Analysis of *EDNRA*: a gene connected to atherosclerosis

Ragnar Óli Vilmundarson

Thesis for the degree of Master of Science
The University of Iceland
Faculty of Medicine
School of Health Sciences

HÁSKÓLI ÍSLANDS

Greining á *EDNRA*: geni tengdu æðakölkun

Ragnar Óli Vilmundarson

Ritgerð til meistaragráðu í Líf- og læknavísindum

Umsjónarkennari: Albert V. Smith

Meistaranámsnefnd: Thor Aspelund og Jóhanna Jakobsdóttir

Læknadeild

Heilbrigðisvísindasvið Háskóla Íslands

Febrúar 2014

Analysis of *EDNRA*: a gene connected to atherosclerosis

Ragnar Óli Vilmundarson

Thesis for the degree of Master of Science

Supervisor: Albert V. Smith

Master's committee: Thor Aspelund and Jóhanna Jakobsdóttir

Faculty of Medicine

School of Health Sciences

The University of Iceland

February 2014

Prentun: Háskólaprent

Reykjavík, Ísland 2014

Ágrip

Hjarta og æðasjúkdómar þar sem æðakölkun er undirliggjandi ástæða eru helsta dánarorsök um allan heim. Æðakölkun er sjúkdómur æðaveggjarins og er þróun hans vel skilgreind. Uppsöfnun kólesteróls og dauðra fruma í æðaveggnum sem er meðal annars afleiðing af samsöfnun á átfrumum í innlagi æðaveggjarins sem leiðir til myndunar æðakölkunarskella (e. plaque). Þar sem æðakölkun er kerfissjúkdómur og finnst í flestum slagæðum líkamans þá má greina hvort einstaklingur hafi æðakölkun með B-mode ómskoðun til dæmis á hálsslagæðum. Mæling á tveim innstu lögum æðaveggjarins kölluð innlags og miðlags þykkt (e. intima media thickness, IMT) hefur verið notuð sem mælikvarði á einkennalausa æðakölkun og verið sett fram sem aðferð til að spá fyrir um áhættu einstaklinga á að fá hjarta eða heilaáföll. Önnur mæling er meira afgerandi tengd æðakölkun og með meira forspárgildi en það er mæling á æðaskellunum sjálfum í æðunum. Rannsóknir á lykil genum (e. candidate genes) hafa fundið fjölmörg tengsl við báðar þessar mælingarnar í manninum. Niðurstöðurnar hafa hins vegar verið misvísandi. CHARGE vinnuhópurinn, sem Öldrunarrannsókn Hjartaverndar (AGES Reykjavik study) tilheyrir lagði saman gögn frá mörgum langvarandi framskyggnum rannsóknum og framkvæmdi safngreiningu til að leita að stöðum í erfðamenginu sem væru tengd æðakölkun í hæálsslagæðum. Nokkur erfðamörk (e. SNPs) tengdust báðum hálsslagæðasvipgerðunum. *EDNRA* setið reyndist sérstaklega áhugavert. Það sýndi tengsl við báðar hálsslagæðasvipgerðirnar auk þess sem að tengjast kransæðasjúkdómi. Það skráir fyrir æðaþelsviðtaka A (endothelin type A receptor) sem bindur endothelin 1 sem veldur samdrætti æða. Saman valda þau samdrætti sléttra vöðvafruma í æðaveggnum og hefur það ferli verið margrannsakað með tilliti til lykilgena. Þar sem erfðamörkin sem fundust í CHARGE rannsókninni höfðu enga þekkta virkni má ályktа að erfðamörkin gætu verið í tengslaójavfnvægi (e. linkage disequilibrium, LD) við annan eða aðra starfræna erfðabreytileika. Hér í þessari rannsókn var notuð 1000 erfðamengja (1000 Genomes) tilreiknun (e. imputation) ásamt setraða (haplotype) rannsóknum til að skilgreina frekar CHARGE *EDNRA* aðal erfðamarkið (rs 1878406) til að reyna að finna starfræna(n) erfðabreytileika.

Rauntíma PCR mæliaðferð var hönnuð til að greina aðalerfðamark EDNRA gensins og voru 5,521 einstaklingar frá Áhættuþáttakönnun Hjartaverndar (REFINE Reykjavik study) erfðamarkagreindir. Aðhvarfsgreining var notuð til að kanna samband milli aðal erfðamarks og hálsslagæðasvipgerða; IMT og æðaskella. LD myndir voru útbúnar. Fyrst fyrir allt EDNRA genið þar sem Haploview forritið var notað og síðan fyrir þau erfðamörk sem voru í sterku LD við aðalerfðamarkið þar sem forritið SNAP var notað til að útbúa tengslamyndir. Erfðamarka aðhvarfsgreining var framkvæmd á erfðamörkum sem voru arfgerðagreind með beinum hætti sem og á 1000 erfðamengja tilreiknuðu erfðamörkunum í Öldrunarrannsókn Hjartaverndar (3,219 einstaklingar). Þar voru könnuð tengsl milli framangreindra erfðamarka og hálsslagæðasvipgerðanna. Tilreiknuð og beint arfgerðargreind erfðamörk voru innan svæðis chr4:148,392,069-148,476,106 (Hg19) sem nær yfir 84kb og inniheldur allt *EDNRA* genið auk um 10kb fyrir framan og aftan genið. Tilreiknuðu erfðamörk með minni gæði tilreiknunar en r^2=0.6, voru sigtuð út fyrir útreikninga. FastPHASE forritið var notað til að útbúa setraðir fyrir 11 erfðamörk í EDNRA geninu sem vor erfðamerkt með beinum hætti í Öldrunarrannsókninni. Setraðir voru útbúnar og samband milli þeirra og æðaskella kannað með aðhvarfsgreiningu. Fylgni milli erfðamarka og milli

erfðamarka og setraða var fundin með Pearson's fylgnistuðlum. Virknitenging (e. functional annotation) fyrir tölfræðilega marktæk erfðamörk var gerð með RegulomeDB og SNPnexus.

Samband aðalerfðamarksins og bæði æðaskella (OR= 1.25, 95% CI; 1.05-1.49) og IMT (β=0.0126, 95% CI; 0.00588-0.0193) var sannreynt í óháðu Íslensku þýði, Áhættuþáttakönnun Hjartaverndar. LD rannsókn á *EDNRA* geninu sýndi að aðalerfðamarkið var innan tengslasvæðis sem innihélt tvær fyrstu tjáningaraðir *EDNRA* gensins sem og hluta af innröð 2 og svæðis fyrir framan *EDNRA* genið. Frekari fíngreining á LD með LD myndum og nálægðar (e. proxy) LD prófunum fann að röðin fyrir framan nær 28kb framfyrir *EDNRA* genið og 23kb inní genið. Aðhvarfsgreiningin fann að 15 erfðamörk voru marktækt tengd hálsæðaskellum OR, 95%CI (alpha=0.01) og 63 erfðamörk marktækt tengd IMT (alpha=0.01). Þrettán erfðamörk voru marktækt tengd bæði hálsslagæðaskellum og IMT. Ein setröð fyrir *EDNRA* tengdist mögulega við hálsslagæðaskellu. Að auki fannst vísbending um aðra tengingu við hálsslagæðaskellur í innröð 6 í *EDNRA* geninu sem þarfnast frekari rannsóknar. Setraðarannsókn fann að þessi nýju tengsl í innröð 6 voru sterkt tengd við erfðamark sem sýndi sterkara OR 1.95 (95% CI; 1.01-3.76) heldur en aðalerfðamarkið OR 1.30 (95% CI; 1.10-1.53) fyrir hálsslagæðaskellur. Rannsóknin greindi því setröð sem er mögulega tengd við hálsslagæðaskellur og fangar bæði tengslin í innröð 6 sem og erfðamarkið fyrir framan *EDNRA* genið. Þessi rannsókn fann að aðalerfðamarkið (rs 1878406) er í sterku LD við annað erfðamark (rs6841581) þar sem áður hefur verið sýnt fram á tjáningamun á milli seta í *EDNRA* geninu. Þetta erfðamark hafði sambærileg tenglsl við bæði hálsslagæða - skellu og þykknun.

Abstract

Cardiovascular disease (CVD) is the leading cause of death worldwide and an underlying cause of CVD is atherosclerosis. Atherosclerosis is a disease of the arterial wall and its progression is well characterized. The build-up of cholesterol and dead cells in the arterial wall, which is the result of, among other things, lipid carrying macrophages collecting in the intima of the arterial wall and lead to a cascade of cellular events which lead to the formation of an atheromatous plaque, more commonly referred to as a plaque. As atherosclerosis is a systemic disease, found in numerous different arteries throughout the body, then it is possible to identify whether an individual has atherosclerosis through the B-mode ultrasonography technique that allows for the imaging of, for example, the carotid artery. A measure of two layers of the arterial wall, termed the intima-media thickness, has long been used as a measure of subclinical atherosclerosis and has been held as a measurement that can be used to predict for an individual's risk of future cardiovascular events. The other measure is more representative of the progression of atherosclerosis as it measures the level of plaque within an artery. The candidate gene studies, have found numerous associations between these two measures of atherosclerosis and variations within the human genome, however, the results have proved to be inconsistent. The CHARGE consortium, including the AGES Reykjavik study incorporated numerous long-term prospective population-based studies from around the world and conducted a meta-analysis of GWASs with the aim of finding those sites of the human genome that were associated with atherosclerosis of the carotid artery. There were numerous loci that associated with the carotid atherosclerosis phenotypes but of particular interest was the *EDNRA* locus. It associated with both of the carotid atherosclerotic phenotypes as well as CAD, and it encodes the endothelin type A receptor which binds endothelin-1, a vasoconstrictor. Together they are responsible for the contraction of the smooth muscle cells of the arterial wall and the endothelin pathway has been the focus of many candidate gene studies in the past few decades. As the SNP identified in the CHARGE analysis had no known function it was posited that the variant could be in linkage disequilibrium (LD) other truly functional variant(s). This study aimed to use 1000 Genome imputations along with haplotype analysis to refine the CHARGE *EDNRA* lead SNP (rs1878406) signal and attempt to identify the functional variants.

A real-time PCR assay was designed for the lead *EDNRA* SNP and 5,521 individuals of the REFINE Reykjavik Study were genotyped. Regression analysis was performed to test the association between the lead SNP and carotid atherosclerotic measures, common cIMT and carotid plaque in addition to coronary events. LD plots were created, first for the entire *EDNRA* gene using Haploview and then for those SNPs in strong LD with the lead SNP, using SNAP to create a LD regional plot. Single-point SNP regression analysis was then performed on both genotyped and 1000 Genome imputed SNPs in the AGES Reykjavik Study cohort of 3,219 individuals, testing for the association between said SNPs and the carotid atherosclerotic measures. Imputed and genotyped SNPs tested were within the range of chr4:148,392,069-148,476,106 (Hg19) which amounts to approximately 84kb and encompasses the entire *EDNRA* gene in addition to its 10kb upstream and downstream regions. The Imputed SNPs with imputation quality below r^2=0.6 were filtered out prior to analysis. The

fastPHASE program was used to create phased haplotypes from the 11 *EDNRA* SNPs genotyped in the AGES Reykjavik Study. Haplotype counts, similar to SNP dosages/counts, were created and the association between haplotypes and carotid plaque was tested in a single-point manner using regression analysis. Correlation between SNPs, and also between SNPs and haplotypes, was calculated through the use of Pearson's r correlation coefficient. Functional annotation of significantly associated SNPs was carried out through the RegulomeDB resource as well as SNPnexus.

The lead SNP's association with both carotid plaque (OR=1.25, 95% CI; 1.05-1.49) and common cIMT (β=0.0126, 95% CI; 0.00588-0.0193) was confirmed in an independent Icelandic cohort, REFINE. The LD structure of *EDNRA* analysis showed that the lead SNP lies within an LD block that covers the first two exons of *EDNRA* in addition to a part of intron 2 and also an upstream region of *EDNRA*. Further fine mapping of LD through regional LD plots and proxy LD testing showed that the upstream region extends to 28kb upstream of *EDNRA* and 23kb into the gene itself. The single-point regression analysis results showed 15 SNPs to be significantly associated with carotid plaque (alpha=0.01) and 63 SNPs significantly associated with common cIMT (alpha=0.01). There were 13 SNPs that were significantly associated with both carotid plaque and common cIMT. One *EDNRA* haplotype was suggestively associated with carotid plaque. In addition, there was an additional signal for carotid plaque in intron 6 of *EDNRA* that requires further investigation. The haplotype analysis showed this additional signal in intron 6 correlates highly with SNPs showing a higher OR 1.95 (95% CI; 1.01-3.76) than the lead SNP OR 1.30 (95% CI; 1.10-1.53) for carotid plaque. This study has identified a haplotype that is suggestively associated with carotid plaque and captures both the additional signal in intron 6 and the lead SNP signal in the upstream region of *EDNRA*. This study found that the lead SNP is in strong LD with a SNP (rs6841581) that has previously been shown to have an allelic difference in the transcription of *EDNRA*. This SNP shared the same signal for both the carotid plaque and common cIMT atherosclerotic phenotypes and is therefore a strong candidate for the SNP in EDNRA that mediates the function associated with carotid atherosclerosis.

Acknowledgements

It has been a great experience working at the Icelandic Heart Association and I have had an opportunity to work with, and under the supervision of, people that are at the forefront of scientific discovery, both within Iceland and also the greater international scientific community. I would like to thank my supervisor Albert V. Smith for his guidance throughout the project in addition to insightful comments and a willingness to answer any and all questions. Thor Aspelund has been of great help to me as the project developed and has always been open to assist me with any and all matters regarding the project, or general questions regarding the field, and for that I will always be grateful. Jóhanna Jakobsdóttir's comments on this thesis and the project as a whole have helped bring into focus the aspects that matter most. The comments always lead to a more streamlined work process and have allowed me to save energy for critical parts of the thesis and analysis. It has truly been a pleasure to work with her and I'm sure she will be an integral part of the Icelandic Heart Association in the years to come. Elena Losievskaja provided me with essential help when I was first starting out on the single-point regression analysis and helped me greatly to understand the functions and scripts I needed to use, for that I am very thankful. Valur Emilsson provided pivotal insight, advice and guidance in the beginning stages of the project and helped shape the direction that the analysis went in, and for that I extend my sincere gratitude. Valur was always willing to discuss the project as well as share his extensive knowledge with me and that was a large part of what made my work at the Icelandic Heart Association enjoyable. None of the research carried out here would be possible if it was not for the contribution of the individuals that have taken part in both the AGES Reykjavik Study and the REFINE Reykjavik Study. I am indebted to them and grateful for their participation.

And to my father, Vilmundur Guðnason, who was the driving force of this project, its ideas and direction, I will always be eternally grateful for the opportunity he gave me to work alongside him and the great people at the Icelandic Heart association.

The AGES Reykjavik Study was funded by the National Institute on Aging, the National Institute on Health, the Icelandic Parliament and the Icelandic Heart Association. The REFINE Reykjavik Study is funded by the Icelandic Heart Association, the post genomics-biomedicine/nanoscience-nanotechnology research program grant from the Icelandic Centre for Research (RANNIS) 72012001 and from the Icelandic Research Fund 90452011.

Declaration

The large majority of work carried out in this thesis was my own. This included the genotyping work carried out in the lab for the lead SNP and the DNA extraction from several hundred REFINE Reykjavik Study samples. The data handling and analysis performed in a combination of SAS and R was carried out by myself under the expert supervision of Albert V. Smith and Thor Aspelund. The imputation of variants using the 1000 Genome Project data was carried out by Albert V. Smith and he created the datasets of imputed SNPs that I used in this project. The design of the lead SNP assay was carried out by Haukur Guðnason.

Table of Contents

List of figures

List of tables

List of abbreviations

1000G – 1000 Genomes Project (data)

AGES – Age, Gene/Environment Susceptibility

BP – Blood Pressure

CABG – Coronary Artery Bypass Graft

CAC – Coronary Artery Calcium

CAD – Coronary Artery Disease

CARDIoGRAM - Coronary Artery Disease Genome-Wide Replication and Meta-Analysis

CE – Coronary Events

CEQ - Coronary Event by Questionnaire

CHARGE – Cohorts for Heart and Aging Research in Genomic Epidemiology

CHD – Coronary Heart Disease

Chr – Chromosome

CI – Confidence Interval

cIMT – carotid Intima-Media Thickness

CT – Computerized Tomography

CVD – Cardiovascular Disease

EDNRA – Endothelin Receptor Type A

GWAS – Genome-Wide Association Study

HDL – High Density Lipoprotein

IHA – Icelandic Heart Association

kb – kilobase(s)

LD – Linkage Disequilibrium

LDL – Low Density Lipoprotein

MAF – Minor Allele Frequency

MI – Myocardial Infarction

OR - Odds Ratio

oxLDL – oxidized Low Density Lipoprotein

PCI – Percutaneous Coronary Intervention

PCR – Polymerase Chain Reaction

REFINE – Risk Evaluation For Infarct Estimates

SD – Standard Deviation

SE – Standard Error

SNP – Single Nucleotide Polymorphism

1 Introduction

Each year cardiovascular disease (CVD) claims the lives of millions around the world. The leading cause of death worldwide, as of 2011, was ischemic heart disease, often referred to as coronary artery disease (CAD), which claimed the lives of approximately 7 million people (1). After ischemic heart disease the next largest cause of death worldwide in 2011 was stroke, which caused 6.2 million deaths (1). In Iceland CVD is also the leading cause of death with 38% of the deaths in 2006 caused by the disease (2). Both of these cardiovascular events that are claiming millions of lives each year throughout the world share a common underlying disease, atherosclerosis.

1.1 Atherosclerosis

Atherosclerosis is systemic disease of the cardiovascular system, specifically the walls of arteries. It presents as an atheroma, which is the swelling of the arterial wall due to the build-up of lipid carrying macrophages covered with a fibrous cap. In the case of atherosclerosis the atheroma is referred to as an atheromatous plaque or, more commonly, simply plaque. In addition to the diseased state of the arterial wall, atherosclerosis can involve a hardening of the arterial wall caused by calcification which is sometimes seen in advanced plaques (3). The disease can occur in arteries such as the carotid and coronary, although it is often seen first in the largest artery in the body, the aorta (4). A typical healthy artery can be divided into three different layers (3), the tunica intima, tunica media and tunica adventitia, often referred to as the tunica externa, Figure 1. Tunica, meaning a coat or layer, is often left out of the equation and the different layers are referred to simply by the names that distinguish them: the intima, media and adventitia.

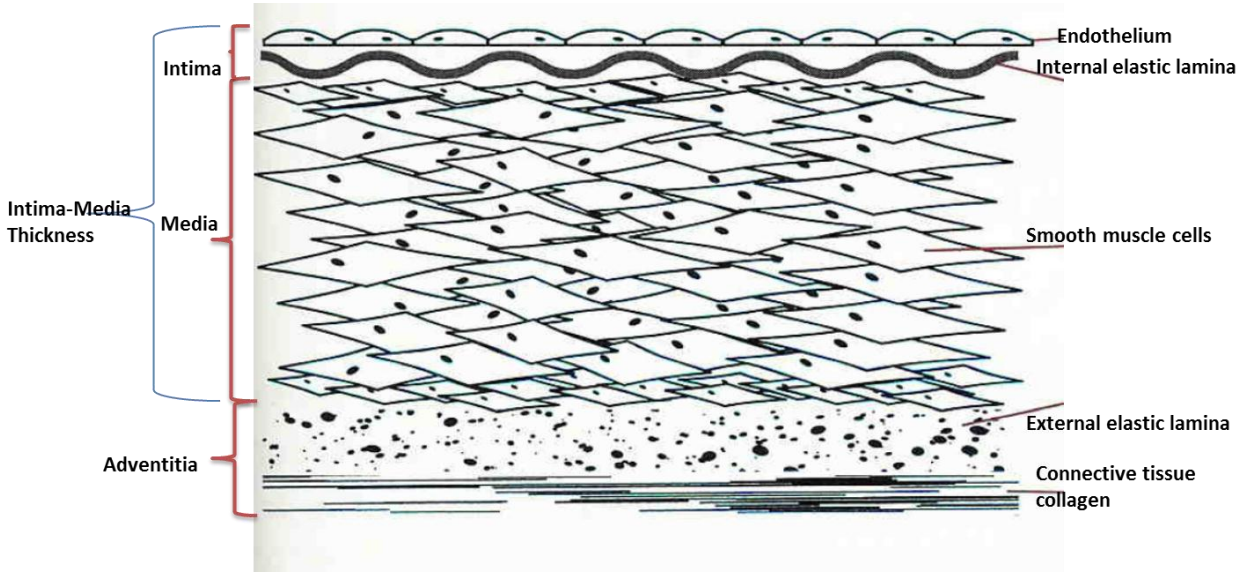

Figure 1. Illustration of a typical arterial wall. The intima, media and adventitia are labelled and the layers they are comprised of shown. The Intima-Media Thickness (IMT) is labelled and it is comprised of both the tunica intima and tunica media layers of the arterial wall. Modified from the book, Biology of Disease (3).

The intima is comprised of two distinct parts; first there is the endothelial cell layer which is in contact with the flow of blood through the lumen, and this endothelium gets its structural support from

the internal elastic lamina, which as the name indicates is an elastic tissue layer. The majority of the intima layer's width is due to the internal elastic lamina which can itself be separated into several layers. The internal elastic lamina varies a great deal in width between the smallest and largest arteries, with large arteries, such as the aorta, having quite thick internal elastic lamina layers. The next arterial wall layer, after the intima, is the media. The media is the middle layer of the arterial wall and is comprised of smooth muscle cells. It is the media layer that is the main determinant of the difference in size between the small and large arteries. The media layer differs between the larger and smaller arteries with regard to elastic lamellae as well. Larger arteries have a higher number of elastic lamellae, which are comprised of sheets of elastic fibres. The measurement of both the intima and media layers of an artery is often referred to as the intima-media thickness. The outermost layer of the arterial wall is the adventitia which is comprised of the external elastic lamina layer and a layer of collagen fibres that serves as a connective tissue layer bringing stability to the blood vessel through connection to organs in close proximity.

1.2 Atherosclerosis progression over time

The progression of atherosclerosis, called atherogenesis, involves plaque development in the arterial wall over an individual's lifetime (5, 6). The first stage of atherogenesis is damage to the intima of the arterial wall, specifically the endothelial cell layer that is in contact with the blood flowing though the lumen. This damage can be exacerbated and influenced by a wide range of irritants, such as toxins from cigarette smoke, hypertension, hyperlipidaemia (excess of LDL) and irritants that cause inflammation. All of these are well established risk factors for atherosclerosis and cardiovascular disease (7, 8). The damage to the endothelium from these irritants makes the layer more permeable and allows the LDL molecules to traverse the cell layer and make their way into the intima of the arterial wall. Whilst this is occurring, the irritated endothelial cells begin to produce leukocyte adhesion molecules which are used to recruit monocytes from the lumen (9). The recruited monocytes migrate into the intima and become activated, leading to their change into macrophages. It is possible that oxidized LDL (oxLDL) could be the cause of the leukocyte adhesion molecule expression (5, 6). The macrophages, once in the intima, express scavenger receptors which bind the oxLDL and the macrophages then draw the oxLDL molecules into themselves and do so to the extent that their appearance, seen through a microscope, becomes foam-like, thereby giving them their name, foam cells. The macrophages that have been loaded with oxLDL and reside in the intima present as fatty streaks along the inside of the arterial wall. When atherosclerosis has reached this stage it is still asymptomatic and these fatty streaks are common in the general population and found in individuals of all ages, even children (6).

These foam cells constitute the beginnings of atherosclerotic plaque formation, along with smooth muscle cells that are recruited from the media layer of the arterial wall and make their way into the intima (5, 6, 9). The smooth muscle cells that are found in the plaque within the intima are no longer functioning muscle cells and they provide no contraction as the typical smooth muscle cells do that are found within the media layer of an artery. It is the smooth muscle cells that have become activated, and are a part of the plaque, that thrive and multiply in addition to being responsible for the creation of

new proteins, such as collagen and elastin. The smooth muscle cells, and the complex of proteins expressed by them, form a hard layer called a fibrous cap. This cap covers what is called the plaque, made up of both living and dead foam cells, the dead ones having either gone through programmed cell death (called apoptosis) or having been starved of oxygen due to their location within the plaque. The dead foam cells release their contents into the extracellular matrix that constitutes the plaque, and these dead cells along with their released lipids make up the necrotic core found in advanced atherosclerotic plaques. For an overview of atherogenesis see Figure 2.

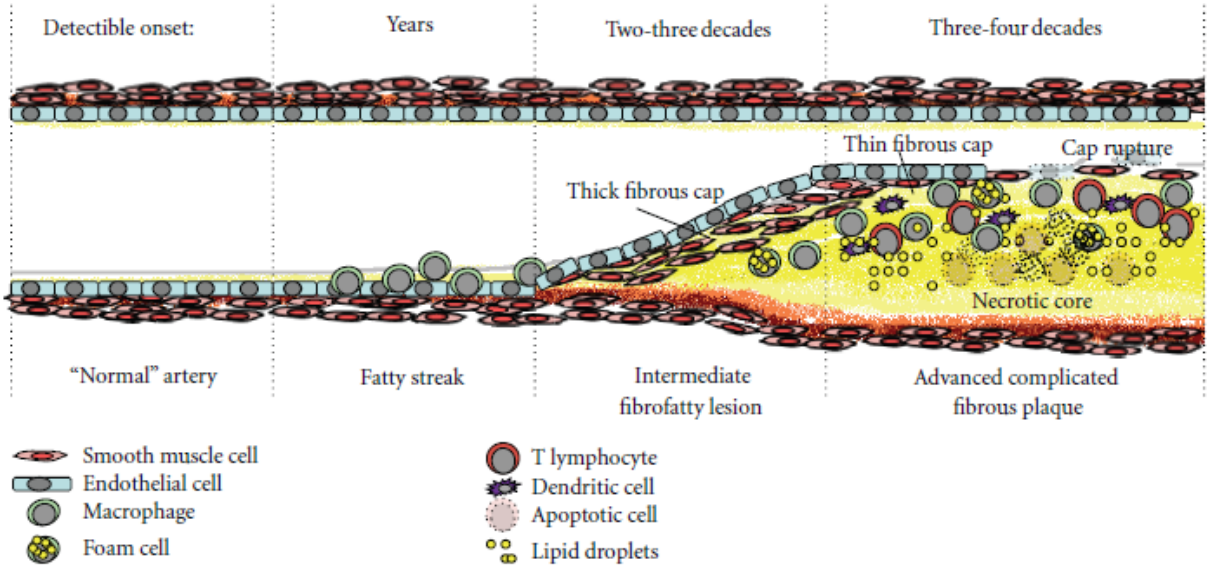

Figure 2. Progress of atherosclerosis over time, figure taken from Autieri 2012 (6). The LDL particles move into the intima of the arterial wall, there they become oxidized. The endothelial cells produce leukocyte adhesion molecules which are used to recruit the monocytes in the lumen of the artery. The monocytes then migrate into the intima and become activated macrophages. The macrophages express scavenger receptors which bind oxidized LDL and mediate its uptake. When the macrophages have taken up the oxidized LDL they are unable to transition out of the intima layer and become stuck there. Macrophages in that state are referred to as foam cells. Smooth muscle migration occurs once foam cells are present and the resulting collection of smooth muscle cells and foam cells are referred to as plaque (9).

1.3 Consequences of atherosclerotic plaque rupture

Once the plaque has reached an advanced stage the most significant danger to an individual's health is not the stenosis that the plaque causes and the consequences of that, but the increased risk that an advanced plaque has of rupturing (5, 6, 9). The consequences of plaque rupture can best be described with an example of a location in the cardiovascular system that often sees plaque formation. One such location is the site of bifurcation in the carotid artery where the common carotid artery splits into two branches, the internal and external carotid arteries. The reason that this site, and sites like it, are prone to plaque formation is that bifurcations result in localized disturbance to the shear stress of blood and such disturbances have been shown to associate with the formation of atherosclerotic plaques (10). There have been, however, certain difficulties proving that disturbances in the lumen's geometry are responsible for the formation of atherosclerotic plaques, although certain progress is being made. For example, a recent study showed that the bifurcation of the carotid artery can be used as a weak independent indicator for future changes in the arterial wall of the carotid, specifically with

regard to wall thickening (11). The difficulties in finding consistent association between vascular remodeling and disturbances in blood flow have been posited to relate to the inherent differences between the way studies both measure flow disturbance and also where in the artery the thickness of the wall is measured (11). However, those difficulties aside, the sites of bifurcation are more likely to see plaque development and it's the more advanced plaques that are of significant risk. For example, if an advanced plaque in the carotids were to rupture the body's response would be to seal the wound through clotting of the blood. The formed blood clot is called a thrombus and the overall process of its formation is called thrombosis (3). There are two possible outcomes from the formation of a thrombus. One possible outcome is that the thrombus grows so large than it blocks the artery completely preventing blood from flowing through it, Figure 3a. If a thrombus blocked off the internal carotid artery, for example, then it could result in an ischemic stroke if the body is unable to compensate for the loss of blood flow to the brain. On the other hand if a thrombus formed in a coronary artery it would cause reduced blood flow to the heart and could lead to a myocardial infarction (MI), more commonly referred to as a heart attack. Another possible outcome of the formation of a thrombus is that it breaks up and parts of it are carried off by the blood stream to other parts of the body, Figure 3b. When a thrombus detaches and becomes free moving then it is called an embolus. An embolus can travel through the blood stream to smaller arteries, such as those found in the brain, and cause a blockage there. A blocked artery in the brain could lead to reduced blood flow to the brain and if the body is unable to compensate then it would result in an ischemic stroke. Therefore, ischemic strokes are the result of reduced or blocked blood flow to the brain and can, in the context of atherosclerosis, be caused by either a thrombus blocking an artery that leads to the brain or due to an embolus that travels to a cerebral artery and causes a blockage there.

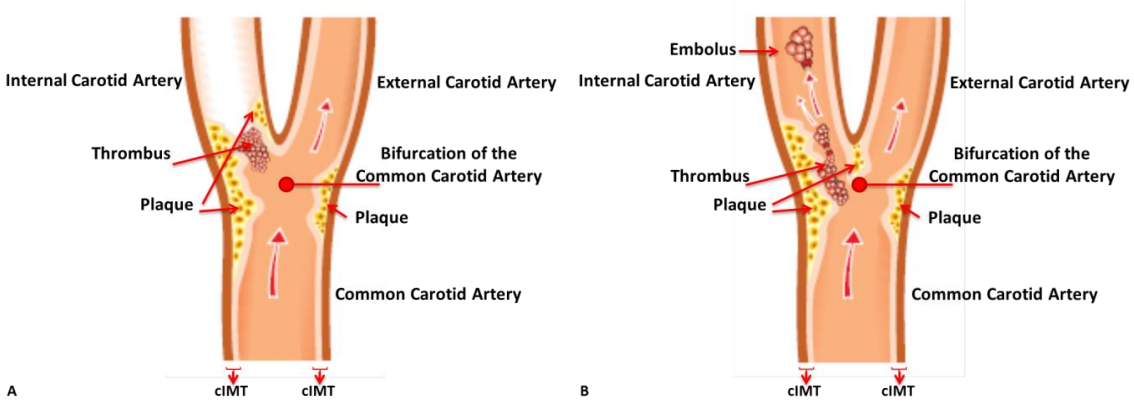

Figure 3. The common, internal and external carotid arteries. Both figures A and B show possible consequences of plaque formation in the carotid artery. A: Part of the blood clot (thrombus) that has formed around the plaque damaged arterial wall detaches and thus an embolus is created. B: The thrombus grows larger and completely occludes the internal carotid artery, blocking blood flow to the brain. The cIMT measurement is labelled in both pictures. Direction of blood flow indicated with arrows. Modified from (12).

1.4 Measuring atherosclerosis

Atherosclerosis can be prevented from progressing further when it is in its early stages. This can be accomplished through the lowering of an individual's cholesterol levels, either through a change in diet or through the use of medication, such as statins. As advanced atherosclerosis can lead to

catastrophic events, such as an MI or stroke, it is therefore important to be able to assess whether individuals are at risk of developing atherosclerosis in order to be able to prevent the development of the disease before the damage caused becomes irreversible. This can be accomplished through risk calculation using standard risk factors for CVD (7), however, risk calculation may be improved through the incorporation of genetic risk factors. Therefore, in research attempting to find genetic associations with atherosclerosis it is pivotal to have markers that are indicatory of the disease.

Through imaging of arteries it is possible to capture atherosclerosis at various stages of progression, all the way from the early thickening of the intima-media to a clearer infiltration of the arterial wall by atherosclerosis measured as a thickening of the IMT or to clearer manifestation of atherosclerosis as atherosclerotic plaque. The most commonly used method to image the arteries is B-mode ultrasonography and the most common location of this imaging is the carotid artery. This measurement captures the carotid intima-media thickness (cIMT) as well as carotid plaque and is typically carried out in the common carotid artery due to the comparative ease of access of that carotid artery over the internal and external arteries (13). A measure of cIMT where the atherosclerotic process is not occurring has little to no predictability of future cardiovascular events (14). When cIMT is measured in areas of plaque, or where the plaque process is underway, then the predictability of future cardiovascular events is increased considerably (15). The measurement of cIMT differs between different studies and it is that difference that has led to controversy over the measurement's use as an indicator of atherosclerosis and predictor of future cardiovascular events, such as MI or stroke (15). When the measurement is carried out to identify plaque and plaque area then the greatest predictability of future events is seen (14). There is considerable correlation between atherosclerosis measured in the carotids and atherosclerosis measured elsewhere, for instance in the coronaries (16). This is indicatory of atherosclerosis being a systemic disease and as such, the measurement of atherosclerosis in one arterial location is sufficient to determine whether an individual is suffering from atherosclerosis (15, 17). This is important when addressing the contribution of risk factors, including genetic factors, to atherosclerosis and its consequences. Applying CT measurements to the coronaries can capture atherosclerosis as it can assess the level of calcification (18). As calcification is an indication of advanced atherosclerotic plaques then this can be an important quantification of how advanced the atherosclerosis in an individual has become (18). However, calcification is not an entirely reliable indicator to use for coronary artery atherosclerosis because it isn't present in all individuals with atherosclerotic plaques (19). Also, a drawback of CT scanning compared to the B-mode ultrasonography is that it is not possible to measure coronary calcification in younger individuals within population-based studies due to the level of radiation involved in the CT scanning process. CT imaging would only be performed in younger individuals if they have symptoms of atherosclerosis or coronary heart disease (CHD). Both IMT and plaque can also be measured in the femoral arteries using ultrasonography, however, this is uncommon in epidemiological studies. An overview of the different locations where atherosclerosis is measured and also the techniques used to quantify and measure atherosclerosis can be seen in Figure 4.

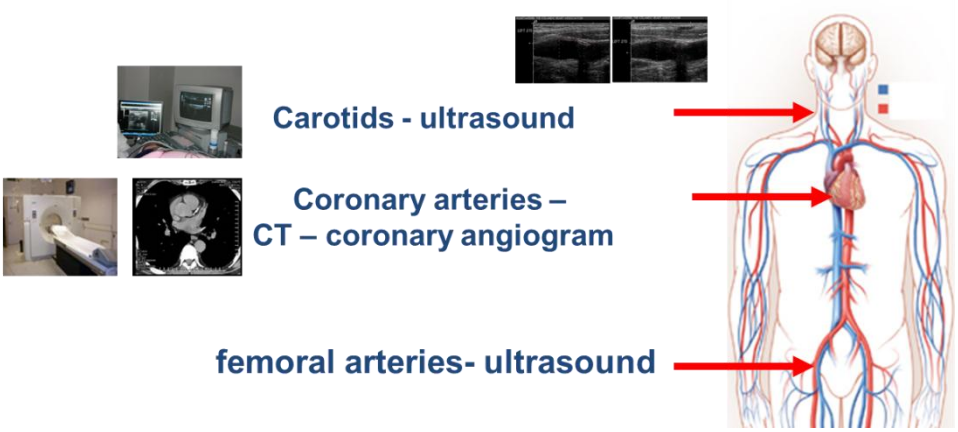

Figure 4. Atherosclerosis as a systemic disease. Measurement sites in the carotid, coronary and femoral arteries and the technique used for each site. Modified and used with the permission of the Icelandic Heart Association.

1.5 Genetic studies of carotid imaging phenotypes

There is ample evidence for the genetic contribution to atherosclerosis including cIMT and carotid plaque. For example, the heritability of carotid plaque was investigated in a recent international study of 192 monozygotic and 83 dizygotic twin pairs. It was found that there was a high heritability of 78% for the presence of carotid plaque (20). This was a considerably higher heritability value for the presence of carotid plaque than seen in previous studied, such as the Erasmus Rucphen family study which found a 28% heritability of carotid plaque (21) and the San Antonio family heart study which found a heritability value of 23% to 28% (22). It has been posited that this difference may be explained by methodological limitations (20).

In addition to heritability studies of carotid plaque, there have been heritability studies of cIMT as well. Those studies involving twins showed heritability values of 31% (23), 36% (24) and 59% (25) for cIMT, whereas the heritability in families recruited based on disease status found higher levels of heritability, ranging from 32% to 86%, reviewed in Manolio et al 2004 (26). In all of these studies the adjustment for external factors, such as CVD risk factors, had miniscule effect on the heritability of carotid IMT.

Numerous studies of the effects of genetic polymorphisms on carotid phenotypes have been carried out and have shown largely inconsistent small-effect findings, reviewed in Manolio et al 2004 (26). The realization that large numbers of individuals are needed to address genetic effects of moderate to small size has led to the rise of large international consortia where multiple cohorts from different countries around the world join forces and collaborate in large projects and meta-analysis. One such consortium is CHARGE, which stands for Cohorts for Heart and Aging Research in Genomic Epidemiology (27). There are five founding cohorts in the CHARGE consortium and they are all prospective studies that follow a population-based sample over a period of time and involve extensive phenotype measurements of all participating individuals (27). The data collected includes a multitude of risk factors for disease, disease endpoints, such as MI or stroke, and indicators of disease, such as the atherosclerosis traits of cIMT and carotid plaque. A strength of the CHARGE consortium is that all participating study individuals are of European ancestry and this reduces the risk

of false-positive results due to the effect that population stratification can have (28). Population stratification is the difference in allele frequencies that is often witnessed between populations of different ancestry and it can be a confounder when testing for association, although there are robust methods to correct for it when performing a GWAS (29).

A meta-analysis of genome-wide association studies (GWASs) was performed by the CHARGE consortium for the carotid plaque as well as common and internal cIMT phenotypes (28). In the face of multiple lines of evidence for the heritable nature of the three carotid phenotypes (28) there is still a lack of understanding when it comes to the genes or pathways involved in atherosclerosis progression, development and its pathology. Therefore, one motivation for the project was to clear up the inconsistent association between cIMT and genetic variation that has been previously shown through studies of cIMT candidate genes and genome-wide linkage scans (26, 28). The overall aim of the study can be said to be that of finding novel variants associated with carotid plaque and both internal and common cIMT. The discovery of any novel variants would deepen the understanding of potential genes and pathways involved in atherosclerosis. There were a total of nine different study cohorts that took part, giving a total sample size of 31,211 individuals in the meta-analysis, although the total number varied slightly depending on the trait being analysed.

There were three SNPs (rs17398575, rs445925 and rs1878406) that were shown through the meta-analysis to be associated with both the carotid plaque and common cIMT phenotype. The first, rs17398575, is located on the q22.3 band of chromosome 7 and lies approximately 96kb upstream of the *PIK3CG* gene. The second, rs445925, is located on the p12.32 band of chromosome 19 and is 2.3kb upstream of the *APOC1* gene. And finally rs1878406 which is located on the q31.22 band of chromosome 4 and is situated 8.4kb upstream of the *EDNRA* gene. Although the associations did pass the genome-wide significance threshold of 10^{-8}, each SNP reached genome-wide significance for only one of the two phenotypes and the association with the other phenotype was suggestive. In the case of the *PIK3CG* (rs17398575) and *EDNRA* (rs1878406) SNPs, their association with carotid plaque reached genome-wide significance but their association with common cIMT was only suggestive, $3x10^{-6}$ and $7x10^{-7}$ respectively. On the other hand the *APOC1* (rs445925) SNP reached genome-wide significance for its association with common cIMT but its association with carotid plaque was only suggestive, $4x10^{-6}$. There were no SNPs that showed statistically significant association with the internal cIMT phenotype as the analysis was underpowered. This was because only three of the nine studies had measured the internal cIMT and resulted in data available for analysis amounting to approximately ten thousand individuals, which is relatively low compared to the 30 thousand individuals that were available for the other two carotid phenotypes.

In the same CHARGE study, the SNPs that were found to be significantly or suggestively associated with carotid plaque and common cIMT were independently tested for their association with CAD in the CARDIoGRAM consortium through a meta-analysis of 81,804 case-control samples (28). There were eight SNPs tested in total, with an even split between the two carotid phenotypes, and of those eight there were four that showed a significant association with CAD (28). The phenotype split of these CAD associated SNPs was one cIMT versus three carotid plaque. The *EDNRA* (rs1878406)

SNP showed the strongest association with CAD with a p-value of $2x10^{-06}$. Further exploration of the *EDNRA* signal is the substance of this thesis.

1.6 Endothelin receptor type A (*EDNRA*)

The *EDNRA* gene is located on the q31.22 band of chromosome 4 in the human genome and reaches approximately 64kb in length when all introns and exons are included. There are eight exons in total in the *EDNRA* gene, seven of which are transcribed and are therefore coding exons (30). The transcription start site is located early in exon 2 (31). As can be seen through the AceView database (32) there are seven different splice variants of *EDNRA,* with their mRNA lengths ranging from 422bp to 4.3kb. The 4.3kb mRNA has been found in many tissues throughout the body but the aorta has been shown to contain the highest level of *EDNRA* mRNA (30). The *EDNRA* gene encodes the endothelin receptor type A which is an integral membrane protein comprised of 427 amino acids and belongs to the rhodopsin-like receptor family, which is a large subgroup of the G protein-coupled receptor family (33).

EDNRA binds all three isoforms of the endothelin protein, named ET-1, ET-2 and ET-3. EDNRA has the strongest binding affinity for ET-1, followed by ET-2 and the receptor has weak binding affinity for ET-3 (34). The ET-1 isoform is 21 amino acids long and is responsible for vasoconstriction through binding to EDNRA (35). The majority of ET-1 protein originates from the endothelial cells of the artery wall and it acts in a paracrine manner with regard to vasoconstriction as the receptor it binds to, EDNRA, is not found in the endothelial layer of the arterial wall (34, 35). The *EDRNA* protein is actually found in the membrane of smooth muscle cells and has been shown to take part in smooth muscle contraction as well as the increase in both cell volume and number, termed hypertrophy and hyperplasia respectively (35).

EDNRA is an interesting candidate for follow-up as it has been connected to a wide array of cell functions that could relate to atherosclerosis pathology. It has been shown, for example, that ET-1 can stimulate DNA synthesis in coronary smooth muscle cell cultures that have had their growth stopped by artificial means (36). In addition to increasing DNA synthesis, ET-1 also causes further protein synthesis and an increase in cell number, mediated through EDNRA (36). Through the use of inhibitors for EDNRA it was determined that these functions were being conducted by ET-1 through the endothelin type A receptor as these effects were not seen when ET-1 binding to EDNRA was prevented (36). This proliferation of smooth muscle cells mediated by EDNRA is interesting as the progression of atherosclerosis involves the proliferation and migration of smooth muscle cells from the media into the intima (36, 37).

The role of *EDNRA* in the endothelin-1 pathway, in addition to its function with regard to vasoconstriction, led to it being examined in many candidate gene studies for diseases associated with dysfunction of the arterial wall. One such example would be the study of EDNRA in association with migraine, which is a disease that has been thought to be connected to vascular dysfunction (38). Migraine had originally been connected to endothelin and studies sought to further investigate the endothelin pathway, which includes *EDNRA*. The −231G>A SNP which lies in the non-coding upstream region of the *EDNRA* gene was found to be associated with migraine in a single study which

was later confirmed in a follow-up meta-analysis of three separate studies (38, 39). Another *EDNRA* variant, rs5333, which is a synonymous mutation (His323His), has been shown to be significantly associated with pulmonary arterial hypertension through a case-control study comparing patients suffering from pulmonary arterial hypertension with healthy individuals that were from the same area (40). In addition, the rs5335 *EDNRA* variant has also been associated with hypertension through a study of families of hypertensive probands (41). Also, *EDNRA* has seen extensive study with regard to hypertension as experiments measuring the level of expression in human mammary arteries have shown a high level of *EDNRA* expression in hypertensive patients (42). The endothelin pathway genes responsible for the expression of ET-1 and EDNRA were again a part of a candidate gene case-control study for heart failure (43) as they had been shown to be produced and present in cardiomyocytes with EDNRA shown to have particular importance with regard to function (44). The case-control study showed that the His323His variant, mentioned previously, was associated with an increased risk of heart failure (43). The His323His variant was also found to be associated with a significant difference in mean IMT in male probands in a case-control study that looked at different measures of atherosclerosis progression and their association with genetic variation in the endothelin pathway, including the genes that code for the ET-1 and EDNRA proteins (45). The same study, however, did not find any association between the variants they examined and plaque (45).

The ET-1 vasoconstriction pathway, which involves EDNRA, has also been posited to be a major driver of renal diseases and so the association between *EDNRA* variants and autosomal dominant polycystic kidney disease (ADPKD) has been examined in a case-control study (46). It was found that the C1363T *EDNRA* variant, which has since been assigned the rs number of rs5343 and is located in the untranslated 3' end of *EDNRA*'s eight exon, was associated with ADPKD (46). There is also suggestive evidence that ADPKD individuals have a higher prevalence of intracranial aneurysm than witnessed in the general population (47). This suggestive evidence is interesting as it has been shown recently, through GWAS of case-control individuals, that a variant upstream of the 5' end of *EDNRA* is significantly associated with intracranial aneurysm (48).

There is, therefore, a wide array of experimental evidence for the association between arterial pathology and variation within, and in close proximity to, *EDNRA*. The majority of the experiments that have shown associations between *EDNRA* and diseases with pathology relating to the arterial wall have been case-control in nature, and that is where the strength of the CHARGE consortium's finding comes in. The CHARGE consortium has shown that a variant upstream of *EDNRA* significantly associated with measures of atherosclerosis in population-based cohorts (28). The amount of evidence available suggests that *EDNRA* could play a pivotal role in the progression of atherosclerosis and makes it a prime candidate for a follow-up study.

As the rs1878406 SNP lying upstream of *EDNRA* has no known function in itself, it is therefore most likely that it represents another variant that it is in linkage disequilibrium (LD) with. In an attempt to identify the functional variant(s) responsible for the effect of *EDNRA* this study will first see whether the effect seen in CHARGE is seen in an independent Icelandic cohort. The attempted replication will seek to confirm that the association is true not only for the AGES Reykjavik Study cohort, which was a part of CHARGE, but also for the rest of the Icelandic population. Then imputations created using

1000 Genome project data will be used in an attempt to refine the CHARGE *EDNRA* signal. Haplotype analysis will also be used as a further attempt at signal refinement.

2 Aims

2.1 General aim

To investigate further the CHARGE consortium's lead SNP (rs1878406) hit for carotid plaque, cIMT as well as CAD. This will be accomplished through refinement of the signal with an aim of finding functional variant(s) hypothesized to be responsible for the lead SNP signal.

2.2 Specific aims

1. To further explore the association of the lead SNP with atherosclerotic phenotypes in the Icelandic population. This will be accomplished through typing the lead SNP marker and attempting replication of the original CHARGE analysis signal in the REFINE Reykjavik Study cohort. It was the AGES Reykjavik Study GWAS cohort that was a part of the CHARGE analysis and as the REFINE Reykjavik Study is not connected to AGES then the lead SNP signal will be investigated in an independent, and also markedly younger, cohort. This will allow for the investigation of whether the signal is found in the general Icelandic population.

2. To refine the lead SNP signal at the *EDNRA* locus in the AGES Reykjavik Study cohort. This will be accomplished by testing the association between SNPs and the following phenotypes: common cIMT, carotid plaque and coronary event endpoints. The SNPs tested will be both genotyped and imputed using 1000 Genomes Project data examined in the context of the LD structure of *EDNRA*.

3. Further refining of the lead SNP signal will be carried out through analysis of *EDNRA* haplotypes. The haplotypes will be phased using genotyped SNPs within, and in close proximity to, *EDNRA*. The testing for association between phased haplotypes and atherosclerotic phenotypes, mentioned previously, may allow identification of functional variant carrying chromosomes.

3 Material and methods

3.1 Cohorts

The two cohorts used for this project are the Age, Gene/Environment Susceptibility (AGES) Reykjavik Study and the Risk Evaluation For INfarct Estimates (REFINE) Reykjavik Study. Both are cohorts that have been recruited by The Icelandic Heart Association and there are currently DNA samples available for both AGES and REFINE individuals totalling 5,764 and 6,941 for each respective study.

The AGES cohort was established as a follow-up study of the original Reykjavik Study cohort which itself is a total population-based study carried out in the Greater Reykjavik area, Iceland and initiated in 1967. The Reykjavik Study featured the recruitment of men and women living in the greater Reykjavik area that were born between the years 1907 and 1935. In total there were 30,795 randomly-sampled originally in the Reykjavik Study, with 27,281 randomly chosen and invited to participate in the study. A total of 19,381 participated which gave the study a recruitment rate of 71.0%. Before recruitment the individuals were split into six groups (B, C, A, D, E, and F). The F group was selected as a control group and did not take part in examinations until much later, specifically in the year 1991. The participants were recruited in five stages between 1967 and 1991. Then a sixth stage was added between 1991 and 1996 for members of the B and F groups who were over the age of 70 at the time of recruitment. Group B was selected for longitudinal follow-up and was invited to participate in all stages of the study.

The AGES Reykjavik Study set about to assess the genetic risk factors associated with disease states that become prevalent with increased, and specifically old, age. This resulted in a multitude of organ systems being assessed and a wealth of phenotypic information accumulated. The AGES study began in 2002 and at that time there were 11,548 surviving Reykjavik Study participants, including 4,799 men and 6,749 women. Of those surviving Reykjavik Study participants, 8,030 were invited to take part in the AGES Reykjavik Study. From 2002 to 2006 a group of 5,764 individuals (42% male and 58% female) that had previously taken part in the Reykjavik Study were re-examined and this gave the AGES Reykjavik Study a 71.8% recruitment rate. The result was that phenotypes which were measured in old age became available and open to comparison with midlife measurements taken during the Reykjavik Study. The AGES Study received approval from both the National Bioethics Committee (VSN: 00-063) and the Data Protection Authority of Iceland.

There were 3,664 individuals randomly selected from the AGES cohort that were genotyped for a GWAS using the Illumina 370CNV BeadChip array. Genotype data on 3,219 individuals became available after filtering for sample failure and other quality values. Imputation was done with the HapMap data from release 22 (built on NCBI build 36). The GWAS was conducted on all 2,533,153 genotyped and imputed SNPs. More detailed information on how the genotyping and imputation was conducted for the GWAS in the AGES Reykjavik Study can be found in the CHARGE meta-analysis supplementary (28). The rationale behind the AGES Reykjavik Study as well as a thorough overview of its design and interim data (2,300 individuals) has been covered previously (49).

The REFINE Reykjavik Study's aim is to develop a way to more accurately predict risk for CAD in individuals, especially focusing on those who are deemed to be at a low or medium level of risk. This

will be accomplished through the use of both new cardiovascular risk measurements and by investigating interactions among previously established risk factors. An emphasis was and will be placed on making any assessment method simple to use and as non-invasive as possible, thereby allowing their use in risk assessment through individual screening. The emphasis for this study will be placed on an in-depth phenotypic measurements and having a large group of individuals to increase statistical power for observing risk factor relationships and associations.

In total there were 9,478 randomly-sampled individuals selected into the REFINE Reykjavik Study cohort. They were all born between the years 1935 and 1985 and as such the age range for these individuals was 25 to 70 years old. In addition, they were all living in the greater Reykjavik area as of November 2005, and recruitment took place from 2006 until 2011. Of those 9,478 selected into the cohort, 8,836 individuals were invited to participate in the REFINE Reykjavik Study and a total of 6,941 took part, resulting in a recruitment rate of 78.6%. These risk factors measured were much akin to those measured in the AGES Reykjavik Study (49). The REFINE Reykjavik Study has received the formal approval of the National Bioethics Committee (VSN: 05-112) as well as the Data Protection Authority of Iceland.

The approval of the National Bioethics Committee and the Data Protection Authority of Iceland, given for both the AGES and REFINE Reykjavik Studies, allows for genetic analysis of the DNA samples and said approval extends to this project. The individuals within the AGES and REFINE cohorts have all provided informed consent for the use of their genetic and phenotypic information for the purposes of research.

3.2 Atherosclerotic phenotypes: Measurements and methodology

This work examines the atherosclerotic phenotypes of the carotids obtained as described in a detail below.

3.2.1 Carotid artery measurements: Plaque and IMT

Carotid plaque and common cIMT were both measured using B-mode ultrasonography and in accordance to a standardized protocol that was developed for the AGES Reykjavik Study. The protocol was designed collaboratively by the IHA and the Vascular Imaging Center of the Julius Center for Health Sciences and Primary Care (University Medical Centre Utrecht, Utrecht, The Netherlands).

Atherosclerotic plaque is examined in both the left and right carotid bifurcation and also the internal carotid artery using the aforementioned B-mode ultrasonography. They were quantified whilst the ultrasound examination occurred and the most severe lesion per segment was assigned a semi-quantitative value. The possibilities for the values assigned to the atherosclerotic plaques are either none, minimal, moderate, severe or semi-occluded.

Measurements of common cIMT are obtained from B-mode ultrasound images. The images are taken of the left and also the right common carotid artery and they encompass a predefined 10mm segment in each artery. They are taken from four different interrogation angles that are standardized between individuals using the Meijer's Carotid Arc (Meijer Medical Ultrasound, Voorschoten, The Netherlands). The cIMT of the near and far walls are measured across all the images of the common

carotid arteries and the average value is taken which gives the common cIMT measurement. An ACUSON Sequoia C256 Echocardiography System (Siemens Healthcare formally known as Siemens Medical Systems, Erlangen, Germany) was used to capture the B-mode ultrasound images mentioned above.

The same measurements detailed above were carried out on the REFINE Reykjavik Study cohort utilizing the same equipment and in accordance with the protocol mentioned with respect to the AGES Reykjavik Study. In both studies two trained technicians analysed the image at the same time. This was done to provide increased assurances about the consistency of the measurements as well as the quality. Technicians were subjected to regular examinations by third-party experts to assure sustained efficiency and proficiency.

3.3 Analysis variables: definitions and assigned values explained

The majority of the risk factors that were measured in both the AGES and REFINE cohorts are standardized. However, there are several variables that have been used in the statistical analyses that require further explanation to clarify exactly what they represent. They are presented in the following text.

There are data available on smoking status in both the AGES Reykjavik and REFINE Reykjavik Study cohorts. The smoking status variable within these two cohorts is split into three groups: those that have never smoked are assigned the value 0, former smokers that have had at least 100 cigarettes or 20 cigars in their lifetime are assigned the value 1 and current smokers have the value 2.

The Coronary Event by Questionnaire (CEQ) variable gives information on whether the individual has experienced at least one of the following events: MI, coronary bypass graft and/or angioplasty of coronary arteries. Those that answer no are assigned the value 0 and those that answered positively are assigned the value 1. Therefore this is a dichotomous variable that is based on responses to a questionnaire.

The coronary event variable is defined as an individual having either suffered MI, undergone Percutaneous Coronary Intervention (PCI) or coronary artery bypass graft (CABG) prior to their taking part in the AGES Reykjavik study. It's a dichotomous variable with the value 1 assigned to those that have suffered a coronary event prior to the study. It is built upon information taken from hospital records.

The carotid plaque variable used for analysis in this project is a dichotomous variable. Those individuals that have no or minimal plaque are grouped together and assigned the value 0. Alternatively, the individuals with moderate, severe or semi-occluded plaque are grouped together and assigned the value 1.

For the common cIMT variable the average of the maximum thickness of the near wall and far wall is used. This quantitative variable is named cIMTMAX and all analyses are performed using the natural-logarithm transformed values due to the skewed nature of the variable.

3.4 DNA whole-blood extraction and isolation

The DNA of both the AGES and REFINE cohorts were isolated using a standard salting out method (50). Once the DNA has been extracted from the whole-blood samples it is placed into a TE buffer solution and placed in a tilt-rocking mixer for 24 hours to assist the DNA in dissolving. From there it is stored at 4°C for a period of 4 weeks to ensure the DNA has completely dissolved into the TE buffer. After this period the DNA concentration is quantified using a Spectromax M2 Microplate Reader (Molecular Devices, CA, USA) and at that point the samples are diluted if necessary. At this point the samples are place into -20°C storage and from there they are ready for experimental use.

3.5 SNP genotyping: assay design and reagents used

The primers and probes for the lead SNP (rs1878406) assay were designed online using the RealTimeDesign (RTD) software from Biosearch Technologies (Petaluma, CA, USA). The primers were, however, ordered from a different company, Sigma-Aldrich (St. Louis, MO, USA), in advance of any order being placed for the Biosearch Technologies' probes. This was done as a precautionary measure to ensure that the target site could be amplified before the real-time PCR genotyping work was underway. The oligonucleotide sequences of the primers and probes, as well as their melting temperature (referred to as Tm) and the fluorescent dyes used in the probes can be found in Table 1 which is located further down.

Table 1. General design information for lead SNP (rs1878406) primer and probes. The information was retrieved from the RTD software on Biosearch Technologies website. Oligo is short for oligonucleotide and Tm is the melting temperate of the oligonucleotide and the 5' pos is short for the 5' position of the oligonucleotides on the target sequence. The positions given are relative to the target sequence selected which was only a short distance either side of rs1878406.

Oligo	Fluorescent dye	Tm	5' Pos	Length	Sequence
Forward Primer	NA	66.0	180	20	AGTGGCAGTTGCTGTCTTCA
Reverse Primer	NA	66.6	31	20	TCATTTGGAACAGGGGCTTC
Allele 1 Probe	FAM	61.1	115	18	ACGAGATAACAGCTAACA
Allele 2 Probe	CAL560	61.0	116	18	TACGAGATAACAGCTGAC

In addition to the information contained within the table, a visual overview of how the primers and probes anneal to the target sequence can be seen in Figure 5, with the primers and probes colour-coordinated for easy reference.

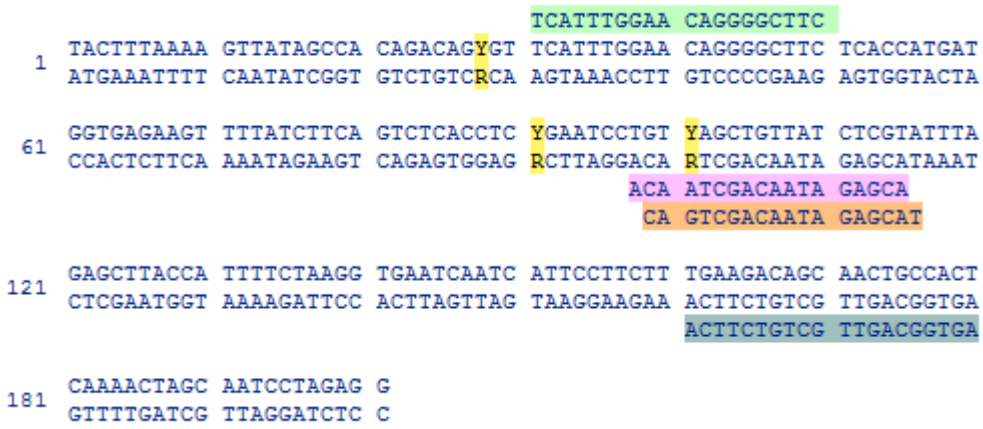

Figure 5. A visual overview of the lead SNP (rs1878406) assay hybridizing to target DNA. This was taken from Biosearch Technologies' RTD software. The green/grey highlighted oligonucleotides are the primers and the pink/orange highlighted oligonucleotides are the probes. The bases highlighted in yellow are the sites of variation within the target region. They are displayed using ambiguity SNP codes, where a C/T variation is represented by a Y and an A/G variation is represented by an R.

Different annealing temperatures as well as a wide range of reagent protocols and mixtures were tested and the success of the PCR was investigated using a standard gel-electrophoresis method. The PCR protocol for the lead SNP primers was deemed to be complete once a single amplicon (151bp in length) could be identified through gel electrophoresis that had been run for 30 minutes at 125mV on a 1.5% agarose gel. After the primer optimization was completed the probes were ordered from Biosearch Technologies and the focus was shifted to the optimization of the real-time PCR genotyping.

The basic PCR protocol was transitioned into the real-time PCR protocol with all reagents, excluding the primer mix, staying at the same ratios in a 20µL reaction. The inclusion of the probes, ROX and $MgCl_2$ resulted in a reduction in the total amount of dH_2O per sample. The primer annealing temperature during normal PCR that was found to provide the strongest amplicon bands was 60°C. This annealing temperature was used in the real-time PCR experiment going forward. The complete thermocycling protocol for the real-time PCR lead SNP assay was as follows: A pre-PCR read (holding stage) at 60°C for 1 minute, a holding stage at 95°C for 45 seconds followed by a cycling stage divided into two steps, 95°C for 15 seconds and then 60°C for one minute with this stage being repeated 40 times. The final stage is the post-PCR read (holding stage) which is 60°C for 1 minute.

When the real-time PCR genotyping assay was run on multiple 96-well plates concurrently, a single plate would be run on the 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) and the fluorescence would be read in real time. Any additional plates that were started at the same time would be run as a normal PCR in either an Applied Biosystems Veriti 96-Well Thermal Cycler or an Applied Biosystems 2720 96-Well Thermal Cycler, with the post-PCR read of probe fluorescence levels being conducted on the Applied Biosystems 7500 Real-Time PCR System. This worked well as the SNP was common enough that all three genotypes were highly likely to be represented on a single 96-well plate and that makes confident genotype calling based on post-PCR reads a simple matter.

Optimization of the real-time PCR protocol occurred in three stages. First, a number of test-DNA reactions were run with the total sample volume of 20µL and differing amounts of 5µM probe-mix, specifically: 0.8µL, 0.5µL, 0.3µL and 0.2µL. From this first stage it was determined that 0.3µL of probe-mix per 20µL reaction would suffice for confident genotype calling. For the testing during this first stage the One-Taq Polymerase (New England Biolabs Inc., Ipswich, MA, USA. Catalogue number: M0480L) was used.

The second stage was conducted to determine whether the standard Taq or One-Taq Polymerase should be used. The difference between the two being that the One-Taq Polymerase has increased reliability and more robust amplification. This is due to the incorporation of 3' to 5' exonuclease activity from the Deep Vent DNA Polymerase into the standard Taq DNA polymerase (51). For the lead SNP assay it was determined that the combination of One-Taq Polymerase with the Standard Taq Buffer (10X concentration) provided slightly increased amplification and as such said combination would be used until the One-Taq Polymerase was depleted. At which point, the combination of standard Taq Polymerase and standard Taq Buffer was used, as it was the more cost-efficient option and the improved amplification of One-Taq was not necessary for confident genotype-calling of the lead SNP.

The third and final stage was to increase cost-efficiency of the reactions by reducing the total reaction volume. The total reaction volumes tested were 20µL, 15µL and 10µL. From these tests the decision was taken to have a total volume of 15µL as this provided the sought after balance between total efficiency and confident lead SNP calling. The DNA volume in the 15µL reaction was not scaled down to the same degree as the other reagents due to the inefficiency of pipetting a volume smaller than 0.5µL. To compensate for the increased volume of DNA the volume of dH_2O was decreased. The final real-time PCR protocol for the self-designed lead SNP assay can be found in Table 2 below.

Table 2. Real-time PCR reagent protocol for a single sample. Note that the standard Taq Polymerase was used for the large majority of the real-time PCR genotyping, however, the first few 96-well plates were genotyped using One-Taq polymerase. REFINE DNA plates are set to the relatively high concentration of 50ng/µL due to requirements of other projects at the IHA. The dNTP was diluted to 2mM down from the stock concentration of 10mM. NEB stands for New England Biolabs Inc. Invitrogen (Carlsbad, CA, USA).

Reagent	Manufacturer	Catalogue #	Volume (µL) - 1 sample
Standard Taq reaction buffer (10X)	NEB	B9014S	1.5
dNTP (2mM)	NEB	N0447L	1.5
Taq Polymerase (Standard)	NEB	M0273X	0.075
Primer mix (5µM)	Sigma-Aldrich	N/A	0.75
Probe mix (5µM)	Biosearch Technologies	N/A	0.225
MgCl2 (25mM)	NEB	B9021S	0.15
DNA (50ng/µL)	N/A	N/A	0.5
ROX(1/10 dilution)	Invitrogen	12223-012	0.15
dH2O	N/A	N/A	10.15
Total Volume	N/A	N/A	15

3.6 Statistical analyses and data handling

Cross-sectional analyses, i.e. logistic and linear regressions, were performed using R (version 2.15.2) (52) for both the SNPs and haplotypes. A logistic regression model was used for dichotomous variables, i.e. carotid plaque and coronary event and a linear regression model was used for quantitative variables, such as common cIMT. All regression associations performed were adjusted for age and gender. When correcting for other CVD risk factors the following factors are correcting for: BMI, Systolic blood pressure, HDL, LDL, triglyceride, type two diabetes, smoking, hypertension medication, statins and C-reactive protein (CRP). To account for multiple testing the alpha threshold of 0.01 was used, both for single-point SNP and haplotype analysis. The haplotype frequency was calculated in the same manner as SNP MAF.

Data handling was conducted in SAS Enterprise Guide (version 4.3). PLINK (version 1.07 for Linux) (53) was used to extract subsets of genotypic data for specific SNPs.

The Pearson's correlation coefficient was used to calculate the correlation between any two SNPs (54). The same method was applied to calculate the correlation between a haplotype and SNP or SNPs.

The test of interaction (effect modification) by cohort was conducted by combining the AGES and REFINE cohort datasets and creating a cohort indicator variable called "AGES". The AGES variable discriminated between the two cohorts in a dichotomous manner, with the value 1 assigned to AGES individuals and the value 0 assigned to REFINE individuals. The "AGES" cohort variable was included in the regression model as a main effect, as was the interaction between the "AGES" cohort variable and lead SNP coded for the minor allele. The interaction variable was a multiplication of the two variables having their interaction tested (AGES x rs1878406_T). The interpretation of the regression coefficient for the interaction variable is the difference between the effect of the SNP in the cohorts.

3.7 Imputation

SNPs across the entire genome of AGES GWAS individuals were imputed using 1000 Genomes Project data (Build: Hg19, 1000G v3 phaseI) and the Mach/minimac program. The SNPs used in this study for regression analysis were within the range chr4:148,392,069-148,476,106 which ranges from 10kb either side of the *EDNRA* gene.

Imputed SNPs were filtered for their imputation quality (r^2 value) and the threshold of 0.6 was used, which is slightly more conservative than typical for 1000 Genomes imputation (55). Filtering for imputation quality prior to running any analyses reduces the number of SNPs being tested and that will lessen the effects of multiple testing. Also, possible false positives that could be the result of poor imputation quality will be avoided.

The first imputation of the AGES Reykjavik Study GWAS was done with HapMap data from release 22 (built on NCBI build 36). The CHARGE meta-analysis GWAS was conducted on all 2,533,153 genotyped and imputed SNPs. More detailed information on how the genotyping and imputation was conducted for the GWAS in the AGES Reykjavik Study can be found in the CHARGE meta-analysis supplementary (28).

3.8 Visualizing LD structures of genomic locations and showing LD between SNPs

LD structure can be visualized through the BROAD Institute's Haploview program (15). The LD blocks, based on the CEU population, were defined based confidence limits of the LD measure D' and recombination levels between SNP pairs (56). D' quantifies the difference between expected haplotype frequency under linkage equilibrium (independence) versus the observed haplotype frequency.

To examine and visualize the LD between SNPs the SNP the Annotation and Proxy Search tool (SNAP) from the BROAD Institute (version 2.2) was used (57). The proxy SNPs to certain target SNPs were determined using the proxy search functionality. The pair-wise LD tool allowed for the LD between SNPs to be investigated. Regional association plots and regional LD plots were constructed to visualize regression analysis results as well as show SNPs in certain levels of LD with a target SNP.

3.9 Haplotype phasing: The principles and use of fastPHASE

The form that SNP information is given in pertains to the minor and major allele, with minor and major being determined by their frequency in a given population. From SNP data it is only possible to know which individuals have which combination of alleles for multiple SNPs, and not which chromosomes particular alleles are found on. The genotyped SNPs are therefore unphased, in that the exact chromosome that their alleles lie on is undetermined. It is possible to determine which alleles from multiple SNPs are most likely to lie together on the same chromosome. This process is called haplotype phasing and there is a wide variety of computational methodologies available that serve this purpose (58).

Here, haplotype phasing was performed through the use of the fastPHASE program (version 1.2) (59). FastPHASE requires a PHASE file formatted input dataset. The input file needs to contain information about the number of individuals to be analysed, the number of SNPs to be included in the phasing and the genotypes of those SNPs for each individual. To extract genotypic information about a group of individuals for a subset of SNPs the PLINK program is used. As PLINK outputs files that are not in fastPHASE compatible formats, a format conversion program is required. IPGWAS (version 1.4) is such a program and converts PLINK files, amongst others, into the PHASE format (60). The physical positions of the SNPs are not used in the default model options and were therefore excluded from the PHASE formatted input file. An overview of the genotyped *EDNRA* SNPs used in the fastPHASE haplotype phasing see Table 3.

Table 3. Genotyped *EDNRA* SNPs used in haplotype phasing, ordered by their chromosome 4 co-ordinates according to the GRCh37/hg19 assembly of the reference Human genome.

SNP	Major/Minor Allele	Chromosome 4 co-ordinates
rs6842241	C/A	148,400,819
rs10024834	A/G	148,420,636
rs6537483	G/A	148,426,451
rs4563479	C/T	148,429,750
rs6827096	C/T	148,435,446
rs1878404	G/A	148,440,625
rs10008744	A/C	148,445,429
rs4639051	A/G	148,448,870
rs5334	G/A	148,461,073
rs5343	C/T	148,464,992
rs10028838	T/G	148,471,332

The fastPHASE program estimates probable haplotypes from unphased SNP genotype datasets. There are a variety of options available for the fastPHASE EM algorithm that make it more applicable to the number of SNPs and individuals in the input SNP dataset. For instance, the number of SNPs in the dataset can vary as well as the number of individuals. For the purposes of this project's haplotype-phasing the algorithm options were kept at their default values (59). This project makes use of the EM algorithm to minimize switch error rather than individual error. The following options for the EM algorithm were used:

The number of random starts of the EM algorithm kept at the default value of 20.

The number of iterations of the EM algorithm kept at the default value of 25.

The number of haplotypes sampled from the posterior distribution from a particular random start of the algorithm set to 20.

The option to estimates haplotypes by minimizing individual error was not invoked.

Once haplotype phasing was completed, using fastPHASE and the above mentioned EM algorithm options, an output file was produced. Formatting of the fastPHASE output file was required as the default format was unworkable with regards to analysis. Formatting was accomplished through the use of a Perl script which converted the output file into a more familiar and workable format, specifically the row names become the individual sample IDs with two separate haplotype columns. To be able to perform regression analyses on the data a haplotype count - analogous to SNP dosage – was created. The COUNTIF function in Excel was used to create a tally of each haplotype, split into separate columns and displayed on a per-individual basis. The haplotype counts file was then imported into R and regression analyses were carried out.

In addition to haplotype phasing of all 11 genotyped SNPs, haplotypes were phased in a sliding-windows manner. Sliding windows analysis involves phasing a small number of SNPs at a time and then testing the smaller phased haplotypes for association with a particular trait or phenotype (61). This requires less computational power than phasing a large number of SNPs at the same time and is more informative of which particular genomic regions associate with a particular phenotype or trait.

The sliding-window analysis in this study involved phasing five SNPs at a time. The first sliding window included SNPs 1 to 5, the second window included SNPs 2 to 6, the third window included SNPs 3 to 7 and so on. In total there were seven different sliding windows and regression analyses were run on all phased haplotypes from each sliding window.

3.10 Functional annotation

3.10.1 RegulomeDB

RegulomeDB is a database that allows for variants, dbSNP or novel, to be annotated with regard to regulatory information (62). It makes use of a number of public datasets, including GEO, the ENCODE project as well as data available in the public literature (62). The RegulomeDB scoring system assigns different scores depending on the amount of different experimental evidence each variant has, Figure 6.

Score	Supporting data
1a	eQTL + TF binding + matched TF motif + matched DNase Footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase Footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding / DNase peak
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak
2b	TF binding + any motif + DNase Footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
4	TF binding + DNase peak
5	TF binding or DNase peak
6	other

Figure 6. The RegulomeDB scoring system. Taken from (63).

3.10.2 SNPnexus

The SNPnexus database allows for the annotation of dbSNP variants, in addition to novel variants (64, 65). The variants can be annotated for gene or protein changes that they are responsible for based on information from publically available resources (RefSeq, Ensembl, AceView, UCSC Genome Browser etc). If the variants lie within a DNA region that codes for protein then the variants effect on the protein function can be assessed through resources such as SIFT or PolyPhen. The variants can be annotated for regulatory function, their MAF in different populations using HapMap data, their conservation score, any phenotype or disease association from publically published GWAS and also structural variants, such as copy number variants (CPV) and insertions or deletions, can be annotated.

4 Results

4.1 Basic characteristics - AGES and REFINE

Basic characteristics of the AGES and REFINE are shown in Table 4 and Table 5. Gender-specific basic characteristic tables for both cohorts are shown in Table 15, Table 16, Table 17 and Table 18. The basic characteristics are limited to the number of genotyped individuals in each cohort. The AGES cohort is made up of GWAS individuals (N=3,219) and the REFINE cohort contains lead SNP genotyped individuals (N=5,521). DNA samples for the entire REFINE Reykjavik Study (N=6,971) were not available at the time of genotyping.

4.1.1 Cardiovascular risk factors

Conventional cardiovascular risk factors show several differences between cohorts and genders.

The average age is higher in the AGES cohort. It is 76.4 years in AGES and 49.4 years in REFINE. The men's age in both cohorts is minutely higher than that of the women. In AGES, men have an average age of 76.5 and women have an average age of 76.3. In REFINE, men have an average age of 49.6 and women have an average age of 49.3.

The average systolic BP is higher in the AGES cohort. It is 143 mmHg in AGES and 122 mmHg in REFINE. The difference between men and women in AGES is minimal, with values of 143 mmHg and 142 mmHg respectively. In REFINE the difference is more pronounced with the systolic BP average of 127 mmHg in men and 118 mmHg in women.

The average cholesterol is higher in the AGES cohort. It is 5.6 mmol/L in AGES and 5.2 mmol/L in REFINE. In AGES, the average cholesterol level is higher in women than in men (6.0 vs. 5.2 mmol/L). In REFINE, the average cholesterol is the same in both genders (5.2 mmol/L).

The average HDL is higher in the AGES cohort. It is 1.58 mmol/L in AGES and 1.49 mmol/L in REFINE. In both cohorts, the average HDL level is higher in women. In AGES, the average HDL level is 1.40 mmol/L in men and 1.72 mmol/L in women. In REFINE, the average HDL level is 1.32 mmol/L in men and 1.65 mmol/L in women.

The average LDL is higher in the AGES cohort. It is 3.51 mmol/L in AGES and 3.18 mmol/L in REFINE. In AGES, the average level of LDL is higher in women (3.70 vs. 3.23 mmol/L). The opposite is seen in REFINE, where the average LDL level in men is 3.26 mmol/L and is 3.1 mmol/L in women.

The median TG is higher in the AGES cohort. It is 1.05 mmol/L in AGES and 1.01 in REFINE. In AGES, the median TG is higher in women (1.07 vs. 1.01 mmol/L). The opposite is seen in REFINE, where the average LD level in men is 0.90 mmol/L and 1.13 mmol/L in men.

The prevalence of T2D is higher in the AGES cohort. It is 11.5% in AGES and 4.5% in REFINE. In both cohorts, the prevalence of T2D is higher in men. In AGES, 15.1% of men and 8.8% of women have T2D. In REFINE, 6.2% of men and 2.9% of women have T2D.

The percentage of individuals taking statins is higher in the AGES cohort. In AGES there are 22.7% that are taking statins and in REFINE it is 9.1%. In both cohorts, the percentage of men taking statins

is higher than that of women. In AGES, 29.4% of men and 17.8% of women are taking statins. In REFINE, 12.3% of men and 5.9% of women are taking statins.

The percentage of individuals taking hypertension medication is higher in the AGES cohort. Individuals taking hypertension medication amount to 63.8% in AGES and 23.4% in REFINE. In AGES, there are fewer men taking hypertension medication compared to women (62.7% vs. 64.7%). In REFINE, there are more men taking hypertension medication compared to women (24.0% vs. 22.7%)

The prevalence of smokers, both current and former, is lower in the AGES cohort. In AGES, current and former smokers amount to 12.7% and 43.5% respectively whereas in REFINE current smokers amount to 21.5% with former smokers totalling 37.8%. In AGES, a higher percentage of men identify themselves as former smokers compared to women (59.4% vs. 35.0%). The opposite is true for current smokers, with a lower percentage of men identifying themselves as current smokers compared to women (12% vs. 13.2%). Overall, smoking is more prevalent in men than women in the AGES cohort (71.4% vs. 48.2%). In REFINE, a higher percentage of men identify themselves as former smokers compared to women (39.8% vs. 35.7%). The same holds true for current smokers (21.7% vs. 21.3%). Overall, in the REFINE cohort current and former smokers are more prevalent among men (61.5% vs. 57.0%).

The median value of CRP is higher in the AGES cohort. On the whole it is 1.90 in AGES and 1.32 in REFINE. In both cohorts, the median CRP value is higher in women than in men. In AGES, it is 1.80 in men and 1.90 in women. In REFINE, it is 1.28 in men and 1.40 in women.

The prevalence of coronary events is higher in the AGES cohort. In AGES, 15.0% have suffered a coronary event whereas in REFINE 3.8% have. In both cohorts, coronary events are more prevalent among men. Specifically, the prevalence in AGES is 25.4% in men and 7.4% in women. And the prevalence in REFINE is 5.8% in men and 1.94% in women.

4.1.2 Subclinical measures of atherosclerosis

Included in the basic characteristics tables are the two separate measures of subclinical atherosclerosis, carotid plaque and common cIMT.

The prevalence of carotid plaque is higher in the AGES cohort. It is 66.9% in AGES and 11.0% in REFINE. In both cohorts the prevalence of carotid plaque is higher in men. In AGES, 68.7% of men and 65.6% of women have carotid plaque. In REFINE, 13.2% of men and 9.0% of women have carotid plaque.

The average common cIMT is higher in the AGES cohort. The average thickness is 1.13 mm in AGES and 0.86 in REFINE. The common cIMT is higher in men in both cohorts. In AGES, the average thickness for men and women is 1.17 mm and 1.11 mm respectively. In REFINE, the average thickness for men and women is 0.88 mm and 0.83 mm respectively.

Table 4. Basic characteristics of the AGES cohort GWAS individuals (N=3,219).

Characteristic	Mean, median or percentage	Range	N (Events)	N (Total)	N (Missing data)
Age (Years)	76.4 ± 5.5	66 - 95		3,219	0
BMI (kg/m²)	27.1 ± 4.4	14.8 - 48.5		3,216	3
Systolic BP (mmHg)	142.6 ± 20.3	92 - 253		3,218	1
Diastolic BP (mmHg)	74.1 ± 9.6	29 - 145		3,218	1
Pulse pressure (mmHg)	68.5 ± 18.2	17 - 168		3,218	1
Common cIMT (mm)	1.13 ± 0.16	0.7 - 1.93		3,073	146
Cholesterol (mmol/L)	5.6 ± 1.2	1.8 - 10.5		3,219	0
HDL (mmol/L)	1.58 ± 0.45	0.35 - 4.05		3,219	0
LDL (mmol/L)	3.51 ± 1.04	0.43 - 8.53		3,217	2
Triglycerides (mmol/L)	1.05 (0.78-1.45)	0.3 - 12.99		3,219	0
Fasting glucose (mmol/L)	5.75 ± 1.11	3.5 - 19		3,219	0
Coronary calcium (Agaston)	283 (43-899)	0 - 8,673		3,176	43
Hba1c (%)	5.69 ± 0.51	4 - 12		2,962	257
Type 2 Diabetes (%)	11.5		369	3,215	4
Statins (%)	22.7		729	3,219	0
Coronary event (%)	15.0		477	3,188	31
Never smoked (%)	42.0		1,353	3,219	0
Former smoker (%)	45.3		1,457		
Current smoker (%)	12.7		409		
Carotid plaque (%)	66.9		2,047	3,058	161
Hypertension medication (%)	63.8		63.81	3,219	0

The mean is displayed with the standard deviation and the median with the inter-quartile range. LDL levels estimated using the Friedewald formula (LDL-chol= Total chol - HDL-chol - TG/2.2). BP stands for Blood Pressure. Carotid plaque represents individuals with minimal, moderate, severe or semi-occluded levels of plaque in the carotid artery.

Table 5. Basic characteristics of REFINE cohort individuals genotyped for rs1878406 (N=5,521).

Characteristic	Mean, median or percentage	Range	N (Events)	N (Total)	N (Missing data)
Age (Years)	49.4 ± 12.1	20 - 73		5,521	0
BMI (kg/m²)	27.4 ± 4.9	15.0 - 65.8		5,518	3
Systolic Blood Pressure (mmHg)	122.3 ± 16.8	77 - 226		5,520	1
Diastolic Blood Pressure (mmHg)	71.6 ± 10.3	36 - 148		5,520	1
Pulse Pressure (mmHg)	50.7 ± 13.5	7.0 - 148		5,520	1
Common cIMT (mm)	0.86 ± 0.17	0.37 - 1.76		5,521	0
Cholesterol (mmol/L)	5.2 ± 1.0	2.3 - 12.7		5,509	12
HDL (mmol/L)	1.49 ± 0.41	0.31 - 3.98		5,509	12
LDL (mmol/L)	3.18 ± 0.93	0.13 - 10.69		5,468	53
Triglycerides (mmol/L)	1.01 (0.73/1.44)	0.16 - 11.64		5,509	12
Fasting Glucose (mmol/L)	5.48 ± 1.00	1.96 - 22.61		5,509	12
Type 2 Diabetes (%)	4.5		248	5,521	0
Statins (%)	9.1		500	5,516	5
History of MI in Family (%)	36.1		1,946	5,384	137
CEQ (%)	3.8		210	5,490	31
Never smoked (%)	40.8		2,238	5,489	32
Former smoker (%)	37.8		2,072		
Current smoker (%)	21.5		1,179		
Plaque (%)	11.0		609	5,521	0
Gender - Male (%)	49.4		2,728	5,521	0
Gender - Female (%)	50.6		2,793		

The mean is displayed with the standard deviation and the median with the inter-quartile range. LDL levels estimated using the Friedewald formula (LDL-chol= Total chol - HDL-chol - TG/2.2). BP stands for Blood Pressure. CEQ stands for Coronary Event by Questionnaire. Carotid plaque represents individuals with minimal, moderate, severe or semi-occluded levels of plaque in the carotid artery.

4.2 Replication of the lead SNP (rs1878406) CHARGE signal in REFINE

To address the first specific project aim of replicating the CHARGE meta-analysis of GWAS result in another independent Icelandic cohort, genotyping was carried out in the REFINE Reykjavik Study cohort.

4.2.1 Genotyping the lead SNP (rs1878406)

The available assay material was enough to genotype the lead SNP in 73 plates, each plate containing 96 samples. A total of 7,008 reactions were run and 5,611 (80%) samples were successfully genotyped. Failed reactions amounted to 1,324 (19%). There were 73 (1%) negative controls overall.

In seven plates (672 samples) failed amplification was related to a faulty 10x Standard Buffer batch from NEB. Poor differentiation of the genotype clusters in post-PCR reads of assay fluorescence levels also resulted in failed genotyping of samples. Random amplification failures with no determinable cause accounted for the rest. An example of a successful post-PCR read of assay fluorescence with good clustering of the three lead SNP genotypes can be seen in Figure 7,

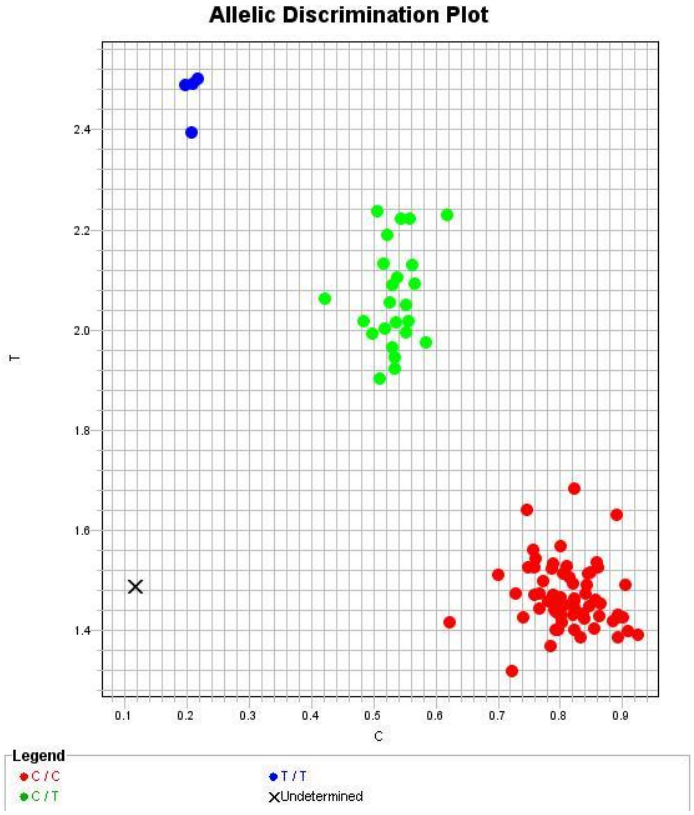

Figure 7. Real-time PCR allelic discrimination plot of lead SNP genotype clustering.96 samples. Image was taken from an Applied Biosystems 7500 Real-Time PCR machine.

Of the 5,521 genotyped individuals there were 4,005 homozygous for the major allele C, 1,401 CT heterozygous and 115 TT homozygous. The lead SNP had a MAF of 0.148 in REFINE.

4.2.2 Association in REFINE Reykjavik Study

The lead SNP (rs1878406) was shown to be significantly associated with carotid plaque and common cIMT in the CHARGE meta-analysis paper and was also found to be associated with CAD in the CARDIoGRAM consortium (28). The AGES Reykjavik Study GWAS was a part of the CHARGE meta-analysis. Replication of these results was attempted in an independent Icelandic cohort, the REFINE Reykjavik study. Also the signal strength in the AGES Reykjavik Study GWAS samples was examined. The lead SNP was imputed in the AGES GWAS samples using 1000 Genomes data in this instance and not HapMap as was carried out previously. The lead SNP has an imputation quality score (r^2) value of 0.971 in AGES.

The association between the dichotomous variable carotid plaque and the lead SNP was tested in both cohorts, Table 6. The lead SNP is significantly associated with carotid plaque in both cohorts, with an OR of 1.25 (95% CI; 1.05-1.49) in REFINE and an OR of 1.30 (95% CI; 1.10-1.52) in AGES. Test of interaction (effect modification) by cohort was not statistically significant (p=0.675). The lead SNP is not significantly associated with coronary events in either cohort or when the cohorts were combined (N=8,687). In the combined AGES and REFINE analysis the p value was 0.316 and the OR was 0.92 (95% CI; 0.781-1.08)

Table 6. Results of association between the lead SNP with carotid plaque and coronary event in AGES and REFINE. The coded lead SNP allele for carotid plaque was the minor allele. The coded lead SNP allele for coronary events was the major allele. Logistic regression analysis was adjusted for age and gender.

Phenotype	Cohort	β	SE	OR (95% CI)	p	N (Events)	N (Total)
Carotid plaque	AGES	0.260	0.0828	1.30 (1.10-1.53)	0.00172	2,047	3,058
	REFINE	0.221	0.0897	1.25 (1.05-1.49)	0.0137	609	5,521
Coronary event	AGES	-0.119	0.102	0.888 (0.727-1.08)	0.245	477	3,188
	REFINE	0.333	0.324	1.39 (0.739-2.63)	0.305	210	5,490

The association between the quantitative variable common cIMT and the lead SNP was tested in both cohorts, Table 7. The lead SNP is significantly associated with common cIMT in both cohorts, with a beta of 0.0126 (95% CI; 0.00588-0.0193) in the REFINE cohort and 0.0187 (95% CI; 0.00925-0.0281) in the AGES cohort. Test of interaction (effect modification) by cohort was not statistically significant (p=0.395).

Table 7. Results of association between the lead SNP with common cIMT in AGES and REFINE. The coded Lead SNP allele for cIMT was the minor allele. Linear regression analysis performed adjusted for age and gender.

Phenotype	Cohort	β (95% CI)	SE	p	N (Total)
Common cIMT	AGES	0.0187 (0.00925-0.0281)	0.00482	1.06E-04	3,073
	REFINE	0.0126 (0.00588-0.0193)	0.00343	2.54E-04	5,521

The lead SNP's signals for carotid plaque and common cIMT, shown as significant in both cohorts, were examined further with respect to potentially confounding cardiovascular risk factors. In addition to age and gender that were included previously, the following risk factors were adjusted for: systolic

blood pressure, HDL, LDL, triglycerides, type 2 diabetes, hypertension medication, statins and CRP. Due to some of the measurements having missing values the original association test where only age and gender were corrected for was rerun on a dataset that had no missing values for any of the additional cardiovascular risk factors listed above. The REFINE cohort went from 5,521 to 5,424 when missing values were removed and the AGES cohort went from 3,219 to 3,047. The number of carotid plaque events was reduced by 10 in REFINE and six in AGES.

The lead SNP remains significantly associated with carotid plaque when the cardiovascular risk factors mentioned above are corrected for, Table 8. The signal in the AGES cohort becomes less significant with the p-value increasing from 0.00163 to 0.00511 and the OR is reduced, going from 1.30 (95% CI; 1.10-1.53) to 1.27 (95% CI; 1.07-1.51). The standard error increases from 0.0830 to 0.0861. In the REFINE cohort the signal becomes more statistically significant with the p-value decreasing from 0.0166 to 0.0138. The OR for the lead SNP signal in REFINE increases from 1.24 (95% CI; 1.04-1.49) to 1.26 (95% CI; 1.05-1.52) with the standard error also seeing an increase, going from 0.0908 to 0.0948. Test of interaction (effect modification) by cohort was not statistically significant (p=0.918).

Table 8. Carotid plaque logistic regression results. The coded lead SNP allele was the minor allele. Only individuals with measurements in all risk factors were included in the analyses.

Cohort	β	SE	OR (95% CI)	p	N (Events)	N (Total)
AGES	0.261	0.0830	1.30 (1.10-1.53)	0.00163*	2,041	3,047
	0.241	0.0861	1.27 (1.07-1.51)	0.00511**		
REFINE	0.218	0.0908	1.24 (1.04-1.49)	0.0166*	599	5,424
	0.233	0.0948	1.26 (1.05-1.52)	0.0138**		

* Corrected for age and gender. ** Corrected for age, gender, systolic BP, HDL, LDL, triglycerides, type 2 diabetes, hypertension medication, statins and CRP.

The association between the lead SNP and common cIMT remains statistically significant when the additional cardiovascular risk factors are corrected for, Table 9. The signal in the AGES cohort becomes less significant with the p-value increasing from 0.00013 to 0.00115 and the beta decreasing from 0.0184 (95% CI; 0.00901-0.0279) to 0.0136 (95% CI; 0.00711-0.0202). The standard error decreases, going from 0.00346 to 0.00333. The signal in the REFINE cohort becomes more statistically significant with the p-value decreasing from 0.000215 to 4.27×10^{-05}. The beta increases from 0.0128 (95% CI; 0.00604-0.0196) to 0.0136 (95% CI; 0.00711-0.0202) with the standard error decreasing from 0.00346 to 0.00333. Test of interaction (effect modification) by cohort was not statistically significant (p=0.814).

Table 9. Common cIMT linear regression results. The coded lead SNP allele was the minor allele. Only individuals with measurements for all risk factors were included in the analyses.

Cohort	β (95% CI)	SE	p	N (Total)
AGES	0.0184 (0.00901-0.0279)	0.00481	0.00013*	3,047
	0.0152 (0.00604-0.0243)	0.00467	0.00115**	
REFINE	0.0128 (0.00604-0.0196)	0.00346	0.000215*	5,424
	0.0136 (0.00711-0.0202)	0.00333	4.27E-05**	

* Corrected for age and gender. ** Corrected for age, gender, systolic BP, HDL, LDL, triglycerides, type 2 diabetes, hypertension medication, statins and CRP.

4.3 Comparing the HapMap and 1000 Genomes imputations

There is essentially no difference between the results seen for the lead SNP when comparing the 1000 Genomes imputation used in this study and the HapMap imputation that was used in the CHARGE meta-analysis. There the lead SNP gave a p-value of 0.00207 and an OR of 1.29 (95% CI; 1.10-1.52) for carotid plaque and a p-value of 7.71×10^{-05} and a beta of 0.0193 (95% CI; 0.00972-0.0288). The lead SNP MAF in the 1000 Genomes imputation is equal to 0.141 and the HapMap imputation gave a MAF of 0.138. The imputation quality of the lead SNP is very high in both the 1000 Genomes and HapMap imputation. The r^2 value is 0.9712 in HapMap which is similar to the 1000 Genomes imputation value of 0.9708.

4.4 Refining the lead SNP signal

To address specific aim 2 the lead SNP signal was examined within the context of the LD structure of *EDNRA*. Other *EDNRA* SNPs, both genotyped and imputed, were investigated in an attempt to refine the lead SNP signal.

4.4.1 LD structure of EDNRA

The LD structure of EDNRA, specifically the region between 148,611,000 and 148,695,000 on chromosome 4 (NCBI36/hg18), can be seen in Figure 8. There are four LD blocks that cover *EDNRA* and a fifth that is downstream of its eighth exon. These LD blocks are split up by three recombination peaks. The 10kb upstream region of *EDNRA* in addition to its first two exons is covered by a single LD block. A part of the first LD block is in relatively strong LD with the second and third LD blocks. The LD between the first block and the fourth is weak but not non-existent. The downstream block is in weak LD with the fourth block but there is little to no LD between the fifth LD block and the first three. This is explained by the large recombination peak that separates them. The first LD block is the largest, followed by the fourth, with the second, third and fifth being comparatively small. The third *EDNRA* exon lies within the third LD block, with the fourth and fifth lying within the fourth LD block. The sixth and seventh exons lie within the recombination peak and the final, and largest exon, lies on the other side of said peak.

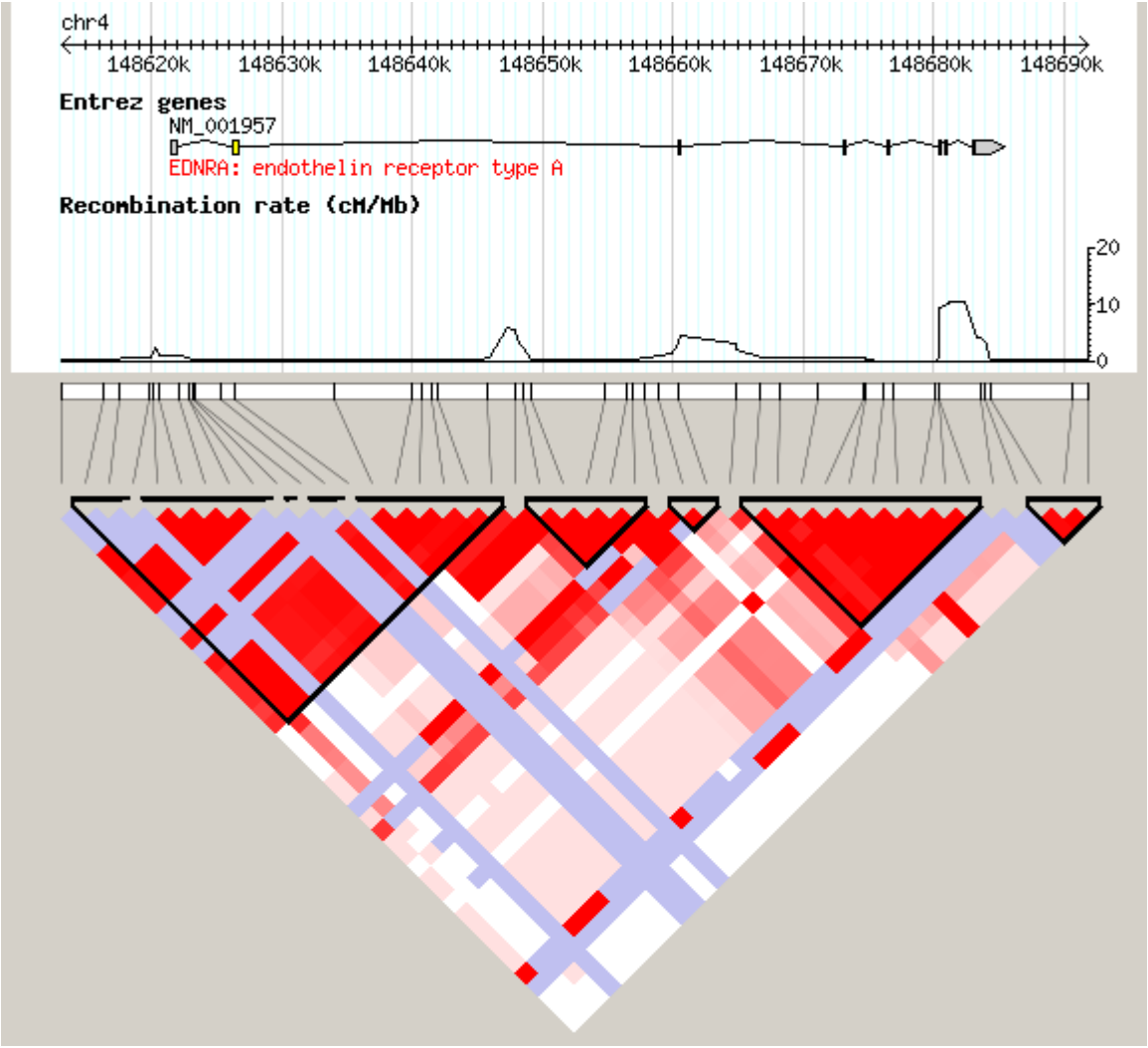

Figure 8. LD plot of *EDNRA*. The *EDNRA* gene, introns and exons are shown, along with the recombination rate across the gene and the upstream and downstream regions. Taken from Haploview and the LD data used was from HapMap version 3 release 27 from the CEU analysis panel. Co-ordinates 10kb either side of *EDNRA*, chr4:148,611,000-148,695,000 in the NCBI36/hg18 assembly of the Human genome. D'/LOD colour scheme.

4.4.2 Regional LD plot for the lead SNP

For a clearer picture of the LD between the lead SNP and surrounding SNPs a regional LD plot was generated based on denser SNP maps that are available from the 1000 Genomes project data, Figure 9. The strong LD (r^2=0.8) boundary is marked with a vertical dotted line and covers the region containing exon one and two of EDNRA as well as the first intron and part of the second. The upstream region of *EDNRA* that contains SNPs in strong LD the lead SNP extends to approximately 28kb from the start of *EDNRA*. The exact chromosomal co-ordinates of this strong LD region are chr4:148,363,840-148,414,651 (NCBI37/hg19). In addition to seeing the SNPs in strong LD with the lead SNP it is also possible to see that there is moderate LD further into *EDNRA*. If the hypothesis that there is a truly functional variant behind the lead SNP is correct then it is likely that it will be a variant in strong LD with the lead SNP and therefore this region is of high interest.

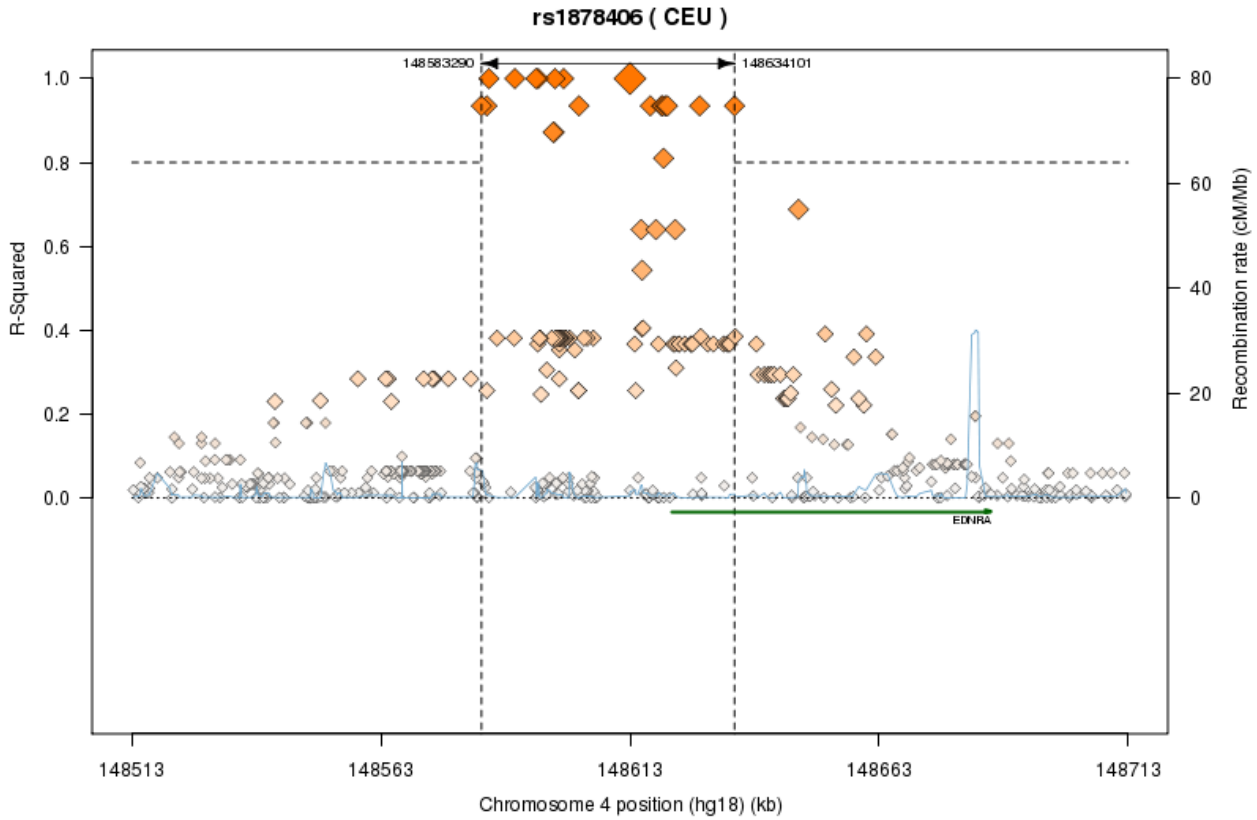

Figure 9. Regional LD plot of rs1878406 and 1000 Genomes Pilot 1 SNPs. Vertical dotted lines mark the region where pairwise LD is equal to or higher than 0.8 r^2 between target SNP rs1878406 and 1000 Genomes SNPs. Plotted with SNAP (57).

4.5 Association between *EDNRA* SNPs and both carotid plaque and common IMT

Examination of both genotyped and 1000G imputed SNPs from the AGES Reykjavik Study GWAS was conducted.

4.5.1 Association between carotid plaque and genotyped SNPs in AGES

There are 11 genotyped SNPs that lie within or in close proximity to *EDNRA*. Their distribution can be seen in Figure 10 and exact chromosomal coordinates for each can found in Table 3. The association between carotid plaque and these 11 genotyped SNPs was tested, Table 10. Of those SNPs tested one proved to be significantly associated with carotid plaque (alpha=0.01). It was the rs6842241 SNP which had an OR of 1.30 (95% CI; 1.10-1.53). Another SNP, rs10008744, showed a suggestive association with carotid plaque and had an OR of 1.16 (95% CI; 1.01-1.32).

Table 10. Single-point logistic regression results for carotid plaque. N=3,058. The SNPs are ordered by their chromosomal coordinates, from lowest to highest.

SNP	Minor allele	β	SE	OR (95% CI)	p	MAF
rs6842241**	A	0.260	0.0831	1.30 (1.10-1.53)	0.00178	0.133
rs10024834	G	0.0987	0.0614	1.10 (0.979-1.24)	0.108	0.301
rs6537483	A	0.100	0.0614	1.10 (0.979-1.25)	0.105	0.301
rs4563479	T	-0.0638	0.0939	0.938 (0.780-1.13)	0.496	0.092
rs6827096	T	-0.0691	0.0763	0.933 (0.804-1.08)	0.366	0.153
rs1878404	A	0.0476	0.0708	1.05 (0.913-1.20)	0.502	0.199
rs10008744*	C	0.143	0.0671	1.15 (1.01-1.32)	0.0332	0.228
rs4639051	G	0.0767	0.0760	1.08 (0.930-1.25)	0.313	0.160
rs5334	A	0.109	0.0654	1.12 (0.981-1.27)	0.0941	0.245
rs5343	T	-0.101	0.0585	0.904 (0.806-1.01)	0.0847	0.331
rs10028838	G	-0.0697	0.0571	0.933 (0.834-1.04)	0.222	0.375

* p≤0.05 **p≤0.01

A regional association plot was constructed to further examine the relationship between the lead SNP and the genotyped SNPs shown to be significantly associated with carotid plaque, Figure 10. The lead SNP is in strong LD with rs6842241 (r^2=0.935, D'=1) and weak LD with rs10008744 (r^2=0.052, D'=0.278). There is weak between rs6842241 and rs10008744 (r^2=0.065, D'=0.322). The lead SNP and rs6842241 represent the same signal, where their beta and OR are equal (β=0.260, OR=1.30) and they are in strong LD with one another.

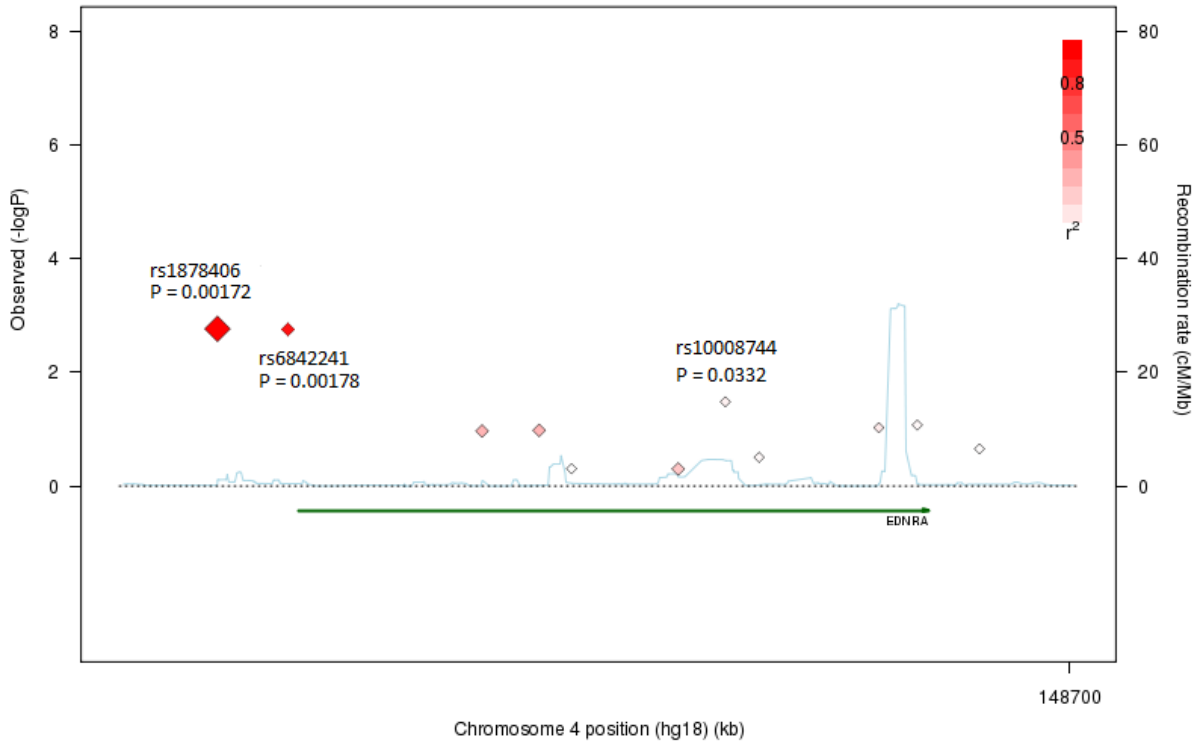

Figure 10. Carotid plaque regional association plot for lead SNP and the genotyped SNPs. The r^2 shown represents the LD between the lead SNP and all other SNPs. Plotted with SNAP using the CEU population panel from the 1000 Genome Pilot 1 dataset (57).

4.5.2 Association between common cIMT and genotyped SNPs in AGES

The association between common cIMT and the 11 genotyped *EDNRA* SNPs was tested, Table 11. Four SNPs were significantly associated with common cIMT (alpha=0.05). The rs6842241 SNP had a beta of 0.0193 (95% CI; 0.00980-0.0288), the rs1878404 SNP had a beta of 0.0138 (95% CI; 0.00547-0.0221), the rs10024834 SNP had a beta of 0.0110 (95% CI; 0.00380-0.0182) and the rs6537483 SNP had a beta of 0.0111 (95% CI; 0.00390-0.0183).

Table 11. Single-point linear regression results for common cIMT. N=3,073. The SNPs are ordered by their chromosomal coordinates, from lowest to highest.

SNP	Minor allele	β (95% CI)	SE	p	MAF
rs6842241**	A	0.0193 (0.00980/0.0288)	0.00483	6.21E-05	0.133
rs10024834**	G	0.0110 (0.00380/0.0182)	0.00368	0.00279	0.301
rs6537483**	A	0.0111 (0.00390/0.0183)	0.00368	0.00266	0.301
rs4563479	T	0.00573 (-0.00548/0.0169)	0.00572	0.317	0.0921
rs6827096	T	-0.00261 (-0.0117/0.00648)	0.00464	0.574	0.153
rs1878404**	A	0.0138 (0.00547/0.0221)	0.00425	0.00118	0.199
rs10008744	C	-0.00104 (-0.00886/0.00678)	0.00399	0.795	0.228
rs4639051	G	-0.00493 (-0.0138/0.00397)	0.00454	0.278	0.160
rs5334	A	0.00124 (-0.00640/0.00888)	0.00390	0.751	0.245
rs5343	T	-0.00237 (-0.00933/0.00459)	0.00355	0.504	0.331
rs10028838	G	-0.000708 (-0.00747/0.00605)	0.00345	0.837	0.375

** p≤0.01

A regional association plot was constructed to further examine the relationship between the lead SNP and the genotyped SNPs shown to be significantly associated with common cIMT, Figure 11. The lead SNP is in strong LD with rs6842241 (r^2=0.935, D'=1), the same weak LD with both rs10024834 and rs6536483 (r^2=0.294, D'=0.895) and also weak LD with rs1878404 (r^2=0.221, D'=0.544).

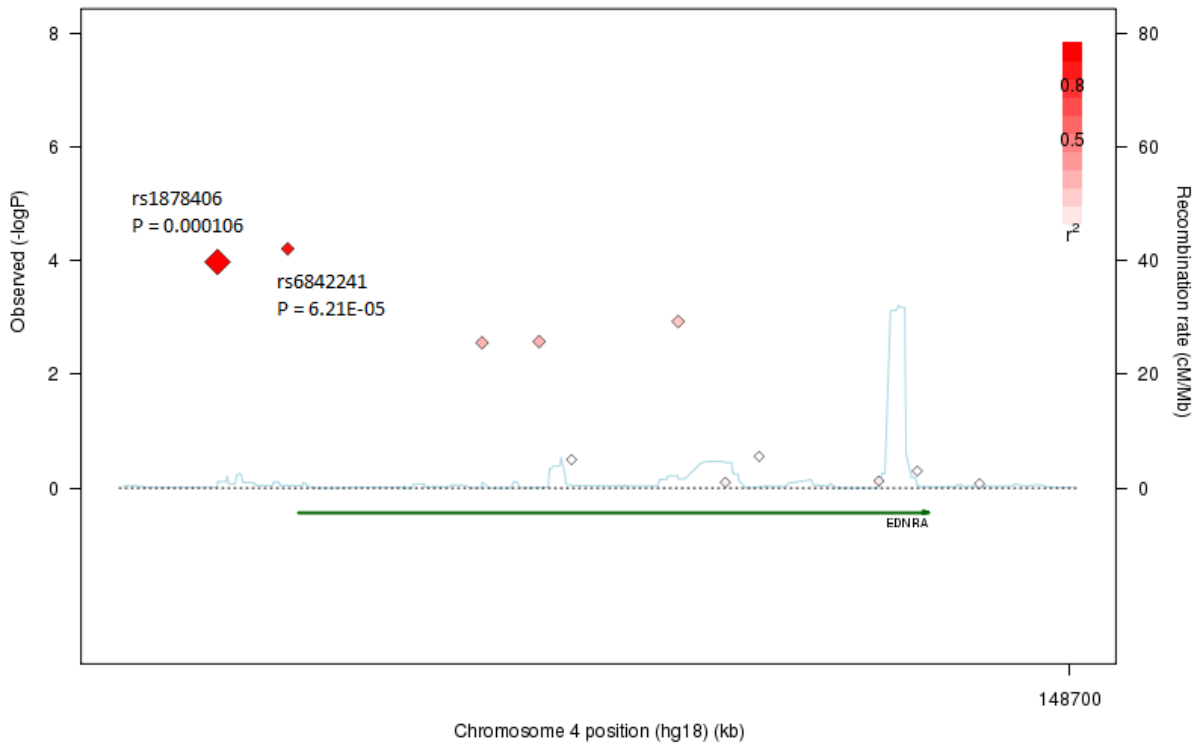

Figure 11. Common cIMT regional association plot for lead SNP and the genotyped SNPs. The r^2 shown represents the LD between the lead SNP and all other SNPs. Plotted with SNAP (57) using the CEU population panel from the 1000 Genome Pilot 1 dataset.

4.5.3 Association between carotid plaque and 1000G imputed variants in AGES

The association between 1000G imputed *EDNRA* variants and carotid plaque was tested, Table 12. Of the 149 variants tested, 15, including the lead SNP, were significantly associated with carotid plaque (alpha=0.01). Within those 15 variants there appear to be two separate signals. The complete results for single-point variant associations are in Table 19 in the Appendix. No SNPs passed Bonferroni correction.

Signal one: The ORs of the 13 variants differ only slightly and range from 1.28 (95% CI; 1.1-1.49) to 1.33 (95% CI; 1.12-1.59). Their MAF are also similar and range from 0.125 to 0.188.

Signal two: The variants rs6854578 and rs10305929 share an OR of 2.25 (95% CI; 1.27-3.98) and a p-value of 0.00548, with their MAF only differing slightly (0.681 vs. 0.680).

Table 12. Single-point variant logistic regression results for carotid plaque, adjusted for age and gender (alpha=0.01). The results are ordered by their chromosomal position and only results with a p-value below 0.01 are included.

SNP	Minor allele	β	SE	OR (95% CI)	p	MAF
rs17612742	C	0.283	0.0880	1.33 (1.12-1.58)	0.00128	0.129
rs78049276	C	0.287	0.0897	1.33 (1.12-1.59)	0.00136	0.125
rs6841473	T	0.276	0.0865	1.32 (1.11-1.56)	0.00142	0.130
rs7659823	G	0.245	0.0778	1.28 (1.1-1.49)	0.00163	0.188
rs6820938	G	0.245	0.0782	1.28 (1.1-1.49)	0.00171	0.188
rs1878406	T	0.260	0.0828	1.3 (1.1-1.53)	0.00172	0.141
rs6855875	T	0.248	0.0791	1.28 (1.1-1.5)	0.00174	0.185
rs73855814	T	0.263	0.0839	1.3 (1.1-1.53)	0.00175	0.133
rs10305839	T	0.262	0.0837	1.3 (1.1-1.53)	0.00178	0.132
rs6841581	A	0.261	0.0836	1.3 (1.1-1.53)	0.00178	0.133
rs72957606	G	0.261	0.0836	1.3 (1.1-1.53)	0.00179	0.133
rs10305838	C	0.261	0.0836	1.3 (1.1-1.53)	0.00179	0.133
rs6823438	T	0.259	0.0833	1.3 (1.1-1.53)	0.00187	0.142
rs6854578	C	0.810	0.291	2.25 (1.27-3.98)	0.00548	0.0187
rs10305929	A	0.810	0.292	2.25 (1.27-3.98)	0.00548	0.0187

A regional association plot was constructed to further examine the relationship between the lead SNP and the imputed variants shown to be significantly associated with carotid plaque, Figure 12. All 149 SNPs tested are included in the plot. Of the 15 significant SNPs there are 13 that have LD information available in the 1000 Genomes pilot 1 dataset and therefore can be plotted using SNAP (57).

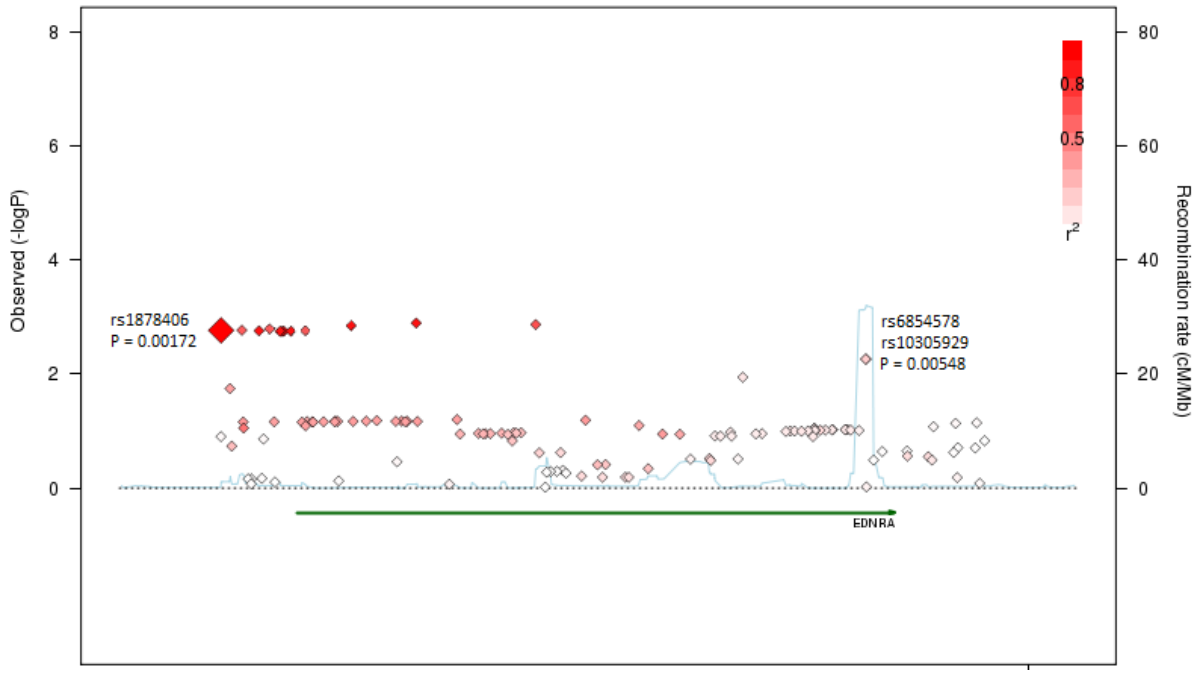

Figure 12. Carotid plaque regional association plot of logistic regression p-values for all imputed SNPs. Target SNP is rs1878406. Plotted with SNAP (57) using the CEU population panel from the 1000 Genome Pilot 1 dataset.

Of the 13 SNPs in signal one there are 11 that are present in the dataset. All 11 SNPs are in strong LD with the lead SNP with variations occurring mostly in their r^2 values. There are six SNPs, rs73855814, rs72957606, rs10305838, rs6841581, rs6841473 and rs17612742 that share the same lead SNP pairwise LD values (r^2=0.935, D'=1) and also have strong LD amongst themselves (with the r^2 ranging from 0.602 to 1 and all D' being equal to 1). The rs10305839 SNP also has its own pairwise LD values (r^2=0.810, D'=1). Three SNPs, rs6820938, rs7659823 and rs6855875 share the same pairwise LD values (r^2=0.640, D'=0.927) and are all in 'perfect' LD with one another (r^2=1, D'=1). The rs78049276 SNP has its own pairwise LD values (r^2=0.688, D'=1).

The second signal is represented in the regional association plot by both SNPs, rs6854578 and rs10305929. They are in equally strong LD with the lead SNP (r^2=0.195, D'=1), although the weak r^2 value reflects the low MAF of both SNPs.

4.5.4 Association between common cIMT and 1000G imputed variants in AGES

The association between 1000G imputed *EDNRA* variants and common cIMT was tested, Table 13. Of the 149 imputed variants tested, including the lead SNP, there were 60 that were significantly associated (alpha=0.01) with common cIMT. Those 60 variants have positive betas ranging from 0.011 (95% CI; 0.00379-0.0182) to 0.0200 (95% CI; 0.01-0.03). There are 18 SNPs that have betas ranging between 0.0190 and 0.0200, and their p-values range from 0.000928 to 2.90×10^{-05}. The genotyped and imputed SNPs tested total 160 and as that is the total number of tests performed it is used to calculate the Bonferroni correction (alpha=0.05/160=0.000313). There were a total of 16 SNPs that passed Bonferroni correction. The complete results for single-point variant associations are in Table 20 in the Appendix.

Table 13. Single-point variant linear regression results for common cIMT, adjusted for age and gender. The results are ordered by their chromosomal position and only results with a p-value below 0.01 are included. Variants below the Bonferroni correction alpha threshold of 0.000313 are labelled with *.

SNP	Minor allele	β (95% CI)	SE	p	MAF
rs6853791*	T	0.0143 (0.00658-0.022)	0.00394	0.000292	0.301
rs6823438*	T	0.0187 (0.00919-0.0282)	0.00485	0.000111	0.142
rs1987147	C	0.0119 (0.00422-0.0196)	0.00392	0.00237	0.424
rs1808039	C	0.0117 (0.00419-0.0192)	0.00383	0.00233	0.414
rs1808038*	C	0.0146 (0.00682-0.0224)	0.00397	0.000236	0.306
rs1878406*	T	0.0187 (0.00925-0.0281)	0.00482	0.000106	0.141
rs13144507*	C	0.015 (0.00741-0.0226)	0.00387	0.000104	0.286
rs12509416	C	0.0119 (0.00463-0.0192)	0.00371	0.00131	0.391
rs6820938*	G	0.019 (0.01-0.028)	0.00458	3.14E-05	0.188
rs6537480	A	0.0133 (0.00536-0.0212)	0.00405	0.00102	0.286
rs6537481	G	0.0129 (0.00459-0.0212)	0.00424	0.00240	0.231
rs73855814*	T	0.0195 (0.00995-0.029)	0.00487	6.44E-05	0.133
rs7659823*	G	0.019 (0.0101-0.0279)	0.00455	2.90E-05	0.188
rs9991584	G	0.0131 (0.00526-0.0209)	0.00400	0.00105	0.291
rs72957606*	G	0.0194 (0.00987-0.0289)	0.00486	6.38E-05	0.133

rs10305838*	C	0.0194 (0.00987-0.0289)	0.00486	6.39E-05	0.133
rs10305839*	T	0.0195 (0.00997-0.029)	0.00486	6.34E-05	0.132
rs6841581*	A	0.0194 (0.00987-0.0289)	0.00486	6.45E-05	0.133
rs1801708	A	0.0127 (0.00484-0.0206)	0.00401	0.00151	0.288
rs6855875*	T	0.0192 (0.0101-0.0283)	0.00463	3.37E-05	0.185
rs4835083	G	0.0127 (0.00484-0.0206)	0.00401	0.00156	0.290
rs4835084	A	0.0128 (0.00496-0.0206)	0.00400	0.00138	0.288
rs984457	T	0.0128 (0.00496-0.0206)	0.00400	0.00140	0.287
rs984458	T	0.0128 (0.00494-0.0207)	0.00401	0.00143	0.287
rs56015601	T	0.0128 (0.00496-0.0206)	0.00400	0.00143	0.287
rs10305860	A	0.0127 (0.00486-0.0205)	0.00400	0.00145	0.287
rs9308218	T	0.0127 (0.00486-0.0205)	0.00400	0.00144	0.288
rs10305862	T	0.0127 (0.00486-0.0205)	0.00400	0.00142	0.287
rs6841473*	T	0.0199 (0.0101-0.0297)	0.00501	7.00E-05	0.130
rs6823537	G	0.0126 (0.00478-0.0204)	0.00399	0.00165	0.287
rs702757	A	0.0125 (0.00468-0.0203)	0.00399	0.00176	0.287
rs4835410	G	0.0123 (0.0045-0.0201)	0.00398	0.00203	0.288
rs7663891	C	0.0122 (0.0044-0.02)	0.00398	0.00225	0.287
rs952402	C	0.0122 (0.0044-0.02)	0.00398	0.00224	0.287
rs1996374	A	0.0121 (0.00432-0.0199)	0.00397	0.00227	0.285
rs1996373	G	0.0121 (0.00432-0.0199)	0.00397	0.00229	0.286
rs17612742*	C	0.02 (0.01-0.03)	0.00509	8.36E-05	0.129
rs1517134	T	0.012 (0.00424-0.0198)	0.00396	0.00245	0.285
rs13143677	A	0.0116 (0.0039-0.0193)	0.00393	0.00326	0.282
rs1400554	T	0.0113 (0.00403-0.0186)	0.00371	0.00226	0.301
rs6840756	C	0.011 (0.00379-0.0182)	0.00368	0.00274	0.300
rs4835086	A	0.011 (0.00377-0.0182)	0.00369	0.00275	0.300
rs17474816	A	0.0111 (0.00387-0.0183)	0.00369	0.00270	0.300
rs6848108	G	0.0111 (0.00387-0.0183)	0.00369	0.00270	0.300
rs7655670	C	0.0111 (0.00389-0.0183)	0.00368	0.00268	0.300
rs17474823	G	0.011 (0.00375-0.0183)	0.00370	0.00300	0.299
rs57886575	A	0.011 (0.00373-0.0183)	0.00371	0.00288	0.306
rs10011966	A	0.011 (0.00379-0.0182)	0.00368	0.00270	0.301
rs6819740	T	0.0111 (0.00389-0.0183)	0.00368	0.00266	0.300
rs6537482	A	0.0111 (0.00389-0.0183)	0.00368	0.00266	0.301
rs78049276*	C	0.0197 (0.00955-0.0299)	0.00518	0.000147	0.125
rs6537487	C	0.0145 (0.00574-0.0233)	0.00447	0.00113	0.189
rs139452637	T	0.0195 (0.00798-0.031)	0.00588	0.000928	0.0999
rs75736224	G	0.0196 (0.00813-0.0311)	0.00585	0.000799	0.0999
rs6537489	T	0.0144 (0.0058-0.023)	0.00439	0.00101	0.192
rs6821368	T	0.0144 (0.0058-0.023)	0.00439	0.00101	0.192
rs4835411	C	0.0195 (0.00815-0.0308)	0.00579	0.000775	0.101
rs7674137	G	0.0143 (0.00577-0.0228)	0.00435	0.00104	0.194
rs4835412	A	0.0191 (0.00787-0.0303)	0.00573	0.000874	0.102
rs908581	T	0.0191 (0.00789-0.0303)	0.00572	0.000859	0.102

A regional association plot was constructed to further examine the relationship between the lead SNP and the imputed variants shown to be significantly associated with common cIMT, Figure 13. The plot includes the p-values of all 123 SNPs that have LD information available in the 1000 Genomes pilot 1 dataset and therefore can be plotted using SNAP and have their pairwise LD values with the lead SNP displayed (57). There are 52 out of the 60 common cIMT significantly associated (p≤0.01) SNPs included in the plot.

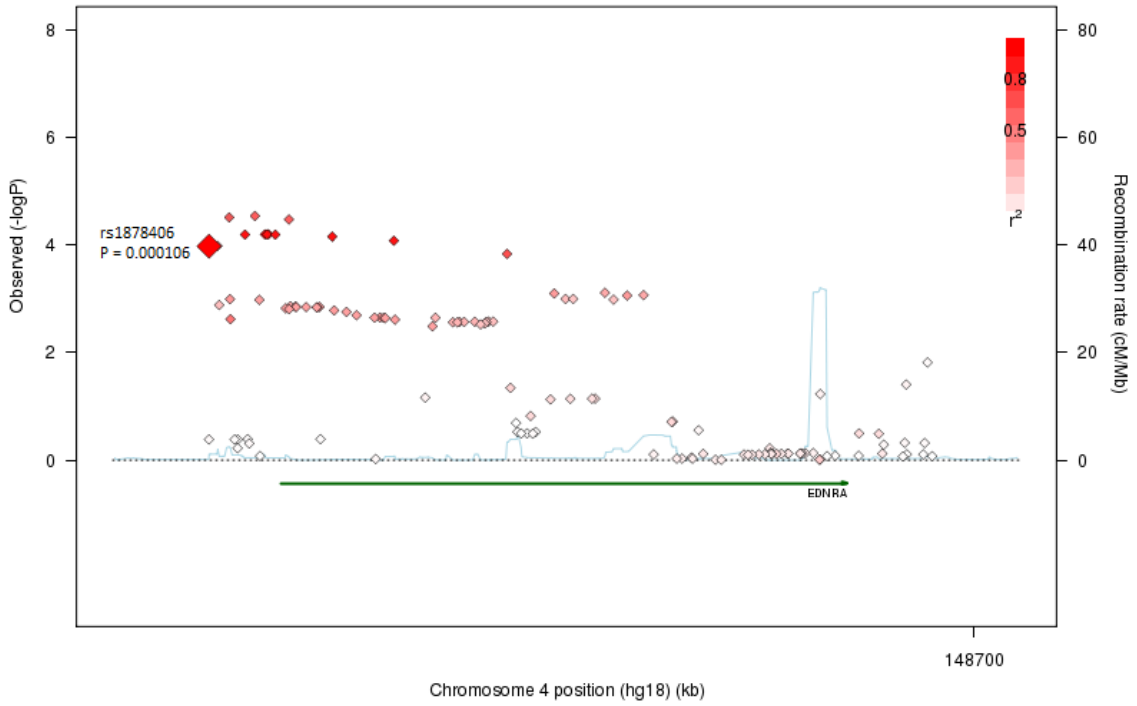

Figure 13. Common cIMT regional association plot of linear regression p-values for all imputed variants. Target SNP is rs1878406. Plotted with SNAP (57) using the CEU population panel from the 1000 Genome Pilot 1 dataset.

The lead SNP is in LD with all of the significantly associated SNPs with r^2 ranging from 0.221 to 0.935 and D' ranging from 0.544 to 1. The SNPs that are in strongest LD with the lead SNP are those with the strongest effect sizes and the lowest p-values. Of the 18 SNPs that have an effect size similar to the lead SNP (beta ranging from 0.0190 to 0.0200) there are 15 that have LD information available. The majority of these 15 SNPs are in strong LD with the lead SNP, with r^2 values ranging from 0.640 to 0.935 and D' values ranging from 0.927 to 1. Four of the 15 SNPs are in weaker LD with the lead SNP with r^2 values ranging from 0.336 to 0.391 and D' values ranging from 0.855 to 0.871. These four SNPs are located within a cluster of SNPs that lie around 30 to 40 thousands bases downstream of the lead SNP (according to Hg18).

4.6 Analysis of phased haplotypes

4.6.1 Association between haplotypes and carotid plaque

In accordance with the third specific aim, the 53 predicted haplotypes were tested for their association with carotid plaque to determine whether any of the haplotypes are capturing additional signal across *EDNRA*. Haplotypes were available for all AGES Reykjavik Study GWAS individuals (N=3,219). The

association between haplotype counts and carotid plaque was tested, Table 21 in the Appendix. A single haplotype, haplotype 10 (AGACCACAGCG, frequency=0.0089) showed suggestive association (p=0.0453) with carotid plaque giving an OR of 1.95 (95% CI; 1.01-3.76) when adjusting for age and gender.

4.6.2 Sliding-window haplotype analysis

The first sliding window consisted of the rs6842241, rs10024834, rs6537483, rs4563479 and rs6827096 SNPs. Of the eight predicted haplotypes one (AGACC) was significantly associated with carotid plaque. The AGACC haplotype, containing the rs6842241 carotid plaque minor allele A, gave an OR of 1.29 (95% CI; 1.09-1.53), a p-value of 0.00275 and had a frequency of 0.125. The CAGTT haplotype, containing the rs6842241 major allele C, gave an OR of 0.574 (95% CI; 0.358-0.921), had a frequency of 0.0116 and showed a suggestive association with a p-value of 0.0213.

Sliding-window two consisted of the rs10024834, rs6537483, rs4563479, rs6827096 and rs1878404 SNPs. None of the 9 phased haplotypes showed statistically significant association with carotid plaque.

Sliding-window three consisted of the rs6537483, rs4563479, rs6827096, rs1878404 and rs10008744 SNPs. One of the 14 phased haplotypes showed a suggestive association with carotid plaque. The ACCGC haplotype gave an OR of 1.82 (95% CI; 1.04-3.19), a p-value of 0.0372 and had a frequency of 0.0120.

Sliding-window four consisted of the rs4563479, rs6827096, rs1878404, rs10008744 and rs4639051 SNPs. None of the 17 phased haplotypes showed a statistically significant association with carotid plaque.

Sliding-window five consisted of the rs6827096, rs1878404, rs10008744, rs4639051 and rs5334 SNPs. None of the 22 phased haplotypes showed a statistically significant association with carotid plaque.

Sliding-window six consisted of the rs1878404, rs10008744, rs4639051, rs5334 and rs5343 SNPs. One of the 21 phased haplotypes showed a suggestive association with carotid plaque. The ACAGC haplotype gave an OR of 1.88 (95% CI; 1.04-3.44), a p-value of 0.0418 and had a frequency of 0.0106.

Sliding-window seven consisted of the rs10008744, rs4639051, rs5334, rs5343 and rs10028838 SNPs. One of the 15 phased haplotypes showed a suggestive association with carotid plaque. The CAGCG haplotype gave an OR of 1.98 (95% CI; 1.08-3.62), a p-value of 0.0266 and had a frequency of 0.0109.

4.6.3 Correlation between carotid plaque associated haplotype and SNPs

To test whether haplotype 10 is capturing the lead SNP signal or other signal within *EDNRA*, correlation analysis was performed.

Haplotype 10 shows weak positive correlation with the lead SNP, Pearson's r equal to 0.24 ($p<0.0001$). The majority, except for two, of the imputed and genotyped SNPs that were significantly

associated with carotid plaque showed a very similar level of correlation to that shown by the lead SNP, Pearson's r ranged between 0.22 and 0.27 ($p<0.0001$). The two SNPs that differed from the lead SNP were the 2.25 OR SNPs (rs6854578, rs10305929). They both showed the same strong positive correlation with the haplotype, Pearson's r equal to 0.82 ($p<0.0001$).

4.7 Functional Annotation

4.7.1 RegulomeDB

The SNPs shown to be significant for common cIMT and carotid plaque, including the lead SNP, were annotated with regard to regulatory information through the use of RegulomeDB, Table 14. A total of 64 common cIMT SNPs and 16 carotid plaque SNPs were annotated. There were 27 SNPs that had no evidence of an effect of TF binding, with 21 of those being common cIMT SNPs and 6 being carotid plaque SNPs. The rs139452637 SNP, significantly associated with common cIMT, was not annotated in RegulomeDB due to an unknown error that rejected its inclusion in any queries. There were 14 SNPs that were significantly associated with both carotid atherosclerotic phenotypes. Of those 14 SNPs there were six that had no evidence of an effect on TF binding. The highest RegulomeDB score that a SNP is annotated with here is 2b, which indicates that a SNP lies within a region that has evidence for TF binding, any TF motif, a DNase footprint and a DNase peak. There is a single SNP with this RegulomeDB score and it is rs1400554.

There are three other scores, in addition to the 2b and 'no data' scores, that the significant SNPs receive from the RegulomeDB annotation. Those scores are 4, 5 and 6. The score of 4 denotes that the SNP lies within a region that has evidence for both the binding of a TF and that lies within a DNase peak. The score of 5 denotes that the region which the SNP lies in does not have both TF binding and a DNase peak, but one of the two. A score of 6 denotes that there is other evidence of the regions involvement in regulation.

	Significantly associated SNPs		
RegulomeDB Score	common cIMT	carotid plaque	Both phenotypes
2b	1	0	0
4	3	1	1
5	23	6	6
6	15	3	1
No data	21	6	6
Total	63	16	14

Table 14. Summary table for RegulomeDB functional annotation of significant SNPs, both for common cIMT and carotid plaque. Significantly associated SNPs are split into three groups: those that associated with common cIMT, those that associated with carotid plaque and those that associated with both common cIMT and carotid plaque. The rs139452637 SNP was not annotated in RegulomeDB as it caused an error to occur.

4.7.2 SNPnexus

SNPnexus was used to annotate both genotyped and imputed SNPs from the AGES Reykjavik Study GWAS that were significantly associated ($p \leq 0.01$) with either carotid plaque or common cIMT (64, 65). In total there were 66 SNPs that were significantly associated with either phenotype, including the lead SNP.

Through SNPnexus the SNPs were annotated with regard to their genomic location. The NCBI reference sequence (RefSeq) database was used to this purpose and informed whether a SNP was within a coding, non-coding, intronic or intergenic region of the genome. There was RefSeq information available for 52 of the SNPs. There was a single 5' UTR SNP, five SNPs in the 5' upstream region and the other 46 SNPs were intronic. The NM_001957 NCBI Reference sequence was used which is the sequence for Human *EDNRA* transcript variant 1 mRNA. No difference was seen in the number of SNPs or their annotation if other reference sequences were used.

The SNPs were also annotated with regard to transcription factor binding sites and it was found that one of the SNPs, rs17612742, lies within a transcription factor binding site for the transcription factor MEF-2A which stands for MEF2A myocyte enhancer factor 2A.

Conservation was assessed using both GERP++ and PHAST methods through SNPnexus and according to the Phast conversation method the rs1517134 and rs78049276 SNPs are conserved. A lod score of 54 is given to the former and a lod score of 162 to the latter. The PHASE conservation probability score of 395 was given to the rs1517134 SNP and a score of 513 was given to the rs78049276 SNP. The GERP++ RS score finds the rs1517134 and rs78049276 SNPs also to be conserved with RS scores of 306 and 567 respectively. In addition the GERP++ RS score of 538 was assigned to the rs6821368 SNP.

5 Discussion

5.1 A summary of the study's aim and main findings

The CHARGE consortium's GWAS found that the apparently non-functional rs1878406 SNP, which lies approximately 8.4kb upstream from the *EDNRA* gene, was significantly associated with both carotid plaque and common cIMT in 31,211 individuals (28). The SNP was also shown to be significantly associated with CAD through testing conducted in the CARDIoGRAM consortium (28). The aim of this study was first to replicate the CHARGE meta-analysis results in an independent cohort and then to refine the lead SNP signal with the aim of identifying a functional variant predicted to lie behind the lead SNP signal. This was to be accomplished through analysis of additional SNPs within and around the *EDNRA* gene in addition to investigating phased haplotypes and their association with carotid plaque.

The main findings are that the associations between the lead SNP and the carotid plaque and common cIMT phenotypes were replicated in the REFINE Reykjavik Study cohort, which is independent from the CHARGE cohorts. Multiple candidates for the functional variant have been found and the most promising of these is rs6841581, which was shown to be in strong LD with the lead SNP through systematic testing of pairwise LD. An additional signal for carotid plaque was found in two intronic *EDNRA* SNPs and there appears to be a single haplotype that captures this additional signal.

5.2 Replication of the CHARGE meta-analysis GWAS signal (rs1878406) in REFINE

The association between the lead SNP and both carotid phenotypes was confirmed in the REFINE Reykjavik Study cohort of 5,521 individuals. For carotid plaque the lead SNP was significantly associated (p=0.0137) with an OR of 1.25 (95% CI; 1.05-1.49). For common cIMT the lead SNP was significantly ($p=2.54x10^{-04}$) associated with a beta of 0.0126 (95% CI; 0.00588-0.0193).

In addition to the replication in the REFINE cohort the lead SNP's association was re-tested in the AGES Reykjavik Study GWAS cohort using the 1000 Genomes Project data for imputation. In the AGES cohort the lead SNP was significantly associated with carotid plaque (p=0.00172) with an observed OR of 1.30 (95% CI; 1.10-1.53). The lead SNP was also significantly associated with common cIMT ($p=1.06x10^{-04}$) and gave an effect size of beta=0.0187 (95% CI; 0.00925-0.0281). The lead SNP's result using the 1000 Genomes imputation in AGES did not differ from the HapMap imputation results used for AGES in the CHARGE meta-analysis.

In the CHARGE meta-analysis an OR of 1.22 (95% CI; 1.15-1.29) was observed for carotid plaque and a beta of 0.0087 (95% CI; 0.0052-0.012) for common cIMT. The CHARGE meta-analysis results for both carotid phenotypes are closer to the results seen in REFINE than to AGES as can be seen from above. The differences between the AGES and REFINE cohorts were examined to see whether the differences between them, with regard to the lead SNP signal for common cIMT and carotid plaque, were statistical significant. Both cohorts are samples of the Icelandic population, with the main difference being that the AGES cohort represents the generation born 1907-1935, and the REFINE

cohort represents the generation born 1936-1986. However, when the two cohorts were combined the interaction between the cohort and the lead SNP was not statistically significant when adjusting for age and gender. The interaction was also not statistically significant when adjusting for age and gender in addition to other risk factors, specifically: systolic BP, HDL, LDL, triglycerides, type 2 diabetes, hypertension medication use, statins and CRP. Therefore we have not found evidence in our data that the association between the lead SNP and either carotid plaque or common cIMT is different between the AGES and REFINE cohorts. This underlines the point that the two cohorts are samples of the same population.

In the CHARGE meta-analysis paper the lead SNP was reported to be associated with CAD in the CARDIoGRAM consortium data (28). An attempted replication of this association within REFINE and AGES using coronary event endpoints was not successful. Although the association does not reach statistical significance within AGES, the OR of 0.888 (95% CI; 0.727-1.08) is consistent with that observed in the CARDIoGRAM consortium, which had an OR of 0.91 (95% CI; 0.87-0.95). The CARDIoGRAM result is given for the major allele C while the result for the CHARGE meta-analysis is given for the T allele and so to make the results comparable the ORs were recalculated. The recalculated CARDIoGRAM OR for the T allele is 1.10 (95%CI; 1.05-1.15), which is in line with the AGES T allele OR of 1.13 (95% CI; 0.922-1.38). Therefore the direction of the ORs for CAD in CARDIoGRAM and coronary events in AGES are in line with the increased OR seen for carotid plaque in the CHARGE analysis. The statistical power of the coronary event analysis was increased by combining the AGES and REFINE cohorts. The association did not reach statistical significance and had an OR of 1.09 (95% CI; 0.923-1.28) for the T minor allele. The prevalence of coronary events in AGES and REFINE is 15.0% and 3.8% respectively. The much lower coronary event prevalence in REFINE (3.8%) compared to AGES (15.0%), may explain the direction of the OR in REFINE, which was 0.719 (95% CI;0.380-1.35) for the T minor allele. The wide confidence intervals underline that the statistical analyses are underpowered in REFINE. When taking into consideration the large difference between the number of individuals used in the CARDIoGRAM consortium (N=81,804) compared to those within AGES and REFINE then it is likely that the failure to confirm the CARDIoGRAM consortium's result in the Icelandic population was due to a lack of statistical power.

The lead SNP's association with CAD was shown to reach genome-wide significance in another large case-control study (N>190,000) which was a collaboration between many consortia, including CARDIoGRAM and CHARGE mentioned previously (66). The study was a follow-up to a previous meta-analysis of 14 CARDIoGRAM GWASs that had included 86,995 individuals (22,223 cases versus 64,762 controls), all of European ancestry (67). The aim was therefore to increase the statistical power by increasing the number of individuals included in the analysis. Variants that had shown suggestive association with CAD ($p<0.01$) in the 'discovery' meta-analysis were included in the new study in addition to fine-mapping SNPs and others provided by other consortia taking part in the meta-analysis. In addition to the discovery samples from the first meta-analysis, addition samples from other consortia were added that were of European and South Asian ancestry. Variants shown to be significant or suggestive in the combined analysis of stage 1 and 2 were then included in the third stage which was replication in an independent group of four different studies totalling of 3,630 cases

and 11,983 controls. The lead SNP in *EDNRA* did not reach statistical significance in the first meta-analysis study of 80 thousand individuals, however, in the follow-up study, through the combination of samples from all three analysis stages, the lead SNP reached genome-wide significance. The OR for all the combined analysis of all three stages was not given, however, in the combination of stage 1 and 2 the OR for the lead SNP is 1.06, with no confidence intervals given. This OR is in line with that seen in the CARDIoGRAM data which was 1.10. There is, therefore, corroborating evidence from a larger study for the result seen in the CHARGE paper, that there is a significant association between the lead *EDNRA* SNP and CAD and that the T allele of the SNP gives an increased odds of CAD.

In both the AGES Reykjavik study and the REFINE Reykjavik study coronary heart disease (CHD) is used instead of CAD. CHD reflects an individual who has suffered an MI or has had surgery (CABG, PCI) to treat coronary arteries that have narrowed to the point of putting an individual at risk of MI. The CAD encompasses CHD, in that individuals who have had an MI or preventative surgery are included, but also includes individuals that have atherosclerosis in their coronary arteries. Therefore CAD is a measure of endpoints as well as disease state (atherosclerosis), and its definition is highly varied between different studies, which can be seen, for instance, in the CARDIoGRAM consortium (67). Therefore as atherosclerosis is a systemic disease and the lead SNP in *EDNRA* has been shown to be significantly associated with atherosclerosis in the carotid arteries (28), it is reasonable to hypothesize that the association with CAD is a reflection of the association with atherosclerosis in the coronary arteries rather than a definitive association with MI, which results from an occlusion of the coronary artery by a thrombus. Evidence for this can be found in the multi-consortia study mentioned previously (66) where all 46 SNPs that reached genome-wide significance for CAD were also tested for their association with MI. In that analysis the lead SNP in *EDNRA* did not reach genome-wide significance for MI ($p=3.96x10^{-03}$, OR=1.05). A possible avenue of further study could be to investigate whether the lead SNP is significantly associated with coronary plaque levels. Such measurements have not been carried out in either the AGES or REFINE Reykjavik Studies, however, it may be something to consider for future projects where intravascular ultrasound is available or by coronary angiography.

5.3 Refining the lead SNP signal

5.3.1 Carotid plaque associated SNPs

In this study it was shown that there was a total of 15 SNPs significantly associated with carotid plaque ($p \leq 0.01$), in addition to the lead SNP. One of these is genotyped and the rest are imputed in the AGES GWAS cohort. The majority of SNPs that are significantly associated with carotid plaque are in moderate to strong levels of LD with the lead SNP and their association with carotid plaque reflects that, with both OR and p-values in line with that seen for the lead SNP. The exceptions to this are two rare SNPs (MAF=0.0187), rs6854578 and rs10305929, that both have an OR of 2.25 (95% CI; 1.27-3.98) and a p-value equal to 0.00548. These two SNPs are both intronic, lie in the sixth intron of *EDNRA* and are completely inter-dependent in terms of LD ($r^2=1$, D'=1) and therefore, they both represent the same signal. Their r^2 value, denoting correlation adjusted for difference in SNP MAF, shows weak LD ($r^2=0.195$). Due to the SNPs being rare, the chance of them being in perfect LD with

the lead SNP due to random chance is high and the r^2 value reflects that. RegulomeDB indicates that both of the 2.25 OR SNPs show minimal evidence for having an effect on transcription factor (TF) binding.

Interestingly, the 2.25 OR SNPs are in strong LD with two other intron six SNPs that have a suggestive association with carotid plaque. The SNPs, rs78046355 and rs77249653, are rare (MAF=0.013) and represent the same signal and both have an OR of 2.11 (95% CI; 1.18-3.77) and a p-value equal to 0.0114. They are of particular interest as one of them, rs78047355, was annotated in RegulomeDB as being likely to affect TF binding. This is supported through multiples lines of evidence including data from ChIP-seq experiments conducted by the ENCODE consortium. The ChIP-seq data shows that the CTCF TF binds in the region of the rs78047355 SNP in multiple cell lines.

The rs10008744 SNP, genotyped in the AGES GWAS cohort, did not reach the statistical significance threshold (alpha=0.01) but did show a suggestive association with carotid plaque (p=0.0332) with a relatively modest OR of 1.15 (95% CI; 1.01-1.32). The SNP is in weak LD with the lead SNP, which could indicate that its significant association with carotid plaque is due to its LD with the lead SNP. However, it shows no association with common cIMT (p=0.795) and could be independent from the lead SNP signal. There is no information, seen through RegulomeDB, that this SNP is likely to have an effect on transcription factor binding or regulation in general.

5.3.2 Common cIMT associated SNPs

In this study there were 63 SNPs significantly associated with common cIMT ($p \leq 0.01$), in addition to the lead SNP. Of those 63 SNPs, 4 are genotyped and 59 are imputed in the AGES GWAS cohort. For the 53 SNPs with LD information available it can be seen that their levels of LD with the lead SNP vary to a degree. In general terms, the SNPs that are in strong LD with the lead SNP have the strongest association with common cIMT and the largest effect sizes. However, the effect sizes are small for these SNPs with the betas ranging from 0.0110 to 0.0200. Therefore for at least for 53 of the 63 SNPs significantly associated with common cIMT there is no clearly independent signal identified, with all SNPs at the very least showing weak levels of LD with the lead SNP.

One of the SNPs significantly associated with common cIMT, rs1400554, has the most evidence for an effect on TF binding of all the carotid phenotype significantly associated SNPs, according to its RegulomeDB score. RegulomeDB predicts that according to the available evidence that the SNP affects TF binding. The lines of evidence include the SNP overlapping a DNase I hypersensity site, as well as overlapping a TF binding sites according to ChIP-seq experiments conducted in three different cell types and it lies in close proximity to a H3K27Ac histone modification mark that often overlaps or is found in close proximity to active regulatory elements (68). The rs1400554 SNP is in strong LD with the lead SNP when D' is considered (D'=0.946), however the r^2 value suggest a weaker LD (r^2=0.299) likely because the rs1400554 SNP is more common than the lead SNP with MAF of 0.301 and 0.141 respectively.

5.3.3 SNPs associated with both carotid phenotypes

In addition to the lead SNP there are 13 SNPs that are significantly associated with both carotid phenotypes. All of the SNPs are in strong LD with the lead SNP with the r^2 ranging from 0.666 to 0.959 and the D' ranging from 0.891 to 0.99. Therefore these SNPs are most likely an indication of the same signal as the lead SNP, as is reflected in their OR which range from 1.28 to 1.33. Of these SNPs there are three with ORs higher than the lead SNP, which itself had an OR of 1.30. They are rs17612742, rs78049276 and rs6841473, and their ORs, in respective order, are 1.33, 1.33 and 1.32. All three SNPs lie further away from the lead SNP than the other SNPs that give a similar signal, with distances of 21kb, 34kb and 14kb respectively. The rs17612742 SNP is of particular interest as it shows a higher OR than the lead SNP, albeit only marginally (1.33 vs. 1.30), in addition to having a lower p-value (0.00128 vs. 0.00172) and has some evidence of a possible functional role. Using RegulomeDB it is possible to see that there is ChIP-seq evidence in multiple cell types for the SNP lying in close proximity to a H327Ac histone mark, which as mentioned previously can be an indicator for active regulatory elements, such as enhancers (68). As the SNP is in strong LD with the lead SNP, in addition to being conserved and lying within a transcription factor binding site through functional annotation, it would be a candidate for inclusion in any follow-up studies. Possible experiments could involve assessing whether the SNP affects the binding of the MEF-2A TF and whether the variant has an effect on the transcription of the *EDNRA* gene or possibly the level of protein expression.

The rs6842241 and rs6841581 SNPs are also of interest because they are proxies for the lead SNP. The rs6842241 SNP is genotyped in the AGES GWAS cohort and shows significant association both with plaque (p=0.00178) and common cIMT ($p=6.21x10^{-05}$). The rs6841581 SNP is also significantly associated with carotid plaque (p=0.00178) and cIMT ($p=6.45x10^{-05}$). Both SNPs are in strong LD with the lead SNP (r^2=0.935, D'=1) and perfect LD (r^2=1, D'=1) with each other. They are of particular interest because, along with the lead SNP, they were identified through a GWAS to be significantly associated with intracranial aneurysm in the Japanese population (69). The rs6842241 SNP was genotyped and reached genome-wide significance (p=9.58E-9) for intracranial aneurysm. The lead SNP (rs1878406) and rs6841581 were imputed in the Japanese population and showed similar association with intracranial aneurysm as seen with rs6842241. Within the same study functional analysis was performed and the rs6841581 and rs1878406 SNPs were tested. The allelic differences, in both the level of *EDNRA* transcription activity and the nuclear protein binding affinity, were examined in human embryonic kidney (HEK293) cells. The difference in transcriptional activity was measured using dual-luciferase reporter assays. The major alleles of both SNPs had a high nuclear binding affinity with the minor alleles showing little to no binding affinity. This translated to a difference in transcriptional activity but only for the rs6841581 SNP, where the major allele G was associated with a significantly lower level of *EDNRA* transcription compared to both the minor allele A and an empty vector. The vectors containing either the minor or major allele of rs1878406 did not show any statistically significant difference between themselves or the empty vector.

With regard to functional annotation, rs6841581 lies within a DNase I Hypersensity peak (62). DNase I hypersensitivity peaks are regions where the DNase I enzyme can cleave the double stranded DNA and as such it is an indication of genomic regions that lack nucleosomes. Such sites

are often sites of promoter, enhancers or other DNA regions which function in the regulation of genes. According to the mammalian promoter database the rs6841581 SNP lies within an *EDNRA* promoter region (70).

Typically in GWAS if there are several SNPs at the same locus that all reach the threshold for genome-wide significance then it is the SNP with the lowest p-value that is presented as the lead marker. As both the rs6842241 and rs6841581 SNPs were present in the CHARGE meta-analysis of GWASs then it follows that they had a higher p-value than the lead SNP. In the analysis of genotyped and imputed SNPs in this study the same result is found, that the lead SNP had a lower p-value than the other two SNPs, both for carotid plaque and common cIMT. If the rs6841581 SNP, shown to be functional, is truly the explanation of the signal then logically it should show a more statistically significant result than the lead SNP. This isn't always the case, however, as other factors can play a part such as statistical variation or that the functional SNP only explains a part of the signal.

The SNPs that are significantly associated with carotid plaque and common cIMT are generally representative of the LD structure seen in and upstream of *EDNRA*, Figure 8. The SNPs with the strongest signal are generally proxies, in terms of LD, for the lead SNP and the SNPs in strongest LD with the lead SNP lie in the same LD block as the lead SNP. For example when viewing the common cIMT regional association plot (Figure 13) for imputed variants and comparing the different signals seen there with the overall LD structure of *EDNRA* (Figure 8) then the third signal, discussed in the results chapter, can be seen to lie within the adjacent LD block. The weak LD between these two blocks could explain why the third signal is being picked up. Those SNPs lie with the variant on some haplotypes and their association with common cIMT is therefore a reflection of that rather than they themselves being associated with common cIMT. This is not an absolute statement, however, judging from the weak LD between those SNPs and the lead SNP and the LD structure of the *EDNRA* gene it is the most likely explanation. Although the LD block that the lead SNP lies within is a good indicator for which SNPs are inherited together, it is not to say that all SNPs within that block are in strong LD with the lead SNP. As can be seen in the regional LD plot for the lead SNP (Figure 9), the boundary for SNPs in strong LD (r^2>0.8) with the lead SNP covers a similar area to that seen for the lead SNP's LD block (Figure 8), however, there are a great number of SNPs that lie within that strong LD region that are not in strong LD with the lead SNP. Therefore LD is more nuanced than LD blocks indicate and this may explain a possible second signal seen for common cIMT in the imputed SNPs, where SNPs within the same LD block as the lead SNP are in varying degrees of LD with the lead SNP and by extension their effect sizes for common cIMT reflect this difference in LD.

5.4 Haplotype analysis

The haplotype analysis found that haplotype 10 (AGACCACAGCG) was suggestively associated with carotid plaque. Through correlation analysis it was found that the haplotype correlated weakly with the lead SNP signal but strongly with the 2.25 OR SNPs, rs6854578 and rs10305929. The haplotype therefore appears to be capturing much of the signal from these rare 2.25 OR SNPs. In addition, the lead SNP's proxy SNP, rs6842241, which has a similar OR to the lead SNP and is in strong LD with it, has been shown with this haplotype phasing and analysis to lie on the same

haplotype as two SNPs that have a stronger association with carotid plaque and give an OR of 2.25. The sliding-window analysis made it apparent that the rs6842241 risk allele A is broken up into three different haplotypes in the first sliding window of five SNPs. Only one of these haplotypes is significantly associated with carotid plaque and gave an OR of 1.29 (95% CI; 1.09-1.53), a p-value of 0.00275 and had a frequency of 0.125. The frequency of this haplotype capturing the rs6842241 risk allele signal is lower than the risk allele itself, which was 0.133. The other two haplotypes (AAGCC and AAGTT) were extremely rare, with frequencies of 0.0073 and 0.000777, and had p-values of 0.594 and 0.956 respectively. The extremely rare AAGTT haplotype could be an artifact of the phasing, however, these two haplotypes show that the risk allele is not always associated with carotid plaque and therefore suggest that the lead SNP signal could be broken up and may be represented by a functional variant that is less frequent than the lead SNP or its proxy, rs6842241.

Although this is interesting, the low frequency of the haplotype (0.0089) along with the wide 95% CI for the OR (95% CI; 1.01-3.76) and p-value of 0.0453 could mean that this signal is a false positive. This finding warrants further investigation although a cautious approach to follow-up studies would be recommended.

5.5 Strengths and weaknesses

This study has certain strengths and weaknesses. One of the strengths is that the studies used here, the AGES and REFINE, are both large and population-based. The size of the studies allows for weaker effect sizes to be detected and also for the analysis of dichotomous variables where there is a loss of power. Also, population-based studies do not have the same biases as are seen in case-control studies. By having a random sampling of the population it is possible to avoid selection bias. For instance, if a case-control study involves individuals that were treated at a hospital then the study could be biased for individuals that sought out medical assistance, which in the context of this study would be individuals that had suffered an MI or a stroke. In addition, the extensive data available from both studies is a strength as more robust analyses can be performed where all the major risk factors for CVD are adjusted for.

The weaknesses of this study are few but important to note. FastPHASE, used in this study to create phased haplotypes from EDNRA SNPs has several improvements over the standard PHASE program. FastPHASE, as the name indicates, allowed for the same accuracy of haplotype phasing as other methods but in a much shorter time (59). This made the phasing of large SNP datasets, such as from GWAS, possible. However, when phasing a SNP dataset that contains a large number of individuals, as was done in this study, then it has been noted that the reduction in computational time comes as the cost of a reduced accuracy in the phased haplotypes (58). Therefore, the haplotype phasing may not have been as accurate as it would have been if other methods had been used. Another weakness of this study may be the significance threshold that was used in the single-point SNP regression analyses. The alpha value of 0.01 was determined to be sufficiently strict to account for the increased likelihood of false-positives that accompanies multiple testing, however, some may find this threshold lenient. It was deemed that Bonferroni correction, often used in data driven studies such as GWAS, was too strict for this study as it is not data driven but rather hypothesis driven.

6 Conclusions and future directions

This study found that the lead SNP from the CHARGE meta-analysis (rs1878406) is in strong LD with a SNP (rs6841581) that shares the same signal for both carotid plaque and common cIMT and has been shown to be functional through an allelic difference in *EDNRA* transcription (69). The rs6841581 SNP is therefore a prime candidate for a functional follow-up study where transcriptional activity of the *EDNRA* protein can be assessed between individuals with different genotypes and also case and controls for atherosclerosis traits such as carotid plaque.

There are also several other SNPs that could be investigated further. For example, the rs78047355 SNP, which was suggestively associated with carotid plaque and is in strong LD with the 2.25 OR SNPs, could be investigated to see whether it has any effect on *EDNRA* transcription in particular cell types, such as smooth muscle cells. Another target could be the rs1400554 SNP. Its significant association with common cIMT and non-association with carotid plaque could be of interest for follow-up as a greater understanding of the function of this SNP may give some insight into how the atherosclerotic phenotypes differ. Also, its predicted effect on TF binding by RegulomeDB warrants further study.

Further analysis is required of the 63 SNPs that have been shown to be significantly associated with common cIMT. In this study their association with common cIMT is described and basic analysis of their relationship with the lead SNP is detailed, however, the future direction would be a more in-depth analysis of the LD relationship between the lead SNP and the other common cIMT significantly associated SNPs. In addition, the SNPs could be studied to determine how they relate to each other in terms of LD and whether there is an additional signal among them, independent of the lead SNP. The 1000 Genomes Pilot 1 dataset that is used by SNAP to assess LD is lacking information for certain SNPs. Therefore with a more up-to-date dataset, that has information on all of the significant SNPs, it would be possible to accomplish a more thorough analysis of the LD relationship between the lead SNP and other common cIMT associated SNPs. Another possibility is to perform a conditional analysis with the lead SNP as a covariate.

If any of these potentially functional SNPs are shown through further functional analysis to be truly functional and have an effect on the transcription of *EDRNA* in, for example, smooth muscle cells, then the next possible step could be to genotype the SNPs in the REFINE cohort as that would be an AGES independent attempt to replicate their effect in the Icelandic population.

Here, in this study, the focus of the study was the *EDRNA* gene itself, however, as can be seen in Figure 9, the range of SNPs in strong LD ($r^2 \geq 0.8$) with the lead SNP extends 28kb upstream of the lead SNP itself. In future studies this extensive region upstream of *EDNRA* could be investigated with regard to additional signal. There are, for example, H3K27Ac histone marks upstream of the lead SNP and outside the region examined here (71). In addition, the haplotype analysis can be performed again adjusting for more risk factors as was done with the lead SNP in this study. It would also be interesting to perform the association testing between the haplotypes and common cIMT. As the large majority of carotid plaque associated SNPs are also associated with common cIMT it could be worthwhile to see if there are any haplotypes that show an exclusive association with common cIMT. There are a large

number of SNPs that associate only with cIMT and not carotid plaque and as common cIMT has been suggested to be a reflection of the effect of blood pressure on the arterial wall it would be interesting to see whether adjusting for blood pressure and additional risk factors has any effect of the strength of the common cIMT signal for these SNPs.

Eventually, a selection may occur of individuals with haplotypes that may better capture the functional signal behind the lead SNP's association with carotid plaque and common cIMT. These selected individuals could then be sequenced with the aim of attempting to determine whether the functional variants discussed here are the cause of the lead SNP's signal or there are other, unknown, variants present.

Bibliography

1. WHO. The 10 leading causes of death in the world, 2000 and 2011. WHO; 2013. Cited: 29 December 2013. Accessed from:http://www.who.int/mediacentre/factsheets/fs310/en/index.html
2. Iceland S. Health at a glance. 2009. Cited: 29 December 2013. Accessed from:http://www.hagstofa.is/Pages/95?NewsID=5070.
3. Ahmed N, Dawson M, Smith C, Wood E. Biology of disease: Taylor & Francis; 2007.
4. Holman RL, Mc GH, Jr., Strong JP, Geer JC. The natural history of atherosclerosis: the early aortic lesions as seen in New Orleans in the middle of the of the 20th century. Am J Pathol. 1958 Mar-Apr;34(2):209-35.
5. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. Nature. 2011 May 19;473(7347):317-25.
6. Autieri MV. Pro- and Anti-Inflammatory Cytokine Networks in Atherosclerosis. ISRN Vascular Medicine. 2012;2012.
7. Aspelund T, Thorgeirsson G, Sigurdsson G, Gudnason V. Estimation of 10-year risk of fatal cardiovascular disease and coronary heart disease in Iceland with results comparable with those of the Systematic Coronary Risk Evaluation project. Eur J Cardiovasc Prev Rehabil. 2007 Dec;14(6):761-8.
8. Kaptoge S, Di Angelantonio E, Pennells L, Wood AM, White IR, Gao P, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. New Engl J Med. 2012 Oct 4;367(14):1310-20.
9. Faxon DP, Fuster V, Libby P, Beckman JA, Hiatt WR, Thompson RW, et al. Atherosclerotic Vascular Disease Conference: Writing Group III: pathophysiology. Circulation. 2004 Jun 1;109(21):2617-25.
10. Cunningham KS, Gotlieb AI. The role of shear stress in the pathogenesis of atherosclerosis. Lab Invest. 2005 Jan;85(1):9-23.
11. Bijari PB, Wasserman BA, Steinman DA. Carotid Bifurcation Geometry Is an Independent Predictor of Early Wall Thickening at the Carotid Bulb. Stroke. 2013 Dec 19.
12. Sverrisdottir A, Sigurdsson G, Sigfusson N, Sveinbjornsdottir S, Agnarsson U. Heilablóðfall, Háþrýstingur hvað er til ráða. Bæklingur Hjartaverndar. 2005. Accessible from: http://www.hjarta.is/.
13. Johnsen SH, Mathiesen EB, Joakimsen O, Stensland E, Wilsgaard T, Lochen ML, et al. Carotid atherosclerosis is a stronger predictor of myocardial infarction in women than in men: a 6-year follow-up study of 6226 persons: the Tromso Study. Stroke. 2007 Nov;38(11):2873-80.
14. Inaba Y, Chen JA, Bergmann SR. Carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: a meta-analysis. Atherosclerosis. 2012 Jan;220(1):128-33.
15. Spence JD. Carotid plaque measurement is superior to IMT Invited editorial comment on: carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: a meta-analysis-Yoichi Inaba, M.D., Jennifer A. Chen M.D., Steven R. Bergmann M.D., Ph.D. Atherosclerosis. 2012 Jan;220(1):34-5.
16. Taylor AJ, Bindeman J, Le TP, Bauer K, Byrd C, Feuerstein IM, et al. Progression of calcified coronary atherosclerosis: relationship to coronary risk factors and carotid intima-media thickness. Atherosclerosis. 2008 Mar;197(1):339-45.
17. Ibrahimi P, Jashari F, Nicoll R, Bajraktari G, Wester P, Henein MY. Coronary and carotid atherosclerosis: How useful is the imaging? Atherosclerosis. 2013 Dec;231(2):323-33.
18. Criqui MH, Denenberg JO, Ix JH, McClelland RL, Wassel CL, Rifkin DE, et al. Calcium Density of Coronary Artery Plaque and Risk of Incident Cardiovascular Events. JAMA. 2013 Nov 18.
19. Iwasaki K, Matsumoto T, Aono H, Furukawa H, Samukawa M. Prevalence of subclinical atherosclerosis in asymptomatic patients with low-to-intermediate risk by 64-slice computed tomography. Coron Artery Dis. 2011 Jan;22(1):18-25.
20. Tarnoki AD, Baracchini C, Tarnoki DL, Lucatelli P, Boatta E, Zini C, et al. Evidence for a strong genetic influence on carotid plaque characteristics: an international twin study. Stroke. 2012 Dec;43(12):3168-72.
21. Sayed-Tabatabaei FA, van Rijn MJ, Schut AF, Aulchenko YS, Croes EA, Zillikens MC, et al. Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. Stroke. 2005 Nov;36(11):2351-6.

22. Hunt KJ, Duggirala R, Goring HH, Williams JT, Almasy L, Blangero J, et al. Genetic basis of variation in carotid artery plaque in the San Antonio Family Heart Study. Stroke. 2002 Dec;33(12):2775-80.
23. Swan L, Birnie DH, Inglis G, Connell JM, Hillis WS. The determination of carotid intima medial thickness in adults--a population-based twin study. Atherosclerosis. 2003 Jan;166(1):137-41.
24. Jartti L, Ronnemaa T, Kaprio J, Jarvisalo MJ, Toikka JO, Marniemi J, et al. Population-based twin study of the effects of migration from Finland to Sweden on endothelial function and intima-media thickness. Arterioscler Thromb Vasc Biol. 2002 May 1;22(5):832-7.
25. Zhao J, Cheema FA, Bremner JD, Goldberg J, Su S, Snieder H, et al. Heritability of carotid intima-media thickness: a twin study. Atherosclerosis. 2008 Apr;197(2):814-20.
26. Manolio TA, Boerwinkle E, O'Donnell CJ, Wilson AF. Genetics of ultrasonographic carotid atherosclerosis. Arterioscler Thromb Vasc Biol. 2004 Sep;24(9):1567-77.
27. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet. 2009 Feb;2(1):73-80.
28. Bis JC, Kavousi M, Franceschini N, Isaacs A, Abecasis GR, Schminke U, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. Nat Genet. 2011 Oct;43(10):940-7.
29. Liu L, Zhang D, Liu H, Arendt C. Robust methods for population stratification in genome wide association studies. BMC Bioinformatics. 2013;14:132.
30. Hosoda K, Nakao K, Tamura N, Arai H, Ogawa Y, Suga S, et al. Organization, structure, chromosomal assignment, and expression of the gene encoding the human endothelin-A receptor. J Biol Chem. 1992 Sep 15;267(26):18797-804.
31. Yang H, Tabuchi H, Furuichi Y, Miyamoto C. Molecular characterization of the 5'-flanking region of human genomic ETA gene. Biochem Biophys Res Commun. 1993 Jan 29;190(2):332-9.
32. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome biology. 2006;7 Suppl 1:S12 1-4.
33. Hayzer DJ, Rose PM, Lynch JS, Webb ML, Kienzle BK, Liu EC, et al. Cloning and expression of a human endothelin receptor: subtype A. Am J Med Sci. 1992 Oct;304(4):231-8.
34. Kedzierski RM, Yanagisawa M. Endothelin system: the double-edged sword in health and disease. Annu Rev Pharmacol Toxicol. 2001;41:851-76.
35. Takahashi M. The role of endothelin-1 in vascular remodeling in vivo. Cardiovasc Res. 2006 Jul 1;71(1):4-5.
36. Hafizi S, Allen SP, Goodwin AT, Chester AH, Yacoub MH. Endothelin-1 stimulates proliferation of human coronary smooth muscle cells via the ET(A) receptor and is co-mitogenic with growth factors. Atherosclerosis. 1999 Oct;146(2):351-9.
37. Ross R. Atherosclerosis - An Inflammatory Disease. New Engl J Med. 1999;340(2):115-26.
38. Miao J, Wang F, Fang Y. Association of 231G>A polymorphism of endothelin type A receptor gene with migraine: a meta-analysis. J Neurol Sci. 2012 Dec 15;323(1-2):232-5.
39. Tzourio C, El Amrani M, Poirier O, Nicaud V, Bousser MG, Alperovitch A. Association between migraine and endothelin type A receptor (ETA -231 A/G) gene polymorphism. Neurology. 2001 May 22;56(10):1273-7.
40. Calabro P, Limongelli G, Maddaloni V, Vizza CD, D'Alto M, D'Alessandro R, et al. Analysis of endothelin-1 and endothelin-1 receptor A gene polymorphisms in patients with pulmonary arterial hypertension. Intern Emerg Med. 2012 Oct;7(5):425-30.
41. Rahman T, Baker M, Hall DH, Avery PJ, Keavney B. Common genetic variation in the type A endothelin-1 receptor is associated with ambulatory blood pressure: a family study. J Hum Hypertens. 2008;22(4):282-8.
42. Hasegawa K, Fujiwara H, Doyama K, Inada T, Ohtani S, Fujiwara T, et al. Endothelin-1-selective receptor in the arterial intima of patients with hypertension. Hypertension. 1994;23(3):288-93.
43. Colombo MG, Ciofini E, Paradossi U, Bevilacqua S, Biagini A. ET-1 Lys198Asn and ET(A) receptor H323H polymorphisms in heart failure. A case-control study. Cardiology. 2006;105(4):246-52.
44. Pönicke K, Vogelsang M, Heinroth M, Becker K, Zolk O, Böhm M, et al. Endothelin receptors in the failing and nonfailing human heart. Circulation. 1998;97(8):744-51.
45. Yasuda H, Kamide K, Takiuchi S, Matayoshi T, Hanada H, Kada A, et al. Association of single nucleotide polymorphisms in endothelin family genes with the progression of atherosclerosis in patients with essential hypertension. J Hum Hypertens. 2007 Nov;21(11):883-92.

46. Reiterova J, Merta M, Stekrova J, Maixnerova D, Obeidova H, Kebrdlova V, et al. The influence of endothelin-A receptor gene polymorphism on the progression of autosomal dominant polycystic kidney disease and IgA nephropathy. Folia Biol (Praha). 2007;53(4):134-7.
47. Chapman AB, Rubinstein D, Hughes R, Stears JC, Earnest MP, Johnson AM, et al. Intracranial aneurysms in autosomal dominant polycystic kidney disease. New Engl J Med. 1992 Sep 24;327(13):916-20.
48. Yasuno K, Bakircioglu M, Low SK, Bilguvar K, Gaal E, Ruigrok YM, et al. Common variant near the endothelin receptor type A (EDNRA) gene is associated with intracranial aneurysm risk. P Natl Acad Sci USA. 2011 Dec 6;108(49):19707-12.
49. Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G, et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. Am J Epidemiol. 2007 May 1;165(9):1076-87.
50. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988 Feb 11;16(3):1215.
51. Barnes WM. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. P Natl Acad Sci USA. 1994 Mar 15;91(6):2216-20.
52. R Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
53. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75.
54. Laird N, Lange, C. The Fundamentals of Modern Statistical Genetics; 2011. ISBN 978-1-4419-7338-2
55. Anon. MaCH: 1000 Genomes Imputation Cookbook. 2011. Cited: 29 December 2013. Accessed from: http://genome.sph.umich.edu/wiki/MaCH:_1000_Genomes_Imputation_Cookbook
56. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002 Jun 21;296(5576):2225-9.
57. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics (Oxford, England). 2008 Dec 15;24(24):2938-9.
58. Browning SR, Browning BL. Haplotype phasing: existing methods and new developments. Nat Rev Genet. 2011 Oct;12(10):703-14.
59. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 2006 Apr;78(4):629-44.
60. Fan YH, Song YQ. IPGWAS: an integrated pipeline for rational quality control and association analysis of genome-wide genetic studies. Biochem Biophys Res Commun. 2012 Jun 8;422(3):363-8.
61. Zhao H, Pfeiffer R, Gail MH. Haplotype analysis in population genetics and association studies. Pharmacogenomics. 2003 Mar;4(2):171-8.
62. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome research. 2012 Sep;22(9):1790-7.
63. RegulomeDB. RegulomeDB score. 2011. Cited: 29 December 2013. Accessed from: http://regulome.stanford.edu/help#score.
64. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics (Oxford, England). 2009 Mar 1;25(5):655-61.
65. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res. 2012 Jul;40(Web Server issue):W65-70.
66. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013 Jan;45(1):25-33.
67. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011 Apr;43(4):333-8.
68. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. P Natl Acad Sci USA. 2010 Dec 14;107(50):21931-6.

69. Low SK, Takahashi A, Cha PC, Zembutsu H, Kamatani N, Kubo M, et al. Genome-wide association study for intracranial aneurysm in the Japanese population identifies three candidate susceptible loci and a functional genetic variant at EDNRA. Hum Mol Genet. 2012 May 1;21(9):2102-10.
70. Institute TW. The mammalian promoter database. 2009. Cited: 29 December 2013. Accessed from: http://mpromdb.wistar.upenn.edu/.
71. Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, et al. The UCSC Genome Browser database: 2014 update. Nucleic Acids Res. 2013 Nov 21.

7 Appendix

7.1 Basic characteristics of the AGES and REFINE cohorts split by gender.

Table 15. Basic characteristics of AGES Reykjavik Study GWAS men (N=1,352).

Characteristic	Mean, median or percentage	Range	N (Events)	N (Total)	N (Missing data)
Age (Years)	76.5 ± 5.3	67 - 94		1,352	0
BMI (kg/m²)	27 ± 3.8	15.9 - 40.6		1,351	1
Systolic BP (mmHg)	143.1 ± 19.9	97 - 253		1,352	0
Diastolic BP (mmHg)	76.3 ± 9.4	48 - 112		1,352	0
Pulse Pressure (mmHg)	66.8 ± 17.1	17 - 168		1,352	0
Common cIMT (mm)	1.17 ± 0.16	0.75 - 1.93		1,300	52
Cholesterol (mmol/L)	5.2 ± 1.1	1.8 - 9.1		1,352	0
HDL (mmol/L)	1.4 ± 0.39	0.35 - 3.83		1,352	0
LDL (mmol/L)	3.25 ± 0.99	0.43 - 6.59		1,351	1
Triglycerides (mmol/L)	1.01 (0.75-1.42)	0.3 - 12.99		1,352	0
Fasting Glucose (mmol/L)	5.91 ± 1.22	3.5 - 19		1,352	0
Coronary Calcium (Agaston)	616 (161-1455)	0 - 8673		1,333	19
Hba1c (%)	5.7 ± 0.56	4 - 12		1,254	98
Type 2 Diabetes (%)	15.1		204	1,349	3
Statins (%)	29.4		397	1,352	0
Coronary Event (%)	25.4		340	1,341	1,878
Never Smoked (%)	28.6		387	1,352	0
Former Smoker (%)	59.4		803		
Current Smoker (%)	12		162		
Carotid plaque (%)	68.7		888	1,292	60
Hypertension Medication (%)	62.7		847	1,352	0

The mean is displayed with the standard deviation and the median with the inter-quartile range. LDL levels estimated using the Friedewald formula. BP stands for Blood Pressure.

Table 16. Basic characteristics of AGES Reykjavik Study GWAS women (N=1,867).

Characteristic	Mean, median or percentage	Range	N (Events)	N (Total)	N (Missing data)
Age (Years)	76.3 ± 5.6	66 - 95		1,867	0
BMI (kg/m²)	27.2 ± 4.9	14.8 - 48.5		1,865	2
Systolic BP (mmHg)	142.2 ± 20.5	92 - 250		1,866	1
Diastolic BP (mmHg)	72.4 ± 9.5	29 - 145		1,866	1
Pulse Pressure (mmHg)	69.7 ± 18.8	17 - 167		1,866	1
Common cIMT (mm)	1.11 ± 0.15	0.7 - 1.77		1,773	94
Cholesterol (mmol/L)	6 ± 1.1	2.3 - 10.5		1,867	0
HDL (mmol/L)	1.72 ± 0.44	0.59 - 4.05		1,867	0
LDL (mmol/L)	3.7 ± 1.03	0.9 - 8.53		1,866	1
Triglycerides (mmol/L)	1.07 (0.81-1.48)	0.32 - 7.48		1,867	0
Fasting Glucose (mmol/L)	5.64 ± 1.02	4.1 - 15.2		1,867	0
Coronary Calcium (Agaston)	141 (13-542)	0 - 6635		1,843	24
Hba1c (%)	5.68 ± 0.46	4.3 - 9.3		1,708	159
Type 2 Diabetes (%)	8.8		165	1,866	1
Statins (%)	17.8		332	1,867	0
Coronary Event (%)	7.4		137	1,847	20
Never Smoked (%)	51.7		966	1,867	0
Former Smoker (%)	35		654		
Current Smoker (%)	13.2		247		
Carotid plaque (%)	65.6		1159	1,766	101
Hypertension Medication (%)	64.7		1207	1,867	0

The mean is displayed with the standard deviation and the median with the inter-quartile range. LDL levels estimated using the Friedewald formula. BP stands for Blood Pressure.

Table 17. Basic characteristics of REFINE Reykjavik Study men genotyped for rs1878406 (N=2,728).

Characteristic	Mean, median or percentage	Range	N (Events)	N (Total)	N (Missing Data)
Age (Years)	49.6 ± 12.0	21 - 73		2,728	0
BMI (kg/m²)	28.0 ± 4.4	15.0 - 58.6		2,727	1
Systolic BP (mmHg)	126.8 ± 15.9	88 - 226		2,727	1
Diastolic BP (mmHg)	72.9 ± 10.8	40 - 148		2,727	1
Pulse Pressure (mmHg)	53.9 ± 13.7	20 - 148		2,727	1
Common cIMT (mm)	0.88 ± 0.17	0.49 - 1.76		2,728	0
Cholesterol (mmol/L)	5.2 ± 1.0	2.4 - 12.7		2,724	4
HDL (mmol/L)	1.32 ± 0.33	0.54 - 3.98		2,724	4
LDL (mmol/L)	3.26 ± 0.92	0.39 - 10.69		2,695	33
Triglycerides (mmol/L)	1.13 (0.81/1.65)	0.24 - 11.64		2,724	4
Fasting Glucose (mmol/L)	5.69 ± 1.10	3.03 - 22.61		2,724	4
Type 2 Diabetes (%)	6.2		168	2,728	0
Statins (%)	12.3		335	2,723	5
History of Family MI (%)	33.6		894	2,662	66
CEQ (%)	5.8		156	2,713	15
Smoking Status - Never (%)	38.5		1,044	2,712	16
Smoking Status - Former (%)	39.8		1,080		
Smoking Status - Current (%)	21.7		588		
Plaque (%)	13.2		359	2,728	0

The mean is displayed with the standard deviation and the median with the inter-quartile range. LDL levels estimated using the Friedewald formula. CEQ stands for Coronary Event by Questionnaire. BP stands for Blood Pressure.

Table 18. Basic characteristics of REFINE Reykjavik Study women genotyped for rs1878406 (N=2,793).

Characteristic	Mean, median or percentage	Range	N (Events)	N (Total)	N (Missing Data)
Age (Years)	49.3 ± 12.2	20 - 73		2,793	0
BMI (kg/m²)	26.9 ± 5.3	15.6 - 65.8		2,791	2
Systolic Blood Pressure (mmHg)	117.9 ± 16.5	77 - 200		2,793	0
Diastolic Blood Pressure (mmHg)	70.2 ± 9.6	36 - 110		2,793	0
Pulse Pressure (mmHg)	47.7 ± 12.6	7 - 112		2,793	0
Common cIMT (mm)	0.83 ± 0.15	0.37 - 1.64		2,793	0
Cholesterol (mmol/L)	5.2 ± 1.0	2.3 - 11.2		2,785	8
HDL (mmol/L)	1.65 ± 0.42	0.31 - 3.58		2,785	8
LDL (mmol/L)	3.1 ± 0.94	0.13 - 9.22		2,773	20
Triglycerides (mmol/L)	0.90 (0.66 - 1.27)	0.16 - 10.19		2,785	8
Fasting Glucose (mmol/L)	5.28 ± 0.83	1.96 - 18.58		2,785	8
Type 2 Diabetes (%)	2.86		80	2,793	0
Statins (%)	5.91		165	2,793	0
History of Family MI (%)	38.65		1,052	2,722	71
CEQ (%)	1.94		54	2,777	16
Smoking Status - Never (%)	43.0		1,194	2,777	16
Smoking Status - Former (%)	35.7		992		
Smoking Status - Current (%)	21.3		591		
Plaque (%)	9.0		250	2,793	0

The mean is displayed with the standard deviation and the median with the inter-quartile range. LDL levels estimated using the Friedewald formula. CEQ stands for Coronary Event by Questionnaire. BP stands for Blood Pressure.

7.2 Single-point regression results tables for 1000 Genome imputed SNPs – full tables

7.2.1 Carotid plaque logistic regression results

Table 19. Single-point logistic regression results for carotid plaque, adjusted for age and gender. Regression was performed on the minor allele of each SNP and results are ordered by SNP chromosome position. Variants lacking an rs number are labelled with their genomic location. Imputation quality given as the r^2.

SNP	Minor allele	OR (95% CI)	p	MAF	r^2
rs17612742	C	1.33 (1.12-1.58)	0.00128	0.129	0.931
rs78049276	C	1.33 (1.12-1.59)	0.00136	0.125	0.922
rs6841473	T	1.32 (1.11-1.56)	0.00142	0.130	0.949
rs7659823	G	1.28 (1.1-1.49)	0.00163	0.188	0.875
rs6820938	G	1.28 (1.1-1.49)	0.00171	0.188	0.868
rs1878406	T	1.3 (1.1-1.53)	0.00172	0.141	0.971
rs6855875	T	1.28 (1.1-1.5)	0.00174	0.185	0.859
rs73855814	T	1.3 (1.1-1.53)	0.00175	0.133	0.985
rs10305839	T	1.3 (1.1-1.53)	0.00178	0.132	0.991
rs6841581	A	1.3 (1.1-1.53)	0.00178	0.133	0.992
rs72957606	G	1.3 (1.1-1.53)	0.00179	0.133	0.992
rs10305838	C	1.3 (1.1-1.53)	0.00179	0.133	0.992
rs6823438	T	1.3 (1.1-1.53)	0.00187	0.142	0.952
rs6854578	C	2.25 (1.27-3.98)	0.00548	0.0187	0.681
rs10305929	A	2.25 (1.27-3.98)	0.00548	0.0187	0.680
rs78047355	G	2.11 (1.18-3.77)	0.0114	0.0127	0.952
rs77249653	T	2.11 (1.18-3.77)	0.0114	0.0130	0.931
rs1808038	C	1.18 (1.03-1.34)	0.0154	0.306	0.825
rs6853791	T	1.17 (1.03-1.34)	0.0157	0.301	0.847
rs13144507	C	1.17 (1.03-1.33)	0.0180	0.286	0.904
rs13121745	T	1.16 (0.999-1.36)	0.0514	0.185	0.840
rs139452637	T	1.21 (0.996-1.47)	0.0554	0.0999	0.898
rs13143677	A	1.13 (0.993-1.29)	0.0629	0.282	0.913
rs75736224	G	1.2 (0.989-1.46)	0.0646	0.0999	0.909
rs4835410	G	1.13 (0.992-1.29)	0.0656	0.288	0.879
rs9308218	T	1.13 (0.991-1.29)	0.0668	0.288	0.870
rs702757	A	1.13 (0.991-1.29)	0.0670	0.287	0.874
rs952402	C	1.13 (0.992-1.29)	0.0670	0.287	0.882
rs1996373	G	1.13 (0.991-1.29)	0.0675	0.286	0.887
rs7663891	C	1.13 (0.992-1.29)	0.0676	0.287	0.880
rs10305862	T	1.13 (0.991-1.29)	0.0677	0.287	0.870
rs6823537	G	1.13 (0.991-1.29)	0.0678	0.287	0.874
rs1517134	T	1.13 (0.991-1.29)	0.0679	0.285	0.891
rs4835084	A	1.13 (0.991-1.29)	0.0681	0.288	0.867
rs1996374	A	1.13 (0.991-1.29)	0.0686	0.285	0.887
rs984457	T	1.13 (0.991-1.29)	0.0688	0.287	0.867

rs56015601	T	1.13 (0.991-1.29)	0.0688	0.287	0.868
rs9991584	G	1.13 (0.991-1.29)	0.0689	0.291	0.860
rs10305860	A	1.13 (0.991-1.29)	0.0689	0.287	0.868
rs6537480	A	1.13 (0.99-1.29)	0.0694	0.286	0.847
rs984458	T	1.13 (0.991-1.29)	0.0697	0.287	0.866
rs150948757	T	0.558 (0.297-1.05)	0.0697	0.0103	0.639
rs1801708	A	1.13 (0.991-1.29)	0.0699	0.288	0.861
rs35993155	A	0.9 (0.803-1.01)	0.0718	0.330	0.998
rs62341322	A	0.9 (0.803-1.01)	0.0734	0.330	0.998
rs192252324	G	1.87 (0.936-3.75)	0.0755	0.0101	0.735
rs4835411	C	1.19 (0.98-1.44)	0.0800	0.101	0.919
rs4835083	G	1.12 (0.986-1.28)	0.0809	0.290	0.861
rs11721708	G	0.904 (0.806-1.01)	0.0840	0.331	0.999
rs6537481	G	1.13 (0.982-1.3)	0.0886	0.231	0.872
rs10305919	A	1.12 (0.982-1.28)	0.0901	0.232	0.959
rs10305918	C	1.12 (0.982-1.28)	0.0905	0.232	0.959
rs5333	C	1.12 (0.981-1.27)	0.0943	0.245	0.999
rs6840375	A	1.12 (0.981-1.27)	0.0945	0.245	0.999
rs6841799	G	1.12 (0.981-1.27)	0.0947	0.245	0.999
rs2884340	G	1.12 (0.981-1.27)	0.0950	0.245	0.998
rs2884339	A	1.12 (0.981-1.27)	0.0952	0.245	0.998
rs2292765	A	1.12 (0.981-1.27)	0.0961	0.247	0.983
rs2884337	C	1.12 (0.981-1.27)	0.0963	0.245	0.998
rs7677315	G	1.12 (0.981-1.27)	0.0968	0.245	0.997
rs10305921	C	1.12 (0.981-1.27)	0.0973	0.245	0.997
rs10305920	G	1.11 (0.98-1.27)	0.0976	0.245	0.997
rs7690347	A	1.12 (0.98-1.27)	0.0977	0.246	0.983
rs3756022	C	1.11 (0.98-1.27)	0.0997	0.245	0.996
rs1400556	A	1.11 (0.98-1.27)	0.100	0.245	0.994
rs9307838	C	1.11 (0.98-1.27)	0.100	0.245	0.994
rs10305912	A	1.11 (0.98-1.27)	0.101	0.245	0.996
rs9307837	T	1.11 (0.979-1.27)	0.102	0.245	0.993
rs143614835	T	1.89 (0.876-4.07)	0.104	0.00919	0.770
rs11936340	T	1.11 (0.978-1.27)	0.105	0.243	0.977
rs10011966	A	1.1 (0.979-1.25)	0.107	0.301	1.000
rs6537482	A	1.1 (0.979-1.25)	0.107	0.301	1.000
rs7655670	C	1.1 (0.978-1.24)	0.109	0.300	0.998
rs6819740	T	1.1 (0.978-1.25)	0.109	0.300	0.998
rs6840756	C	1.1 (0.978-1.24)	0.110	0.300	0.998
rs17474816	A	1.1 (0.978-1.24)	0.111	0.300	0.997
rs6848108	G	1.1 (0.978-1.24)	0.111	0.300	0.996
rs2048894	A	1.11 (0.976-1.26)	0.111	0.248	0.992
rs4835086	A	1.1 (0.977-1.24)	0.112	0.300	0.997
rs10029799	C	1.11 (0.976-1.26)	0.112	0.248	0.992
rs143184430	A	0.854 (0.702-1.04)	0.113	0.105	0.788

rs1400554	T	1.1 (0.977-1.25)	0.113	0.301	0.981
rs4835412	A	1.17 (0.964-1.41)	0.113	0.102	0.944
rs908581	T	1.17 (0.964-1.41)	0.114	0.102	0.944
rs17474823	G	1.1 (0.976-1.24)	0.115	0.299	0.993
rs6843446	G	1.11 (0.974-1.26)	0.122	0.247	0.993
rs10003447	T	1.11 (0.974-1.26)	0.123	0.247	0.995
rs10028507	A	1.11 (0.973-1.26)	0.123	0.247	0.996
rs7673178	T	0.859 (0.707-1.04)	0.125	0.105	0.795
rs10305916	G	1.12 (0.97-1.28)	0.125	0.194	0.983
rs1568137	C	0.86 (0.704-1.05)	0.138	0.101	0.780
rs1987147	C	1.1 (0.969-1.25)	0.141	0.424	0.759
rs9884563	A	0.92 (0.822-1.03)	0.148	0.398	0.959
rs57886575	A	1.09 (0.969-1.23)	0.149	0.306	0.975
rs1808039	C	1.09 (0.964-1.24)	0.166	0.414	0.799
rs12509416	C	1.09 (0.962-1.22)	0.184	0.391	0.865
rs4835087	A	0.929 (0.83-1.04)	0.196	0.375	0.997
rs10032496	G	0.929 (0.831-1.04)	0.198	0.375	0.999
rs185348394	A	0.86 (0.674-1.1)	0.222	0.0802	0.652
rs6831880	G	0.933 (0.834-1.04)	0.222	0.375	0.999
rs9307839	T	0.933 (0.834-1.04)	0.224	0.375	0.998
rs5342	G	0.934 (0.834-1.04)	0.229	0.375	0.995
rs11734418	T	0.935 (0.836-1.05)	0.238	0.375	0.997
rs80127554	C	1.26 (0.856-1.86)	0.239	0.0227	0.907
rs6537484	G	1.07 (0.954-1.21)	0.241	0.375	0.929
rs190510692	A	1.26 (0.855-1.86)	0.243	0.0227	0.906
rs5344	G	0.763 (0.474-1.23)	0.264	0.0151	0.789
rs140087311	T	1.77 (0.643-4.86)	0.270	0.00506	0.649
rs9307840	C	1.16 (0.888-1.52)	0.275	0.0443	0.987
rs35513874	T	1.16 (0.886-1.52)	0.282	0.0445	0.990
rs189991953	A	9.85E+05 (8.46E-06-1.15E+17)	0.286	0.00106	0.631
rs6537487	C	1.08 (0.935-1.25)	0.290	0.189	0.924
rs7655892	T	1.11 (0.912-1.35)	0.304	0.0876	0.986
rs147979664	G	1.31 (0.783-2.18)	0.306	0.0139	0.878
rs62345688	C	1.08 (0.931-1.25)	0.309	0.160	0.998
rs116759404	T	0.783 (0.489-1.26)	0.311	0.0142	0.855
rs79831593	C	1.16 (0.863-1.56)	0.323	0.0370	0.950
rs5335	G	0.942 (0.837-1.06)	0.324	0.400	0.870
rs11938394	G	1.1 (0.907-1.34)	0.328	0.0863	0.987
rs3814415	G	0.924 (0.784-1.09)	0.346	0.157	0.816
rs6821368	T	1.07 (0.923-1.23)	0.387	0.192	0.947
rs6537489	T	1.06 (0.922-1.23)	0.391	0.192	0.948
rs1589056	T	0.814 (0.506-1.31)	0.396	0.0180	0.676
rs116468624	A	1.09 (0.874-1.35)	0.454	0.0786	0.877
rs7674137	G	1.06 (0.916-1.22)	0.456	0.194	0.956
rs35961327	C	0.937 (0.779-1.13)	0.489	0.0917	0.995

rs7666105	T	0.94 (0.782-1.13)	0.509	0.0924	0.997
rs4591581	A	0.94 (0.782-1.13)	0.513	0.0925	0.995
rs7657903	A	0.943 (0.785-1.13)	0.528	0.0998	0.944
rs75092505	T	0.875 (0.567-1.35)	0.548	0.0210	0.732
rs10305845	T	1.29 (0.534-3.09)	0.575	0.00542	0.855
rs2357924	G	0.953 (0.805-1.13)	0.576	0.166	0.764
rs1568136	T	1.03 (0.908-1.18)	0.615	0.252	0.949
rs6812093	T	1.03 (0.906-1.17)	0.645	0.253	0.957
rs10305895	G	1.03 (0.906-1.17)	0.646	0.253	0.956
rs6822565	C	1.03 (0.906-1.17)	0.646	0.254	0.957
rs72721939	C	0.968 (0.84-1.12)	0.654	0.213	0.873
rs9998357	T	1.07 (0.791-1.44)	0.676	0.0553	0.635
rs4591580	C	1.06 (0.789-1.43)	0.691	0.0547	0.642
rs4333164	T	1.06 (0.789-1.43)	0.691	0.0547	0.642
rs10305863	C	1.05 (0.775-1.43)	0.747	0.0517	0.636
rs190507340	A	1.13 (0.53-2.42)	0.748	0.00748	0.811
rs77192844	A	1.06 (0.69-1.63)	0.786	0.0222	0.731
rs80119650	C	0.975 (0.778-1.22)	0.825	0.108	0.601
rs116639478	A	1.07 (0.559-2.04)	0.843	0.00970	0.734
rs115341767	C	1.09 (0.423-2.82)	0.857	0.00528	0.674
rs150852865	A	1.05 (0.4-2.74)	0.927	0.00564	0.608
rs114781555	T	1.04 (0.262-4.14)	0.952	0.00255	0.621
rs6537485	A	1 (0.824-1.22)	0.964	0.0883	0.924

7.2.2 Common cIMT linear regression results

Table 20. Single-point logistic regression results for common cIMT, adjusted for age and gender. Regression was performed on the minor allele of each SNP and the results are sorted by *P* value, from lowest to highest. Variants lacking an rs number are labelled with their genomic location. Imputation quality given as the r^2.

SNP	Minor allele	β (95% CI)	SE	p	MAF	r^2
rs143184430	A	-0.0049 (-0.0169/0.0071)	0.00612	0.424	0.105	0.788
rs185348394	A	-0.000328 (-0.0152/0.0145)	0.00759	0.966	0.0802	0.652
rs6853791	T	0.0143 (0.00658/0.022)	0.00394	0.000292	0.301	0.847
rs6823438	T	0.0187 (0.00919/0.0282)	0.00485	0.000111	0.142	0.952
rs1987147	C	0.0119 (0.00422/0.0196)	0.00392	0.00237	0.424	0.759
rs1808039	C	0.0117 (0.00419/0.0192)	0.00383	0.00233	0.414	0.799
rs1808038	C	0.0146 (0.00682/0.0224)	0.00397	0.000236	0.306	0.825
rs7673178	T	-0.00499 (-0.0169/0.00697)	0.00610	0.413	0.105	0.795
rs1878406	T	0.0187 (0.00925/0.0281)	0.00482	0.000106	0.141	0.971
rs150852865	A	0.0518 (-0.00426/0.108)	0.0286	0.0695	0.00564	0.608
rs13144507	C	0.015 (0.00741/0.0226)	0.00387	0.000104	0.286	0.904
rs12509416	C	0.0119 (0.00463/0.0192)	0.00371	0.00131	0.391	0.865
rs6820938	G	0.019 (0.01/0.028)	0.00458	3.14E-05	0.188	0.868
rs6537480	A	0.0133 (0.00536/0.0212)	0.00405	0.00102	0.286	0.847
rs6537481	G	0.0129 (0.00459/0.0212)	0.00424	0.00240	0.231	0.872
rs4591580	C	0.00744 (-0.0104/0.0253)	0.00912	0.414	0.0547	0.642
rs116639478	A	-0.0102 (-0.0484/0.028)	0.0195	0.600	0.00970	0.734
rs4333164	T	0.00744 (-0.0104/0.0253)	0.00912	0.414	0.0547	0.642
rs73855814	T	0.0195 (0.00995/0.029)	0.00487	6.44E-05	0.133	0.985
rs9998357	T	0.00757 (-0.0103/0.0254)	0.00912	0.407	0.0553	0.635
rs1568137	C	-0.00432 (-0.0166/0.00793)	0.00625	0.490	0.101	0.780
rs7659823	G	0.019 (0.0101/0.0279)	0.00455	2.90E-05	0.188	0.875
rs9991584	G	0.0131 (0.00526/0.0209)	0.00400	0.00105	0.291	0.860
rs77192844	A	0.00271 (-0.0232/0.0286)	0.0132	0.837	0.0222	0.731
rs72957606	G	0.0194 (0.00987/0.0289)	0.00486	6.38E-05	0.133	0.992
rs10305838	C	0.0194 (0.00987/0.0289)	0.00486	6.39E-05	0.133	0.992
rs10305839	T	0.0195 (0.00997/0.029)	0.00486	6.34E-05	0.132	0.991
rs6841581	A	0.0194 (0.00987/0.0289)	0.00486	6.45E-05	0.133	0.992
rs10305845	T	-0.00735 (-0.0573/0.0426)	0.0255	0.773	0.00542	0.855
rs1801708	A	0.0127 (0.00484/0.0206)	0.00401	0.00151	0.288	0.861
rs6855875	T	0.0192 (0.0101/0.0283)	0.00463	3.37E-05	0.185	0.859
rs4835083	G	0.0127 (0.00484/0.0206)	0.00401	0.00156	0.290	0.861
rs4835084	A	0.0128 (0.00496/0.0206)	0.00400	0.00138	0.288	0.867
rs984457	T	0.0128 (0.00496/0.0206)	0.00400	0.00140	0.287	0.867
rs984458	T	0.0128 (0.00494/0.0207)	0.00401	0.00143	0.287	0.866
rs189991953	A	0.119 (-0.00566/0.244)	0.0636	0.0616	0.00106	0.631
rs56015601	T	0.0128 (0.00496/0.0206)	0.00400	0.00143	0.287	0.868
rs147979664	G	0.0239 (-0.0055/0.0533)	0.0150	0.111	0.0139	0.878
rs10305860	A	0.0127 (0.00486/0.0205)	0.00400	0.00145	0.287	0.868

rs9308218	T	0.0127 (0.00486/0.0205)	0.00400	0.00144	0.288	0.870
rs10305862	T	0.0127 (0.00486/0.0205)	0.00400	0.00142	0.287	0.870
rs10305863	C	0.0078 (-0.0106/0.0262)	0.00941	0.407	0.0517	0.636
rs6841473	T	0.0199 (0.0101/0.0297)	0.00501	7.00E-05	0.130	0.949
rs6823537	G	0.0126 (0.00478/0.0204)	0.00399	0.00165	0.287	0.874
rs702757	A	0.0125 (0.00468/0.0203)	0.00399	0.00176	0.287	0.874
rs4835410	G	0.0123 (0.0045/0.0201)	0.00398	0.00203	0.288	0.879
rs7663891	C	0.0122 (0.0044/0.02)	0.00398	0.00225	0.287	0.880
rs3814415	G	-0.00031 (-0.0103/0.00971)	0.00511	0.952	0.157	0.816
rs952402	C	0.0122 (0.0044/0.02)	0.00398	0.00224	0.287	0.882
rs1996374	A	0.0121 (0.00432/0.0199)	0.00397	0.00227	0.285	0.887
rs1996373	G	0.0121 (0.00432/0.0199)	0.00397	0.00229	0.286	0.887
rs17612742	C	0.02 (0.01/0.03)	0.00509	8.36E-05	0.129	0.931
rs1517134	T	0.012 (0.00424/0.0198)	0.00396	0.00245	0.285	0.891
rs115341767	C	0.051 (-0.00388/0.106)	0.0280	0.0688	0.00528	0.674
rs150948757	T	-0.0118 (-0.0516/0.028)	0.0203	0.562	0.0103	0.639
rs13143677	A	0.0116 (0.0039/0.0193)	0.00393	0.00326	0.282	0.913
rs1400554	T	0.0113 (0.00403/0.0186)	0.00371	0.00226	0.301	0.981
rs6840756	C	0.011 (0.00379/0.0182)	0.00368	0.00274	0.300	0.998
rs4835086	A	0.011 (0.00377/0.0182)	0.00369	0.00275	0.300	0.997
rs17474816	A	0.0111 (0.00387/0.0183)	0.00369	0.00270	0.300	0.997
rs6848108	G	0.0111 (0.00387/0.0183)	0.00369	0.00270	0.300	0.996
rs7655670	C	0.0111 (0.00389/0.0183)	0.00368	0.00268	0.300	0.998
rs17474823	G	0.011 (0.00375/0.0183)	0.00370	0.00300	0.299	0.993
rs143614835	T	0.0258 (-0.015/0.0666)	0.0208	0.215	0.00919	0.770
rs57886575	A	0.011 (0.00373/0.0183)	0.00371	0.00288	0.306	0.975
rs10011966	A	0.011 (0.00379/0.0182)	0.00368	0.00270	0.301	1.00
rs6819740	T	0.0111 (0.00389/0.0183)	0.00368	0.00266	0.300	0.998
rs6537482	A	0.0111 (0.00389/0.0183)	0.00368	0.00266	0.301	1.00
rs190510692	A	0.0165 (-0.00584/0.0388)	0.0114	0.149	0.0227	0.906
rs78049276	C	0.0197 (0.00955/0.0299)	0.00518	0.000147	0.125	0.922
rs6537484	G	0.00726 (0.000145/0.0144)	0.00363	0.0455	0.375	0.929
rs190507340	A	-0.00287 (-0.047/0.0412)	0.0225	0.899	0.00748	0.811
rs6537485	A	0.00776 (-0.0042/0.0197)	0.00610	0.204	0.0883	0.924
rs7657903	A	0.00591 (-0.00524/0.0171)	0.00569	0.299	0.0998	0.944
rs4591581	A	0.00571 (-0.00552/0.0169)	0.00573	0.319	0.0925	0.995
rs6537487	C	0.0145 (0.00574/0.0233)	0.00447	0.00113	0.189	0.924
rs139452637	T	0.0195 (0.00798/0.031)	0.00588	0.000928	0.0999	0.898
rs7666105	T	0.00572 (-0.00551/0.017)	0.00573	0.318	0.0924	0.997
rs80127554	C	0.0164 (-0.00594/0.0387)	0.0114	0.151	0.0227	0.907
rs35961327	C	0.00569 (-0.00558/0.017)	0.00575	0.322	0.0917	0.995
rs75092505	T	0.0143 (-0.0126/0.0412)	0.0137	0.299	0.0210	0.732
rs1568136	T	0.00711 (-0.000691/0.0149)	0.00398	0.0744	0.252	0.949
rs75736224	G	0.0196 (0.00813/0.0311)	0.00585	0.000799	0.0999	0.909
rs6537489	T	0.0144 (0.0058/0.023)	0.00439	0.00101	0.192	0.948

rs10305895	G	0.00712 (-0.000661/0.0149)	0.00397	0.0725	0.253	0.956
rs6821368	T	0.0144 (0.0058/0.023)	0.00439	0.00101	0.192	0.947
rs6812093	T	0.00712 (-0.000642/0.0149)	0.00396	0.0721	0.253	0.957
rs6822565	C	0.00713 (-0.000632/0.0149)	0.00396	0.0720	0.254	0.957
rs4835411	C	0.0195 (0.00815/0.0308)	0.00579	0.000775	0.101	0.919
rs7674137	G	0.0143 (0.00577/0.0228)	0.00435	0.00104	0.194	0.956
rs4835412	A	0.0191 (0.00787/0.0303)	0.00573	0.000874	0.102	0.944
rs908581	T	0.0191 (0.00789/0.0303)	0.00572	0.000859	0.102	0.944
rs116759404	T	-0.00411 (-0.0335/0.0253)	0.0150	0.784	0.0142	0.855
rs116468624	A	0.00867 (-0.00434/0.0217)	0.00664	0.192	0.0786	0.877
rs192252324	G	-0.00757 (-0.046/0.0308)	0.0196	0.699	0.0101	0.735
rs7655892	T	0.00771 (-0.00393/0.0194)	0.00594	0.194	0.0876	0.986
rs11938394	G	0.0078 (-0.00392/0.0195)	0.00598	0.192	0.0863	0.987
rs6843446	G	-0.000314 (-0.00794/0.00731)	0.00389	0.936	0.247	0.993
rs10003447	T	-0.000315 (-0.00794/0.00731)	0.00389	0.935	0.247	0.995
rs78047355	G	-0.00485 (-0.035/0.0253)	0.0154	0.753	0.0127	0.952
rs11936340	T	0.00052 (-0.0072/0.00824)	0.00394	0.895	0.243	0.977
rs10028507	A	-0.000307 (-0.00793/0.00732)	0.00389	0.937	0.247	0.996
rs62345688	C	-0.00493 (-0.0138/0.00399)	0.00455	0.278	0.160	0.998
rs77249653	T	-0.00459 (-0.0348/0.0256)	0.0154	0.766	0.0130	0.931
rs10029799	C	-0.0000846 (-0.00771/0.00754)	0.00389	0.983	0.248	0.992
rs2048894	A	-0.0000657 (-0.00769/0.00756)	0.00389	0.987	0.248	0.992
rs13121745	T	0.000904 (-0.00821/0.01)	0.00465	0.846	0.185	0.840
rs9307837	T	0.00106 (-0.0066/0.00872)	0.00391	0.787	0.245	0.993
rs1400556	A	0.00102 (-0.00664/0.00868)	0.00391	0.795	0.245	0.994
rs9307838	C	0.00103 (-0.00663/0.00869)	0.00391	0.793	0.245	0.994
rs2357924	G	-0.0109 (-0.0211/-0.000708)	0.00520	0.0361	0.166	0.764
rs10305912	A	0.00108 (-0.00658/0.00874)	0.00391	0.783	0.245	0.996
rs3756022	C	0.0011 (-0.00656/0.00876)	0.00391	0.779	0.245	0.996
rs10305916	G	-0.0022 (-0.0105/0.00609)	0.00423	0.603	0.194	0.983
rs10305918	C	0.00135 (-0.00659/0.00929)	0.00405	0.738	0.232	0.959
rs10305919	A	0.00135 (-0.00659/0.00929)	0.00405	0.739	0.232	0.959
rs10305920	G	0.00115 (-0.00651/0.00881)	0.00391	0.769	0.245	0.997
rs10305921	C	0.00115 (-0.00651/0.00881)	0.00391	0.768	0.245	0.997
rs7677315	G	0.00117 (-0.00647/0.00881)	0.00390	0.765	0.245	0.997
rs2884337	C	0.00118 (-0.00646/0.00882)	0.00390	0.762	0.245	0.998
rs2884339	A	0.00121 (-0.00643/0.00885)	0.00390	0.756	0.245	0.998
rs2884340	G	0.00121 (-0.00643/0.00885)	0.00390	0.756	0.245	0.998
rs6841799	G	0.00122 (-0.00642/0.00886)	0.00390	0.754	0.245	0.999
rs6840375	A	0.00123 (-0.00641/0.00887)	0.00390	0.753	0.245	0.999
rs5333	C	0.00123 (-0.00641/0.00887)	0.00390	0.752	0.245	0.999
rs2292765	A	0.00128 (-0.00642/0.00898)	0.00393	0.744	0.247	0.983
rs7690347	A	0.00131 (-0.00639/0.00901)	0.00393	0.738	0.246	0.983
rs6854578	C	-0.000365 (-0.03/0.0292)	0.0151	0.981	0.0187	0.681
rs10305929	A	-0.00037 (-0.03/0.0292)	0.0151	0.980	0.0187	0.680

rs114781555	T	-0.0781 (-0.159/0.00304)	0.0414	0.0591	0.00255	0.621
rs5335	G	-0.000731 (-0.00789/0.00642)	0.00365	0.841	0.400	0.870
rs5342	G	-0.000755 (-0.00754/0.00603)	0.00346	0.827	0.375	0.995
rs5344	G	-0.000926 (-0.0307/0.0289)	0.0152	0.951	0.0151	0.789
rs9307839	T	-0.00075 (-0.00751/0.00601)	0.00345	0.828	0.375	0.998
rs9307840	C	0.00809 (-0.00783/0.024)	0.00812	0.319	0.0443	0.987
rs35513874	T	0.00798 (-0.00788/0.0238)	0.00809	0.324	0.0445	0.990
rs6831880	G	-0.000728 (-0.00749/0.00603)	0.00345	0.833	0.375	0.999
rs79831593	C	0.00276 (-0.0147/0.0202)	0.00889	0.756	0.0370	0.950
rs11721708	G	-0.0023 (-0.00926/0.00466)	0.00355	0.516	0.331	0.999
rs140087311	T	0.0273 (-0.029/0.0836)	0.0287	0.341	0.00506	0.649
rs11734418	T	-0.000628 (-0.00741/0.00615)	0.00346	0.856	0.375	0.997
rs62341322	A	-0.00253 (-0.00949/0.00443)	0.00355	0.476	0.330	0.998
rs72721939	C	-0.00902 (-0.0176/-0.000435)	0.00438	0.0394	0.213	0.873
rs4835087	A	-0.000963 (-0.00774/0.00582)	0.00346	0.781	0.375	0.997
rs1589056	T	0.002 (-0.0276/0.0316)	0.0151	0.894	0.0180	0.676
rs10032496	G	-0.000946 (-0.00771/0.00582)	0.00345	0.784	0.375	0.999
rs35993155	A	-0.00251 (-0.00947/0.00445)	0.00355	0.480	0.330	0.998
rs80119650	C	-0.0168 (-0.0304/-0.00324)	0.00692	0.0153	0.108	0.601
rs9884563	A	-0.000679 (-0.0075/0.00614)	0.00348	0.845	0.398	0.959

7.3 Haplotype association for carotid plaque

Table 21. Haplotype logistic regression analysis results for carotid plaque, adjusted for age and gender. Table is ordered by haplotype composition and haplotypes are numbered accordingly. First SNP in each haplotype is rs6842241.

Haplotype	#	β	SE	OR (95% CI)	p	Frequency
AAGCCGAAGCT	1	12.8	198	349000 (4.04E-164-3.01E+174)	0.949	0.00109
AAGCCGAAGTG	2	-0.144	3.17E-01	0.866 (0.465-1.61)	0.65	0.00715
AAGTTACGACT	3	12.1	184	177000 (1.6E-152-1.97E+162)	0.948	0.000466
AGACCAAAGCT	4	12.2	229	193000 (5.1E-190-7.31E+199)	0.958	0.000311
AGACCAAAGTG	5	0.158	0.283	1.17 (0.672-2.04)	0.577	0.0103
AGACCAAGACT	6	-1.59	9.37E-01	0.203 (0.0324-1.27)	0.0889	0.000932
AGACCACAACG	7	-0.0241	0.221	0.976 (0.633-1.5)	0.913	0.016
AGACCACAACT	8	0.248	0.177	1.28 (0.906-1.81)	0.161	0.0269
AGACCACAATG	9	0.338	0.485	1.4 (0.542-3.62)	0.486	0.00388
AGACCACAGCG	10	0.67	0.335	1.95 (1.01-3.76)	0.0453	0.00885
AGACCACAGTG	11	0.208	0.688	1.23 (0.32-4.74)	0.762	0.00171
AGACCACGACT	12	0.165	0.237	1.18 (0.742-1.87)	0.486	0.0144
AGACCGAAGCT	13	0.148	0.253	1.16 (0.707-1.9)	0.559	0.0124
AGACCGAAGTG	14	0.478	0.263	1.61 (0.964-2.7)	0.0686	0.0141
AGACCGAGACT	15	0.358	0.356	1.43 (0.712-2.87)	0.315	0.00715
AGACCGAGGCT	16	11.7	229	121000 (3.18E-190-4.63E+199)	0.959	0.000311
AGACCGCAACT	17	0.236	1.17	1.27 (0.128-12.5)	0.84	0.000621
AGACCGCAGCG	18	11.8	222	140000 (2.51E-184-7.75E+193)	0.957	0.000311
AGACCGCGACT	19	0.244	0.366	1.28 (0.623-2.61)	0.505	0.0059
AGATTACGACT	20	11.6	221	104000 (3.95E-184-2.75E+193)	0.958	0.000311
CAGCCACAACT	21	11.7	2.29E+02	119000 (2.23E-190-6.36E+199)	0.959	0.000466
CAGCCACAATG	22	-0.33	0.355	0.719 (0.359-1.44)	0.352	0.00606
CAGCCACGACT	23	-0.496	1.23	0.609 (0.0544-6.82)	0.687	0.000621
CAGCCACGATG	24	-0.285	1.27	0.752 (0.0618-9.15)	0.823	0.000466
CAGCCGAAACT	25	0.931	1.09	2.54 (0.302-21.3)	0.391	0.00109
CAGCCGAAGCT	26	-0.0283	0.0573	0.972 (0.869-1.09)	0.621	0.38
CAGCCGAAGTG	27	-0.0827	0.0737	0.921 (0.797-1.06)	0.262	0.171
CAGCCGAGACT	28	0.0276	0.187	1.03 (0.712-1.48)	0.883	0.023
CAGCCGCAACG	29	-0.0204	0.331	0.98 (0.512-1.87)	0.951	0.00699
CAGCCGCAACT	30	0.0119	0.254	1.01 (0.616-1.66)	0.962	0.0116
CAGCCGCAATG	31	0.9	1.13	2.46 (0.27-22.4)	0.425	0.000932
CAGCCGCAGCG	32	12.4	196	240000 (1.47E-162-3.93E+172)	0.95	0.00109
CAGCCGCAGTG	33	-0.927	1.51	0.396 (0.0206-7.62)	0.539	0.000311
CAGCCGCGACT	34	0.43	0.279	1.54 (0.89-2.66)	0.123	0.0113
CAGCCGCGGTG	35	-0.071	0.472	0.931 (0.37-2.35)	0.88	0.00326
CAGCTGAAGCT	36	-1.18	6.77E-01	0.309 (0.0818-1.16)	0.0827	0.00155
CAGCTGCAACT	37	-0.0133	0.879	0.987 (0.176-5.52)	0.988	0.00109
CAGCTGCGACT	38	0.017	0.135	1.02 (0.781-1.32)	0.899	0.0463
CAGCTGCGATG	39	-0.21	0.247	0.81 (0.5-1.31)	0.394	0.0115
CAGTTACGACT	40	-0.44	0.254	0.644 (0.391-1.06)	0.0829	0.0106

CGACCACAACG	41	-0.709	6.18E-01	0.492 (0.147-1.65)	0.251	0.00186
CGACCACAATG	42	0.169	8.81E-01	1.18 (0.211-6.66)	0.847	0.000932
CGACCACAGCG	43	-0.945	1.42E+00	0.389 (0.0239-6.33)	0.507	0.000311
CGACCGAAACT	44	0.774	0.805	2.17 (0.448-10.5)	0.336	0.00155
CGACCGAAGCG	45	-0.433	0.645	0.648 (0.183-2.3)	0.502	0.00155
CGACCGAAGCT	46	0.0998	0.198	1.1 (0.749-1.63)	0.615	0.0205
CGACCGAAGTG	47	-0.16	0.109	0.852 (0.688-1.05)	0.142	0.0669
CGACCGCAATG	48	1.17	1.08	3.22 (0.389-26.6)	0.278	0.0014
CGATTAAAGCT	49	-0.166	0.192	0.847 (0.582-1.23)	0.386	0.0197
CGATTAAAGTG	50	-0.201	0.158	0.818 (0.6-1.12)	0.205	0.03
CGATTACAACG	51	0.345	0.348	1.41 (0.714-2.79)	0.321	0.00683
CGATTACAACT	52	-0.669	1.02E+00	0.512 (0.0696-3.77)	0.511	0.000621
CGATTACGACT	53	0.318	0.195	1.37 (0.937-2.01)	0.103	0.0227