory variable and the quantitative trait as the response variable. The association is adjusted for age and gender, where appropriate, by including corresponding terms in the regression analysis as explanatory variables.

Correction for relatedness of the subjects and Genomic Control

Some of the individuals in both the Icelandic patient and control groups are related to each other, causing the chi-square test statistic to have a mean >1 and median >0.675 (Devlin, B & Roeder, K., *Biometrics* **55**, 997 (1999)). We estimated the inflation factor for the genome-wide association by calculating the average of the 305,953 chi-square statistics, which was a method of genomic control (Devlin, B & Roeder, K., *Biometrics* **55**, 997 (1999)) to adjust for both relatedness and potential population stratification. The inflation factor was estimated as 1.129 and the results presented from the genome-wide association are based on adjusting the chi-square statistics by dividing each of them by 1.129. For the Icelandic replication cohort and the combined Icelandic replication and discovery cohort, we used a previously described procedure where we simulated genotypes through the genealogy of 708,683 Icelanders to estimate the adjustment factor (S9). The corresponding adjustment factors were 1.092 and 1.029, respectively.

PCR screening of cDNA libraries

cDNA libraries were constructed from whole blood (pool of 90 individuals), EBV-transformed human lymphoblastoid cells (pool of 90 individuals), human cardiac myocyte cells (Sciencell, Cat.no. 6200), human aortic smooth muscle cells (Sciencell, Cat.no. 6110), human cardiac fibroblast ventricular cells (Sciencell Cat.no. 6310) and human primary umbilical vein endothelial cells (HUVEC) (pool of 4 individuals). Total RNA was isolated using the RNeasy RNA isolation kit (Qiagen, Cat.no. 75144), the RNeasy RNA isolation from whole blood kit (Qiagen, Cat.no. 52304) or the mirVana RNA isolation kit, using the total RNA isolation procedure (Ambion Inc. Cat.no. 1560) according to manufacturer's recommendations. cDNA libraries were prepared at deCODE using the High Capacity cDNA Archive Kit with random primers (Applied Biosystems PN 4322171). In addition to the libraries above, two commercial cDNA libraries from whole heart (Clontech-639304) and aorta (Clontech-639325) were screened.

PCR screening was carried out using the Advantage2® Polymerase mix (Clontech cat.no. 639202) according to manufacturer's instructions with primers from Operon Biotechnologies. The PCR reactions were carried out in 10 µl volume at a final concentration of 3.5 µM of forward and reverse primers, 2 mM dNTP, 1x Advantage 2 PCR buffer, 0.2 µl of Advantage enzyme and 0.5 µl of cDNA library. (See **Table 23**). Expression was detected for all of the ESTs in several of the libraries tested (**Table 24**). None of the ESTs have an open reading frame larger than 77 bp. Many of them overlap with a recently reported antisense non-coding RNA whose expression has been shown to cocluster with p14/ARF (Pasmant, E., *et al.*, *Cancer Res* **67**, 3963 (2007)).

Sequencing of CDKN2A and CDKN2B

PCR amplifications and sequencing reactions were set up on Zymark SciClone ALH300 robotic workstations and amplified on MJR Tetrads. PCR products were verified for correct length by agarose gel electrophoresis and purified using AMPure (Agencourt Bioscience). Purified products were sequenced using an ABI PRISM Fluorescent Dye Terminator system, repurified using CleanSEQ (Agencourt), and resolved on Applied Biosystems 3730 capillary sequencers. SNP calling from primary sequence data was carried out using deCODE Genetics Sequence Miner software. All *CDKN2A* and *CDKN2B* variants identified by the automated systems were confirmed by manual inspection of primary signal traces. Samples from 96 early onset MI patients were sequenced using primers indicated in **Table 25** and a list of the SNPs identified is provided in **Table 26**.

Surveying for candidate regulatory variants in the candidate region

The University of California Santa Cruz genome browser (genome.ucsc.edu) was used to extract positions of SNPs and conserved TF binding sites for a 600 kb surrounding the MI region (hg release 17, chromosome 9, bases 21800000 to 22400000). The two tables were cross referenced with Python scripts and SNPs that resided in binding sites were interrogated for LD with rs1333040 in the CEU sample of Hapmap (release 22). The analyses were implemented for release 18 of the human genome, and the results converted to hg 17 coordinates.

This bioinformatic analysis of 600 kb surrounding the MI region yielded 16 SNPs which coincide with conserved binding sites for transcription factors (**Table 27**). Lack of LD to SNPs tagging the MI haplotype enabled exclusion of a half of the 16 SNPs from this candidate list. The remaining polymorphisms could impact gene function by altering conserved TF binding sites. In parallel we looked for correlation between SNPs located in conserved blocks (based on Multiz alignments available through the UCSC genome browser, release hg 18) and the MI haplotype tagging SNPs. While about half of the 74 SNPs are represented in HapMap, we found none that were highly correlated with the MI haplotype (data not shown). In sum, these analyses do not identify particular SNPs as putative regulatory variants on the MI haplotypes. Note of caution, the analysis will only detect a proportion of the functional candidates in the region because; i) the MI haplotypes have not been sequenced fully, ii) several candidate SNPs are not typed in

Hapmap, hence it is unknown whether they sit on the risk conferring haplotypes. In addition, it is a realistic possibility that polymorphisms in less conserved regions are the functional MI variants.

Example 1.

Genome-wide association study

We successfully genotyped 1570 Icelandic myocardial infarction patients and 7088 population control individuals without known history of coronary artery disease (Cohort A) using the Illumina 330K chip. We performed a genome-wide scan for association to MI, testing individually each of the 309,091 SNPs that was successfully genotyped. Three markers (rs10116277, rs1333040, rs2383207), all located in a single LD block (denoted herein as LD block C09) on chromosome 9p show strong association to MI (see Figure 1 and Table 1a). All three markers are strongly correlated (Table 2) and the population frequency of the at-risk variants range from 42% to 49%, and the corresponding relative risk is approximately 1.2. The risk alleles of the same three markers also showed a significant correlation to lower age at onset within the MI patient – individuals carrying the at-risk variant are at significant risk of developing MI at a younger age than individuals who are non-carriers of the at-risk allele (Table 8a).

The LD-block LD block C09 containing the three associated marker is flanked and defined by two recombination hot-spots – one at approximately 21,920,000bp in Build 34, the other approximately 22,150,000bp on chromosome 9 (*Nature* **437**, 1299-1320 (27 October 2005))). Investigating other genetic markers in LD block C09, we identified two micro-satellite markers, D9S1870 and D9S1814, that are strongly correlated with the at-risk alleles of the markers rs10116277, rs1333040, and rs2383207. Table 1b shows the association of each of the alleles of the microsatellite markers to MI in the same cohort of MI patients and controls that was used in the genome-wide association scan. For marker D9S1814 the associated allele was allele 0, however for D9S1870 multiple alleles (alleles -4, -2 and 0) showed increased risk of MI. By investigating the correlation between the different alleles of D9S1870 to the at-risk alleles of the markers rs10116277, rs1333040 and rs2383207, observed that by pooling together all alleles of D9S1870 shorter than 2 (alleles -6, -4, -2 and 0 respectively), the composite allele, denoted X, was strongly correlated to the original at-risk alleles of the SNP (**Table 2**). Composite allele X of D9S1870 and allele 0 of D9S1814 show similar association to MI as the at-risk alleles of rs10116277, rs1333040 and rs2383207 (**Table 1c**) and all five at-risk alleles are highly correlated (**Table 2**).

Further investigation of all SNPs in the HapMap v9 CEU dataset that are located in the chromosomal region defined by LD block C09 identified further 88 markers that are strongly correlated with the at-risk allele of at least one of the five markers rs10116277, rs1333040, rs2383207, D9S1814 and D9S1870 (**Table 3**) and hence those markers could also be used as surrogate markers to tag the observed association to MI in LD block C09.

We genotyped the micro-satellite marker D9S1870 in a large cohort of over 70,000 Icelanders that included, among others, 668 additional MI cases and 58,643 additional controls without known history of coronary artery disease (Cohort B). In addition we typed the microsatellite D9S1870 in three replication cohorts from US incdluding; 549 MI patients and 606 controls from Cleveland; 580 MI cases and 404 controls from UPenn; and 400 MI cases and 477 controls from Emory. All individuals in the US cohorts are of Caucasian origin. We tested the composite allele X for association to MI in all four cohorts (**Table 4a and b**), and all but one (Cleveland) showed significant association. Combining the results from the four replication cohorts (**Table 4c**) yielded a combined P-value = 2.65×10^{-8}. Combined with the original Icelandic cohort used in the genome-wide association (Cohort A) the P-value is 1.44×10^{-12} and, assuming a multiplicative model, each allele X confers an estimated relative risk (RR) of 1.214 [95%CI:1.151-1.281] per copy carried, compare with the risk for non-carriers. The corresponding combined population attributable risk (PAR) is 17.1%.

If we investigate separately the risk conferred for individuals heterozygous for the risk allele X and individuals homozygous for X, relative to individuals that do not carry X (**Table 7a**), the estimated genotype relative risk (GRR) for heterozygous individuals is 1.204 [CI:1.094-1.324] and for homozygous individuals GRR is 1.507 [CI:1.360-1.670]. This is consistent with the multiplicative model, i.e an additive contribution of the allele X to the risk of MI.

We further investigated the correlation of the risk allele X to age of onset of MI in all four cohorts. Restricting the analysis to early-onset MI cases, defined as a MI event before the age of 50 for males and before the age of 60 for females, the relative risk for the cohorts combined increases to 1.331 [CI:1.223-1.449; P = 3.96×10^{-11}; PAR = 24.7%] compared to 1.214 for all MI cases (**Table 6c**). Correspondingly the genotype relative risk increases to 1.314 [CI:1.105-1.562] and 1.790 [CI:1.517-2.113] for heterozygous and homozygous carriers of allele X respectively (**Table 7b**). Alternatively, we tested using multiple regression the correlation between the number of copies of X carried by individuals in the MI case group and the age of onset of MI (**Table 8b**). Combining results from all four cohorts, we observed that the mean age of onset decreased by 0.95 year [SE=0.25] for each copy of X carried by the MI individuals (P = 0.000099).

Among individuals typed for the marker D9S1870 in the Icelandic and the Emory cohort are individuals that have undergone percutaneous transluminal coronary angioplasty (PTCA) or coronary artery bypass graft surgery (CABG), both indicative of severe coronary artery disease (CAD). We tested if those individuals also had an increased frequency of the risk allele X compared to controls, speculating that the variant might predispose individuals to a more general coronary artery disease than just MI (**Table 5a and c**). In both cohorts we observed similarly increased risk for PTCA and CABG as for MI, if not stronger, that was very significant in the Icelandic cohort, although for the Emory cohort only the association to PTCA was significant. It should be noted, however, that the number of individuals with PTCA and CABG in the Emory

cohort is small. This association remained in the Icelandic cohort even after removing known MI cases from the PTCA and CABG groups (**Table 5b**).

In addition, in the Icelandic cohort we investigated the association to other diseases related to coronary artery disease, such as peripheral artery disease (PAD) and stroke phenotypes such as infarct or transient ischemic attack (TIA). For 1661 PAD cases we observed a very significant association, $P = 5.36\times10^{-5}$ and RR = 1.154 [CI:1.074-1.239] (**Table 5a**) - this association remained significant after removing MI cases from the PAD cohort in the analysis although the effect was somewhat weaker (**Table 5b**). For 1678 individuals with infarct or TIA we did not observe significant association to X, however, for individuals diagnosed with large vessel diseases (LVD) - the stroke sub-phenotype that is most closely related to coronary artery disease - we observed an increased risk, RR =1.120 or 1.172 if we include or exclude MI cases respectively (**Table 5a and b**). There are however only 197 individuals in the Icelandic cohort diagnosed with LVD and this association in not statistically significant.

We investigated the frequency of the risk variant X of marker D9S1870 in a group of 454 Icelandic individuals for which we had information on in-stent restenosis and that are genotyped for the variant. The cohort was divided into individuals with severe in-stent restenosis (50% or greater) and individuals with mild in-stent restenosis (less than 50%). As all those indivduals have undergone PTCA, and hence have coronary artery disease, both groups have significantly higher frequency of the variant X than is observed in controls (Table 9a). However, the frequency in the group of individuals with severe restenosis is higher than in the group with mild restenosis, RR = 1.067 [CI:0.827-1.376], and although this difference is not significant (**Table 9b**), this suggests that the variant could be indicative of the severity of in-stent restenosis in coronary artery disease patients that have undergone PT

Example 2.

A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction

Coronary artery disease (CAD), including acute myocardial infarction (MI), is the leading cause of death worldwide (Thom, T., *et al., Circularion* **113:**e85 (2006)). Identification of the underlying genetic architecture of heart disease may provide improved risk assessment and better measures for prevention and treatment.

To this end we conducted a genome-wide association study on Icelandic patients with MI, using the Illumina Hap300 chip. After quality filtering, 305,953 SNPs were tested for association to MI in a sample of 1607 cases, with age at onset before 70 in males and 75 in females, and 6728 controls without a history of CAD (Helgadottir, A., *et al., Science* **316:**1491 (2007)). The results were adjusted for relatedness between individuals and potential population stratification using a method of genomic control (*3*). Although none of the SNPs were significant after adjusting for the number of tests performed more signals bordering on significance were

observed than expected by chance. Hence, we further explored the SNPs that were closest to genome wide significance.

The strongest association to MI was observed with three correlated SNPs, rs1333040, rs2383207 and rs10116277, each with odds ratio (OR) around 1.22 for the risk allele and *P* of approximately 1×10^{-6} (**Table 15**). All three SNPs are located within the linkage disequilibrium (LD) block on chromosome 9p21 denoted herein as LD block C09 (Fig. 1). Apart from these three SNPs, eleven other SNPs in the same LD block showed nominally significant association to MI. The associations to these SNPs tended to become weaker after accounting for the association to the three SNPs mentioned above (**Table 15**). After adjustment, a few remained nominally significant ($P < 0.05$), but none had a $P < 0.01$.

To replicate the observed associations we genotyped the three SNPs, rs1333040, rs2383207 and rs10116277, in an additional 665 Icelandic MI cases and 3533 controls and in three case-control sample sets of European descent from three cities from the United States: Philadelphia, Atlanta, and Durham (*2*). For consistency we used the same age at onset criteria in the association analysis as for the discovery group. The association to MI was replicated with significance in all four groups (**Table 16**). When the replication sets were combined using a Mantel-Haenszel model (Mantel, N. & Haenszel, *J. Natl. Cancer Inst.* **22:**719 (1959)), all three SNPs showed highly significant association to MI ($P < 1\times10^{-8}$), with ORs comparable to those observed in the Icelandic discovery samples. When all groups were combined, rs2383207 showed the most significant association ($P = 2.0\times10^{-16}$), with an OR of 1.25 (95% CI 1.18-1.31) for the risk allele G. It is noted that rs2383207 and rs10116277 are highly correlated ($r^2 = 0.90$) and their effects could not be reliably distinguished from each other in these data. The SNP rs1333040 is also substantially correlated with rs2383207 and rs10116277 ($r^2 = 0.57$ and 0.67 respectively). In an attempt to refine this association signal, we identified the SNPs that are substantially correlated with rs2383207 ($r^2 > 0.5$) based on the Hapmap CEU data and are not part of the Illumina Hap300 chip. Among the 36 such SNPs, we selected eight to be genotyped. Each of the 36 SNPs was either one of the eight or it had a very good surrogate among them ($r^2 > 0.90$) (**Table 21**). With data from all case-control groups combined, allele G of the refinement SNP rs10757278 showed the strongest association to the disease (OR = 1.28, $P = 1.2\times10^{-20}$; **Tables 12 and 16**). Furthermore, while rs2383207 was no longer significant after adjusting for rs10757278 ($P = 0.25$), rs10757278 remained significant with adjustment for rs2383207 ($P = 2.0\times10^{-5}$). Among the SNPs in this region that showed very significant association to the disease when tested individually, none was significant after adjustment for rs10757278 with the exception of the refinement SNP rs13330406, which was marginally significant ($P = 0.044$) with adjustment (**Table 22**). Henceforth, for simplicity of presentation, we focus on the most significant SNP rs10757278 in the main text but additional results for other SNPs in the region are provided **Tables 16 to 20**.

To investigate the mode of inheritance in more detail, we computed genotype specific ORs for rs10757278. With results from all groups combined, relative to non-carriers, the ORs for

heterozygous and homozygous carriers of the risk allele G were 1.26 and 1.64, respectively (**Table 13**). Assuming a frequency of 45.3% for the allele, the average of the frequencies in Iceland and the US, the corresponding PAR is 21%.

Because the impact of genetic factors on CAD has been shown to be greater at early ages (*5*) we investigated the correlation of allele G of rs10757278 to age at onset of MI. Note that in this analysis we used all cases with a known age at onset including those who had onset after the age of 70 or 75 for males and females, respectively. This added a total of 973 cases to the study groups compared to what was used in the case-control analyses. Regressing the age at onset on the number of risk alleles showed that, for each copy of the risk allele, the age at onset of MI was on average reduced by approximately one year ($P = 2.9\times10^{-7}$) (**Table 18**). Alternatively, restricting the case-control analysis to early onset MI, defined as an MI before the age of 50 for males and before the age of 60 for females, the allelic OR for rs10757278 G in all groups combined increased to 1.42 (95% CI 1.31 - 1.53) (**Table 19**). Relative to non-carriers, genotype specific OR for early onset MI is 1.49 and 2.02 for heterozygous and homozygous carriers of the risk allele, respectively (**Table 13**).

Having established that allele G of rs10757278 is associated to MI, we explored its impact on the broader phenotype of CAD (**Table 14**). To eliminate bias that could have arisen from the selection of the most significant variants in the initial genome-wide study, the cases and controls from the Icelandic discovery group (Iceland A) were not included here. We do note that if the latter were included, there would be little change to the estimated effects, but the results would become more significant due to the larger sample sizes. Also, the group from Durham did not have CAD cases without MI. As expected, rs10757278 was associated with high significance to CAD (OR = 1.29, $P = 3.6\times10^{-14}$ for the groups combined). After removal of MI cases from the analyses, the associations remained significant for the groups from Iceland and Atlanta, but not in the Philadelphia group. Combining results from the three groups gave an OR of 1.24 ($P = 0.000011$).

The variants on chromosome 9q21 associated to MI are located in an LD block that contains the *CDKN2A* and *CDKN2B* genes. The proteins encoded by these genes, called $p16^{INK4a}$, ARF and $p15^{INK4b}$ have a critical role in regulating cell proliferation, cell aging/senescence, and apoptosis in many cell types (Kim, W.Y. & Sharpless, N.E. *Cell* **127:**265 (2006)). These are all important features of atherogenesis, the underlying cause of MI and CAD (Lusis, A.J. *Nature* **407:**233 (2000); Minamino, T. & Komuro, I. *Circ Res* **100:**15 (2007)). Sequencing of 93 early onset MI patients across exons, exon-intron junctions, and regulatory regions of *CDKN2A* and *CDKN2B* did not reveal obvious candidates for functional variants or other variants that could account for the observed association to rs10757278 (**Tables 25 and 26**). In addition to *CDKN2A* and *CDKN2B* genes, the LD block contains two exons of the mRNA transcript AF109294, a hypothetical methylthioadenosine phosphorylase fusion protein mRNA and several ESTs that are expressed in various tissues (Helgadottir, A., *et al., Science* **316:**1491 (2007)). The functional relevance of the variants of this genomic region to MI/CAD remains to be elucidated.

In summary, we have shown that a common genetic variant located in the vicinity of the tumor suppressor genes *CDKN2A* and *CDKN2B* on chromosome 9p21 associate to MI. This is the first common variant discovered to consistently confer substantial risk (OR > 1.20) of MI in multiple case-control groups of European descent. Due to its high frequency, the population attributable risk of the variant is approximately 21% for MI in general and approximately 31% for early onset cases, which is substantial from a public health point of view. However, as the relative risks are not extremely high, it explains only a small fraction of the familial clustering of the disease and would not generate large linkage scores. Hence, others susceptibility variants remain to be identified and some could be located in candidate regions identified by genome-wide linkage scans (Zintzaras, E. & Kitsios, G., *J Hum Genet* **51:**1015 (2006); Wang Q., *et al., Am J Hum Genet* **74:**262 (2004); Samani, N.J., *et al., Am J Hum Genet* **77:**1011 (2005)). There is evidence supporting that the variant identified here could increase the risk of CAD in general in addition to their impact on MI, an observation that warrants further investigation. The mechanism whereby the genetic variants exert their effects in the pathogenesis of MI remains to be elucidated.

Example 3.

Genotyping of polymorphic markers identified through sequencing

Sequencing of the exons of CDKN2A and CDKN2B genes, the exon-intron junctions and potential regulatory regions using the primers as indicated in Table 25 resulted in the identification of a number of SNPs, as shown in Table 26. Flanking seqeunces for three of those SNPs that were not found in public databases are indicated in Table 31. As it is possible that SNP markers or other polymorphisms in LD with the markers found to be associating to MI in this region of chromosome 9 show association with a higher risk, we genotyped these additional markers by sequencing, as indicated in Table 28. Several of the markers show association to MI with RR values as high as 1.7-1.8, in particular markers SG09S291 and rs2069416.

Example 4

Association to related cardiovascular disorders

We have investigated association of the at-risk variants of the invention to the related disorders peripheral artery disease (PAD), abdominal aorta aneurysm (AAA) and large vessel disease stroke (LVD) for three of the markers giving signal on Chromosome 9 as presented herein. As can be seen in **Table 29**, these markers are associated with these related disorders. The association is particularly compelling for AAA, wherein significant association is observed for a large number of markers in addition to these three, as shown in Table 30. These results

illustrate that the markers and haplotypes of the invention are indeed reflective of disorders related to coronary artery disease, MI and in-stent restenosis, such as abdominal aorta aneurysm.

Example 5

Further refinement of association to the arterial phenotypes AAA, IA and Stroke.

To investigate the effect of rs10757278 on other cardiovascular diseases in more detail, we further explored the association to abdominal aortic aneurysm (AAA) and Stroke, and also investigated the arterial disorder intrachranial aneurysm (IA).

Methods

Study Populations

Coronary artery disease groups

The coronary artery disease groups from Iceland and the United States were as described above (see also Helgadottir, A., *et.al., Science* **316:**1491-3 (2007))

Icelandic controls

The 14278 Icelandic controls used in the association study were selected among individuals who have participated in various GWA studies and were recruited as part of genetic programs at deCODE. The medical histories of the controls were unknown unless they had also participated in one or more of the CVD genetic programs (i.e. MI, stroke, PAD, T2D, obesity, familial combined hyperlipidemia, coronary restenosis, and hypertension). Individuals with known MI, stroke, PAD or CAD, or with T2D were excluded as controls. Of the 14259 controls 9202 overlap with those used in our previous GWA study in MI (Helgadottir, A., *et.al., Science* **316:**1491-3 (2007)). The controls included 5615 males and 8644 females and their mean age was 55.2 (SD 21.7). The breakdown of the control group into the various genetic programs was approximately (with the frequency of the two variants, rs10757278 allele G and rs10811661 allele T in parenthesis): Schizophrenia 500 (0.428/0.825), Prostate cancer 900 (0.447/0.815), Breast Cancer 1300 (0.433/0.817), Colon Cancer 700 (0.413/0.817), Addiction 2600 (0.444/0.814), Anxiety 900 (0.442/0.824), Infectious diseases 1200 (0.434/0.821), Population Controls 700 (0.427/0.830), Microarray expression studies 400 (0.445/0.817), Longevity 1100 (0.450/0.819), Migraine 1100 (0.446/0.818), Restless Leg Syndrome 400 (0.439/0.812), Alzheimer disease 350 (0.457/0.822), Asthma 1300 (0.419/0.819), Dyslexia 600 (0.438/0.830). No significant differences in frequencies were observed between the disease groups for either of the two variants (P = 0.52 and P = 0.99 for rs10757278 and rs10811661, respectively).

Stroke groups

Icelandic stroke patients were recruited from a registry of over 4000 individuals which includes individuals diagnosed with ischemic stroke or TIA at the major hospital in Reykjavik, the Landspitali University Hospital, during the years 1993 to 2002. Stroke patients have been enrolled over the past nine years through the cardiovascular disease (CVD) genetics program at deCODE. Swedish patients with ischemic stroke or TIA attending the stroke unit or the stroke outpatient clinic at Karolinska University Hospital, Huddinge unit in Stockholm, Sweden were recruited from 1996 to 2002 as part of an ongoing genetic epidemiology study, the South Stockholm Ischemic Stroke Study (SSISS). All patients from Iceland and Sweden had clinically relevant investigations performed, including brain imaging with computed tomography (CT) or/and magnetic resonance imaging (MRI) as well as ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries, echocardiography, Holter monitoring, MR-angiography, CT-angiography and standardized blood tests. Patients were classified into ischemic subtypes according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification by a physician reviewing original imaging and data (Adams, H.P. Jr., *et al.*, *Stroke* **24:**35-41 (1993)). Patients classified with cardioembolic stroke and documented atrial fibrillation were excluded from the analysis. The Swedish controls used in this study are population-based controls recruited from the same region in central Sweden as the patients, representing the general population in this area. The individuals were either blood donors (recruited in 2001) or healthy volunteers (collected in 1990-1994) recruited by the Clinical Chemistry Department at the Karolinska University Hospital to represent a normal reference population. These stroke studies from Iceland and Sweden were approved by relevant Institutional Review Boards or ethics committees and all participants provided written informed consent.

Intracranial aneurysm groups

Icelandic IA patients were identified through an inpatient database from 1994-2006 at the Landspitali University Hospital, which is the only hospital with a neurosurgical service in the country. All patients in the years 1996-2006 with the ICD10 diagnosis I60.0-7 (aneurysmal subarachnoid hemorrhage), I67.1 (ruptured cerebral aneurysm) and I69.0 (sequele of subarachnoid haemorrhage) were enrolled, as well as patients with the ICD9 diagnosis 430 (subarachnoid hemorrhage from ruptured cerebral aneurysm) in the years 1994-1996. This totalled 367 IA patients. All patients had clinically relevant investigations performed, including CT scan of the head and or conventional cerebral angiogram, CT- angiogram or MRi angiogram. DNA samples were available for 170 of the 367 patients.

Dutch patients with ruptured (91.5%) or unruptured (8.5%) IA admitted to the University Medical Center Utrecht were used for the study. Ruptured intracranial aneurysms were defined by symptoms suggestive of subarachnoidal hemorrage (SAH) combined with subarachnoid blood on CT and a proven aneurysm at angiography (conventional angiogram, CT- or MR-angiogram) and unruptured intracranial aneurysms were identified by CT or MR angiography or conventional angiography. Multiple intracranial aneurysms were found in 20.5%

of cases. Mean age at time of SAH was 49.5 years (range 10-84) and 66.1% of the patients were females. The controls were healthy Dutch blood bank donors of European origin.

Finnish IA patients admitted for treatment of intracranial aneurysm at either the University Hospital of Kuopio, or University Hospital of Helsinki, in Finland, were used for the study. This study group and the Finnish controls used have been described previously (Weinsheimer, S. *et al., Stroke* **38:**2670-6 (2007)).

The Icelandic, Dutch and Finnish IA studies were approved by relevant Institutional Review Boards or ethics committees and all participants provided written informed consent.

Peripheral arterial disease groups

Icelandic patients with PAD were recruited from a registry of individuals diagnosed with PAD at the major hospital in Reykjavik, the Landspitali University Hospital, during the years 1983 to 2006. The PAD diagnosis was confirmed by vascular imaging or segmental pressure measurements. PAD patients have been enrolled over the past nine years as part of the CVD genetics program at deCODE.

Italian patients and controls were recruited among subjects consecutively admitted to the Department of Internal Medicine and Angiology of the A. Gemelli University Hospital of Rome, from 2000 to 2001. Inclusion criteria for the PAD group were European descent and presence of PAD. Diagnosis of PAD was performed in accordance with established criteria (*J Vasc Surg* **4:**80-94 (1986)). All patients had an ankle/arm pressure index lower than 0.8 and were at Fontaine's stage II, with intermittent claudication and no rest pain or trophic lesions. Inclusion criteria for the control group were European descent, absence of PAD and CAD and no relationship with cases. Additional, exclusion criteria from the study were tumours, chronic inflammatory diseases, and autoimmune diseases (Flex, A., *et al., Eur J Vasc Endovasc Surg* **24:** 264-8 (2002)).

Swedish PAD patients and controls were recruited at the Department of Vascular Diseases at Malmö University Hospital, a single referral centre for all patients with critical limb ischemia in the three southernmost health-care districts in Sweden (723,750 inhabitants in 2001). The diagnosis of critical limb ischemia was made in accordance with TransAtlantic Inter-Society Consensus scientific criteria of ulceration, gangrene, or rest pain caused by PAD proven by ankle pressure (<50 to 70 mm Hg), reduced toe pressure (<30 to 50 mm Hg), or reduced transcutaneous oxygen tension (Dormandy, J.A. & Rutherford, R.B., *J Vasc Surg* **31:**S1-S296 (2000)). Diagnosis was confirmed by an experienced vascular surgery consultant and toe pressure measurements in patients with arteries in the affected leg that were noncompressible and the ankle pressure was >50 to 70 mm Hg. The control group consisted of healthy individuals included in a health screening programme for a preventive medicine project. None of those had symptomatic PAD (Barani, J., *et al., J Vasc Surg* **42:**75-80 (2005)).

New Zealand PAD patients were recruited from the Otago-Southland region of the country, the vast majority (>97%) being of Anglo-European ancestry as reported previously (Jones G.T., *et al., Clin Chem* **53:**679-85 (2007)). PAD was confirmed by an ankle brachial index<0.7, pulse volume recordings and angiography/ultrasound imaging. The control group consisted of elderly individuals with no previous history of vascular disease from the same

geographical region. Controls were asymptomatic for PAD and had ankle brachial indexes >1. An abdominal ultrasound scan excluded concurrent AAA from both the PAD and control groups.

The Icelandic, Italian, Swedish, and New Zealand PAD studies were approved by relevant Institutional Review Boards or ethics committees and all participants provided written informed consent.

Abdominal aortic aneurysm groups

Icelandic patients with AAA were recruited from a registry of individuals who were admitted either for emergency repair of symptomatic or ruptured AAA or for an elective surgery to the Landspitali, University Hospital, in Reykjavik, Iceland in the years 1980-2005. Subjects with AAA were enrolled over the last nine years as part of the CVD genetics program at deCODE. In some of the analyses AAA cases that overlapped with a comprehensive list of CAD patients (Helgadottir, A., *et.al., Science* **316:**1491-3 (2007)) diagnosed in Iceland in the years 1981-2006 were excluded. Of the 397 (288 males and 109 females, mean age 75.3 (SD 8.7)) AAA cases, 208 overlapped with the CAD patients. Of the 189 (131 males and 58 females, mean age 75.5 (SD 9.3)) remaining CAD, information was not available for 138 cases and 51 individuals reported in a questionnaire as not having been diagnosed with CAD.

UK patients with AAA referred to vascular surgeons at 93 UK hospitals were entered into UK Small Aneurysm Trial. For the purpose of the current study those randomised to surveillance in the UK Small Aneurysm Trial with AAA diameter 4.0-5.5 cm were selected as a patient group, although some patients had been monitored before their aneurysm reached the 4.0cm threshold for the trial. Mean AAA diameter at baseline was 4.5 cm (3.2-5.5 cm) (Eriksson, P., *et al., Br J Surg* **92:**1372-6 (2005)). Information on the occurrence of CAD was available for 97% (466 out of 479) of AAA cases. History of CAD was regarded as positive if the subject was under treatment for angina, had a previous MI, coronary artery bypass graft surgery or angioplasty or if ECG coding had any indications of ischeamia, as judged by two independent expert observers. Among those with this information, the frequency of CAD amongst the AAA subjects was 52%. Controls were of European descent, recruited from England.

Belgian and Canadian patients with AAA who were admitted either for emergency repair of ruptured AAA or for an elective surgery to the University Hospital of Liège in Belgium and to Dalhousie University Hospital in Halifax Canada, respectively, were used for this study. Details of these case-control sets have been previously reported (Ogata, T., *et al., J Vasc Surg* **41:**1036-42 (2005)). All patients were of European descent and had a diameter of infrarenal aorta ≥ 3 cm. Thirty-five patients were diagnosed with AAA using ultrasonography and did not undergo surgery because of old age or because the aneurysm was relatively small. Approximately 40% of AAA patients had a family history of AAA. For Belgian AAA patients, information on CAD history was ascertained for those who underwent surgery through interviews as well as from medical files. In addition, all patients underwent cardiologic explorations such as transthoracic echography, stress tests and coronary angiography if CAD was suspected. CAD information for this study was available for 45% (79 out of 176) of AAA cases from Belgium. Among those with this information, the frequency of CAD amongst the AAA subjects was 29%. Control samples (51%

males) of European descent were obtained from spouses of AAA patients or from individuals admitted to the same hospitals for reasons other than AAA.

Patients admitted to the University Hospital of Pittsburgh for either elective or emergency surgery for AAA were selected for the study (St Jean, P.L., *et al., Ann Hum Genet* **59:**17-24 (1995)). History of CAD was self-reported and was available for 86% (87 out of 101). Among those with this information, the frequency of CAD amongst the AAA subjects was 48%. Controls were selected from participants of the PENN CATH study program at the University of Pennsylvania Medical Center Philadelphia. The control group represents individuals who were without significant luminal stenosis on coronary angiography (luminal stenosis less than 50%) and did not have a history of MI. These are the same controls as were used in the association analysis for the CAD samples from Pennsylvania (Helgadottir, A., *et.al., Science* **316:**1491-3 (2007)).

New Zealand patients with AAA were recruited from the Otago-Southland region of the country, the vast majority (>97%) being of Anglo-European ancestry as reported previously (Jones, G.T., *et al., Clin Chem* **53:**679-85 (2007)). Approximately 80% of patients had undergone surgical AAA repair (typically AAA's >50mm in diameter). Controls were the same vascular disease free individuals as described for comparison with the New Zealand PAD group. CAD information was available for 98% (575 out of 588) of the AAA patients. Of those with information the frequency of CAD was 40%.

AAA sample set from the Netherlands was recruited from 8 centres in the country, mostly when patients visited their vascular surgeon or in rare cases during hospital admission. The controls were healthy Dutch blood donors of European origin. Information on other CVD was self-reported and available for 69% (330 out of 480). Treatment for angina pectoris, previous MI, coronary bypass surgery or stent insert was considered as CAD. Of the 330 with information, 96 had CAD (29%).

These AAA studies from Iceland, UK, Belgium, Canada, Pennsylvania, The Netherlands and New Zealand were approved by relevant Institutional Review Boards or ethics committees and all participants provided written informed consent.

SNP Genotyping.

SNP genotyping for all samples was carried out at deCODE genetics in Reykjavik, Iceland. Individual SNP genotyping was carried out with the Centaurus (Nanogen) platform (Kutyavin, I.V., *et al., Nucleic Acids Res* **34:**e128 (2006)). The quality of each Centaurus SNP assay was evaluated by genotyping each assay in the CEU and/or YRI HapMap samples and comparing the results to the HapMap data. The key markers rs10757278 and rs10811661 were re-genotyped on more than 10% of samples and a mismatch was observed in less than 0.5% of samples. For some of the samples we had previously genotyped the SNPs rs1333040, rs2383207, and rs10116277 either with the Illumina 317K Bead chip or with the Centaurus method. These SNPs are highly correlated with rs10757278 (r^2 = 0.57, 0.87, and 0.90, respectively, in the HapMap CEU dataset) and were used to impute the genotypes for rs10757278 where they were missing. In addition, for a large number of the Icelandic samples the SNP rs2383208, which is present on

the Illumina 317K Bead chip, was previously genotyped. This SNP is a perfect surrogate for the SNP rs10811661 (r^2 = 1 in the HapMap CEU dataset) and was used to impute genotypes for rs10811661.

The SNPs did not deviate from Hardy Weinberg Equilibrium in any of study cohorts used for the analyses.

Association Analysis

We used a standard likelihood ratio statistics, implemented in the NEMO software (to calculate two-sided *P* values and odds ratio (OR) for each individual allele, assuming a multiplicative model for risk, i.e., that the risk of the two alleles a person carries multiply. Allelic frequencies, rather than carrier frequencies are presented for the markers, and, for the Icelandic study groups, *P* values are given after adjustment for the relatedness of the subjects by simulating genotypes through the genealogy of 708,683 Icelanders as previously described (Stefansson, H., *et al., Nat Genet* **37:**129-37 (2005)). When estimating genotype specific OR (**Table 34**) genotype frequencies in the population were estimated assuming HWE. Heterogeneity tests were performed assuming that the estimates of OR for various groups have log-normal distributions. A likelihood ratio chi-square test was used with associated degrees of freedom equal to the number of groups compared minus one.

In general, allele/haplotype frequencies are estimated by maximum likelihood and tests of differences between cases and controls are performed using a generalized likelihood ratio test. This method is particularly useful in situations where there are some missing genotypes for the marker of interest and genotypes of another marker, which is in strong LD with the marker of interest, are used to provide some partial information. To handle uncertainties with phase and missing genotypes, maximum likelihood estimates, likelihood ratios and *P* values are computed directly for the observed data, and hence the loss of information due to uncertainty in phase and missing genotypes is automatically captured by the likelihood ratios.

The correlation of rs10757278 allele G and rs10811661 allele TT to both age and sex was tested in the Icelandic control population. Neither of the alleles demonstrated significant association to these covariates. In addition, no significant difference was detected in the frequency of the variants between males and females within the AAA cases (data not shown). Furthermore, including age and sex as covariates in the association analysis of rs10757278 allele G to AAA in the Icelandic samples had negligible impact on the results. Thus, for simplicity, the association analysis is presented without adjustment for age and sex.

The possibility that the association results observed for rs10757278 allele G to AAA was influenced by population stratification was addressed for the UK AAA cases and controls by typing 13 SNPs identified by the WTCCC as showing strong evidence for geographic differentiation in the WTCCC samples (*Nature* **447:**661-78 (2007)). Only one of those SNPs showed nominally significant difference between the UK AAA cases and controls (P = 0.017), which is not significant if we adjust for having tested 13 SNPs. If we adjust for this SNP in the case-control analysis, the association of rs10757278 allele G to AAA in the UK case-control group

is not affected (P = 0.0052 and OR = 1.36 instead of P = 0.0063 and OR = 1.35). In the Icelandic case-control analysis we have adjusted for the relatedness of the study individuals. This adjustment has been shown to agree very well with the adjustment based on genomic control, which would include adjustment for any population stratification, in recent publications of genome-wide association studies in the Icelandic population (Steinthorsdottir, V., *et al., Nat Genet* **39:**770-5 (2007); Gudmundsson, J. *et al., Nat Genet* **39:**631-7 (2007)). Most importantly, the very similar allelic odds ratios obtained from the five AAA data sets makes it highly unlikely that population stratification has any substantial impact on the estimates of the effect.

Results from multiple case-control groups were combined using a Mantel-Haenszel model in which the groups were allowed to have different population frequencies for alleles, haplotypes and genotypes but were assumed to have common relative risks (Mantel, N. & Haenszel, W. *J Natl Cancer Inst* **22:** 719-48 (1959)).

Results

The results shown in **Table 32** shows results from the Icelandic IA cohort, replication cohorts, and a combined analysis for the cohorts. When data from the multiple case-control groups studied were combined separately for IA, AAA, PAD and LAA/cardiogenic stroke, rs10757278 allele G showed significant association to all of the four phenotypes. However, the estimated effect size differed substantially and was strongest for AAA (combined analysis, OR= 1.31, $P = 1.2 \times 10^{-12}$) and IA (combined analysis, OR= 1.29, $P = 2.5 \times 10^{-6}$). In addition to the high overall statistical significance, it is also important to note that the estimated risk conferred by rs10757278 allele G to IA and AAA was very similar across the three IA sample sets from Iceland (OR=1.36), Finland (OR=1.33), and the Netherlands (OR=1.24) (P_{het}, the P-value for the test of heterogeneity = 0.75), and the seven AAA sample sets from Iceland (OR=1.37), Belgium (OR=1.21), Canada (OR=1.29), Pennsylvania US (OR=1.39), United Kingdom (UK) (OR=1.35), Netherlands (OR=1.31), and New Zealand (OR=1.25), (P_{het} = 0.98). The effect of rs10757278 allele G on IA and AAA is comparable to that previously reported for CAD (Helgadottir, A., *et al. Science* **316:**1491-3 (2007)) (**Table 32**).

Because of high co-morbidity between AAA and CAD, we explored the nature of the effect of the variant on the two conditions, by repeating the association analysis for AAA after removing cases with evidence of CAD. As shown in **Table 33**, the effect of rs10757278 allele G on AAA without evidence of CAD was only slightly smaller than that for the whole sample sets, or OR=1.3 for the Icelandic, 1.31 for UK, 1.19 for Pennsylvania, 1.20 for Belgian, 1.25 for The Netherlands, and 1.18 for the New Zealand sample sets. For the six different groups with available CAD information, after removing known CAD cases the combined OR was 1.25 and P = 3.0×10^{-6}, indicating that the association to AAA is not simply a consequence of the association between rs10757278 allele G and CAD. To our knowledge there is no evidence in the literature suggesting co-segregation of IA and CAD. Furthermore, the gender ratio in IA is also different

from that for the atherosclerotic diseases such as CAD; IA is more frequent in females than in males, and the peak incidence is also at a younger age than for CAD (Schievink, W.I., *N Engl J Med* **336:**28-40 (1997)). The effect of rs10757278 allele G on IA is thus not mediated through CAD.

When genotype-specific effects were studied based on data from all seven AAA and the three IA sample sets the ORs for heterozygous and homozygous carriers of the risk allele G were estimated to be 1.36 and 1.74, respectively for AAA and 1.38 and 1.72, respectively, for IA (**Table 34**). Assuming a population frequency of 47.5% for the G allele, the corresponding population attributable risk is about 26% for both AAA and IA. It is noted that rs10757278 allele G is the first common sequence variant described that affects the risk of IA or AAA.

The prevalence of AAA (defined as > 3 cm aortic diameter) has been reported to be 4.3% and 1.0% in men and women over 50 years of age, respectively (Lederle, F.A., *et al, J Vasc Surg* **34:**122-6 (2001); Lederle, F.A:, *et al. Arch Intern Med* **160:**1425-30 (2000)), and 2-5% of the general population have IA (Brisman, J.L, *et al., N Engl J Med* **355:**928-39 (2006)). Both AAA and IA represent a degenerative process of the arteries leading to their enlargement that is usually asymptomatic with natural history culminating in either a therapeutic intervention or rupture. Rupture of IA leads to subarachnoid haemorrhage, and rupture of both IA and AAA have high morbidity and mortality (Brisman, J.L, *et al., N Engl J Med* **355:**928-39 (2006); Thompson, R.w. *Cardiovasc Surg* **10:** 389-94 (2002)). In the case of AAA the rupture risk increases with the growth rate as well as the size of the aneurysm. While patients from the UK study group included only those with small asymptomatic aneurysms (aortas <5.5cm diameter), the other study groups included mainly patients undergoing aneurysm repair and are therefore likely to be biased towards larger and symptomatic aneurysms (aortas >5.5 cm diameter). Despite recruitment differences the ORs were similar, suggesting that the variant does not confer direct risk of the growth of aortic aneurysms. This concept was further investigated in the sample set from UK, where the subjects with small asymptomatic AAA had been followed with sequential aortic aneurysm size measurements (Eriksson, P. *et al., Br J Surg* **92:**1372-6 (2005)). As shown in **Table 35**, there is no evidence of an association between rs10757278 allele G and either aneurysm growth or rupture. Rather there is some indication that allele G is correlated with slow growth. The difference in average growth rates between the homozygous GG group and the heterozygous AG group is -0.46 mm/year and is nominally significant (P = 0.05), an observation that warrants further investigation. If confirmed, this inverse association would echo previous findings where slower aneurysm growth rates were observed in patients with low ankle/brachial pressure index, a marker of generalised atherosclerosis (Brady A.R., *et al., Circulation* **110:**16-21 (2004)). These data suggest that the sequence variant leads to increased susceptibility of developing aneurysm rather than increasing the risk of rapid aneurysm progression.

The effect of rs10757278 allele G on the risk of PAD and LAA/cardiogenic stroke appeared to be weaker than that for AAA, IA and CAD (**Table 32**). In the Icelandic samples there was no

difference in the frequency of rs10757278 allele G between the AAA and IA cases, however the frequency was lower in both the PAD and LAA/cardiogenic stroke cases than in the combined group of AAA and IA cases (P=0.012 and P= 0.052, respectively). Furthermore, after excluding PAD and LAA/cardiogenic stroke subjects with known CAD from the analysis the effect was reduced even further, particularly for PAD (**Table 33**).

We tested the association of several variants in the LD block C09 region with MI in African American samples. As shown by the results presented in **Table 36**, the effect in African Americans is comparable in magnitude, as measured by Relative Risk, to the effect in Caucasian samples. The lack of nominal statistical significance of the association (p-value less than 0.05) in the African American samples for many of the markers is due to the relatively small samples size. Most importantly, the association observed in African Americans, together with reported association in Asian samples originating from Japan and Korea with comparable risk to that determined for Caucasians (see *Arterioscler Thromb Vasc Biol.* 2008 Feb;**28(2)**:360-5. Epub 2007 Nov 29, and *J Hum Genet*. Epub 2008 Feb 9) shows that the genetic effects originally discovered in Caucasian samples from Iceland manifests itself in other human populations. The effect therefore has implications for cardiovascular disease in all major human populations..

These data demonstrate that rs10757278 allele G has less effect on the atherosclerotic diseases, PAD and LAA/cardiogenic stroke, than on CAD. In contrast, the effect on the two aneurysmal diseases, AAA with weaker association to atherosclerosis and IA with no such relationship, was comparable to that on CAD, suggesting that the variant plays a role in a pathophysiological component common to these arterial phenotypes. This may involve abnormal vascular remodelling and/or repair which has been identified as a key in the pathogenesis CAD, AAA and IA (Chatzidisis, Y.S. *et al., J Am Coll Cardiol* **49:**2379-93 (2007); Hashimoto, T. *et al., Neurol Res* **8:**372-80 (2006); Moore, J.E., Jr., *et al, Atherosclerosis* **110:**225-40 (2006)). The sequence variant rs10757278 allele G on chromosome 9p21, and/or variants in linkage disequilibrium with rs10757278, may thus function as a genetic determinant of the tissue response to unfavourable conditions that prevail in the lower abdominal aorta, in the circle of Willis where IAs occur, and in those regions of the coronary tree that are prone to develop unstable, rupture prone atherosclerotic plaques.

TABLES

Table 1: *a*) Results for the 12 SNPs in LD block C09 that showed nominally significant association to myocardial infarction in the Icelandic discovery cohort (Cohort A) of 1570 MI cases and 7088 controls. Shown is the frequency of the risk allele in cases and controls, the corresponding relative risk (RR), the unadjusted *P*-value and the *P*-value after adjusting for the relatedness of cases and controls. Also included are results for a test of association to MI for the same SNPs conditioned on the association of the SNP rs1333040. *b*) Association to MI for the different alleles of the micro-satellites D9S1814 and D9S1870 in Cohort A. *c*) Association to MI of the at-risk allele 0 for D9S1814 and of the composite at-risk allele X of D9S1870 in Cohort A. The composite allele X includes alleles -6, -4, -2 and 0 of D9S1870.

				Frequency					Conditioned on rs1333040		
	SNP	Allele	Position	Cases	Controls	RR	*P*	*P**	RR**	*P***	*r2* ***
a)	rs7041637	A	21951866	0.251	0.232	1.104	0.031	0.044	1.003	0.96	0.12
	rs2811712	A	21988035	0.890	0.874	1.171	0.010	0.016	1.146	0.040	0.05
	rs3218018	A	21988139	0.894	0.878	1.174	0.010	0.016	1.149	0.038	0.03
	rs3217992	A	21993223	0.376	0.344	1.149	0.00071	0.0015	1.015	0.80	0.28
	rs2069426	C	21996273	0.895	0.880	1.159	0.019	0.028	1.134	0.063	0.02
	rs2069422	A	21998026	0.891	0.874	1.174	0.0092	0.015	1.149	0.037	0.05
	rs1333034	A	22034122	0.890	0.874	1.168	0.012	0.018	1.143	0.043	0.05
	rs1011970	G	22052134	0.801	0.772	1.189	0.00034	0.00081	1.114	0.050	0.13
	rs10116277	T	22071397	0.463	0.419	1.195	7.05×10^{-6}	2.63×10^{-5}	1.054	0.52	0.67
	rs1333040	T	22073404	0.538	0.491	1.209	1.56×10^{-6}	6.98×10^{-6}	NA	NA	NA
	rs2383207	G	22105959	0.502	0.456	1.204	2.64×10^{-6}	1.11×10^{-5}	1.102	0.15	0.57
	rs1333050	T	22115913	0.694	0.672	1.107	0.017	0.025	1.005	0.92	0.27
b)	D9S1814	-2	22078225	0.019	0.019	1.049	0.79	0.81	-	-	-
	D9S1814	0	22078225	0.500	0.451	1.217	1.35×10^{-6}	6.17×10^{-6}	-	-	-
	D9S1814	2	22078225	0.359	0.385	0.895	0.015	0.023	-	-	-
	D9S1814	4	22078225	0.112	0.149	0.720	4.00×10^{-6}	1.60×10^{-5}	-	-	-
	D9S1870	-6	22093010	0.019	0.020	0.961	0.79	0.81	-	-	-
	D9S1870	-4	22093010	0.039	0.029	1.352	0.0090	0.015	-	-	-
	D9S1870	-2	22093010	0.376	0.339	1.177	0.00011	0.00031	-	-	-
	D9S1870	0	22093010	0.044	0.043	1.038	0.72	0.74	-	-	-
	D9S1870	2	22093010	0.072	0.074	0.974	0.74	0.76	-	-	-
	D9S1870	4	22093010	0.166	0.200	0.797	2.96×10^{-5}	9.31×10^{-5}	-	-	-
	D9S1870	6	22093010	0.104	0.114	0.905	0.14	0.17	-	-	-
	D9S1870	8	22093010	0.079	0.080	0.989	0.89	0.89	-	-	-
	D9S1870	10	22093010	0.072	0.077	0.931	0.37	0.40	-	-	-
	D9S1870	12	22093010	0.021	0.018	1.152	0.34	0.37	-	-	-
c)	D9S1814	0	22078225	0.500	0.451	1.217	1.35×10^{-6}	6.17×10^{-6}	1.118	0.099	0.53
	D9S1870	X	22093010	0.486	0.438	1.211	2.52×10^{-6}	1.07×10^{-5}	1.113	0.081	0.55

* P-value adjuste for relatedness of cases and controls.

** P-value adjusted for relatedness and conditioned on the association to rs133304, and the corresponding relative risk.

*** Pair-wise correlation r2 between the SNP and the at-risk variant rs1333040

Table 2: Pair-wise correlation among the 5 markers, 3 SNPs and 2 micro-satellites, that show strongest association to myocardial infarction in LD block C09, based on the HapMap v19 CEU dataset. In the upper right corner are shown values for the correlation coefficient r^2, while in the lower left corner are values for D'.

		r^2				
	Marker	**rs10116277**	**rs1333040**	**D9S1814**	**D9S1870**	**rs2383207**
	rs10116277	-	0.667	0.806	0.839	0.905
	rs1333040	1.000	-	0.528	0.550	0.569
D'	**D9S1814**	0.964	0.829	-	0.651	0.743
	D9S1870	1.000	1.000	0.954	-	0.779
	rs2383207	1.000	0.879	0.893	1.000	-

Table 3: List of all SNPs (from HapMap v19 CEU dataset) in LD block C09 that are correlated, with correlation coefficient $r^2 \geq 0.2$, with at least one of the five markers (rs10116277, rs1333040, D9S1814, D9S1870 or rs2383207). For each SNP shown the table includes the position (in Build 34 and in SEQ ID NO:94) and the correlation coefficient r^2 to each of the five at-risk markers.

SNP	**Position[a]**	**Position[b]**	**D9S1814**	**D9S1870**	**rs1333040**	**rs2383207**	**rs10116277**
rs7041637	21951866	31720	0.18	0.12	0.20	0.21	0.15
rs3218020	21987872	67726	0.46	0.33	0.37	0.55	0.41
rs3217992	21993223	73077	0.43	0.28	0.34	0.53	0.38
rs1063192	21993367	73221	0.22	0.09	0.27	0.28	0.16
rs2069418	21999698	79552	0.19	0.06	0.25	0.25	0.14
rs2069416	22000004	79858	0.41	0.24	0.32	0.50	0.36
rs573687	22001642	81496	0.23	0.07	0.28	0.20	0.17
rs545226	22002422	82276	0.31	0.18	0.24	0.40	0.27
rs10811640	22003411	83265	0.21	0.09	0.29	0.29	0.19
rs10811641	22004137	83991	0.41	0.27	0.33	0.50	0.36
rs2106120	22007101	86955	0.23	0.10	0.31	0.32	0.21
rs2106119	22007550	87404	0.23	0.10	0.31	0.32	0.21
rs643319	22007836	87690	0.22	0.09	0.29	0.32	0.20
rs7044859	22008781	88635	0.23	0.10	0.31	0.32	0.21
rs523096	22009129	88983	0.17	0.05	0.22	0.23	0.12
rs10757264	22009732	89586	0.21	0.09	0.29	0.30	0.19
rs10965212	22013795	93649	0.29	0.14	0.38	0.40	0.27
rs1292137	22014023	93877	0.24	0.09	0.31	0.34	0.21
rs1292136	22014351	94205	0.30	0.16	0.37	0.41	0.28
rs10811644	22015067	94921	0.25	0.11	0.33	0.34	0.22
rs7035484	22015240	95094	0.24	0.10	0.32	0.33	0.22
rs10738604	22015493	95347	0.49	0.35	0.41	0.60	0.44
rs615552	22016077	95931	0.17	0.06	0.21	0.20	0.10
rs543830	22016639	96493	0.23	0.08	0.28	0.31	0.17
rs1591136	22016834	96688	0.29	0.14	0.38	0.40	0.27

SNP	Position[a]	Position[b]	D9S1814	D9S1870	rs1333040	rs2383207	rs10116277
rs7049105	22018801	98655	0.29	0.14	0.38	0.40	0.27
rs679038	22019080	98934	0.23	0.08	0.28	0.31	0.17
rs10965215	22019445	99299	0.27	0.13	0.35	0.38	0.25
rs564398	22019547	99401	0.21	0.07	0.26	0.29	0.15
rs7865618	22021005	100859	0.24	0.10	0.29	0.31	0.18
rs10115049	22022119	101973	0.29	0.14	0.38	0.40	0.27
rs634537	22022152	102006	0.24	0.08	0.29	0.33	0.20
rs2157719	22023366	103220	0.26	0.11	0.32	0.33	0.19
rs2151280	22024719	104573	0.27	0.15	0.36	0.38	0.25
rs1008878	22026112	105966	0.24	0.12	0.30	0.32	0.18
rs1556515	22026367	106221	0.26	0.13	0.32	0.35	0.20
rs1333037	22030765	110619	0.28	0.12	0.34	0.36	0.21
rs1360590	22031443	111297	0.31	0.15	0.40	0.43	0.29
rs1412829	22033926	113780	0.23	0.08	0.28	0.31	0.17
rs1360589	22035317	115171	0.30	0.12	0.36	0.40	0.23
rs7028570	22038683	118537	0.31	0.14	0.40	0.43	0.29
rs944801	22041670	121524	0.30	0.12	0.36	0.40	0.23
rs10965219	22043687	123541	0.30	0.13	0.40	0.48	0.27
rs7030641	22044040	123894	0.30	0.12	0.36	0.40	0.23
rs10120688	22046499	126353	0.38	0.18	0.47	0.51	0.35
rs2184061	22051562	131416	0.50	0.24	0.59	0.44	0.40
rs1537378	22051614	131468	0.50	0.24	0.59	0.44	0.40
rs8181050	22054391	134245	0.47	0.22	0.55	0.40	0.38
rs8181047	22054465	134319	0.33	0.49	0.38	0.28	0.21
rs10811647	22055002	134856	0.70	0.49	0.59	0.83	0.62
rs1333039	22055657	135511	0.47	0.22	0.55	0.40	0.38
rs10965224	22057276	137130	0.45	0.21	0.53	0.40	0.35
rs10811650	22057593	137447	0.70	0.49	0.59	0.83	0.63
rs10811651	22057830	137684	0.48	0.23	0.56	0.41	0.39
rs4977756	22058652	138506	0.47	0.22	0.55	0.40	0.38
rs10757269	22062264	142118	0.94	0.66	0.80	0.78	0.84
rs9632884	22062301	142155	0.93	0.65	0.80	0.78	0.84
rs1412832	22067543	147397	0.32	0.48	0.37	0.28	0.22
rs10116277	22071397	151251	-	0.67	0.81	0.84	0.90
rs10965227	22071796	151650	0.28	0.02	0.13	0.18	0.31
rs6475606	22071850	151704	1.00	0.67	0.81	0.84	0.90
rs1333040	22073404	153258	0.67	-	0.53	0.55	0.57
rs1537370	22074310	154164	1.00	0.67	0.81	0.84	0.90
rs7857345	22077473	157327	0.36	0.55	0.43	0.32	0.26
rs10738607	22078094	157948	1.00	0.69	0.81	0.84	0.90
rs10757272	22078260	158114	1.00	0.67	0.81	0.84	0.90
rs4977574	22088574	168428	1.00	0.67	0.81	0.84	0.90
rs2891168	22088619	168473	1.00	0.67	0.81	0.84	0.90
rs1537371	22089568	169422	1.00	0.67	0.81	0.84	0.90
rs1556516	22090176	170030	1.00	0.67	0.81	0.84	0.90
rs6475608	22091702	171556	0.36	0.54	0.42	0.31	0.25
rs7859727	22092165	172019	1.00	0.66	0.80	0.84	0.90
rs1537373	22093341	173195	1.00	0.67	0.81	0.84	0.90
rs1333042	22093813	173667	0.97	0.63	0.81	0.81	0.94
rs7859362	22095927	175781	0.90	0.57	0.74	0.78	1.00
rs1333043	22096731	176585	0.94	0.60	0.74	0.81	0.97
rs1412834	22100131	179985	0.90	0.57	0.74	0.78	1.00
rs7341786	22102241	182095	0.88	0.54	0.71	0.75	0.97
rs10511701	22102599	182453	0.88	0.54	0.71	0.75	0.97

SNP	Position[a]	Position[b]	D9S1814	D9S1870	rs1333040	rs2383207	rs10116277
rs10733376	22104469	184323	0.90	0.57	0.74	0.78	1.00
rs10738609	22104495	184349	0.90	0.57	0.74	0.78	1.00
rs2383206	22105026	184880	0.90	0.57	0.74	0.78	1.00
rs944797	22105286	185140	0.90	0.57	0.74	0.78	1.00
rs1004638	22105589	185443	0.90	0.57	0.74	0.78	1.00
rs2383207	22105959	185813	0.90	0.57	0.74	0.78	-
rs1537374	22106046	185900	0.90	0.57	0.74	0.78	1.00
rs1537375	22106071	185925	0.90	0.57	0.74	0.78	1.00
rs1333045	22109195	189049	0.69	0.37	0.52	0.60	0.65
rs10738610	22113766	193620	0.94	0.60	0.77	0.78	0.97
rs1333046	22114123	193977	0.94	0.60	0.77	0.78	0.97
rs10757278	22114477	194331	0.90	0.57	0.74	0.75	0.87
rs1333047	22114504	194358	0.90	0.59	0.72	0.74	0.88
rs4977575	22114744	194598	0.90	0.59	0.72	0.74	0.88
rs1333048	22115347	195201	0.97	0.63	0.77	0.81	0.94
rs1333049	22115503	195357	0.90	0.59	0.72	0.74	0.88
rs1333050	22115913	195767	0.17	0.27	0.06	0.15	0.20

[a]Position with respect to NCBI Build 34, Build 35 or Build 36
[b]Position with respect to SEQ ID NO:94 (LD Block C09)

Table 4: *a*) Association of the composite at-risk allele X of the micro-satellite D9S1870 to myocardial infarction (MI) in a replication cohort (Cohort B) of 668 Icelandic MI cases and 58,543 controls and in the combined Icelandic cohort (Cohort A+B). Included in the table is the number of cases *n* and controls *m*, the allele frequency in cases and controls, the corresponding P value adjusted for relatedness of the study individuals, the relative risk (RR) with 95% confidence interval (CI), and the population attributed risk (PAR). Known CAD or CVD cases have been excluded from the control cohort. *b*) Association of allele X to MI in three distinct US replication cohorts. All cases and controls in those three cohorts are of Caucasian origin. *c*) Combined association results, using a Mantel-Haenzel model, for the composite at-risk allele X in all replication cohorts combined, and in all cohorts combined, respectively.

	Frequency					
Cohort (n/m)	**Cases**	**Controls**	***P****	**RR**	**CI**	**PAR**
***a*)** Iceland						
Cohort B (668/58643)	0.501	0.441	1.39×10^{-5}	1.274	[1.142,1.420]	0.203
Cohort A+B (2238/65731)	0.491	0.441	8.91×10^{-10}	1.221	[1.145,1.301]	0.169
***b*)** US replication cohorts						
Cleveland (549/606)	0.501	0.478	0.27	1.097	[0.932,1.292]	0.087
UPenn (580/404)	0.539	0.483	0.014	1.252	[1.046,1.499]	0.205
Emory (400/477)	0.548	0.482	0.0064	1.299	[1.076,1.569]	0.236
***c*)** Combined**						
Replication cohorts (2197/60130)			2.65×10^{-8}	1.231	[1.144,1.324]	0.183
All cohorts (3767/67218)			1.44×10^{-12}	1.214	[1.151,1.281]	0.171

* P-values for the Icelandic cohorts are adjusted for relatedness using simulations (see methods)
** Mantel-Haenzel model was used to combine the results from the different cohorts.

Table 5: *a*) Association of the composite at-risk allele X of D9S1870 in Icelandic cases with various cardiovascular disease phenotypes. Included in the table are association results for individuals who have undergone PTCA or CABG, for the combined MI and CAD (PTCA and CABG) cohort, individuals with PAD or that have had infarct or TIA, and stroke patients diagnosed with LVD. In all tests the same set of 65,731 Icelandic controls without known CAD complications is used. *b*) Association of allele X in Iceland to the same coronary artery disease phenotypes as in *a*) after excluding all known cases of MI from the list of cases. C) Association of allele X to PTCA, CABG and a combined phenotype MI and other CAD, in the Emory cohort.

	Frequency						
Phenotype (*n*)	**Cases**	**Controls**	**RR**	**CI**	***P***	***P****	**PAR**
***a*)** Iceland							
PTCA (1791)	0.489	0.441	1.211	[1.129,1.298]	2.69E-08	8.60E-08	0.163
CABG (642)	0.514	0.441	1.338	[1.189,1.506]	3.38E-07	1.31E-06	0.243
MI + other CAD (3513)	0.488	0.441	1.209	[1.148,1.273]	2.57E-14	6.24E-13	0.162
PAD(1661)	0.477	0.441	1.154	[1.074,1.239]	5.36E-05	8.58E-05	0.123
Stroke, Infarct + TIA (1678)	0.448	0.441	1.027	[0.958,1.102]	0.45	0.45	0.023
Stroke, LVD (197)	0.469	0.441	1.120	[0.915,1.370]	0.27	0.27	0.098
***b*)** Iceland, excl. MI cases							
PTCA (941)	0.483	0.441	1.183	[1.077,1.300]	0.00033	0.00045	0.144
CABG (221)	0.501	0.441	1.270	[1.044,1.545]	0.013	0.017	0.202
PAD(1322)	0.464	0.441	1.096	[1.012,1.186]	0.021	0.025	0.079
Stroke, Infarct + TIA (1390)	0.450	0.441	1.037	[0.960,1.119]	0.35	0.36	0.032
Stroke, LVD (153)	0.480	0.441	1.172	[0.934,1.470]	0.17	0.17	0.136
***c*)** Emory							
PTCA (141)	0.567	0.482	1.408	[1.079,1.839]	0.012	NA	0.302
CABG (112)	0.549	0.482	1.308	[0.977,1.751]	0.071	NA	0.242
MI + other CAD (476)	0.551	0.482	1.320	[1.103,1.581]	0.0025	NA	0.250

* P-value adjusted for relatedness using simulations.

Table 6: Association of the composite at-risk allele X of D9S1870 to early-onset MI or CAD *a*) in Iceland, *b*) in the three US replication cohorts, *c*) in all early-onset MI cohorts combined. Early-onset MI is defined as a MI event before the age 50 for males and before the age of 60 for females.

	Frequency						
Phenotype (*n*)	**Cases**	**Controls**	**RR**	**CI**	***P***	***P****	**PAR**
a) Iceland							
Early onset MI (646/65731)	0.514	0.441	1.339	[1.193,1.503]	3.63E-07	7.51E-07	0.243
Early onset CAD (320/65731)	0.532	0.441	1.439	[1.227,1.687]	6.12E-06	7.40E-06	0.298
Early onset MI+CAD (877/65731)	0.511	0.441	1.325	[1.200,1.464]	9.87E-09	2.70E-08	0.235
b) US cohorts							
Emory EO MI (222/477)	0.559	0.482	1.359	[1.084,1.703]	0.0078	NA	0.273
Cleveland EO MI (183/606)	0.514	0.478	1.155	[0.914,1.459]	0.2281	NA	0.133
Upenn EO MI (275/490)	0.569	0.473	1.469	[1.191,1.811]	0.00033	NA	0.330
c) All cohorts combined							
Early onset MI (1326/67304)	-	-	1.331	[1.223,1.449]	3.96E-11	NA	0.247

* P-value adjusted for relatedness using simulations.

Table 7: Model-free estimates of the genotype relative risks (GRR) of the composite allele X of D9S1870. Included is the risk for heterozygous carriers (0X) and homozygous carriers (XX) compared with risk for non-carriers (00) and the corresponding 95% confidence intervals (CI). Results are shown for *a*) all MI cases versus controls and *b*) for early-onset MI cases versus controls for the whole Icelandic cohort, the three US replication cohorts and for all the cohorts combined using a Mantel-Haenzel model.

	Genotype relative risk (GRR)			
Phenotype (n/m)	**0X**	**CI**	**XX**	**CI**
***a*)** All MI cases				
Iceland (2238/65731)	1.173	[1.060,1.298]	1.507	[1.338,1.697]
Cleveland(549/606)	1.077	[0.804,1.443]	1.204	[0.860,1.686]
UPenn (679/490)	1.443	[1.074,1.938]	1.728	[1.228,2.431]
Emory (400/477)	1.213	[0.836,1.760]	1.669	[1.144,2.435]
Combined (3866/67304)	1.204	[1.094,1.324]	1.507	[1.360,1.670]
***b*)** Early onset MI cases				
Iceland (641/65731)	1.191	[0.982,1.444]	1.803	[1.453,2.237]
Cleveland (183/606)	1.432	[0.932,2.201]	1.867	[1.176,2.964]
UPenn (275/490)	1.692	[1.134,2.524]	2.237	[1.440,3.476]
Emory (222/477)	1.069	[0.567,2.017]	1.328	[0.827,2.133]
Combined (1321/67304)	1.314	[1.105,1.562]	1.790	[1.517,2.113]

Table 8: *a*) Association of the risk alleles of the markers rs10116277, rs1333040 and rs2383207 to age of onset of MI in the Icelandic Cohort A. The analysis is done using multiple regression where age of onset of MI is taken as the response variable and the number of copies of X is used as explanatory variable. The analysis is adjusted for gender by include sex of the individuals as an explanatory variable. Included in the table is the effect size, the standard error of the mean (SE) and the corresponding P-value. *b*) Association of the composite at-risk allele X of D9S1870 to age-of-onset of MI in the combined Icelandic MI patient cohort, and for the three US replication cohorts. Results for the four cohorts are combined weighting the contribution from each cohort proportional to the inverse of the standard error. In all tests only cases with known age-of-onset are included.

Cohort/Marker (*n*)	Effect	SE	P
a) SNPs in Iceland A			
rs10116277 (1293)	-0.94	0.32	0.0039
rs1333040 (1293)	-1.01	0.32	0.0018
rs2383207 (1293)	-0.79	0.32	0.015
b) D9S1870 allele X			
Iceland (2127)	-0.93	0.29	0.0016
Cleveland (424)	-0.47	0.79	0.55
UPenn (646)	-1.19	0.69	0.083
Emory (403)	-1.38	0.87	0.11
Combined (3600)	-0.95	0.25	0.000099

Table 9: *a*) Association of the composite risk allele X of D9S1870 to either mild (<50%) or severe (≥50%) in-stent restenosis in coronary artery disease patients that have undergone PTCA compared to controls without known history of CAD. *b*) Comparision of the frequency of the risk variant X in patients with severe in-stent restenosis to patients with mild in-stent restenosis.

	Frequency				
Phenotype (*n,m*)	Cases	Controls	RR	CI	*P*
a) Restenosis patients vs controls					
Mild restenosis, < 50% (323/65731)	0.492	0.441	1.228	[1.050,1.436]	0.010
Severe restenosis (≥50% (193/65731)	0.509	0.441	1.313	[1.071,1.608]	0.0086
b) Severe vs mild restenosis (193/323)	0.509	0.492	1.067	[0.827,1.376]	0.62

Table 10.

A. SNP markers within LD Block C09 (Between 21,920,147 and 22,149,982 on C09; NCBI Build 34, Build 35 and Build 36 (SEQ ID NO:94)).

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs7864029	C/G	21920147	+	1
rs2518714	A/C	21920303	+	157
rs10965187	A/G	21920505	+	359
rs2811715	G/T	21920571	+	425
rs2518715	C/T	21920700	+	554
rs12002708	A/C	21920762	+	616
rs10811639	A/G	21920803	+	657
rs11998828	C/T	21920812	+	666
rs7868374	A/G	21921609	+	1463
rs4977570	G/T	21921642	+	1496
rs7390564	G/T	21921642	+	1496
rs7869004	G/T	21921896	+	1750
rs974679	A/G	21922245	+	2099
rs2891010	A/G	21922261	+	2115
rs12003714	A/G	21922366	+	2220
rs28756321	C/G	21923029	+	2883
rs12000395	A/G	21923096	+	2950
rs2188126	C/G	21923125	+	2979
rs12003027	C/T	21924442	+	4296
rs7023582	A/C	21924684	+	4538
rs28613732	A/G	21925807	+	5661
rs10965188	C/G	21925874	+	5728
rs12552975	C/T	21926059	+	5913
rs7875199	C/T	21926381	+	6235
rs12236992	A/C	21927299	+	7153
rs7021848	C/T	21927351	+	7205
rs7036999	A/G	21927875	+	7729
rs12345808	A/G	21928367	+	8221
rs35148759	C/T	21929306	+	9160
rs7872310	A/G	21929387	+	9241
rs2811716	C/T	21930017	+	9871
rs10965189	A/C	21930588	+	10442
rs4308829	C/T	21931196	+	11050
rs10965190	A/G	21931385	+	11239
rs10965191	A/G	21932535	+	12389
rs12345373	A/G	21933199	+	13053
rs4977750	A/C	21934317	+	14171
rs10965193	A/G	21934544	+	14398
rs4246843	C/T	21934627	+	14481
rs4382559	A/G	21934818	+	14672
rs4360371	A/C	21935562	+	15416
rs16938595	A/T	21935605	+	15459
rs10965194	A/G	21936214	+	16068
rs2811717	C/T	21936322	+	16176
rs6475598	C/T	21936322	+	16176
rs2811718	A/T	21936891	+	16745
rs10965196	C/G	21937472	+	17326
rs2811719	C/G	21937472	+	17326
rs2811720	C/G	21937957	+	17811
rs6475599	C/G	21937957	+	17811
rs11521166	C/T	21938376	+	18230
rs2518718	A/T	21938489	+	18343
rs10965197	C/T	21938666	+	18520
rs2106117	G/T	21939527	+	19381
rs2106118	A/T	21939528	+	19382
rs2427852	A/T	21940432	+	20286
rs2157716	C/T	21940446	+	20300

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs2263145	C/T	21940452	+	20306
rs2427853	C/T	21940472	+	20326
rs2382894	C/G	21940523	+	20377
rs2427854	A/T	21940721	+	20575
rs7047211	A/T	21940722	+	20576
rs11532909	A/T	21940723	+	20577
rs13297154	A/T	21940725	+	20579
rs13298455	A/T	21940741	+	20595
rs2891092	A/G	21940782	+	20636
rs6415737	A/G	21940782	+	20636
rs4503179	A/G	21940879	+	20733
rs12349081	A/C	21941902	+	21756
rs2518722	C/T	21942926	+	22780
rs10757260	A/G	21943137	+	22991
rs28369665	C/T	21944487	+	24341
rs10965199	C/T	21944653	+	24507
rs10757261	A/G	21944953	+	24807
rs12335859	A/G	21945460	+	25314
rs12335941	A/G	21945669	+	25523
rs2027938	A/G	21946078	+	25932
rs9657608	C/T	21946230	+	26084
rs2518716	A/G	21946470	+	26324
rs2027939	A/G	21946492	+	26346
rs7869996	C/T	21946914	+	26768
rs10965200	C/G	21948142	+	27996
rs717326	A/G	21948524	-	28378
rs35116241	A/C	21949093	+	28947
rs2518717	C/T	21949751	+	29605
rs3948753	A/G	21949860	+	29714
rs2106115	C/T	21949900	+	29754
rs10965201	C/G	21949950	+	29804
rs2106116	C/T	21949966	+	29820
rs10965202	A/G	21951180	+	31034
rs1985742	A/T	21951227	+	31081
rs7041637	A/C	21951866	+	31720
rs2263146	C/T	21953048	+	32902
rs35732310	A/C	21953841	+	33695
rs4518744	A/C	21954015	+	33869
rs2188127	C/G	21955232	+	35086
rs3731257	C/T	21956221	-	36075
rs3731256	C/T	21956617	-	36471
rs11793581	G/T	21957446	+	37300
rs28695347	A/C	21957706	+	37560
rs3731255	C/G	21957855	-	37709
rs3731253	C/G	21957952	-	37806
rs3088440	A/G	21958159	+	38013
rs11515	C/G	21958199	-	38053
rs34011899	A/C	21958712	+	38566
rs2255962	A/G	21959827	+	39681
rs2518719	A/G	21960427	+	40281
rs3731249	A/G	21960916	-	40770
rs6413464	G/T	21960979	-	40833
rs34170727	C/T	21960988	-	40842
rs6413463	A/T	21960989	-	40843
rs35741010	A/G	21961054	-	40908
rs34886500	C/T	21961063	-	40917
rs4987127	A/G	21961085	-	40939
rs11552822	G/T	21961108	-	40962
rs34968276	A/C	21961109	-	40963
rs11552823	C/T	21961116	-	40970
rs3731247	C/T	21961352	-	41206

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs12377672	C/T	21961418	+	41272
rs3731246	C/G	21961989	-	41843
rs3731245	A/G	21962445	-	42299
rs3731244	A/G	21962813	-	42667
rs3731243	A/G	21963050	-	42904
rs2811708	G/T	21963422	+	43276
rs3731241	A/G	21963767	-	43621
rs13288666	C/T	21963857	+	43711
rs3731240	A/G	21964131	-	43985
rs3731239	C/T	21964218	-	44072
rs3814960	A/C/G/T	21965017	+	44871
rs3731238	A/G	21965561	-	45415
rs3731237	A/G	21965728	-	45582
rs3731236	C/T	21966976	-	46830
rs3731235	G/T	21967450	-	47304
rs12350633	C/G	21967553	+	47407
rs3731234	A/C	21967579	-	47433
rs3731233	A/T	21968358	-	48212
rs3731232	C/G	21968443	-	48297
rs2518720	C/T	21968979	+	48833
rs3731230	C/T	21969163	-	49017
rs2518721	A/G	21969204	+	49058
rs3731229	A/C	21969497	-	49351
rs3731228	A/G	21969602	-	49456
rs2811709	A/G	21970151	+	50005
rs3731227	C/T	21970744	-	50598
rs3731226	C/T	21970792	-	50646
rs13297747	C/G	21970941	+	50795
rs36170221	C/G	21970941	+	50795
rs36153543	C/T	21970944	+	50798
rs7874405	C/T	21970944	+	50798
rs13302595	C/T	21971034	+	50888
rs13301751	A/G	21971039	+	50893
rs13302611	C/T	21971068	+	50922
rs13302761	C/T	21971100	+	50954
rs13302792	C/T	21971168	+	51022
rs3731225	A/G	21971351	-	51205
rs3731224	C/T	21971411	-	51265
rs4074785	A/G	21971583	+	51437
rs3731223	A/G	21973834	-	53688
rs3731222	A/G	21973914	-	53768
rs3731221	A/G	21974010	-	53864
rs3731220	A/G	21974019	-	53873
rs3731219	A/G	21974086	-	53940
rs3731218	C/T	21974331	-	54185
rs3731217	G/T	21974661	-	54515
rs3731216	G/T	21975576	-	55430
rs3731215	C/T	21975771	-	55625
rs3731214	A/G	21975968	-	55822
rs3731213	A/G	21976218	-	56072
rs3731212	C/T	21976271	-	56125
rs3731211	A/T	21976847	-	56701
rs3731210	A/G	21976859	-	56713
rs3731208	C/T	21977155	-	57009
rs3731207	C/T	21977353	-	57207
rs12376353	C/G	21977433	+	57287
rs3731206	A/G	21977472	-	57326
rs3731205	A/G	21977522	-	57376
rs3731204	A/G	21977584	-	57438
rs10757262	A/T	21977874	+	57728
rs3731202	A/G	21978800	-	58654

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs3731201	A/G	21978896	-	58750
rs3731199	A/G	21979330	-	59184
rs3731198	A/G	21979477	-	59331
rs7036656	C/T	21980457	+	60311
rs7867492	C/T	21981016	+	60870
rs3731197	A/G	21981371	-	61225
rs3731196	A/G	21981652	-	61506
rs3731195	C/T	21981695	-	61549
rs3731194	C/G	21981752	-	61606
rs2811710	C/T	21981923	+	61777
rs3731192	G/T	21982274	-	62128
rs3731191	C/T	21983048	-	62902
rs2811711	C/T	21983964	+	63818
rs3731190	C/T	21984282	-	64136
rs2518723	A/G	21985882	-	65736
rs7860185	C/T	21986986	+	66840
rs3218024	A/G	21987437	-	67291
rs3218023	C/T	21987597	-	67451
rs3218022	A/G	21987723	-	67577
rs3218021	G/T	21987752	-	67606
rs3218020	C/T	21987872	-	67726
rs3218019	A/G	21987904	-	67758
rs2811712	A/G	21988035	+	67889
rs3218018	A/C	21988139	-	67993
rs3218016	A/G	21988273	-	68127
rs3218015	C/G	21988392	-	68246
rs3218013	C/T	21988556	-	68410
rs3218012	C/T	21988660	-	68514
rs3218011	A/G	21988676	-	68530
rs3218010	A/G	21988733	-	68587
rs3218009	C/G	21988757	-	68611
rs10965208	A/G	21988965	+	68819
rs2811713	A/G	21989328	+	69182
rs2811714	C/T	21989334	+	69188
rs3218007	A/G	21989800	-	69654
rs3218006	G/T	21989980	-	69834
rs3218005	A/G	21990247	-	70101
rs3218004	C/G	21990687	-	70541
rs3218003	C/G	21990770	-	70624
rs3218002	C/T	21990841	-	70695
rs3218001	C/G	21990959	-	70813
rs3218000	C/T	21991078	-	70932
rs3217999	C/T	21991572	-	71426
rs3217998	A/C	21991667	-	71521
rs3217997	A/G	21992316	-	72170
rs3217996	C/T	21992322	-	72176
rs3217994	A/G	21992864	-	72718
rs3217993	C/T	21993169	-	73023
rs3217992	A/G	21993223	-	73077
rs1063192	C/T	21993367	-	73221
rs3217990	A/C	21993521	-	73375
rs3217989	A/G	21993790	-	73644
rs3217988	A/G	21994082	-	73936
rs2285329	C/T	21994153	-	74007
rs10965209	A/G	21994669	+	74523
rs3217987	C/T	21995061	-	74915
rs11792943	A/G	21995123	+	74977
rs11792944	G/T	21995127	+	74981
rs3217986	A/C	21995330	-	75184
rs3217985	C/T	21995453	-	75307
rs3217984	C/G	21995493	-	75347

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs3217983	C/T	21995623	-	75477
rs3217982	C/T	21995647	-	75501
rs3217981	A/G	21996269	-	76123
rs2069426	A/C	21996273	-	76127
rs974336	A/G	21996348	-	76202
rs3217980	C/T	21996607	-	76461
rs2069425	C/T	21996793	-	76647
rs2285328	C/T	21996966	-	76820
rs2285327	A/G	21997048	-	76902
rs3217979	A/G	21997187	-	77041
rs3217978	G/T	21997330	-	77184
rs2069423	C/T	21997771	-	77625
rs2069422	A/C	21998026	-	77880
rs2069421	A/G	21998313	-	78167
rs3217976	A/C	21998439	-	78293
rs2069420	A/T	21998504	-	78358
rs2069419	A/C/T	21999337	-	79191
rs2069418	C/G	21999698	-	79552
rs3217974	A/C	21999908	-	79762
rs3217973	C/T	21999960	-	79814
rs2069416	A/C/T	22000004	-	79858
rs495490	C/T	22000412	-	80266
rs3808845	A/G	22000575	+	80429
rs3808846	A/G	22000946	+	80800
rs575427	C/T	22001477	-	81331
rs573687	C/T	22001642	-	81496
rs13298881	C/T	22002051	+	81905
rs16935753	A/G	22002229	+	82083
rs16935754	C/T	22002236	+	82090
rs545226	C/T	22002422	-	82276
rs7032979	C/T	22002457	+	82311
rs10811640	G/T	22003411	+	83265
rs10757263	C/T	22003805	+	83659
rs10811641	C/G	22004137	+	83991
rs7042051	C/T	22004758	+	84612
rs7027610	C/T	22004872	+	84726
rs7045307	C/T	22005057	+	84911
rs1101330	G/T	22005465	-	85319
rs13295358	A/C	22005465	+	85319
rs1101329	A/G	22005997	-	85851
rs1633381	A/G	22005997	-	85851
rs10217269	C/T	22006173	+	86027
rs10217281	A/G	22006617	+	86471
rs10965211	C/T	22006891	+	86745
rs28451206	C/G	22006921	+	86775
rs2157718	G/T	22007025	-	86879
rs2106120	A/C	22007101	-	86955
rs16905562	A/T	22007425	+	87279
rs2106119	C/T	22007550	-	87404
rs643319	G/T	22007836	-	87690
rs575237	C/T	22008108	-	87962
rs642323	C/T	22008121	-	87975
rs7044859	A/T	22008781	+	88635
rs523096	C/T	22009129	-	88983
rs518394	C/G	22009673	-	89527
rs10757264	A/G	22009732	+	89586
rs490005	C/T	22010493	-	90347
rs7858261	C/T	22010757	+	90611
rs34623146	A/G	22010822	+	90676
rs597816	A/G	22011172	-	91026
rs7048912	A/G	22011425	+	91279

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs568447	C/T	22011615	-	91469
rs567453	C/G	22011737	-	91591
rs7018665	A/G	22011819	+	91673
rs11515247	C/T	22012269	+	92123
rs581876	A/G	22012376	-	92230
rs7039304	A/T	22012786	+	92640
rs11789770	C/G	22012982	+	92836
rs10965212	A/T	22013795	+	93649
rs1292137	A/T	22014023	-	93877
rs504318	A/T	22014023	-	93877
rs1292136	A/G	22014351	-	94205
rs496892	A/G	22014351	-	94205
rs647188	A/C	22014965	-	94819
rs10811643	A/G	22014966	+	94820
rs10811644	A/T	22015067	+	94921
rs7035484	C/G	22015240	+	95094
rs10738604	A/G	22015493	+	95347
rs11791383	C/G	22015814	+	95668
rs615552	A/G	22016077	-	95931
rs1591137	C/T	22016483	-	96337
rs613312	C/T	22016594	-	96448
rs543830	A/T	22016639	-	96493
rs1591136	C/G	22016834	-	96688
rs10965214	A/T	22017274	+	97128
rs599452	C/T	22017402	-	97256
rs598664	A/G	22017551	-	97405
rs7049105	A/G	22018801	+	98655
rs679038	C/T	22019080	-	98934
rs10965215	A/G	22019445	+	99299
rs564398	A/G	22019547	-	99401
rs4977753	C/T	22020027	+	99881
rs662463	C/T	22020438	-	100292
rs7865618	A/G	22021005	+	100859
rs649436	C/T	22021085	-	100939
rs10115049	A/G	22022119	+	101973
rs634537	A/C	22022152	-	102006
rs2157719	A/G	22023366	-	103220
rs1759417	A/G	22023389	-	103243
rs1633383	A/G	22023532	-	103386
rs2151280	C/T	22024719	-	104573
rs1008878	G/T	22026112	+	105966
rs7029531	C/G	22026170	+	106024
rs1556515	A/G	22026367	-	106221
rs35975148	A/T	22027071	+	106925
rs17694478	C/G	22027171	+	107025
rs12376000	C/T	22029426	+	109280
rs1333037	A/G	22030765	-	110619
rs7028469	G/T	22031342	+	111196
rs1360590	A/G	22031443	-	111297
rs17694493	C/G	22031998	+	111852
rs12352425	A/G	22032086	+	111940
rs12686542	A/C	22032227	+	112081
rs10965216	C/T	22032879	+	112733
rs1412830	A/G	22033612	-	113466
rs1333036	A/G	22033819	-	113673
rs1412829	C/T	22033926	-	113780
rs1333035	C/T	22034059	-	113913
rs1333034	A/G	22034122	-	113976
rs10965217	A/T	22034317	+	114171
rs13290048	C/T	22034804	+	114658
rs28419335	A/G	22035035	+	114889

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs28621545	G/T	22035037	+	114891
rs1360589	A/G	22035317	-	115171
rs12338105	C/G	22035344	+	115198
rs1333033	A/G	22035653	-	115507
rs17834131	A/G	22036168	+	116022
rs12340618	A/G	22037050	+	116904
rs7851706	C/T	22037437	+	117291
rs12683931	A/G	22037916	+	117770
rs10120806	C/T	22037945	+	117799
rs7027950	C/T	22038391	+	118245
rs7028268	A/G	22038414	+	118268
rs7028570	A/G	22038683	+	118537
rs10757265	C/T	22038859	+	118713
rs10738605	C/G	22039130	+	118984
rs10757266	C/T	22039555	+	119409
rs10811645	A/G	22039656	+	119510
rs2151279	C/T	22039845	-	119699
rs6475603	A/G	22040612	+	120466
rs944799	A/G	22040613	+	120467
rs944800	A/G	22040898	+	120752
rs17694555	A/G	22041295	+	121149
rs944801	C/G	22041670	+	121524
rs6475604	C/T	22042734	+	122588
rs10757267	C/G	22042810	+	122664
rs4433231	G/T	22043244	+	123098
rs11790231	A/G	22043591	+	123445
rs7854869	A/C	22043651	+	123505
rs10965219	A/G	22043687	+	123541
rs7027048	A/G	22043709	+	123563
rs17756311	A/G	22043895	+	123749
rs7030641	C/T	22044040	+	123894
rs17694572	A/G	22044356	+	124210
rs7874604	C/T	22044690	+	124544
rs2383204	A/G	22045048	+	124902
rs7036489	A/G	22045992	+	125846
rs7039467	A/G	22046213	+	126067
rs7866660	A/T	22046233	+	126087
rs10965220	A/G	22046279	+	126133
rs7853090	C/T	22046295	+	126149
rs7866783	A/G	22046359	+	126213
rs10120688	A/G	22046499	+	126353
rs13292618	A/G	22047339	+	127193
rs10121501	C/G	22047390	+	127244
rs13299593	C/T	22048918	+	128772
rs7021816	A/C	22049277	+	129131
rs10757268	C/T	22049905	+	129759
rs2095144	C/T	22050136	-	129990
rs2383205	A/G	22050935	+	130789
rs2184061	G/T	22051562	-	131416
rs1537378	C/T	22051614	-	131468
rs4977754	A/C	22052012	+	131866
rs1011970	G/T	22052134	+	131988
rs10965221	G/T	22052999	+	132853
rs8181050	A/G	22054391	+	134245
rs8181047	A/G	22054465	+	134319
rs10811647	C/G	22055002	+	134856
rs1333038	A/G	22055572	+	135426
rs4144664	C/T	22055656	+	135510
rs1333039	C/G	22055657	+	135511
rs4977755	A/T	22056363	+	136217
rs28557075	A/G	22056572	+	136426

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs10965223	A/G	22057004	+	136858
rs10965224	A/T	22057276	+	137130
rs10811648	C/T	22057542	+	137396
rs10811649	C/T	22057554	+	137408
rs10811650	A/G	22057593	+	137447
rs10811651	A/G	22057830	+	137684
rs16905597	A/G	22058074	+	137928
rs1412831	A/G	22058646	+	138500
rs4977756	A/G	22058652	+	138506
rs16905599	A/G	22059144	+	138998
rs34871414	A/C	22059537	+	139391
rs7042970	A/G	22059580	+	139434
rs4451405	C/T	22061750	+	141604
rs4645630	A/G	22061751	+	141605
rs12555547	C/G	22062040	+	141894
rs10757269	A/G	22062264	+	142118
rs9632884	C/G	22062301	+	142155
rs9632885	A/G	22062638	+	142492
rs10757270	A/G	22062719	+	142573
rs17761197	C/T	22062730	+	142584
rs10965226	G/T	22063170	+	143024
rs16923583	A/T	22063334	+	143188
rs1855185	G/T	22063996	+	143850
rs7855162	C/T	22064793	+	144647
rs1831733	C/T	22066071	+	145925
rs1831734	C/T	22066208	+	146062
rs10757271	A/G	22066795	+	146649
rs10811652	A/C	22067085	+	146939
rs1412832	C/T	22067543	+	147397
rs7855660	C/T	22068305	+	148159
rs6475605	C/G	22069020	+	148874
rs16905613	A/G	22070363	+	150217
rs7858034	A/T	22070791	+	150645
rs12347950	A/G	22071128	+	150982
rs1412833	A/G	22071346	+	151200
rs10116277	G/T	22071397	+	151251
rs10965227	A/G	22071796	+	151650
rs6475606	C/T	22071850	+	151704
rs1547704	A/G	22072340	+	152194
rs1547705	A/C	22072375	+	152229
rs10965228	A/G	22072380	+	152234
rs7853953	A/C	22073017	+	152871
rs1333040	C/T	22073404	+	153258
rs1537370	C/T	22074310	+	154164
rs10122192	G/T	22074633	+	154487
rs1970112	C/T	22075598	+	155452
rs10120722	A/C	22076840	+	156694
rs16905635	C/T	22076883	+	156737
rs7857345	C/T	22077473	+	157327
rs10738606	A/T	22078090	+	157944
rs10738607	A/G	22078094	+	157948
rs10757272	C/T	22078260	+	158114
rs16905640	A/G	22078556	+	158410
rs13300638	G/T	22078937	+	158791
rs13284693	A/T	22079014	+	158868
rs12235973	A/T	22079193	+	159047
rs10757273	A/C	22080301	+	160155
rs10965230	C/T	22080416	+	160270
rs9644859	A/G	22080521	+	160375
rs7019916	C/T	22080683	+	160537
rs7020031	C/T	22080753	+	160607

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs7034707	C/T	22080811	+	160665
rs34597771	C/T	22081731	+	161585
rs34555767	G/T	22081924	+	161778
rs7866503	G/T	22081924	+	161778
rs7869527	A/G	22082097	+	161951
rs2210538	A/G	22082257	+	162111
rs7870178	A/G	22082551	+	162405
rs34184423	A/G	22082924	+	162778
rs34168773	A/T	22083299	+	163153
rs9722878	G/T	22083462	+	163316
rs7848875	A/G	22084281	+	164135
rs35537809	A/G	22084330	+	164184
rs4977757	A/G	22084330	+	164184
rs7388840	A/G	22084330	+	164184
rs10738608	A/C	22084796	+	164650
rs35869261	A/T	22085567	+	165421
rs35062160	G/T	22085730	+	165584
rs2891167	C/T	22085851	+	165705
rs10757274	A/G	22086055	+	165909
rs16905644	C/T	22087022	+	166876
rs6475607	A/G	22087693	+	167547
rs7037832	A/G	22088038	+	167892
rs1333041	C/G	22088374	+	168228
rs4977574	A/G	22088574	+	168428
rs2891168	A/G	22088619	+	168473
rs10965231	A/G	22088674	+	168528
rs11787814	A/G	22088683	+	168537
rs1537371	A/C	22089568	+	169422
rs7856476	A/T	22089940	+	169794
rs1556516	C/G	22090176	+	170030
rs12238050	A/C	22090726	+	170580
rs10965232	C/T	22091120	+	170974
rs13292938	G/T	22091259	+	171113
rs7028026	A/G	22091435	+	171289
rs6475608	C/T	22091702	+	171556
rs10125231	A/G	22092128	+	171982
rs7859727	C/T	22092165	+	172019
rs1537372	G/T	22093183	+	173037
rs10965233	A/G	22093314	+	173168
rs1537373	G/T	22093341	+	173195
rs7022719	C/T	22093748	+	173602
rs1333042	A/G	22093813	+	173667
rs4336695	C/G	22094450	+	174304
rs7872591	A/C	22095595	+	175449
rs7859362	C/T	22095927	+	175781
rs10757275	A/G	22096225	+	176079
rs6475609	A/G	22096271	+	176125
rs1333043	A/T	22096731	+	176585
rs7855190	C/T	22098069	+	177923
rs10217720	A/G	22098942	+	178796
rs10217426	C/G	22099387	+	179241
rs1412834	C/T	22100131	+	179985
rs17761319	G/T	22100478	+	180332
rs16905648	A/G	22101973	+	181827
rs7341786	A/C	22102241	+	182095
rs7341791	A/G	22102427	+	182281
rs10511701	C/T	22102599	+	182453
rs17834367	C/T	22102606	+	182460
rs7032115	A/G	22102943	+	182797
rs13301964	C/G	22103324	+	183178
rs16905652	A/T	22103924	+	183778

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs10733376	C/G	22104469	+	184323
rs10738609	A/G	22104495	+	184349
rs2383206	A/G	22105026	+	184880
rs10965234	G/T	22105078	+	184932
rs10965235	A/C	22105105	+	184959
rs4990722	G/T	22105217	+	185071
rs944796	C/G	22105285	+	185139
rs944797	C/T	22105286	+	185140
rs1004638	A/T	22105589	-	185443
rs10965236	C/G	22105633	+	185487
rs2383207	A/G	22105959	+	185813
rs1537374	A/G	22106046	+	185900
rs1537375	C/T	22106071	+	185925
rs1537376	C/T	22106220	+	186074
rs7851006	A/G	22107669	+	187523
rs17834457	C/T	22108026	+	187880
rs17761446	G/T	22108102	+	187956
rs7854631	A/C	22108378	+	188232
rs4977758	A/T	22108481	+	188335
rs4977759	C/T	22108885	+	188739
rs1333044	A/G	22109128	+	188982
rs1333045	C/T	22109195	+	189049
rs12685422	A/C	22111167	+	191021
rs10217586	A/T	22111349	+	191203
rs7860589	C/T	22111353	+	191207
rs7020671	A/T	22112193	+	192047
rs10965237	A/C	22112530	+	192384
rs13285121	A/T	22112912	+	192766
rs7869069	G/T	22113590	+	193444
rs10738610	A/C	22113766	+	193620
rs7854016	A/C	22113967	+	193821
rs1333046	A/T	22114123	+	193977
rs7857118	A/T	22114140	+	193994
rs17761458	A/G	22114368	+	194222
rs10757277	A/G	22114450	+	194304
rs10811656	C/T	22114472	+	194326
rs10757278	A/G	22114477	+	194331
rs1333047	A/T	22114504	+	194358
rs10757279	A/G	22114630	+	194484
rs4977575	C/G	22114744	+	194598
rs1333048	A/C	22115347	+	195201
rs1333049	C/G	22115503	+	195357
rs1333050	C/T	22115913	+	195767
rs12345199	A/G	22116454	+	196308
rs12336106	A/G	22116885	+	196739
rs10757281	C/T	22117613	+	197467
rs10811657	A/G	22117641	+	197495
rs17834529	C/G	22117777	+	197631
rs1889086	C/T	22117879	-	197733
rs10965238	A/G	22117883	+	197737
rs10965239	A/G	22117965	+	197819
rs12379111	C/G	22118180	+	198034
rs10811658	A/G	22118600	+	198454
rs12347779	C/G	22118709	+	198563
rs10965240	C/T	22119164	+	199018
rs7020996	C/T	22119579	+	199433
rs10965241	C/G	22119594	+	199448
rs10965243	A/G	22120065	+	199919
rs10965244	A/T	22120389	+	200243
rs10965245	A/G	22120515	+	200369
rs2891169	A/G	22121825	+	201679

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs4977576	A/C	22121861	+	201715
rs2383208	A/G	22122076	+	201930
rs10965246	C/T	22122698	+	202552
rs10965247	A/G	22122729	+	202583
rs10965248	C/T	22122878	+	202732
rs10965249	C/T	22123131	+	202985
rs7045889	A/G	22123251	+	203105
rs10965250	A/G	22123284	+	203138
rs10217762	C/T	22123645	+	203499
rs10811659	C/T	22123716	+	203570
rs12686509	A/T	22123767	+	203621
rs10757282	C/T	22123984	+	203838
rs10965251	A/G	22124029	+	203883
rs10811660	A/G	22124068	+	203922
rs10811661	C/T	22124094	+	203948
rs10757283	C/T	22124172	+	204026
rs10811662	A/G	22124253	+	204107
rs7019437	C/G	22124302	+	204156
rs7019472	C/T	22124395	+	204249
rs7019778	A/C	22124651	+	204505
rs13287212	G/T	22125071	+	204925
rs10965252	A/G	22125919	+	205773
rs12555274	C/G	22126440	+	206294
rs1333051	A/T	22126489	+	206343
rs10965253	G/T	22126687	+	206541
rs7018475	G/T	22127685	+	207539
rs9969854	G/T	22127710	+	207564
rs11791416	A/G	22128105	+	207959
rs10757284	C/G	22128458	+	208312
rs4977761	C/T	22128762	+	208616
rs35660019	A/G	22128865	+	208719
rs10811663	A/G	22129220	+	209074
rs10965254	A/G	22129485	+	209339
rs10965255	C/T	22130019	+	209873
rs2065501	A/C	22130224	+	210078
rs2065503	A/T	22130336	+	210190
rs7866021	C/T	22130339	+	210193
rs7866410	G/T	22130627	+	210481
rs12340450	A/T	22130678	+	210532
rs7854629	A/G	22131034	+	210888
rs7026735	C/T	22131269	+	211123
rs2065504	G/T	22131552	+	211406
rs2065505	A/G	22131790	+	211644
rs4977577	C/T	22131875	+	211729
rs6475610	C/T	22131894	+	211748
rs12376511	C/T	22132756	+	212610
rs10811664	A/G	22132907	+	212761
rs7859532	A/C	22132956	+	212810
rs7862936	A/G	22133133	+	212987
rs7849199	A/T	22133293	+	213147
rs7849302	C/T	22133377	+	213231
rs10757287	A/T	22133570	+	213424
rs2151285	A/T	22134255	-	214109
rs2151284	C/G	22134276	-	214130
rs2151283	G/T	22134305	-	214159
rs2151282	A/G	22134316	-	214170
rs7867100	A/G	22134460	+	214314
rs7853656	G/T	22134530	+	214384
rs2065500	A/G	22135694	+	215548
rs7030345	A/T	22135739	+	215593
rs13298423	A/C	22136196	+	216050

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs13298664	C/T	22136202	+	216056
rs10811665	C/T	22136604	+	216458
rs7022662	C/G	22137715	+	217569
rs13286296	A/T	22137761	+	217615
rs13285137	A/C	22137863	+	217717
rs28752115	A/G	22137944	+	217798
rs12341394	C/T	22138055	+	217909
rs28539212	A/G	22138269	+	218123
rs4614078	C/G	22140034	+	219888
rs7856219	C/T	22140261	+	220115
rs7043398	A/C	22140707	+	220561
rs12337417	C/T	22140792	+	220646
rs12115577	A/T	22140863	+	220717
rs10811666	A/T	22140867	+	220721
rs13293520	C/T	22140897	+	220751
rs7873930	C/T	22141050	+	220904
rs6475611	A/G	22141139	+	220993
rs7047414	A/C	22141412	+	221266
rs10965256	A/G	22141465	+	221319
rs10965257	C/T	22141528	+	221382
rs7849231	C/T	22142401	+	222255
rs4097833	C/G	22142520	-	222374
rs6475612	A/G	22142580	+	222434
rs6475613	G/T	22142643	+	222497
rs7021554	C/T	22142884	+	222738
rs6475614	C/T	22143265	+	223119
rs7853123	A/G	22143360	+	223214
rs10965258	A/G	22143663	+	223517
rs7853621	A/T	22143714	+	223568
rs7045424	C/T	22144009	+	223863
rs6475615	A/G	22144408	+	224262
rs12001831	G/T	22144411	+	224265
rs13295528	A/T	22144432	+	224286
rs7046009	C/T	22144458	+	224312
rs10965259	C/T	22144489	+	224343
rs10965260	A/T	22144539	+	224393
rs10965261	C/T	22144585	+	224439
rs10965262	C/T	22144592	+	224446
rs7030057	C/T	22144628	+	224482
rs10965263	C/G	22144648	+	224502
rs10965264	C/G	22144655	+	224509
rs7046289	C/T	22144664	+	224518
rs7029976	A/T	22144682	+	224536
rs10965265	G/T	22144683	+	224537
rs7046298	C/T	22144684	+	224538
rs7030278	C/G	22144691	+	224545
rs12343752	A/G	22145075	+	224929
rs7033903	A/G	22145090	+	224944
rs944802	A/G	22145709	-	225563
rs1930590	C/T	22146785	-	226639
rs7028213	G/T	22147360	+	227214
rs7856172	C/T	22147489	+	227343
rs7042842	A/G	22147512	+	227366
rs7856274	G/T	22147532	+	227386
rs7869966	C/G	22147546	+	227400
rs7869852	A/T	22147555	+	227409
rs7870203	C/G	22147567	+	227421
rs7870099	C/T	22147624	+	227478
rs7856433	C/T	22147671	+	227525
rs7856749	G/T	22147902	+	227756
rs1333052	A/C	22147908	+	227762

Marker	Type of polymorphism	Location (NCBI Build 34/35/36)	Orientation	Location (SEQ ID NO:1)
rs12238587	A/T	22148168	+	228022
rs10738611	A/G	22148598	+	228452
rs10122243	C/T	22148924	+	228778
rs10757288	C/T	22149416	+	229270
rs10811667	A/C	22149982	+	229836

B. Microsatellite markers markers within LD Block on C09 (between 21,920,147 and 22,149,982; NCBI Build 34/35/36).

5

Marker	Start position	End position	strand	Forward primer	Reverse Primer
DG9S762	21977346	21977478	+	TTATTACGAGCCTGGTCTGGA	CTGTTCGTGCAGGATGAATG
DG9S761	21980412	21980677	+	CCCATCTAAGGGTAGAGAAGC	AAGCAAGATTCCAAACAGTAAACA
DG9S760	21994905	21995264	+	AGCAATCTAGGCGTTTGCAC	TGCTGGCCTTTGCTCTTACT
DG9S746	22034929	22035109	+	TGCTAAATGATCTATTTCCACCAT	CCTTTGCATAGGGAGACCAC
D9S1814	22078225	22078501	+	CTTCGATTGCTGGGATTATG	GGGCCTGTGAACCTACTGAC
D9S1870	22093010	22093220	-	TGGGTATGGTTTTCTGG	TTGAGGCAGGTCAAATAA

Table 11. Amplimers for surrogate SNPs for rs10116277, rs1333040, and/or rs2383207 in the CEU population, as listed in Table 3.

rs7041637
TTTCGCAATGCTTATTTTCAATTTCTTCAGAAATGCCTTAAAGATATTAATGGAGGTAACAACTTAA
TCTCAAATAGTAATCCATAGACAGAATATGTAA[A/C]AGCAATGTTCTCTGATCTGTTCTTTGGCTTCTATT
CCCTAGAGAAATAGTTCTCTAAGACCAAACAGTCTATAGATAGAATTGTAGCAACAGTCAATTAT

rs3218020
GTGGAGAGAAAATGATTATACTTTGAGCTATATGGCTCCAATAAACAAAGATAGATCCCTCAATTT
AAATTTGATCCTCAGAAAACTGAGGGTCAGAGAA[C/T]CCCTCAGGCATGACGGGATAATGTGACAGTTA
ATTTGGTATGTCAACTTGGCTAGGCTGTGGTACCCAGTGTTTGAGTCAAACACCAGTCTAAATATTGC

rs3217992
AGTACTATATTACACTGTTTTTTTTGTTTGTTTTGTTAGTTTTTTTTATTTAAAGCAAACCTCAAACA
TTATTGGGTATCAATTACCACCTGGTTGTATT[A/G]AAATAGTAACTTATCAATGCCATGTAAAAATTAATTC
CATTTTCGAAGCCACCTGGCAGACAGGTTTAGCTGTTTCATCAGCAGCCTAATATATACTGTT

rs1063192
CATTATACTGGGTCATGAAAAATTATCCCTTGAAATAGATATGAAACATGTTACTTCATTTCTGGTT
TAAATAACTTGTGGAATCTTTCCTAATGACAAC[C/T]TGATATTAAGGGAAACTAAAGAAAATGTTATTGTG
GATCCCACAGTACTATATTACACTGTTTTTTTTGTTTGTTTTGTTAGTTTTTTTTATTTAAAGCA

rs2069418
TGATACAAGTTATGAAACTTGTGAAGCCCAAGTACTGCCTGGGGATGAATTTAACTTGTATGACA
GGTGCAGAGCTGTCGCTTTCAGACATCTTAAGAAA[C/G]ACGGAGTTATTTTGAATGACTTTCTCTCGGTC
ACAAGGGAGCCACCAACGTCTCCACAGTGAAACCAACTGGCTGGCTGAAGGAACAGAAATCCTCTGCT

rs2069416
AAATAAAAATAAGATACCTGACAAAGTGGGTTTAAATAGGTAAGAGTGCAAACAAAGATTTACTGT
ACAAATATGATGAAACTGGGATCTCAGATTCTTA[A/C/T]AGTATAATTTTTTTTTGTCTTATGTGTGCCAGG
TTGCCACTCTCAATCTCGAACTAGTTTTTTTCTCTTTTAAGGGTTGTATCCATAATGCAAAAATGGA

rs573687
GTCCAAGACAAATGTGCTATTGTATTACATGTGAAATGTCATCTTTGAAGTCTGGTAAGGGTGTGC
TGTGAGGTGAGCCATCTGGAAAACACAGTGTAGA[C/T]TGAAAAATAATTATAAGCCAGTTTATTACTTTTT
TCCAGTTAAGCCTACCATGACAGCTGCTAAAAAAAACACTATGTAGTATAAAGGGTAAAAAGACTC

rs545226
GGGGTGCAGGTTGTTGGTGTGGCCACACTTCTTCTTGCGGCAATTGACAGCATAGGGGTGCAGG
AGAGCATAGCGCTTATGGCAGATCATCTTGTTTCAG[C/T]TGTATTTCTAGGTGAGCTGGAAGAGTGAAGG
CTCAATAATGCCACCTCGCAGGTGCAGCACCAGGTGCGGGGTGGGCTGTTTCTGGACGTTGTAGTCTGA

rs10811640
TCTTAATTTTTACACATTTTACTTTTCATTTCTTTTTAAACTGTTATTAATAATTTATTCATTTGAATA
AGGATTAAAATAAGGCTAGGATATTGAAATT[G/T]GTTGAAATTGCTACAGTCTCTTGTATCTCTCTCTCTC
TCTTTTTTTTCTTATAAGGGACAGGTTTCATTCACCTTGTCGACCAGGCTGGAGTGCAATACT

rs10811641
TGTGATTCTAGCAGCCATGGATAATTATTTCATAGATTATTATTTTCTTGGGGATGGCAAAATGGT
GATATTCTAATTTTACTATTCCTTCATTTACTAG[C/G]TGGAATGTCTTTTTAAATTATTTATTTATTTATTTA
TTATTTGAGACAGAGTCTTGCACTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCATAGCTCAC

rs2106120
TGCGCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCTCGCCAACTTC
CTTATTTTAAATGCCATTTCCCACTAAAAATAAAAC[A/C]AGTAATTCTTTGAAAAAAAGTTAATATTATGTA
TAGGACTGGAAGTATATAAGATAAAACTGGAATATATTGTCATACCAGAAATCAAAGATTTTGTCAA

rs2106119
ATTACTGATGTGACAAGGTACACAAGCCAATGTTGACATAATGTTTTCAAAATGGGGTGTCTGCTG
TAACTGAACTAAATATAATAACTTTATTCAAGAA[C/T]GAGTTTCAATGATAGGACAAAACTTGATAAAATG
AATAAATAAATAATTATATGCCAGAGTTCAGTAAACCCTGTGTGTACACCTGAAAAAGCTCAAACT

rs643319
GAACAGAGCAGAAGAGAGTCTGGATACACAAATTTCACAATTATTGGCTCCCATCAACATATCTAA
CTCAAGCATAAAGTTGTTTCAGCAGTAGTTTAAG[G/T]TTGGTTACTAATGCAACACCTCTTTGCATGCAAT
GGCCCATTAAATTATCTTCAACTTTAAAAGGTTCCTTTGTTTTTAAATGCTTATAATGAACAAATA

rs7044859
TATGTGATGTAAAGAGCGCCAACATGTTTATATCCTCCTATTTCAATCTACTTTTACTTCATCTACA
TTTTTAGCAATAATGTGAACATGAAATCTTGAA[A/T]AATTAGCTATCTGTAATATATTTACTCATCCACTCA
AAATATTGAGCCCCCCCAATAAATATCATACACTATATTCTAGGTACAGGTGATAAACAATTCA

rs10757264
CCAGTGAAAGTTAGAGAGAGGGGTCTAGAGCTCAGGGAGGAGTGTGTATCCCTAATTTAGACTA
ATTTGCATTAACAGCTGTAGTAATGCAATTTTCTCT[A/G]TACTGAAATGCAGACATTTGAGTATAGAAAAT
TAGCAGACATTTGAATATAGAAGAAAGATTTACTTTCCTTCAGAAAAAGAATAGTAGAGTATAAAGAA

rs10965212
TCGTGCCACTGCACTCCAGCCTGGGTGAGTTGAGAGTCCGTATCAAAAGCAAACCTACACATTTT
TGTGGCCTGTTTTTAGCTCTATCAAGTCAGTTACA[A/T]TCTTCTGTATTCTAGCTTTTTTATCTGTAAGCTC
CTGCAATGCTTTATTAGAATTTTTAGTTCCTTTGCATTTTGTTACAACATACTCCTATAACTCTGC

rs1292137
TAAAAGAAAAAATAATCCAAATGTCAGCAACCTCAAAGATTGAAGGTAGATGAGCCCACAAAGAT
AAGAAATAATCAGCACAAGAACACTGAAAACTCAA[A/T]AATCAAGAGTACCTTCTTTCCTCCAAAGGACC
ACATCACCTCTCTGGCAAGCGTTCAGTACCAGGCTGAGGCTGAGATGGCTGAAATGATAGAAGCAGAA

rs10811644
CAAGACCACTCTGCAGATATCAAGTCTGAAAATTCCCCTAGGGCCAAAGTCTATTATGGGAGCAA
GTTGAGCCTAGAGGGATCGCCATCCCTGTCCATGC[A/T]CTGCTGTAGACACTCCTGCACTAAACCCTCTG
GGCTCCACATCAGCTGGCTTGCTGCTCTACCACTTTGCTTGTCTCTTGGGGGCTCCACCCCAGAGAGA

rs7035484
GTCTCTTGGGGGCTCCACCCCAGAGAGATGTGGGTCAGCAATCATTCAGTTCAATCAGCCCAGGA
TGGAGAGTCTGTGCTATGGGCCCAAGCCAGGGGCT[C/G]TCTGTCTGGTGACGAGCAGCTGGGGGGTGG
GGTGGGACCTGTGGGAGATGGACTGGCCTCCTCTCCTTGAGTCAACTGCTGCTTATTGGAGGTGTGGATG

rs10738604
CCATTGCAGAGGTAGTGGCAGAGAGGCTTTCAGTTGTCCGTGGCGGCTCTGTCCAGGGAGTTGC
TGAGTTGCTATTGGCTTGATATCTCTGGTGGGGTGT[A/G]GCTAGAGGTCCAGGCCTGGAGGACCTGCTC
GTTGAAGAGATGTGGGAATGGGCACCCACATAACAGTCTGTTCACTTTTCCATAGGGCTGCTGTAGTATG

rs615552
GGGGAGCTTAAGCAGGGGTGGAAGGGGAACCCCAGCACAGAACAGCTGCCCGACAAAAATGTG
GCCACACTGCTTTTTTTTTTTTTTTAGCAGGTCCCCA[A/G]TCCCGTTCCTCATCACTGGGTAGAGTCTCCC
ATCTGGGGTCCCCGGCTACCCCAACTGGTGTTCTCTGACTGAGAGAGGTTTCAGACTTCCATGGGACCT

rs543830
TGGGCAAAAGTAGCTGCGACTTTGGCAAAGCTGGAGGTTAGACCCCCAAACATACCCCAGGAGG
TTAGACTCCCATACATACCCCAGAAAAGAGGCTGAA[A/T]CCAGCGAGATCAGCAGAGAAGGTCTACAGG
CCCCACTTCCACAGCGCCTCACAGGATAAGACCCACTGGCTTGGAATTCCAGCTAGCCACCAGTAGCAAC

rs1591136
AAGTGGTAAGTGAATGCGCAACTCTGGGAATCCACACTGCTCTCACAGATCTTTGCAAACCTCAG
ATCAGGAGATCCCCTTGTGAACTCACTCCATTAGG[C/G]CCTTCACACACAGCCACGTGGAGTCTCAGCA
GAGCAGCCACTCAGGCATGCATGGAGACCCAAGAGCTTTAGCTACTCCAGCTTTCTGGGTGTCTGGGCA

rs7049105
CTTAACAGCAAAGTATCAGATTCATTTATAAAACAATGTGACTGATCTTTATGTATGGTTTGTGAAA
CATTTATGCAGTGTCACTTCAGAAAACTCTGCC[A/G]TTATAGATTTGAATTGATTAAGGATATCCACTCCT
TTCCTTGGCATGATACAAATAAATTACTAAAGTATAATTGTAACAATGATAAATATAAGTGACAA

rs679038
CACCCTTTGGGGAAGGGGATCAAAATATAGTGATTTTCTAATTCTAAGATTCGTTCTGTGTATATT
AGCTGCTATTCTTCTAGAAAGAAGACTTTCACCA[C/T]CGTCCCTTGGATTTTTTTTTTTAATTTTTGTATGG
TTTAAATTGCATCATCATTCTTTTTGATGTCCAAATTGTCCCAAATTAAGCCAGTTAGAGAAACA

rs10965215
TAATGGGATTCCTGATGGAATGTTTAGTCTGAATCTAATCACATAGAGACTTGTCTGACAAATCCA
GATTTTTTGGATGTTTTGCAGGACTATTTGCCAC[A/G]ACATTTCAAAGGATTCCAAGAGAGAATATTGGT
GTCCATGCTGTGATGATTCCTCAGCTCCTCTCATCTGATCTCCGTCCTGGCCCCCATGACTTTCTTT

rs564398
CTCTTCTTTTATCACACAGACCTGAAAGATGATGGTTTCCCAAACAGCACTTACAGCAATAGGTGT
GGGCCTCAGTGGCACATACCACACCCTAACTACC[A/G]CAAAGAAAGTCATGGGGGCCAGGACGGAGAT
CAGATGAGAGGAGCTGAGGAATCATCACAGCATGGACACCAATATTCTCTCTTGGAATCCTTTGAAATG

rs10115049
CAAATCTTTCCCTAAGGGAGATTTCAGATGAGGCCCCCAGCCTTGGTAGACACCTTGATTGCAGT
CTTGTGAGAGATTGTGAAGCAGAGCTATTCTCAGA[A/G]TTCATGGGTGTTTTAAGCTATGAGGTTTTTTTT
GGGGGGGAGAATGGTCATTTGTGATTCAGCTATACATAAGTCTACAAAAGTCATTCCAGAAGTGATTC

rs634537
TGGGGCTCACAACCACATGATTCTACCTCCACTGAATCACTTCTGGAATGACTTTTGTAGACTTAT
GTATAGCTGAATCACAAATGACCATTCTCCCCCC[A/C]AAAAAAACCTCATAGCTTAAAACACCCATGAATT
CTGAGAATAGCTCTGCTTCACAATCTCTCACAAGACTGCAATCAAGGTGTCTACCAAGGCTGGGGG

rs2157719
TTTTTATAGTGTGACTCATTTACATATGCATGTGTATGTTTAGGTGCTATTATTAAATTTTGCTGGC
ATATAGTGAGGAAATTGTGATTCAAATTCGTCC[A/G]TATGTACTCCTCCCCCACCATCTGCTCTGCCCCTC
CATTTACCAGAAGGCTAGCTTTAGCTACTTGTGCATGTAAAACAGAAGCAAGCAACACTGTGAAA

rs2151280
GAACATAGATACTCCTTCATTCATGTATTGTCTATGGGTGAGTCTTTATTACAACAGCAGAGATGA
GTAGTTGTGACAGAAACTCGATGGCCCTCAAAAG[C/T]GAAACAAGCTACTATCAGGACCTCTATAGAAAA
AGTTTGCCAACCTCTACACTGTAGTATGCCTTAAGGATTTTTAGAAGATTGAGTATGATAAACACTT

rs1008878
GAAGGGATGATCAGTCCTTCCCTCCTCTATTTTCTTGAGCCCCGTTTTTCACCTTTCTTTTTCTCTC
TCCTTTCTTATCATGAAGAATAAAGACAAATGA[G/T]AACAGATCTACCTTAGGCTGATACAGGGCAGGGA
ATCCATTTAATAATAAAACGTGGGTCAAAATTCACTTTTCTCCTTTTGAATTGAAATTATATTGTG

rs1556515
ATTTGATGCAACTTACACACTGTTGTTATACCTCTAGAATTAAAATGACAATTTTTTAAATAATTTTG
GGGGGCCTAGATTTGCTATTTAACCTATCAAA[A/G]AATTGTGTCTTACAGTATTATTCAAATGTAGTGTGT
AAAGACTTATACTATTGGTCCTAAGCACTACTGGTTGTTTTAGGCTTTTTCTCTTTCTCTGTAG

rs1333037
ACATTTATATAATTAAGATGCTAACACTGACTGACAGAAATGTCAGTAAGATGAAATCAGACTGCA
TGGGAGTTTTATGTTACATTAATTTGTAAATTGT[A/G]TATCTCTGTATTCATGTGAGTGTGGCTATCATGA
TGTTAGACATCCAGCTACAAAGGAGGCATTCGTGCACACACACAGTCTCCAATCTTCTGTTTACCT

rs1360590
TAGACAAAGTTTAATGTTCCCTTTTATATGTTTTCCTGGTAAACAAAAATTGTCTCAGGGTTATTAT
GCATATATGATATTGTCAAGAAACTTTCTGGGT[A/G]CTGTGGGGCAAAGTCTTCTCCATAAATAAGCTAG
GGTTGATTGGAGTTTTCACTTTGAAAAATATCGCACAGGAGGATCTCAACAGCTAGACAATTTCCA

rs1412829
AAATTAAATGACATACGTAAAGTCCTCAATAAATAATAGCTCTTATTACCATTGCTATGGTTACTAT
CACTATTTCTGTATTTTCTTTTGCCATTCCTCA[C/T]GCTTGAATATGAATCTCATGGGTAGAGTTTCCCAAA
GCATGATATGTGTAGTACTACTAAAGGCAAGATTTTGGGTGCATACAGACAAAAAAATAATTTA

rs1360589
AAACCAGAGGAGTAAAATTCTACTTTCACCAGTAATTAGCAGTGTGGAGTTGAGTAAATAAACCTC
TCTAAGTCTCAGTTTCTACATCTACTAAATCTAA[A/G]CAAATTCAAAACAGTGATTATTTCATTAGATTAGA
TATTTTGATTAGTCTTAAATGTCTAATATATAATAAACACTCAACAGGTAGTAGCTATTCTATGT

rs7028570
TCATTTCTTCTTTTAAGGTCTTAATTTCTTGTTTTTCGAGTTTCATGGGAGATATCCAGTCACCAAT
CCAATCCATATCGGGGAAAAGTACAACAAATGA[A/G]TGAAATTTGTAACCAACCTTGGATGATGGAATAA
GACATTTGGGAGAACACAGGAGAAGTGGGGAGGTTAAGGAGGGATAGCTCTGTGAAAATTTTGCAT

rs944801
AACTAATTCTCCAAATTTGCAATTTGGCAGCATCCTACTGGGACTCTAGAAGGCTGATAAATCATG
GAGAGTAGGTATTCATATAGGAACTATGAAAGCT[C/G]TATGTAGTAAACACTACTTAAGAAGGCCTTACA
TTTCATAAAAAGTTGGAGATTTTTGTGGAGACTCATAAAATGCATCCTTTATATCAGTGAAGTTTTT

rs10965219
ACTCGTAGCCAGAGCTACCTTCCAGATGACTTCTTTCTACCACTTTCTTTCTTCCCAGTGTAAGAG
AATGCAAGTATATGCTGATGTTTGGAGCAAGAAC[A/G]TTCAAAAATTTTCTTATTAACATAACTTCTAATG
GAAATACAGTATACTACTATGGTGCATACAAAGAAGAAATAGCAACATATATTTGTTTTAGACCTG

rs7030641
ATAAGCAGCCTTAAATTAAAAAAAAAAAAGTTAACTCATAACTAACTGTGTGACCTGGGATAAGTT
ACTGACCCTCTTTAGGGCTTAGGGTCCTAATCTG[C/T]AAAACGGAAATTATAATAATAACCTTAGCTAGCA
TTTCTTGTGCACATACTATAAGCTGGTGATAAACAATTTATACACACTATCTCATTTAATCCTCAC

rs10120688
TCCAATGCAAAAGAATAATAGGAGCAAAAGCACAGTGGTGAGAAATTGGAGGGGAACTGTGAAA
ATTGCCACATAGATTAGAGGCAGGAAAATAAAGGAC[A/G]GCTAAGTTTATATAGTGAACAGTGAGCCGC
ATGGACACAGGTGACTGTTTTCTCCTTTTTGAACCCCTGCTTACTCCAGAGTCACCACCTCTCCTGGCTT

rs2184061
TATTCTGAGTATTAATTCCTGTTTCCAAATAGATTACTCTTTTAAACATAGCACTACTACTTACCTAA
TGAAATTTAGTTGCTATTAATGGATGAATTTT[G/T]TATCTAACAGGCTTGATTTTGATTATGCATTTTAAAT
GTCAGTCAGACACATATTAATAATGATCCATGTTTGTAGCTAATAGGCCCAATATATACTTTT

rs1537378
GTAAGGGCTGGGACAAATAAACACAAGTAATTTTCAAATATATTAATAATAATATTCTGAGTATTAA
TTCCTGTTTCCAAATAGATTACTCTTTTAAACA[C/T]AGCACTACTACTTACCTAATGAAATTTAGTTGCTAT
TAATGGATGAATTTTGTATCTAACAGGCTTGATTTTGATTATGCATTTTAAATGTCAGTCAGAC

rs8181050
TGGTGGTCCTAAAGTGGCATTAAGGAGCCAATAAATTGTCATTCCTACCTTAGCTCTGTGTCAGAT
GAAATACACAGCATAGTGTGGGGAGAAAATGTTG[A/G]GCTTATTGGGGATGGGGTCTTTCACATAAAGG
AAGAAGGTTTCAGAAGGCATAGTGGTATGAAAAGAGGAGAAACCAAAGGGAGGAAGGTCAATAAAGGG

rs8181047
CAGCATAGTGTGGGGAGAAAATGTTGGGCTTATTGGGGATGGGGTCTTTCACATAAAGGAAGAA
GGTTTCAGAAGGCATAGTGGTATGAAAAGAGGAGAA[A/G]CCAAAGGGAGGAAGGTCAATAAAGGGTTA
AGAACGAGGGGAGGCAAATTGACTTTCTTTCAGCATATGAGGATTATAGGAATGGAAACCTTAATTGGAAT

rs10811647
GAGGATTTAATGCAATTGTTTGTGGGAAAGCACTTTAACAACTCTAAATTACGATATATATGCTAG
GTTTTATTGTTACCCACACCTTTGATGTATTTCT[C/G]TTTGTACTCTTCACTGTATCTGTAACACATTCCCT
AGGATAATTAGGGCTACCCTTTAACAAAGCCAAGATTCTATTTATAGTGGTAAGCTGGCACCTGG

rs1333039
CTTCTGCTATTGAACGAACTTTTTGTTAAGGTAGCTCCCAAGCAGGTTCAGTAGCTTTGTTCTATTA
TCACTTTTCTACTGACAGTGATTTTTTTCCTTT[C/G]AAGGCCTGGGACATGGAGACTGCTTTTCTGCAGAA
ACCACATCCCTTGGAGTAATGAGCTACACCTACCTCAATTATTCAGTGCAGTACAACACTCCAGG

rs10965224
ACATTGTGCACATGTACCCTAGAACTTAAAGTATAATAATAAACAAAAAAAAACCACTGCACAATCT
CTAGTATTCAGATGGAGACTAAGCATGATTTTTC[A/T]TATAAAAGAGCAGATCAGAATGTTGTATCTTTTA
TTCAGAAGACTGGAGTTAATCACTGTTATCTTTAGTACTTAGTGCTGCCAAGGCTGTGTGTTCACA

rs10811650
ACAGAGTGCTTATTTAAAGAAAAATAAAAAGAACACACACACACACGCACGCACACACACACGCA
CGCACACACACACACATGTAGCTACATGTCTAGGA[A/G]GGATGTGGAGAGCTGAAATATGAAGGCAAAA
TAAAACATCTTTTTCAAAGTATACAGCCTACAGTGGTTAGCACAGAGCTGGCCACATAGCAGGGGTTTC

rs10811651
CAATAATGTAGAAGCAAAGAGCCTAAAGTGTTTTCATAAATCTTAAGTGGTAGCTTTATGTTCCAG
TTCAGCAAAACACAAATTTGAAGGCAATCTGTAC[A/G]TTAGGGTTCAGGTGAAGAAGGCAAAGGAATCA
ATGAAATTGTAAAAGCTTTCCAAATTTGCCTTTTCTCTTAAGATTGTCTTTCTCTCATTCTCTTCTCC

rs4977756
TTTCTGTAGCAGAAGTGTAAGGGTGTTACTCGTAGGAGGCCTCTATTGAACTCTTTTCCAGTGAC
GTAGTGTGTGGTCTTTAAGTGGCTTTGCAATGATA[A/G]TAAGATCAGCATTGCATTACTGAATGAGCTCC
TTTAGTAAACGTGGATATGTGCTTTCTGAATCTATTTGTTTGTTTTTCCCAAGTCATAAACAGTGAAT

rs10757269
CCATTTAGAGTACTTGCCTCTGAGGGAAATAAAAATTTGCTAGCAATTTTCTCTAAATGACATTATC
ATAGGCACTTAATTCCTTGATAGGTTCTTTTAG[A/G]TAATTTTTTTATAATGAAGCAATTAATTTGATTCAC
GAAAGTAAGTTTCTAGTTTATATAAAGACCAGATCTGGCCTATTTCTTAGCTTGTCTACATTTG

rs9632884
TGCTAGCAATTTTCTCTAAATGACATTATCATAGGCACTTAATTCCTTGATAGGTTCTTTTAGATAA
TTTTTTTATAATGAAGCAATTAATTTGATTCAC[C/G]AAAGTAAGTTTCTAGTTTATATAAAGACCAGATCTG
GCCTATTTCTTAGCTTGTCTACATTTGAGTAGTTCCATTGCTGGAAAATGACCCTGGAGCTTTT

rs1412832
AAGTTGCTCTCAACATACTTAAAGTTTTCCAATAACTGAATTAAATATCAGTTTATCAGTTTAATATA
AACAATTAGGGTAAATGAAAATAAAATTTCAG[C/T]TCTTTGGTTCCATTAGCCATGGTTCAGGAGCAGAAT
AGTCACCTGAGGCTAGTGACAACGCTTTTGGATCACAGGAAAGAAGAAAAAAAATCAAAATAAT

rs10116277
TTATAACTAATGAGGCAATGTGTCTTGAGTATTTTGAATTAACTCTCTAGAATCGATTCTTGGGGA
GGTTATTTACTTTGAAGTGATGGACAGAGTGTAG[G/T]AGATTTATGAGTGAACTCTTGTCTGATTTGGAA
ATATAGAGTTGTTTAGGCTAGGTATTACCAACCCAAAGTTGACACTTGAGTCACCTAAGTTCTTCTC

rs10965227
TATTAGTTGTGTAATCTTGAAAAAATCTCTGACACTTTTCCCTCTGACTCAGTTTCCCCATCTGGCA
CCCAATCTTTTACAGTGTTATGAAAAATAGGGA[A/G]AATGTAGAAAGGAAGAACATGGCACCCAATCCTT
AATGGACACTCAGTGAAAGCTGGCTATCATCATCATTTTTGGGGTTGTTGTGTTCTACAAATGTAT

rs6475606
TCCCCATCTGGCACCCAATCTTTTACAGTGTTATGAAAAATAGGGAAAATGTAGAAAGGAAGAAC
ATGGCACCCAATCCTTAATGGACACTCAGTGAAAG[C/T]TGGCTATCATCATCATTTTTGGGGTTGTTGTG
TTCTACAAATGTATTTTCCCAGGAGTTTTTTTTACTCTGTCTCCTCTTTCCTTCATATACCCCCAGCC

rs1333040
TGGGAAGGATGAATTAATGGGATGGAGTGCAGGGGATGCAGAGTGCCCACTTATGGAATGATTT
CATTCAAGAGAGACAGGAGGGTCAGAGGTAAGAATG[C/T]TACCGCTGGGACAGAGAGGAAGGTACAGA
TATGAGATATGGTAAGAAGGTATACTACAACAGTGGCTCCCAAATCTCAATGAGTAGCCAGTTCTCATGGA

rs1537370
TTTGAGCCCAAGTCTCTTTCTGACTCTAGGCTTAGAGCTTTAGGGCTATTTCACAAAAGGGCTGTT
CCTAGGTCAGGCATGACAACTTCTATATTACCTT[C/T]GTAAAAGAAGCAATATAATCTACCACTATTAAAT
TTTGCAGGTTAATTTTATATTATGTTTAAATACAGAAAACTTTATTTAAAACTCAGTTGAATTTCT

rs7857345
GAAACTGGATCCCTGATGACATTGAACCATTGACTGAATCTACCCTGGAACCATCAGGAAATAAT
CCTTAGTTTTTTAAAGATGCTTTTAGTTGTGTTTT[C/T]TATTACAAGTACCTGAAAGCATCCTAACTAATCA
ATGCTAAATGCATCTCTCACAGTTTATGCTTATTTTTCAGAAATGCCTAGTGGAAATTTCTATTGC

rs10738607
TCCTTCCACTGACTGAGACTATTTCCTTGCCACAATCAGAAGAACTAAAAGAAAGGAGGATATCTG
TTAATATATGAATTTATCTAAATGTCATGCAGTG[A/G]CTTCTAAAATCATCTGGTGTGCTCTGTTTCCCCTT
GGAGGTGACTTAGGCCTGGCATCCCAAACAATACATACTGGAGTGAAGCTCCAGGAAACCCTGAG

rs10757272
ATACATACTGGAGTGAAGCTCCAGGAAACCCTGAGGAGAAGAGAAGGGCTTAAAGAGCAATCAG
CCTTCGATTGCTGGGATTATGAAAGGTCGTAAGAAG[C/T]GAATGTTGCAATGTTTTATTATACTTGATATT
GAAGCAAGGACAAGTAATAATTTATTATTCTCTCCATGTCAGTGGTATTTACCTTTTTGGAATCATGT

rs4977574
ATAAAAATAAAATAAAATAAAAAATGAAAAACAAAGTCCACTTGTAACCACATGTCAGTAGCATGT
TTGCTTTCAGGGTACATCAAATGCATTCTATAGC[A/G]CAGGATGTTCCAGTCACTCTAACAAAAGATGTC
CTGTTTGGAACACCAACTCTGTATCAGTTACTTCAGACACTTTCTCTCATTGAGTCCCTTCAGCAAG

rs2891168
AACCACATGTCAGTAGCATGTTTGCTTTCAGGGTACATCAAATGCATTCTATAGCACAGGATGTTC
CAGTCACTCTAACAAAAGATGTCCTGTTTGGAAC[A/G]CCAACTCTGTATCAGTTACTTCAGACACTTTCTC
TCATTGAGTCCCTTCAGCAAGCCCTTTTAGGTTTATGTTCTTAGATGAGGAAACCAAGTCTTAGAA

rs1537371
ATATTTTCTTGTTTTTAGATGCACATATACGTACTTTTTTAGCTGGTCATTTCTTTCTGAAATTGGAA
TGAATCTTACAATCAATGGCATGTTATAATTT[A/C]ATTGGCAGCATTATTTGTCTCTTAAGGGCCCCCAAA
TAATAGTGTGTCACATAACTGATAGCATCTCAAATTAGATGAAATACAGTAGTCCAGGCAAGAA

rs1556516
GTTATGGGATAAAGGCGATAGTATTTTATTGACTATATTTTATTCTTTTAATTATTCCTCTAATTTCT
TAAAACAACTTTATTGAGGTATAACTTCCACG[C/G]TATAATTTCACCCATTTTAAGTGCATGAATTCAGTG
ATTTTTAGTAGAGTCATTGAGTAGTGTAACCATTCCTACAATGGTTATAGCACATTTTTATCAT

rs6475608
TGTTTTCTTTTATTTTTCCTTCTAAAATAATCACACGTTTCATTGCAACCCTAACCCTCTTCAACACA
CACACACACACACACACACACACACACACACA[C/T]GGCTTCTAGATTCTACATGTACAAGAGTGCAAATC
AAACTACCATAGAAAAACTAAGAAGAGAGGCCTAGAAGCAAGAGGCTGATACACTATCTCAGGCT

rs7859727
AACCAACACTTAAAATGCAGGGAATTTAAGATAAAAATTTGATAAAAATGGGAAGATTTGGCCGTA
TTGGGCTCATGGTAACTGAGATGCATCTGAATGA[C/T]AGGCATTCCTTTGAATTGCACATTTGCTCTTGTT
TTTACTATAGGCCACTCTCACTTTCTGTTTTTTTCCCCGGCTTTGAAACGATCAGTTTTAGTACTG

rs1537373
ATGACTGGGCAATTATGTCATTATCACCACTGATATATAGCTGGAAGAGTTTAGTGTTGCCCTGCT
AAGATCTGGATTTTCTTTTCTGGAGCTTGGCTAT[G/T]GGGGCATTGAGAAGTCCAGCCAGGAGGTTGGT
CAGAGGCTAACCCAAAAAGCTTTGCTTAACTCTGGGCTACAGCTGGGGGTTGCCAGAGAGAAGTGCCT

rs1333042
AATTTATTTGAGTAGACAGCCAACCCCCTGTATTGTACTCCTTTAAAAAATATTTTAGGCTTTTTAA
ATGCTGAGGCAAGGGGACATACCAAACACTAAC[A/G]GGCACATTGGGGTTTTCTGGCTATTGAAATAAA
AATGTCCTTACATAACACTGATGTACTGGAATAGCACTGCGTTCCAGTGACGGTTATTGCAACTCAG

rs7859362
GGAGCATGATGTGCTTTGATTTCAACTATGGGCTTTATTACTTACTAACTGGGTTACTTTGGTTAA
GTTGTTTGACTCTTGTTTTTTGAGATGGAGTCAG[C/T]CTGGGAGACTCCAGCTCTGTCGCCCAGGCCGGA
GTGCAATGGCACGATCTTGGCTCACTGCAACCTCTGCCTCCCGGATACAAGCGATTCTCATGCCTTA

rs1333043
CTGGCACAAAGTAGGCACTTCATATATAAAAGCTGTGATTATTGATGAACCAGTAGTGAGGTACAT
AACTGGGGAAGGAGAAGGGGCCAGTTTGTGGGAA[A/T]GCTTTTTTAGTTATTAATAGTAAGGTGGTAAA
ATAATAATAGTAATAATAACCAAAAGTTACTGAAAACTAAATACAGTGCTAAACTCTTTAAAAGGAGT

rs1412834
CTTCTTTAGCTGTAAATAAGTAATTGTATGAGGTGATGGTTAAGGTGATTTACTAATTTTACAATTC
TATTATTTTATGAATAGACCCTAGTTAGGATAG[C/T]TTGAAATAGATACTTAATCCACTATTATTCTCTCTT
CTAAGATATAGTTACTAGTTGATCATACTTTTCCTTAAAGGCTGAACTGAATTCTCTGATATCA

rs7341786
GAAACATACTGGTTAATGGAATTCCAGAAAGGACTGAACAATCAAACCATTTTGAAGGACAGCAT
AGAGCTGGACTCTAGAACAGCCAAAACAAGGGGTT[A/C]AACCACTGCGAGGGATCTCTCTCCAACTCTT
GCTCAGGCTTTTCTCCCTGGCTTGACTTTCTTCTCTTTCACTGTAGATTGGCTTCTCTCACATGGCAAG

rs10511701
TATTATGTGTGGCTGACTATATAAAAACATGGATATTTTCTTGGAATCACTTGGTTTGACTGGGAG
AAGACCATTCTCAAAACAAAGGAAGTGCAATTTA[C/T]AGAAGGTAGTAGAATAGGCAGATAAAACAATAA
TTCTTCACTATATTGCTCAAATAATCCCCATGACATTTTTAGTATATTATAAAGAGAGTTCTAAAGT

rs10733376
GAGAAAGATGTTAAGATGAAATTAGATGTGCAAGAGATTCGCCGAGGTAAACCTTGTGGGAGAAA
ATGGAGAGGTACATAGAGGAGCCTGGGCAGACTGT[C/G]TGGCTACTATGTAAGACTCATCCCCATGAAG
GAGAAAGGAGAGGAAGGCAAAGAAGAAAAACCTTAAGATTTCAATTCTAAGAACGTTTTGACAAAGCTG

rs10738609
TGTGCAAGAGATTCGCCGAGGTAAACCTTGTGGGAGAAAATGGAGAGGTACATAGAGGAGCCTG
GGCAGACTGTGTGGCTACTATGTAAGACTCATCCCC[A/G]TGAAGGAGAAAGGAGAGGAAGGCAAAGAA
GAAAAACCTTAAGATTTCAATTCTAAGAACGTTTTGACAAAGCTGATTAGGAGTATTTAAGGCAAAGCTGC

rs2383206
AAATACTTTAACTCATGGCCCGATGATTTTCAGTTAACCAAATTCTCCCTTACTATCCTGGTTGCCC
CTTCTGTCTTTTCCTTAGAAATGTTATTGTAGT[A/G]TTTGCAAGATGGCCTGAATCCTGAACCCCCCATCT
TCAATGAGCACCAAATGGTAATTATAGATTCCCAGCTGTAGAGCTATGTCAGACAAAGGAAACTT

rs944797
CTCTTCCTTGGTGGCTTAAAGTTAGGCTGAAGAAGATTTACATTATGTTGTGCATGACCTCTTTAG
TTTGGTTCTACTTATACTTTCAAGGAGGGAAGAC[C/T]GGGGAAGGTGTCCCTTAGTGAGCATATTTTGTA
CAAATGAAAACAGGGTACTAACACTTATGCCAGGACGCATGCATAAACTAGGATGGTTCTGAGAAAA

rs1004638
AGAACCTTAATGGGAGCACAGGTCCCACCCACCCCTTGCTACCCCATGTACTTGTTCCCATCTTCA
CCCAAGAGAGGAAACACTCTGGAACTAGGGCAGC[A/T]TAAGTGAAGCAGAGTGAAAAGGAATGTGAAG
TTTTGAGAAGAAAGAAAAGGCTAAAGTGTCTATCTTTCCACATTGCTTTTTTCAGGTTTCTCTTCGGAA

rs2383207
GATGAAAAATTCATATTCATCTGAATTTTATAAGTGAATCATGAGAACTCAAAGATACTTAGCCCTT
GGGACCATTTTTTACTCCTGTTCGGATCCCTTC[A/G]GCTAAGCATGATTATTTACTATTTTCAGCTATTAG
TTATGTCTTGTTGAAAAAGTATGAAAAGAGCTGCCCAATAAATTAGAGTGTATGCTCAACATTCT

rs1537374
TTCGGATCCCTTCAGCTAAGCATGATTATTTACTATTTTCAGCTATTAGTTATGTCTTGTTGAAAAA
GTATGAAAAGAGCTGCCCAATAAATTAGAGTGT[A/G]TGCTCAACATTCTCTTAGCTTCTTTATCTCTTTCC
AAAATTGGATCAAATGACATTGGACATGATCAACTTCTTACTGTTTTGACAAACATCTGAGGATA

rs1537375
TTATTTACTATTTTCAGCTATTAGTTATGTCTTGTTGAAAAAGTATGAAAAGAGCTGCCCAATAAAT
TAGAGTGTATGCTCAACATTCTCTTAGCTTCTT[C/T]ATCTCTTTCCAAAATTGGATCAAATGACATTGGAC
ATGATCAACTTCTTACTGTTTTGACAAACATCTGAGGATACTTTTATAATTGATAATTTGGACTA

rs1333045
TTTTGTGCCTCAGTTTCCTCATTCAATATGGGTGTAATAACTGTGCCTGTCTTGTAGGATTATTGTG
AGGCCCAAGTGCAATAATATATAGTACACTGTG[C/T]CTGGCATCTAGTAAGCATTCATTAAGATGACATG
AAGATAACACAGATATATCTTAACATGTAATTATGATTTTGCTTATTCAAGGCCAAGCATTCCAAT

rs10738610
GAAGAAGAAGACAGTCAGAGAGAAGTGAGGGCTTACTTTTCATGTTTAAAGTCTGTTATGTGGTA
AAGGGATTAGATTTATCTGTGTTGTTCCAGGGGAC[A/C]GAAATAGGACAAATGGATGCAAATAGAGTGA
GGAAGATTTAAAACAAATGGAGAAGACATTCTAAAATCAACTACAATGAGCGTAAACAATGACAACGGA

rs1333046
TCATATGCATAGACAAATACACCAAACTGATGAATATTTGCCTTGTATAATCTTTTTGTAGTTTTTTT
ATGAACATATATTACTCAAACAATTTAGAACA[A/T]TTGGCAATATATATATATTTCATTTATAAAAGGTTAG
GAAGATTAATTACACTTTCTGAGGTCGCAACTAAAAGCCAAGATTTTAATCCATTTCTATTTG

rs10757278
AGTGTCACTGGAAAGTGACAAAGAGGACAGTTAAGTTAGTTGGAACTGAACTGAGGCCAGACAG
GGCTGTGGGACAAGTCAGGGTGTGGTCATTCCGGTA[A/G]GCAGCGATGCAGAATCAAGACAGAGTAGT
TTCTCCTTCTCTCTCTCTCTTTAATTGTAACGCCTTTTATAACAAACAAATATTATGCTTATTTCTGTCTT

rs1333047
CAGTTAAGTTAGTTGGAACTGAACTGAGGCCAGACAGGGCTGTGGGACAAGTCAGGGTGTGGTC
ATTCCGGTAAGCAGCGATGCAGAATCAAGACAGAGT[A/T]GTTTCTCCTTCTCTCTCTCTTTAATTGTAA
CGCCTTTTATAACAAACAAATATTATGCTTATTTCTGTCTTTAAATTTTTTGTAGTAATTTCTCATCA

rs4977575
TTTTCTAGTTGAGCTATCATTCATATTTATTATGTGGAACTAGAGGTAGTCCTGGCTACTTGGGAA
CAGCGTGGAGTCTAGCCATGTCAGGGCCAGAAGT[C/G]GTCTCAGCTAAGTTAGAATGTGATACCATTGT
TTACACAAGTGTGGCCTGCCTTCAAGATAGGGTGAGGTGTTTTATGACCACAGGCTTTATGAGTTATA

rs1333048
TGACTCTGAAGATCATACCCGAAGTAGAGCTGCAAAGATATTTGGAATATTGGTAATATCCAATAA
AGAATGACCTTCATGCTATTTTGAGGAGATGTTT[A/C]AATGTCGAATTATTGAAATATTTATAAAATACAA
ATAAACTAACTCTGCTTCATATTCCAACTTGTGTATGACACTTCTTAGGCTATCATTTCATTCCAA

rs1333049
TTCCAACTTGTGTATGACACTTCTTAGGCTATCATTTCATTCCAAATTTATGGTCACTACCCTACTG
TCATTCCTCATACTAACCATATGATCAACAGTT[C/G]AAAAGCAGCCACTCGCAGAGGTAAGCAAGATATA
TGGTAAATACTGTGTTGACAAAAGTATGCAGAAGCAGTCACATTTATACAGTAGTGAAGGAAATGT

rs1333050
AATTACAGTATATCTAAAAAAAGAATAATATATAACAACTGAAAAAATAAAATAGTTGATATAAGCA
GATATTCCAAGATCTGCCAGACATATTGTTAAA[C/T]GAAAAATCTAGATACAAAATTGTTTATAGTTCTCTT
TCATACTATAGCCAAAGAAAATTCAGAAAAAACTACTTACAGTTGATCCTTGAATAATGCAGCA

Table 12: Association results for the MI phenotype for rs2383207 (G) and rs10757278 (G), in 9p21 in Iceland and the US. Results are shown for the initial Icelandic discovery MI case-control group (Iceland A), an independent Icelandic replication group (Iceland B) and for the three US replication groups of Caucasian origin. Also included are the results for the MI case-control groups combined.

Study population (n/m)[a]	Frequency			
Variant (allele)	Controls	Cases	OR (95% CI)	*P*
Iceland A (1607/6728)				
rs2383207 (G)	0.455	0.506	1.22 (1.13-1.33)	1.4×10^{-6}
rs10757278 (G)	0.434	0.489	1.25 (1.15-1.36)	1.5×10^{-7}
Iceland B (665/3533)				
rs2383207 (G)	0.462	0.525	1.29 (1.15-1.45)	0.000026
rs10757278 (G)	0.436	0.503	1.31 (1.16-1.47)	0.000014
Atlanta (596/1284)				
rs2383207 (G)	0.541	0.593	1.23 (1.07-1.42)	0.0030
rs10757278 (G)	0.484	0.551	1.31 (1.14-1.50)	0.00015
Philadelphia (582/504)				
rs2383207 (G)	0.524	0.602	1.37 (1.16-1.63)	0.00026
rs10757278 (G)	0.470	0.550	1.38 (1.17-1.64)	0.00019
Durham (1137/718)				
rs2383207 (G)	0.513	0.559	1.20 (1.05-1.37)	0.0060
rs10757278 (G)	0.460	0.521	1.28 (1.12-1.46)	0.00027
Combined Iceland[b] (2274/10260)				
rs2383207 (G)	0.458	0.511	1.24 (1.16-1.33)	3.3×10^{-10}
rs10757278 (G)	0.435	0.493	1.26 (1.18-1.35)	5.3×10^{-12}
US groups[c] (2315/2508)				
rs2383207 (G)	0.526	0.585	1.25 (1.15-1.36)	1.1×10^{-7}
rs10757278 (G)	0.471	0.541	1.31 (1.21-1.43)	1.5×10^{-10}
Replication groups[d] (2980/6041)				
rs2383207 (G)	0.494	0.555	1.27 (1.18-1.36)	1.4×10^{-11}
rs10757278 (G)	0.454	0.522	1.31 (1.22-1.40)	1.0×10^{-14}
All groups[b,d] (4589/12768)				
rs2383207 (G)	0.492	0.548	1.25 (1.18-1.31)	2.0×10^{-16}
rs10757278 (G)	0.453	0.517	1.28 (1.22-1.35)	1.2×10^{-20}

[a] Number of MI cases (*n*) and controls (*m*). [b] When combining the Icelandic cohorts they are analysed together and the results adjusted for relatedness in the combined group. [c] For the combined groups OR and *P* value are calculated using a Mantel-Haenszel model and the frequency in cases and controls is a simple average over the frequency in the individual groups. [d]When combining Icelandic and US groups, the frequency in cases and controls is the average over the two populations.

Table 13: Genotype specific odds ratio for the risk allele of rs10757278. Shown is the risk for heterozygous carriers (0X) and homozygous carriers (XX) compared to the risk for non-carriers (00), together with 95% confidence intervals (CI), and the population attributable risk (PAR). The lower part of the table includes the corresponding values when the analysis is restricted to early-onset MI cases.

Study population (*n*/*m*)[a]	Genotype specific Odds Ratio[b]			
Variant (allele)	**00**	**0X (95% CI)**	**XX (95% CI)**	**PAR[b]**
Iceland[c] (2272/10261)				
rs10757278 (G)	1	1.25 (1.12-1.39)	1.58 (1.38-1.81)	0.19
US groups (2315/2508)				
rs10757278 (G)	1	1.28 (1.14-1.45)	1.72 (1.45-2.03)	0.23
All groups (4587/12769)				
rs10757278 (G)	1	1.26 (1.16-1.36)	1.64 (1.47-1.82)	0.21
Early onset MI (< 50 for males; < 60 for females)				
Iceland[c] (621/10261)				
rs10757278 (G)	1	1.38 (1.13-1.69)	1.94 (1.53-2.46)	0.27
US groups (1080/2508)				
rs10757278 (G)	1	1.56 (1.32-1.85)	2.08 (1.69-2.58)	0.34
All groups (1701/12769)				
rs10757278 (G)	1	1.49 (1.31-1.69)	2.02 (1.72-2.36)	0.31

[a]Number of MI cases (*n*) and control (*m*). [b]Genotype specific odds ratio for heterozygous (0X) and homozygous carrier (XX) compared to non-carriers (00). [c]Population attributable risk (PAR). [d]For the Icelandic groups, *P* values and OR were adjusted for relatedness using simulations.

Table 14: Association of the G allele of rs10757278 to coronary artery disease (CAD). The association results are shown for CAD, both including and excluding known MI cases. Results are shown for the Icelandic case-control group (excluding the discovery group), for two of the US groups, and for the groups combined.

Study population (n1/n2/m)[a]		All CAD cases			Excluding MI cases		
Variant (allele)	Cont.frq	Case.frq	OR (95% CI)	*P*	Case.frq	OR (95% CI)	*P*
Iceland[b] (1563/773/3533)							
rs10757278 (G)	0.439	0.496	1.26 (1.15-1.37)	1.9×10^{-7}	0.490	1.22 (1.09-1.37)	0.00050
Atlanta (724/128/1284)							
rs10757278 (G)	0.484	0.552	1.31 (1.15-1.50)	0.000036	0.557	1.34 (1.04-1.73)	0.026
Philadelphia (709/126/504)							
rs10757278 (G)	0.470	0.547	1.36 (1.16-1.60)	0.00019	0.528	1.26 (0.96-1.66)	0.10
Combined							
US groups[c] (1433/254/1788)							
rs10757278 (G)	0.477	0.550	1.33 (1.20-1.47)	2.7×10^{-8}	0.542	1.30 (1.08-1.57)	0.0059
All groups[c] (2996/1027/5321)							
rs10757278 (G)	0.458	0.523	1.29 (1.21-1.38)	3.6×10^{-14}	0.525	1.24 (1.13-1.37)	0.000011

[a] Number of all cases (n_1), cases excluding MI patients (n_2), and controls (m). [b] Individuals used in the initial discovery group have been excluded both from cases and controls. [c] For the combined groups, the allelic frequency in cases and controls is a simple average over the individual groups or, when combining Icelandic and US groups, the average over the two populations.

Table 15: Association to MI on 9p21

Shown are all SNPs in the region 21.92 to 22.12 (NCBI build 34) on 9p21 that show nominally significant association to MI in the genome-wide association study. Results are shown for 1607 MI cases and 6728 controls from the Icelandic discovery cohort. Also included are the corresponding results if the association test is done adjusting for the observed association to the three SNPs, rs1333040, rs10116277 and rs2383207 (indicated in bold).

			Frequency					
SNP	**Allele**	**Position**	**Controls**	**MI cases**	**OR**	***P*[a]**	**OR[b]**	***P*[b]**
rs10757260	A	21943137	0.606	0.627	1.09	0.041	1.10	0.037
rs7041637	A	21951866	0.233	0.253	1.12	0.021	0.99	0.91
rs2811712	A	21988035	0.875	0.893	1.19	0.0076	1.16	0.022
rs3218018	A	21988139	0.879	0.896	1.20	0.0074	1.17	0.021
rs3217992	A	21993223	0.344	0.378	1.16	0.00055	0.98	0.75
rs2069426	C	21996273	0.881	0.897	1.18	0.0146	1.15	0.038
rs2069422	A	21998026	0.875	0.894	1.20	0.0059	1.17	0.018
rs2151280	T	22024719	0.464	0.486	1.09	0.038	0.96	0.43
rs1333034	A	22034122	0.875	0.893	1.19	0.0073	1.17	0.021
rs1011970	G	22052134	0.772	0.806	1.22	9.2×10^{-5}	1.14	0.020
rs10116277	T	22071397	0.418	0.468	1.22	1.9×10^{-6}	na	na
rs1333040	T	22073404	0.490	0.542	1.23	6.1×10^{-7}	na	na
rs2383207	G	22105959	0.455	0.506	1.22	1.4×10^{-6}	na	na
rs1333050	T	22115913	0.671	0.693	1.11	0.020	0.97	0.61

[a]*P* value adjusted using genomic controls. [b]*P* value and OR adjusted for the observed association to rs1333040, rs2383207 and rs10116277.

Table 16: Association to MI. Shown is the association for the risk alleles of the three SNPs from the genome-wide study, rs1333040, rs2383207 and rs10116277, and the most significant refinement SNP, rs10757278, to MI in the combined Icelandic case-control group and in the three US case-controls groups.

Study population (*n*/*m*)[a]	Controls		Cases			
Variant (allele)	AA/Aa/aa	Frq.	AA/Aa/aa	Frq	OR (95% CI)	*P*
Iceland A (1607/6728)						
rs1333040 (T)	1770/3315/1636	0.490	342/783/478	0.542	1.23 (1.14-1.34)	6.1×10^{-7}
rs2383207 (G)	2022/3280/1418	0.455	389/811/408	0.506	1.22 (1.13-1.33)	1.4×10^{-6}
rs10116277 (T)	2305/3212/1208	0.418	454/805/350	0.468	1.22 (1.13-1.33)	1.9×10^{-6}
rs10757278 (G)	592/869/318	0.434	413/770/376	0.489	1.25 (1.15-1.36)	1.5×10^{-7}
Iceland B (665/3533)						
rs1333040 (T)	893/1750/889	0.499	135/312/188	0.541	1.18 (1.05-1.33)	0.0065
rs2383207 (G)	1016/1770/746	0.462	146/319/171	0.525	1.29 (1.15-1.45)	0.000026
rs10116277 (T)	1160/1770/602	0.421	178/317/148	0.480	1.27 (1.12-1.43)	0.00010
rs10757278 (G)	224/366/128	0.436	160/329/161	0.503	1.31 (1.16-1.47)	0.000014
Atlanta (596/1284)						
rs1333040 (T)	190/588/369	0.573	63/253/230	0.648	1.37 (1.19-1.58)	0.000016
rs2383207 (G)	273/603/381	0.541	100/270/206	0.593	1.23 (1.07-1.42)	0.0030
rs10116277 (T)	296/571/310	0.503	114/273/190	0.565	1.28 (1.12-1.47)	0.00041
rs10757278 (G)	341/618/287	0.484	119/291/175	0.551	1.31 (1.14-1.50)	0.00015
Philadelphia (582/504)						
rs1333040 (T)	80/225/172	0.585	55/263/232	0.661	1.38 (1.16-1.65)	0.00031
rs2383207 (G)	105/250/127	0.524	86/274/197	0.602	1.37 (1.16-1.63)	0.00026
rs10116277 (T)	120/222/125	0.505	86/262/178	0.587	1.39 (1.18-1.65)	0.00013
rs10757278 (G)'	137/254/103	0.470	116/281/169	0.550	1.38 (1.17-1.64)	0.00019
Durham (1137/718)						
rs1333040 (T)	101/364/230	0.588	159/520/427	0.618	1.14 (0.99-1.30)	0.067
rs2383207 (G)	156/377/176	0.513	230/535/353	0.559	1.20 (1.05-1.37)	0.0060
rs10116277 (T)	166/366/174	0.504	256/526/334	0.534	1.13 (0.99-1.29)	0.076
rs10757278 (G)	189/370/134	0.460	261/545/304	0.521	1.28 (1.12-1.46)	0.00027
Combined						
Iceland[b] (2274/10260)						
rs1333040 (T)	2663/5065/2525	0.493	477/1095/666	0.542	1.21 (1.14-1.30)	1.6×10^{-8}
rs2383207 (G)	3038/5050/2164	0.458	535/1130/579	0.511	1.24 (1.16-1.33)	3.3×10^{-10}
rs10116277 (T)	3465/4982/1810	0.419	632/1122/498	0.471	1.23 (1.15-1.32)	1.1×10^{-9}
rs10757278 (G)	816/1235/446	0.435	573/1099/537	0.493	1.26 (1.18-1.35)	5.3×10^{-12}
US groups[c] (2315/2508)						
rs1333040 (T)		0.582		0.642	1.27 (1.17-1.39)	3.6×10^{-8}
rs2383207 (G)		0.526		0.585	1.25 (1.15-1.36)	1.1×10^{-7}
rs10116277 (T)		0.504		0.562	1.24 (1.14-1.35)	3.1×10^{-7}
rs10757278 (G)		0.471		0.541	1.31 (1.21-1.43)	1.5×10^{-10}
Replication groups[d] (2980/6041)						
rs1333040 (T)		0.541		0.592	1.24 (1.16-1.33)	1.3×10^{-9}
rs2383207 (G)		0.494		0.555	1.27 (1.18-1.36)	1.4×10^{-11}
rs10116277 (T)		0.463		0.521	1.25 (1.17-1.34)	1.3×10^{-10}
rs10757278 (G)		0.454		0.522	1.31 (1.22-1.40)	1.0×10^{-14}
All groups[b,d] (4589/12768)						
rs1333040 (T)		0.538		0.592	1.24 (1.17-1.30)	4.1×10^{-15}
rs2383207 (G)		0.492		0.548	1.25 (1.18-1.31)	2.0×10^{-16}
rs10116277 (T)		0.492		0.548	1.24 (1.17-1.30)	1.8×10^{-15}
rs10757278 (G)		0.453		0.517	1.28 (1.22-1.35)	1.2×10^{-20}

[a] Number of MI cases (*n*) and controls (*m*). [b] When combining the Icelandic cohorts they are analysed together and the results adjusted for relatedness in the combined group. [c] For the combined groups OR and *P* value are calculated using a Mantel-Haenszel model and the frequency in cases and controls is a simple average over the frequency in the individual groups. [d]When combining Icelandic and US groups, the frequency in cases and controls is the average over the two populations.

Table 17: Genotype specific odds ratio.

The upper part shows the genotype specific odds ratios for the risk alleles of the three SNPs from the genome-wide study, rs1333040, rs2383207 and rs10116277, and the most significant refinement SNP, rs10757278, for all MI cases. Shown is the risk for heterozygous carriers (0X) and homozygous carriers (XX) compared to the risk for non-carriers (00), together with 95% confidence intervals (CI). Also included is the population attributed risk (PAR). The lower part of the table includes the corresponding values when the analysis is restricted to early-onset MI cases.

Study population (*n*/*m*)[a]	Genotype specific Odds Ratio[b]			
Variant (allele)	**00**	**0X (95% CI)**	**XX (95% CI)**	**PAR[c]**
Iceland[d] (2272/10261)				
rs1333040 (T)	1	1.18 (1.05-1.32)	1.46 (1.28-1.68)	0.17
rs2383207 (G)	1	1.26 (1.13-1.40)	1.53 (1.34-1.76)	0.19
rs10116277 (T)	1	1.23 (1.11-1.37)	1.52 (1.32-1.74)	0.17
rs10757278 (G)	1	1.25 (1.12-1.39)	1.58 (1.38-1.81)	0.19
US groups (2315/2508)				
rs1333040 (T)	1	1.34 (1.16-1.55)	1.65 (1.38-1.97)	0.28
rs2383207 (G)	1	1.18 (1.04-1.34)	1.54 (1.30-1.82)	0.19
rs10116277 (T)	1	1.16 (1.03-1.31)	1.52 (1.29-1.79)	0.17
rs10757278 (G)	1	1.28 (1.14-1.45)	1.72 (1.45-2.03)	0.23
All groups (4587/12768)				
rs1333040 (T)	1	1.24 (1.14-1.35)	1.52 (1.37-1.69)	0.22
rs2383207 (G)	1	1.22 (1.13-1.32)	1.54 (1.39-1.71)	0.20
rs10116277 (T)	1	1.20 (1.11-1.30)	1.53 (1.38-1.69)	0.18
rs10757278 (G)	1	1.26 (1.16-1.36)	1.64 (1.47-1.82)	0.21
Early onset MI (< 50 for males; < 60 for females)				
Iceland[d] (621/10261)				
rs1333040 (T)	1	1.28 (1.01-1.63)	1.94 (1.50-2.50)	0.27
rs2383207 (G)	1	1.30 (1.04-1.62)	1.80 (1.40-2.32)	0.24
rs10116277 (T)	1	1.32 (1.06-1.63)	1.86 (1.44-2.40)	0.24
rs10757278 (G)	1	1.38 (1.13-1.69)	1.94 (1.53-2.46)	0.27
US groups (1080/2508)				
rs1333040 (T)	1	1.58 (1.29-1.94)	2.00 (1.58-2.53)	0.38
rs2383207 (G)	1	1.40 (1.16-1.67)	1.88 (1.52-2.33)	0.31
rs10116277 (T)	1	1.50 (1.26-1.79)	1.90 (1.53-2.35)	0.32
rs10757278 (G)	1	1.56 (1.32-1.85)	2.08 (1.69-2.58)	0.34
All groups (1701/12769)				
rs1333040 (T)	1	1.46 (1.25-1.70)	1.95 (1.65-2.32)	0.34
rs2383207 (G)	1	1.36 (1.18-1.56)	1.84 (1.56-2.17)	0.28
rs10116277 (T)	1	1.43 (1.25-1.64)	1.87 (1.58-2.20)	0.29
rs10757278 (G)	1	1.49 (1.31-1.69)	2.02 (1.72-2.36)	0.31

[a]Number of MI cases (*n*) and control (*m*). [b]Genotype specific odds ratio for heterozygous (0X) and homozygous carrierr (XX) compared to non-carriers (00). [c]Population attibuted risk (PAR). [d]For the Icelandic groups, *P* values and OR were adjusted for relatedness using simulations.

Table 18: Association to age at onset of MI.

Shown is the association of the risk alleles of the three SNPs from the genome-wide study, rs1333040, rs2383207 and rs10116277, and the most significant refinement SNP, rs10757278, to age at onset of MI. The results are based on regressing the sex adjusted age at onset on the number of risk alleles an individual carries. The combined analysis is done by including a cohort indicator as a explanatory variable in the regression. All MI cases with known age at onset, including late-onset MI, from the four study groups are included in the analysis; this adds 973 MI cases to the study groups compared to the case-controls analysis.

Study population (*n/m*)[a]	Effect (s.e.m.)	*P*
Iceland (2896/750)		
rs1333040 (T)	-1.20 (0.31)	0.00012
rs2383207 (G)	-1.04 (0.31)	0.00080
rs10116277 (T)	-1.05 (0.31)	0.00069
rs10757278 (G)	-1.08 (0.31)	0.00042
Atlanta (611/40)		
rs1333040 (T)	-1.36 (0.69)	0.050
rs2383207 (G)	-1.16 (0.65)	0.075
rs10116277 (T)	-1.47 (0.64)	0.023
rs10757278 (G)	-1.35 (0.65)	0.038
Philadelphia (555/82)		
rs1333040 (T)	-0.85 (0.79)	0.28
rs2383207 (G)	-1.01 (0.76)	0.19
rs10116277 (T)	-1.01 (0.76)	0.18
rs10757278 (G)	-1.25 (0.74)	0.092
Durham (1213/101)		
rs1333040 (T)	-0.89 (0.48)	0.19
rs2383207 (G)	-1.11 (0.46)	0.017
rs10116277 (T)	-1.14 (0.46)	0.013
rs10757278 (G)	-1.19 (0.46)	0.0098
Combined		
US groups (2379/223)		
rs1333040 (T)	-0.97 (0.36)	0.0014
rs2383207 (G)	-1.09 (0.34)	0.0014
rs10116277 (T)	-1.19 (0.34)	0.00039
rs10757278 (G)	-1.26 (0.34)	0.00019
All groups (5275/793)		
rs1333040 (T)	-1.10 (0.23)	2.7×10^{-6}
rs2383207 (G)	-1.06 (0.23)	3.5×10^{-6}
rs10116277 (T)	-1.12 (0.23)	9.2×10^{-7}
rs10757278 (G)	-1.16 (0.23)	2.9×10^{-7}

[a] *n* is the number of MI cases used in the regression and *m* is the number of MI cases used that were not included in the case-control analysis.

Table 19: Association to early-onset MI.

Shown is the association of the risk alleles of the three SNPs from the genome-wide study, rs1333040, rs2383207 and rs10116277, and the most significant refinement SNP, rs10757278, to early-onset MI in the combined Icelandic case-control group and in the three US case-controls groups. Early-onset MI is defined as a MI event before the age of 50 for males and before the age of 60 for females.

Study population (*n*/*m*)[a]	Controls		Cases			
Variant (allele)	AA/Aa/aa	Frq.	AA/Aa/aa	Frq	OR (95% CI)	*P*
Iceland[b] (621/10261)						
rs1333040 (T)	2663/5065/2525	0.493	114/293/210	0.576	1.40 (1.24-1.57)	1.9×10^{-8}
rs2383207 (G)	3038/5050/2164	0.458	133/310/176	0.533	1.35 (1.20-1.52)	3.4×10^{-7}
rs10116277 (T)	3465/4982/1810	0.419	156/308/153	0.496	1.36 (1.21-1.53)	1.9×10^{-7}
rs10757278 (G)	816/1235/446	0.435	142/299/166	0.518	1.40 (1.24-1.57)	3.5×10^{-8}
Atlanta (305/1284)						
rs1333040 (T)	190/588/369	0.573	27/131/121	0.659	1.44 (1.20-1.74)	0.00011
rs2383207 (G)	273/603/381	0.541	45/145/105	0.600	1.27 (1.06-1.52)	0.0082
rs10116277 (T)	296/571/310	0.504	47/150/97	0.584	1.38 (1.16-1.65)	0.00035
rs10757278 (G)	341/618/287	0.484	53/161/86	0.558	1.35 (1.13-1.61)	0.00099
Philadelphia (211/504)						
rs1333040 (T)	80/225/172	0.585	17/102/83	0.661	1.38 (1.09-1.75)	0.0075
rs2383207 (G)	105/250/127	0.524	30/95/75	0.618	1.47 (1.17-1.85)	0.0011
rs10116277 (T)	120/222/125	0.505	27/97/66	0.593	1.42 (1.13-1.79)	0.0026
rs10757278 (G)	137/254/103	0.470	39/103/63	0.561	1.44 (1.15-1.81)	0.0017
Durham (564/720)						
rs1333040 (T)	101/364/230	0.588	64/249/234	0.651	1.31 (1.11-1.54)	0.0012
rs2383207 (G)	156/377/176	0.513	91/271/192	0.596	1.40 (1.20-1.64)	0.000026
rs10116277 (T)	166/366/174	0.504	97/278/178	0.572	1.31 (1.12-1.53)	0.00070
rs10757278 (G)	189/370/134	0.459	105/278/168	0.559	1.49 (1.28-1.75)	5.0×10^{-7}
Combined						
US groups[c] (1080/2508)						
rs1333040 (T)		0.582		0.657	1.37 (1.23-1.52)	1.6×10^{-8}
rs2383207 (G)		0.526		0.605	1.37 (1.23-1.52)	4.8×10^{-9}
rs10116277 (T)		0.504		0.583	1.36 (1.22-1.51)	9.5×10^{-9}
rs10757278 (G)		0.471		0.559	1.43 (1.29-1.59)	1.7×10^{-11}
All groups[c] (1701/12769)						
rs1333040 (T)		0.538		0.617	1.38 (1.28-1.50)	1.6×10^{-15}
rs2383207 (G)		0.492		0.569	1.36 (1.26-1.47)	8.2×10^{-15}
rs10116277 (T)		0.462		0.540	1.36 (1.26-1.47)	9.2×10^{-15}
rs10757278 (G)		0.453		0.539	1.42 (1.31-1.53)	3.3×10^{-18}

[a] Number of MI cases (*n*) and controls (*m*). [b] For the Icelandic group the *P* value and CI are adjusted for relatedness using simulations. [c] For the combined groups, the frequency in cases and controls is a simple average over the frequency in the individual group or, when combining Icelandic and US groups, the average over the two populations.

Table 20: Association to coronary artery disease.

Association of the risk alleles of the three SNPs from the genome-wide study, rs1333040, rs2383207 and rs10116277, and the most significant refinement SNP, rs10757278, to coronary artery disease (CAD) in an Icelandic group of CAD patients and in groups of CAD patients from two of the US study groups. The study group from Durham does not include any CAD patients in addition to the MI patients and was excluded from this part of the analysis. Also included are the corresponding results if all known MI cases are excluded from the CAD patient group.

Study population ($n_1/n_2/m$)[a]	Controls		All CAD cases				Excluding MI cases			
Variant (allele)	AA/Aa/aa	Frq.	AA/Aa/aa	Frq.	OR (95% CI)	*P*	AA/Aa/aa	Frq.	OR (95% CI)	*P*
Iceland[b] (1563/773/3533)										
rs1333040 (T)	893/1750/889	0.499	308/697/413	0.533	1.15 (1.05-1.25)	0.0020	152/328/192	0.527	1.12 (1.00-1.25)	0.057
rs2383207 (G)	1016/1770/746	0.462	353/732/387	0.513	1.23 (1.13-1.34)	2.2×10^{-6}	186/353/185	0.500	1.17 (1.04-1.31)	0.0067
rs10116277 (T)	1160/1770/602	0.421	409/707/333	0.474	1.24 (1.14-1.35)	9.5×10^{-7}	204/333/158	0.468	1.21 (1.08-1.35)	0.0010
rs10757278 (G)	224/366/128	0.439	393/745/376	0.496	1.26 (1.15-1.37)	1.9×10^{-7}	203/365/188	0.490	1.22 (1.09-1.37)	0.00050
Atlanta (724/128/1284)										
rs1333040 (T)	190/588/369	0.572	75/310/281	0.649	1.38 (1.21-1.58)	2.4×10^{-6}	12/57/51	0.655	1.42 (1.09-1.86)	0.010
rs2383207 (G)	273/603/381	0.541	117/335/249	0.595	1.25 (1.09-1.42)	0.00093	17/65/43	0.606	1.30 (1.00-1.69)	0.046
rs10116277 (T)	296/571/310	0.503	130/341/228	0.571	1.31 (1.15-1.50)	0.000038	16/68/38	0.598	1.47 (1.13-1.90)	0.0039
rs10757278 (G)	341/618/287	0.484	139/362/207	0.552	1.31 (1.15-1.50)	0.000036	20/71/32	0.557	1.34 (1.04-1.73)	0.026
Philadelphia (709/126/504)										
rs1333040 (T)	80/225/172	0.585	59/273/235	0.648	1.31 (1.11-1.55)	0.00170	24/49/47	0.588	1.02 (0.76-1.35)	0.92
rs2383207 (G)	105/250/127	0.524	89/285/200	0.600	1.36 (1.15-1.60)	0.00023	25/52/46	0.587	1.29 (0.98-1.71)	0.072
rs10116277 (T)	120/222/125	0.504	90/274/180	0.582	1.37 (1.16-1.61)	0.00017	26/52/43	0.556	1.23 (0.93-1.62)	0.14
rs10757278 (G)	137/254/103	0.470	146/339/208	0.547	1.36 (1.16-1.60)	0.00019	30/57/38	0.528	1.26 (0.96-1.66)	0.10
Combined										
US groups[c] (1433/254/1788)										
rs1333040 (T)		0.579		0.649	1.35 (1.22-1.50)	1.9×10^{-8}		0.621	1.23 (1.01-1.51)	0.044
rs2383207 (G)		0.533		0.598	1.29 (1.17-1.43)	9.8×10^{-7}		0.597	1.30 (1.08-1.57)	0.0068
rs10116277 (T)		0.504		0.577	1.34 (1.21-1.48)	2.1×10^{-8}		0.577	1.35 (1.12-1.63)	0.0018
rs10757278 (G)		0.477		0.550	1.33 (1.20-1.47)	2.7×10^{-8}		0.542	1.30 (1.08-1.57)	0.0059
All groups[c] (2996/1027//5321)										
rs1333040 (T)		0.539		0.591	1.22 (1.14-1.31)	3.2×10^{-9}		0.590	1.14 (1.04-1.26)	0.0082
rs2383207 (G)		0.497		0.555	1.25 (1.17-1.34)	1.3×10^{-11}		0.565	1.20 (1.09-1.32)	0.00019
rs10116277 (T)		0.462		0.525	1.28 (1.20-1.37)	1.7×10^{-13}		0.541	1.25 (1.13-1.37)	7.8×10^{-6}
rs10757278 (G)		0.458		0.523	1.29 (1.21-1.38)	3.6×10^{-14}		0.525	1.24 (1.13-1.37)	0.000011

[a] Number of all cases (n_1), cases excluding MI patients (n_2), and controls (m). [b] Individuals used in the initial discovery group have been excluded both from cases and controls. [c] For the combined groups, the allelic frequency in cases and controls is a simple average over the individual groups or, when combining Icelandic and US groups, the average over the two populations.

Table 21: Markers correlated with the at-risk signal.

All SNPs in the LD-block (based on the HapMap v19 CEU dataset) that are correlated, with $r^2 \geq 0.5$, to at least one of the three SNPs, rs1333040, rs10116277 and rs2383207, together with the correlation coefficients *D'* and r^2. Additional markers selected for typing on both the Icelandic and all the US case/control groups are indicated in bold italic.

				rs1333040		rs10116277		rs2383207	
SNP	**Position[a]**	**Position[b]**	**Frq**	***D'***	r^2	***D'***	r^2	***D'***	r^2
rs10811647	22055002	134856	0.449	0.95	0.49	0.92	0.70	0.92	0.62
rs10811650	22057593	137447	0.450	0.95	0.49	0.93	0.70	0.92	0.63
rs9632884	22062301	142155	0.521	0.96	0.65	1	0.93	0.93	0.84
rs10116277	22071397	151251	0.500	1	0.67	1	1	1	0.90
rs6475606	22071850	151704	0.500	1	0.67	1	1	1	0.90
rs1333040	22073404	153258	0.600	1	1	1	0.67	0.88	0.57
rs1537370	22074310	154164	0.500	1	0.67	1	1	1	0.90
rs7857345	22077473	157327	0.733	1	0.55	1	0.36	0.81	0.26
rs10738607	22078094	157948	0.504	1	0.69	1	1	1	0.90
rs10757272	22078260	158114	0.500	1	0.67	1	1	1	0.90
rs4977574	22088574	168428	0.500	1	0.67	1	1	1	0.90
rs2891168	22088619	168473	0.500	1	0.67	1	1	1	0.90
rs1537371	22089568	169422	0.500	1	0.67	1	1	1	0.90
rs1556516	22090176	170030	0.500	1	0.67	1	1	1	0.90
rs6475608	22091702	171556	0.737	1	0.54	1	0.36	0.8	0.25
rs7859727	22092165	172019	0.496	1	0.66	1	1	1	0.90
rs1537373	22093341	173195	0.500	1	0.67	1	1	1	0.90
rs1333042	22093813	173667	0.508	0.96	0.63	1	0.97	1	0.94
rs7859362	22095927	175781	0.525	0.88	0.57	1	0.90	1	1
rs1333043	22096731	176585	0.517	0.92	0.60	1	0.94	1	0.97
rs1412834	22100131	179985	0.525	0.88	0.57	1	0.90	1	1
rs7341786	22102241	182095	0.533	0.84	0.54	1	0.88	1	0.97
rs10511701	22102599	182453	0.533	0.84	0.54	1	0.88	1	0.97
rs10733376	22104469	184323	0.525	0.88	0.57	1	0.90	1	1
rs10738609	22104495	184349	0.525	0.88	0.57	1	0.90	1	1
rs2383206	22105026	184880	0.525	0.88	0.57	1	0.90	1	1
rs944797	22105286	185140	0.525	0.88	0.57	1	0.90	1	1
rs1004638	22105589	185443	0.525	0.88	0.57	1	0.90	1	1
rs2383207	22105959	185813	0.525	0.88	0.57	1	0.90	1	1
rs1537374	22106046	185900	0.525	0.88	0.57	1	0.90	1	1
rs1537375	22106071	185925	0.525	0.88	0.57	1	0.90	1	1
rs1333045	22109195	189049	0.548	0.68	0.37	0.92	0.69	0.85	0.65
rs10738610	22113766	193620	0.517	0.92	0.60	1	0.94	1	0.97
rs1333046	22114123	193977	0.517	0.92	0.60	1	0.94	1	0.97
rs10757278	22114477	194331	0.491	0.95	0.57	0.96	0.90	1	0.87
rs1333047	22114504	194358	0.492	0.96	0.59	0.97	0.90	1	0.88
rs4977575	22114744	194598	0.492	0.96	0.59	0.97	0.90	1	0.88
rs1333048	22115347	195201	0.508	0.96	0.63	1	0.97	1	0.94
rs1333049	22115503	195357	0.492	0.96	0.59	0.97	0.90	1	0.88

[a] Base-pair location in NCBI Build 34, Build 35 and Build 36.

[b]Position in SEQ ID NO:94 (LD Block C09)

Table 22a: Association to MI for additional markers typed in the LD block C09.

Association to MI of the three SNPs from the genome-wide association study and of the 10 highly correlated refinement markers. The association is calculated for the combined Icelandic and US case-control groups with OR and P-values combined using a Mantel-Haenszel model. Also included are the corresponding adjusted P values for each marker when the association is tested conditional on the observed association of each of the other markers.

			Frequency[b]		Unadjusted		Adjusted *P* values										
SNP	Allele	Position[a]	Controls	Cases	OR	*P*	rs10811650	rs10116277	rs1333040	rs10738607	rs4977574	rs6475608	rs2383207	rs1333045	rs1333046	rs10757278	rs1333048
rs10811650	G	22057593	0.398	0.449	1.21	3.0×10^{-12}	na	0.79	0.10	0.43	0.47	1.6×10^{-6}	0.47	0.063	1.00	0.51	0.86
rs10116277	T	22071397	0.461	0.516	1.24	1.8×10^{-15}	1.2×10^{-4}	na	0.011	0.47	0.83	3.0×10^{-9}	0.20	0.0046	0.38	0.67	0.65
rs1333040	T	22073404	0.537	0.592	1.24	4.1×10^{-15}	6.6×10^{-5}	0.040	na	0.18	0.11	1.6×10^{-8}	0.011	2.3×10^{-4}	0.027	0.15	0.050
rs10738607	G	22078094	0.463	0.525	1.27	2.1×10^{-19}	6.2×10^{-9}	2.5×10^{-5}	4.2×10^{-6}	na	0.056	7.1×10^{-13}	2.8×10^{-4}	8.0×10^{-6}	0.0015	0.61	0.0079
rs4977574	G	22088574	0.465	0.525	1.27	1.1×10^{-18}	1.6×10^{-8}	1.8×10^{-4}	1.7×10^{-5}	0.43	na	2.7×10^{-12}	0.0021	7.3×10^{-5}	0.033	0.61	0.076
rs6475608	C	22091702	0.700	0.737	1.18	6.3×10^{-8}	0.058	0.52	0.84	0.71	0.72	na	0.70	0.83	0.93	0.73	0.93
rs2383207	G	22105959	0.492	0.548	1.25	2.0×10^{-16}	9.2×10^{-6}	0.017	3.8×10^{-4}	0.42	0.73	3.2×10^{-10}	na	0.0020	0.56	0.25	0.78
rs1333045	C	22109195	0.508	0.563	1.24	6.3×10^{-15}	3.4×10^{-5}	0.024	4.1×10^{-4}	0.92	0.85	1.2×10^{-8}	0.15	na	0.50	0.17	0.62
rs1333046	A	22114123	0.468	0.526	1.25	2.5×10^{-17}	6.2×10^{-7}	0.0036	1.5×10^{-4}	0.20	0.68	5.0×10^{-11}	0.038	5.0×10^{-4}	na	0.044	0.74
rs10757278	G	22114477	0.453	0.517	1.28	1.2×10^{-20}	4.8×10^{-10}	2.7×10^{-6}	4.8×10^{-7}	0.039	0.0041	5.0×10^{-14}	2.0×10^{-5}	2.6×10^{-7}	2.4×10^{-5}	na	1.1×10^{-4}
rs1333048	C	22115347	0.472	0.532	1.26	6.0×10^{-18}	1.6×10^{-7}	9.5×10^{-4}	4.6×10^{-5}	0.39	0.94	1.2×10^{-11}	0.017	2.4×10^{-4}	0.24	0.10	na

[a] Base-pair location in NCBI Build 34. [b] The frequency in cases and controls is a simple average over the frequency in Iceland and in US.

Table 22b: Genotype count for additional markers typed in the LD block C09.

Genotype counts in cases and controls for the eight additional refinement SNPs typed in the LD block C09 and for the three SNPs from the genome-wide study, rs1333040, rs2383207 and rs10116277. Genotype counts are shown for the combined Icelandic case-control group (Iceland A+B) and for the three US replication cohorts. For each SNP counts are shown for the risk allele a and the wild type allele A.

			Iceland A+B		Philadelphia		Atlanta		Durham	
			Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases
SNP	Risk Allele	Position[a]	AA/aA/aa	AA/aA/aa	AA/aA/aa	AA/aA/aa	AA/aA/aa	AA/aA/aa	AA/aA/aa	AA/aA/aa
rs10811650	G	22057593	819/1015/279	652/878/336	143/193/74	107/239/119	293/394/147	165/271/133	202/380/112	344/527/222
rs10116277	T	22071397	3465/4982/1810	632/1122/498	120/222/125	86/262/178	296/571/310	114/273/190	166/366/174	256/526/334
rs1333040	T	22073404	2663/5065/2525	477/1095/666	80/225/172	55/263/232	190/588/369	63/253/230	101/364/230	159/520/427
rs10738607	G	22078094	1502/2326/933	558/1099/547	131/244/116	105/278/173	332/603/312	114/283/186	187/373/137	261/552/312
rs4977574	G	22088574	1507/2335/964	554/1105/556	130/246/119	103/286/180	332/597/325	115/274/188	187/383/144	267/549/316
rs6475608	C	22091702	235/980/1135	170/870/1059	36/200/250	26/199/342	101/476/616	24/210/330	58/276/353	67/415/588
rs2383207	G	22105959	3038/5050/2164	535/1130/579	105/250/127	86/274/197	273/603/381	100/270/206	156/377/176	230/535/353
rs1333045	C	22109195	610/1264/638	433/1102/655	115/262/114	95/289/180	286/605/347	93/290/200	159/378/169	218/548/345
rs1333046	A	22114123	1519/2266/959	583/1078/554	123/244/120	102/274/181	315/586/330	114/273/191	182/378/154	251/539/323
rs10757278	G	22114477	816/1235/446	573/1099/537	137/254/103	116/281/169	341/618/287	119/291/175	189/370/134	261/545/304
rs1333048	C	22115347	1372/2202/926	473/947/478	119/247/120	99/274/184	210/418/231	108/286/188	175/379/154	247/532/325

[a] Base-pair location in NCBI Build 34.

Table 23: Primers in the AF109294 gene and ESTs used for PCR screening of cDNA libraries.

ESTs*	Forward primer	Reverse primer
AF109294	TTGGTGTCCATGCTGTGATGATT	GGTTGGGGACCCCTGGTGTA
CN277071	GGTTCAAGCATCACTGTTAGGTGT	GAGGCGGGCGAATCACGA
AW169296	GCTCAGAGCAATTCCAGTGCAAG	GGTTCCAGTCCTGGTTCTGC
BX100299	TCTCATTGGGGATACGAAGCTCT	TCTGGCCCTAGCCTCCATGT
ESTs*	**Nested forward primer**	**Nested reverse primer**
AF109294	AACTCCAAAGAAACCATCAGAGG	TGGGGACCCCTGGTGTAGTG
CN277071	CTTTCCCGAGTCAGTACTGCTTTCT	CGGGCGAATCACGAGGTC
AW169296	TTCCAGTGCAAGTATGGTCTGTGA	CCAGTCCTGGTTCTGCCACA
BX100299	TCATTGGGGATACGAAGCTCTACA	CAGAAAGCTGCAAAGGCCTCA

ESTs* names are from NCBI Build 36.

Table 24. Expression analysis of ESTs and AF109294 in various cDNA libraries by PCR screening

ESTs	cDNA libraries						
	Whole blood	EBV transf. lymphoblasts	Whole heart	Aorta	Cardiac fibro-myocytes	Ventricular fibroblasts	Endothelial cells (HUVEC)
AF109294	positive	nd	nd	positive	positive	nd	nd
CN277071	positive	positive	nd	positive	positive	positive	positive
AW169296	positive	positive	positive	positive	positive	positive	positive
BX100299	positive	positive	nd	positive	positive	positive	positive

nd: not detected. positive: generated PCR products that were confirmed by sequencing

Table 25: Primers used for sequencing of *CDKN2A* and *CDKN2B*.

Primer alias	Forward primer	Primer alias	Reverse primer
CDKA.1e-f.F	AAAGAAGCCAGACACGGAAG	CDKA.1e-f.R	GTAACTGAATCCAGCCAACC
CDKA.1f.F	GGATGAGGCAGCGTGGAC	CDKA.1f.R	AAGCCGTGTCTCAAGATCG
CDKA.1g.F	TCCGGTTTGGCAGCAGTC	CDKA.1g.R	CTAGCAAATGGCAGAACCA
CDKA.1h.F	CAACAGTGTCAGAAACGATGC	CDKA.1h.R	ATCAGTCACCGAAGGTCCTA
CDKA.3b.F	CTTGATCTCCCAAAGTGAAGG	CDKA.3b.R	CGACTCTGGAGGACGAAGTT
CDKA.4d.F	AGATCTCGGAACGGCTCT	CDKA.4d.R	GAGGCGTGCAGCGGTTTA
CDKA.4e.F	GGAAGAAAGGAAAGCGAGGT	CDKA.4e.R	CGGGATCAAGGGGAGTCG
CDKA.4f.F	TCCTCGCGTAGAATGGTTGT	CDKA.4f.R	AGCCCGCGAGGTTTAGGAC
CDKA.4g.F	CCTGAGCGCGGTCTAAGC	CDKA.4g.R	CGTTTTGTCTTGGGTTTGTACC
CKDN2A.1.F	CCCCTTCAGATCTTCTCAGC	CKDN2A.1.R	AGCACCGGAGGAAGAAAGAG
CKDN2A.2.F	CCCGCACCTCCTCTACCC	CKDN2A.2.R	AGTGAACGCACTCAAACACG
CKDN2A.3.F	TTGGCAAGGAAGGAGGACTG	CKDN2A.3.R	TACCAGGCAATGTACACGTC
CKDN2A.4.F	GGTTCACTAAGTCAGAAACCCTAGT	CKDN2A.4.R	AGCTTAGGATGTGTGGCACT
CKDN2A.5.F	AGTCTTCATTGCTCCGCAGT	CKDN2A.5.R	GACACGCTGGTGGTGCTG
CKDN2A.6.F	ATCTATGCGGGCATGGTTAC	CKDN2A.6.R	ACAGTGCTCTCTGCCTGTGAC
CKDN2A.7.F	CAAAATGCTTGTCATGAAGTCG	CKDN2A.7.R	GTGAAGCCATTGCGAGAA
CKDN2A.8.F	TTTCAATCGGGGATGTCTGC	CKDN2A.8.R	CCACTGAGACTCATTATATAACACTCGTT
p14.1.F	ATTCCCACCCAGGATATTCG	p14.1.R	GGTCCCAGTCTGCAGTTAAG
p14.2.F	CTGCGCACCATGTTCTCG	p14.2.R	CGAGCAGCACCAGAATCC
CDKB.06.F	CCCTACTGACTATTACATATCAATGC	CDKB.06.R	CAGAAAATTAAATATACCTGTTAAGTTCG
CDKB.07.F	TTTTAACCATTTAAGGCATAGGA	CDKB.07.R	GCAAACCTCAAACATTATTGG
CDKB.08.F	CTGCTGATGAAACAGCTAAACC	CDKB.08.R	GCACTCAATCATTAGAGGCTACA
CDKB.09.F	TCTTGGAATTTAAGATATAGAGGTCAA	CDKB.09.R	TGCACAAAGAAGTGCATCTAGT
CDKB.10.F	GTTAGAGAAAGAAAAGCCACCTTAG	CDKB.10.R	ACAAGTCATTTGAGAGTGGAGAC
CDKB.11.F	AACATATGCTCTGATTCTCAACTAAC	CDKB.11.R	GGGATTTAATTTCCAGGGTTG
CDKB.12.F	CAAACATTGAGAGAAGGGAACC	CDKB.12.R	GGAAGAACTACAGCTCTTAAATGTAGC
CDKB.13.F	TCTGCACCCTGAGACACTCTA	CDKB.13.R	GGAGACCCTCGCCCAACT
CDKB.14.F	TAAGAGCAAAGGCCAGCATCC	CDKB.14.R	CACTCACCATGAAGCGAAAC
CDKB.15.F	TAATCACTGCCTTCTCCCACTC	CDKB.15.R	GGAGGGCTTCCTGGACAC
CDKB.16.F	GGGTGGGAAATTGGGTAA	CDKB.16.R	GGAAAGTGGATTGCATCAGC
CDKB.17.F	GGCAGGTATGGGAGATGC	CDKB.17.R	TCTCCCCTAAACCATTACTCC
CDKB.23.F	ACAATACAACAGATTTCATATAGTAGCTTAG	CDKB.23.R	TAGTGGAGAAGGTGCGACAG
CDKB.24.F	TAGGTTCCAGCCCCGATCC	CDKB.24.R	GGCTGGCTCCCCACTCTG
CDKB.25.F	TTCCTGGCGCTCAAGAACC	CDKB.25.R	CACAAGGGAGCCACCAAC
CDKB.26.F	CACTGCCCTCAGCTCCTA	CDKB.26.R	CCTGACAAAGTGGGTTTAAATAGGT
CDKB.27.F	TGCATTATGGATACAACCCTTA	CDKB.27.R	TCTTCCTCAGCACTCCGAAC
CDKB.28.F	CGGATGCTACATTGGATAGG	CDKB.28.R	GGCTCAAGAATTGGGTCA
CDKB.a29.F	GAAGGGAACCGGGTAGCA	CDKB.a29.R	CCATAATGTCCTTTCTATTTGACG

Table 26: Sequencing variants in *CDKN2A* and *CDKN2B*.

Shown are all SNPs identified through sequencing of *CDKN2A* and *CDKN2B* for 93 early onset MI cases using primers in **Table 25**. Many of the SNPs identified in the sequencing effort are rare, and have low correlation with rs10757278. These SNPs cannot account for the correlation of rs10757278 to the disease. Two common SNPs, rs3217992 and rs2069416 have modest correlation with rs10757278 (r^2 = 0.36 and 0.37 respectively). rs3217992 is a part of the Illumina Hap300 chip. For Iceland A, rs10757278 gave a *P* value of 1.5×10^{-7}, while rs3217992 gave a *P* value of 5.4×10^{-4}. Hence rs3217992 cannot account for the association of rs10757278. Neither can rs2069416 since it is highly correlated with rs3217992 ($r^2 > 0.8$ both in HapMap CEU and Iceland). SNP rs1063192, which has r^2 of 0.23 with rs10757278 in these sequenced individuals, is also an Illumina SNP and did not even show nominal significance in Iceland A ($P > 0.05$). Rs2069418 is highly correlated with rs1063192.

MAF[a]	A	a	Position[b]	Position[d]	rs names	D' [c]	r^2 [c]	Location
CDKN2A								
0.069	A	G	21958159	38013	rs3088440	0.5	0.02	Exon3
0.176	C	G	21958199	38053	rs11515	0.44	0.03	Exon3
0.005	C	T	21960674	40528		1	0	Intron2
0.042	T	C	21960916	40770	rs3731249	1	0.04	Exon2
0.005	A	G	21961188	41042		1	0	Exon2
0.01	G	C	21964859	44713	rs1800586	1	0.01	Exon1
0.356	C	T	21965017	44871	rs3814960	0.18	0.02	5'UTR
0.042	A	T	21965319	45173	SG09S293*	1	0.04	5'UTR
0.036	T	C	21965561	45415	rs3731238	1	0.03	5'UTR
0.057	A	C	21965807	45661	SG09S291*	0.49	0.01	5'UTR
0.094	C	T	21983964	63818	rs2811711	0.42	0.02	5'UTR
CDKN2B								
0.005	A	G	21957014	36868		1	0.01	3'UTR
0.01	A	G	21957207	37061		1	0.01	3'UTR
0.01	G	T	21957291	37145		1	0.01	3'UTR
0.005	C	A	21957479	37333		1	0.01	3'UTR
0.005	G	C	21957838	37692		1	0.01	3'UTR
0.071	A	G	21958159	38013		0.33	0.01	3'UTR
0.38	G	A	21964218	44072	rs3731239	0.41	0.13	3'UTR
0.005	G	C	21964355	44209		1	0.01	3'UTR
0.01	G	T	21985044	64898		1	0.01	3'UTR
0.005	T	C	21985467	65321		1	0.01	3'UTR
0.323	T	C	21985882	65736	rs2518723	0.29	0.07	3'UTR
0.422	T	C	21993223	73077	rs3217992	0.78	0.36	Exon2
0.398	G	A	21993367	73221	rs1063192	0.53	0.23	Exon2
0.005	C	A	21993417	73271		1	0	Exon2
0.005	C	T	21993591	73445		1	0.01	Exon2
0.078	G	T	21995330	75184	rs3217986	0.58	0.03	Exon2
0.005	C	G	21995493	75347	rs3217984	1	0.01	Exon2
0.104	T	G	21996273	76127	rs2069426	0.08	0	Intron1
0.005	A	G	21996303	76157		1	0.01	Intron1
0.104	T	C	21996348	76202	rs974336	0.08	0	Intron1
0.005	G	A	21996536	76390		1	0.01	Intron1
0.382	G	C	21999698	79552	rs2069418	0.55	0.23	5'UTR
0.011	G	A	21999915	79769		1	0.01	5'UTR
0.005	T	C	21999953	79807		1	0.01	5'UTR
0.1	del	A	21999996	79850	rs2069417	0	0	5'UTR
0.395	A	T	22000004	79858	rs2069416	0.8	0.37	5'UTR
0.089	G	A	22000412	80266	rs495490	0.03	0	5'UTR
0.021	A	G	22000681	80535	SG09S492*	1	0.02	5'UTR
0.005	C	G	22001083	80937		1	0.01	5'UTR

MAF[a]	A	a	Position[b]	Position[d]	rs names	D' [c]	r^2 [c]	Location
0.005	C	G	22001158	81012		1	0.01	5'UTR

[a] MAF : minor allele frequency.
[b] Base-pair location in NCBI Build 34.
[c] Correlation to the refinement SNP rs10757278 based on the 93 sequenced MI cases.
[d] Position in SEQ ID NO:94 (LD Block C09).
*Alternate names used herein for non-public SNPs

Table 27: SNPs in conserved TF bindings sites within the MI region.

SNP		TF binding site			LD to rs1333040		
Name	Location	TF Name	Start	End	D'	r^2	P
rs16935754	22002235	PAX2	22002234	22002253	nd	nd	nd
rs35113513	22023540	FOXO4	22023540	22023551	nd	nd	nd
rs35834365	22023550	FOXO4	22023540	22023551	nd	nd	nd
rs17694493	22031997	STAT	22031995	22032004	0.41	0.05	0.03
rs1412830	22033611	FOXO4*	22033601	22033615	nd	nd	nd
rs1412830	22033611	FOXO3*	22033601	22033615	nd	nd	nd
rs4977758	22108480	EVI1	22108480	22108496	nd	nd	nd
rs34974971	22126835	EVI1	22126830	22126839	nd	nd	nd
rs6475610	22131893	AREB6	22131892	22131905	0.11	0.01	0.36
rs10757289#	22150453	MRF2	22150444	22150458	0.14	0.02	0.22
rs10757289#	22150453	SEF1	22150452	22150471	0.14	0.02	0.22
rs1679013	22196986	BACH1	22196977	22196992	0.01	0	0.96
rs1679014	22197036	PAX6	22197025	22197046	0.21	0	0.57
rs10965296	22205659	GATA6	22205657	22205667	1	0.01	0.31
rs7043085	22323165	OCT1	22323163	22323176	nd	nd	nd
rs1969926	22347344	SOX9	22347334	22347348	nd	nd	nd
rs10113901	22364031	HNF1	22364024	22364039	0.06	0	0.64
rs7046709	22366969	MEF2	22366953	22366975	0.13	0.01	0.34

All coordinates are for human genome release 17 (build 35).
*Two related TFs recognize the same motif.
A SNP lands in two partially overlapping TF binding sites.
LD between SNPs and the rs1333040 is summarized by *D'*, r^2 and a *P* value determined by Chi-square tests the CEU hapmap sample. nd: Measures of LD could not be ascertained for SNPs not represented in the CEU hapmap.

Table 28

Shown are all SNP association results from sequencing of CDKN2A and CDKN2B using primers in Table 25 and association results for early onset MI cases and controls.

location[a]	rs names	Allele	RR	#aff	aff.freq	#con	con.freq	p-value	Variation
21958159	rs3088440	A	0.7264	118	0.080508	674	0.107567	0.194719	G/a
21958199	rs11515	C	1.0349	118	0.161017	668	0.156437	0.858926	G/c
21960916	rs3731249	T	1.4053	119	0.037815	680	0.027206	0.384391	C/t
21964218	rs3731239	A	1.0712	96	0.619792	87	0.603448	0.748662	A/g
21965017	rs3814960	C	0.8697	118	0.347458	686	0.379738	0.34197	T/c
21965319	SG09S293	A	1.4498	119	0.037815	682	0.026393	0.344886	T/a
21965561	rs3731238	T	1.2498	120	0.033333	689	0.026851	0.582384	C/t
21965807	SG09S291	A	1.7884	120	0.058333	687	0.033479	0.077647	C/a
21983964	rs2811711	C	0.8364	119	0.088235	675	0.103704	0.458559	T/c
21985882	rs2518723	T	1.0818	65	0.323077	49	0.306122	0.784969	C/t
21993223	rs3217992	T	1.3279	96	0.421875	86	0.354651	0.188816	C/t
21993367	rs1063192	A	1.27	93	0.602151	80	0.54375	0.273278	A/g
21995330	rs3217986	T	1.1274	96	0.921875	86	0.912791	0.753194	T/g
21996273	rs2069426	T	1.0602	96	0.104167	86	0.098837	0.866531	G/t
21996348	rs974336	T	1.0078	96	0.104167	87	0.103448	0.982048	G/t
21999698	rs2069418	C	1.4107	93	0.61828	87	0.534483	0.107569	C/g
22000004	rs2069416	A	1.6529	93	0.387097	85	0.276471	0.026641	T/a/g
22000004	rs2069416	T	0.6509	93	0.569892	85	0.670588	0.050367	T/a/g
22000004	rs2069416	G	0.804	93	0.043011	85	0.052941	0.660944	T/a/g
22000412	rs495490	A	1.2051	95	0.910526	85	0.894118	0.600156	A/g
22000681	SG09S492	A	3.587	94	0.021277	83	0.006024	0.206723	G/a

[a]location applies to NCBI build 34. Allele: the allele shown is the one tested for association to myocardial infarction. RR is the relative risk. #aff: number of affected individuals. Aff.freq: frequency of allele in affected individuals. #con: number of controls. Con.freq.: frequency of allele in controls. Variation*: Shown are the alleles of the SNPs with major allele shown with capital letters.

Table 29: Association to atherosclerosis in other vascular beds.

Shown is the association of the SNPs, rs1333040, rs2383207 and rs10116277 to peripheral artery disease (PAD), abdominal aorta aneurysm (AAA), and to large vessel disease stroke (LVD).

Study population (*n*/*m*)[a]	Frequency			
Variant (allele)	Controls	Cases	*RR* (95% CI)	*P*
Iceland PAD (1504/3533)				
rs1333040 (T)	0.499	0.503	1.01 (0.93-1.11)	0.75
rs2383207 (G)	0.462	0.481	1.08 (0.99-1.18)	0.082
rs10116277 (T)	0.421	0.438	1.07 (0.98-1.17)	0.12
Emory PAD (34/1284)				
rs1333040 (T)	0.573	0.721	1.92 (1.12-3.30)	0.017
rs2383207 (G)	0.541	0.692	1.91 (1.15-3.17)	0.012
rs10116277 (T)	0.504	0.676	2.06 (1.25-3.39)	0.0044
Iceland LVD (154/3533)				
rs1333040 (T)	0.499	0.527	1.12 (0.87-1.44)	0.39
rs2383207 (G)	0.462	0.488	1.11 (0.88-1.41)	0.38
rs10116277 (T)	0.421	0.457	1.16 (0.91-1.48)	0.24
Iceland AAA (287/3533)				
rs1333040 (T)	0.499	0.572	1.34 (1.12-1.60)	0.0012
rs2383207 (G)	0.462	0.536	1.35 (1.13-1.61)	0.00073
rs10116277 (T)	0.421	0.485	1.30 (1.09-1.54)	0.0035

[a] Number of cases (*n*) and controls (*m*). [b] Individuals used in the initial discovery group have been excluded both from cases and controls. [C] For the combined groups, the allelic frequency in cases and controls is the weighted average over the individual groups.

Table 30: Association of refinement markers to MI, early-onset MI and AAA in Iceland.

Shown is the association to MI, early-onset MI and AAA case-control groups for 10 of the markers included in Table 23b. All tests use the same set of 10260 controls.

			MI (2270/10260)				Early onset MI (621/10260)				AAA (323/10260)				CAD[a] (508/10260)			
SNP	Allele	Position	Con.frq	Case.frq	OR	*P*	Con.frq	Case.frq	RR	*P*	Con.frq	Case.frq	RR	*P*	Con.frq	Case.frq	RR	*P*
rs10116277	T	22071397	0.419	0.471	1.23	3.2E-10	0.419	0.496	1.36	1.3E-07	0.419	0.488	1.32	6.0E-04	0.419	0.456	1.16	0.024
rs1333040	T	22073404	0.493	0.541	1.21	6.7E-09	0.493	0.576	1.40	1.3E-08	0.493	0.579	1.41	2.7E-05	0.493	0.519	1.11	0.130
rs10738607	G	22078094	0.441	0.497	1.25	8.4E-12	0.441	0.521	1.38	4.0E-08	0.441	0.522	1.38	5.8E-05	0.441	0.474	1.14	0.043
rs4977574	G	22088574	0.444	0.499	1.25	2.0E-11	0.443	0.522	1.37	8.3E-08	0.444	0.527	1.39	3.8E-05	0.443	0.474	1.13	0.056
rs6475608	C	22091702	0.683	0.709	1.13	1.7E-03	0.689	0.737	1.26	6.3E-04	0.693	0.752	1.35	1.8E-03	0.695	0.722	1.14	0.110
D9S1870	X	22093010	0.440	0.490	1.22	2.5E-09	0.441	0.513	1.34	1.2E-06	0.441	0.507	1.31	1.2E-03	0.441	0.471	1.13	0.073
rs2383207	G	22105959	0.458	0.511	1.24	6.5E-11	0.458	0.535	1.36	1.4E-07	0.458	0.537	1.38	7.4E-05	0.458	0.487	1.13	0.068
rs1333045	C	22109195	0.503	0.551	1.21	2.0E-08	0.506	0.575	1.32	3.5E-06	0.508	0.580	1.34	4.2E-04	0.508	0.554	1.20	0.007
rs1333046	A	22114123	0.439	0.494	1.24	4.6E-11	0.440	0.518	1.37	8.3E-08	0.442	0.528	1.42	1.7E-05	0.440	0.472	1.14	0.053
rs10757278	G	22114477	0.435	0.492	1.26	2.8E-12	0.434	0.518	1.40	1.2E-08	0.436	0.522	1.41	1.8E-05	0.435	0.474	1.17	0.016

[a] Known cases of MI are excluded from the CAD cases. [b] For the microsatellite D9S1870, all alleles smaller than 2 have been combined in to a composite risk allele X.

Table 31. Amplimers for non-public SNPs identified in table 26.

SG09S293

GGAAGCAGCCCTCGCCAGAGCCAGCGTTGGCAAGGAAGGAGGACTGGGCTCCTCCCCACCTGCC
CCCCACACCGCCCTCCGGCCTCCCTGCTCCCAGCCGCGCTCCCCCGCCTGCCAGCAAAGGCGTG
TTTGAGTGCGTTCACTCTGTTAAAAAGAAATCCGCCCCCGCCCCGTTTCCTTCCTCCGCGATACAA
CCTTCC**[T/a]**AACTGCCAAATTGAATCGGGGTGTTTGGTGTCATAGGGAAAGTATGGCTTCTTCT
TTTAATCATAAGAAAAAGCAAAACTATTCTTTCCTAGTTGTGAGAGCCCCACCGAGAATCGAAATC
ACCTGTACGACTAGAAAGTGTCCCCCTACCCCCTCAACCCTTGATTTTCAGGAGCGCGGGGTTCA
CTAAGTCAGAAACCCTAGTTCAAAGGA

SG09S291

ATTGGAAGGACGGACTCCATTCTCAAAGTCATAATTCCTAGACCAGAAAAAGTGCTCAGTGTTCTA
GAAGCAGAGTTG**[C/a]**ACAGTGATCCAAAGACCAGCTTCAAATACTGTCCTGTCTCCTTCACACT
TCTCACATTTCTCTTTCCTACTGAAAATACCTTGCATTTTTCGTAATTATAAAGGGGGAAGGGAATA
TGAGTGCCCCCTGCTTTATAGGGGTTGTTGTGAGTTTAAATGATGTATTAATACATATAAGCCTTAA
GAACAGTGCCACACATCCTAAGCTAATACCTGTTAGCTCTTGAATTATCCGCTTTGAGGACTGGCT
TGCAATCTTGTTTTGAGGCATAGAAAGAAAATGCTTTGGAGCAGGACGCGGTGGCTCACACCTGT
AATCCCAGCACTTTGGGAAGCCGAGGCGGGCA

SG09S492

TGAATCAACATTTATTACTTAAAATATTTAAAACATTTCAGCGGATGCTACATTGGATAGGAAGAGA
ACCGCAAGTTATGGATTTGTTGCCTAAAAACTTTGGTGAGGAACTGCATAAGTGGACCTCTCCTAA
AAGTGAACAaTTTTTGTTTACAGAATCATTTTGGTTCGGAGTGCTGAGGAAGACAAAGTCTTAACA
GGAGGGCAATTGCTTGTGTATTGCAAAATGAGAGTCTTCACATGTTTTTTTTAGGATACCTTAGCT
CTGACTCCTCATCCCCCAAATCCCTGTAGAATTAAAAAAAgCTCTTTCTTTTAAAGGCAGTGGAAGT
GCCACCACCATGGAAGTGCTGGTTAGGGCTGAAAATCTACTGACAGAGCCTCAACAGAGCTGAAA
TCCACCTGGACAGG**[G/a]**AAGGGAACCGGGTAGCATTAATAACAATTTCTTTTTCTTTCCCATCC
AACCCCCATTTCCTAGTCTTCAGTTTCTTAATTTCTCTACCTTTTACTCTTATGCTCTTGTTTTGACC
TTTGAGTTTCTCTGAAACTTATCAGAAAAGTTAGGACAAGATAGTCTGACCCAATTCTTGAGCCATT
TTCTTAGGTAGTAAATATGTCAGAAAAATGAAAGCTGTTTGGAGTTGATAAGGAAATGGAAGATAA
TGTTTTTCTTTGAGGGgGACATAAAGAATGGTGATAGGGAAAGAACCAATGACTAAGTAAAATGAC
TGAGAATCTTGCACGAGGCAGATGTGTGAGCTTCGCGAAGCAAGTTGACTGAATGAAAAACAACT
TTGGGTAGGGAAAACGTTGCCGGGGGCATTCGC

Table 32: Association between rs10757278 allele G and arterial diseases

Phenotype	Frequency			
Study population (*n/m*)	Controls	Cases	OR (95% CI)	*P*
Abdominal Aortic Aneurysm (AAA)				
Iceland (14259/398)	0.437	0.515	1.37 (1.18-1.58)	2.6×10^{-5}
Belgium (267/176)	0.527	0.574	1.21 (0.92-1.58)	0.18
Canada (150/206)	0.470	0.533	1.29 (0.96-1.74)	0.097
Pennsylvania, US (447/101)	0.468	0.549	1.39 (1.02-1.89)	0.037
The Netherlands (915/476)	0.461	0.529	1.31 (1.12-1.53)	0.00078
UK (252/478)	0.470	0.545	1.35 (1.09-1.68)	0.0064
New Zealand (442/588)	0.474	0.530	1.25 (1.05-1.50)	0.012
All groups (16732/2836)			1.31 (1.22-1.42)	1.2×10^{-12}
Intracranial Aneurysm (IA)				
Iceland (14259/170)	0.437	0.514	1.36 (1.10-1.69)	0.0048
The Netherlands (915/644)	0.461	0.516	1.24 (1.08-1.43)	0.0029
Finland (307/320)	0.400	0.469	1.33 (1.06-1.66)	0.015
All groups (15481/1134)			1.29 (1.16-1.43)	2.5×10^{-6}
Peripheral Artery Disease (PAD)				
Iceland (14259/1764)	0.437	0.473	1.16 (1.07-1.25)	0.00014
Italy (181/179)	0.510	0.499	0.96 (0.71-1.29)	0.78
Sweden (143/206)	0.427	0.507	1.38 (1.02-1.87)	0.036
New Zealand (463/450)	0.474	0.491	1.07 (0.89-1.29)	0.47
All groups (15025/2599)			1.14 (1.07-1.22)	6.1×10^{-5}
LAA /Cardiogenic Stroke				
Iceland (14259/415)	0.437	0.473	1.16 (1.00-1.34)	0.046
Sweden (734/290)	0.433	0.468	1.15 (0.95-1.39)	0.16
All groups (15012/705)			1.15 (1.03-1.29)	0.015
Coronary Artery Disease (CAD) [a]				
Iceland (14259/3051)	0.437	0.492	1.25 (1.17-1.32)	1.9×10^{-12}
Atlanta (1246/840)	0.479	0.544	1.30 (1.15-1.48)	0.000033
Philadelphia (447/724)	0.467	0.547	1.38 (1.17-1.63)	0.00017
Durham (614/1201)	0.455	0.521	1.30 (1.13-1.50)	0.00018
All groups (16566/5539)			1.28 (1.22-1.34)	1.2×10^{-23}

Association results for rs10757278 allele G for the arterial diseases: AAA, IA, PAD, combined LAA (Large Artery Atherosclerotic)/cardiogenic stroke, and CAD, and for T2D, in several study populations. Also included are the results for each phenotype after combining the study populations using a Mantel-Haenszel model. Number of controls (n) and cases (m) is shown. The results for the Icelandic population are adjusted for relatedness of the individuals.

[a]The results presented for CAD have been published previously [1] (apart from the Icelandic control group that has been increased) and are presented here for comparison of the results with the other arterial phenotypes.

Table 33. Association between rs10757278 allele G and arterial diseases after excluding known CAD cases from the sample sets

Phenotype	Frequency			
Study population (*n/m*)	Controls	Cases	OR (95% CI)	*P*
Abdominal Aortic Aneurysm (AAA)				
Iceland (14259/190)	0.437	0.503	1.30 (1.06-1.60)	0.013
Belgium (267/156)	0.527	0.573	1.20 (0.91-1.60)	0.200
Pennsylvania, US (447/62)	0.469	0.513	1.19 (0.82-1.74)	0.36
The Netherlands (915/380)	0.461	0.517	1.25 (1.06-1.48)	0.0097
UK (252/220)	0.470	0.538	1.31 (1.02-1.70)	0.038
New Zealand (442/360)	0.474	0.516	1.18 (0.97-1.44)	0.097
All groups (16639/2017)			1.25 (1.14-1.37)	3.0×10^{-6}
Peripheral Artery Disease (PAD)				
Iceland (14259/732)	0.437	0.463	1.12 (1.00-1.25)	0.055
Italy (181/113)	0.509	0.488	0.92 (0.66-1.28)	0.62
New Zealand (463/326)	0.474	0.491	1.07 (0.87-1.31)	0.51
All groups (14882/1171)			1.09 (0.99-1.20)	0.075
LAA/Cardiogenic Stroke				
Iceland (14259/278)	0.437	0.458	1.09 (0.92-1.30)	0.32
Sweden (734/213)	0.433	0.467	1.15 (0.92-1.42)	0.22
All groups (14993/491)			1.11 (0.97-1.27)	0.12

Association results are shown for rs10757278-G for the arterial diseases: AAA, IA, PAD, combined LAA (Large Artery Atherosclerotic)/cardiogenic stroke, after excluding cases with known CAD. Number of controls (n) and cases (m) is shown. The results for the Icelandic population are adjusted for relatedness. No information on CAD was available for the AAA group from Canada and the PAD group from Sweden. Those study groups were excluded from this analysis. Information on the occurrence of CAD among the AAA cases was available for 97% (466 out of 479) of AAA cases from UK, 86% (87 out of 101) of cases from Pennsylvania, 45% (79 out of 176) of cases from Belgium, 69% (330 out of 476) of cases from The Netherlands, and 98% (575 out of 588) of cases from New Zealand. Among those with this information, the frequency of CAD amongst the AAA subjects was 52% in the UK group, 48% in the Pennsylvania group, 29% in the Belgium group, 29% in the Dutch group and 40% in the group from New Zealand.

Table 34: Genotype specific odds ratio for rs10757278 for abdominal aortic aneurysm and intracranial aneurysm

	Genotype Specific Odds Ratio		
Study population (*n/m*)	**AA**	**AG (95% CI)**	**GG (95% CI)**
Abdominal Aortic Aneurysm (AAA)			
Iceland (14259/398)	1	1.22 (0.95-1.56)	1.85 (1.39-2.45)
Belgium (267/176)	1	1.42 (0.91-2.20)	1.52 (0.87-2.66)
Canada (150/206)	1	1.04 (0.69-1.57)	1.62 (0.90-2.91)
Pennsylvania, US (447/101)	1	2.00 (1.17-3.44)	2.06 (1.06-4.03)
The Netherlands (915/476)	1	1.43 (1.12-1.84)	1.76 (1.28-2.43)
UK (252/478)	1	1.60 (1.19-2.15)	1.95 (1.26-3.03)
New Zealand (442/588)	1	1.22 (0.94-1.58)	1.50 (1.04-2.17)
Combined (16732/2836)	1	1.36 (1.21-1.52)	1.74 (1.49-2.02)
Intracranial Aneurysm (IA)			
Iceland (14259/170)	1	1.39 (0.96-2.02)	1.90 (1.24-2.93)
The Netherlands (915/644)	1	1.34 (1.08-1.66)	1.61 (1.20-2.16)
Finland (307/320)	1	1.45 (1.06-1.98)	1.79 (1.12-2.86)
Combined (15481/1134)	1	1.38 (1.18-1.63)	1.72 (1.39-2.13)

Genotype specific odds ratios for rs10757278 for AAA and IA cases versus controls. Shown is the risk for heterozygous carriers (AG) and homozygous carriers (GG) compared to the risk for non-carriers (AA), together with 95% confidence intervals (CI). Results are shown for the AAA case-control groups from Iceland, Belgium, Canada, Pennsylvania, US, UK, The Netherlands and New Zealand and for all the groups combined and for the IA case-control groups from Iceland, The Netherlands and Finland. Number of controls (n) and cases (m) is shown. Tests of heterogeneity showed no significant difference in the genotype specific odds ratio between the different study groups For AAA, P_{het} = 0.38 and P_{het} = 0.95 for the AG and the GG genotype, and for IA, P_{het} = 0.91 and P_{het} = 0.81.

Table 35: Correlation between growth rate of AAA and genotypes for rs10757278 from the UK Small Aneurysm Trial

Genotype of rs10757278	*n*	Mean baseline diameter (mm)	Linear growth rate (mm/year)	Mean difference (95%CI)
AA	79	45.3	3.20	0.03 (-0.38-0.41)
AG	214	44.8	3.15	reference
GG	107	44.7	2.53	-0.46 (-0.93-0.00)

Linear growth rates were determined as previously described[16]. At least three AAA diameter measurements and growth rate were available for 400 patients who had been genotyped for rs10757278 and *n* is the number of individuals with the different genotypes. The largest group (AG) was set as the reference group and then the growth rates in the other genotype groups were compared with the mean of the reference group. This leads to estimation of the mean difference [95%CI] of the growth rates in the homozygous groups. The analysis of the average difference was adjusted for age, sex, smoking status, baseline diameter and curvature in growth pattern. In this cohort there were 24 ruptured AAA; 6 with the AA genotype, 14 with AG, and 4 with GG.

Table 36: Association of SNPs in chromosome 9p21 region to MI in African-Americans. We tested 9 SNPs for association with MI in African Americans. These SNPs included 2 SNPs from the genome-wide scan on MI in Icelanders (rs10116277 and rs2383207) and rs10757278, which showed strongest association with MI in Caucasians, as well as six other SNPs that were correlated with rs10757278 in Caucasians. As shown in the Table all, SNPs have greater frequency in cases compared to the control groups and the odds ratios are comparable to that for Caucasians.

			Combined[a]		Philadelphia (93/139)				Durham (262/243)				Cleveland (46/81)				Atlanta (91/357)			
SNP	Allele	Position	RR (95CI)	Pa	Con.frq	Case.frq	OR	P	Con.frq	Case.frq	RR	P	Con.frq	Case.frq	RR	P	Con.frq	Case.frq	RR	P
rs8181050	A	22054391	1.30 (0.91-1.86)	0.14	0.921	0.920	0.99	0.98	0.912	0.926	1.20	0.43	0.887	0.913	1.34	0.51	0.934	0.967	2.08	0.075
rs10116277	T	22071397	1.37 (1.05-1.80)	0.022	0.889	0.893	1.05	0.89	0.870	0.901	1.35	0.13	0.856	0.891	1.39	0.41	0.886	0.934	1.82	0.049
rs10738607	G	22078094	1.12 (0.92-1.35)	0.27	0.238	0.290	1.31	0.22	0.233	0.254	1.12	0.43	0.228	0.244	1.09	0.77	0.226	0.220	0.97	0.87
rs4977574	G	22088574	1.20 (0.97-1.48)	0.096	0.175	0.250	1.57	0.051	0.184	0.204	1.13	0.430	0.160	0.163	1.02	0.95	0.173	0.187	1.10	0.68
rs2383207	G	22105959	1.34 (1.03-1.74)	0.027	0.881	0.877	0.96	0.88	0.872	0.906	1.43	0.078	0.877	0.880	1.04	0.93	0.888	0.945	2.18	0.015
rs1333045	C	22109195	1.22 (1.04-1.43)	0.014	0.455	0.497	1.18	0.38	0.472	0.525	1.24	0.093	0.416	0.435	1.08	0.78	0.473	0.544	1.33	0.094
rs1333046	A	22114123	1.19 (0.99-1.44)	0.069	0.227	0.294	1.42	0.11	0.257	0.291	1.18	0.24	0.204	0.213	1.06	0.86	0.240	0.258	1.10	0.61
rs10757278	G	22114477	1.19 (0.98-1.45)	0.08	0.183	0.257	1.54	0.059	0.194	0.223	1.20	0.250	0.160	0.152	0.94	0.86	0.187	0.198	1.07	0.74
rs1333048	C	22115347	1.16 (0.96-1.40)	0.12	0.277	0.301	1.13	0.57	0.288	0.321	1.17	0.25	0.251	0.304	1.30	0.37	0.283	0.303	1.10	0.61

[a] Results for the four African-American cohorts are combined using a Mantel-Haenzsel model. Shown is the association to MI in four African-American MI case-control groups for 9 markers in the chromosome 9p21LD-block (LD Block C09). These SNPs are correlated in Caucasians. The markers include 2 SNPs from the genome-wide scan (rs10116277 and rs2383207) and rs10757278 that showed strongest association with MI in Caucasians, and six other correlated SNPs.

Table 37. Key to sequence listing.

Sequence	SEQ ID NO:
rs7041637	1
rs3218020	2
rs3217992	3
rs1063192	4
rs2069418	5
rs2069416	6
rs573687	7
rs545226	8
rs10811640	9
rs10811641	10
rs2106120	11
rs2106119	12
rs643319	13
rs7044859	14
rs10757264	15
rs10965212	16
rs1292137	17
rs10811644	18
rs7035484	19
rs10738604	20
rs615552	21
rs543830	22
rs1591136	23
rs7049105	24
rs679038	25
rs10965215	26
rs564398	27
rs10115049	28
rs634537	29
rs2157719	30
rs2151280	31
rs1008878	32
rs1556515	33
rs1333037	34
rs1360590	35
rs1412829	36
rs1360589	37
rs7028570	38
rs944801	39
rs10965219	40
rs7030641	41
rs10120688	42
rs2184061	43
rs1537378	44
rs8181050	45
rs8181047	46
rs10811647	47
rs1333039	48
rs10965224	49
rs10811650	50
rs10811651	51
rs4977756	52
rs10757269	53
rs9632884	54
rs1412832	55
rs10116277	56
rs10965227	57
rs6475606	58
rs1333040	59
rs1537370	60

rs7857345	61
rs10738607	62
rs10757272	63
rs4977574	64
rs2891168	65
rs1537371	66
rs1556516	67
rs6475608	68
rs7859727	69
rs1537373	70
rs1333042	71
rs7859362	72
rs1333043	73
rs1412834	74
rs7341786	75
rs10511701	76
rs10733376	77
rs10738609	78
rs2383206	79
rs944797	80
rs1004638	81
rs2383207	82
rs1537374	83
rs1537375	84
rs1333045	85
rs10738610	86
rs1333046	87
rs10757278	88
rs1333047	89
rs4977575	90
rs1333048	91
rs1333049	92
rs1333050	93
LD Block C09	94

CLAIMS

1. A method for determining a susceptibility to cardiovascular disease in a human individual, 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 derived from the individual, wherein the at least one polymorphic marker is selected from the polymorphic markers set forth in Table 10, and markers in linkage disequilibrium therewith, and wherein the presence of the at least one allele is indicative of a susceptibility to cardiovascular disease.

2. The method according to Claim 1, wherein the at least one polymorphic marker is located within the genomic segment with the sequence as set forth in SEQ ID NO:94.

3. The method according to Claim 1 or Claim 2, wherein the at least one polymorphic marker is in linkage disequilibrium with the CDKN2A and/or CDKN2B genes.

4. The method according to any of the preceding Claims, wherein the at least one polymorphic marker is selected from the markers set forth in Table 3 and Table 21.

5. The method according to any of the preceding Claims, wherein the at least one polymorphic marker is selected from rs10811650, rs10116277, rs1333040, rs10738607, rs4977574, rs6475608, D9S1870, rs2383207, rs1333045, rs1333046, rs10757278 and rs1333048.

6. The method according to any of the preceding Claims, wherein the at least one polymorphic marker is selected from rs1333040, rs10116277, rs2383207 and rs10757278.

7. The method according to any of the preceding Claims, wherein the at least one polymorphic marker is rs10757278.

8. The method according to any of the preceding Claims, further comprising assessing the frequency of at least one haplotype in the individual.

9. The method of any of the preceding claims, wherein the susceptibility conferred by the presence of the at least one allele or haplotype is increased susceptibility.

10. The method according to Claim 9, wherein the presence of allele T in rs1333040, allele T in rs10116277, allele G in rs2383207, allele G in rs10757278, allele X in DG9S1870 or allele 0 in D9S1814 is indicative of increased susceptibility to cardiovascular disease.

11. The method according to Claim 9 or 10, wherein the presence of the at least one allele or haplotype is indicative of increased susceptibility with a relative risk (RR) or odds ratio (OR) of at least 1.2.

12. The method according to Claim 9 or 10, wherein the presence of the at least one allele or haplotype is indicative of increased susceptibility with a relative risk (RR) or odds ratio (OR) of at least 1.3.

13. The method according to any of the claims 1-8, wherein the susceptibility conferred by the presence of the at least one allele or haplotype is decreased susceptibility.

14. The method according to any of the preceding Claims wherein the presence of the marker or haplotype is indicative of a different response rate of the subject to a particular treatment modality for a Cardiovascular disease.

15. The method according to Claim 14, wherein the treatment modality is a coronary stenosis method selected from balloon angioplasty, stenting, cutting balloon angioplasty, percutaneous transluminal coronary angioplasty (PTCA), directional coronary atherectomy, rotational coronary atherectomy, brachytherapy, drug-eluting stent (DES) insertion, metal stent insertion, and coronary artery surgery.

16. The method of claim 14, wherein the treatement modality is adminstration of a therapeutic agent for preventing and/or ameliorating symptoms associated with the Cardiovascular disease..

17. A method of identification of a marker for use in assessing susceptibility to cardiovascular disease, the method comprising

a. identifying at least one polymorphic marker in linkage disequilibrium with at least one of the markers within LD Block C09 (SEQ ID NO:94);

b. determining the genotype status of a sample of individuals diagnosed with, or having a susceptibility to, cardiovascular disease; and

c. determining the genotype status of a sample of control individuals;

wherein a significant difference in frequency of at least one allele in at least one polymorphism in individuals diagnosed with, or having a susceptibility to, cardiovascular disease, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing susceptibility to cardiovascular disease.

18. The method according to Claim 17, wherein an increase in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or

having a susceptibility to, cardiovascular disease, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing increased susceptibility to cardiovascular disease.

19. The method according to Claim 17 or Claim 18, wherein a decrease in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or having a susceptibility to, cardiovascular disease, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing decreased susceptibility to, or protection against, cardiovascular disease.

20. The method according to any of the Claims 17-19, wherein the at least one marker within LD Block C09 is selected from the markers set forth in Table 10.

21. The method according to any of the Claims 17-20, wherein the at least one marker within LD Block C09 is selected from the markers set forth in Table 3 and Table 21.

22. The method according to any of the Claims 17-21, wherein the at least one marker within LD Block C09 is selected from rs1333040, rs10116277, rs2383207 and rs10757278.

23. A method of genotyping a nucleic acid sample obtained from a human individual at risk for, or diagnosed with, cardiovascular disease, comprising determining the presence or absence of at least one allele of at least one polymorphic marker in the sample, wherein the at least one marker is selected from the group consisting of the markers set forth in Table 3 and Table 21, and markers in linkage disequilibrium therewith, and wherein the presence or absence of the at least one allele of the at least one polymorphic marker is indicative of a susceptibility of cardiovascular disease.

24. The method of Claim 23, wherein the at least one marker is selected from rs10811650, rs10116277, rs1333040, rs10738607, rs4977574, rs6475608, D9S1870, rs2383207, rs1333045, rs1333046, rs10757278 and rs1333048.

25. The method according to Claim 23 or 24, wherein genotyping comprises amplifying a segment of a nucleic acid that comprises the at least one polymorphic marker by Polymerase Chain Reaction (PCR), using a nucleotide primer pair flanking the at least one polymorphic marker.

26. The method according to any of the Claims 23 - 25, wherein genotyping is performed using a process selected from allele-specific probe hybridization, allele-specific primer extension, allele-specific amplification, nucleic acid sequencing, 5'-

exonuclease digestion, molecular beacon assay, oligonucleotide ligation assay, size analysis, and single-stranded conformation analysis.

27. The method according to Claim 25, wherein the process comprises allele-specific probe hybridization.

28. The method according to Claim 27, wherein the process comprises DNA sequencing.

29. The method according to any of the Claims 23 – 27, comprising:

4) contacting copies of the nucleic acid with a detection oligonucleotide probe and an enhancer oligonucleotide probe under conditions for specific hybridization of the oligonucleotide probe with the nucleic acid;

wherein

a) the detection oligonucleotide probe is from 5-100 nucleotides in length and specifically hybridizes to a first segment of the nucleic acid whose nucleotide sequence is given by SEQ ID NO:94;

b) the detection oligonucleotide probe comprises a detectable label at its 3' terminus and a quenching moiety at its 5' terminus;

c) the enhancer oligonucleotide is from 5-100 nucleotides in length and is complementary to a second segment of the nucleotide sequence that is 5' relative to the oligonucleotide probe, such that the enhancer oligonucleotide is located 3' relative to the detection oligonucleotide probe when both oligonucleotides are hybridized to the nucleic acid; and

d) a single base gap exists between the first segment and the second segment, such that when the oligonucleotide probe and the enhancer oligonucleotide probe are both hybridized to the nucleic acid, a single base gap exists between the oligonucleotides;

5) treating the nucleic acid with an endonuclease that will cleave the detectable label from the 3' terminus of the detection probe to release free detectable label when the detection probe is hybridized to the nucleic acid; and

6) measuring free detectable label, wherein the presence of the free detectable label indicates that the detection probe specifically hybridizes to the first segment of the nucleic acid, and indicates the sequence of the polymorphic site as the complement of the detection probe.

30. A method of assessing an individual for probability of response to a therapeutic agent for preventing and/or ameliorating symptoms associated with cardiovascular disease, 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, wherein the at least one polymorphic marker is selected from the group consisting of the polymorphic markers set forth in Table 3 and Table 21, and markers in linkage disequilibrium therewith, wherein the presence of the at least one allele of the at least one marker is indicative of a probability of a positive response to the therapeutic agent.

31. A method of predicting prognosis of an individual diagnosed with cardiovascular disease, the method 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, wherein the at least one polymorphic marker is selected from the group consisting of the polymorphic markers set forth in Tables 3 and 21, and markers in linkage disequilibrium therewith, wherein the presence of the at least one allele is indicative of a worse prognosis of the cardiovascular disease in the individual.

32. A method of monitoring progress of treatment of an individual undergoing treatment for cardiovascular disease, the method 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, wherein the at least one polymorphic marker is selected from the markers set forth in Tables 3 and 21, and markers in linkage disequilibrium therewith, wherein the presence of the at least one allele is indicative of the treatment outcome of the individual.

33. The method according to any of the Claims 30-32, wherein the at least one polymorphic marker is selected from rs10811650, rs10116277, rs1333040, rs10738607, rs4977574, rs6475608, D9S1870, rs2383207, rs1333045, rs1333046, rs10757278 and rs1333048, and markers in linkage disequilibrium therewith.

34. The method of any of the preceding Claims, further comprising assessing at least one biomarker in a sample from the individual.

35. The method of Claim 34, wherein the biomarker is a cardiac marker or an inflammatory marker.

36. The method of Claim 34 or Claim 35, wherein the at least one biomarker is selected from creatin kinase, troponin, glycogen phosphorylase, C-reactive protein (CRP), serum amyloid A, fibrinogen, interleukin-6, tissue necrosis factor-alpha, soluble vascular cell adhesion molecules (sVCAM), soluble intervascular adhesion

molecules (sICAM), E-selectin, matrix metalloprotease type-1, matrix metalloprotease type-2, matrix metalloprotease type-3, matrix metalloprotease type-9, serum sCD40L, leukotrienes, leukotriene metabolites, interleukin-6, tissue necrosis factor-alpha, myeloperoxidase (MPO), and N-tyrosine.

37. The method of Claim 36, wherein the leukotriene is selected from LTB4, LTC4, LTD4 and LTE4.

38. The method of any of the preceding claims, further comprising analyzing a sample comprising genomic DNA from a human individual or a genotype dataset derived from a human individual for the presence or absence of at least one at-risk allele of at least one at-risk variant for cardiovascular disease not in linkage disequilibrium with any one of the markers set forth in Table 10.

39. The method of any of the claims 1-6, comprising determining the presence or absence of at least one allele in at least two polymorphic markers, wherein the presence of the at least one allele in the at least two polymorphic markers is indicative of an increased susceptibility to cardiovascular disease.

40. The method of any of the preceding Claims, further comprising analyzing non-genetic information to make risk assessment, diagnosis, or prognosis of the individual.

41. The method of Claim 40, wherein the non-genetic information is selected from age, gender, ethnicity, socioeconomic status, previous disease diagnosis, medical history of subject, family history of cardiovascular disease, biochemical measurements, and clinical measurements.

42. The method of any of the Claims 34 - 41, further comprising calculating combined risk.

43. The method according to any of the preceding Claims, wherein linkage disequilibrium is characterized by a particular numeric cutoff value for the linkage disequilibrium measures r^2 and/or |D'|.

44. The method according to Claim 43, wherein wherein linkage disequilibrium is characterized by a cutoff value for r^2 of greater than 0.2.

45. The method according to Claim 43 or Claim 44, wherein linkage disequilibrium is characterized by a cutoff value for r^2 of greater than 0.5.

46. The method according to Claim 42, wherein linkage disequilibrium is characterized by a cutoff value for r^2 of greater than 0.2 and/or |D'| of greater than 0.8.

47. The method according to any of the preceding Claims, wherein the Cardiovascular disease is at least one of Myocardial Infarcation, Coronary Artery Disease, Percutaneous Transluminal Coronary Angioplasty (PTCA), Coronary Artery Bypass Surgery (CABG), Restenosis, Periperal Arterial Disease, Stroke, Abdominal Aorta Aneurysm and Intracranial Aneurysm.

48. The method according to Claim 47, wherein the Cardiovascular disease is at least one of Myocardial Infarction, Coronary Artery Disease, Restenosis, Intracranial Aneurysm and Abdominal Aorta Aneurysm.

49. The method according to Claim 47, wherein the Stroke is Large Artery Atherosclerotic Stroke or Cardiogenic Stroke.

50. The method according to Claim 47 or 48, wherein Myocardial Infarction is an early onset Myocardial Infarction.

51. The method according to Claim 50, wherein the Cardiovascular Disease is Myocardial Infarction and/or Coronary Artery Disease with an early onset before age 50 for males and 60 for females.

52. A kit for assessing susceptibility to cardiovascular disease in a human individual, the kit comprising reagents for selectively detecting at least one allele of at least one polymorphic marker in the genome of the individual, wherein the polymorphic marker is selected from the markers set forth in Tables 10, and markers in linkage disequilibrium therewith, and wherein the presence of the at least one allele is indicative of a susceptibility to cardiovascular disease.

53. The kit of Claim 52, wherein the at least one polymorphic marker is selected from rs10811650, rs10116277, rs1333040, rs10738607, rs4977574, rs6475608, D9S1870, rs2383207, rs1333045, rs1333046, rs10757278 and rs1333048.

54. The method kit according to Claim 52 or Claim 53, wherein the at least one polymorphic marker is selected from rs1333040, rs10116277, rs2383207 and rs10757278.

55. The kit according to Claim 55, wherein the at least one polymorphic marker is selected from rs10757278 and markers in linkage disequilibrium therewith.

56. The kit according to any of the Claims 52-55, wherein linkage disequilibrium is characterized by a particular numeric cutoff value for the linkage disequilibrium measures r^2 and/or |D'|.

57. The kit according to Claim 56, wherein wherein linkage disequilibrium is characterized by a cutoff value for r^2 of greater than 0.2.

58. The kit according to Claim 56 or Claim 57, wherein linkage disequilibrium is characterized by a cutoff value for r^2 of greater than 0.5.

59. The method according to Claim 56, wherein linkage disequilibrium is characterized by a cutoff value for r^2 of greater than 0.2 and/or |D'| of greater than 0.8.

60. The kit according to any of the Claims 52 - 59, wherein the reagents comprise at least one contiguous oligonucleotide that hybridizes to a fragment of the genome of the individual comprising the at least one polymorphic marker, a buffer and a detectable label.

61 The kit according to any of the Claims 52 - 60, wherein the reagents comprise at least one pair of oligonucleotides that hybridize to opposite strands of a genomic nucleic acid segment obtained from the subject, wherein each oligonucleotide primer pair is designed to selectively amplify a fragment of the genome of the individual that includes one polymorphic marker, and wherein the fragment is at least 30 base pairs in size.

62. The kit according to Claim 60 or 61, wherein the at least one oligonucleotide is completely complementary to the genome of the individual.

63. The kit according to any of the Claims 60 - 62, wherein the oligonucleotide is about 18 to about 50 nucleotides in length.

64. The kit according any of the Claims 60 - 63, wherein the oligonucleotide is 20-30 nucleotides in length.

65. The kit according to any of the Claims 52 - 64, wherein the kit comprises:

a. a detection oligonucleotide probe that is from 5-100 nucleotides in length;

b. an enhancer oligonucleotide probe that is from 5-100 nucleotides in length; and

c. an endonuclease enzyme;

wherein the detection oligonucleotide probe specifically hybridizes to a first segment of the nucleic acid whose nucleotide sequence is set forth in SEQ ID NO:94; and

wherein the detection oligonucleotide probe comprises a detectable label at its 3' terminus and a quenching moiety at its 5' terminus;

wherein the enhancer oligonucleotide is from 5-100 nucleotides in length and is complementary to a second segment of the nucleotide sequence that is 5' relative

to the oligonucleotide probe, such that the enhancer oligonucleotide is located 3' relative to the detection oligonucleotide probe when both oligonucleotides are hybridized to the nucleic acid;

wherein a single base gap exists between the first segment and the second segment, such that when the oligonucleotide probe and the enhancer oligonucleotide probe are both hybridized to the nucleic acid, a single base gap exists between the oligonucleotides; and

wherein treating the nucleic acid with the endonuclease will cleave the detectable label from the 3' terminus of the detection probe to release free detectable label when the detection probe is hybridized to the nucleic acid.

66. Use of an oligonucleotide probe in the manufacture of a reagent for diagnosing and/or assessing susceptibility to cardiovascular disease in a human individual, wherein the probe hybridizes to a segment of a nucleic acid whose nucleotide sequence is set forth in SEQ ID NO:94, wherein the probe is 15-500 nucleotides in length.

67. A computer-readable medium on which is stored:

a. an identifier for at least one polymorphic marker;

b. an indicator of the frequency of at least one allele of said at least one polymorphic marker in a plurality of individuals diagnosed with cardiovascular disease; and

c. an indicator of the frequency of the least one allele of said at least one polymorphic markers in a plurality of reference individuals;

wherein the at least one polymorphic marker is selected from the polymorphic markers set forth in Table 10, and polymorphic markers in linkage disequilibrium therewith.

68. The medium according to Claim 67, wherein the polymorphic marker is selected from rs10811650, rs10116277, rs1333040, rs10738607, rs4977574, rs6475608, D9S1870, rs2383207, rs1333045, rs1333046, rs10757278 and rs1333048, and markers in linkage disequilibrium therewith.

69. The medium according to claim 67 or 68, further comprising information about the ancestry of the plurality of individuals.

70. An apparatus for determining a genetic indicator for cardiovascular disease in a human individual, comprising:

a computer readable memory; and

a routine stored on the computer readable memory;

wherein the routine is adapted to be executed on a processor to analyze marker and/or haplotype information for at least one human individual with respect to at least one polymorphic marker selected from the markers set forth in Table 10, and markers in linkage disequilibrium therewith, and generate an output based on the marker or haplotype information, wherein the output comprises a risk measure of the at least one marker or haplotype as a genetic indicator of cardiovascular disease for the human individual.

71. The apparatus of Claim 70, wherein the routine further comprises an indicator of the frequency of at least one allele of at least one polymorphic marker or at least one haplotype in a plurality of individuals diagnosed with cardiovascular disease, and an indicator of the frequency of at the least one allele of at least one polymorphic marker or at least one haplotype in a plurality of reference individuals, and wherein a risk measure is based on a comparison of the at least one marker and/or haplotype status for the human individual to the indicator of the frequency of the at least one marker and/or haplotype information for the plurality of individuals diagnosed with cardiovascular disease.

72. The apparatus according to Claim 70 or 71, wherein the at leat one polymorphic marker is selected from rs10811650, rs10116277, rs1333040, rs10738607, rs4977574, rs6475608, D9S1870, rs2383207, rs1333045, rs1333046, rs10757278 and rs1333048, and markers in linkage disequilibrium therewith.

73. The apparatus or medium according to any of the claims 67 – 72, wherein linkage disequilibrium is characterized by numerical values of r^2 of at least 0.2 and/or values of |D'| of at least 0.8.

74. The apparatus of any of the Claims 70 - 73, wherein the risk measure is characterized by an Odds Ratio (OR) or a Relative Risk (RR).