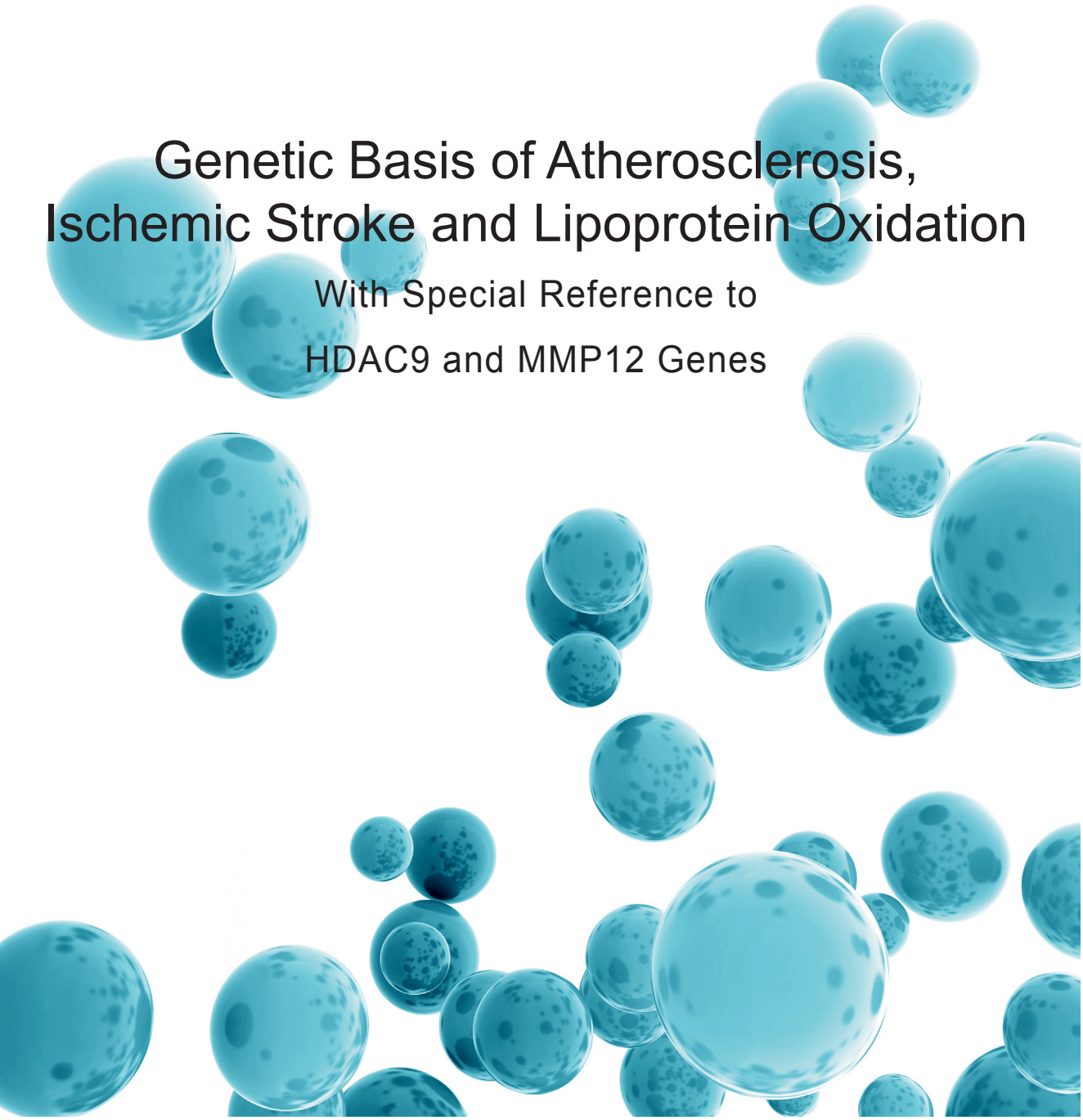
Acta Universitatis Tamperensis 2073

KARI-MATTI MÄKELÄ

# Genetic Basis of Atherosclerosis, Ischemic Stroke and Lipoprotein Oxidation

## With Special Reference to HDAC9 and MMP12 Genes



KARI-MATTI MÄKELÄ

# Genetic Basis of Atherosclerosis, Ischemic Stroke and Lipoprotein Oxidation

## With Special Reference to HDAC9 and MMP12 Genes



ACADEMIC DISSERTATION
To be presented, with the permission of
the Board of the School of Medicine of the University of Tampere,
for public discussion in the Auditorium of
School of Health Sciences, T Building, Medisiinarinkatu 3,
Tampere, on August 21st, 2015, at 12 o'clock.

UNIVERSITY OF TAMPERE

KARI-MATTI MÄKELÄ

# Genetic Basis of Atherosclerosis, Ischemic Stroke and Lipoprotein Oxidation

## With Special Reference to HDAC9 and MMP12 Genes

*Acta Universitatis Tamperensis 2073*
*Tampere University Press*
*Tampere 2015*

UNIVERSITY
OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, School of Medicine
Fimlab Laboratories Ltd.
Finland

*Supervised by*
Professor Terho Lehtimäki
University of Tampere
Finland

*Reviewed by*
Docent Timo Hiltunen
University of Helsinki
Finland
Docent Kari Kervinen
University of Oulu
Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.



Cover design by
Mikko Reinikka

Distributor:
verkkokauppa@juvenesprint.fi
https://verkkokauppa.juvenes.fi

Acta Universitatis Tamperensis 2073
ISBN 978-951-44-9854-1 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1567
ISBN 978-951-44-9855-8 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2015

# Abstract

**Background:** Cardiovascular diseases (CVD) are the number one cause of death and morbidity in the modern world. Oxidized low-density lipoprotein (oxLDL) is considered to be a key factor in the development of atherosclerosis, which leads to CVD such as coronary artery disease (CAD) and ischemic stroke. We used genome-wide association (GWAS) and genome-wide expression approaches in order to find novel genes and their genetic variants associated with the pathogenesis and risk factors of atherosclerosis, ischemic stroke and circulatory oxLDL concentrations.

**Aims of the Study: 1)** To find single nucleotide polymorphisms (SNP) associated with oxLDL by performing a GWAS on serum oxLDL-levels. **2)** Study the association of the lead-SNP affecting LDL oxidation, rs676210 (apolipoprotein-B [apoB] Pro2739Leu), with the prevalence of CAD, MI and ischemic stroke. **3)** Study the association of ischemic stroke risk increasing HDAC9 variants with carotid plaque prevalence and intima-media thickness (cIMT). **4)** Find novel loci associated with ischemic stroke by performing an age-of-onset informed GWAS. **5)** Study the expression of the found MMP12 gene and HDAC9 in clinically significant atherosclerotic arteries. **6)** Study HDAC9 and MMP12 expression in relation to histologically determined severity of atherosclerotic plaques and gene markers for plaque stability, M1/M2 macrophages and smooth muscle cells.

**Materials and Methods:** The artery samples used in the study were from Tampere Vascular Study (TVS, N=96, original communications **III**- **V**) and blood samples from the participants of the Young Finns Study (YFS, N=2080, **I**), the Ludwigshafen Risk and Cardiovascular Health (LURIC, study A, N=2913, with 271 cases and 2642 controls, **I** and **II**), Kooperative Gesundheitsforschung in der Region Augsburg (KORA) study (B, N=1326, **I**), the Finnish Cardiovascular Study (FINCAVAS, N=1118, C, **I**), Angiography and Genes Study (ANGES, N=808, D, **I**), Wellcome Trust Case Control Consortium 2 (WTCCC2, E, N=3548 cases, 5972 controls, **II**- **IV**), CHARGE consortium (F, 25179 plaques and 31210 cIMT, **III**) and METASTROKE consortium (G, 6778 cases and 12095 controls, **IV**). Genotyping and imputation in all studies passed strict quality control (QC) measures. DNA (YFS and studies A-G) and mRNA (TVS) were isolated with appropriate commercial kits and the quality and integrity of RNA was closely examined. Serum levels of oxLDL were measured with Mercodia ELISA oxLDL

assay (**I** and **II**). Statistical analyses were performed using PLINK, ProbABEL, PolyPhen-2 softwares and R statistical programming language.

**Results:** In study **I**, the genetic variant rs676210 (Pro2739Leu) on apoB associated with oxLDL ($p=4.3 \times 10^{-136}$, effect size = 13.2 U/l per allele). Using PolyPhen-2 software Pro2739Leu was predicted to cause functional change in apoB protein structure. It did not associate significantly with CAD (hazard ratio [HR]=1.00 [0.94–1.06] per allele) or MI (HR=1.04 [0.96–1.12]). In study **II,** rs676210 associated with cerebrovascular disease events (p=0.030; odds ratio=1.29 [95% confidence interval 1.03–1.63] for risk allele G). In study E, rs676210 did not associate with the history of ischemic stroke.

In study **III**, both HDAC9 SNPs (rs11984041 and rs2107595) were associated with both common carotid IMT (p=0.00391 and p=0.0018, respectively) and with presence of carotid plaque (p=0.00425 and p=0.0022, respectively). HDAC9 staining was seen in the nuclei and cytoplasm of vascular smooth muscle cells, and in endothelial cells of cerebral and systemic arteries. In TVS, HDAC9 expression was upregulated in carotid plaques compared to atherosclerosis free controls (p=0.00000103). In study **IV**, we found novel MMP12 locus (rs660599) associated with ischemic stroke using an age-of-onset informed GWAS ($p=2.5 \times 10^{-7}$). In TVS, MMP12 gene was upregulated in atherosclerotic plaques compared to control vessels (fold change=336, $p=1.2 \times 10^{-15}$).

In study **V**, HDAC9 and MMP12 expressions increased with plaque severity determined by American Heart Association classification (p=0.00018, and p<0.0001, for trend respectively) in all artery beds. HDAC9 expression correlated significantly with MMP12 expression in carotid plaques (r=0.46, p=0.012, N=29), and control samples (r=-0.44, p=0.034, N=28), but not in other artery beds. MMP12 expression showed positive correlation ($p < 0.05$) with 22% (2/9) M1 macrophage markers, and 79% (11/14) M2 macrophage markers, negative correlation with 72% (18/25) SMC markers, and no correlation with plaque stability markers in the carotid artery plaques.

**Conclusions:** These results give novel insight into the genetic background of atherosclerosis, ischemic stroke and lipoprotein oxidation, and specifically indicate that apoB rs676210 (Pro2739Leu), HDAC9 and MMP12 related gene variations associate with the risk of ischemic stroke. HDAC9 variants may act via promoting atherosclerosis in the carotid artery. Both HDAC9 and MMP12 are overexpressed in atherosclerotic plaques and correlate with plaque severity. In addition, apoB Pro2739Leu is a novel genetic factor regulating circulatory oxLDL levels.

# Tiivistelmä

**Tausta:** Ateroskleroosista eli valtimonkovettumataudista johtuvat sydän- ja verisuonitaudit aiheuttavat suurimman osan kuolemista ja sairastuvuudesta länsimaissa. Hapettunutta ”low-density" –lipoproteiinia (oxLDL) pidetään olennaisena riskitekijänä ateroskleroosin tautiprosessissa. Ateroskleroosivaurion repeämä ja sitä seuraava veritulppa voivat aiheuttaa esimerkiksi sydäninfarktin, sydänperäisen äkkikuoleman tai iskeemisen aivoverenkiertohäiriön. Tässä tutkimuksessa käytimme koko genomin laajuista lähestymistapaa. Tarkoituksenamme oli löytää uusia geenejä tai niiden geenivariantteja, jotka vaikuttavat ateroskleroosin ja iskeemisen aivoverenkiertohäiriön patogeneesiin sekä verenkierron oxLDL:n määrään.

**Tavoitteet: 1)** etsiä perimänlaajuisesti yhden emäksen vaihteluita (SNP, single nucleotide polymorphism), jotka vaikuttavat verenkierron hapettuneen oxLDL:n pitoisuuteen. **2)** Tutkia merkittävimmän oxLDL:n pitoisuuteen vaikuttavan geenivariantin SNP rs676210 (Pro2739Leu) yhteyttä sepelvaltimotaudin, sydäninfarktin ja iskeemisen aivoverenkiertohäiriön esiintyvyyteen. **3)** Tutkia aivoinfarktin riskiä lisäävien HDAC9-geenivarianttien (SNP:t rs11984041 ja rs2107595) vaikutusta kaulasuonten ateroskleroottisten plakkien esiintyvyyteen ja intima-median paksuuteen (cIMT). **4)** Etsiä iskeemiseen aivoverenkiertohäiriöön liittyviä geenimerkkejä iän huomioon ottavalla genomin laajuisella lähestymistavalla. **5)** Tutkia HDAC9-geenin sekä löydetyn MMP12-geenin ilmentymistä kliinisesti merkittävissä verisuonissa. **6)** Tutkia HDAC9- ja MMP12-geenien ilmentymistä suhteessa histologisesti määritettyyn ateroskleroosin vaikeusasteeseen, plakin stabiilisuutta/repeämisherkkyyttä kuvaaviin geenimarkkereihin sekä M1/M2-tyyppisten makrofagien ja sileiden lihassolujen ilmentämiin geenimarkkereihin.

**Aineisto ja menetelmät:** Tutkimuksessa käytetyt valtimonäytteet ovat osa Tampere Vascular Study:a (TVS, N=96, osatyöt **III**-**V**). Verinäytteet ovat peräisin Lasten ja Nuorten Sepelvaltimotaudin riskitekijät –tutkimuksen (YFS, N=2080, osatyö **I**) osallistujilta sekä aineistoista LURIC (A, N=2913, 271 tapausta ja 2642 verrokkia, osatyöt **I**, **II**), KORA (B, N=1326, osatyö **I**), FINCAVAS (C, N=1118, osatyö **I**), ANGES (D, N=808, osatyö **I**), WTCCC2-konsortio (E, N=3548 tapausta, 5972 verrokkia, osatyöt **II**-**IV**), CHARGE-konsortio (F, plakit N=25179 ja cIMT N=31210, osatyö **III**) METASTROKE-konsortio (G, 6778 tapausta ja 12095 verrokkia, **IV**). Kaikkien tutkimusten genotyypityksen laatu kontrolloitiin (QC) ja aineistot imputoitiin yhdenmukaisesti. DNA (YFS, aineistot A-F) ja mRNA

(TVS) eristettiin kaupallisilla eristyskiteillä huomioiden RNA:n laatu ja eheys. Hapettuneen oxLDL:n pitoisuutta mitattiin Mercodian kehittämällä ELISA-menetelmällä (**I** ja **II**). Tilastoanalyysit tehtiin PLINK-, ProbABEL-, PolyPhen-2-ohjelmistoja ja R-ohjelmointikieltä käyttäen.

**Tulokset:** Työssä **I** apoliporoteiini B:n aminohapporakennetta muuttava (Pro2739Leu) SNP rs676210 assosioitui oxLDL pitoisuuteen ($p=4{,}3x10^{-136}$, efektikoko 13,2 U/l per alleeli). PolyPhen-2-ohjelmistolla tehdyn analyysin mukaan (Pro2739Leu) aminohapon muuttuminen aiheuttaa apoB:n proteiinin toiminnallisen muutoksen. Pro2739Leu ei assosioitunut sepelvaltimotautiin (vaarasuhde [HR] = 1,00 [0,94–1,06] per alleeli) eikä sydäninfarktiin (HR=1,04 [0,96–1,12]). Työssä **II** rs676210 assosioitui aivoverenkiertohäiriötapauksiin (p=0,030; riskisuhde=1,29 [95 % luottamusväli 1,03–1,63] alleeli G:lle). Aineistossa E rs676210 ei assosioitunut iskeemiseen aivotapahtumaan.

Työssä **III** rs11984041 ja rs2107595 assosioituivat merkitsevästi cIMT-muuttujaan ja plakin olemassaoloon aineistossa F (rs2107595 p=0,0018 ja p=0,0022; rs11984041 0.00391 ja p=0.00425). Histologisessa tutkimuksessa HDAC9:n värjäytymistä nähtiin sileiden lihassolujen ja endoteelisolujen solulimassa keskimmäisessä aivovaltimossa, sisemmässä kaulavaltimossa, aortassa ja sepelvaltimoissa. TVS:ssa havaittiin, että HDAC9:n ilmentyminen oli suurentunut kaulavaltimon plakeissa kontrolleihin verrattuna (p=0,00000103). Työssä **IV** löysimme uuden iskeemiseen aivoverenkiertohäiriöön assosioituvan geenimerkin rs660599 MMP12:n läheltä ($p=2.5x10^{-7}$). TVS:ssa MMP12 ilmentyi 336-kertaisesti ateroskleroottisessa kaulavaltimossa verrattuna ateroskleroosittomaan suoneen ($p=1.2x10^{-15}$).

Työssä **V** TVS-aineistossa HDAC9:n ja MMP12:n ekspressiot assosioituivat plakin vakavuuteen (p=0,00018 ja p<0,0001, vastaavasti). MMP12:n ilmentyminen korreloi HDAC9:n ilmentymisen kanssa positiivisesti kaulasuonissa (r=0,46, p=0,012, N=29) ja negatiivisesti histologisesti terveissä kontrollisuonissa (r=-0,44, p=0,034, N=28). MMP12:n ilmentyminen korreloi positiivisesti 2/9 M1-makrofagimarkkerin ja 11/14 M2-makrofagimarkkerin kanssa. Korrelaatio oli negatiivinen 18/25 sileälihassolumarkkerin kanssa. Korrelaatiota ei löytynyt plakin stabiliteettimarkkerien kanssa sisemmissä kaulavaltimoissa.

**Johtopäätökset:** Löysimme uusia ateroskleroosin ja iskeemisen aivoverenkiertohäiriön riskiin sekä oxLDL-pitoisuuteen vaikuttavia geneettisiä tekijöitä. HDAC9 edistää kaulasuonen ateroskleroosia. HDAC9 ja MMP12 assosioituvat ateroskleroottisen plakin vakavuusasteeseen ja iskeemisen aivoverenkiertohäiriön riskiin. ApoB Pro2739Leu on uusi seerumin oxLDL-pitoisuuteen vaikuttava geenivariaatio.

# Contents

# List of Original Publications

This thesis is based on the following original communications, referred to in the text with their Roman numeral I-V. These original publications have been reproduced with the permission of the copyright holders.

I. Mäkelä, K.M., Seppälä, I., Hernesniemi, J.A., Lyytikäinen, L.P., Oksala, N., Kleber, M.E., Scharnagl, H., Grammer, T.B., Baumert, J., Thorand, B., Jula, A., Hutri-Kähönen, N., Juonala, M., Laitinen, T., Laaksonen, R., Karhunen, P.J., Nikus, K.C., Nieminen, T., Laurikka, J., Kuukasjärvi, P., Tarkka, M., Viik, J., Klopp, N., Illig, T., Kettunen, J., Ahotupa, M., Viikari, J.S., Kähönen, M., Raitakari, O.T., Karakas, M., Koenig, W., Boehm, B.O., Winkelmann, B.R., März, W. and Lehtimäki, T., 2013. Genome-wide association study pinpoints a new functional apolipoprotein B variant influencing oxidized low-density lipoprotein levels but not cardiovascular events: AtheroRemo Consortium. Circulation. Cardiovascular genetics, 6(1), pp. 73-81.

II. Mäkelä, K.M., Traylor, M., Oksala, N., Kleber, M.E., Seppälä, I., Lyytikäinen, L.P., Hernesniemi, J.A., Kähönen, M., Bevan, S., Rothwell, P.M., Sudlow, C., Dichgans, M., Wellcome Trust Case Control Consortium 2 (WTCCC2), Delgado, G., Grammer, T.B., Scharnagl, H., Markus, H.S., März, W. and Lehtimäki, T., 2014. Association of the novel single-nucleotide polymorphism which increases oxidized low-density lipoprotein levels with cerebrovascular disease events. Atherosclerosis, 234(1), pp. 214-217.

III. Markus, H.S., Mäkelä, K.M., Bevan, S., Raitoharju, E., Oksala, N., Bis, J.C., O'Donnell, C., Hainsworth, A. And Lehtimäki, T., 2013. Evidence HDAC9 genetic variant associated with ischemic stroke increases risk via promoting carotid atherosclerosis. Stroke, 44(5), pp. 1220-1225.

IV. Traylor, M., Mäkelä, K.M., Kilarski, L.L., Holliday, E.G., Devan, W.J., Nalls, M.A., Wiggins, K.L., Zhao, W., Cheng, Y.C., Achterberg, S., Malik, R., Sudlow, C., Bevan, S., Raitoharju, E., Metastroke, International Stroke Genetics Consortium, Wellcome Trust Case Consortium 2 (WTCCC2), Oksala, N., Thijs, V., Lemmens, R., Lindgren, A., Slowik, A., Maguire, J.M., Walters, M., Algra, A., Sharma, P., Attia, J.R., Boncoraglio, G.B., Rothwell, P.M., De Bakker, P.I., Bis, J.C., Saleheen, D., Kittner, S.J., Mitchell, B.D., Rosand, J., Meschia, J.F., Levi, C., Dichgans, M., Lehtimäki, T., Lewis, C.M. And Markus, H.S., 2014. A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach. PLoS genetics, 10(7), pp. e1004469.

V. Mäkelä K.M., Seppälä, I., Raitoharju, E., Lyytikäinen, L.P., Illig, T., Klopp, N., Kholova, I., Oksala, N., and Lehtimäki, T., 2015. Expression and cell specific correlations of histone deacetylase 9 and matrix metalloproteinase 12 in atherosclerotic plaques in Tampere vascular study. Submitted to Atherosclerosis.

# Abbreviations

| | |
|---|---|
| 3VD | three vessel disease |
| AHA | American Heart Association |
| AIC | akaike information criterion |
| ANGES | Angiography and Genes Study |
| ANOVA | analysis of variance |
| apoB | apolipoprotein-B |
| BIC | Bayesian Information Criterion |
| BMI | body mass index |
| CAD | coronary artery disease |
| CCA | common carotid artery |
| CE | cardioembolic stroke |
| CHARGE | Cohorts for Heart and Aging Research in Genomic Epidemiology |
| CT | computed tomography |
| CVD | cardiovascular disease |
| DNA | deoxyribonucleic acid |
| ECG | electrocardiography |
| EDTA | ethylenediaminetetraacetic acid |
| ECM | extracellular matrix |
| ELISA | enzymelinked immunosorbent assay |
| FINCAVAS | The Finnish Cardiovascular Study |
| GWAS | genome-wide association study |
| HDAC9 | histone deacetylase 9 |
| HDL | high-density lipoprotein |
| HWE | Hardy–Weinberg equilibrium |
| ICAM-1 | inter-cellular adhesion molecule 1 |
| IDL | intermediate-density lipoprotein |
| IL | interleukin |
| IMT | intima-media thickness |

| | |
|---|---|
| KORA | Kooperative gesundheitsforschung in der Region Augsburg |
| LAA | large artery atherosclerosis |
| LD | linkage disequilibrium |
| LDL | low-density lipoprotein |
| LURIC | the LUdwigshafen RIsk and Cardiovascular Health |
| MASS | Modern Applied Statistics with S |
| MCA | middle cerebral artery |
| MI | myocardial infarction |
| MMP | matrix metalloproteinase |
| MRI | magnetic resonance imaging |
| NET | neutrophil extracellular trap |
| NSTEMI | non ST segment elevation myocardial infarction |
| PCI | percutaneus coronary intervention |
| oxLDL | oxidized low-density lipoprotein |
| QQ | quantile quantile |
| RNA | ribonucleic acid |
| SMC | smooth muscle cell |
| SNP | single nucleotide polymorphism |
| STEMI | ST segment elevation myocardial infarction |
| SVD | small vessel disease |
| TOAST | Trial of Org 10172 in Acute Stroke Treatment |
| tPA | tissue plasminogen activator |
| TLR | toll-like receptor |
| TIA | transient ischemic attack |
| TVS | Tampere Vascular Study |
| UAP | unstable angina pectoris |
| VCAM-1 | vascular-cell adhesion molecule 1 |
| VLDL | very-low-density lipoprotein |
| VSMC | vascular smooth muscle cell |
| WHO | World Health Organization |
| YFS | the Young Finns Study |
| WTCCC1 | Wellcome Trust Case Control Consortium |

Abbreviations are defined at first mention in the abstract and the review of the literature and used only for concepts that occur more than twice.

# 1 Introduction

Atherosclerosis (Weber, Noels 2011, Libby P, Ridker PM et al. 2011) is considered to be a chronic inflammatory disease of the artery wall. Complications of atherosclerosis, such as myocardial infarction (MI) or ischemic stroke, are the major causes of disability and death in the modern world. According to World Health Organization (WHO) cardiovascular diseases (CVD) caused approximately 17.3 million deaths, which are 30 per cent of all deaths, worldwide in the year 2008. Coronary artery disease (CAD) caused 7.3 million and stroke 6.2 million of these deaths (Figure 1). WHO predicts that the burden of death by cardiovascular disease is increasing, and that in the year 2030 23 million people will die of cardiovascular disease yearly.

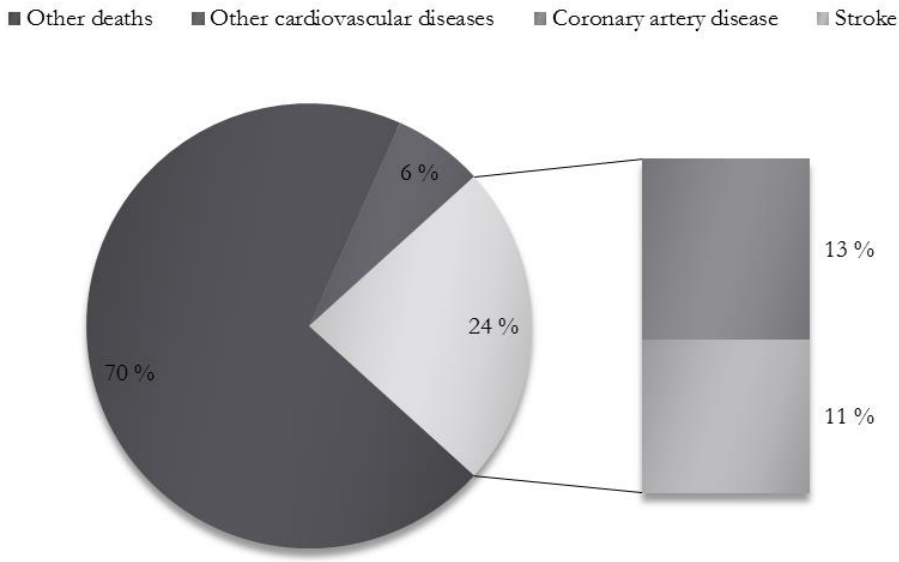



**Figure 1** – Deaths worldwide in the year 2008 (WHO 2012)

In atherosclerotic process (Libby P, Ridker PM et al. 2011), atherogenesis, immunological mechanisms interact with metabolic risk factors in the initiation of the disease process, as well as, in the disease progression and lesion activation (Kovanen, Kaartinen et al. 1995). Rupture of an atherosclerotic lesion, atheroma, leads to clinical manifestations such as MI and ischemic stroke (Hansson 2005). There are multiple hypotheses of the pathophysiology of the atheroma formation, however, the definitive evidence of processes such as lipoprotein oxidation, inflammation and immunity having crucial role in human atherosclerosis is still lacking (Libby P, Ridker PM et al. 2011).

One key event in atherogenesis is considered to be the conversion of low-density lipoprotein (LDL) to oxidized LDL (oxLDL). It is thought that this conversion increases the atherogenic potential of LDL and leads to fatty streak formation (Steinberg 1997). It has been shown in mice, that the removal of circulating oxLDL prevents atherosclerosis (Ishigaki, Katagiri et al. 2008). Moreover, systemic inflammatory state, hemodynamic conditions in different parts of the arterial tree and the dysfunction of the endothelium of arterial wall are considered important factors in atherogenesis (Gimbrone, Garcia-Cardena 2013). Recently, Histone deacetylase 9 (HDAC9) (International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2) et al. 2012) has been associated with the progression of atherosclerosis and with ischemic stroke in genome-wide association studies (GWAS).

When the human genome was discovered (Sachidanandam, Weissman et al. 2001), the era of intense human genetic studying began. The discovery of haplotypes (Gabriel, Schaffner et al. 2002), the correlation between nearby genetic variation, in humans allowed the development of HapMap database (International HapMap Consortium 2005). At the same time, high throughput genotyping methods (Hoheisel 2006) to cost-effectively genotype single nucleotide polymorphisms (SNP) and statistical methods to analyze the data (Marchini, Howie 2010) were developed. In 2005, the first GWAS on age-related macular degeneration was performed (Klein, Zeiss et al. 2005). Since then there have been publications of hundreds of GWASes (Manolio 2010). Atherosclerosis related GWASes have found multiple novel loci on intima-media thickness (IMT) (Bis, Kavousi et al. 2011), on CAD (CARDIoGRAMplusC4D Consortium, Deloukas et al. 2013), MI (Myocardial Infarction Genetics Consortium, Kathiresan et al. 2009). GWASes have not fulfilled the early high expectations (Manolio, Collins et al. 2009), however, they have given a huge amount of new data on

common diseases, and the next important step is to find the clinically significant information from this huge mass of data (Hirschhorn, Gajdos 2011).

Since oxLDL is considered to be a key factor in atherosclerosis, and little is known of the genetic factors affecting the susceptibility of LDL to oxidation, in this thesis a GWAS was performed on oxLDL and the impact of the found single nucleotide polymorphisms (SNP) was tested on the clinical manifestation of atherosclerosis; CAD, MI, and ischemic stroke. Furthermore, an age-of-onset informed GWAS was done on ischemic stroke, and novel stroke gene MMP12 was discovered. Recent GWASes have also found HDAC9 as possible novel gene acting in atherosclerotic lesions and in ischemic stroke. The role of HDAC9 and MMP12 is not clear and these novel ischemic stroke genes were further studied in human atherosclerotic plaques.

# 2 Review of the Literature

## 2.1 Atherosclerosis

### 2.1.1 Pathogenesis

#### 2.1.1.1 Imbalanced Lipid Metabolism and a Maladaptive Immune Response Lead to Chronic Inflammation of the Arterial Wall

The development of ahteromatous plaques in the inner lining of the arteries is called atherogenesis (Libby P, Ridker PM et al. 2011). The atherosclerotic lesions, atheromata, are asymmetric focal thickenings in innermost layer of the artery called intima (Hansson 2005). The atheromata consist of blood-borne immunological cells, connective-tissue elements, lipids, and derbis (Stary, Chandler et al. 1995).

Based on animal experiments and observation in humans, atherogenesis begins as qualitative change in the monolayer of endothelial cells that line the inner arterial surface (Figure 2). The first step in atherogenesis is considered to be the formation of fatty streaks. Fatty streaks are prevalent already in the young; however, they do not proceed to atheromata in all people (Stary, Chandler et al. 1994). It has been shown in animals and humans that hypercholesterolemia causes a focal activation of endothelium in large and medium sized arteries (Hansson 2005).

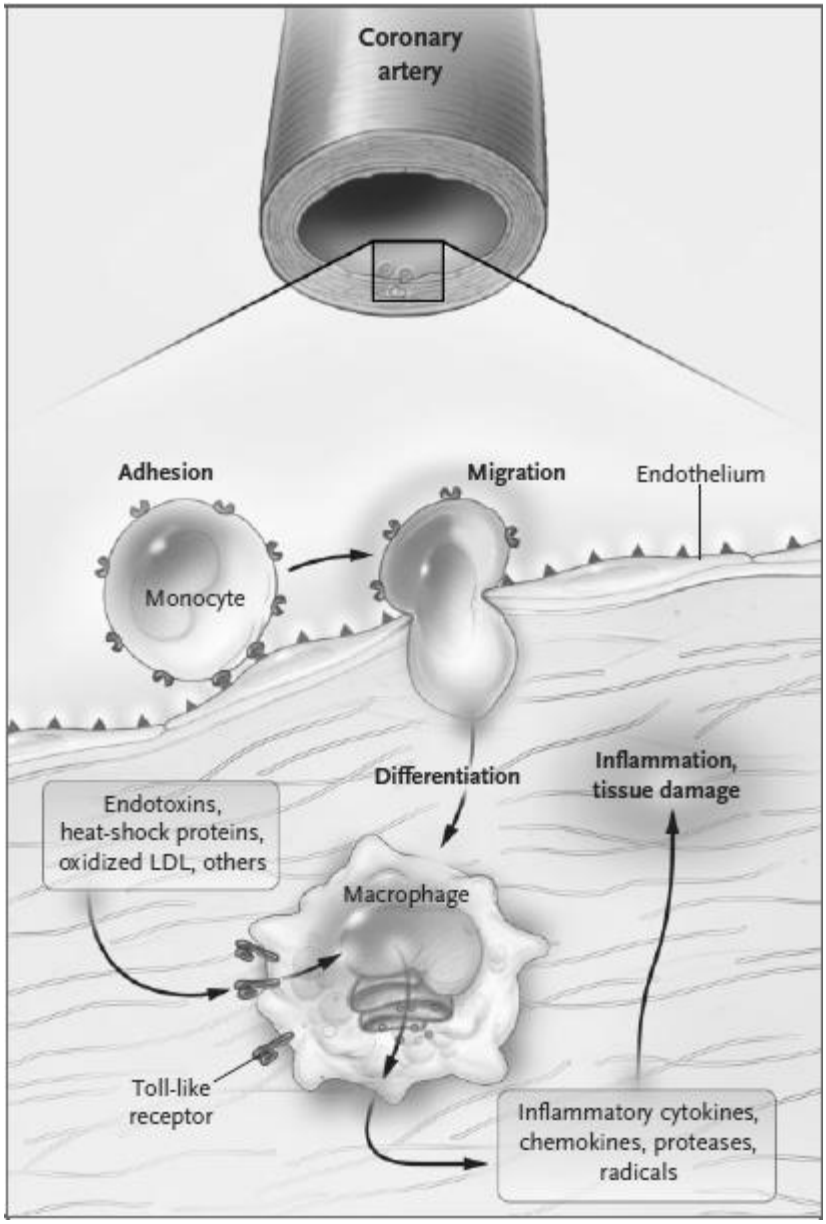



**Figure 2** – Hypercholesterolemia causes a focal activation of the endothelium in large and medium arteries, and the infiltration and retention of low-density lipoprotein (LDL) in the intima initiates an inflammatory response. Reproduced with permission from (Hansson 2005), Copyright Massachusetts Medical Society.

Lipids are insoluble, and therefore need to be transported through the circulation in complexes with proteins (Lusis, Pajukanta 2008). Cholesterol, which is acquired from food (and transported by chylomicrons to the liver) or produced by liver itself, is packed to very-low-density lipoprotein (VLDL) particles formed in the liver (Figure 3) (Yazdanyar, Jiang 2012). Furthermore, apoB particle is formed and incorporated in the VLDL particle. The best known function of apoB is to act as ligand for LDL receptors in various cells (Ooi, Russell et al. 2012). Apolipoprotein E (apoE) is also an important factor in lipoprotein metabolism (Kervinen, Kaprio et al. 1998).

VLDL is released to bloodstream to transport cholesterol and various other substances to cells that require those (Lusis, Pajukanta 2008). After interacting with HDL, or releasing some of the contents to tissues, VLDL gets denser and is called intermediate-density lipoprotein (IDL) (Figure 3). After more of the contents are released the particle gets denser and is called low-density lipoprotein (LDL). In each of the lipoprotein particles one apoB moiety is found (Ooi, Russell et al. 2012).

In normal operation of lipoprotein metabolism LDL is transported by arteries to various tissues that require cholesterol and other contents of the particle for their function (Lusis, Pajukanta 2008). After releasing the contents the particle travels back to the liver where it is incorporated to hepatocytes by LDL-receptor and degraded (Yazdanyar, Jiang 2012). Liver creates new VLDL-particles and the process starts again from the beginning.

In hypercholesterolemia, low-density lipoprotein (LDL) infiltrates the intima, and is retained there, causing inflammatory response in the artery wall (Skalen, Gustafsson et al. 2002). LDL gets modified through oxidation and enzymatic attack in the intima, which leads to release of phospholipids, which activate endothelial cells (Leitinger 2003). The activation of the endothelial cells happens preferentially at the sites of hemodynamic strain, such as arterial branches (Nakashima, Raines et al. 1998). The increased shear stress increases the expression of adhesion molecules and inflammatory genes by the endothelial cells (Dai, Kaazempur-Mofrad et al. 2004). Therefore, it is thought that the combined shear stress and accumulation of lipids begins the inflammatory response in the artery wall (Weber, Noels 2011).

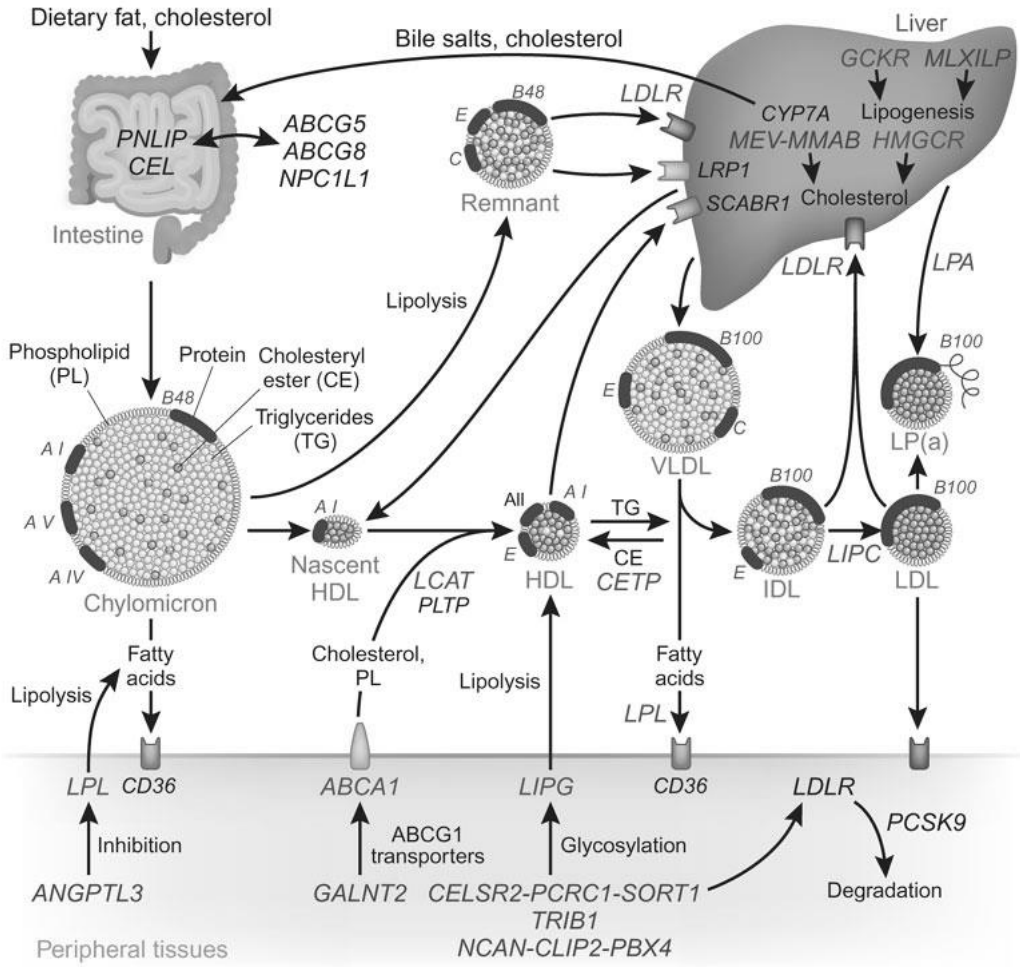



**Figure 3** – Disturbed lipoprotein metabolism is an important factor in atherogenesis. The figure shows lipoproteins and genes currently known to be involved in human lipoprotein metabolism. Reprinted by permission from Macmillan Publishers Ltd: Nature Genetics (Lusis, Pajukanta 2008), copyright 2008.

Platelets are first cells to get in contact with the activated endothelial cells (Badimon, Vilahur 2014). Glycoproteins Ib and IIb/IIIa engage the surface proteins of the endothelial cells (Massberg, Brand et al. 2002). This is thought to contribute to the activation of the endothelia. In hypercholesterolemic mice, inhibition of platelet adhesion reduced leukocyte infiltration and atherosclerosis progression (Massberg, Brand et al. 2002).

The activation of the endothelial cells causes them to express several types of leukocyte adhesion molecules, such as inter-cellular adhesion molecule 1 (ICAM-1) and vascular-cell adhesion molecule 1 (VCAM-1) (Galkina, Ley 2007). For example, VCAM-1 is upregulated in response to hypercholesterolemia (Cybulsky, Gimbrone 1991). Immunological cells, such as monocytes and lymphocytes, carry counter receptors for VCAM-1 and adhere to sites of upregulated VCAM-1 expression (Galkina, Ley 2007). Monocytes tether and roll along the vascular surface and adhere at the site of the activated endothelium (Woollard, Geissmann 2010). After attachment to VCAM-1 intima produces chemokines which stimulate the blood cells to migrate through the endothelial junctions into the subendothelial space (Figure 2) in movement called transmigration and diapedesis. It has been shown in mice, that deletion of the adhesion molecule genes or pharmaceutical blocking of certain chemokines and adhesion molecules for the mononuclear cells inhibit atherosclerosis (Lesnik, Haskell et al. 2003, Lutters, Leeuwenburgh et al. 2004).

In the sub-endothelial space macrophage colony-stimulating factor induces monocytes to differentiate into macrophages (Smith, Trogan et al. 1995). This is a critical step in the process of atherosclerosis. In this process pattern-recognition receptors for innate immunity, such as scavenger receptors and toll-like receptors are upregulated (Peiser, Mukhopadhyay et al. 2002, Janeway, Medzhitov 2002). Elevated levels of circulating cholesterol transported by apolipoprotein-B (apoB) containing LDL get stuck in the intima as apoB binds to negatively charged extracellular matrix proteoglycans (Williams, Tabas 1995) and oxidizes to oxLDL (Sanchez-Quesada, Villegas et al. 2012). Normally LDL is internalized to cells by so called Brown-Goldstein LDL-receptor (Brown, Goldstein 1983). Within this process is a mechanism that controls the internalization so that cells cannot get overfilled with LDL. However, as LDL gets modified in the intima, it loses its typical form and is called oxidized LDL (oxLDL) (Ishigaki, Oka et al. 2009).

Scavenger receptors internalize a broad range of molecules and particles which have pathogen-like molecular patterns, for example, bacterial endotoxins, apoptotic cell fragments (Peiser, Mukhopadhyay et al. 2002). Also, oxLDL is taken up and destroyed through this pathway (Woollard, Geissmann 2010). As there is no negative feedback mechanism, cholesterol gets accumulated in the macrophages as more and more oxLDL is internalized by macrophages (Park 2014). There is some evidence in rabbits that cells could protect themselves from excessive uptake of oxLDL in advanced atherosclerotic lesions by generating scavenger receptors that cannot bind oxLDL (Hiltunen, Gough et al. 2001). Cholesterol forms cytosolic

droplets which eventually transform the cell to foam cell which is filled with cholesterol (Stary, Chandler et al. 1994). The foam cell is prototypical cell in atherosclerosis, and is the basis of for example the fatty streaks seen already in the young (Stary, Chandler et al. 1994). Finally the foam cells start to die apoptotically (Seimon, Nadolski et al. 2010). Cholesterol and other substances form a necrotic core to the intima (Lusis 2012). As this material is foreign to the vessel wall, protective measures are initiated. Smooth muscle cells (SMC) from the outer parts of the artery (adventitia) travel to the scene and start to form a fibrous cap around the foreign material to sequestrate it from the environment (Badimon, Vilahur 2014).

Toll-like receptors (TLR) also bind pathogen-like molecular patterns (Figure 4). In contrast to scavenger receptors, they initiate a signal cascade which leads to cell activation (Janeway, Medzhitov 2002). The activated macrophages produce inflammatory cytokines, proteases, cytotoxic oxygen, and nitrogen radical molecules (Janeway, Medzhitov 2002). Also, dendritic cells, mast cells, and endothelial cells express toll-like receptors and produce similar effects (Bobryshev, Lord 1995). It is thought that plaque inflammation is partly dependent on the toll-like receptor pathway (Weber, Noels 2011). Macrophages can be divided to M1 and M2 classes (Salagianni, Galani et al. 2012). Inflammatory M1 macrophages and M1-associated cytokines are considered to be involved in the development of the vulnerable plaques, whereas M2 macrophages are considered to be protective through paracrine anti-inflammatory effect which they exert on M1 macrophages (Salagianni, Galani et al. 2012). Recently, the loss of macrophage nuclear factor E2-related factor 2 (Nrf2) has been shown to protect against atherogenesis (Ruotsalainen, Inkala et al. 2013).

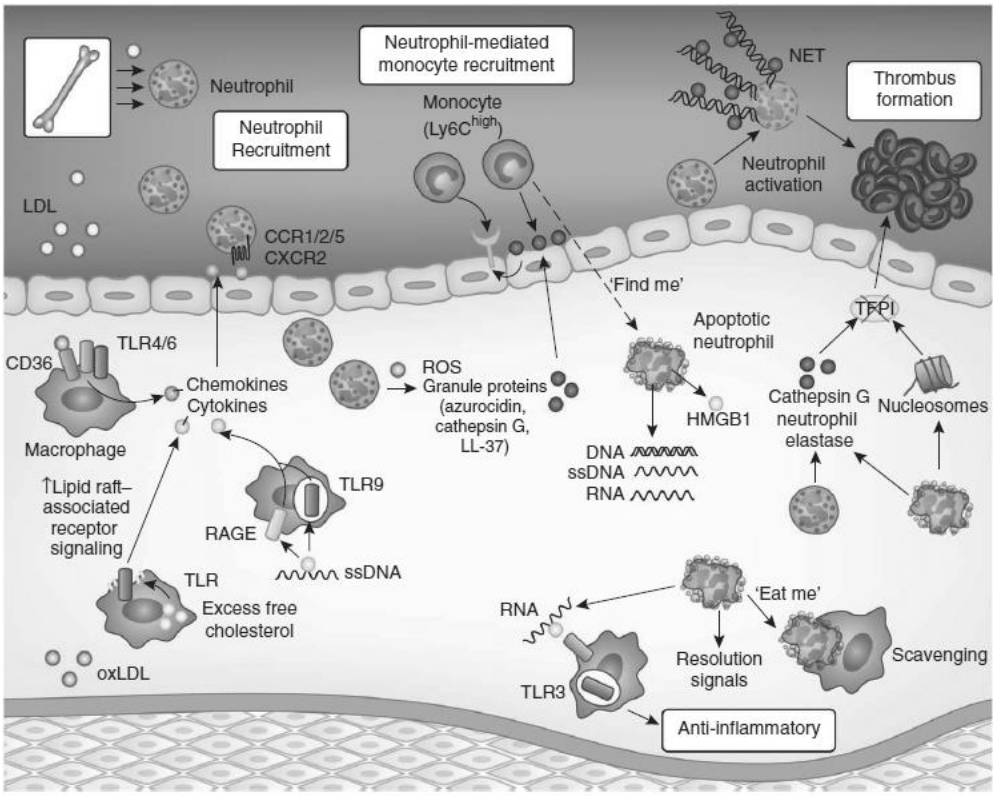



**Figure 4** – The complex role of neutrophils in the atherosclerotic process. Reprinted by permission from Macmillan Publishers Ltd: Nature Medicine (Weber, Noels 2011), copyright 2011.

Moreover, immunological signaling by neutrophils plays an important role in atherogenesis (Figure 4). As neutrophils are present in the inflamed intima, they sustain monocyte recruitment through various find-me and eat-me signals (Soehnlein, Lindbom 2010). As neutrophils are activated, neutrophil protease-mediated proteolysis of tissue pathway inhibitor (Massberg, Grahl et al. 2010) could promote atheroprogression and thrombus growth. Neutrophil extracellular trap (NET) formation upon neutrophil activation (Papayannopoulos, Zychlinsky 2009) and tissue factor pathway inhibitor proteolysis by neutrophil proteases (Massberg, Grahl et al. 2010) could promote the progression of atherosclerosis and thrombus growth. Even though neutrophils can provide resolution signals (Soehnlein, Lindbom 2010) that can trigger antiatherogenic TLR3 signaling (Cole, Navin et al. 2011), they can also provide a chronic inflammatory trigger that sustains atherogenesis. It is not known which factor cause the inflammatory

triggering in chronic atherosclerosis. Without challenge from pathogens, the continued presence of neutrophils in advanced plaques may contribute to large-vessel thrombosis as a trigger for MI and stroke (Weber, Noels 2011).

#### 2.1.1.2 Evolution of the Rupture Prone Plaque

AHA classification of atherosclerotic plaques (Figure 5) divides plaques histologically to II) presence of foam cells in the arterial wall, III) preatheroma, IV) atheroma, V) fibroatheroma and VI) complicated lesion. Fibrous cap composed of collagen and SMCs covers the fibroatheromatous plaque (Badimon, Vilahur 2014). The composition of the fibrous cap is known to be essential in determining how dangerous atherosclerosis is to an individual (Lusis 2012). Thin and inflamed cap is prone to rupture, which manifests as thromboembolic events such as MI or ischemic stroke (Hansson 2005). Fibrous cap is located between the vascular lumen and the necrotic core (Badimon, Vilahur 2014). Autopsy studies have determined that ruptured plaques are extremely thin (<65 micrometers thick), have a low collagen content and have a macrophage density of around 26% (Falk, Nakano et al. 2013).

Complication of atheroma occurs when foam cells release cytokines and growth factors to stimulate vascular smooth muscle cells (VSMC) to migrate from media to intima (Badimon, Vilahur 2014). VSMCs divide and produce extracellular matrix (ECM) components that contribute to the fibrous cap development (Koga, Aikawa 2012). Many of the foam cells undergo apoptosis at early stages of atherosclerosis development and are removed by M2 macrophages in efferocytosis (Tabas 2010). Macrophage death leads to release of lipids, pro-inflammatory and pro-thrombotic mediators (tissue factors) and metalloproteinases (MMP) (Tabas 2010). MMPs digest the ECM scaffold, including the fibrous cap, and this makes the plaques more susceptible to rupture (Lin, Kakkar et al. 2014). Vulnerable necrotic core of the plaque is characterized by lack of supporting collagen determined by fewer VSMCs which are the main source of collagen production (Tabas 2010). Moreover, presence of hemorrhage has been shown to enlarge the necrotic core (Teng, Sadat et al. 2014).

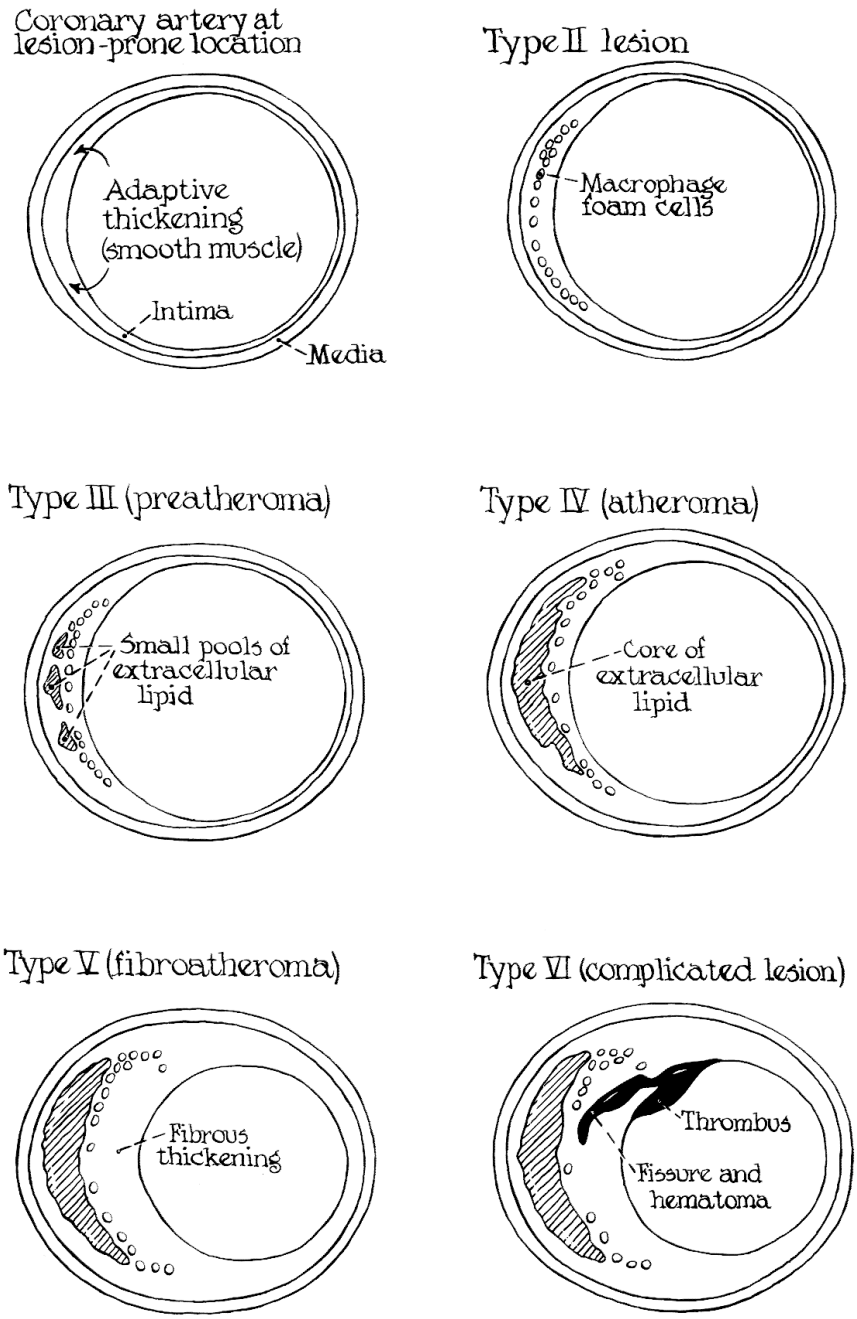



**Figure 5** – American Heart Association (AHA) classification of atheromata. Reprinted by permission from Wolters Kluwer Health: Circulation (Stary, Chandler et al. 1995) copyright 1995.

The distribution of lipid core seems to be essential for plaque instability (Badimon, Vilahur 2014). In vulnerable plaque there is accumulation of free cholesterol in the center and presence of low free-to-esterified cholesterol ratio at the edges (Felton, Crook et al. 1997). Eccentric distribution of the lipid core leads to rearrangement of circumferential stress to the shoulder regions of the plaque, and increases the vulnerability of these sites to rupture (Tabas 2010). Almost 60 per cent of fibrous cap fissures occur in this region (Loree, Kamm et al. 1992, Richardson, Davies et al. 1989). However, in one interesting study of sudden cardiac death, rupture of the plaque occurred in the mid-portion in those persons who were performing intense physical exercise whereas rupture occurred in the shoulder region in those who died at rest (Burke, Farb et al. 1999).

Atherosclerotic lesions with a potential to rupture, i.e. vulnerable plaques, are the primary cause of clinical episodes of atherosclerosis related diseases. One possible therapy for atherosclerosis could be towards stabilization of the plaques for example through apoptosis modulation (Beohar, Flaherty et al. 2004). Interestingly, oxLDL-binding protein has been shown to stabilize plaques in mice (Zeibig, Li et al. 2011).

#### 2.1.1.3 Oxidized Low-Density Lipoprotein

OxLDL is considered to be an important factor in the atherosclerotic process (Goldstein, Ho et al. 1979, Ishigaki, Oka et al. 2009). Oxidative stress modifies the LDL particle stuck in the intima, converting it to oxLDL, which then is incorporated by the scavenger receptors of macrophages (Libby P, Ridker PM et al. 2011). Lipoproteins are multi-molecular by nature (Lusis, Pajukanta 2008). They are formed by (i) a lipid core which contains fat soluble substances such as cholesterol esters, triglycerides, vitamins, etc., (ii) by phospholipid bi-layer which forms the amphipathic surface of the particle separating the hydrophobic core from the hydrophilic environment, and (iii) by apolipoproteins which e.g. guide the particles to right places by acting as ligands for cell surface receptors (Kumar, Butcher et al. 2011).

Due to the multi-molecular nature of LDL, oxLDL can be defined in multiple ways (Brinkley, Nicklas et al. 2009), and therefore multiple different methods exist for its measurement (Itabe, Ueda 2007). The different compartments can oxidize together or separately, and many metabolites that indicate LDL oxidation are also formed (Lehtimäki, Lehtinen et al. 1999).

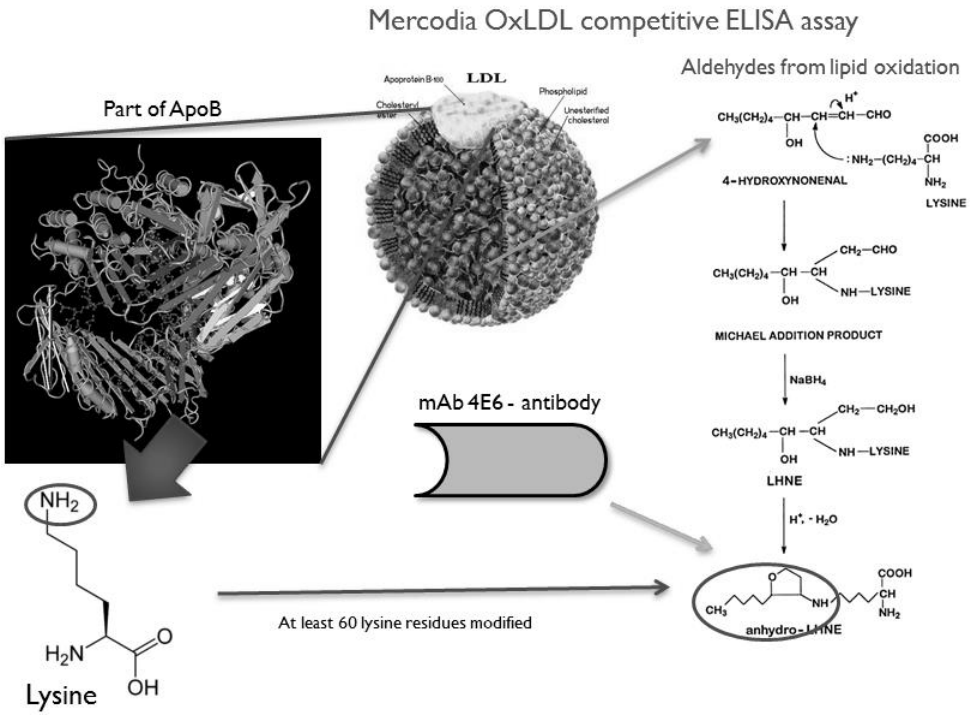



**Figure 6** – The principle of the Mercodia oxLDL assay and the basic structure of a lipoprotein particle.

One of the most widely used method is the Mercodia oxLDL assay (Holvoet, Stassen et al. 1998). The principle of the assay is shown in Figure 6. It measures the modification of the apoB moiety (Segrest, Jones et al. 2001) of LDL. When the LDL particle lies in the hostile environment of arterial intima the lysine residues of apoB get substituted by aldehydes which are recognized by the monoclonal 4E6 antibody of this enzyme linked immunesorbent assay (ELISA) (Holvoet, Stassen et al. 1998).

Of the other methods, e.g. LDL diene conjugation measures the oxidation of the lipid compartment of the LDL particle (Ahotupa, Marniemi et al. 1998). LDL baseline diene conjugation can be measured by determining the level of baseline diene conjugation in lipids extracted from LDL. First, serum LDL is isolated by means of precipitation with buffered heparin. Then lipids are extracted from LDL samples with chloroform-methanol, dried under nitrogen, then redissolved in cyclohexane and analyzed spectrophotometrically at 234 nm (Ahotupa, Marniemi et al. 1998).

Moreover, oxLDL can be measured by for instance antibodies against malondialdehyde-modified LDL (MDA-LDL), and against copper oxLDL (Toshima, Hasegawa et al. 2000). It is extremely important to consider the limitations of the measurement method when interpreting the results.

The Mercodia-assay has been used in multiple studies; however, the results are inconclusive. Some studies show that oxLDL predicts cardiac syndromes (Meisinger, Baumert et al. 2005, Shimada, Mokuno et al. 2004, Tsimikas 2006), coronary artery disease (CAD) severity (Uzun, Zengin et al. 2004) plaque instability (Nishi, Itabe et al. 2002), cerebral infarction (Uno, Kitazato et al. 2003), and restenosis after myocardial infarction (Naruko, Ueda et al. 2006). However, some larger studies have been unable to replicate these results (Ishigaki, Oka et al. 2009). Interestingly the removal of circulating oxLDL has been shown to prevent atherosclerosis in mice (Ishigaki, Katagiri et al. 2008).

#### 2.1.1.4 Histone Deacetylase 9 (HDAC9)

HDAC9 acts as an epigenetic gene expression regulator via deacetylation of previously acetylated histone proteins, and thus modifies the interchanges between relaxed and closed chromatin genes (Arrowsmith, Bountra et al. 2012). Although known as histone deacetylases, these proteins also act on other substrates and lead to both up and down regulation of genes (Haberland, Montgomery et al. 2009). Acetylation is a widespread modification in cell proteins, and in addition to histones, HDACs can deacetylate other proteins as well. There are in total 18 HDACs which are encoded by distinct genes (McKinsey 2011). They are grouped into four classes on the basis of similarity to yeast transcriptional repressors (McKinsey 2011). HDAC9 is a member of the Class IIa HDACs. The class IIa HDACs have been shown to interact with members of the myocyte enhancer factor-2 (MEF2) transcription factor family which are regulators of VSMC proliferation (McKinsey 2011).

Variation in the introns of HDAC9 gene has been associated with the risk of ischemic stroke (International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2) et al. 2012). It has recently been shown that HDAC9 represses cholesterol efflux and alternatively activated macrophages in the development of atherosclerosis (Cao, Rong et al. 2014, Azghandi, Prell et al. 2014).

## 2.1.2 Clinical Manifestations of Atherosclerosis

### 2.1.2.1 Coronary Artery Disease and Myocardial Infarction

Atherosclerosis in the coronary arteries can lead to CAD (Hansson 2005). The classical risk factors for CAD are hypercholesterolemia, smoking, hypertension, diabetes, and old age. Other risk factors include low high density lipoprotein (HDL) concentration, overweight, low exercise, infections and conditions affecting blood coagulation (Hansson 2005).

First clinical symptom of CAD can be angina pectoris, chest pain in exercise (Natarajan 2002). The pain is caused by the occlusion of blood flow to the heart muscle by large atheromatous plaque protruding to lumen of the artery (Nabel, Braunwald 2012). Other symptoms include shortness of breath in exercise and tiredness after exercise. Angina pectoris pain starts slowly when starting to exercise and worsens as exercise continues. The pain is felt in the middle of sternum as wide squeezing sensation. Patient can also feel discomfort in the neck, jaw, shoulder, back or arm. Typical angina pectoris pain stops in rest or after administration of nitroglycerin. Exercise test can be used to diagnose angina pectoris and the severity of stenosis.

Acute coronary syndrome and MI are caused by a rupture of atherosclerotic plaque (Nabel, Braunwald 2012). The unstable plaques with thin fibrous caps are most prone for rupture. Intramural plaques do not cause typical angina pectoris pain and MI or even sudden cardiac death can be the first sign of atherosclerosis (Nabel, Braunwald 2012). The rupture leads to formation of blood clot and the total or partial occlusion of blood flow to the coronary arteries leading to hypoxia of cardiac muscle. The basis of diagnosis includes anamnesis, clinical findings and changes in electrocardiography (ECG). Acute coronary syndrome can be divided to unstable angina pectoris (UAP), MI without ST-elevation (NSTEMI) and MI with ST-elevation (STEMI). Elevation of cardiac muscle enzyme troponin confirms the diagnosis (Nabel, Braunwald 2012).

In Finland, 17 000 patients are treated in hospital yearly due to acute coronary syndrome, and 6 000 people die yearly because of CAD at home or en route to hospital (Current Care Guideline 2014).

#### 2.1.2.2 Ischemic Stroke

Stroke is a clinical condition where disturbance in the blood supply to the brain causes a loss of brain function (Chamorro, Meisel et al. 2012, van der Worp, van Gijn 2007). The clinical presentation depends on the region suffering from loss of blood flow. The same clinical symptoms can be caused by hemorrhage or by clotting of artery by thrombosis or embolism. Therefore, computed tomography (CT) scan is needed to differentiate the underlying cause which is essential in treatment selection (van der Worp, van Gijn 2007).

Ischemic stroke can be classified into five subtypes according to Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria (Adams, Bendixen et al. 1993); 1) large-artery atherosclerosis (LAA), 2) small-vessel disease (SVD), 3) cardioembolic stroke (CE), 4) other aetiology, or 5) unknown aetiology. Transient ischemic attack (TIA) can precede ischemic stroke. Risk factors especially for LAA caused by atherosclerosis are mostly the same as for CAD. CE risk factors include atrial fibrillation, MI, heart failure, mitral valve prolapse, endocarditis, heart myxoma and artificial heart valve (van der Worp, van Gijn 2007). Of elderly patients over 80 years of age one quarter of ischemic stroke is caused by atrial fibrillation. In the young atherosclerosis is rarely the cause. The most important causes in the young are carotid artery dissection or prothrombotic condition especially in patients with patent foramen ovale.

The most common clinical representation of ischemic stroke is caused by occlusion of the middle cerebral artery (MCA) which is a branch of carotid artery (van der Worp, van Gijn 2007). The occlusion of MCA causes sudden hemiplegia and loss of speech, aphasia. If the occlusion is in the vertebrobasillar region, the typical clinical presentation is sudden vertigo, nausea, diplopia and dysphagia.

First line treatment is executed in organized inpatient care, stroke unit. Stroke patients treated in stroke unit are more likely to be alive, independent, and living at home one year after stroke than patients receiving less organized care (Stroke Unit Trialists' Collaboration 2013). Thrombolysis by recombinant tissue plasminogen activator (tPA) within 4.5 hours of symptom onset reduces the proportion of dead and dependent people (Wardlaw, Murray et al. 2014). Stent retriever thrombectomy within 8 hours reduces the severity of post-stroke disability and increases the rate of functional independence in patients with anterior circulation stroke (Jovin, Chamorro et al. 2015). Other first line treatment in stroke unit include monitoring of blood glucose levels, body temperature, blood pressure, cerebral edema, heart arrhythmia, and prevention of pneumonia and deep vein thrombosis (van der Worp, van Gijn 2007).

Early initiation of rehabilitation and rehabilitation in a multidisciplinary setting after stroke improve functional outcome (Cifu, Stewart 1999). Secondary prevention of atherothrombotic stroke is executed by aspirin-dipyridamole or clopidogrel which are equally effective (Sudlow, Mason et al. 2009). Treatment of hypertension is essential. Statins have been shown to be effective independent of hypercholesterolemia (Manktelow, Potter 2009). Carotid endarterectomy in stroke or TIA patients with highly occluded carotid artery reduces the risk of ipsilateral ischemic stroke (Rerkasem, Rothwell 2011). The same life-style changes such as smoking cessation, weight loss, moderate alcohol consumption and exercise are beneficial in prevention of ischemic stroke and CAD (van der Worp, van Gijn 2007).

Atherosclerosis is considered to be a systemic condition. The pathogeneses of carotid and coronary atherosclerosis are mostly concordant, however, there are some subtle differences (Jashari, Ibrahimi et al. 2013). For example, artery-to-artery embolization from carotid plaque is more frequent cause of ischemic stroke than embolization from coronary plaque is as cause of MI. Moreover, cholesterol is considered to be more important risk factor in CAD than in stroke, and hypertension is more important risk factor of ischemic stroke (Kannel, Wolf 2006).

## 2.2 Genome-Wide Association Studies of Atherosclerosis Related Diseases

### 2.2.1 Common Variation in the Human Genome

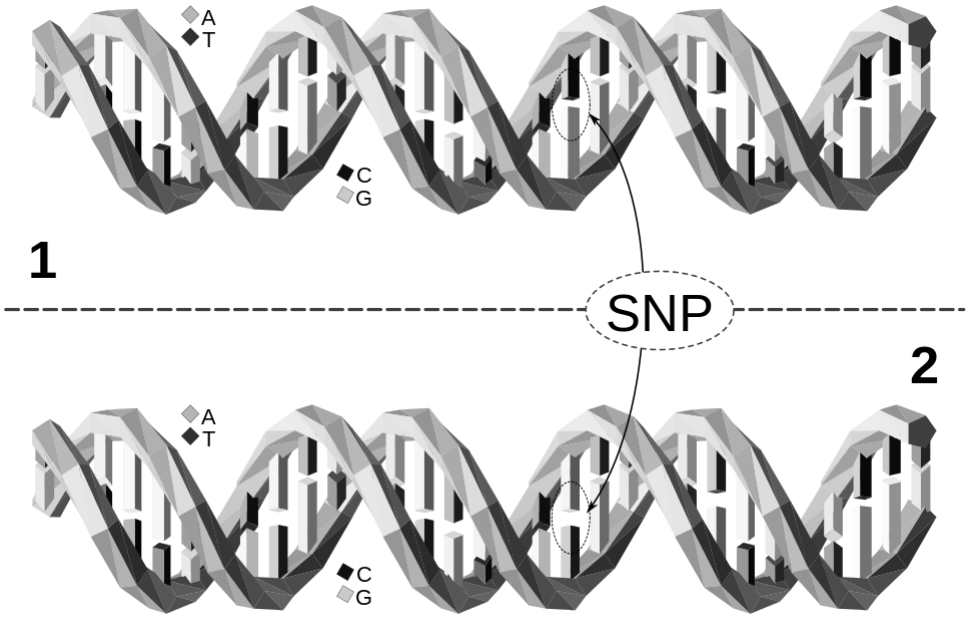



**Figure 7** – Deoxyribonucleic acid (DNA) carries genetic information through generations. Single nucleotide polymorphisms (SNP) are one cause of phenotype variation between individuals.

Since the discovery of human genome (Sachidanandam, Weissman et al. 2001) the understanding of human genetics has taken huge leaps. In 2002 correlation between nearby genetic variation was studied (Gabriel, Schaffner et al. 2002) and the HapMap database was developed in 2005 (International HapMap Consortium 2005). The first idea of genotyping microarray came about in 1986 (Poustka, Pohl et al. 1986). Since then microarray technology has taken huge advances (Hoheisel 2006) and cheap and rapid genotyping of multiple single nucleotide polymorphisms (SNP) has become available to a wide audience. Moreover, methods in bioinformatics were developed (Marchini, Howie 2010) (Marchini, Howie 2010)

which allowed the first genome-wide association study to be performed (Klein, Zeiss et al. 2005).

The theory behind GWAS is the so called common disease – common variant hypothesis (Manolio, Brooks et al. 2008). It states that complex diseases are predisposed by carrying many varied alleles with small effect, with combined large effect, ultimately manifesting as e.g. coronary artery disease. In GWAS the idea is to harvest these areas in studies with thousands of individuals to find all these small variations in the genome giving higher risk to the studied disease.

The main critique towards GWAS (McClellan, King 2010) comes from evolutionary perspective: the rare and harmful mutations have been removed during many generations (Barreiro, Laval et al. 2008). The majority of variation in the human genome is quite recent. Moreover, the majority of GWAS findings are on so called gene-deserts where there is no known mechanism of function. Furthermore, because the inability to find plausible biological explanation to the associations it has been proposed that majority of the finding could be spurious mostly due to unaccounted population stratification (McClellan, King 2010).

However, GWASes are planned to that they tag the most probable areas with association to the studied phenotype (Wang, Bucan et al. 2010, Klein, Xu et al. 2010). This is built-in in the technique because it takes advantage from the linkage disequilibrium (LD) of the human genome. It was discovered in the early 21st century that the genome is most likely structured so that large LD blocks are passed down in generations (Gabriel, Schaffner et al. 2002). This means that nearby single nucleotide polymorphisms (SNPs) are highly correlated with each other. The genotyping arrays were therefore designed so that they capture the variation in the genome by taking one tag from each of these blocks. This way it is not possible to find the exact spot behind the association, merely the most probable area. Hence, further studies are required to find the mechanism behind the associations.

To explain why most associations are found on gene-deserts, there could be a still unknown mechanism behind these associations (Wang, Bucan et al. 2010, Klein, Xu et al. 2010). Moreover, more gene-gene and gene-environment interaction studies will be needed to determine where the missing heritability lies (Zuk, Hechter et al. 2012).

GWASes have so far been a huge success story bringing about huge collaborative projects worldwide to illuminate the largely unknown mechanisms behind complex diseases. The major challenge is translating this knowledge to clinical practice (Fugger, McVean et al. 2012).

Hundreds of loci are known to be involved in lipoprotein metabolism (Lusis, Pajukanta 2008) and pathogenesis of atherosclerosis (Lusis, Fogelman et al. 2004). Huge GWASes have been done to detect multiple new loci in CAD (CARDIoGRAMplusC4D Consortium, Deloukas et al. 2013), and lipid metabolism (Teslovich, Musunuru et al. 2010, Willer, Sanna et al. 2008). Moreover, usage of next generation sequencing in the near future will bring about huge amounts of additional information (Shendure, Ji 2008).

The main motivation for GWAS studies is to find intervention targets in the tagged regions (Fugger, McVean et al. 2012). Moreover, GWAS results could be utilized in genetic testing (Grosse, Khoury 2006). The GWAS results have been tried to use in predicting disease susceptibility but this target has been elusive in practice (Ripatti, Tikkanen et al. 2010).

### 2.2.2 Coronary Artery Disease

The largest GWAS on CAD so far was published in 2013 (CARDIoGRAMplusC4D Consortium, Deloukas et al. 2013). In that study 63,746 CAD cases and 130,681 controls were analyzed. The study identified 15 loci reaching genome-wide significance. Now there are in total 46 susceptibility loci for CAD. These variants explain in total 10.6% of CAD heritability. 12 of the loci associate with lipid trait, 5 with blood pressure, however, none associate with diabetes. In interaction network analysis of 233 candidate genes four most significant pathways were linked to lipid metabolism and inflammation underscoring their causal role in etiology of CAD (CARDIoGRAMplusC4D Consortium, Deloukas et al. 2013).

### 2.2.3 Myocardial Infarction

The largest GWAS for MI was performed in 2009 (Myocardial Infarction Genetics Consortium, Kathiresan et al. 2009). In that study association of SNPs and copy number variants were associated with early onset MI in 2,967 cases and 3,075 controls. The results were replicated in an independent sample. SNPs at nine loci reached genome-wide significance: three were newly identified (21q22 near MRPS6-SLC5A3-KCNE2, 6p24 in PHACTR1 and 2q33 in WDR12) and six replicated prior observations (9p21, 1p13 near CELSR2-PSRC1-SORT1, 10q11

near CXCL12, 1q41 in MIA3, 19p13 near LDLR and 1p32 near PCSK9) (Myocardial Infarction Genetics Consortium, Kathiresan et al. 2009).

In a more recent study (Holmen, Zhang et al. 2014) using exome array of 80,137 coding variants in 5,643 Norwegians novel locus TM6F2 encoding p.Glu167Lys was found as causal variant for total cholesterol and myocardial infarction risk.

### 2.2.4 Ischemic Stroke

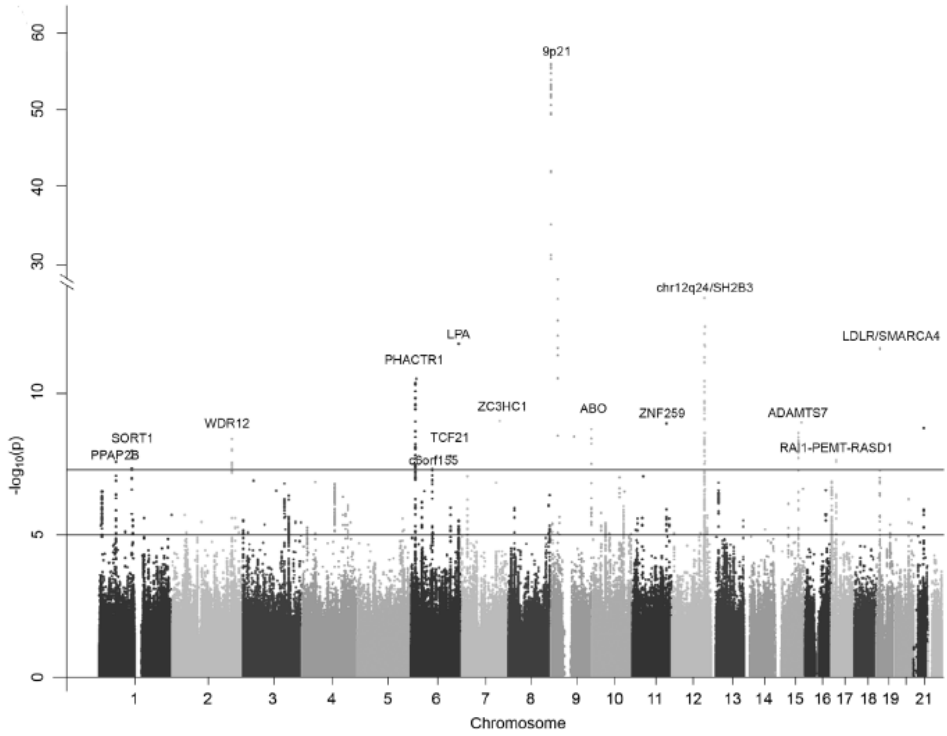



**Figure 8** – Shared loci for coronary artery disease and ischemic stroke. Reprinted by permission from Wolters Kluwer Health: Stroke (Dichgans, Malik et al. 2014), copyright 2014.

Before GWASes, proprotein convertase subtilisin/kexin type 9 (PCSK9) has been shown to associate with ischemic stroke (Abboud, Karhunen et al. 2007). The first

GWAS on ischemic stroke was done as late as in the year 2012 associating HDAC9 (International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2) et al. 2012) with LAA subtype of stroke. Collecting a sufficient sample size a challenge since stroke is a heterogenic condition with multiple etiologies. In a recent GWAS, there has been a novel association at 21q24.12 (Kilarski, Achterberg et al. 2014). These studies have also confirmed previously known atrial fibrillation genes in the etiology of CE subtype of stroke. Furthermore, shared loci between CAD and ischemic stroke have been studied (Figure 8). In that study a substantial overlap between the genetic risk of ischemic stroke (especially LAA) and CAD was found (Dichgans, Malik et al. 2014). Moreover, it was shown that HDAC9 variation associates with both ischemic stroke and CAD.

# 3 Aims of the Study

OxLDL is considered to be essential in the development of atherosclerosis. Genetic regulation for serum oxLDL levels has not been studied before. Genetic variation affecting oxLDL levels could have effect on the risk of CAD, MI or ischemic stroke. HDAC9 has recently been associated with atherosclerosis and ischemic stroke, and its role in atherosclerotic plaque is not clear. The major aims of the study were:

1) Determine possible genome-wide genetic basis for serum oxLDL levels regulation using GWAS (study **I**).

2) Study the association of the oxLDL concentration related functional genetic variant rs676210 causing a missense mutation Pro2739Leu in apoB with CAD and MI in clinical cohorts with coronary angiography patients (**I**).

3) Study the association of oxLDL concentration related functional genetic variant rs676210 causing a missense mutation Pro2739Leu in apolipoprotein B with cerebrovascular disease events and ischemic stroke (**II**).

4) Study the association of HDAC9-GWAS lead SNPs (rs11984041 and rs2107595) with occurrence of early markers of atherosclerosis i.e., asymptomatic carotid plaque and carotid intima-media thickness detected by carotid ultrasound (**III**).

5) Perform age-of-onset informed GWAS on ischemic stroke (**IV**)

6) Study HDAC9 and MMP12 protein/mRNA expression in human atherosclerotic plaques, taken from different clinically important blood vessels (**III-V**).

# 4 Materials and Methods

## 4.1 Clinical Cohorts and the Definition of their Clinical Characteristics

For all cohorts, the recruitment of patients was approved by the relevant local ethics committees, and studies were conducted in accordance with the Declaration of Helsinki. The capacity of all patients with cerebrovascular disease to give an informed consent was assessed by trained medical staff. In LURIC, all participants gave informed consent. No patient was recruited to LURIC study if they did not have the capacity to consent. In WTCCC2, all participants gave informed consent to participate. In WTCCC2 cases where patients had a compromised capacity to consent, consent was obtained from next of kin.

### 4.1.1 The Young Finns Study (YFS)

YFS is a longitudinal Finnish population sample for studying cardiovascular risk factors and the evolution cardiovascular diseases from childhood to adulthood (Raitakari, Juonala et al. 2008). The first cross-sectional study was done in the year 1980 at five centers (Tampere, Helsinki, Turku, Oulu, and Kuopio). 3,596 people were selected randomly from the national population register in age groups of 3, 6, 9, 12, 15, and 18. The subjects have been re-examined in 1983 and 1986 as youngsters and in 2001, 2007, and 2012 as adults. In this study the data from the year 2001 was used.

During the follow-up in 2001, a total of 2,283 participants aged 24–39 years were examined for numerous study variables, including serum lipoproteins, glucose, insulin, obesity indices, blood pressure, lifestyle factors, smoking status, alcohol use, and general health status (Raitakari, Juonala et al. 2008). Genotype and phenotype data for this study were available for 2,080 subjects.

### 4.1.2 The Ludwigshafen Risk and Cardiovascular Health (LURIC)

LURIC study consists of 3,316 Caucasian patients who were referred to coronary angiography due to chest pain at a tertiary care center in Southwest Germany between the years 1997 and 2000 (Winkelmann, Marz et al. 2001). The study aims to provide a well-defined resource for the study of environmental and genetic risk factors, and their interactions, and the study of functional relationships between gene variation and biochemical phenotype (functional genomics) or response to medication (pharmacogenomics). Long-term follow-up on clinical events will allow us to study the prognostic importance of common genetic variants (polymorphisms) and plasma biomarkers. All the necessary covariate and endpoint data were available for 2,912 LURIC patients. They formed the present study population.

### 4.1.3 Kooperative Gesundheitsforschung in der Region Augsburg (KORA)

The MONICA/KORA Augsburg study (KORA) is a series of population-based surveys conducted in the region of Augsburg in Southern Germany (Lowel, Doring et al. 2005). The data for the present study was drawn from a sub-cohort randomly selected by sex and survey from the KORA surveys S1–S3 conducted between 1984 and 1995 (Thorand, Schneider et al. 2005). Out of these, 1,326 subjects had all the required covariate and endpoint data available and were included in the present study.

### 4.1.4 The Finnish Cardiovascular Study (FINCAVAS)

The FINCAVAS population consists of patients who underwent an exercise stress test at Tampere University Hospital, Finland (Nieminen, Lehtinen et al. 2006). From the overall recruited study population, 1,118 individuals had all the necessary angiographic, genetic, and covariate data available and were included in the current study.

The exercise test indications were a diagnosis of CAD, a post-MI assessment, evaluation of drug therapy, arrhythmia, assessment of performance (working capacity), or an evaluation prior to surgery. The purpose of FINCAVAS is to construct a risk profile of individuals at high risk of cardiovascular diseases, events, and deaths. FINCAVAS has an extensive set of data on patient history, genetic

variation (the Metabochip), cardiovascular parameters, ECG markers, and follow-up data on clinical events, hospitalizations, and deaths. Of the patients included, 43.6% also underwent coronary angiography (Nieminen, Lehtinen et al. 2006).

### 4.1.5 Angiography and Genes Study (ANGES)

The ANGES population consisted of 1,000 patients with a symptomatic heart disease referred to coronary angiography to rule out or confirm CAD (Raitoharju, Seppala et al. 2011, Mennander, Kuukasjarvi et al. 2008). The population studied consists of 1,000 Finnish individuals participating in the ongoing ANGES study. Angiographic, genetic, and covariate data was available for 808 individuals (516 men and 292 women; mean age 62 ± 10). The data was collected between September 2002 and July 2005. All patients underwent coronary angiography at Tampere University Hospital due to clinically suspected coronary artery disease. The study is a cross-sectional study, and after the angiography, patients were treated according to the Finnish Current Care Guidelines. Patients were also interviewed by a study nurse, and a questionnaire was used to collect general information—age, sex, body mass index, alcohol consumption, smoking, medication as well as traditional risk factors of atherosclerosis and myocardial infarction (MI) (Mennander, Kuukasjarvi et al. 2008).

### 4.1.6 Wellcome Trust Case Control Consortium 2 (WTCCC2)

Discovery stroke cohorts in WTCCC2 ischaemic stroke GWAS included samples from the UK (a-c) and Germany (d), with a total of 3,548 cases and 5,972 controls (International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2) et al. 2012). Cases were phenotyped and classified into mutually exclusive aetiologic subtypes according to the TOAST classification (Adams, Bendixen et al. 1993).

(a) St George's Stroke Study, London, UK: Ischaemic stroke patients of European descent attending a cerebrovascular service were recruited in 1995–2008. All cases were phenotyped by one experienced stroke neurologist with a review of original imaging. All patients underwent clinically relevant diagnostic examinations, including brain imaging with CT and/or magnetic resonance imaging (MRI) as well as ancillary diagnostic investigations including duplex ultrasonography of the

carotid and vertebral arteries, echocardiography, Holter monitoring, magnetic resonance angiography (MRA), CT angiography (CTA) and blood tests.

(b) The Oxford Vascular Study, UK: Patients of European descent who had suffered an acute ischaemic stroke or TIA with evidence of infarction in brain imaging were recruited during 2002–2008 as a part of a population-based study of all TIA and stroke cases among approximately 91,000 people in Oxfordshire, UK. All cases were phenotyped by one experienced stroke neurologist with a review of original imaging.

(c) The Edinburgh Stroke Study, Scotland, UK: Between 2002 and 2005, consecutive consenting stroke patients who were admitted to or seen as outpatients at the Western General Hospital, Edinburgh, were prospectively recruited. Cases in this study were those with a clinically evident stroke, demonstrated by brain imaging (CT or MRI) to be ischaemic. An experienced stroke physician assessed each patient as soon as possible after the stroke, prospectively recording demographic and clinical details, including vascular risk factors and results of brain imaging and other investigations.

(d) Munich, Germany: White European patients were recruited consecutively from a single dedicated Stroke Unit (Klinikum Großhadern, Ludwig-Maximilians-University of Munich) in 2002–2008. Brain imaging was performed on all patients, with the majority of patients (>80%) undergoing MRI including diffusion-weighted imaging. The diagnosis of ischaemic stroke was based on neurological symptoms in combination with a documented acute infarct visible in neuroimaging. The diagnostic protocol included ECG and duplex ultrasonography of the extracranial arteries in all cases. Transcranial ultrasound, CTA and/or MRA, transthoracic and transesophageal echocardiography as well as Holter monitoring were performed if clinically relevant.

The control data set for the British discovery samples was the WTCCC2 common control set, which includes healthy blood donors from the United Kingdom Blood Service's (UKBS) collection and individuals from the 1958 Birth Cohort dataset (58C). The control data set for the German cases was taken from the MONICA/KORA Augsburg Study's population-based controls from the same region in Germany (International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2) et al. 2012).

### 4.1.7 Tampere Vascular Study (TVS)

For TVS (N=96), carotid, femoral, and aortic atherosclerotic plaques constituting the intima and inner media were prospectively obtained between 2005 and 2009 from patients fulfilling the following inclusion criteria: 1) carotid endarterectomy due to asymptomatic or symptomatic and hemodynamically significant (>70%) carotid stenosis, or 2) femoral or 3) aortic endarterectomy with aortoiliac or aortobifemoral bypass due to symptomatic peripheral arterial disease. Whole thickness left internal thoracic artery (LITA) samples were used as controls and were obtained during coronary artery bypass surgery All open vascular surgical procedures were performed at the Division of Vascular Surgery and Heart Center at Tampere University Hospital (Oksala, Pärssinen et al. 2013).

### 4.1.8 Other Cohorts

Associations with carotid plaque and common carotid artery (CCA) IMT were examined in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium which brings together five population based studies, and four additional community based studies which had collaborated with the CHARGE consortium in a previous GWAS of these phenotypes. All individuals have (GWAS) data; this was used to perform a look-up of the SNPs (Bis, Kavousi et al. 2011).

Measurements of CCA-IMT were available on 31210 participants from 9 studies, and of carotid artery plaque on participants 25179 from seven studies. The individual studies were: Aging Gene-Environment Susceptibility-Reykjavik Study (AGES), Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), the Rotterdam Study I (RS-I), Old Order Amish (Amish) Study, Erasmus Ruchpen Family (ERF) Study, SardiNIA Study, and Study of Health in Pomerania (SHIP), Austrian Stroke Prevention Study (ASPS), Carotid Atherosclerosis Progression Study (CAPS), Gutenberg Heart Study (GHS), MONICA/KORA Augsburg Study (KORA), The Orkney Complex Disease Study (ORCADES), and The Cardiovascular Risk In Young Finns (YFS) study. (Bis, Kavousi et al. 2011).

Results in study **IV** were replicated in a further 10 cohorts from METASTROKE. In METASTROKE, cases were phenotyped and classified into mutually exclusive aetiologic subtypes according to the TOAST classification (Adams, Bendixen et al. 1993). Nine of the centres used a cross-sectional design,

while one was a large prospective, population based cohort (ARIC). Nine of the centres were of European ancestry, while one consisted of individuals of Pakistani ancestry (RACE). All centres used a case-control methodology; centres with a cross sectional design used logistic regression to model the association of genotype dosages from imputation with the dichotomous outcome of ischaemic stroke and prospective cohorts used Cox proportional-hazards models to evaluate time to first stroke, fitting an additive model relating genotype dose to the stroke outcome. European ancestry replication centres were meta-analysed using a fixed effects inverse-variance weighted method. To assess the evidence for association of the SNP for replication samples of all ancestries, we performed a trans-ethnic meta-analysis using a random-effects model to control for any resulting heterogeneity. To evaluate the overall evidence for association, the results of the discovery and replication analyses were combined using Fisher's Method.

### 4.1.9 Definition of Cardiovascular Risk Factors and Candidate SNPs

**Definition of BMI, hypertension and dyslipidemia.** In all studies, the formula body-mass index (BMI) = weight (kg) / height $(m)^2$ was used for BMI calculation and hypertension was diagnosed if the systolic and/or diastolic blood pressure exceeded 140 and/or 90 mmHg, respectively, or if there was a significant history of hypertension (use of antihypertensive medication).

In YFS, FINCAVAS, ANGES, and LURIC, dyslipidemia was defined as usage of lipid-lowering medication or having a triglyceride concentration of over 2 mmol/l, an LDL cholesterol concentration of over 3 mmol/l (LURIC, 4 mmol/l), an HDL cholesterol concentration below 1 mmol/l, or an HDL/total cholesterol of over 1:4.

**Definition of diabetes.** In YFS, ANGES, and FINCAVAS, participants were classified as having type 2 diabetes mellitus (T2DM) if they (i) had a fasting plasma glucose level of ≥ 7.0 mmol/L (≥ 125 mg/dL); or (ii) reported receiving oral hypoglycemic agents and/or insulin injections and did not have type 1 diabetes mellitus; or (iii) reported a history of physician-diagnosed T2DM, which is consistent with the World Health Organization definition. Women who reported having physician-diagnosed diabetes mellitus only during the term of their pregnancy were considered to have had gestational diabetes and were classified as not currently having T2DM, provided that their plasma glucose levels were not ≥7.0 mmol/L (≥125 mg/dL).

In LURIC, Individuals were classified as having diabetes mellitus if their plasma glucose level was > 125 mg/dL (7 mmol/l) in the fasting state or > 200 mg/dL (11 mmol/l) 2 h after the oral glucose load (performed on individuals with no previous diabetes mellitus diagnosis), or if individuals were receiving oral antidiabetics or insulin.

In all studies, subjects were asked to fill out questionnaires that included questions of for instance concerning the use of medications (including lipid-lowering medication [statins]) and usage of tobacco products.

**Selection of candidate gene SNPs**. Rationale for selecting the HADC9 SNPs for work **III** were based on two previous GWAS studies: First WTCCC2 consortium study that found the strongest association with HADC9 rs11984041, while recent GWAS meta-analysis in 12,389 ischaemic stroke individuals and 62,004 controls found the strongest association with rs2107595, which is in LD

with rs11984041. To account for the 2 SNPs a Bonferroni correction was applied and significance level of 0.025 pre-defined.

## 4.2 Genetic Methods

### 4.2.1 Deoxyribonucleic Acid (DNA) Extraction, Genotyping and Quality Control (I-V)

**In YFS**, genomic DNA was extracted from peripheral blood leukocytes using a commercially available kit and the Qiagen BioRobot M48 Workstation according to the manufacturer's instructions (Qiagen, Hilden, Germany). Genotyping was performed on 2,556 samples using a custom-built Illumina Human 670k BeadChip at the Welcome Trust Sanger Institute. Genotypes were called using the Illuminus clustering algorithm. Fifty-six samples failed to meet the Sanger genotyping pipeline QC criteria (i.e. duplicated samples, heterozygosity, low call rate, or Sequenom fingerprint discrepancy). Out of the remaining 2,500 samples, one failed the gender check, three were removed due to low genotyping call rate (< 0.95), and 54 were excluded for possible relatedness (pi-hat > 0.2). Based on the Hardy–Weinberg equilibrium (HWE) test, 11,766 SNPs were excluded ($p \leq 10\text{-}6$), and 7,746 SNPs failed the missingness test (call rate < 0.95) and another 34,596 failed the frequency test (MAF < 0.01). After quality control, 2,442 samples and 546,677 genotyped SNPs were available for further analysis (Smith, Chen et al. 2010).

Imputation of SNPs means using the LD structure of the genome in prediction of the SNPs that have not been genotyped (Marchini, Howie 2010). Genotype imputation was performed using MACH 1.0 (Li, Willer et al. 2009, Li, Willer et al. 2010) and HapMap II CEU (release 22, NCBI build 36, dbSNP 126) samples as a reference. Palindromic A/T and C/G SNPs were removed before imputation. After imputation, 2,543,887 SNPs were available. SNPs with a squared correlation (r2) of ≥ 0.30 between imputed and true genotypes were considered well imputed.

**In FINCAVAS and ANGES**, for the DNA extraction, 9.0 ml ethylenediaminetetraacetic acid (EDTA) whole blood was drawn from the participants and stored at -20 °C. Genomic DNA was extracted from peripheral blood leukocytes by using the QIAamp DNA Blood Midi Kit and automated biorobot M48 extraction (QIAGEN GmbH, Hilden, Germany). Genotyping was completed for 2,824 samples using the Illumina Cardio-Metabo Chip (Illumina

Inc., San Diego, CA, USA) at the Helmholtz Zentrum, München, Germany. The chips were scanned with the Illumina iScan system and genotypes called with Illumina GenomeStudio software. The following quality control filters were applied: GenCall score < 0.2, sample and SNP call rate < 0.95, HWE p value < 10-6, MAF < 0.01, cryptic relatedness (pi-hat > 0.2), and gender check. After quality control, 2,620 samples and 120,721 SNPs were available. Both genotype and clinical data were available for 2,390 samples in FINCAVAS and for 808 in ANGES.

**In LURIC**, genomic DNA was prepared from EDTA anticoagulated peripheral blood using a common salting-out procedure. Genotyping was accomplished for 2,966 samples using a custom-built Illumina 200k BeadChip (Cardio-Metabo Chip) at the Institute of Human Genetics at the Department of Genomics, Life & Brain Center, University of Bonn, Germany. Forty-seven samples failed to meet the QC criteria (i.e. duplicated samples, possible relatedness, low call rate [< 0.95], or gender discrepancy). Out of the SNPs, 5,384 failed to meet the QC criteria (HWE test [p ≤ 10–6], call rate < 0.95). After quality control, 2,919 samples and 191,341 genotyped SNPs were available for further analysis. Both genotype and clinical data were available for 2,912 samples in this study.

**In KORA**, genotyping was accomplished by using the IBC 50K array, which is an Illumina iSelect genotyping array designed to test ~50,000 SNPs identified through genome-wide meta-analyses associated with a range of cardiovascular, metabolic, and inflammatory syndromes (Keating, Tischfield et al. 2008).

The studied top SNP (rs676210) passed QC (call rate > 0.95, MAF > 0.01, HWE p > 10-6) in all cohorts.

**For the WTCCC2 studies and METASTROKE**, 2,858 cases and 5,716 matched controls genotyped using the Immunochip platform; and 3,940 cases genotyped using either the Illumina 610k or 660k platforms matched with 6,379 controls genotyped on the Illumina Human 1.2M Duo (UK), Illumina Human 550k (German) and Illumina 610k platforms (Italian). Bead intensity data were processed and normalized in BeadStudio (Illumina); data for successfully genotyped samples were extracted and genotypes called within the collections using Illuminus. German controls were genotyped on the Illumina Human550k platform, and intensity data were processed and normalised for each sample in GenomeStudio (Illumina) using the Illumina cluster file HumanHap550v3. Standard quality control procedures were undertaken on all centres, before centre-wise imputation to the 1000 Genomes phase 1 integrated variant set (March 2012),

using IMPUTE v2.2.0. SNPs with poor imputation quality (info<0.3) or low minor allele frequency (MAF<0.01) were discarded.

**In CHARGE cohorts**, the nine studies used commercial genotyping platforms available from Illumina and Affymetrix. Each study performed genotyping quality control checks and imputed the approximately 2.5 million polymorphic autosomal SNPs described in the HapMap CEU population for each participant using available imputation methods.

### 4.2.2 Messenger Ribonucleic Acid (mRNA) Isolation, Microarrays and Quantitative Real-Time PCR of mRNAs (III-V)

In TVS, fresh tissue samples were immediately soaked in RNALater solution (Ambion Inc.) and homogenized using an Ultra-Turrax® T80 homogenizer (IKA) (Oksala, Pärssinen et al. 2013). RNA was extracted with the Trizol reagent (Invitrogen) and miRNEasy® Mini-Kit (Qiagen) with the RNase-Free DNase Set (Qiagen) according to the manufacturers' instructions. The RNA isolation protocol was validated by analysing the integrity of the RNA with the RNA 6000 Nano Chip Kit (Agilent).

The quality of the RNA samples was evaluated spectrophotometrically, and the samples were stored in –80°C. The expression levels were analyzed with an Illumina HumanHT-12 v3 Expression BeadChip (Illumina) analyzing 47 000 transcripts of all known genes, gene candidates, and splice variants. 300–500 ng of RNA was reverse transcribed into cRNA and biotin-UTP labelled using the Illumina TotalPrep RNA Amplification Kit (Ambion), and 1500 ng of cRNA was then hybridized to the Illumina HumanHT-12 v3 Expression BeadChip.

The BeadChips were scanned with the Illumina iScan system. After background subtraction, raw intensity data was exported using the Illumina GenomeStudio software. Further data processing was conducted by means of R language and appropriate Bioconductor modules. Data was log2-transformed and robust multichip average and robust spline normalization (rma_rsn) was used.

After background subtraction, raw intensity data were exported using the Illumina GenomeStudio software. Raw expression data were imported into R version 3.1.1 (http://www.r-project.org/), log2 transformed and normalized by the locally estimated scatterplot smoothing normalization method implemented in the R/Bioconductor package Lumi (www.bioconductor.org) (Oksala, Pärssinen et al. 2013). Locally estimated scatterplot smoothing normalization was selected for the

data because it gave the best accuracy in comparison with quantitative reverse transcription polymerase chain reaction data for artery samples (Raitoharju, Seppälä et al. 2013). Data quality control criteria included detection of outlier arrays based on the low number of robustly expressed genes and hierarchical clustering. Artery samples (n=92: 68 plaque, 24 LITA) fulfilled all data quality control criteria. Probes were considered robustly expressed if the detection was $P<0.05$ for at least half of the samples in the data set. Both HDAC9 and MMP12 were robustly expressed in human atherosclerotic plaques and LITAs.

The high MMP12 expression in the samples was also confirmed using reverse transcription polymerase chain reaction (RT-PCR). The accuracy, sensitivity and specificity between the HumanHT-12 v3 Expression BeadChip and TagMan qRT-PCR low-density array (LDA) in TVS have been validated earlier (Raitoharju, Seppälä et al. 2013).

## 4.3 Biochemical and Immunohistological Methods

**Mesurements of the circulating oxLDL.** In YFS, LURIC and KORA studies, the circulating oxLDL levels were measured with the same ELISA utilizing a specific murine monoclonal antibody, mAB 4E6 (Holvoet, Stassen et al. 1998) (Mercodia, Uppsala, Sweden; detection limit < 0.3 U/l). The assay utilizes the phenomenon where a conformational epitope is formed at apoB-moiety of LDL when at least 60 lysine residues are substituted with aldehydes when LDL gets oxidized (Figure 6). The substitute aldehydes can be produced by the peroxidation of LDL lipids, resulting in the generation of oxLDL. Aldehydes released by endothelial cells under oxidative stress or by activated platelets may also induce the oxidative modification of apoB-100 in the absence of the peroxidation of LDL lipids (Paniagua, Lopez-Miranda et al. 2005).

In YFS, LDL baseline diene conjugation was measured by determining the level of baseline diene conjugation in lipids extracted from LDL (Ahotupa, Marniemi et al. 1998). Serum LDL was isolated by means of precipitation with buffered heparin. Lipids were extracted from LDL samples with chloroform-methanol, dried under nitrogen, then redissolved in cyclohexane and analyzed spectrophotometrically at 234 nm.

**Measurements of serum lipids and apolipoprotein B.** In YFS, LURIC, ANGES, and FINCAVAS, venous blood samples were drawn after an overnight fast. In KORA, non-fasting venous blood samples were drawn. Standard enzymatic methods were used for serum total cholesterol, triglycerides, and high-density lipoprotein (HDL) cholesterol. LDL cholesterol was calculated in YFS, KORA, ANGES, and FINCAVAS by the Friedewald formula: LDL cholesterol = total cholesterol – HDL cholesterol – triglyceride-concentration / 2.2 (Friedewald, Levy et al. 1972). In LURIC, lipoproteins were separated by means of a combined ultracentrifugation–precipitation method (ß-quantification) (Winkelmann, Marz et al. 2001).

In YFS, for ApoB analysis a immunoturbidometrical assay was used (Orion Diagnostica, Espoo, Finland); the interassay CV was 2.8% for apoB (Raitakari, Juonala et al. 2008). In LURIC, apoB was analyzed by means of a photometric assay which uses antihuman apoB antibody (apoB Test, Rolf Greiner Bio-chemica, Flacht, Germany).

**Analysis of C-reactive protein (CRP).** In YFS, CRP was analyzed by an automated analyzer (Olympus AU400) with a latex turbidimetric immunoassay kit (CRP-UL assay, Wako Chemicals, Neuss, Germany). The detection limit reported by the manufacturer for the assay was 0.06 mg/l.

In LURIC, sensitive CRP was measured by means of immunonephelometry on a Behring Nephelometer II (N High Sensitivity CRP, Dade Behring, Marburg, Germany) after the completion of the patient recruitment in 2001 on samples stored at -80 °C. In the CRP assay used, the limit of detection for CRP is > 0.17 mg/L; it is linear up to 500 mg/L. The lowest and the highest CRP concentrations encountered in this study were 0.17 mg/L and 269 mg/L, respectively. In KORA, CRP concentrations were measured using a high-sensitivity immunoradiometric assay (IRMA; range, 0.05–10 mg/l; Survey 1: men aged 45–64; Survey 3) or a high-sensitivity latex enhanced nephelometric assay on a BN II analyzer (Dade Behring, Marburg, Germany; Survey 1: men aged 35–44 and all women; Survey 2). Both methods gave similar results when the same samples were analyzed. The respective intra- and inter-assay coefficients of variation for quality control test sera were 4.0% and 12.0% for the IRMA assay, and 2.5% and 5.1% for the nephelometric assay.

**Immunohistochemistry analyses.** Expression of HDAC9 was examined by immunohistochemistry in human large arteries derived from surgical or post

mortem material: aorta (N=7), internal carotid (N=5), middle cerebral (N=5), and coronary arteries(n=5). Tissues were used with ethical approval via the UK National Research Ethics Service.

Anti-HDAC9 antibodies were rabbit polyclonal, 18970 and 59718 (both Abcam, Cambridge, UK). 18970 is raised against the peptide EVPVGLEPISPLDLRT (corresponding to residues 12-27 of human HDAC9 isoform 1) present in human HDAC9 isoforms 1, 3, 5, 6, 7, CRA_g, CRA_i, CRA_j. 59718 is raised against a peptide corresponding to amino acids 541-590 at the C-terminal of Human HDAC9 isoform 6 that is found in human HDAC9 isoforms 3, 6, 7, 8, 9, 10, CRA_h, CRA_i, CRA_j. Other antibodies used for immunohistochemistry were: CD31 (PECAM1), CD45 (leukocyte common antigen; clones 2B11 and PD7/26) and CD68 (clone PG-M1;) all mouse monoclonals from Dako, Ely, UK; and smooth muscle α-actin (mouse monoclonal, clone 1A4) and smooth muscle myosin (SMM; mouse monoclonal, clone h-SMV) from Sigma-Aldrich, Poole, UK. Paraffin wax embedded sections (6μm) were processed for standard immunohistochemical labelling. Endogenous peroxidase activity was quenched byH2O2 (3% v/v, aqueous solution) for 8 min. After high-pressure heat-induced antigen retrieval (30 s, 125oC, in pH 7.8 Tris-citrate buffer) sections were exposed to primary antibodies. HDAC9 primary antibodies ab18970 and ab59718 were applied to human tissues (1:300) and to pig tissue (1:500) overnight at 4oC. Antibody labelling was visualised using a peroxidase-conjugated secondary reagent (Envision® kit, K-5007, Dako, Ely, UK) and diaminobenzidine (DAB) chromogen, then counterstained with Mayer's haematoxylin. Sections were examined on a Zeiss Axioplan-2 microscope driven by Axiovision software (version 4.7).

In TVS, the vascular samples were histologically classified according to the American Heart Association classification (AHA) (Stary, Chandler et al. 1995). The carotid and femoral artery samples were type V or VI, aorta samples were type VI and all control vessels were histologically normal.

## 4.4 Definition of Subclinical and Clinical Cardiovascular Endpoints

**Coronary angiography.** In FINCAVAS, ANGES, and LURIC, coronary angiography (KORA did not have angiographic data) was performed using the standard Judkins technique. Transluminal narrowing of at least 50% in any major

coronary artery (left anterior descending, left circumflex, or right coronary artery) was the criterion for the diagnosis of coronary artery disease. The number of arteries with significant >50% stenosis was used to determine the severity of coronary artery disease. In this study, subjects with no CAD were compared to those with three-vessel disease [3VD].

**Definition of CAD and MI.** In FINCAVAS, the presence of CAD in the patients prior to the exercise test was determined based on patient interviews, resting ECG recordings, and existing hospital records. Of the patients, 43.6% had undergone coronary angiography (as described above) that unambiguously revealed the status of the coronary arteries. The data on the presence of hypertension, valvular conditions, cardiomyopathies, other heart diseases, and diabetes (types 1 and 2) were based on interviews and hospital records.

In LURIC, CAD was assessed by angiography (as described above) using the maximum luminal narrowing estimated by visual analysis. Clinically relevant CAD was defined as the occurrence of at least one 50% stenosis in at least 1 of 15 coronary segments. In LURIC, MI was defined as evidence of any MI (acute, previous, ST-elevation MI [STEMI], or non-ST-elevation MI [nonSTEMI]). In FINCAVAS, a history of prior MI was based on patient interviews, resting ECG recordings, and hospital records. In ANGES, a clinical diagnosis of MI was based on symptoms, electrocardiographic findings, and biochemical marker tests measuring troponin I and creatine kinase. Information concerning previous cardiovascular diseases, surgical procedures, and MIs was collected from patient records at Tampere University Hospital.

**Definition of Ischemic Stroke and Carotid Intima Media Thickness.** In LURIC, previous cerebrovascular disease events were defined as a documented history of a foregoing transient ischemic attack (TIA), prolonged ischemic deficit, or cerebral infarction with or without a remaining neurologic deficit (Winkelmann, Marz et al. 2001). In WTCCC2 and METASTROKE ischemic stroke was defined as a typical clinical syndrome with radiological confirmation; ascertained cases were classified into individual stroke subtypes using the Trial of Org 10172 in acute stroke (TOAST) criteria in all centres (Adams, Bendixen et al. 1993). Age-at-onset was defined as age at first hospital admission for stroke; where this information was unavailable, age at blood draw was used (7.3% of cases). Age-at-onset quantiles were calculated from all the cases from the discovery datasets in the four stroke phenotypes (all IS and the three stroke subtypes: CE, LAA, SVD) and these were used to evaluate associated loci at different age-at-onset thresholds.

In **III**, each study evaluated the carotid arteries using B-mode ultrasonography and previously-described reading protocols. Data was used from the baseline examination, or the first examination in which carotid ultrasonography was obtained. CCA-IMT was typically summarized as the mean of the maximum of several measurements. For most studies, this was an average of multiple measurements of both the left and right arteries. All studies measured the far wall, and in addition several included the near wall. We also examined atherosclerotic thickening of the carotid artery wall, defined in seven of the nine studies by either the presence of plaque (ARIC, AGES, ERF, CHS, RS-I, SHIP, YFS) or the proxy measure of stenosis greater than 25% (FHS).

## 4.5 Statistical Methods

### 4.5.1 GWAS in YFS (I)

For the GWAS analysis, oxLDL was Box-Cox transformed. Residuals were obtained using a linear regression model, in which the variables were adjusted for sex, age, and BMI, as well as principal components (to control population stratification (Price, Patterson et al. 2006)) and apoB. The GWAS was adjusted for apoB to identify SNPs affecting the oxidation process only (each LDL particle has one apoB molecule and the measured oxLDL strongly correlates with apoB). A GWAS was also performed on oxLDL without adjusting for apoB as well as by adjusting for LDL concentrations. Residuals were standardized (mean 0, s.d. 1) and their distributions confirmed to be very close to normal by means of visual Q-Q plot analysis. We also verified that the estimates for the beta coefficients from the GWAS were not driven by a few outliers by plotting leverage versus standardized residuals plots for the residuals.

Tests for additive genetic effects were carried out on a linear scale by means of linear regression. Genotypes were coded as 0, 1, or 2 when the SNP was genotyped and by dosage (scale 0–2) when imputed. In true genotyped SNPs the minor allele was the effect allele. The imputation software (MACH 1.0) used HapMap II as reference to assign the alleles for imputed SNPs. Tests were performed to assess the association of SNPs with the standardized residuals using PLINK (Purcell, Neale et al. 2007) for the genotyped data. ProbABEL (Aulchenko, Struchalin et al. 2010) was employed to fit the linear regression model, taking into account the genotype uncertainty in imputed SNPs. P values were combined from the analysis by favoring genotyped SNPs over imputed ones. Q-Q and Manhattan plots were drawn for the analysis of the results. The p value for genome-wide significance was set at $p < 5 \times 10^{-8}$, corresponding to a target $\alpha$ of 0.05 with a Bonferroni correction for one million independent tests.

The severity (functionality) of mutations was assessed by PolyPhen-2 version 2.1.0 software (Adzhubei, Schmidt et al. 2010).

Further statistical analyses were performed using the R Statistical package v. 2.11.1 (http://www.r-project.org). In order to define associations non-redundantly associated with oxLDL, forward selection algorithm was applied (as described in

(Pare, Chasman et al. 2009)). All the top SNPs with a p value below $5x10^{-8}$ and the covariates were inserted in the same linear model, and a stepwise model selection (Akaike Information Criterion, AIC) algorithm in the R package Modern Applied Statistics with S (MASS) was used with the Bayesian IC (BIC) criterion to leave only the individually associated SNPs and covariates in the model. Linkage disequilibrium (LD) was also analysed visually with the Haploview software with an r2 threshold of 0.8 using HapMap (phase II, release 22 CEU) haplotypes (Barrett, Fry et al. 2005). Moreover, the possible haplotypic effect of the associated SNPs was studied by using the haplo.stats package in R. To assess the proportion of oxLDL explained by the top SNP, r2 was calculated twice—first by using a linear regression model explaining oxLDL with the SNP and all covariates, and secondly only with the covariates. The remainder of these two was considered as r2 for the SNP.

### 4.5.2 Association Studies (I-V)

The SNPs with genome-wide significance (top SNPs) were associated with cardiovascular-disease-related endpoints (angioraphically verified CAD, severity of CAD, and MI) in FINCAVAS, ANGES, and LURIC. The associations were assessed using the appropriate statistical models (chi-squared test, analysis of variance [ANOVA], linear regression, or Cox Proportional-Hazards regression) in R Statistical package v. 2.15.2 (http://www.r-project.org). Meta-analyses were performed using a fixed effects model when the p for cohort heterogeneity was higher than 0.05. KORA did not have angiographic data and was only used for the replication of the oxLDL association. The YFS participants were young (< 39 yrs, average age 31.7 yrs in 2001), still without major clinical endpoints, and it was therefore not possible to include them in these analyses. P values below 0.05 were considered significant.

Statistics for cerebrovascular disease event study in the LURIC study were performed using logistic regression in the R In the WTCCC2, analysis was performed with logistic regression using PLINK (Purcell, Neale et al. 2007) on the separate groups; meta-analysis using an inverse-variance-weighted approach was performed using METAL (Willer, Li et al. 2010).

In CHARGE cohorts, each study independently implemented a predefined GWAS analysis plan. For the continuous measures of CCA-IMT, we evaluated cross-sectional associations of log(IMT) and genome-wide variation using linear

regression models (or linear mixed effects models, in Amish, FHS, and ERF to account for family relatedness). For each of the 2.5 million SNPs, each study fit additive genetic models relating genotype dosage (0 to 2 copies of the variant allele) with the study trait. For the dichotomous outcome of plaque, each study used logistic regression models (or general estimating equations clustering on family to account for familial correlations in FHS and ERF). In our primary analyses all studies adjusted for age and sex. Some studies made additional adjustments including study site (ARIC and CHS), familial structure (Amish, FHS, and ERF), or for whether the DNA had been whole genome amplified (FHS). A meta-analysis of beta estimates and standard errors was conducted from the nine studies using an inverse-variance weighting approach as implemented in METAL (Willer, Li et al. 2010). Prior to meta-analysis, we calculated a genomic inflation factor (λgc) for each study to screen for cryptic population substructure or undiagnosed irregularities that might have inflated the test statistics. Inflation was low, with λgc below 1.09 in all studies. Genomic control was applied to each study whose genomic inflation factor was greater than 1.00 by multiplying all of the standard errors by the square root of the study-specific λgc. For IMT, we express the association of each SNP and log(IMT) as the regression slope(β), its standard error [SE(β)] and a corresponding p-value. For the presence of plaque, meta-analysis odds ratio (OR) was calculated, which represents the increase or decrease in the odds of plaque for each additional copy of the SNP's coded allele.

In study **IV**, The prevalence of ischaemic stroke by age was obtained from a recent publication; (Seshadri, Wolf 2007) gender-specific estimates were averaged, and prevalences within each of the stroke subtypes were assumed to be approximately 20% of the overall total, similar to proportions seen in population-based studies. The phenotype data was modeled using a continuous unobserved quantitative trait called the disease liability, which we used to approximate the effect of age-at-onset on the liability scale, based on estimates of ischaemic stroke prevalence by age from epidemiological data. We developed two models for our analysis; one based on the prevalence rates for all ischaemic stroke cases, and secondly for the three stroke subtypes. We used these models to calculate posterior liabilities after conditioning on age-at-onset and stroke affection status for the four stroke phenotypes separately. Regression was then performed on posterior liabilities by multiplying the number of samples by the squared correlation between the expected genotype dosage and posterior liabilities for each of the discovery cohorts in the four ischaemic stroke phenotypes (CE, LAA, SVD, IS), following a previous approach (Zaitlen, Lindstrom et al. 2012).

The results from each centre were meta-analysed for each of the four phenotypes using Stouffer's Z-score weighted approach, as implemented in METAL (Marchini, Howie 2010). Genomic control was used to correct for any residual inflation due to population stratification. Between-study heterogeneity was assessed using Cochran's Q statistic. We considered only SNPs present in at least 75% of the cases, and with no evidence of heterogeneity (Cochran's Q p-value > 0.001). All SNPs analysed were either genotyped or imputed in both the Immunochip and the genome-wide datasets. After meta-analysis, the resulting p-values were compared with the equivalent values from an unconditioned analysis. For SNPs more significant in the age-at-onset informed analysis and with $p<5x10^{-6}$, we determined the evidence of a true age-at-onset effect by generating 1000 permutations of age-at-onset and rerunning the age-at-onset informed analysis, meta-analysing as previously. We calculated an empirical p-value by dividing the number of permuted observations showing greater significance in the meta-analysis than the observed results by the number of permutations. Any novel SNP with a meta-analysis $p<5x10^{-6}$ and evidence of an age-at-onset effect at $p<0.05$ were taken forward for replication. We set the experiment-wide significance threshold at $p<5x10^{-8}$.

Furthermore, all of the SNPs identified were then investigated using RegulomeDB to determine the evidence that any of the SNPs have a regulatory function (Boyle, Hong et al. 2012). Moreover, a simulation study was performed to evaluate the age-at-onset informed approach, to show that including age at onset information directly led to the increased significance, due solely to inclusion of age-at-onset information at tested SNPs.

In TVS, Statistical analyses were performed using R version 3.1.1 (http://www.r-project.org). HDAC9 and MMP12 were correlated with previously determined expression signature genes (Puig, Yuan et al. 2011, Salagianni, Galani et al. 2012) with nonparametric Spearman correlation. Association of HDAC9 and MMP12 with AHA classification of plaque severity was studied with analysis of variance (ANOVA). Differences were considered significant when $P<0.05$.

# 5 Results

## 5.1 Novel Genetic Variant Affecting the Oxidation of Low-Density Lipoprotein was Discovered in GWAS (I)

The general characteristics of the study populations showed a female predominance in the YFS population and a male predominance in the FINCAVAS, LURIC, and KORA populations (Table 1). Furthermore, there were age differences between the cohorts, with YFS as the youngest and LURIC as the oldest population. The mean oxLDL levels also varied, being highest in KORA and lowest in LURIC.

**Table 1.** General characteristics of the study populations. Modified from study **I**.

| Study | Acronym | N* | Age, years† | % Male |
|---|---|---|---|---|
| The Cardiovascular Risk in Young Finns Study | YFS | 2080 | 31.7 (5.0) | 45.2% |
| The Ludwigshafen Risk and Cardiovascular Health Study | LURIC | 2912 | 62.6 (10.7) | 69.3% |
| Kooperative Gesundheitsforschung in der Region Augsburg (Cooperative Health Research in the Region of Augsburg) Study | KORA | 1326 | 52.7 (10.6) | 53.1% |
| The Finnish Cardiovascular Study | FINCAVAS | 1118 | 59.9 (10.4) | 71.1% |
| The Angiography and Genes Study | ANGES | 808 | 62.9 (10.0) | 63.9% |

**Abbreviations:** n, number. *n for subjects having data required for the present study; †values expressed as means (standard deviations).

In the YFS discovery GWAS, 328 SNPs were associated with oxLDL with genome-wide statistical significance ($p < 5 \times 10\text{-}8$). All of the statistical adjustments applied produced identical top SNP associations. All of these SNPs were within 210 kb from the apolipoprotein B-100 precursor coding region (OMIM 107730) on chromosome 2 (see the Q-Q plot [Figure 9], Manhattan plot [Figure 10], and regional plot [Figure 11]).

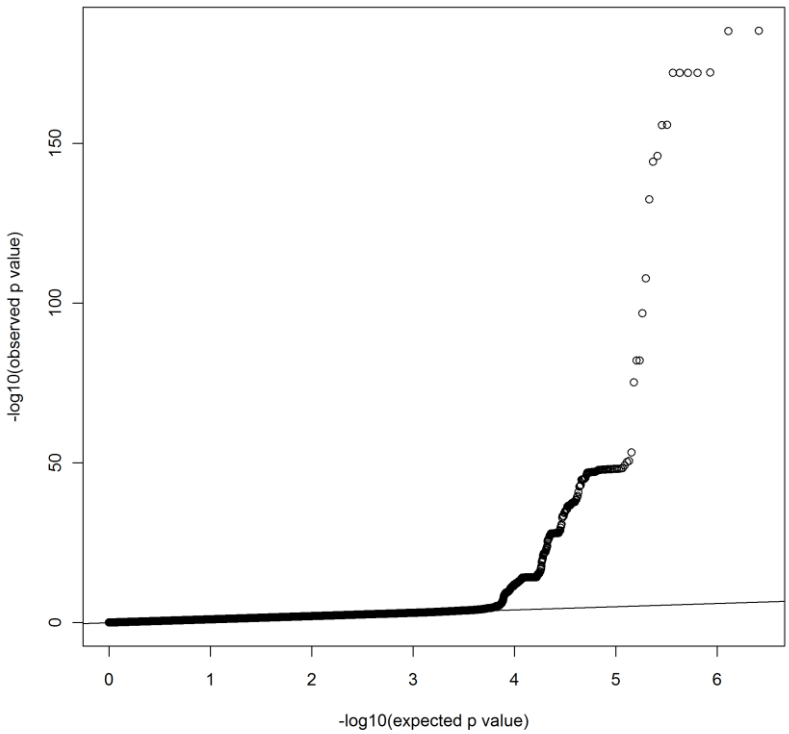



**Figure 9 –** Q-Q plot of the genome-wide association study of oxidized low-density lipoprotein showing a clear deviation from normal distribution (lambda = 0.997). The gray line represents normal distribution and the p-values of the GWAS SNPs are plotted on the y-axis. It shows a clear deviance from normal distribution with the GWAS-results. This means that the results are most probably not due to chance. Modified from study **I**.

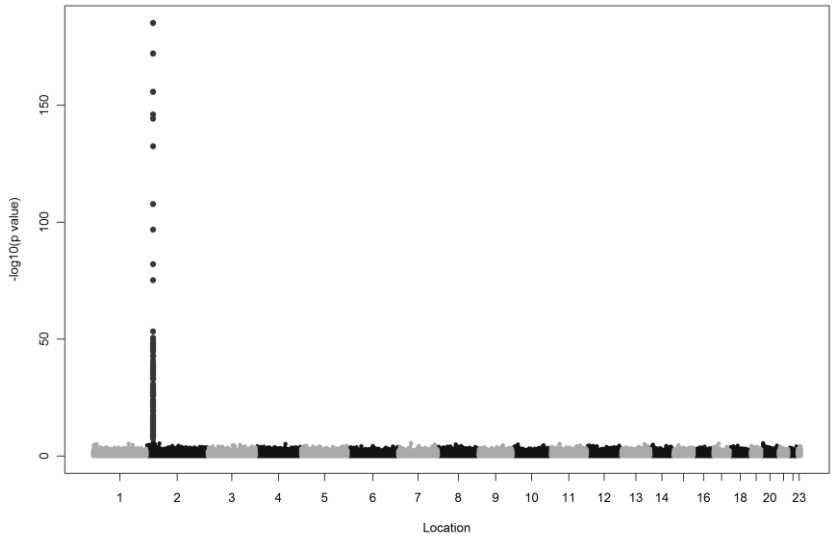



**Figure 10 –** Manhattan plot of the genome-wide association study of oxidized low-density lipoprotein (oxLDL) showing the association of multiple single-nucleotide polymorphisms with oxLDL in chromosome 2. Manhattan plot maps the SNP p values on chromosomes. On the x-axis is the chromosome number (1-22) and the y-axis show the –log of the association p-value. The significant associations are found on chromosome two. Modified from study **I**.

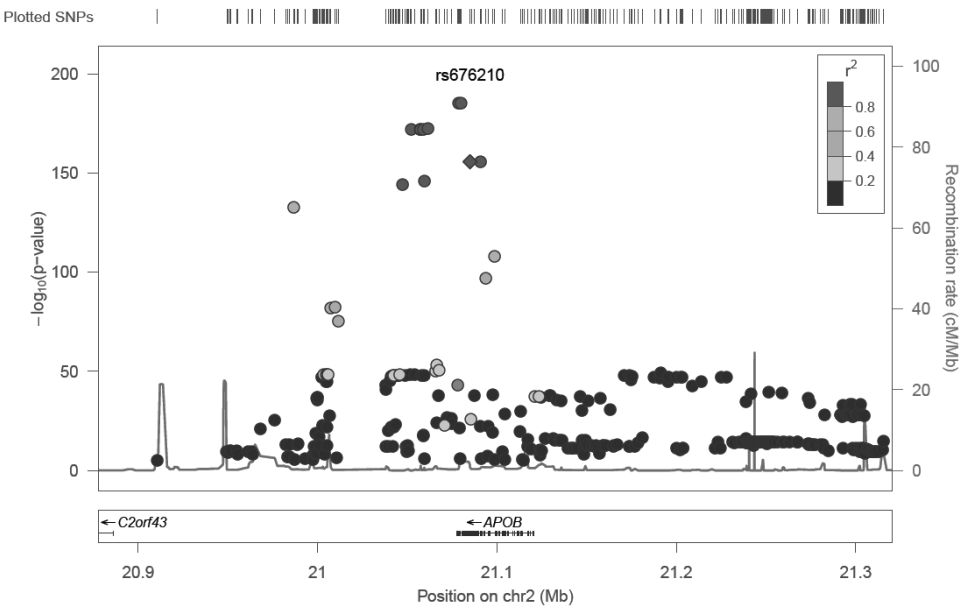



**Figure 11 –** Regional plot of the single nucleotide polymorphisms (SNPs) associated significantly with oxidized low-density lipoprotein. Dots represent -2log10 (p values) of SNPs; the color represents the r2 value of the most significantly associated SNP. The gray line shows recombination rates yielded by the HapMap database. The lower part indicates RefSeq genes in the locus. The plot was drawn using LocusZoom version 1.1 (Pruim, Welch et al. 2010). The main finding of the study, rs676210, has the most significant association with serum oxLDL levels. The SNP is a Pro2739Leu missense mutation located on apolipoprotein B-100 precursor coding region (OMIM 107730) on chromosome 2. Modified from study **I**.

Using a forward selection algorithm with a probability value cut-off of $5 \times 10^{-8}$, only one SNP (rs676210) was independently associated with oxLDL, implicating a role as the proxy for all of the associations. 11 SNPs had $r^2 > 0.5$ with rs676210 (rs1042034, rs6728178, rs6754295, rs673548, rs6711016, rs11902417, rs10184054, rs6544366, rs4564803, rs7557067, and rs2678379) and forward selection algorithm cannot separate which is the true proxy. We chose to include rs676210 in the subsequent analyses since it is biologically most plausible from its LD-block because it causes a missense mutation. We also ran the GWAS by adjusting for rs676210, in addition to other covariates. No other independent associations were found (Figure 12). Moreover, in a haplotype analysis of all associated missense

mutations, only rs676210 (or rs1042034 which is in perfect LD with rs676210) showed an independent effect on oxLDL.

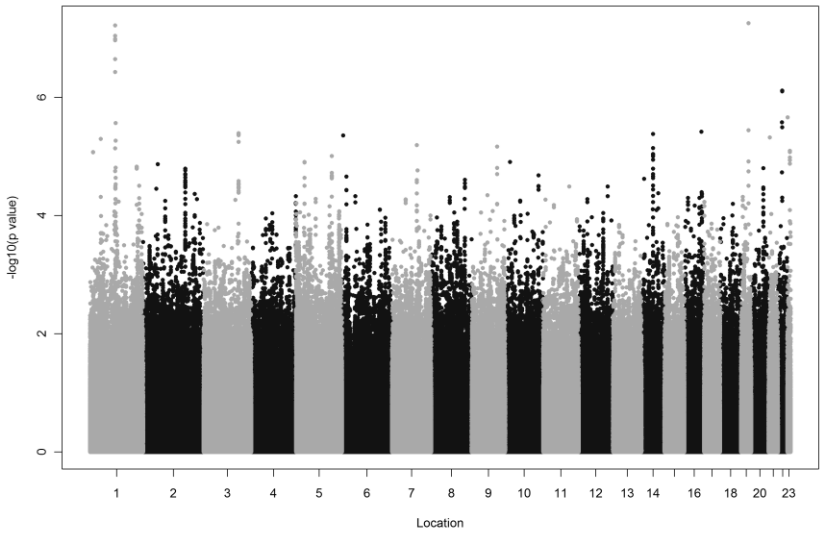



**Figure 12 –** Manhattan plot of the genome-wide association study of oxidized low-density lipoprotein adjusted for the top single-nucleotide polymorphism (rs676210) showing no further independent associations with a $p < 5 \times 10^{-8}$. Modified from study **I**.

**Table 2.** Clinical and biochemical trait profiles of oxidized LDL study cohorts according to rs676210 (Pro2739Leu) genotype. Modified from study **I**.

| | ApoB rs676210 (Pro2739Leu) genotype | | | |
|---|---|---|---|---|
| **Cohort / trait** | **AA** | **GA** | **GG** | **p value** |
| **YFS** | (n = 149) | (n = 806) | (n = 1125) | |
| OxLDL (U/l) | 61.7 (20) | 76.2 (22) | 91.3 (24.6) | 4.3E-136 |
| OxLDL / LDL (U/l/mmol/l) | 19.8 (5.98) | 23.7 (5.12) | 28 (5.22) | 2E-113 |
| OxLDL / apoB (U/l/g/l) | 60.9 (17.1) | 73.4 (13.9) | 85.7 (13.8) | 4.9E-124 |
| Sex (%male) | 67 (45%) | 357 (44.3%) | 516 (45.9%) | 0.789 |
| Hypertension (%) | 5 (3.36%) | 20 (2.48%) | 30 (2.67%) | 0.828 |
| Diabetes (%) | 2 (1.34%) | 3 (0.372%) | 7 (0.622%) | 0.341 |
| Statin use (%) | 0 (0%) | 2 (0.248%) | 5 (0.444%) | 0.582 |
| Age (years) | 31.7 (5.09) | 31.8 (4.9) | 31.7 (5.04) | 0.694 |
| BMI (kg/m$^2$) | 25.3 (3.97) | 25 (4.45) | 25.2 (4.5) | 0.636 |
| Triglycerides (mmol/l) | 1.33 (1.01) | 1.27 (0.747) | 1.37 (0.884) | 0.0483 |
| Total cholesterol (mmol/l) | 5.03 (0.919) | 5.13 (0.964) | 5.17 (0.979) | 0.109 |
| LDL cholesterol (mmol/l) | 3.15 (0.828) | 3.25 (0.844) | 3.28 (0.827) | 0.0801 |
| HDL cholesterol (mmol/l) | 1.32 (0.337) | 1.31 (0.321) | 1.27 (0.308) | 0.00605 |
| Total / HDL cholesterol | 4.05 (1.25) | 4.16 (1.39) | 4.3 (1.37) | 0.00671 |
| apoB (g/l) | 1.02 (0.261) | 1.04 (0.261) | 1.07 (0.266) | 0.00501 |
| **LURIC** | (n = 165) | (n = 986) | (n = 1761) | |
| OxLDL (U/l) | 58.9 (52.6) | 68.6 (26.2) | 80.1 (24.5) | 2.5E-47 |
| OxLDL / LDL (U/l/mg/dl) | 0.507 (0.393) | 0.614 (0.263) | 0.74 (0.312) | 9.7E-36 |
| OxLDL / apoB (U/l/mg/dl) | 0.554 (0.414) | 0.659 (0.223) | 0.768 (0.201) | 4.3E-49 |
| OxLDL / LDL-apoB (U/l / mg/dl) | 0.69 (0.53) | 0.825 (0.309) | 0.981 (0.357) | 7.2E-38 |
| Sex (%male) | 119 (72.1%) | 676 (68.6%) | 1224 (69.5%) | 0.636 |
| Hypertension (%) | 122 (73.9%) | 706 (71.6%) | 1285 (73%) | 0.684 |
| Diabetes (%) | 26 (15.8%) | 175 (17.7%) | 305 (17.3%) | 0.819 |
| Statin use (%) | 80 (48.5%) | 447 (45.3%) | 819 (46.5%) | 0.701 |
| Age (years) | 62.6 (10.8) | 62.9 (10.6) | 62.5 (10.8) | 0.447 |
| BMI (kg/m$^2$) | 26.9 (3.44) | 27.3 (4.12) | 27.5 (4.1) | 0.0501 |

**Table 2 – continued from page 62**

| | | | | |
|---|---|---|---|---|
| Triglycerides (mg/dl) | 154 (84) | 169 (115) | 178 (125) | 0.00487 |
| Total cholesterol (mg/dl) | 194 (35.8) | 192 (39.3) | 192 (38.6) | 0.933 |
| LDL cholesterol (mg/dl) | 119 (33.8) | 116 (34.9) | 115 (33.5) | 0.111 |
| VLDL cholesterol (mg/dl) | 34.3 (25.8) | 36.3 (25.4) | 38.8 (27.5) | 0.0034 |
| HDL cholesterol (mg/dl) | 40.6 (12.5) | 39.1 (10.5) | 38.2 (10.7) | 0.00168 |
| Total / HDL cholesterol | 5.12 (1.63) | 5.31 (3.53) | 5.37 (1.77) | 0.245 |
| apoB (mg/dl) | 104 (22.4) | 104 (25.2) | 105 (24.2) | 0.236 |
| LDL-apoB (mg/dl) | 85.1 (21.3) | 84.7 (22.9) | 84.5 (21.9) | 0.697 |
| | | | | |
| **KORA** | (n = 61) | (n = 462) | (n = 803) | |
| OxLDL (U/l) | 79.9 (22.7) | 88.0 (24.2) | 96.8 (26.7) | 1.05E-11 |
| OxLDL / LDL (U/l / mg/dl) | 0.583 (0.126) | 0.627 (0.154) | 0.674 (0.167) | 0.000000235 |
| Sex (%male) | 33 (54.1%) | 230 (49.8%) | 441 (54.9%) | 0.209 |
| Hypertension (%) | 23 (37.7%) | 194 (42.0%) | 355 (44.2%) | 0.507 |
| Diabetes (%) | 2 (3.3%) | 28 (6.1%) | 48 (6.0%) | 0.675 |
| Statin use (%) | 0 (0%) | 2 (0.43%) | 4 (0.50%) | 0.853 |
| Age (years) | 50.9 (11.0) | 52.5 (10.6) | 52.9 (10.5) | 0.341 |
| BMI ($kg/m^2$) | 26.7 (4.1) | 27.1 (3.9) | 27.3 (4.2) | 0.513 |
| Triglycerides (mg/dl) | 145.6 (83.2) | 167.8 (107.5) | 180.4 (127.0) | 0.334 |
| Total cholesterol (mg/dl) | 232.7 (40.7) | 237.1 (42.2) | 238.8 (46.4) | 0.524 |
| LDL cholesterol (mg/dl) | 142.8 (40.2) | 147.7 (41.7) | 150.0 (43.9) | 0.455 |
| HDL cholesterol (mg/dl) | 55.5 (15.0) | 57.7 (17.3) | 55.1 (16.5) | 0.03 |
| Total / HDL cholesterol | 4.52 (1.59) | 4.50 (1.86) | 4.73 (1.84) | 0.079 |

**Abbreviations:** YFS, The Cardiovascular Risk in Young Finns Study; LURIC, The Ludwigshafen Risk and Cardiovascular Health Study; KORA, Kooperative Gesundheitsforschung in der Region Augsburg Study; OxLDL, oxidized low-density lipoprotein (LDL); apoB, apolipoprotein-B; BMI, body mass index; HDL, high-density lipoprotein; VLDL, very-low-density lipoprotein; **Statistics:** Values are numbers (percentages) in cases of categorical data and means (standard deviations) in cases of continuous data; p values (difference between rs676210 genotype groups) calculated with chi-square test for categorical data, with a linear regression model for oxLDL, oxLDL/LDL, and oxLDL/apoB, and with analysis of variance (ANOVA) for other continuous data. The OxLDL and oxLDL/LDL models have been adjusted for age, sex, BMI, HDL, APOB, triglycerides, and current smoking in YFS and LURIC. The OxLDL and oxLDL/LDL models have been

adjusted for age, sex, survey number, and BMI in KORA. The OxLDL/apoB models have been adjusted for age, sex, BMI, HDL, and triglycerides. **Conversion factors:** total cholesterol, LDL cholesterol, and HDL cholesterol 1 mg/dl = 0.0259 mmol/l; triglycerides: 1 mg/dl = 0.0113 mmol/l; apoB: 1mg/dl = 0.01 g/l.

The variant rs676210 causes a missense mutation (change of proline to leucine at position 2739) in apoB. This mutation was predicted to be damaging in the analysis performed with the PolyPhen-2 software. This supports the notion of the biological functionality of the amino acid change caused by the SNP. The SNP in perfect LD with rs676210 (rs1042034 [Ser4338Asn]) was predicted to be benign, which further supports the role of rs676210 as the functional variant. Two other nearby missense mutations (rs533617 [His1923Arg] and rs679899 [Ala618Val]) were also predicted to be probably damaging to apoB but showed no independent effect on oxLDL in the haplotype analyses and were therefore not studied further. The other significantly associated missense mutations (rs1801695 [Ala4481Thr], rs1367117 [Thr98Ile], and rs1042031 [Glu4181Lys]) were predicted to be benign, and did not show independent effect on oxLDL in the haplotype analyses. Since rs676210 was the most probable SNP behind all the found associations, further analyses were conducted with this SNP only.

In YFS, rs676210 was associated with oxLDL with a p value of $4.3 \times 10^{-136}$ and an effect size of 13.2 U/l oxLDL per allele. The major (risk) allele carriers had significantly higher levels of oxLDL (Table 2). In addition to conventional risk factors, rs676210 explained 11% of the variation in oxLDL ($r2 = 0.11$). The association was replicated in two independent cohorts—in LURIC with a p value of $2.5 \times 10^{-47}$ and an effect size of 10.5 U/l, and in KORA with a p value of $1.1 \times 10^{-11}$ and an effect size of 7.8 U/l. The association was also strong with the oxLDL/LDL, oxLDL/apoB, and oxLDL/LDL-apoB ratios (Table 2).

In a linear regression model adjusted for age, sex, and BMI in the YFS, rs676210 also associated significantly with the apoB-epitope-structure-independent measurement of oxLDL—LDL diene conjugation—with a p value of 0.028 and an effect size of 0.73 μmol/L, confirming the effect of the studied SNP on LDL oxidation.

Furthermore, rs676210 was significantly associated with apoB in YFS; with triglyceride concentrations in YFS and LURIC; with very low-density lipoprotein (VLDL) cholesterol concentration in LURIC; with high-density lipoprotein (HDL) cholesterol concentrations in YFS, LURIC, and KORA; and with total cholesterol/HDL cholesterol concentrations in YFS and LURIC (Table 2).

## 5.2 The ApoB Genetic Variant Pro2739Leu Did Not Associate with Coronary Artery Disease or Myocardial Infarction (I)

**Table 3.** Risk profiles of the additional meta-analysis study cohorts according to apoB rs676210 (Pro2739Leu) genotype. Modified from study **I**.

| | ApoB rs676210 (Pro2739Leu) genotype | | | |
|---|---|---|---|---|
| **Cohort/trait** | **AA** | **AG** | **GG** | **p-value** |
| **FINCAVAS** | (n = 99) | (n = 408) | (n = 608) | |
| Age (years) | 60.2 (10.4) | 60 (10.7) | 59.9 (10.2) | 0.858 |
| Sex (% female) | 26 (26.3%) | 118 (28.9%) | 178 (29.3%) | 0.828 |
| BMI (kg/m$^2$) | 28.2 (4.14) | 27.6 (4.36) | 27.8 (4.56) | 0.814 |
| Hypertension (%) | 88 (88.9%) | 357 (87.5%) | 540 (88.8%) | 0.856 |
| Diabetes (%) | 17 (17.2%) | 50 (12.3%) | 69 (11.3%) | 0.285 |
| Statin use (%) | 53 (53.5%) | 222 (54.4%) | 348 (57.2%) | 0.597 |
| LDL cholesterol (mmol/l) | 2.63 (0.791) | 2.67 (0.934) | 2.75 (0.931) | 0.112 |
| HDL cholesterol (mmol/l) | 1.32 (0.444) | 1.28 (0.409) | 1.27 (0.379) | 0.337 |
| Triglycerides (mmol/l) | 1.42 (0.837) | 1.42 (0.742) | 1.61 (1.97) | 0.0813 |
| Total cholesterol (mmol/l) | 4.58 (0.879) | 4.59 (1.04) | 4.73 (1.15) | 0.0475 |
| **ANGES** | (n = 54) | (n = 309) | (n = 441) | |
| Age (years) | 65.1 (9.09) | 62.8 (10.6) | 62.6 (9.61) | 0.217 |
| Sex (% female) | 18 (33.3%) | 119 (38.5%) | 155 (35.1%) | 0.573 |
| BMI (kg/m$^2$) | 28.2 (4.07) | 27.6 (4.1) | 28.4 (4.46) | 0.0908 |
| Hypertension (%) | 51 (94.4%) | 286 (92.6%) | 408 (92.5%) | 0.873 |
| Diabetes (%) | 14 (25.9%) | 89 (28.8%) | 126 (28.6%) | 0.894 |
| Statin use (%) | 37 (68.5%) | 198 (64.1%) | 304 (68.9%) | 0.368 |
| LDL cholesterol (mmol/l) | 2.57 (0.656) | 2.79 (0.783) | 2.77 (0.826) | 0.434 |
| HDL cholesterol (mmol/l) | 1.19 (0.373) | 1.19 (0.351) | 1.16 (0.311) | 0.297 |
| Triglycerides (mmol/l) | 1.5 (0.923) | 1.52 (0.806) | 1.58 (0.972) | 0.314 |
| Total cholesterol (mmol/l) | 4.42 (0.993) | 4.5 (1.02) | 4.5 (0.938) | 0.719 |

**Abbreviations:** FINCAVAS, The Finnish Cardiovascular Study; ANGES, The Angiography and Genes Study; BMI, body mass index; LDL, low-density lipoprotein; HDL, high-density lipoprotein. **Statistics:** Values are numbers (percentages) in cases of categorical data and means (standard deviations) in cases of continuous data; p values are calculated with chi-squared for categorical variables and with analysis of variance (ANOVA) for continuous variables. **Conversion factors:** total cholesterol, LDL cholesterol, and HDL cholesterol 1 mg/dl = 0.0259 mmol/l; triglycerides: 1 mg/dl = 0.0113 mmol/l

We found statistically significant differences in traits between genotypes in the angiographic cohorts (FINCAVAS, ANGES, and LURIC; see Tables 2 and 3). Therefore, and because of the pleiotropic effect of rs676210 with lipids, the meta-analyses were performed adjusting for age, sex, BMI, LDL, HDL, and triglycerides. We also performed the same analyses without adjusting for LDL, HDL, and triglycerides to reveal which proportion of the effects is caused by the pleiotropic effects of rs676210. The apoB Pro2739Leu (rs676210 allele G, also the allele causing high oxLDL levels) was not associated with CAD, three vessel CAD or MI after adjustment for age, sex, BMI, statin use, LDL, HDL, and triglycerides (Figure 13). Results without adjustment for LDL, HDL, and triglycerides were also not significant in the meta-analysis (Figure 14).

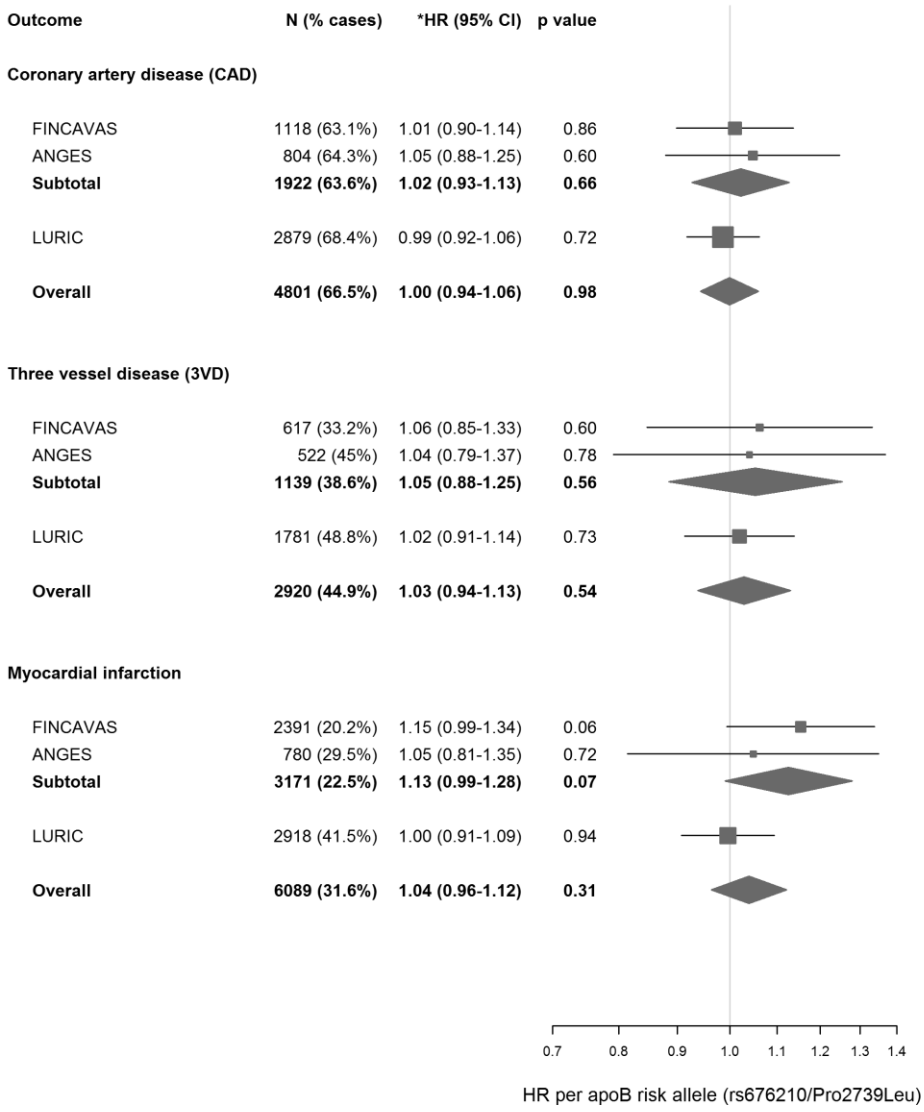



**Figure 13 –** Meta-analyses of the association between apoB alleles (rs676210/Pro2739Leu) and the age of onset of cardiovascular outcomes in three independent clinical study cohorts. **Statistics:** *Hazard ratios (HR; gray squares, sizes indicate the relative weight of each study) and 95% confidence intervals (CI; horizontal lines) were estimated by a per-risk-allele (G) additive Cox Proportional-Hazards regression using age as time variable adjusted for sex, body mass index (BMI), low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides, and statin use. The fixed effect meta-analysis (p value for heterogeneity > 0.05 for all) estimates are shown as gray diamonds. Definitions: a) coronary artery disease (CAD), with or without

angiographically verified 50% stenosis; b) three vessel disease (3VD), with or without angiographically verified 50% stenosis in three vessels; c) myocardial infarction (MI), with or without MI **Abbreviations:** N, number; FINCAVAS, The Finnish Cardiovascular Study; ANGES, The Angiography and Genes Study; LURIC, The Ludwigshafen Risk and Cardiovascular Health Study. Modified from study **I**.

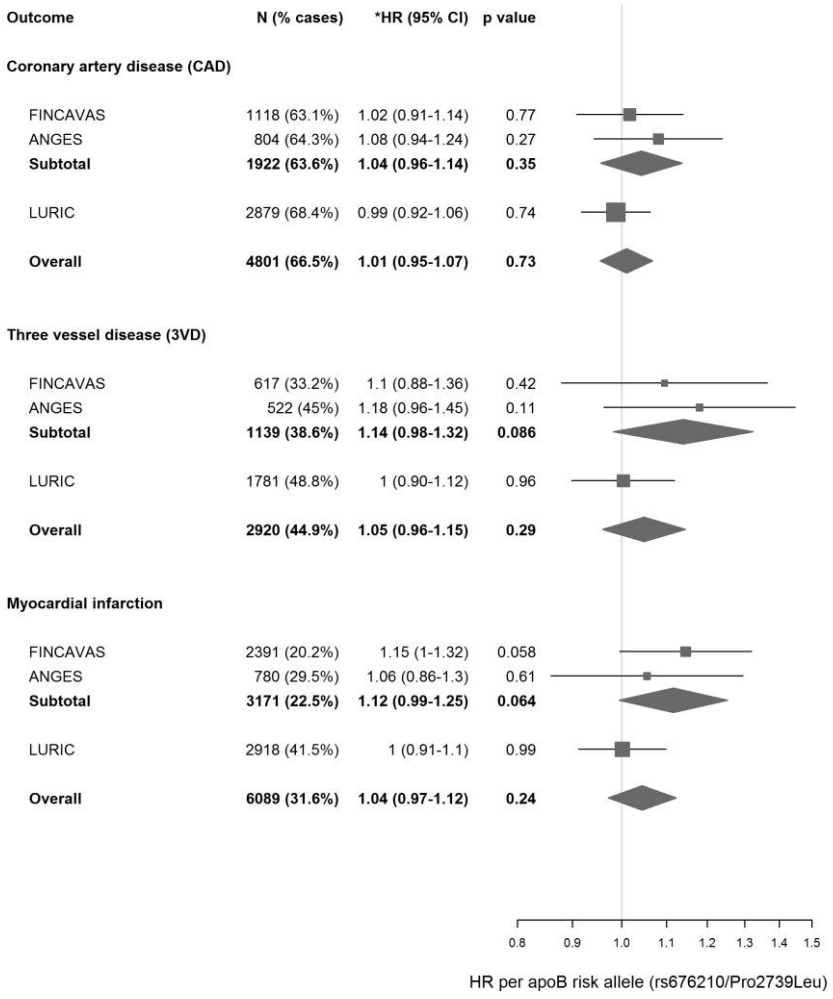



**Figure 14 –** Meta-analyses of the association between apoB alleles (rs676210/Pro2739Leu) and the age of onset of cardiovascular outcomes in three independent clinical study cohorts. **Statistics:** *Hazard ratios (HR; gray squares, sizes indicate the relative weight of each study) and 95% confidence intervals (CI; horizontal lines) were estimated by a per-risk-allele (G)

additive Cox Proportional-Hazards regression using age as time variable adjusted for sex, body mass index (BMI), and statin use. The fixed effect meta-analysis (p value for heterogeneity > 0.05 for all) estimates are shown as blue diamonds. Definitions: a) coronary artery disease (CAD), with or without angiographically verified 50% stenosis; b) three vessel disease (3VD), with or without angiographically verified 50% stenosis in three vessels; c) myocardial infarction (MI), with or without MI. **Abbreviations**: N, number; FINCAVAS, The Finnish Cardiovascular Study; ANGES, The Angiography and Genes Study; LURIC, The Ludwigshafen Risk and Cardiovascular Health Study. Modified from study **I**.

## 5.3 Association of ApoB Pro2739Leu with Cerebrovascular Disease Events was Found (II)

**Table 4.** Characteristics of cerebrovascular disease (CVD) event cases and controls in LURIC. Modified from study **II**.

| Variables | All (N=2913) | Controls (N=2642) | CVD event (N=271) | p-value |
|---|---|---|---|---|
| Sex (female) | 893 (30.7%) | 805 (30.5%) | 88 (32.5%) | 0.541 |
| Hypertension (yes) | 1706 (58.6%) | 1509 (57.1%) | 197 (72.7%) | <<0.001 |
| Diabetes (yes) | 941 (32.3%) | 809 (30.6%) | 132 (48.7%) | <<0.001 |
| Carotid stenosis >50% (yes) | 134 (4.6%) | 93 (3.52%) | 41 (15.1%) | <<0.001 |
| Atrial fibrillation (yes) | 362 (12.4%) | 308 (11.7%) | 54 (19.9%) | <<0.001 |
| CAD (yes) | 1964 (67.4%) | 1759 (66.6%) | 205 (75.6%) | <<0.001 |
| MI (yes) | 1207 (41.4%) | 1063 (40.2%) | 144 (53.1%) | <<0.001 |
| Smoking (yes, ever) | 1856 (63.7%) | 1671 (63.2%) | 185 (68.3%) | 0.116 |
| BMI ($kg/m^2$) | 27.4 (4.07) | 27.4 (4.05) | 27.6 (4.34) | 0.406 |
| LDL cholesterol (mg/dl) | 116 (34) | 116 (34) | 112 (33.6) | 0.0442 |
| OxLDL (U/l) | 75 (28.3) | 75 (28.6) | 75.1 (24.9) | 0.955 |

**Abbreviations**: LURIC, Ludwigshafen Risk and Cardiovascular Health study; CAD, coronary artery disease (over 50% stenosis); MI, myocardial infarction; BMI, body-mass index; LDL, low-density lipoprotein; oxLDL, oxidized LDL. **Statistics:** Values are numbers (percentages) in cases of categorical data and means (standard deviations) in cases of continuous data; p values (difference between rs676210 genotype groups) calculated with chi-square test for categorical data, and with analysis of variance (ANOVA) for continuous data.

The general characteristics of the LURIC cohort and the difference between CVD event cases and controls are displayed in Table 4. More than 60% of the LURIC population was diagnosed with CAD, and CAD and MI were more prevalent among CVD event cases. Out of the known ischaemic stroke risk factors, hypertension, diabetes, carotid stenosis, atrial fibrillation and LDL cholesterol were

associated with history of CVD events. OxLDL did not associate with CVD events in LURIC (p=0.955).

**Table 5.** The association of the apolipoprotein-B Pro2739Leu missense mutation rs676210 with cerebrovascular disease (CVD) events and its risk factors in LURIC. Modified from study **II**.

| | rs676210 genotype | | | |
|---|---|---|---|---|
| | **AA** | **AG** | **GG** | |
| **Variables** | (N=165) | (N=987) | (N=1761) | **p-value** |
| CVD event (yes) | 11 (6.67%) | 80 (8.11%) | 180 (10.2%) | 0.030 |
| OxLDL (U/l) | 58.9 (52.6) | 68.7 (26.2) | 80.1 (24.5) | <<0.001 |
| Sex (female) | 46 (27.9%) | 310 (31.4%) | 537 (30.5%) | 0.643 |
| Hypertension (yes) | 95 (57.6%) | 562 (56.9%) | 1049 (59.6%) | 0.392 |
| Diabetes (yes) | 52 (31.5%) | 327 (33.1%) | 562 (31.9%) | 0.787 |
| Carotid stenosis (yes) | 12 (7.27%) | 42 (4.26%) | 80 (4.54%) | 0.229 |
| Atrial fibrillation (yes) | 17 (10.3%) | 127 (12.9%) | 218 (12.4%) | 0.633 |
| CAD (yes) | 119 (72.1%) | 662 (67.1%) | 1183 (67.2%) | 0.469 |
| MI (yes) | 62 (37.6%) | 423 (42.9%) | 722 (41%) | 0.373 |
| Smoking (yes, ever) | 116 (70.3%) | 607 (61.5%) | 1133 (64.3%) | 0.0643 |
| BMI ($kg/m^2$) | 26.9 (3.44) | 27.3 (4.12) | 27.5 (4.1) | 0.0501 |
| LDL cholesterol (mg/dl) | 119 (33.8) | 116 (34.9) | 115 (33.5) | 0.111 |

**Abbreviations:** LURIC, Ludwigshafen Risk and Cardiovascular Health study; CAD, coronary artery disease (over 50% stenosis); MI, myocardial infarction; BMI, body-mass index; LDL, low-density lipoprotein; oxLDL, oxidized LDL. **Statistics:** Values are numbers (percentages) in cases of categorical data and means (standard deviations) in cases of continuous data; p values (difference between rs676210 genotype groups) calculated with chi-square test for categorical data, with logistic regression for CVD events and with analysis of variance (ANOVA) for other continuous data. The CVD event model is adjusted for hypertension, diabetes, carotid stenosis, atrial fibrillation and low-density lipoprotein cholesterol.

In the logistic regression model with no covariates, rs676210 associated with CVD events (p=0.030, odds ratio [OR] = 1.28 [95% confidence interval, CI = 1.03–1.60] for risk allele G). In the logistic regression model adjusted for the significant risk factors (hypertension, diabetes, carotid stenosis, atrial fibrillation and LDL

cholesterol), rs676210 remained significantly associated with CVD events (p=0.030, OR=1.29 [1.03–1.63] for risk allele G). The studied LURIC variables according to rs676210 genotype are displayed in Table 5. In addition to oxLDL and CVD events, only BMI and smoking were borderline-significantly associated with rs676210.

**Table 6.** The general characteristics of ischaemic stroke cases and controls in the WTCCC2 study populations. Modified from study **II**.

| Cohort | All cases | Subtypes | | | Controls |
|---|---|---|---|---|---|
| - | - | CE | LAA | SVD | - |
| | | | **Number** | | |
| WTCCC2-Germany | 1,174 | 330 | 346 | 106 | 797 |
| WTCCC2-UK | 2,374 | 460 | 498 | 474 | 5,175 |
| *Total* | *3,548* | *790* | *844* | *580* | *5,972* |
| | | | **Age (mean ± SD)** | | |
| WTCCC2-Germany | 66.9 ± 12.9 | 71.7 ± 12.1 | 65.8 ± 10.8 | 65.9 ± 11.4 | 62.7 ± 10.9 |
| WTCCC2-UK | 72.2 ± 12.3 | 77.0 ± 12.9 | 70.0 ± 10.2 | 70.4 ± 11.7 | 52* |
| | | | **Male sex (%)** | | |
| WTCCC2-Germany | 61.9 | 52.1 | 70.7 | 72.6 | 51.4 |
| WTCCC2-UK | 53.7 | 59.4 | 63.3 | 49.8 | 50.5 |

**Abbreviations:** WTCCC2, Wellcome Trust Case-control consortium II; IS, all ischaemic stroke cases; SVD, small-vessel disease; CE, cardioembolic source; LAA, large-artery atherosclerosis**.** **Note *,** a subset of controls from WTCCC2-UK were from the 1958 birth cohort; the age of all other participants was unknown.

We attempted to replicate the association of rs676210 with CVD events in LURIC in the meta-analysis of WTCCC2 ischaemic stroke cohorts. The general characteristics as well as number of cases and controls in WTCCC2 are displayed in Table 6. Logistic regression without adjustments was used as in LURIC. There was no significant association of rs676210 with all types of ischemic stroke (p=0.81, OR=1.00 [0.93-1.09]), LAA (p=0.85, OR=0.99 [0.87-1.12]), CE (p=0.66, OR=1.03 [0.90-1.18]), or SVD (p=0.65, OR=0.97 [0.83-1.13]).

## 5.4 HDAC9 Promotes Carotid Atherosclerosis (III)

**Table 7.** General characteristics of the TVS population. Modified from study **III**.

| | **Carotid plaque** | **Aortic plaque** | **Femoral plaque** | **Control arteries** | **All** |
|---|---|---|---|---|---|
| N | 29 | 15 | 24 | 28 | 96 |
| Age, median, year (SD) | 70 (9.5) | 61.0 (10.8) | 76.0 (9.4) | 69.0 (9.6) | 69.0 (10.2) |
| Males, % | 62.1 | 73.3 | 70.8 | 82.1 | 71.9 |
| BMI, median $kg/m^2$ (SD) | 25.6 (3.4) | 25.9 (4.2) | 26.7 (4.3) | 28.2 (5.1) | 27.0 (4.4) |
| Dyslipidemia, % | 75.9 | 46.7 | 70.8 | 85.7 | 72.9 |
| Statins, % | 100.0 | 40.0 | 62.5 | 82.1 | 76 |
| Hypertension, % | 79.3 | 80.0 | 87.5 | 100.0 | 87.5 |
| Blood pressure medication, % | 82.8 | 80.0 | 79.2 | 92.9 | 84.4 |
| History of smoking % | 65.5 | 100.0 | 70.8 | 64.3 | 71.9 |
| AHA Class V-VI, %, of the atherosclerotic arteries | 82.8 | 73.3 | 62.5 | NA | 74.6 |

**Abbreviations:** N, Number; BMI, body-mass index; AHA, American Heart Association

Both SNPs (rs11984041 and rs 2107595) were associated with both CCA-IMT and with presence of carotid plaque (Table 9) in CHARGE cohorts (Table 8). The strongest associations were seen for SNPs rs2107595; CCA-IMT p=0.0018; carotid plaque p=0.0022.

In all arterial beds strong HDAC9 labelling was seen in nuclei and cytoplasm of VSMC, and in endothelial cells, where intact endothelia were visible (example in Fig 1A). In the medial layer a high fraction of cells were labelled (80-90%) with a similar distribution to the VSMC-specific marker SMM (Figure 15). A minority of medial cells, with fibroblast-like morphology, were negative for HDAC9 and SMM, consistent with normal incidence of structural fibroblasts. Similar results were obtained with the two different anti-HDAC9 antibodies used.

Distinct patterns of labelling were observed with primary antibodies to leukocyte common antigen (CD45) and a lysosomal marker for macrophage/monocytic cells (CD68) at similar titre. Immunolabeling was absent in adjacent negative control sections treated without primary antibody (Figure 16).

Gene expression was analysed from 29 carotid, 15 abdominal aorta plaques and 24 femoral plaques, and 28 atherosclerosis free LITA controls. Demographics and AHA plaque grading for the different plaques are shown in Table 7.

HDAC9 expression was upregulated in carotid plaques compared to LITA controls. (p=0.00000103, FC=3.06). It was also upregulated in aortic plaques (p=0.0038, FC=1.76) and femoral plaques (p=0.038, FC=1.57) compared to controls. HDAC9 mRNA expression was greater in carotid compared with femoral plaques (p=0.0038, FC=1.76), although there was no significant difference between carotid and aortic plaques (p=0.90, FC=1.19).

**Table 8.** The general characteristics of CHARGE cohorts used in association analyses. Modified from study **III**.

| Characteristic | AGES N=3073 | Amish N=1,054 | ARIC N= 7,767 | CHS N= 3,261 | ERF N= 1,809 | FHS N=3,004 | RS-I N= 4,699 | SardiNIA N=4,235 | SHIP N= 2,309 |
|---|---|---|---|---|---|---|---|---|---|
| Age, years | 76.4(5.4) | 48.1(15.9) | 54.3(5.7) | 72.3(5.4) | 48.5(14.5) | 58.5(9.7) | 68.9(8.70) | 43.5(17.5) | 61.8(9.5) |
| Women, % | 57.7% | 49.4% | 53% | 61% | 56.5 | 53.3% | 59.3% | 56.2% | 48.6% |
| Hypertension, % | 80.6% | 9.3% | 27% | 51% | 51.4% | 40.5% | 55.9% | 29.1% | 72.4% |
| Diabetes, % | 11.6% | 2.1% | 8% | 12% | 6.1% | 8.6% | 10% | 4.8% | 10.1% |
| Current smoker, % | 12.6% | 9.4% | 25% | 11% | 39.4% | 15.6% | 23.4% | 20.2% | 19.2% |
| Total cholesterol, mg/dL | 217.9(44.5) | 211.3(48.1) | 214.7(40.5) | 213.0(38.9) | 214.4(42.6) | 205.9(39.7) | 256.0(46.8) | 208.6(42.1) | 234.3(47.9) |
| HDL cholesterol, mg/dL | 61.0(17.1) | 55.7(14.8) | 50.7 (16.8) | 55.3(15.8) | 49.5(14.1) | 51.1(16.1) | 51.8(13.9) | 64.4(14.9) | 55.3(17.8) |

| **Table 8. Continued** | **from page** | **76.** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Triglyceride, mg/dL | 107.0 (59.0) | 74.9(47.1) | 136.0(89.5) | 140.4(76.4) | 118.6(68.1) | 142.3(138.6) | N/A | 87.2(61.4) | 177.6(134.8) |
| BMI, kg/m$^2$ | 27.1(4.5) | 26.9(4.7) | 26.9(4.7) | 26.3(4.5) | 26.8(4.7) | 27.9(5.1) | 26.3(3.7) | 25.3(4.7) | 28.5(4.6) |
| Prevalent CVD | 21.9% | 6.9% | 5% | 0% | 3.1% | 10.4% | 30.8% | 1.7% | 8.4% |
| IMT common carotid | 0.97(0.1) | 0.74(0.2) | 0.77(0.2) | 1.03(0.2) | 0.82 (0.2) | 0.74(0.2) | 1.02(0.2) | 0.54(0.1) | 0.93(0.2) |

**Abbreviations:** N, Number; BMI, body-mass index; HDL, high-density lipoprotein; CVD, cardiovascular disease; IMT, intima-media thickness.
**Conversion factors:** total cholesterol, HDL cholesterol 1 mg/dl = 0.0259 mmol/l; triglycerides: 1 mg/dl = 0.0113 mmol/l.

**Table 9.** Association of the studied SNPs with carotid IMT and plaque in meta-analysis. Modified from study **III**.

A. CCC IMT

| SNP | Allele1 | Allele 2 | Freq1.z | Effect | StdErr | P.value | Direction | N | N_eff | P.value.z |
|---|---|---|---|---|---|---|---|---|---|---|
| rs11984041 | t | c | 0.088 | 0.0077 | 0.0027 | 0.00391 | +++++++-+ | 31210 | 26187.95 | 0.006243 |
| rs2107595 | a | g | 0.146 | 0.0065 | 0.0021 | 0.00184 | +-+++++++ | 31210 | 27035.76 | 0.001833 |

B. CAROTID PLAQUE

| SNP | Allele1 | Allele 2 | Freq1.z | Effect | StdErr | P.value | Direction | N | N_eff | P.value.z |
|---|---|---|---|---|---|---|---|---|---|---|
| rs11984041 | t | c | 0.097 | 0.1069 | 0.0374 | 0.00425 | +++-+++ | 25179 | 21616.84 | 0.002554 |
| rs2107595 | a | g | 0.159 | 0.0911 | 0.0298 | 0.00222 | +++++++ | 25179 | 22257.60 | 0.001395 |

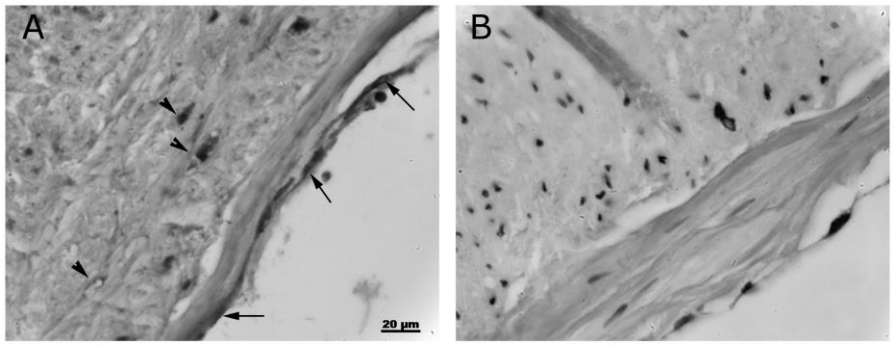



**Figure 15 –** Immunohistochemical labelling of HDAC9 in human middle cerebral artery Post mortem sample of human middle cerebral artery labelled for HDAC9 (panel A) or with no primary antibody (B). A: cellular HDAC9 immunoreactivity (brown) is evident in endothelial cells (examples marked with arrows) and in myoctes within the medial layer (examples marked with arrowheads). B: labelling was absent from neighbouring sections treated identically but with no primary antibody. Scale bars 20 µm. Modified from study **III**.

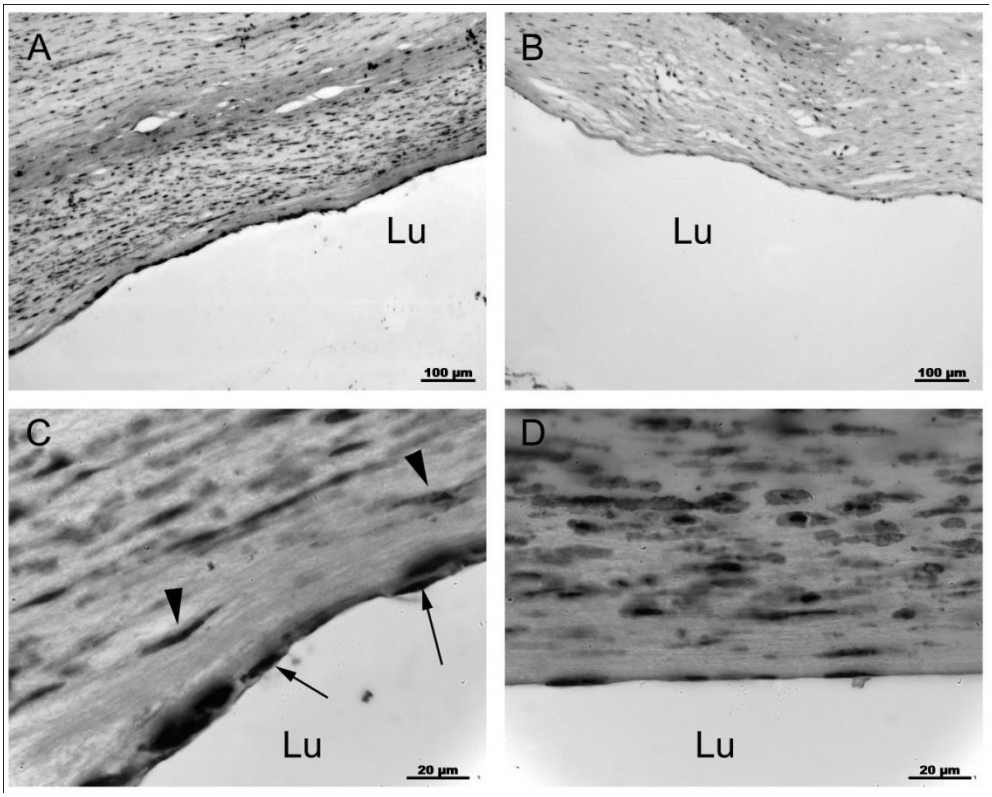



**Figure 16 –** Immunohistochemical labelling of HDAC9 in human aorta. Cellular HDAC9 immunoreactivity (brown) in adult human aorta (A) was absent from a neighbouring negative control section (B). C, higher magnification shows HDAC9 in VSMC (arrowheads) and endothelial cells (arrows). In VSMC a similar pattern of labelling is seen with smooth muscle myosin (SMM; D). Haematoxylin counterstain (blue) labels nuclear chromatin. Scale bars 100 µm (A,B), 20 µm (C,D). Modified from study **III**.

## 5.5 Novel Association with Ischemic Stroke Was Found Near MMP12 Locus in Age-of-Onset Informed GWAS (IV)

As prevalence rates in ischemic stroke are markedly affected by age, younger onset cases may have higher genetic predisposition. Therefore, we investigated whether an age-at-onset informed approach could detect novel associations with ischemic stroke and its subtypes; CE, LAA and SVD. Regression analysis to identify SNP associations was performed on posterior liabilities after conditioning on age-at-onset and affection status. We identified a novel association with an MMP12 locus in LAA (Figure 17, rs660599; $p=2.5x10^{-7}$), with independent replication in METASTROKE population (p=0.004, OR(95% CI)=1.18(1.05–1.32); meta-analysis $p=2.5x10^{-8}$).

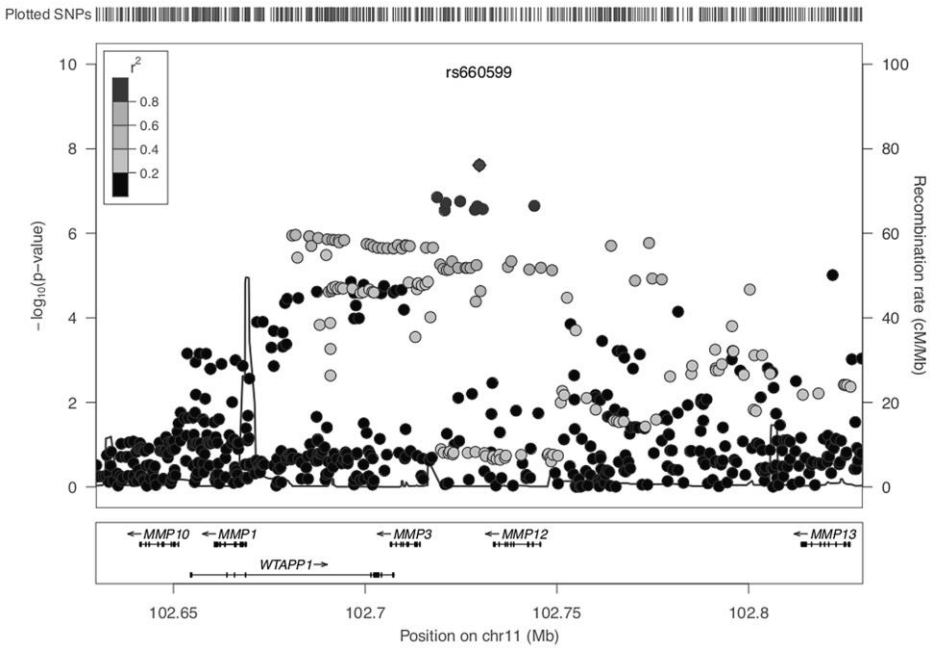



**Figure 17 –** LocusZoom plot of MMP12 association using age-at-onset informed approach. SNPs are colored based on their correlation (r2) with the labeled top SNP, which has the smallest P value in the region. The fine-scale recombination rates estimated from 1000 Genomes (EUR) data are marked in light gray, with genes marked below by horizontal gray lines. Arrows on the horizontal gray lines show the direction of transcription, and rectangles are

exons. SNP p-values are from the discovery meta-analysis only with the exception of rs660599, for which the given p-value indicates the overall evidence for association from the discovery and replication cohorts. Modified from study **IV**.

The nearby gene, MMP12, was significantly overexpressed in carotid plaques compared to atherosclerosis-free control arteries ($p=1.2x10^{-15}$; fold change=335.6) in TVS.

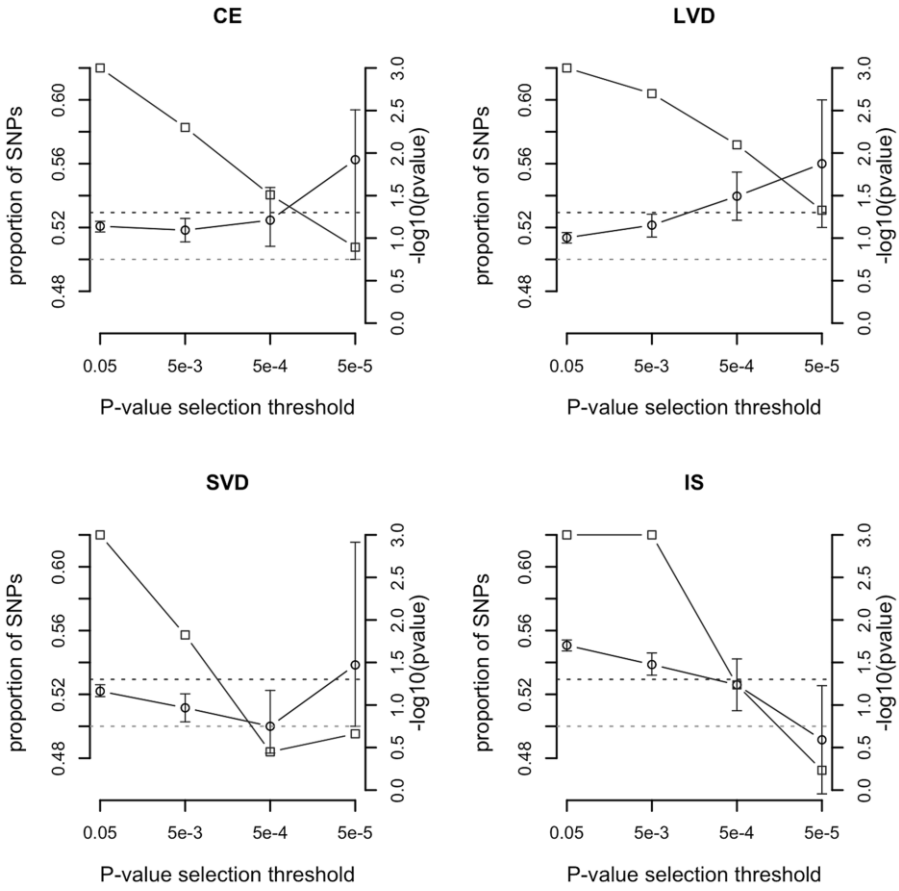



**Figure 18 –** Evaluation of evidence genome-wide for SNPs exhibiting greater significance using the age-at-onset informed approach compared to permutations. -log10(p value) from permutations for evidence of age-at-onset effect at given SNP p-value selection threshold shown in light gray; median proportion of SNPs (with IQR) more significant in observed age-at-onset informed meta-analysis compared to permutations shown in dark gray; horizontal line at p=0.05; horizontal line at median proportion of SNP=0.5; IS, all ischaemic stroke; CE,

cardioembolic stroke; LAA, large artery atherosclerotic stroke; SVD, small vessel disease. Modified from study **IV**.

Permutation analyses demonstrated a gain of power for detecting associations when accounting for age-at-onset in all four stroke phenotypes (Figure 18, $p<0.001$). Our results show that a covariate-informed design, by adjusting for age-at-onset of stroke, can detect variants not identified by conventional GWAS.

Furthermore, using RegulomeDB, we found further evidence from ENCODE that one of these SNPs (rs2276109) is indeed functional, giving evidence that the associated locus in this analysis is likely to affect MMP12 expression through altered transcription. We also compared the sum of the meta-analysis Z scores from all SNPs with $p<0.05$ in the observed age at onset informed meta-analysis with those from permutations. At this p-value selection threshold, we found strong evidence ($p<0.001$) for genome-wide age-at-onset effects in each of the stroke phenotypes, with consistently increased summed Z scores in the observed age-at-onset informed meta-analysis compared to the permutations.

## 5.6 HDAC9 and MMP12 Expression in the Atherosclerotic Plaque (V)

The general characteristics of the TVS study population are shown in Table 7. Expression of HDAC9 and MMP12 were correlated with the expression signatures of M1/M2 macrophages, and plaque stability in atherosclerotic plaques in three artery sites; carotid artery (Figure 19), femoral artery (Figure 21), abdominal aorta (Figure 22), and control artery LITA (Figure 23).

HDAC9 expression correlated significantly with MMP12 expression in carotid artery plaques ($R=0.46$, $p=0.012$), and LITA controls ($r=-0.44$, $p=0.034$). The correlation was not statistically significant in femoral ($r=0.04$, $p=0.84$) or abdominal ($r=-0.10$, $p=0.72$) plaques.

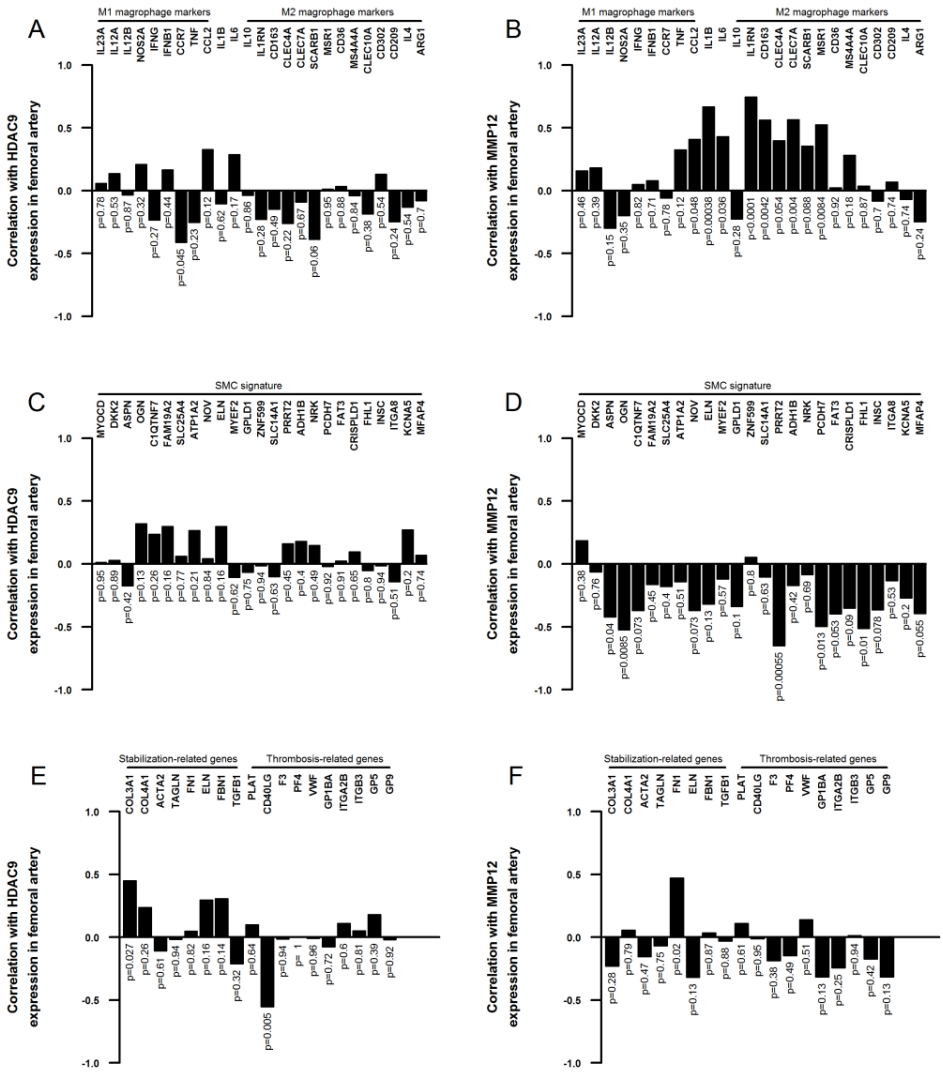



**Figure 19 –** Correlation among Histone deacetylase 9 (HDAC9) and matrix metalloproteinase 12 (MMP12) mRNA levels and established M1/M2 macrophage markers (panels A and B) in human atherosclerotic plaque, top 25 genes differentially expressed between regions of human atherosclerotic plaque enriched in smooth muscle cells (SMCs, panels C and D), and established markers for plaque stability (panels E and F) in carotid artery samples. Spearman correlation coefficient (r) and statistical significance values (P) are shown. Modified from study **V**.

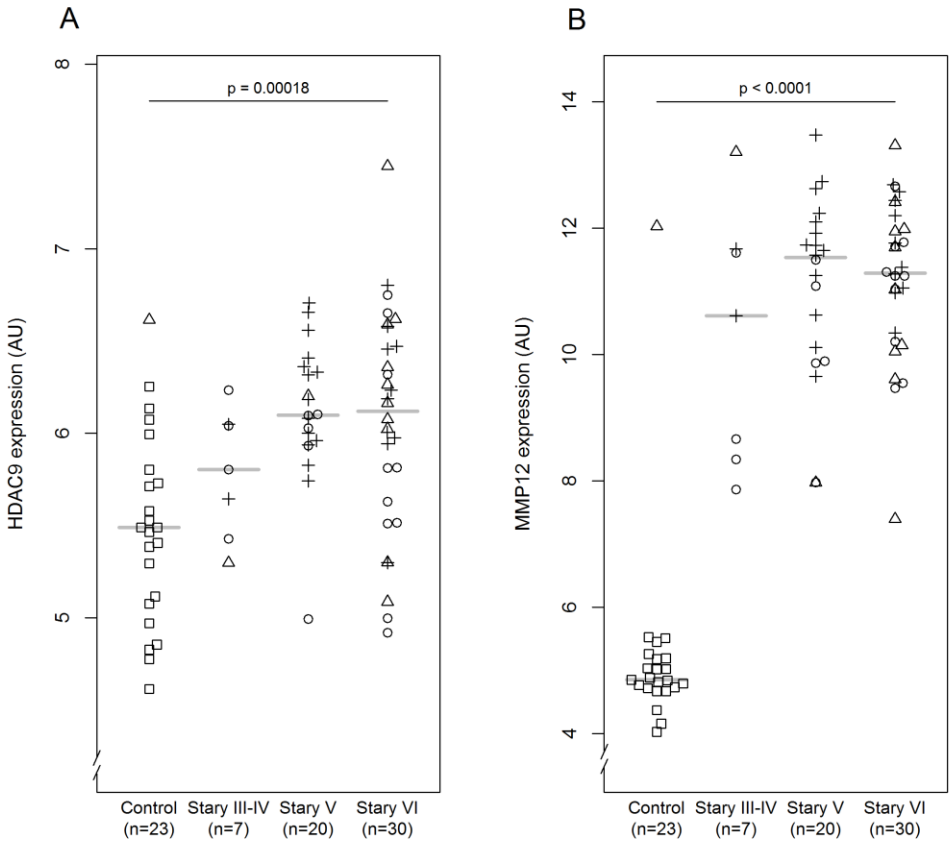



**Figure 20 –** The association of Histone deacetylase 9 (HDAC9, panel A), and matrix metalloproteinase 12 (MMP12, panel B) mRNA levels with the American Heart Association classification of plaque severity in carotid (cross), femoral (sphere), abdominal (triangle), and left internal thoracic artery (LITA, square) arteries. Horizontal line shows median expression level. AU, artificial unit. P values calculated with analysis of variance (ANOVA). Modified from study **V**.

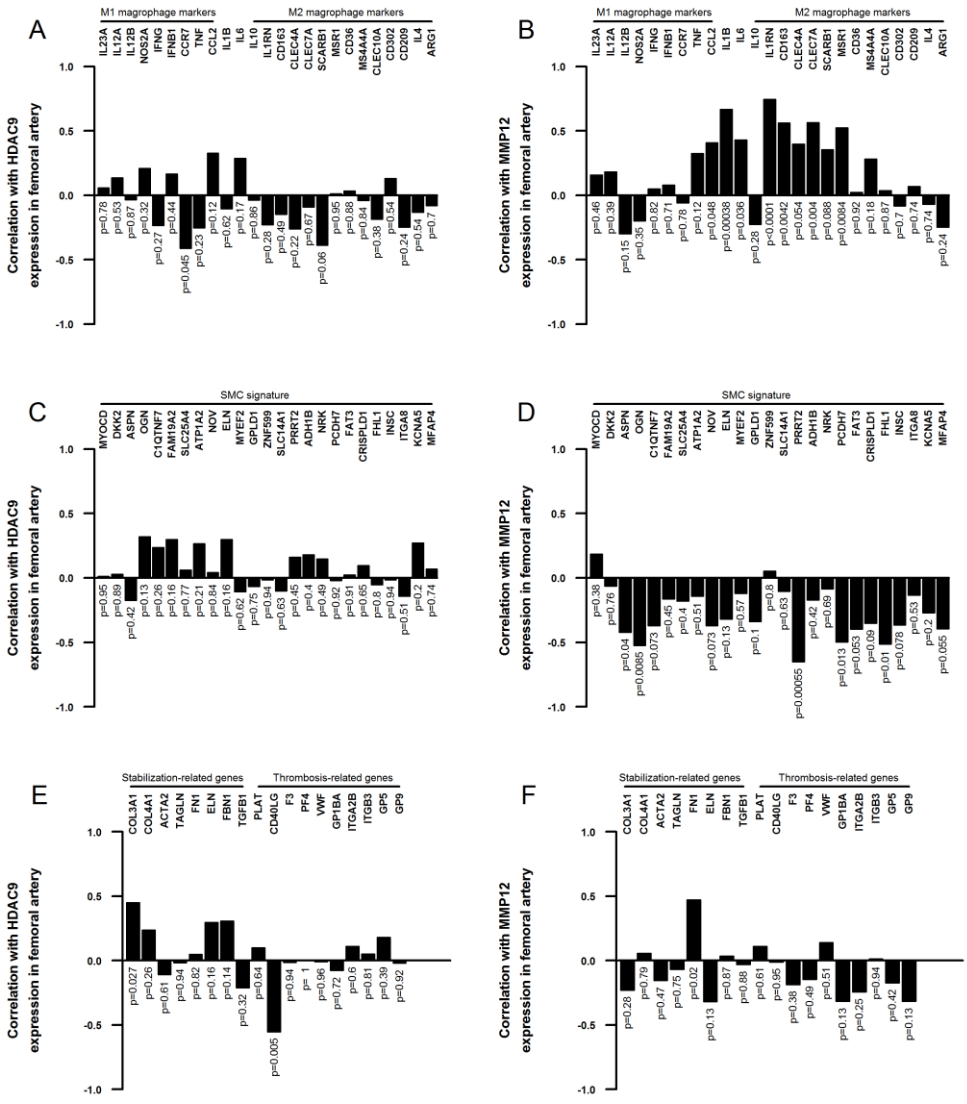



**Figure 21 –** Correlation among Histone deacetylase 9 (HDAC9) and matrix metalloproteinase 12 (MMP12) mRNA levels and established M1/M2 macrophage markers (panels A and B) in human atherosclerotic plaque, top 25 genes differentially expressed between regions of human atherosclerotic plaque enriched in smooth muscle cells (SMCs, panels C and D), and established markers for plaque stability (panels E and F) in femoral artery samples. Spearman correlation coefficient (r) and statistical significance values (P) are shown. Modified from study **V**.

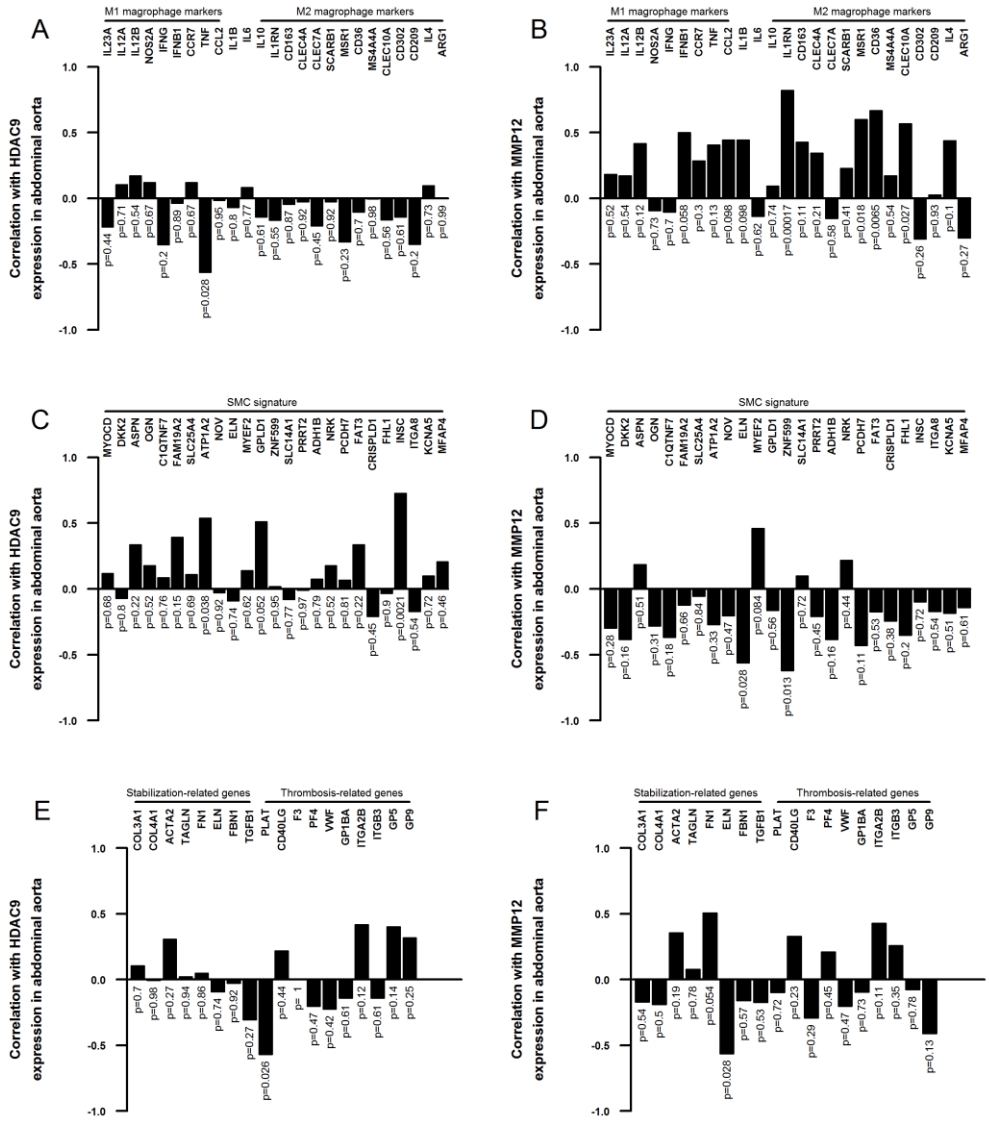



**Figure 22 –** Correlation among Histone deacetylase 9 (HDAC9) and matrix metalloproteinase 12 (MMP12) mRNA levels and established M1/M2 macrophage markers (panels A and B) in human atherosclerotic plaque, top 25 genes differentially expressed between regions of human atherosclerotic plaque enriched in smooth muscle cells (SMCs, panels C and D), and established markers for plaque stability (panels E and F) in abdominal aorta samples. Spearman correlation coefficient (r) and statistical significance values (P) are shown. Modified from study **V**.

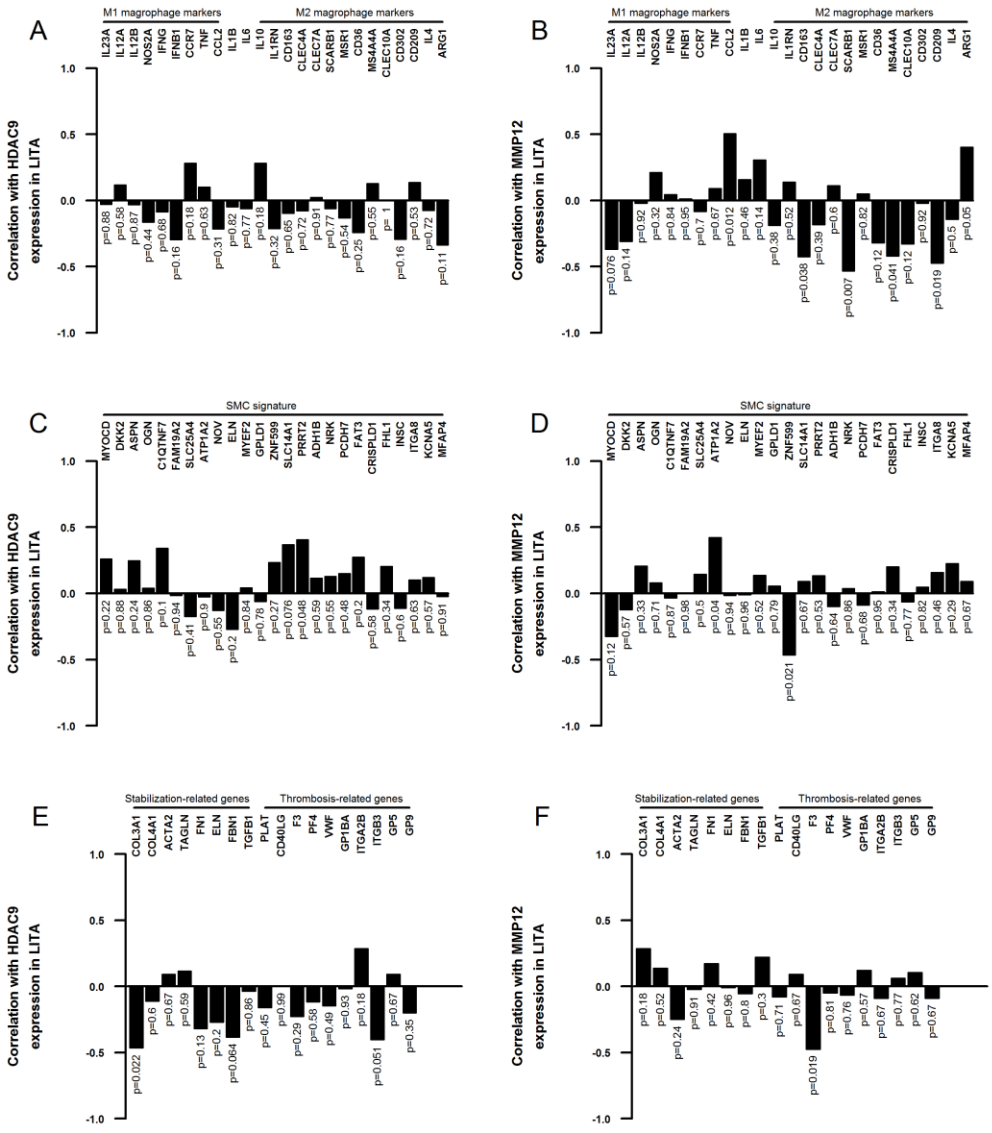



**Figure 23 –** Correlation among Histone deacetylase 9 (HDAC9) and matrix metalloproteinase 12 (MMP12) mRNA levels and established M1/M2 macrophage markers (panels A and B) in human atherosclerotic plaque, top 25 genes differentially expressed between regions of human atherosclerotic plaque enriched in smooth muscle cells (SMCs), and established markers for plaque stability (panels E and F) in left internal thoracic artery (LITA) controls. Spearman correlation coefficient (r) and statistical significance values (P) are shown. Modified from study **V**.

In carotid artery (Figure 19) MMP12 showed positive correlation ($p<0.05$) with M2 polarized macrophages (panel B, 22 per cent [2/9] of M1 markers, and 79 per cent [11/14] of M2 markers), negative correlation toward SMC markers (panel D, 72 per cent [18/25]), and no correlation with plaque stability (panel F) in the carotid artery plaques. There was a similar, but not as significant, pattern of correlation of MMP12 in the femoral artery (Figure 21) and abdominal aorta (Figure 22) plaques, and no correlation in the control arteries (Figure 23). There was no clear correlation pattern for HDAC9 expression with the markers in any artery beds (panels A, C, and E).

HDAC9 (Figure 20, panel A) and MMP12 (Figure 20, panel B) expressions increased significantly with plaque severity (AHA classification, ANOVA for trend, $p=0.00018$, and $p<0.0001$, respectively) in all artery beds.

# 6 Discussion

## 6.1 Genetics of LDL Oxidation (I)

Of the 328 SNPs associated with circulating oxLDL levels in a healthy Caucasian adult population (YFS), only one missense mutation leading to a proline-to-leucine interchange in ApoB (SNP rs676210, Pro2739Leu) remained significant in further analysis. ApoB is coded by the gene apolipoprotein B-100 precursor coding region (OMIM 107730) on chromosome 2. The association of apoB rs676210 with oxLDL was convincingly replicated in the LURIC and KORA cohorts. We also tested the association of rs676210 with cardiovascular end points in a meta-analysis of three independent clinical cohorts but did not find significant associations.

The YFS subjects had higher mean oxLDL levels than the LURIC subjects even though the LURIC subjects were older and had more co-morbidity. This could simply be due to different storage conditions or differences in the use of statin medication (Homma, Michishita et al. 2010) or the levels of oxidants or anti-oxidants. In LURIC, there was a higher proportion of statin users in comparison to the healthier YFS population (mean oxLDL levels were lower among statin users than among non-users, data not shown).

There are few previous reports about the genetics of LDL oxidation. The interleukin-1b (IL-1B) gene has been associated with oxLDL levels (Manica-Cattani, Duarte et al. 2012), supporting the role of genetic variation in LDL oxidation. This study is the first GWAS on circulating oxLDL. Employing a forward selection algorithm, performing haplotype analyses, assessing the damage-producing probability of the SNPs, and running the GWAS with top-SNP adjustment strongly suggest that apoB rs676210 is the most probable functional variant and the proxy for all of the found 328 associations.

Each LDL particle contains one apoB moiety. In the present study, the missense mutation of proline to leucine (rs676210, Pro2739Leu) in apoB increased plasma oxLDL levels in a stepwise manner in the genotype order of AA (Leu/Leu, the minor allele), GA (Pro/Leu), and GG (Pro/Pro). The results are supported by the fact that the variation in apoB probably changes the 3D structure of apoB (Kumar, Butcher et al. 2011) in a way that makes LDL less prone to oxidation in

homozygous apoB (Leu/Leu) minor allele carriers. It could be that LDL is more easily stuck in the proteoglycans in the matrix of the intima (Weber, Noels 2011), and therefore the LDL particle could be exposed to the oxidative processes longer.

There seem to be some pleiotropic effects for this variation as rs676210 was also associated with HDL cholesterol in all three cohorts and with triglyceride levels in YFS and LURIC. The oxLDL association, however, was adjusted for these, so the main effect seems to be on the oxLDL levels.

There are a few studies related to the genetic variation rs676210. In one previous study, the rs676210 minor allele (A) was associated with lower triglyceride, total cholesterol, and LDL cholesterol levels, and with higher HDL cholesterol levels, with a p value of $< 5x10\text{-}8$, in comparison to major allele (G) carriers (Teslovich, Musunuru et al. 2010). These results are in accordance with this study. In another earlier report, rs676210 was found to associate with VLDL-related fractions, triglycerides, and mean VLDL/LDL size (Chasman, Pare et al. 2009). In that study, minor allele carriers had larger VLDL/LDL particles and lower VLDL cholesterol and triglyceride concentrations. These findings are also in accordance with our results, showing a linear trend for the minor allele carriers in LURIC to have lower VLDL cholesterol levels ($p=0.0034$). VLDL cholesterol was not measured in YFS with this method.

Interestingly, the minor allele of rs676210 (A) has been linked to an improved response to fenofibrate treatment (Wojczynski, Gao et al. 2010)—the treatment was reported to lower triglyceride levels by 24.7%, 28.3%, and 34.5% according to the rs676210 genotypes GG, GA, and AA, respectively. The allele we found to decrease LDL oxidation also seems to improve the response to fenofibrate. Furthermore, minor allele carriers had lower triglyceride levels in YFS and LURIC when compared to major risk allele carriers. Moreover, fenofibrates have been shown to have effect on oxLDL (Hogue, Lamarche et al. 2008). There is no data on fenofibrate use in the studied cohorts.

## 6.2 Association of ApoB Pro2739Leu (rs676210) with CAD, MI, and Ischemic Stroke (I and II)

Since apoB Pro2739Leu (rs676210) associates strongly with oxLDL in both the young healthy population and elderly angiography patients, it could be used as a substitute for the effect of the lifetime risk of increased oxLDL levels on CAD, MI or ischemic stroke.

It seems, however, that the oxLDL levels associated apoB Pro2739Leu mutation is not associated with cardiovascular end points. This was observed across the cohorts studied in the combined meta-analysis. The association of serum oxLDL levels measured with Mercodia ELISA assay with CAD is controversial. Many smaller studies report oxLDL as a predictor of CAD, however, larger studies have been negative after adjusting for standard lipid variables (Wu, Willett et al. 2006). Our results are in line with these larger non-genetic studies by showing that serum oxLDL levels associated gene variant is not associated with the cardiovascular end points.

Association of rs676210 was also studied with the history of cerebrovascular disease events (TIA, or ischemic stroke) in LURIC and the result was sought to replicate the results in a large GWAS meta-analysis of history of ischaemic stroke and its pathophysiological subtypes. An association between the oxLDL-levels-increasing apolipoprotein-B missense mutation Pro2739Leu (rs676210) and history of cerebrovascular events was found in LURIC. The result was not, however, replicated in a large meta-analysis of ischaemic stroke cohorts.

The evident difference between the LURIC and WTCCC2 cohorts could explain the distinct association of rs676210 with cerebrovascular events in LURIC and ischaemic stroke in WTCCC2. The WTCCC2 cohorts were a collected as series of stroke patients enrolled to hospital, whereas LURIC is a series of patients referred for coronary angiography, most of whom were diagnosed with CAD. The expected association between oxLDL and ischaemic stroke would be for the LAA subtype of stroke due to the known involvement of oxLDL in the atherosclerotic process (Yla-Herttuala 1998). However, the absence of an association between rs676210 and CAD or MI and the controversial evidence of circulating oxLDL associating with CAD suggest that this would not explain the observed association of the SNP with cerebrovascular events.

One explanation could be found in the evidence of oxLDL binding to thrombocytes and increasing their adhesion to vessel walls (Stellos, Sauter et al. 2012). There is also evidence of higher serum oxLDL levels especially in CAD patients (Weinbrenner, Cladellas et al. 2003). This higher oxidative stress burden in CAD patients could be the key to explaining the increase in the risk of at least the LAA subtype of stroke in the cases of CAD among the LURIC patients. Unfortunately, we found no association for this subtype in WTCCC2. The cerebrovascular event definition in LURIC includes TIAs which also could be one explanation for the distinct association between the studies, as increased oxLDL levels in LURIC could also increase the risk for TIA.

The predictive value of cardiovascular disease by oxLDL biomarkers is considered controversial (Tsimikas 2006) and, accordingly, there is no clinical use for them at present. A genetic variant such as rs676210 which has a high impact on oxLDL levels would be an interesting alternative for oxLDL biomarkers, as it describes the lifetime risk for high oxLDL levels in comparison to measurements of oxLDL at one point in time. Unfortunately, the evidence of circulating oxLDL associating with ischaemic stroke is scarce (Uno, Kitazato et al. 2003, Serebruany, Sani et al. 2011, Vibo, Korv et al. 2007) and our current results are inconsistent. In our results, the cross-sectional circulating oxLDL levels did not associate with history of cerebrovascular events even though rs676210 did. Therefore, rs67210 could have predictive value for ischaemic stroke in a specific population of CAD risk patients. This needs to be verified in another similar sample.

## 6.3 HDAC9 and MMP12 in Atherosclerosis and Ischemic Stroke (III-V)

The results in study **III** show the HDAC9 7p21.1 locus, previously associated with large artery stroke, is also associated with asymptomatic carotid plaque and carotid IMT in community populations. This is consistent with a mechanism related with acceleration of the progression of atherosclerosis. HDAC9 is the most likely gene underlying this association. Consistent with this it was demonstrated that HDAC9 is expressed in VSMC and endothelium of healthy human adult large arteries, including both cerebral and systemic arteries. A similar pattern was obtained with two antibodies raised against two non-overlapping HDAC9-specific protein sequences. Consistent with a potential role in atherosclerosis we found increased expression of HDAC9 mRNA in carotid atherosclerotic plaques in TVS.

While canonical HDACs are ubiquitously expressed, Class IIa HDACs (including HDAC9) have more restricted expression. Expression in heart, pancreatic islets, spinal cord and brain of mouse embryos has been demonstrated, and human tissue lysates for HDAC9 mRNA show high expression in skeletal muscle and brain (Milde, Oehme et al. 2010). There are recent reports of HDAC9 protein expression using immunohistochemical labelling in cerebral medulloblastoma tumours (using one of the antibodies we used, ab59718) (Milde, Oehme et al. 2010) and in teeth, using a different antibody (Klinz, Korkmaz et al. 2012).

Since its discovery as a risk factor for stroke, a recent very large GWAS meta-analysis in 63,746 coronary artery disease (CAD) cases and 130,681 controls has found an association of the HDAC9 locus with CAD but with a much smaller effect size (CARDIoGRAMplusC4D Consortium, Deloukas et al. 2013); the odds ratio was 1.09, compared with an odds ratio of 1.42 with large artery ischaemic stroke in WTCCC2 (International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2) et al. 2012). This suggests this locus predisposes to large artery disease in the carotid arteries to a much greater extent than to CAD.

Interestingly we found HDAC9 mRNA expression was greater in carotid compared with femoral plaques. How such a risk factor would preferentially increase risk of carotid plaque is uncertain. One possible factor is flow dependent mechanisms dependent on local anatomy; local haemodynamic factors, and the

anatomy of the carotid bifurcation, are known to be related to early atherosclerotic changes (Sitzer, Puac et al. 2003).

Given the VSMC expression of HDAC9, increased risk of large vessel disease could be via promotion of atherosclerosis as a consequence of HDAC9 mediated increased VSMC proliferation – an action impeded by HDAC9 inhibition in vitro (Okamoto, Fujioka et al. 2006).

In study **IV**, we found novel MMP12 locus associated with ischemic stroke. Moreover, it was shown in our results that MMP12 is overexpressed in atherosclerotic plaques compared to atherosclerosis free controls. MMP12, also known as macrophage metalloelastase, belongs to the MMP family of proteases which are capable of degrading extracellular matrix proteins, and have an important role in atherosclerosis (Halpert, Sires et al. 1996). MMP12 has also earlier been shown to be overexpressed in atherosclerotic plaques (Levula, Oksala et al. 2012) and has been localized to macrophages at the border of the lipid core (Halpert, Sires et al. 1996), and is suggested to be involved in late stage plaque instability (Morgan, Rerkasem et al. 2004, Yamada, Wang et al. 2008). In accordance with our results, Chehaibi and others showed association of MMP12 polymorphism with ischemic stroke in diabetic patients (Chehaibi, Hrira et al. 2014).

Furthermore, we demonstrated in study **V** that HDAC9 and MMP12 expressions correlate with each other and associate with plaque severity according to AHA classification. The results also suggest anti-inflammatory M2 macrophages as a possible source of expression for MMP12 in advanced human atherosclerotic plaques.

It has been shown that HDAC inhibition affects MMP expression (Young, Lakey et al. 2005) in chondrosarcoma cell line. In our results HDAC9 expression correlated negatively with MMP12 expression, which implies that HDAC9 could have effect on MMP12 activity in atherosclerotic plaque as well. We showed that MMP12 correlates positively with the M2 signature and negatively with the SMC signature in plaques, with no correlation with these in the control arteries. We also showed that MMP12 expression is increased by the severity of the plaque. These results are in line with the previous studies which localize the expression of MMP12 in advanced atherosclerotic plaques (Morgan, Rerkasem et al. 2004, Yamada, Wang et al. 2008), and suggest M2 macrophages as a source of expression of MMP12 in adipose tissue of mice (Lee, Pamir et al. 2014). Inflammatory M1 macrophages and M1-associated cytokines are considered to be involved in the development of the vulnerable plaques, whereas M2 macrophages are considered to be protective through paracrine anti-inflammatory effect which they exert on

M1 macrophages (Salagianni, Galani et al. 2012). In our results, MMP12 expression associated mostly with M2 signature, and not with the plaque instability markers, which suggest that MMP12 exerts anti-inflammatory effect in the plaque. This is supported by results, where MMP12 has been shown to exert anti-inflammatory effect also in synovitis (Bellac, Dufour et al. 2014). Since MMP12 variation associates with increased risk of ischemic stroke (original communication **IV**), it might be that this variation inactivates the protective role of MMP12. However, it has also been shown that MMP12 initiates atherosclerosis and stimulates the progression of fatty streaks to fibrous plaques in transgenic rabbits (Yamada, Wang et al. 2008), which suggest opposite effect for MMP12. It has been suggested that selective inhibition of MMP12 could be used as therapeutic application in atherosclerosis (Devel, Garcia et al. 2010). If MMP12 was protective, the inhibition of MMP12 might not be effective.

There is a suggestive positive correlation of HDAC9 expression toward M2 macrophages in the carotid artery. This result is in line with previous studies, where the deletion of HDAC9 in murine models showed an attenuation of atherosclerosis with minimal effect on lipid concentrations (Cao, Rong et al. 2014). In that study, macrophages polarized toward M2 when HDAC9 was deleted systematically or specifically in macrophages. The effect on the attenuation of atherosclerosis has also been shown to be allele specific (Azghandi, Prell et al. 2014). It might be that the HDAC9 expressed by M2 macrophages could exert negative feedback preventing further M2 polarization. Cao and others showed that the attenuation of atherosclerosis was location dependent in the aorta (Cao, Rong et al. 2014). In our results there was suggestive, but not significant, reverse pattern of expression when comparing carotid and femoral plaques, which also suggest distinct effect depending on artery site. This is also supported by our result where the correlation between HDAC9 and MMP12 was only significant in carotid artery plaques.

Interestingly, there was a significant interaction with HDAC9 and MMP12 association in relation to case control status. In controls HDAC9 correlated inversely (-0.44) with MMP12 expression while in atherosclerotic plaque in the carotid artery the correlation was in opposite direction. Possibly, HDAC9 related MMP12 methylation regulation (Young, Lakey et al. 2005) could be able to regulate MMP12 gene expression in healthy tissue while this regulation does not work in inflamed atherosclerotic tissue.

Taken together the results are consistent with the HDAC9 locus acting as a risk factor for atherosclerosis. Such an association with large artery stroke could be via increasing plaque development, or by mechanisms which results in plaque

instability (Badimon, Vilahur 2014) and increase the risk of subsequent thromboembolism, the major cause of stroke in large artery disease. The association with asymptomatic carotid plaque, plaques which have not yet become unstable and symptomatic, would support the former mechanism. An association with carotid IMT was also found which would be consistent with increased risk occurring at the earlier stages of plaque formation. Increased carotid IMT is believed to occur with both early atherosclerosis and also vascular remodelling (Mathiesen, Johnsen 2009). Our results also show that expression of HDAC9 and MMP12 correlate with each other positively in carotid plaques and negatively in control samples, associate with plaque severity, and suggest anti-inflammatory M2 macrophages as a possible source for expression of MMP12 in advanced human atherosclerotic plaques.

## 6.4 Limitations and Strengths of the Study

In all studies the study subjects were Caucasians and the results cannot be generalized to other populations.

In original communication **I**, the main reason for not finding additional signals could be the small number of study subjects. In the largest GWASes hundreds of thousands of individuals are required to significantly find the weakest signals. An obvious strength of the study is the replication of the major results in multiple independent cohorts. There were significant differences in sex distributions and mean ages between the cohorts. However, the association of rs676210 with oxLDL concentration remained significant in spite of these and other differences in the background populations.

The association of rs676210 with LDL oxidation is convincing. The oxidation of LDL was, however, measured by structural changes in the apoB protein and does not fully account for the lipid component of these apoB-containing particles. To confirm that our results are not due to the possibility of the Pro2739Leu substitution altering the binding of the used monoclonal antibody to apoB, we also applied a second, apoB-epitope-structure-independent assay to assess the effect of the SNP on the LDL oxidation, with parallel results. In this independent assay, the SNP rs676210 was significantly associated with LDL diene conjugation. These results put together show that our findings are not due to the oxLDL assay employed and imply a true biological function for rs676210 in LDL oxidation.

The step-wise algorithm and top-SNP-adjusted GWAS strongly suggest that rs676210 is the proxy SNP. We also carried out manual haplotype analyses, arriving at the same conclusion of rs676210 as the implied proxy. This information, together with the predicted damaging change caused in apoB (proline-to-leucine interchange), indicates that out of the available SNPs, this is the most likely functional top-SNP candidate. The SNP in perfect LD with rs676210 was predicted to be benign. Two other nearby missense mutations were also predicted to be damaging but showed no independent effect in the haplotype analysis nor with the forward-selection algorithm. Other associated missense-mutations were predicted not to be damaging. These results strongly suggest that rs676210 is the true functional variant. However, it is important to note that the true proxy cannot be found with 100% certainty by means of bioinformatic methods, and, therefore, we performed the analyses with the most probable one.

In comparison, the cohorts used for the clinical endpoint association assessments were quite similar. The current analyses related to oxLDL did not include the FINCAVAS or ANGES study populations because oxLDL was not measured in them. The clinical implications of rs676210 leading to Pro2739Leu missense mutation require further investigation—however, we did not find significant associations in a meta-analysis of three independent studies.

In original communication **II**, the obvious limitation of this study is the lack of replication of the association of rs676210 with ischaemic stroke. To verify the association, it needs to be replicated in a sample similar to LURIC. Moreover, the definitions for cerebrovascular disease differ between LURIC and WTCCC2. The cerebrovascular event definition in LURIC includes all TIAs and strokes. The absence of replication of this association in a general ischemic stroke setting means that there is not clear clinical use for the studied SNP per se in, for example, the prediction of stroke in the general population. However, if the result could be replicated in another cohort of angiography patients, there might be some predictive value for this SNP in patients with suspected CAD. LURIC does not have stroke subtype definition available to verify whether the association is with a specific subtype of ischaemic stroke.

In original communication **III**, there are a number of potential limitations to the study. Not all patients who had IMT measured also had carotid plaque measured. However this would have tended to reduce power to detect any statistical association with plaque, and we found such an association. The CHARGE consortium includes a number of different populations which introduces heterogeneity; therefore we analysed using a meta-analysis approach and

the associations we found were consistent across almost all populations. In the mRNA expression studies there were relatively small sample sizes, although the upregulation of HDAC9 was still detected in atherosclerotic plaque.

In **IV**, we used imputed data from the Immunochip platform, meaning we only had access to ~40% of the genome across all centres. Secondly, cases were drawn from a number of international centres, meaning that despite efforts to standardize phenotyping, we cannot rule out differences in screening and clinical ascertainment.

In original communication **V**, even though the control arteries were microscopically free of atherosclerosis, the control subjects have atherosclerosis at least in cardiac arteries since they are undergoing coronary artery bypass surgery. The bioinformatic method used in the study is only suggestive tool for localization and for example confocal microscopy or double staining immunohistochemistry would be needed to confirm the results. Moreover, the studied plaques are mostly at advanced stage and the role of the studied genes in the early process of atherosclerosis is not known. The advantage of the used method is that in comparison to other methods it a cost effective way to give rough estimation of localization of expression of genes.

## 6.5 Future Prospects

The Pro2379Leu missense mutation rs676210 might have use as a determinant of life time exposure to oxLDL since it associated convincingly with circulating oxLDL-concentration in a young healthy population, and also in a sample of patients undergoing coronary angiography. However, since the mutation did not associate with CAD or MI it might not have clinical significance in their risk assessment. We found association of Pro2739Leu with cerebrovascular disease events in the population of angiography patients, however, the result did not replicate in a large meta-analysis of ischemic stroke patients. The variant might have prognostic value for ischemic stroke in a specific subpopulation.

Interestingly, it has been shown that rs676210 modifies the response to cholesterol lowering fenofibrate treatment (Wojczynski, Gao et al. 2010). Therefore, Pro2739Leu might have use in pharmacogenomics in determining patients who benefit the most from fenofibrate treatment. Moreover, there might be other gene-by-environment interactions for this variant, such as smoking which increases oxidative stress. In further studies with larger populations it should be determined whether Pro2739Leu could predict CAD, MI or ischemic stroke in smokers only.

HDAC9 and MMP12 are novel susceptibility genes for ischemic stroke. Their variation could be use as part of ischemic stroke risk prediction in the future. Moreover, especially HDAC9 could function as a therapeutical target in prevention of atherosclerosis, since specific HDAC-inhibitors are being developed and there is evidence of non-specific inhibitors, such as natrium valproate, reducing atherosclerosis in animal models (Dregan, Charlton et al. 2014). Furthermore, selective MMP12 inhibition has been suggested as a therapeutic application for atherosclerosis (Devel, Garcia et al. 2010) acting possibly via stabilization of vulnerable plaques.

These novel genetic findings also give new information of the pathophysiology of atherosclerosis. It is interesting, that the effect of HDAC9 seems to be especially for the development of atherosclerosis in the carotid artery. Further studies are needed to determine which factors cause the location specific action for this gene in the atherosclerotic process.

# 7 Summary and Conclusions

The present study was conducted to analyse the genetic background of human atherosclerosis plaques, cardiovascular diseases especially focusing on ischemic stroke and to investigate the role of oxLDL related gene variation, MMP12 and HSCD9 genes in their pathogenesis. Figure 24 bellow summarizes the major results of this study.

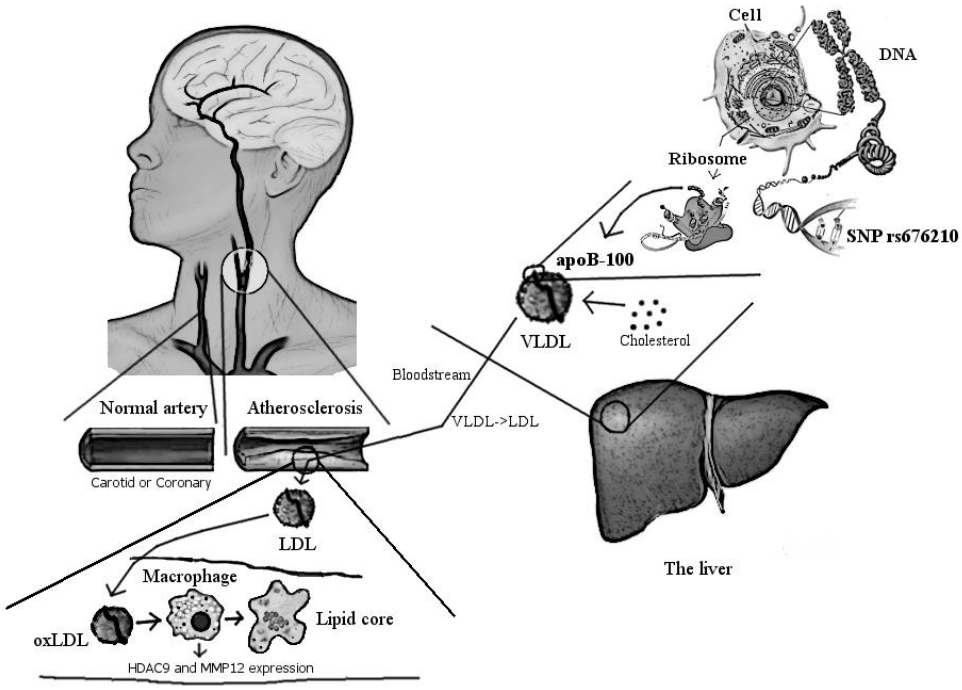



**Figure 24 –** Summary of the results. Novel SNP rs676210 causing Pro2739Leu missense mutation in apoB affecting the oxidation of LDL was found in study **I**. Clear association of rs676210 with CAD or MI was not found in **I**. In **II** rs676210 associated with cerebrovascular disease events in patients undergoing coronary angiography. In study **III** ischemic stroke associated HDAC9 was found to be acting through promoting carotid atherosclerosis. In study

**IV** we found novel MMP12 locus associated with ischemic stroke, and found that MMP12 is overexpressed in atherosclerotic plaque compared to atherosclerosis free control vessels. In **V** HDAC9 and MMP12 expression in human atherosclerotic plaques were further studied and M2 macrophages pinpointed as a possible source for MMP12 expression.

The main findings of the study were (Figure 24):

1. In discovery GWAS using YFS samples and related replication analysis with two independent cohorts (LURIS and KORA cohorts) we found a novel genetic marker SNP rs676210 for oxLDL concentration (in YFS $p=4.3x10^{-136}$, effect size 13.2 U/l per allele). The variant rs676210 causes missense mutations i.e., change of proline to leucine at position 2739 in apoB. This mutation was predicted to be functional in the analysis performed with the PolyPhen-2 software (**I**).

2. The apoB (Pro2739Leu) mutation proofed to be important factor regulating oxLDL levels but not a major genetic factor for coronary artery disease or myocardial infarction (**I**).

3. In **II**, the functional apoB (Pro2739Leu) mutation (rsr67210) associated with cerebrovascular disease events in patients undergoing coronary angiography, however, the result did not replicate in larger meta-analysis of ischemic stroke patients participating in WTCCC2 consortium.

4. In **III**, both HDAC9 GWAS top-SNPs rs11984041 and rs2107595, associated significantly with carotid artery intima-media thickness measured by ultrasound and presence of atherosclerotic plaques (p-values for rs2107595 were p=0.0018 and p=0.0022 and for rs11984041 p=0.00391 and p=0.00425 for cIMT and plaques, respectively). HDAC9 was showed to be acting in the risk of stroke via promoting atherosclerosis in the carotid artery.

5. In **IV,** MMP12 locus was found to be associated with ischemic stroke and atherosclerosis using age-of-onset informed GWAS.

6. In **V**, it was shown in TVS samples that HDAC9 and MMP12 expressions associate with plaque severity, i.e., HDAC9 and MMP12 expressions were

increased with plaque severity ($p=0.00018$, and $p<0.0001$, for trend respectively) in all artery beds. HDAC9 and MMP12 correlated with each other, and the results suggested anti-inflammatory M2 macrophages as a possible source of expression for MMP12 in advanced human atherosclerotic plaques.

These results give novel insight into the genetic background of lipoprotein oxidation, atherosclerosis and ischemic stroke.

# 8 Acknowledgements

This Doctoral thesis was done at the Department of Clinical Chemistry, University of Tampere School of Medicine, and the Fimlab laboratories, Tampere, during the years 2010-2015.

I want to thank my teacher and supervisor Professor Terho Lehtimäki who presented the initial idea of the study and introduced me to biochemical and genetic research of atherosclerosis and cardiovascular research. I am grateful to all of my colleagues at our research group at Fimlab laboratories for all the help received. Especially Ilkka Seppälä, Leo-Pekka Lyytikäinen, Emma Raitoharju, Nina Mononen, Niku Oksala, Jussi Hernesniemi, and Professor Reijo Laaksonen were essential during the thesis work. I also want to thank all my co-authors and the people involved with the Young Finns and FINCAVAS studies for creating an excellent study population which will help us further understand the complex nature of atherosclerosis in the following decades. Especially Academy professor Olli Raitakari and Professor Jorma Viikari and Mika Kähönen are gratefully acknowledged.

This work would not have been possible without Ludwigshafen Risk and Cardiovascular Health (LURIC) study and I want to thank Professor Winfried März and Doctor Marcus Kleber for the possibility of including their excellent study population.

Working with professor Hugh Markus and Doctor Matthew Traylor made the latter part of the thesis possible and I want to thank for the fruitful collaboration.

Finally, I have huge gratitude for having loving and supporting family and friends. Without them making of this thesis would not have been possible.

The thesis was funded by grants awarded to Kari-Matti Mäkelä by Kliinisen Kemian Tutkimussäätiö, Yrjö Jahnsson foundation, Laboratoriolääketieteen edistämissäätiö, Oskar Öflund foundation, Ida Montini foundation, Ahvenainen foundation, Kalle Kaiharin sydäntutkimuksen apuraha, Urho Känkänen foundation and Orion-Farmos research foundation.

This work received funding from the European Union 7th Framework Programme funding for the AtheroRemo project (201668), the Tampere University Hospital Medical Fund grant 9N035 and X51001, the Emil Aaltonen Foundation

(T.L., N.O.), the Finnish Foundation of Cardiovascular Research, and the Tampere Tuberculosis Foundation (T.L.). The expert technical assistance in data management and statistical analyses by Irina Lisinen and Ville Aalto in YFS are gratefully acknowledged.

We thank the participants of the LURIC study and LURIC study team who were either temporarily or permanently involved in patient recruitment as well as sample and data handling, in addition to the laboratory staff at the Ludwigshafen General Hospital and the Universities of Freiburg, Ulm, and Graz, Germany. We would like to thank Gerlinde Trischler (University of Ulm) for her expert technical assistance in KORA study.

The Cardiovascular Risk in Young Finns Study has been financially supported by the Academy of Finland (grants 126925, 121584, 124282, 129378 [Salve], 117787 [Gendi], 41071 [Skidi], and 286284); the Social Insurance Institution of Finland; the Kuopio, Tampere, and Turku University Hospital Medical Funds (grant 9M048 and 9N035 for TeLeht); the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Finnish Foundation of Cardiovascular Research; the Finnish Cultural Foundation; as well as the Tampere Tuberculosis Foundation and the Emil Aaltonen Foundation (T.L).

The MONICA/KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, and supported by grants from the German Federal Ministry of Education and Research (BMBF). A part of the work was financed by the German National Genome Research Network (NGFNPlus, project number 01GS0834), the German Research Foundation (TH-784/2-2), and additional funds from the University of Ulm. Furthermore, the research was supported within the Munich Center of Health Sciences (MC Health) as a part of the LMU innovative.

LURIC received funding through the 6th Framework Programme (integrated project Bloodomics, grant LSHM-CT-2004-503485) and the 7th Framework Programme (integrated project AtheroRemo, grant agreement number 201668 and RiskyCAD, grant agreement number 305739) of the European Union. Mercodia AB, Uppsala, provided the reagents for the determination of oxLDL free of charge, but did not assume any other role in the conducting of this study.

FINCAVAS and ANGES have been financially supported by the Tampere University Hospital Medical Fund (grant 9M048 and 9N035 for TeLeht), the Tampere Tuberculosis Foundation, and the Emil Aaltonen Foundation (T.L.).

The Edinburgh Stroke Study was supported by the Wellcome Trust (clinician scientist award to Dr. Sudlow) and the Binks Trust (research fellowship to Dr.

Jackson). The collection of the Munich cases and data analysis were supported by the Vascular Dementia Research Foundation. The Oxford Vascular Study is funded by the Medical Research Council, the Stroke Association, the Dunhill Medical Trust, the National Institute of Health Research (NIHR) and the NIHR Biomedical Research Centre, Oxford. Segal, Burgess and Poole are funded by the NIHR Biomedical Research Centre, Oxford. The collection of some of the St George's stroke cohort was supported by project grant support from the Stroke Association. The Aberdeen cases were collected with support from the Stroke Association and the Scottish Chief Scientist's Office.

# 9 References


ABBOUD, S., KARHUNEN, P.J., LUTJOHANN, D., GOEBELER, S., LUOTO, T., FRIEDRICHS, S., LEHTIMAKI, T., PANDOLFO, M. and LAAKSONEN, R., 2007. Proprotein convertase subtilisin/kexin type 9 (PCSK9) gene is a risk factor of large-vessel atherosclerosis stroke. *PloS one,* **2**(10), pp. e1043.

ADAMS, H.P.,Jr, BENDIXEN, B.H., KAPPELLE, L.J., BILLER, J., LOVE, B.B., GORDON, D.L. and MARSH, E.E.,3rd, 1993. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke; a journal of cerebral circulation,* **24**(1), pp. 35-41.

ADZHUBEI, I.A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V.E., GERASIMOVA, A., BORK, P., KONDRASHOV, A.S. and SUNYAEV, S.R., 2010. A method and server for predicting damaging missense mutations. *Nature methods,* **7**(4), pp. 248-249.

AHOTUPA, M., MARNIEMI, J., LEHTIMAKI, T., TALVINEN, K., RAITAKARI, O.T., VASANKARI, T., VIIKARI, J., LUOMA, J. and YLA-HERTTUALA, S., 1998. Baseline diene conjugation in LDL lipids as a direct measure of in vivo LDL oxidation. *Clinical biochemistry,* **31**(4), pp. 257-261.

ARROWSMITH, C.H., BOUNTRA, C., FISH, P.V., LEE, K. and SCHAPIRA, M., 2012. Epigenetic protein families: a new frontier for drug discovery. *Nature reviews.Drug discovery,* **11**(5), pp. 384-400.

AULCHENKO, Y.S., STRUCHALIN, M.V. and VAN DUIJN, C.M., 2010. ProbABEL package for genome-wide association analysis of imputed data. *BMC bioinformatics,* **11**, pp. 134-2105-11-134.

AZGHANDI, S., PRELL, C., VAN DER LAAN, S.W., SCHNEIDER, M., MALIK, R., BERER, K., GERDES, N., PASTERKAMP, G., WEBER, C., HAFFNER, C. and DICHGANS, M., 2014. Deficiency of the Stroke Relevant HDAC9 Gene Attenuates Atherosclerosis in Accord With Allele-Specific Effects at 7p21.1. *Stroke; a journal of cerebral circulation,* .

BADIMON, L. and VILAHUR, G., 2014. Thrombosis formation on atherosclerotic lesions and plaque rupture. *Journal of internal medicine,* **276**(6), pp. 618-632.

BARREIRO, L.B., LAVAL, G., QUACH, H., PATIN, E. and QUINTANA-MURCI, L., 2008. Natural selection has driven population differentiation in modern humans. *Nature genetics,* **40**(3), pp. 340-345.

BARRETT, J.C., FRY, B., MALLER, J. and DALY, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England),* **21**(2), pp. 263-265.

BELLAC, C.L., DUFOUR, A., KRISINGER, M.J., LOONCHANTA, A., STARR, A.E., AUF DEM KELLER, U., LANGE, P.F., GOEBELER, V., KAPPELHOFF, R., BUTLER, G.S., BURTNICK, L.D., CONWAY, E.M., ROBERTS, C.R. and

OVERALL, C.M., 2014. Macrophage matrix metalloproteinase-12 dampens inflammation and neutrophil influx in arthritis. *Cell reports,* **9**(2), pp. 618-632.

BEOHAR, N., FLAHERTY, J.D., DAVIDSON, C.J., MAYNARD, R.C., ROBBINS, J.D., SHAH, A.P., CHOI, J.W., MACDONALD, L.A., JORGENSEN, J.P., PINTO, J.V., CHANDRA, S., KLAUS, H.M., WANG, N.C., HARRIS, K.R., DECKER, R. and BONOW, R.O., 2004. Antirestenotic effects of a locally delivered caspase inhibitor in a balloon injury model. *Circulation,* **109**(1), pp. 108-113.

BIS, J.C., KAVOUSI, M., FRANCESCHINI, N., ISAACS, A., ABECASIS, G.R., SCHMINKE, U., POST, W.S., SMITH, A.V., CUPPLES, L.A., MARKUS, H.S., SCHMIDT, R., HUFFMAN, J.E., LEHTIMAKI, T., BAUMERT, J., MUNZEL, T., HECKBERT, S.R., DEHGHAN, A., NORTH, K., OOSTRA, B., BEVAN, S., STOEGERER, E.M., HAYWARD, C., RAITAKARI, O., MEISINGER, C., SCHILLERT, A., SANNA, S., VOLZKE, H., CHENG, Y.C., THORSSON, B., FOX, C.S., RICE, K., RIVADENEIRA, F., NAMBI, V., HALPERIN, E., PETROVIC, K.E., PELTONEN, L., WICHMANN, H.E., SCHNABEL, R.B., DORR, M., PARSA, A., ASPELUND, T., DEMISSIE, S., KATHIRESAN, S., REILLY, M.P., TAYLOR, K., UITTERLINDEN, A., COUPER, D.J., SITZER, M., KAHONEN, M., ILLIG, T., WILD, P.S., ORRU, M., LUDEMANN, J., SHULDINER, A.R., EIRIKSDOTTIR, G., WHITE, C.C., ROTTER, J.I., HOFMAN, A., SEISSLER, J., ZELLER, T., USALA, G., ERNST, F., LAUNER, L.J., D'AGOSTINO RB, S., O'LEARY, D.H., BALLANTYNE, C., THIERY, J., ZIEGLER, A., LAKATTA, E.G., CHILUKOTI, R.K., HARRIS, T.B., WOLF, P.A., PSATY, B.M., POLAK, J.F., LI, X., RATHMANN, W., UDA, M., BOERWINKLE, E., KLOPP, N., SCHMIDT, H., WILSON, J.F., VIIKARI, J., KOENIG, W., BLANKENBERG, S., NEWMAN, A.B., WITTEMAN, J., HEISS, G., DUIJN, C., SCUTERI, A., HOMUTH, G., MITCHELL, B.D., GUDNASON, V., O'DONNELL, C.J. and CARDIOGRAM CONSORTIUM, 2011. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nature genetics,* **43**(10), pp. 940-947.

BOBRYSHEV, Y.V. and LORD, R.S., 1995. Ultrastructural recognition of cells with dendritic cell morphology in human aortic intima. Contacting interactions of Vascular Dendritic Cells in athero-resistant and athero-prone areas of the normal aorta. *Archives of Histology and Cytology,* **58**(3), pp. 307-322.

BOYLE, A.P., HONG, E.L., HARIHARAN, M., CHENG, Y., SCHAUB, M.A., KASOWSKI, M., KARCZEWSKI, K.J., PARK, J., HITZ, B.C., WENG, S., CHERRY, J.M. and SNYDER, M., 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research,* **22**(9), pp. 1790-1797.

BROWN, M.S. and GOLDSTEIN, J.L., 1983. Lipoprotein receptors in the liver. Control signals for plasma cholesterol traffic. *The Journal of clinical investigation,* **72**(3), pp. 743-747.

BURKE, A.P., FARB, A., MALCOM, G.T., LIANG, Y., SMIALEK, J.E. and VIRMANI, R., 1999. Plaque rupture and sudden death related to exertion in men with coronary artery disease. *Jama,* **281**(10), pp. 921-926.

CAO, Q., RONG, S., REPA, J.J., ST CLAIR, R., PARKS, J.S. and MISHRA, N., 2014. Histone deacetylase 9 represses cholesterol efflux and alternatively activated macrophages in atherosclerosis development. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **34**(9), pp. 1871-1879.

CARDIOGRAMPLUSC4D CONSORTIUM, DELOUKAS, P., KANONI, S., WILLENBORG, C., FARRALL, M., ASSIMES, T.L., THOMPSON, J.R., INGELSSON, E., SALEHEEN, D., ERDMANN, J., GOLDSTEIN, B.A., STIRRUPS, K., KONIG, I.R., CAZIER, J.B., JOHANSSON, A., HALL, A.S., LEE, J.Y., WILLER, C.J., CHAMBERS, J.C., ESKO, T., FOLKERSEN, L., GOEL, A., GRUNDBERG, E., HAVULINNA, A.S., HO, W.K., HOPEWELL, J.C., ERIKSSON, N., KLEBER, M.E., KRISTIANSSON, K., LUNDMARK, P., LYYTIKAINEN, L.P., RAFELT, S., SHUNGIN, D., STRAWBRIDGE, R.J., THORLEIFSSON, G., TIKKANEN, E., VAN ZUYDAM, N., VOIGHT, B.F., WAITE, L.L., ZHANG, W., ZIEGLER, A., ABSHER, D., ALTSHULER, D., BALMFORTH, A.J., BARROSO, I., BRAUND, P.S., BURGDORF, C., CLAUDI-BOEHM, S., COX, D., DIMITRIOU, M., DO, R., DIAGRAM CONSORTIUM, CARDIOGENICS CONSORTIUM, DONEY, A.S., EL MOKHTARI, N., ERIKSSON, P., FISCHER, K., FONTANILLAS, P., FRANCO-CERECEDA, A., GIGANTE, B., GROOP, L., GUSTAFSSON, S., HAGER, J., HALLMANS, G., HAN, B.G., HUNT, S.E., KANG, H.M., ILLIG, T., KESSLER, T., KNOWLES, J.W., KOLOVOU, G., KUUSISTO, J., LANGENBERG, C., LANGFORD, C., LEANDER, K., LOKKI, M.L., LUNDMARK, A., MCCARTHY, M.I., MEISINGER, C., MELANDER, O., MIHAILOV, E., MAOUCHE, S., MORRIS, A.D., MULLER-NURASYID, M., MUTHER CONSORTIUM, NIKUS, K., PEDEN, J.F., RAYNER, N.W., RASHEED, A., ROSINGER, S., RUBIN, D., RUMPF, M.P., SCHAFER, A., SIVANANTHAN, M., SONG, C., STEWART, A.F., TAN, S.T., THORGEIRSSON, G., VAN DER SCHOOT, C.E., WAGNER, P.J., WELLCOME TRUST CASE CONTROL CONSORTIUM, WELLS, G.A., WILD, P.S., YANG, T.P., AMOUYEL, P., ARVEILER, D., BASART, H., BOEHNKE, M., BOERWINKLE, E., BRAMBILLA, P., CAMBIEN, F., CUPPLES, A.L., DE FAIRE, U., DEHGHAN, A., DIEMERT, P., EPSTEIN, S.E., EVANS, A., FERRARIO, M.M., FERRIERES, J., GAUGUIER, D., GO, A.S., GOODALL, A.H., GUDNASON, V., HAZEN, S.L., HOLM, H., IRIBARREN, C., JANG, Y., KAHONEN, M., KEE, F., KIM, H.S., KLOPP, N., KOENIG, W., KRATZER, W., KUULASMAA, K., LAAKSO, M., LAAKSONEN, R., LEE, J.Y., LIND, L., OUWEHAND, W.H., PARISH, S., PARK, J.E., PEDERSEN, N.L., PETERS, A., QUERTERMOUS, T., RADER, D.J., SALOMAA, V., SCHADT, E., SHAH, S.H., SINISALO, J., STARK, K., STEFANSSON, K., TREGOUET, D.A., VIRTAMO, J., WALLENTIN, L., WAREHAM, N., ZIMMERMANN, M.E., NIEMINEN, M.S., HENGSTENBERG, C., SANDHU, M.S., PASTINEN, T., SYVANEN, A.C., HOVINGH, G.K., DEDOUSSIS, G., FRANKS, P.W., LEHTIMAKI, T., METSPALU, A., ZALLOUA, P.A., SIEGBAHN, A., SCHREIBER, S., RIPATTI, S., BLANKENBERG, S.S., PEROLA, M., CLARKE, R., BOEHM, B.O., O'DONNELL, C., REILLY, M.P., MARZ, W., COLLINS, R., KATHIRESAN, S., HAMSTEN, A., KOONER, J.S., THORSTEINSDOTTIR, U., DANESH, J., PALMER, C.N., ROBERTS, R., WATKINS, H., SCHUNKERT, H. and SAMANI, N.J., 2013. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics,* **45**(1), pp. 25-33.

CHAMORRO, A., MEISEL, A., PLANAS, A.M., URRA, X., VAN DE BEEK, D. and VELTKAMP, R., 2012. The immunology of acute stroke. *Nature reviews.Neurology,* **8**(7), pp. 401-410.

CHASMAN, D.I., PARE, G., MORA, S., HOPEWELL, J.C., PELOSO, G., CLARKE, R., CUPPLES, L.A., HAMSTEN, A., KATHIRESAN, S., MALARSTIG, A., ORDOVAS, J.M., RIPATTI, S., PARKER, A.N., MILETICH, J.P. and RIDKER, P.M., 2009. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS genetics,* **5**(11), pp. e1000730.

CHEHAIBI, K., HRIRA, M.Y., NOUIRA, S., MAATOUK, F., BEN HAMDA, K. and SLIMANE, M.N., 2014. Matrix metalloproteinase-1 and matrix metalloproteinase-12 gene polymorphisms and the risk of ischemic stroke in a Tunisian population. *Journal of the neurological sciences,* **342**(1-2), pp. 107-113.

CIFU, D.X. and STEWART, D.G., 1999. Factors affecting functional outcome after stroke: a critical review of rehabilitation interventions. *Archives of Physical Medicine and Rehabilitation,* **80**(5 Suppl 1), pp. S35-9.

COLE, J.E., NAVIN, T.J., CROSS, A.J., GODDARD, M.E., ALEXOPOULOU, L., MITRA, A.T., DAVIES, A.H., FLAVELL, R.A., FELDMANN, M. and MONACO, C., 2011. Unexpected protective role for Toll-like receptor 3 in the arterial wall. *Proceedings of the National Academy of Sciences of the United States of America,* **108**(6), pp. 2372-2377.

CURRENT CARE GUIDELINE, 2014. Update on current care guideline: acute coronary syndrome--unstable angina and myocardial infarction without ST elevation. *Duodecim; laaketieteellinen aikakauskirja,* **130**(17), pp. 1764-1766.

CYBULSKY, M.I. and GIMBRONE, M.A.,Jr, 1991. Endothelial expression of a mononuclear leukocyte adhesion molecule during atherogenesis. *Science (New York, N.Y.),* **251**(4995), pp. 788-791.

DAI, G., KAAZEMPUR-MOFRAD, M.R., NATARAJAN, S., ZHANG, Y., VAUGHN, S., BLACKMAN, B.R., KAMM, R.D., GARCIA-CARDENA, G. and GIMBRONE, M.A.,Jr, 2004. Distinct endothelial phenotypes evoked by arterial waveforms derived from atherosclerosis-susceptible and -resistant regions of human vasculature. *Proceedings of the National Academy of Sciences of the United States of America,* **101**(41), pp. 14871-14876.

DEVEL, L., GARCIA, S., CZARNY, B., BEAU, F., LAJEUNESSE, E., VERA, L., GEORGIADIS, D., STURA, E. and DIVE, V., 2010. Insights from selective non-phosphinic inhibitors of MMP-12 tailored to fit with an S1' loop canonical conformation. *The Journal of biological chemistry,* **285**(46), pp. 35900-35909.

DICHGANS, M., MALIK, R., KONIG, I.R., ROSAND, J., CLARKE, R., GRETARSDOTTIR, S., THORLEIFSSON, G., MITCHELL, B.D., ASSIMES, T.L., LEVI, C., O'DONNELL, C.J., FORNAGE, M., THORSTEINSDOTTIR, U., PSATY, B.M., HENGSTENBERG, C., SESHADRI, S., ERDMANN, J., BIS, J.C., PETERS, A., BONCORAGLIO, G.B., MARZ, W., MESCHIA, J.F., KATHIRESAN, S., IKRAM, M.A., MCPHERSON, R., STEFANSSON, K., SUDLOW, C., REILLY, M.P., THOMPSON, J.R., SHARMA, P., HOPEWELL, J.C., CHAMBERS, J.C., WATKINS, H., ROTHWELL, P.M., ROBERTS, R., MARKUS, H.S., SAMANI, N.J., FARRALL, M., SCHUNKERT, H., METASTROKE CONSORTIUM, CARDIOGRAM CONSORTIUM, C4D CONSORTIUM and INTERNATIONAL STROKE GENETICS CONSORTIUM, 2014. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke; a journal of cerebral circulation,* **45**(1), pp. 24-36.

DREGAN, A., CHARLTON, J., WOLFE, C.D., GULLIFORD, M.C. and MARKUS, H.S., 2014. Is sodium valproate, an HDAC inhibitor, associated with reduced risk of stroke and myocardial infarction? A nested case-control study. *Pharmacoepidemiology and drug safety,* **23**(7), pp. 759-767.

FALK, E., NAKANO, M., BENTZON, J.F., FINN, A.V. and VIRMANI, R., 2013. Update on acute coronary syndromes: the pathologists' view. *European heart journal,* **34**(10), pp. 719-728.

FELTON, C.V., CROOK, D., DAVIES, M.J. and OLIVER, M.F., 1997. Relation of plaque lipid composition and morphology to the stability of human aortic plaques. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **17**(7), pp. 1337-1345.

FRIEDEWALD, W.T., LEVY, R.I. and FREDRICKSON, D.S., 1972. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical chemistry,* **18**(6), pp. 499-502.

FUGGER, L., MCVEAN, G. and BELL, J.I., 2012. Genomewide association studies and common disease--realizing clinical utility. *The New England journal of medicine,* **367**(25), pp. 2370-2371.

GABRIEL, S.B., SCHAFFNER, S.F., NGUYEN, H., MOORE, J.M., ROY, J., BLUMENSTIEL, B., HIGGINS, J., DEFELICE, M., LOCHNER, A., FAGGART, M., LIU-CORDERO, S.N., ROTIMI, C., ADEYEMO, A., COOPER, R., WARD, R., LANDER, E.S., DALY, M.J. and ALTSHULER, D., 2002. The structure of haplotype blocks in the human genome. *Science (New York, N.Y.),* **296**(5576), pp. 2225-2229.

GALKINA, E. and LEY, K., 2007. Vascular adhesion molecules in atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **27**(11), pp. 2292-2301.

GIMBRONE, M.A.,Jr and GARCIA-CARDENA, G., 2013. Vascular endothelium, hemodynamics, and the pathobiology of atherosclerosis. *Cardiovascular pathology : the official journal of the Society for Cardiovascular Pathology,* **22**(1), pp. 9-15.

GOLDSTEIN, J.L., HO, Y.K., BASU, S.K. and BROWN, M.S., 1979. Binding site on macrophages that mediates uptake and degradation of acetylated low density lipoprotein, producing massive cholesterol deposition. *Proceedings of the National Academy of Sciences of the United States of America,* **76**(1), pp. 333-337.

GROSSE, S.D. and KHOURY, M.J., 2006. What is the clinical utility of genetic testing? *Genetics in medicine : official journal of the American College of Medical Genetics,* **8**(7), pp. 448-450.

HABERLAND, M., MONTGOMERY, R.L. and OLSON, E.N., 2009. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature reviews.Genetics,* **10**(1), pp. 32-42.

HALPERT, I., SIRES, U.I., ROBY, J.D., POTTER-PERIGO, S., WIGHT, T.N., SHAPIRO, S.D., WELGUS, H.G., WICKLINE, S.A. and PARKS, W.C., 1996. Matrilysin is expressed by lipid-laden macrophages at sites of potential rupture in atherosclerotic lesions and localizes to areas of versican deposition, a proteoglycan substrate for the enzyme. *Proceedings of the National Academy of Sciences of the United States of America,* **93**(18), pp. 9748-9753.

HANSSON, G.K., 2005. Inflammation, atherosclerosis, and coronary artery disease. *The New England journal of medicine,* **352**(16), pp. 1685-1695.

HILTUNEN, T.P., GOUGH, P.J., GREAVES, D.R., GORDON, S. and YLA-HERTTUALA, S., 2001. Rabbit atherosclerotic lesions express scavenger receptor AIII mRNA, a naturally occurring splice variant that encodes a non-functional,

dominant negative form of the macrophage scavenger receptor. *Atherosclerosis,* **154**(2), pp. 415-419.

HIRSCHHORN, J.N. and GAJDOS, Z.K., 2011. Genome-wide association studies: results from the first few years and potential implications for clinical medicine. *Annual Review of Medicine,* **62**, pp. 11-24.

HOGUE, J.C., LAMARCHE, B., TREMBLAY, A.J., BERGERON, J., GAGNE, C. and COUTURE, P., 2008. Differential effect of atorvastatin and fenofibrate on plasma oxidized low-density lipoprotein, inflammation markers, and cell adhesion molecules in patients with type 2 diabetes mellitus. *Metabolism: clinical and experimental,* **57**(3), pp. 380-386.

HOHEISEL, J.D., 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nature reviews.Genetics,* **7**(3), pp. 200-210.

HOLMEN, O.L., ZHANG, H., FAN, Y., HOVELSON, D.H., SCHMIDT, E.M., ZHOU, W., GUO, Y., ZHANG, J., LANGHAMMER, A., LOCHEN, M.L., GANESH, S.K., VATTEN, L., SKORPEN, F., DALEN, H., ZHANG, J., PENNATHUR, S., CHEN, J., PLATOU, C., MATHIESEN, E.B., WILSGAARD, T., NJOLSTAD, I., BOEHNKE, M., CHEN, Y.E., ABECASIS, G.R., HVEEM, K. and WILLER, C.J., 2014. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nature genetics,* **46**(4), pp. 345-351.

HOLVOET, P., STASSEN, J.M., VAN CLEEMPUT, J., COLLEN, D. and VANHAECKE, J., 1998. Oxidized low density lipoproteins in patients with transplant-associated coronary artery disease. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **18**(1), pp. 100-107.

HOMMA, Y., MICHISHITA, I., HAYASHI, H., SHIGEMATSU, H. and KANAGAWA LIPID RESEARCH GROUP, 2010. Effects of low-dose simvastatin on the distribution of plasma cholesterol and oxidized low-density lipoprotein in three ultra-centrifugally separated low-density lipoprotein subfractions: 12- month, open-label trial. *Journal of Atherosclerosis and Thrombosis,* **17**(10), pp. 1049-1053.

INTERNATIONAL HAPMAP CONSORTIUM, 2005. A haplotype map of the human genome. *Nature,* **437**(7063), pp. 1299-1320.

INTERNATIONAL STROKE GENETICS CONSORTIUM (ISGC), WELLCOME TRUST CASE CONTROL CONSORTIUM 2 (WTCCC2), BELLENGUEZ, C., BEVAN, S., GSCHWENDTNER, A., SPENCER, C.C., BURGESS, A.I., PIRINEN, M., JACKSON, C.A., TRAYLOR, M., STRANGE, A., SU, Z., BAND, G., SYME, P.D., MALIK, R., PERA, J., NORRVING, B., LEMMENS, R., FREEMAN, C., SCHANZ, R., JAMES, T., POOLE, D., MURPHY, L., SEGAL, H., CORTELLINI, L., CHENG, Y.C., WOO, D., NALLS, M.A., MULLER-MYHSOK, B., MEISINGER, C., SEEDORF, U., ROSS-ADAMS, H., BOONEN, S., WLOCH-KOPEC, D., VALANT, V., SLARK, J., FURIE, K., DELAVARAN, H., LANGFORD, C., DELOUKAS, P., EDKINS, S., HUNT, S., GRAY, E., DRONOV, S., PELTONEN, L., GRETARSDOTTIR, S., THORLEIFSSON, G., THORSTEINSDOTTIR, U., STEFANSSON, K., BONCORAGLIO, G.B., PARATI, E.A., ATTIA, J., HOLLIDAY, E., LEVI, C., FRANZOSI, M.G., GOEL, A., HELGADOTTIR, A., BLACKWELL, J.M., BRAMON, E., BROWN, M.A., CASAS, J.P., CORVIN, A., DUNCANSON, A., JANKOWSKI, J., MATHEW, C.G., PALMER, C.N., PLOMIN, R., RAUTANEN, A., SAWCER, S.J., TREMBATH, R.C., VISWANATHAN, A.C., WOOD, N.W., WORRALL, B.B.,

KITTNER, S.J., MITCHELL, B.D., KISSELA, B., MESCHIA, J.F., THIJS, V., LINDGREN, A., MACLEOD, M.J., SLOWIK, A., WALTERS, M., ROSAND, J., SHARMA, P., FARRALL, M., SUDLOW, C.L., ROTHWELL, P.M., DICHGANS, M., DONNELLY, P. and MARKUS, H.S., 2012. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nature genetics,* **44**(3), pp. 328-333.

ISHIGAKI, Y., KATAGIRI, H., GAO, J., YAMADA, T., IMAI, J., UNO, K., HASEGAWA, Y., KANEKO, K., OGIHARA, T., ISHIHARA, H., SATO, Y., TAKIKAWA, K., NISHIMICHI, N., MATSUDA, H., SAWAMURA, T. and OKA, Y., 2008. Impact of plasma oxidized low-density lipoprotein removal on atherosclerosis. *Circulation,* **118**(1), pp. 75-83.

ISHIGAKI, Y., OKA, Y. and KATAGIRI, H., 2009. Circulating oxidized LDL: a biomarker and a pathogenic factor. *Current opinion in lipidology,* **20**(5), pp. 363-369.

ITABE, H. and UEDA, M., 2007. Measurement of plasma oxidized low-density lipoprotein and its clinical implications. *Journal of Atherosclerosis and Thrombosis,* **14**(1), pp. 1-11.

JANEWAY, C.A.,Jr and MEDZHITOV, R., 2002. Innate immune recognition. *Annual Review of Immunology,* **20**, pp. 197-216.

JASHARI, F., IBRAHIMI, P., NICOLL, R., BAJRAKTARI, G., WESTER, P. and HENEIN, M.Y., 2013. Coronary and carotid atherosclerosis: similarities and differences. *Atherosclerosis,* **227**(2), pp. 193-200.

JOVIN, T.G., CHAMORRO, A., COBO, E., DE MIQUEL, M.A., MOLINA, C.A., ROVIRA, A., ROMAN, L.S., SERENA, J., ABILLEIRA, S., RIBO, M., MILLAN, M., URRA, X., CARDONA, P., LOPEZ-CANCIO, E., TOMASELLO, A., CASTANO, C., BLASCO, J., AJA, L., DORADO, L., QUESADA, H., RUBIERA, M., HERNANDEZ-PEREZ, M., GOYAL, M., DEMCHUK, A.M., VON KUMMER, R., GALLOFRE, M., DAVALOS, A. and REVASCAT TRIAL INVESTIGATORS, 2015. Thrombectomy within 8 Hours after Symptom Onset in Ischemic Stroke. *The New England journal of medicine,* .

KANNEL, W.B. and WOLF, P.A., 2006. Peripheral and cerebral atherothrombosis and cardiovascular events in different vascular territories: insights from the Framingham Study. *Current atherosclerosis reports,* **8**(4), pp. 317-323.

KEATING, B.J., TISCHFIELD, S., MURRAY, S.S., BHANGALE, T., PRICE, T.S., GLESSNER, J.T., GALVER, L., BARRETT, J.C., GRANT, S.F., FARLOW, D.N., CHANDRUPATLA, H.R., HANSEN, M., AJMAL, S., PAPANICOLAOU, G.J., GUO, Y., LI, M., DEROHANNESSIAN, S., DE BAKKER, P.I., BAILEY, S.D., MONTPETIT, A., EDMONDSON, A.C., TAYLOR, K., GAI, X., WANG, S.S., FORNAGE, M., SHAIKH, T., GROOP, L., BOEHNKE, M., HALL, A.S., HATTERSLEY, A.T., FRACKELTON, E., PATTERSON, N., CHIANG, C.W., KIM, C.E., FABSITZ, R.R., OUWEHAND, W., PRICE, A.L., MUNROE, P., CAULFIELD, M., DRAKE, T., BOERWINKLE, E., REICH, D., WHITEHEAD, A.S., CAPPOLA, T.P., SAMANI, N.J., LUSIS, A.J., SCHADT, E., WILSON, J.G., KOENIG, W., MCCARTHY, M.I., KATHIRESAN, S., GABRIEL, S.B., HAKONARSON, H., ANAND, S.S., REILLY, M., ENGERT, J.C., NICKERSON, D.A., RADER, D.J., HIRSCHHORN, J.N. and FITZGERALD, G.A., 2008. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PloS one,* **3**(10), pp. e3583.

KERVINEN, K., KAPRIO, J., KOSKENVUO, M., JUNTUNEN, J. and KESANIEMI, Y.A., 1998. Serum lipids and apolipoprotein E phenotypes in identical twins reared apart. *Clinical genetics,* **53**(3), pp. 191-199.

KILARSKI, L.L., ACHTERBERG, S., DEVAN, W.J., TRAYLOR, M., MALIK, R., LINDGREN, A., PARE, G., SHARMA, P., SLOWIK, A., THIJS, V., WALTERS, M., WORRALL, B.B., SALE, M.M., ALGRA, A., KAPPELLE, L.J., WIJMENGA, C., NORRVING, B., SANDLING, J.K., RONNBLOM, L., GORIS, A., FRANKE, A., SUDLOW, C., ROTHWELL, P.M., LEVI, C., HOLLIDAY, E.G., FORNAGE, M., PSATY, B., GRETARSDOTTIR, S., THORSTEINSDOTTIR, U., SESHADRI, S., MITCHELL, B.D., KITTNER, S., CLARKE, R., HOPEWELL, J.C., BIS, J.C., BONCORAGLIO, G.B., MESCHIA, J., IKRAM, M.A., HANSEN, B.M., MONTANER, J., THORLEIFSSON, G., STEFANSON, K., ROSAND, J., DE BAKKER, P.I., FARRALL, M., DICHGANS, M., MARKUS, H.S., BEVAN, S. and GARNET COLLABORATIVE RESEARCH GROUP, WELLCOME TRUST CASE CONTROL CONSORTIUM 2, AUSTRALIAN STROKE GENETIC COLLABORATIVE, THE METASTROKE CONSORTIUM, AND THE INTERNATIONAL STROKE GENETICS CONSORTIUM, 2014. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology,* **83**(8), pp. 678-685.

KLEIN, R.J., XU, X., MUKHERJEE, S., WILLIS, J. and HAYES, J., 2010. Successes of genome-wide association studies. *Cell,* **142**(3), pp. 350-1; author reply 353-5.

KLEIN, R.J., ZEISS, C., CHEW, E.Y., TSAI, J.Y., SACKLER, R.S., HAYNES, C., HENNING, A.K., SANGIOVANNI, J.P., MANE, S.M., MAYNE, S.T., BRACKEN, M.B., FERRIS, F.L., OTT, J., BARNSTABLE, C. and HOH, J., 2005. Complement factor H polymorphism in age-related macular degeneration. *Science (New York, N.Y.),* **308**(5720), pp. 385-389.

KLINZ, F.J., KORKMAZ, Y., BLOCH, W., RAAB, W.H. and ADDICKS, K., 2012. Histone deacetylases 2 and 9 are coexpressed and nuclear localized in human molar odontoblasts in vivo. *Histochemistry and cell biology,* .

KOGA, J. and AIKAWA, M., 2012. Crosstalk between macrophages and smooth muscle cells in atherosclerotic vascular diseases. *Vascular pharmacology,* **57**(1), pp. 24-28.

KOVANEN, P.T., KAARTINEN, M. and PAAVONEN, T., 1995. Infiltrates of activated mast cells at the site of coronary atheromatous erosion or rupture in myocardial infarction. *Circulation,* **92**(5), pp. 1084-1088.

KUMAR, V., BUTCHER, S.J., OORNI, K., ENGELHARDT, P., HEIKKONEN, J., KASKI, K., ALA-KORPELA, M. and KOVANEN, P.T., 2011. Three-dimensional cryoEM reconstruction of native LDL particles to 16A resolution at physiological body temperature. *PloS one,* **6**(5), pp. e18841.

LEE, J.T., PAMIR, N., LIU, N.C., KIRK, E.A., AVERILL, M.M., BECKER, L., LARSON, I., HAGMAN, D.K., FOSTER-SCHUBERT, K.E., VAN YSERLOO, B., BORNFELDT, K.E., LEBOEUF, R.C., KRATZ, M. and HEINECKE, J.W., 2014. Macrophage metalloelastase (MMP12) regulates adipose tissue expansion, insulin sensitivity, and expression of inducible nitric oxide synthase. *Endocrinology,* **155**(9), pp. 3409-3420.

LEHTIMÄKI, T., LEHTINEN, S., SOLAKIVI, T., NIKKILÄ, M., JAAKKOLA, O., JOKELA, H., YÄ-HERTTUALA, S., LUOMA, J.S., KOIVULA, T. and NIKKARI, T., 1999. Autoantibodies against oxidized low density lipoprotein in

patients with angiographically verified coronary artery disease. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **19**(1), pp. 23-27.
LEITINGER, N., 2003. Oxidized phospholipids as modulators of inflammation in atherosclerosis. *Current opinion in lipidology,* **14**(5), pp. 421-430.
LESNIK, P., HASKELL, C.A. and CHARO, I.F., 2003. Decreased atherosclerosis in CX3CR1-/- mice reveals a role for fractalkine in atherogenesis. *The Journal of clinical investigation,* **111**(3), pp. 333-340.
LEVULA, M., OKSALA, N., AIRLA, N., ZEITLIN, R., SALENIUS, J.P., JÄRVINEN, O., VENERMO, M., PARTIO, T., SAARINEN, J., SOMPPI, T., SUOMINEN, V., VIRKKUNEN, J., HAUTALAHTI, J., LAAKSONEN, R., KÄHÖNEN, M., MENNANDER, A., KYTÖMÄKI, L., SOINI, J.T., PARKKINEN, J., PELTO-HUIKKO, M. and LEHTIMÄKI, T., 2012. Genes involved in systemic and arterial bed dependent atherosclerosis--Tampere Vascular study. *PloS one,* **7**(4), pp. e33787.
LI, Y., WILLER, C., SANNA, S. and ABECASIS, G., 2009. Genotype imputation. *Annual review of genomics and human genetics,* **10**, pp. 387-406.
LI, Y., WILLER, C.J., DING, J., SCHEET, P. and ABECASIS, G.R., 2010. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology,* **34**(8), pp. 816-834.
LIBBY P, RIDKER PM and HANSSON GK, 2011. Progress and challenges in translating the biology of atherosclerosis. *Nature,* **473**(7347), pp. 317-325.
LIN, J., KAKKAR, V. and LU, X., 2014. Impact of matrix metalloproteinases on atherosclerosis. *Current Drug Targets,* **15**(4), pp. 442-453.
LOREE, H.M., KAMM, R.D., STRINGFELLOW, R.G. and LEE, R.T., 1992. Effects of fibrous cap thickness on peak circumferential stress in model atherosclerotic vessels. *Circulation research,* **71**(4), pp. 850-858.
LOWEL, H., DORING, A., SCHNEIDER, A., HEIER, M., THORAND, B., MEISINGER, C. and MONICA/KORA STUDY GROUP, 2005. The MONICA Augsburg surveys--basis for prospective cohort studies. *Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)),* **67 Suppl 1**, pp. S13-8.
LUSIS, A.J., 2012. Genetics of atherosclerosis. *Trends in genetics : TIG,* **28**(6), pp. 267-275.
LUSIS, A.J., FOGELMAN, A.M. and FONAROW, G.C., 2004. Genetic basis of atherosclerosis: part I: new genes and pathways. *Circulation,* **110**(13), pp. 1868-1873.
LUSIS, A.J. and PAJUKANTA, P., 2008. A treasure trove for lipoprotein biology. *Nature genetics,* **40**(2), pp. 129-130.
LUTTERS, B.C., LEEUWENBURGH, M.A., APPELDOORN, C.C., MOLENAAR, T.J., VAN BERKEL, T.J. and BIESSEN, E.A., 2004. Blocking endothelial adhesion molecules: a potential therapeutic strategy to combat atherogenesis. *Current opinion in lipidology,* **15**(5), pp. 545-552.
MANICA-CATTANI, M.F., DUARTE, M.M., RIBEIRO, E.E., DE OLIVEIRA, R. and MANICA DA CRUZ, I.B., 2012. Effect of the interleukin-1B gene on serum oxidized low-density lipoprotein levels. *Clinical biochemistry,* **45**(9), pp. 641-645.
MANKTELOW, B.N. and POTTER, J.F., 2009. Interventions in the management of serum lipids for preventing stroke recurrence. *The Cochrane database of systematic reviews,* **(3):CD002091. doi**(3), pp. CD002091.
MANOLIO, T.A., 2010. Genomewide association studies and assessment of the risk of disease. *The New England journal of medicine,* **363**(2), pp. 166-176.

MANOLIO, T.A., BROOKS, L.D. and COLLINS, F.S., 2008. A HapMap harvest of insights into the genetics of common disease. *The Journal of clinical investigation,* **118**(5), pp. 1590-1605.

MANOLIO, T.A., COLLINS, F.S., COX, N.J., GOLDSTEIN, D.B., HINDORFF, L.A., HUNTER, D.J., MCCARTHY, M.I., RAMOS, E.M., CARDON, L.R., CHAKRAVARTI, A., CHO, J.H., GUTTMACHER, A.E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C.N., SLATKIN, M., VALLE, D., WHITTEMORE, A.S., BOEHNKE, M., CLARK, A.G., EICHLER, E.E., GIBSON, G., HAINES, J.L., MACKAY, T.F., MCCARROLL, S.A. and VISSCHER, P.M., 2009. Finding the missing heritability of complex diseases. *Nature,* **461**(7265), pp. 747-753.

MARCHINI, J. and HOWIE, B., 2010. Genotype imputation for genome-wide association studies. *Nature reviews.Genetics,* **11**(7), pp. 499-511.

MASSBERG, S., BRAND, K., GRUNER, S., PAGE, S., MULLER, E., MULLER, I., BERGMEIER, W., RICHTER, T., LORENZ, M., KONRAD, I., NIESWANDT, B. and GAWAZ, M., 2002. A critical role of platelet adhesion in the initiation of atherosclerotic lesion formation. *The Journal of experimental medicine,* **196**(7), pp. 887-896.

MASSBERG, S., GRAHL, L., VON BRUEHL, M.L., MANUKYAN, D., PFEILER, S., GOOSMANN, C., BRINKMANN, V., LORENZ, M., BIDZHEKOV, K., KHANDAGALE, A.B., KONRAD, I., KENNERKNECHT, E., REGES, K., HOLDENRIEDER, S., BRAUN, S., REINHARDT, C., SPANNAGL, M., PREISSNER, K.T. and ENGELMANN, B., 2010. Reciprocal coupling of coagulation and innate immunity via neutrophil serine proteases. *Nature medicine,* **16**(8), pp. 887-896.

MATHIESEN, E.B. and JOHNSEN, S.H., 2009. Ultrasonographic measurements of subclinical carotid atherosclerosis in prediction of ischemic stroke. *Acta neurologica Scandinavica.Supplementum,* **(189):68-72. doi**(189), pp. 68-72.

MCCLELLAN, J. and KING, M.C., 2010. Genetic heterogeneity in human disease. *Cell,* **141**(2), pp. 210-217.

MCKINSEY, T.A., 2011. Targeting inflammation in heart failure with histone deacetylase inhibitors. *Molecular medicine (Cambridge, Mass.),* **17**(5-6), pp. 434-441.

MEISINGER, C., BAUMERT, J., KHUSEYINOVA, N., LOEWEL, H. and KOENIG, W., 2005. Plasma oxidized low-density lipoprotein, a strong predictor for acute coronary heart disease events in apparently healthy, middle-aged men from the general population. *Circulation,* **112**(5), pp. 651-657.

MENNANDER, A., KUUKASJARVI, P., LAURIKKA, J., NIKUS, K., KARHUNEN, P.J., TARKKA, M. and LEHTIMAKI, T., 2008. Diagnostic performance of plasma high sensitive C-reactive protein in detecting three-vessel coronary artery disease: modification by apolipoprotein E genotype. *Scandinavian Journal of Clinical and Laboratory Investigation,* **68**(8), pp. 714-719.

MILDE, T., OEHME, I., KORSHUNOV, A., KOPP-SCHNEIDER, A., REMKE, M., NORTHCOTT, P., DEUBZER, H.E., LODRINI, M., TAYLOR, M.D., VON DEIMLING, A., PFISTER, S. and WITT, O., 2010. HDAC5 and HDAC9 in medulloblastoma: novel markers for risk stratification and role in tumor cell growth. *Clinical cancer research : an official journal of the American Association for Cancer Research,* **16**(12), pp. 3240-3252.

MORGAN, A.R., RERKASEM, K., GALLAGHER, P.J., ZHANG, B., MORRIS, G.E., CALDER, P.C., GRIMBLE, R.F., ERIKSSON, P., MCPHEAT, W.L., SHEARMAN, C.P. and YE, S., 2004. Differences in matrix metalloproteinase-1 and matrix metalloproteinase-12 transcript levels among carotid atherosclerotic plaques with different histopathological characteristics. *Stroke; a journal of cerebral circulation,* **35**(6), pp. 1310-1315.

MYOCARDIAL INFARCTION GENETICS CONSORTIUM, KATHIRESAN, S., VOIGHT, B.F., PURCELL, S., MUSUNURU, K., ARDISSINO, D., MANNUCCI, P.M., ANAND, S., ENGERT, J.C., SAMANI, N.J., SCHUNKERT, H., ERDMANN, J., REILLY, M.P., RADER, D.J., MORGAN, T., SPERTUS, J.A., STOLL, M., GIRELLI, D., MCKEOWN, P.P., PATTERSON, C.C., SISCOVICK, D.S., O'DONNELL, C.J., ELOSUA, R., PELTONEN, L., SALOMAA, V., SCHWARTZ, S.M., MELANDER, O., ALTSHULER, D., ARDISSINO, D., MERLINI, P.A., BERZUINI, C., BERNARDINELLI, L., PEYVANDI, F., TUBARO, M., CELLI, P., FERRARIO, M., FETIVEAU, R., MARZILIANO, N., CASARI, G., GALLI, M., RIBICHINI, F., ROSSI, M., BERNARDI, F., ZONZIN, P., PIAZZA, A., MANNUCCI, P.M., SCHWARTZ, S.M., SISCOVICK, D.S., YEE, J., FRIEDLANDER, Y., ELOSUA, R., MARRUGAT, J., LUCAS, G., SUBIRANA, I., SALA, J., RAMOS, R., KATHIRESAN, S., MEIGS, J.B., WILLIAMS, G., NATHAN, D.M., MACRAE, C.A., O'DONNELL, C.J., SALOMAA, V., HAVULINNA, A.S., PELTONEN, L., MELANDER, O., BERGLUND, G., VOIGHT, B.F., KATHIRESAN, S., HIRSCHHORN, J.N., ASSELTA, R., DUGA, S., SPREAFICO, M., MUSUNURU, K., DALY, M.J., PURCELL, S., VOIGHT, B.F., PURCELL, S., NEMESH, J., KORN, J.M., MCCARROLL, S.A., SCHWARTZ, S.M., YEE, J., KATHIRESAN, S., LUCAS, G., SUBIRANA, I., ELOSUA, R., SURTI, A., GUIDUCCI, C., GIANNINY, L., MIREL, D., PARKIN, M., BURTT, N., GABRIEL, S.B., SAMANI, N.J., THOMPSON, J.R., BRAUND, P.S., WRIGHT, B.J., BALMFORTH, A.J., BALL, S.G., HALL, A., WELLCOME TRUST CASE CONTROL CONSORTIUM, SCHUNKERT, H., ERDMANN, J., LINSEL-NITSCHKE, P., LIEB, W., ZIEGLER, A., KONIG, I., HENGSTENBERG, C., FISCHER, M., STARK, K., GROSSHENNIG, A., PREUSS, M., WICHMANN, H.E., SCHREIBER, S., SCHUNKERT, H., SAMANI, N.J., ERDMANN, J., OUWEHAND, W., HENGSTENBERG, C., DELOUKAS, P., SCHOLZ, M., CAMBIEN, F., REILLY, M.P., LI, M., CHEN, Z., WILENSKY, R., MATTHAI, W., QASIM, A., HAKONARSON, H.H., DEVANEY, J., BURNETT, M.S., PICHARD, A.D., KENT, K.M., SATLER, L., LINDSAY, J.M., WAKSMAN, R., KNOUFF, C.W., WATERWORTH, D.M., WALKER, M.C., MOOSER, V., EPSTEIN, S.E., RADER, D.J., SCHEFFOLD, T., BERGER, K., STOLL, M., HUGE, A., GIRELLI, D., MARTINELLI, N., OLIVIERI, O., CORROCHER, R., MORGAN, T., SPERTUS, J.A., MCKEOWN, P., PATTERSON, C.C., SCHUNKERT, H., ERDMANN, E., LINSEL-NITSCHKE, P., LIEB, W., ZIEGLER, A., KONIG, I.R., HENGSTENBERG, C., FISCHER, M., STARK, K., GROSSHENNIG, A., PREUSS, M., WICHMANN, H.E., SCHREIBER, S., HOLM, H., THORLEIFSSON, G., THORSTEINSDOTTIR, U., STEFANSSON, K., ENGERT, J.C., DO, R., XIE, C., ANAND, S., KATHIRESAN, S., ARDISSINO, D., MANNUCCI, P.M., SISCOVICK, D., O'DONNELL, C.J., SAMANI, N.J., MELANDER, O., ELOSUA, R., PELTONEN, L., SALOMAA,

V., SCHWARTZ, S.M. and ALTSHULER, D., 2009. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nature genetics,* **41**(3), pp. 334-341.

NABEL, E.G. and BRAUNWALD, E., 2012. A tale of coronary artery disease and myocardial infarction. *The New England journal of medicine,* **366**(1), pp. 54-63.

NAKASHIMA, Y., RAINES, E.W., PLUMP, A.S., BRESLOW, J.L. and ROSS, R., 1998. Upregulation of VCAM-1 and ICAM-1 at atherosclerosis-prone sites on the endothelium in the ApoE-deficient mouse. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **18**(5), pp. 842-851.

NARUKO, T., UEDA, M., EHARA, S., ITOH, A., HAZE, K., SHIRAI, N., IKURA, Y., OHSAWA, M., ITABE, H., KOBAYASHI, Y., YAMAGISHI, H., YOSHIYAMA, M., YOSHIKAWA, J. and BECKER, A.E., 2006. Persistent high levels of plasma oxidized low-density lipoprotein after acute myocardial infarction predict stent restenosis. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **26**(4), pp. 877-883.

NATARAJAN, M., 2002. Unstable angina. *Clinical evidence,* **(7)**(7), pp. 214-226.

NIEMINEN, T., LEHTINEN, R., VIIK, J., LEHTIMAKI, T., NIEMELA, K., NIKUS, K., NIEMI, M., KALLIO, J., KOOBI, T., TURJANMAA, V. and KAHONEN, M., 2006. The Finnish Cardiovascular Study (FINCAVAS): characterising patients with high risk of cardiovascular morbidity and mortality. *BMC cardiovascular disorders,* **6**, pp. 9.

NISHI, K., ITABE, H., UNO, M., KITAZATO, K.T., HORIGUCHI, H., SHINNO, K. and NAGAHIRO, S., 2002. Oxidized LDL in carotid plaques and plasma associates with plaque instability. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **22**(10), pp. 1649-1654.

OKAMOTO, H., FUJIOKA, Y., TAKAHASHI, A., TAKAHASHI, T., TANIGUCHI, T., ISHIKAWA, Y. and YOKOYAMA, M., 2006. Trichostatin A, an inhibitor of histone deacetylase, inhibits smooth muscle cell proliferation via induction of p21(WAF1). *Journal of Atherosclerosis and Thrombosis,* **13**(4), pp. 183-191.

OKSALA, N., PÄRSSINEN, J., SEPPÄLÄ, I., RAITOHARJU, E., KHOLOVA, I., HERNESNIEMI, J., LYYTIKÄINEN, L.P., LEVULA, M., MÄKELÄ, K.M., SIORIS, T., KÄHÖNEN, M., LAAKSONEN, R., HYTÖNEN, V. and LEHTIMÄKI, T., 2013. Association of neuroimmune guidance cue netrin-1 and its chemorepulsive receptor UNC5B with atherosclerotic plaque expression signatures and stability in human(s): Tampere Vascular Study (TVS). *Circulation.Cardiovascular genetics,* **6**(6), pp. 579-587.

OOI, E.M., RUSSELL, B.S., OLSON, E., SUN, S.Z., DIFFENDERFER, M.R., LICHTENSTEIN, A.H., KEILSON, L., BARRETT, P.H., SCHAEFER, E.J. and SPRECHER, D.L., 2012. Apolipoprotein B-100-containing lipoprotein metabolism in subjects with lipoprotein lipase gene mutations. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **32**(2), pp. 459-466.

PANIAGUA, J.A., LOPEZ-MIRANDA, J., PEREZ-MARTINEZ, P., MARIN, C., VIDA, J.M., FUENTES, F., FERNANDEZ DE LA PUEBLA, R.A. and PEREZ-JIMENEZ, F., 2005. Oxidized-LDL levels are changed during short-term serum glucose variations and lowered with statin treatment in early Type 2 diabetes: a study of endothelial function and microalbuminuria. *Diabetic medicine : a journal of the British Diabetic Association,* **22**(12), pp. 1647-1656.

PAPAYANNOPOULOS, V. and ZYCHLINSKY, A., 2009. NETs: a new strategy for using old weapons. *Trends in immunology,* **30**(11), pp. 513-521.

PARE, G., CHASMAN, D.I., PARKER, A.N., ZEE, R.R., MALARSTIG, A., SEEDORF, U., COLLINS, R., WATKINS, H., HAMSTEN, A., MILETICH, J.P. and RIDKER, P.M., 2009. Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. *Circulation.Cardiovascular genetics,* **2**(2), pp. 142-150.

PARK, Y.M., 2014. CD36, a scavenger receptor implicated in atherosclerosis. *Experimental & molecular medicine,* **46**, pp. e99.

PEISER, L., MUKHOPADHYAY, S. and GORDON, S., 2002. Scavenger receptors in innate immunity. *Current opinion in immunology,* **14**(1), pp. 123-128.

POUSTKA, A., POHL, T., BARLOW, D.P., ZEHETNER, G., CRAIG, A., MICHIELS, F., EHRICH, E., FRISCHAUF, A.M. and LEHRACH, H., 1986. Molecular approaches to mammalian genetics. *Cold Spring Harbor symposia on quantitative biology,* **51 Pt 1**, pp. 131-139.

PRICE, A.L., PATTERSON, N.J., PLENGE, R.M., WEINBLATT, M.E., SHADICK, N.A. and REICH, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics,* **38**(8), pp. 904-909.

PRUIM, R.J., WELCH, R.P., SANNA, S., TESLOVICH, T.M., CHINES, P.S., GLIEDT, T.P., BOEHNKE, M., ABECASIS, G.R. and WILLER, C.J., 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England),* **26**(18), pp. 2336-2337.

PUIG, O., YUAN, J., STEPANIANTS, S., ZIEBA, R., ZYCBAND, E., MORRIS, M., COULTER, S., YU, X., MENKE, J., WOODS, J., CHEN, F., RAMEY, D.R., HE, X., O'NEILL, E.A., HAILMAN, E., JOHNS, D.G., HUBBARD, B.K., YEE LUM, P., WRIGHT, S.D., DESOUZA, M.M., PLUMP, A. and REISER, V., 2011. A gene expression signature that classifies human atherosclerotic plaque by relative inflammation status. *Circulation.Cardiovascular genetics,* **4**(6), pp. 595-604.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M.A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P.I., DALY, M.J. and SHAM, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* **81**(3), pp. 559-575.

RAITAKARI, O.T., JUONALA, M., RONNEMAA, T., KELTIKANGAS-JARVINEN, L., RASANEN, L., PIETIKAINEN, M., HUTRI-KAHONEN, N., TAITTONEN, L., JOKINEN, E., MARNIEMI, J., JULA, A., TELAMA, R., KAHONEN, M., LEHTIMAKI, T., AKERBLOM, H.K. and VIIKARI, J.S., 2008. Cohort profile: the cardiovascular risk in Young Finns Study. *International journal of epidemiology,* **37**(6), pp. 1220-1226.

RAITOHARJU, E., SEPPALA, I., LEVULA, M., KUUKASJARVI, P., LAURIKKA, J., NIKUS, K., HUOVILA, A.P., OKSALA, N., KLOPP, N., ILLIG, T., LAAKSONEN, R., KARHUNEN, P.J., VIIK, J., LEHTINEN, R., PELTO-HUIKKO, M., TARKKA, M., KAHONEN, M. and LEHTIMAKI, T., 2011. Common variation in the ADAM8 gene affects serum sADAM8 concentrations and the risk of myocardial infarction in two independent cohorts. *Atherosclerosis,* **218**(1), pp. 127-133.

RAITOHARJU, E., SEPPÄLÄ, I., LYYTIKÄINEN, L.P., LEVULA, M., OKSALA, N., KLOPP, N., ILLIG, T., LAAKSONEN, R., KÄHÖNEN, M. and LEHTIMÄKI, T., 2013. A comparison of the accuracy of Illumina HumanHT-12 v3 Expression

BeadChip and TaqMan qRT-PCR gene expression results in patient samples from the Tampere Vascular Study. *Atherosclerosis,* **226**(1), pp. 149-152.

RERKASEM, K. and ROTHWELL, P.M., 2011. Carotid endarterectomy for symptomatic carotid stenosis. *The Cochrane database of systematic reviews,* **(4):CD001081. doi**(4), pp. CD001081.

RICHARDSON, P.D., DAVIES, M.J. and BORN, G.V., 1989. Influence of plaque configuration and stress distribution on fissuring of coronary atherosclerotic plaques. *Lancet,* **2**(8669), pp. 941-944.

RIPATTI, S., TIKKANEN, E., ORHO-MELANDER, M., HAVULINNA, A.S., SILANDER, K., SHARMA, A., GUIDUCCI, C., PEROLA, M., JULA, A., SINISALO, J., LOKKI, M.L., NIEMINEN, M.S., MELANDER, O., SALOMAA, V., PELTONEN, L. and KATHIRESAN, S., 2010. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet,* **376**(9750), pp. 1393-1400.

RUOTSALAINEN, A.K., INKALA, M., PARTANEN, M.E., LAPPALAINEN, J.P., KANSANEN, E., MAKINEN, P.I., HEINONEN, S.E., LAITINEN, H.M., HEIKKILA, J., VATANEN, T., HORKKO, S., YAMAMOTO, M., YLA-HERTTUALA, S., JAUHIAINEN, M. and LEVONEN, A.L., 2013. The absence of macrophage Nrf2 promotes early atherogenesis. *Cardiovascular research,* **98**(1), pp. 107-115.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S.C., KAKOL, J.M., STEIN, L.D., MARTH, G., SHERRY, S., MULLIKIN, J.C., MORTIMORE, B.J., WILLEY, D.L., HUNT, S.E., COLE, C.G., COGGILL, P.C., RICE, C.M., NING, Z., ROGERS, J., BENTLEY, D.R., KWOK, P.Y., MARDIS, E.R., YEH, R.T., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L., HILLIER, L., WATERSTON, R.H., MCPHERSON, J.D., GILMAN, B., SCHAFFNER, S., VAN ETTEN, W.J., REICH, D., HIGGINS, J., DALY, M.J., BLUMENSTIEL, B., BALDWIN, J., STANGE-THOMANN, N., ZODY, M.C., LINTON, L., LANDER, E.S., ALTSHULER, D. and INTERNATIONAL SNP MAP WORKING GROUP, 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* **409**(6822), pp. 928-933.

SALAGIANNI, M., GALANI, I.E., LUNDBERG, A.M., DAVOS, C.H., VARELA, A., GAVRIIL, A., LYYTIKAINEN, L.P., LEHTIMAKI, T., SIGALA, F., FOLKERSEN, L., GORGOULIS, V., LENGLET, S., MONTECUCCO, F., MACH, F., HEDIN, U., HANSSON, G.K., MONACO, C. and ANDREAKOS, E., 2012. Toll-like receptor 7 protects from atherosclerosis by constraining "inflammatory" macrophage activation. *Circulation,* **126**(8), pp. 952-962.

SANCHEZ-QUESADA, J.L., VILLEGAS, S. and ORDONEZ-LLANOS, J., 2012. Electronegative low-density lipoprotein. A link between apolipoprotein B misfolding, lipoprotein aggregation and proteoglycan binding. *Current opinion in lipidology,* **23**(5), pp. 479-486.

SEGREST, J.P., JONES, M.K., DE LOOF, H. and DASHTI, N., 2001. Structure of apolipoprotein B-100 in low density lipoproteins. *Journal of lipid research,* **42**(9), pp. 1346-1367.

SEIMON, T.A., NADOLSKI, M.J., LIAO, X., MAGALLON, J., NGUYEN, M., FERIC, N.T., KOSCHINSKY, M.L., HARKEWICZ, R., WITZTUM, J.L., TSIMIKAS, S., GOLENBOCK, D., MOORE, K.J. and TABAS, I., 2010. Atherogenic lipids and

lipoproteins trigger CD36-TLR2-dependent apoptosis in macrophages undergoing endoplasmic reticulum stress. *Cell metabolism,* **12**(5), pp. 467-482.

SEREBRUANY, V., SANI, Y., EISERT, C., SCHEVCHUCK, A., FONG, A. and HANLEY, D., 2011. Effects of Aggrenox and aspirin on plasma endothelial nitric oxide synthase and oxidised low-density lipoproteins in patients after ischaemic stroke. The AGgrenox versus aspirin therapy evaluation (AGATE) biomarker substudy. *Thrombosis and haemostasis,* **105**(1), pp. 81-87.

SESHADRI, S. and WOLF, P.A., 2007. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *The Lancet.Neurology,* **6**(12), pp. 1106-1114.

SHENDURE, J. and JI, H., 2008. Next-generation DNA sequencing. *Nature biotechnology,* **26**(10), pp. 1135-1145.

SHIMADA, K., MOKUNO, H., MATSUNAGA, E., MIYAZAKI, T., SUMIYOSHI, K., MIYAUCHI, K. and DAIDA, H., 2004. Circulating oxidized low-density lipoprotein is an independent predictor for cardiac event in patients with coronary artery disease. *Atherosclerosis,* **174**(2), pp. 343-347.

SITZER, M., PUAC, D., BUEHLER, A., STECKEL, D.A., VON KEGLER, S., MARKUS, H.S. and STEINMETZ, H., 2003. Internal carotid artery angle of origin: a novel risk factor for early carotid atherosclerosis. *Stroke; a journal of cerebral circulation,* **34**(4), pp. 950-955.

SKALEN, K., GUSTAFSSON, M., RYDBERG, E.K., HULTEN, L.M., WIKLUND, O., INNERARITY, T.L. and BOREN, J., 2002. Subendothelial retention of atherogenic lipoproteins in early atherosclerosis. *Nature,* **417**(6890), pp. 750-754.

SMITH, E.N., CHEN, W., KAHONEN, M., KETTUNEN, J., LEHTIMAKI, T., PELTONEN, L., RAITAKARI, O.T., SALEM, R.M., SCHORK, N.J., SHAW, M., SRINIVASAN, S.R., TOPOL, E.J., VIIKARI, J.S., BERENSON, G.S. and MURRAY, S.S., 2010. Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS genetics,* **6**(9), pp. e1001094.

SMITH, J.D., TROGAN, E., GINSBERG, M., GRIGAUX, C., TIAN, J. and MIYATA, M., 1995. Decreased atherosclerosis in mice deficient in both macrophage colony-stimulating factor (op) and apolipoprotein E. *Proceedings of the National Academy of Sciences of the United States of America,* **92**(18), pp. 8264-8268.

SOEHNLEIN, O. and LINDBOM, L., 2010. Phagocyte partnership during the onset and resolution of inflammation. *Nature reviews.Immunology,* **10**(6), pp. 427-439.

STARY, H.C., CHANDLER, A.B., DINSMORE, R.E., FUSTER, V., GLAGOV, S., INSULL, W.,Jr, ROSENFELD, M.E., SCHWARTZ, C.J., WAGNER, W.D. and WISSLER, R.W., 1995. A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **15**(9), pp. 1512-1531.

STARY, H.C., CHANDLER, A.B., GLAGOV, S., GUYTON, J.R., INSULL, W.,Jr, ROSENFELD, M.E., SCHAFFER, S.A., SCHWARTZ, C.J., WAGNER, W.D. and WISSLER, R.W., 1994. A definition of initial, fatty streak, and intermediate lesions of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Circulation,* **89**(5), pp. 2462-2478.

STEINBERG, D., 1997. Low density lipoprotein oxidation and its pathobiological significance. *The Journal of biological chemistry,* **272**(34), pp. 20963-20966.

STELLOS, K., SAUTER, R., FAHRLEITNER, M., GRIMM, J., STAKOS, D., EMSCHERMANN, F., PANAGIOTA, V., GNERLICH, S., PERK, A., SCHONBERGER, T., BIGALKE, B., LANGER, H.F. and GAWAZ, M., 2012. Binding of oxidized low-density lipoprotein on circulating platelets is increased in patients with acute coronary syndromes and induces platelet adhesion to vascular wall in vivo--brief report. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **32**(8), pp. 2017-2020.

STROKE UNIT TRIALISTS' COLLABORATION, 2013. Organised inpatient (stroke unit) care for stroke. *The Cochrane database of systematic reviews,* **9**, pp. CD000197.

SUDLOW, C.L., MASON, G., MAURICE, J.B., WEDDERBURN, C.J. and HANKEY, G.J., 2009. Thienopyridine derivatives versus aspirin for preventing stroke and other serious vascular events in high vascular risk patients. *The Cochrane database of systematic reviews,* **(4):CD001246. doi**(4), pp. CD001246.

TABAS, I., 2010. Macrophage death and defective inflammation resolution in atherosclerosis. *Nature reviews.Immunology,* **10**(1), pp. 36-46.

TENG, Z., SADAT, U., BROWN, A.J. and GILLARD, J.H., 2014. Plaque hemorrhage in carotid artery disease: pathogenesis, clinical and biomechanical considerations. *Journal of Biomechanics,* **47**(4), pp. 847-858.

TESLOVICH, T.M., MUSUNURU, K., SMITH, A.V., EDMONDSON, A.C., STYLIANOU, I.M., KOSEKI, M., PIRRUCCELLO, J.P., RIPATTI, S., CHASMAN, D.I., WILLER, C.J., JOHANSEN, C.T., FOUCHIER, S.W., ISAACS, A., PELOSO, G.M., BARBALIC, M., RICKETTS, S.L., BIS, J.C., AULCHENKO, Y.S., THORLEIFSSON, G., FEITOSA, M.F., CHAMBERS, J., ORHO-MELANDER, M., MELANDER, O., JOHNSON, T., LI, X., GUO, X., LI, M., SHIN CHO, Y., JIN GO, M., JIN KIM, Y., LEE, J.Y., PARK, T., KIM, K., SIM, X., TWEE-HEE ONG, R., CROTEAU-CHONKA, D.C., LANGE, L.A., SMITH, J.D., SONG, K., HUA ZHAO, J., YUAN, X., LUAN, J., LAMINA, C., ZIEGLER, A., ZHANG, W., ZEE, R.Y., WRIGHT, A.F., WITTEMAN, J.C., WILSON, J.F., WILLEMSEN, G., WICHMANN, H.E., WHITFIELD, J.B., WATERWORTH, D.M., WAREHAM, N.J., WAEBER, G., VOLLENWEIDER, P., VOIGHT, B.F., VITART, V., UITTERLINDEN, A.G., UDA, M., TUOMILEHTO, J., THOMPSON, J.R., TANAKA, T., SURAKKA, I., STRINGHAM, H.M., SPECTOR, T.D., SORANZO, N., SMIT, J.H., SINISALO, J., SILANDER, K., SIJBRANDS, E.J., SCUTERI, A., SCOTT, J., SCHLESSINGER, D., SANNA, S., SALOMAA, V., SAHARINEN, J., SABATTI, C., RUOKONEN, A., RUDAN, I., ROSE, L.M., ROBERTS, R., RIEDER, M., PSATY, B.M., PRAMSTALLER, P.P., PICHLER, I., PEROLA, M., PENNINX, B.W., PEDERSEN, N.L., PATTARO, C., PARKER, A.N., PARE, G., OOSTRA, B.A., O'DONNELL, C.J., NIEMINEN, M.S., NICKERSON, D.A., MONTGOMERY, G.W., MEITINGER, T., MCPHERSON, R., MCCARTHY, M.I., MCARDLE, W., MASSON, D., MARTIN, N.G., MARRONI, F., MANGINO, M., MAGNUSSON, P.K., LUCAS, G., LUBEN, R., LOOS, R.J., LOKKI, M.L., LETTRE, G., LANGENBERG, C., LAUNER, L.J., LAKATTA, E.G., LAAKSONEN, R., KYVIK, K.O., KRONENBERG, F., KONIG, I.R., KHAW, K.T., KAPRIO, J., KAPLAN, L.M., JOHANSSON, A., JARVELIN, M.R., JANSSENS, A.C., INGELSSON, E., IGL, W., KEES HOVINGH, G., HOTTENGA, J.J., HOFMAN, A., HICKS, A.A., HENGSTENBERG, C., HEID, I.M., HAYWARD, C., HAVULINNA, A.S., HASTIE, N.D., HARRIS, T.B., HARITUNIANS, T.,

HALL, A.S., GYLLENSTEN, U., GUIDUCCI, C., GROOP, L.C., GONZALEZ, E., GIEGER, C., FREIMER, N.B., FERRUCCI, L., ERDMANN, J., ELLIOTT, P., EJEBE, K.G., DORING, A., DOMINICZAK, A.F., DEMISSIE, S., DELOUKAS, P., DE GEUS, E.J., DE FAIRE, U., CRAWFORD, G., COLLINS, F.S., CHEN, Y.D., CAULFIELD, M.J., CAMPBELL, H., BURTT, N.P., BONNYCASTLE, L.L., BOOMSMA, D.I., BOEKHOLDT, S.M., BERGMAN, R.N., BARROSO, I., BANDINELLI, S., BALLANTYNE, C.M., ASSIMES, T.L., QUERTERMOUS, T., ALTSHULER, D., SEIELSTAD, M., WONG, T.Y., TAI, E.S., FERANIL, A.B., KUZAWA, C.W., ADAIR, L.S., TAYLOR, H.A.,Jr, BORECKI, I.B., GABRIEL, S.B., WILSON, J.G., HOLM, H., THORSTEINSDOTTIR, U., GUDNASON, V., KRAUSS, R.M., MOHLKE, K.L., ORDOVAS, J.M., MUNROE, P.B., KOONER, J.S., TALL, A.R., HEGELE, R.A., KASTELEIN, J.J., SCHADT, E.E., ROTTER, J.I., BOERWINKLE, E., STRACHAN, D.P., MOOSER, V., STEFANSSON, K., REILLY, M.P., SAMANI, N.J., SCHUNKERT, H., CUPPLES, L.A., SANDHU, M.S., RIDKER, P.M., RADER, D.J., VAN DUIJN, C.M., PELTONEN, L., ABECASIS, G.R., BOEHNKE, M. and KATHIRESAN, S., 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature,* **466**(7307), pp. 707-713.

THORAND, B., SCHNEIDER, A., BAUMERT, J., DORING, A., MAROWSKY-KOPPL, M., HEIER, M., MEISINGER, C. and LOWEL, H., 2005. Case-cohort studies: an effective design for the investigation of biomarkers as risk factors for chronic diseases--demonstrated by the example of the MONICA/KORA Augsburg Case-Cohort Study 1984-2002. *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany)),* **67 Suppl 1**, pp. S98-102.

TOSHIMA, S., HASEGAWA, A., KURABAYASHI, M., ITABE, H., TAKANO, T., SUGANO, J., SHIMAMURA, K., KIMURA, J., MICHISHITA, I., SUZUKI, T. and NAGAI, R., 2000. Circulating oxidized low density lipoprotein levels. A biochemical risk marker for coronary heart disease. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **20**(10), pp. 2243-2247.

TSIMIKAS, S., 2006. Oxidative biomarkers in the diagnosis and prognosis of cardiovascular disease. *The American Journal of Cardiology,* **98**(11A), pp. 9P-17P.

UNO, M., KITAZATO, K.T., NISHI, K., ITABE, H. and NAGAHIRO, S., 2003. Raised plasma oxidised LDL in acute cerebral infarction. *Journal of neurology, neurosurgery, and psychiatry,* **74**(3), pp. 312-316.

UZUN, H., ZENGIN, K., TASKIN, M., AYDIN, S., SIMSEK, G. and DARIYERLI, N., 2004. Changes in leptin, plasminogen activator factor and oxidative stress in morbidly obese patients following open and laparoscopic Swedish adjustable gastric banding. *Obesity Surgery,* **14**(5), pp. 659-665.

VAN DER WORP, H.B. and VAN GIJN, J., 2007. Clinical practice. Acute ischemic stroke. *The New England journal of medicine,* **357**(6), pp. 572-579.

VIBO, R., KORV, J., ROOSE, M., KAMPUS, P., MUDA, P., ZILMER, K. and ZILMER, M., 2007. Acute phase proteins and oxidised low-density lipoprotein in association with ischemic stroke subtype, severity and outcome. *Free radical research,* **41**(3), pp. 282-287.

WANG, K., BUCAN, M., GRANT, S.F., SCHELLENBERG, G. and HAKONARSON, H., 2010. Strategies for genetic studies of complex diseases. *Cell,* **142**(3), pp. 351-3; author reply 353-5.

WARDLAW, J.M., MURRAY, V., BERGE, E. and DEL ZOPPO, G.J., 2014. Thrombolysis for acute ischaemic stroke. *The Cochrane database of systematic reviews,* **7**, pp. CD000213.

WEBER, C. and NOELS, H., 2011. Atherosclerosis: current pathogenesis and therapeutic options. *Nature medicine,* **17**(11), pp. 1410-1422.

WEBER, C. and NOELS, H., 2011. Atherosclerosis: current pathogenesis and therapeutic options. *Nature medicine,* **17**(11), pp. 1410-1422.

WEINBRENNER, T., CLADELLAS, M., ISABEL COVAS, M., FITO, M., TOMAS, M., SENTI, M., BRUGUERA, J. and MARRUGAT, J., 2003. High oxidative stress in patients with stable coronary heart disease. *Atherosclerosis,* **168**(1), pp. 99-106.

WILLER, C.J., LI, Y. and ABECASIS, G.R., 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England),* **26**(17), pp. 2190-2191.

WILLER, C.J., SANNA, S., JACKSON, A.U., SCUTERI, A., BONNYCASTLE, L.L., CLARKE, R., HEATH, S.C., TIMPSON, N.J., NAJJAR, S.S., STRINGHAM, H.M., STRAIT, J., DUREN, W.L., MASCHIO, A., BUSONERO, F., MULAS, A., ALBAI, G., SWIFT, A.J., MORKEN, M.A., NARISU, N., BENNETT, D., PARISH, S., SHEN, H., GALAN, P., MENETON, P., HERCBERG, S., ZELENIKA, D., CHEN, W.M., LI, Y., SCOTT, L.J., SCHEET, P.A., SUNDVALL, J., WATANABE, R.M., NAGARAJA, R., EBRAHIM, S., LAWLOR, D.A., BEN-SHLOMO, Y., DAVEY-SMITH, G., SHULDINER, A.R., COLLINS, R., BERGMAN, R.N., UDA, M., TUOMILEHTO, J., CAO, A., COLLINS, F.S., LAKATTA, E., LATHROP, G.M., BOEHNKE, M., SCHLESSINGER, D., MOHLKE, K.L. and ABECASIS, G.R., 2008. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature genetics,* **40**(2), pp. 161-169.

WILLIAMS, K.J. and TABAS, I., 1995. The response-to-retention hypothesis of early atherogenesis. *Arteriosclerosis, Thrombosis, and Vascular Biology,* **15**(5), pp. 551-561.

WINKELMANN, B.R., MARZ, W., BOEHM, B.O., ZOTZ, R., HAGER, J., HELLSTERN, P., SENGES, J. and LURIC STUDY GROUP (LUDWIGSHAFEN RISK AND CARDIOVASCULAR HEALTH), 2001. Rationale and design of the LURIC study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics,* **2**(1 Suppl 1), pp. S1-73.

WOJCZYNSKI, M.K., GAO, G., BORECKI, I., HOPKINS, P.N., PARNELL, L., LAI, C.Q., ORDOVAS, J.M., CHUNG, B.H. and ARNETT, D.K., 2010. Apolipoprotein B genetic variants modify the response to fenofibrate: a GOLDN study. *Journal of lipid research,* **51**(11), pp. 3316-3323.

WOOLLARD, K.J. and GEISSMANN, F., 2010. Monocytes in atherosclerosis: subsets and functions. *Nature reviews.Cardiology,* **7**(2), pp. 77-86.

WU, T., WILLETT, W.C., RIFAI, N., SHAI, I., MANSON, J.E. and RIMM, E.B., 2006. Is plasma oxidized low-density lipoprotein, measured with the widely used antibody 4E6, an independent predictor of coronary heart disease among U.S. men and women? *Journal of the American College of Cardiology,* **48**(5), pp. 973-979.

YAMADA, S., WANG, K.Y., TANIMOTO, A., FAN, J., SHIMAJIRI, S., KITAJIMA, S., MORIMOTO, M., TSUTSUI, M., WATANABE, T., YASUMOTO, K. and SASAGURI, Y., 2008. Matrix metalloproteinase 12 accelerates the initiation of

atherosclerosis and stimulates the progression of fatty streaks to fibrous plaques in transgenic rabbits. *The American journal of pathology,* **172**(5), pp. 1419-1429.

YAZDANYAR, A. and JIANG, X.C., 2012. Liver phospholipid transfer protein (PLTP) expression with a PLTP-null background promotes very low-density lipoprotein production in mice. *Hepatology (Baltimore, Md.),* **56**(2), pp. 576-584.

YLA-HERTTUALA, S., 1998. Is oxidized low-density lipoprotein present in vivo? *Current opinion in lipidology,* **9**(4), pp. 337-344.

YOUNG, D.A., LAKEY, R.L., PENNINGTON, C.J., JONES, D., KEVORKIAN, L., EDWARDS, D.R., CAWSTON, T.E. and CLARK, I.M., 2005. Histone deacetylase inhibitors modulate metalloproteinase gene expression in chondrocytes and block cartilage resorption. *Arthritis research & therapy,* **7**(3), pp. R503-12.

ZAITLEN, N., LINDSTROM, S., PASANIUC, B., CORNELIS, M., GENOVESE, G., POLLACK, S., BARTON, A., BICKEBOLLER, H., BOWDEN, D.W., EYRE, S., FREEDMAN, B.I., FRIEDMAN, D.J., FIELD, J.K., GROOP, L., HAUGEN, A., HEINRICH, J., HENDERSON, B.E., HICKS, P.J., HOCKING, L.J., KOLONEL, L.N., LANDI, M.T., LANGEFELD, C.D., LE MARCHAND, L., MEISTER, M., MORGAN, A.W., RAJI, O.Y., RISCH, A., ROSENBERGER, A., SCHERF, D., STEER, S., WALSHAW, M., WATERS, K.M., WILSON, A.G., WORDSWORTH, P., ZIENOLDDINY, S., TCHETGEN, E.T., HAIMAN, C., HUNTER, D.J., PLENGE, R.M., WORTHINGTON, J., CHRISTIANI, D.C., SCHAUMBERG, D.A., CHASMAN, D.I., ALTSHULER, D., VOIGHT, B., KRAFT, P., PATTERSON, N. and PRICE, A.L., 2012. Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS genetics,* **8**(11), pp. e1003032.

ZEIBIG, S., LI, Z., WAGNER, S., HOLTHOFF, H.P., UNGERER, M., BULTMANN, A., UHLAND, K., VOGELMANN, J., SIMMET, T., GAWAZ, M. and MUNCH, G., 2011. Effect of the oxLDL binding protein Fc-CD68 on plaque extension and vulnerability in atherosclerosis. *Circulation research,* **108**(6), pp. 695-703.

ZUK, O., HECHTER, E., SUNYAEV, S.R. and LANDER, E.S., 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America,* **109**(4), pp. 1193-1198.

Original Article

# Genome-Wide Association Study Pinpoints a New Functional Apolipoprotein B Variant Influencing Oxidized Low-Density Lipoprotein Levels But Not Cardiovascular Events

## AtheroRemo Consortium

Kari-Matti Mäkelä, BM, BSc; Ilkka Seppälä, MSc; Jussi A. Hernesniemi, MD, PhD; Leo-Pekka Lyytikäinen, MD; Niku Oksala, MD, PhD, DSc; Marcus E. Kleber, PhD; Hubert Scharnagl, PhD; Tanja B. Grammer, MD; Jens Baumert, PhD; Barbara Thorand, PhD; Antti Jula, MD, PhD; Nina Hutri-Kähönen, MD, PhD; Markus Juonala, MD, PhD; Tomi Laitinen, MD, PhD; Reijo Laaksonen, MD, PhD; Pekka J. Karhunen, MD, PhD; Kjell C. Nikus, MD, PhD; Tuomo Nieminen, MD, PhD, MSc; Jari Laurikka, MD, PhD; Pekka Kuukasjärvi, MD, PhD; Matti Tarkka, MD, PhD; Jari Viik, PhD; Norman Klopp, PhD; Thomas Illig, PhD; Johannes Kettunen, PhD; Markku Ahotupa, PhD; Jorma S.A. Viikari, MD, PhD; Mika Kähönen, MD, PhD; Olli T. Raitakari, MD, PhD; Mahir Karakas, MD; Wolfgang Koenig, MD, PhD; Bernhard O. Boehm, MD; Bernhard R. Winkelmann, MD; Winfried März, MD; Terho Lehtimäki, MD, PhD

***Background*—**Oxidized low-density lipoprotein may be a key factor in the development of atherosclerosis. We performed a genome-wide association study on oxidized low-density lipoprotein and tested the impact of associated single-nucleotide polymorphisms (SNPs) on the risk factors of atherosclerosis and cardiovascular events.

***Methods and Results*—**A discovery genome-wide association study was performed on a population of young healthy white individuals (N=2080), and the SNPs associated with a $P<5\times10^{-8}$ were replicated in 2 independent samples (A: N=2912; B: N=1326). Associations with cardiovascular endpoints were also assessed with 2 additional clinical cohorts (C: N=1118; and D: N=808). We found 328 SNPs associated with oxidized low-density lipoprotein. The genetic variant rs676210 (Pro2739Leu) in apolipoprotein B was the proxy SNP behind all associations ($P=4.3\times10^{-136}$, effect size=13.2 U/L per allele). This association was replicated in the 2 independent samples (A and B, $P=2.5\times10^{-47}$ and $1.1\times10^{-11}$, effect sizes=10.3 U/L and 7.8 U/L, respectively). In the meta-analyses of cohorts A, C, and D (excluding cohort B without angiographic data), the top SNP did not associate significantly with the age of onset of angiographically verified coronary artery disease (hazard ratio=1.00 [0.94–1.06] per allele), 3-vessel coronary artery disease (hazard ratio=1.03 [0.94–1.13]), or myocardial infarction (hazard ratio=1.04 [0.96–1.12]).

***Conclusions*—**This novel genetic marker is an important factor regulating oxidized low-density lipoprotein levels but not a major genetic factor for the studied cardiovascular endpoints. **(*Circ Cardiovasc Genet*. 2013;6:73-81.)**

**Key Words:** atherosclerosis ■ coronary artery disease ■ genome-wide association study ■ lipoproteins ■ oxidative stress

Atherosclerosis, a major cause of disability and death, is a multifactorial disease with a strong hereditary and environmental background.[1]

**Clinical Perspective on p 81**

The conversion of low-density lipoprotein (LDL) to oxidized LDL (oxLDL) increases the atherogenic potential of LDL and is regarded as a key event in the development of fatty streaks, the early atherosclerotic lesions.[2] The removal of oxLDL has been shown to prevent atherosclerosis in mice.[3]

At present, little is known about the genetic factors affecting the susceptibility of LDL to oxidation. As oxLDL plays an important role in the development of atherosclerosis, we wanted to elucidate the possible genetic factors influencing its formation by performing a genome-wide association study (GWAS) on a population of healthy white adults from the Cardiovascular Risk in Young Finns Study (YFS).[4] The genome-wide significant associations were replicated in 2 additional samples: the Ludwigshafen Risk and Cardiovascular Health (LURIC) study[5] and the Kooperative Gesundheitsforschung in der Region Augsburg (Cooperative Health Research in the Augsburg Region, ie, KORA) study.[6,7] The effect of single-nucleotide polymorphisms (SNPs) with a $P<5\times10^{-8}$ on cardiovascular endpoints was further analyzed

Received July 25, 2012; accepted November 16, 2012.

**The online-only Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.111.964965/-/DC1.**

Correspondence to Kari-Matti Mäkelä, Department of Clinical Chemistry, Finn-Medi 2, PO Box 2000, FI-33521 Tampere, Finland. E-mail kari-matti.makela@uta.fi

in 2 additional clinical cohorts, the Finnish Cardiovascular Study (FINCAVAS)[8] and the Angiography and Genes Study (ANGES).[9] Meta-analyses of the association between the top hits and cardiovascular outcomes were carried out on 3 independent cohorts (FINCAVAS, ANGES, and LURIC; KORA did not have angiographic data).

The current study comprises 8244 subjects with measured early subclinical (YFS) or more advanced atherosclerosis-related clinical phenotypes, that is, cardiovascular disease endpoints (FINCAVAS, ANGES, KORA, and LURIC). The mentioned studies were engaged in cooperation with the ongoing AtheroRemo Consortium and have been widely involved with other international genetic consortia (online-only Data Supplement references).

## Materials and Methods

For more details, please see the online-only Data Supplement.

### Study Populations and Ethical Statements

All studies were conducted according to the guidelines of the Declaration of Helsinki, and the study protocols were approved by local ethics committees. All participants gave an informed consent.

The YFS cohort is a Finnish longitudinal population study sample on the evolution of cardiovascular risk factors from childhood to adulthood.[4] The first cross-sectional study was conducted in 1980 at 5 different centers. These subjects were reexamined in 1983 and 1986 as young individuals and in 2001 and 2007 as adults. In the current study, we used the variables measured in 2001. Genotype, risk factor, and phenotype data were available for 2080 subjects, and they formed the current study population.

The LURIC study comprises 3316 white patients who were referred to coronary angiography due to chest pain at a tertiary care center in Southwest Germany between 1997 and 2000.[5] All the necessary covariate and endpoint data were available for 2912 LURIC patients, who were hence included in the present study.

The Monitoring Trends and Determinants in Cardiovascular Disease/KORA Augsburg study (KORA) is a series of population-based surveys conducted in the region of Augsburg in Southern Germany.[7] The data for the current study were drawn from a subcohort randomly selected on the basis of sex and survey from the KORA surveys S1–S3 conducted between 1984 and 1995.[6] Of these, 1326 subjects had all the required covariate and end point data available and were included in the current study.

The FINCAVAS population consists of patients who underwent an exercise stress test at Tampere University Hospital, Finland.[8] From the overall recruited study population, 1118 individuals had all the necessary angiographic, genetic, and covariate data available and were included in the current study.

The ANGES population consisted of 1000 patients with a symptomatic heart disease referred to as coronary angiography to rule out or confirm coronary artery disease (CAD).[9,10] Angiographic, genetic, and covariate data were available for 808 individuals (online-only Data Supplement Table I, and section A1).

### Clinical and Biochemical Characteristics of the Participating Study Cohorts

#### *OxLDL, Lipoprotein, and Apolipoprotein B Measurements*

In YFS, LURIC, and KORA, the circulating oxLDL levels were measured using the same immunoassay. In brief, circulating serum oxLDL levels were assayed with a competitive ELISA utilizing a specific murine monoclonal antibody, mAB 4E6[11] (Mercodia, Uppsala, Sweden; detection limit <0.3 U/L).

In YFS, LDL baseline diene conjugation was measured by determining the level of baseline diene conjugation in lipids extracted from LDL.[12] In all cohorts, standard methods were used for serum cholesterol and apolipoprotein B (apoB) analyses (online-only Data Supplement section A2).

#### *Cardiovascular Endpoint Definitions in FINCAVAS, ANGES, and LURIC*

##### Angiographically Verified CAD

In FINCAVAS, ANGES, and LURIC, coronary angiography (KORA did not have angiographic data) was performed using the standard Judkins technique. Transluminal narrowing of at least 50% in any major coronary artery (left anterior descending, left circumflex, or right coronary artery) was the criterion for the diagnosis of CAD. The number of arteries with significant >50% stenosis was used to determine the severity of CAD. In this study, subjects with no CAD were compared with those with 3-vessel disease.

##### Myocardial Infarction

In LURIC, myocardial infarction (MI) was defined as evidence of any MI (acute, previous, ST-elevation MI, or non-ST-elevation MI).[5] In FINCAVAS, a history of prior MI was based on patient interviews, resting ECG recordings, and hospital records.[8] In ANGES, the clinical diagnosis of MI was based on symptoms, electrocardiographic findings, and biochemical marker tests measuring troponin I and creatine kinase (online-only Data Supplement section A3).[9]

### Genotyping and Quality Control in the Different Cohorts

In brief, in YFS, genotyping was carried out by using a custom-built Illumina Human 670k BeadChip at the Welcome Trust Sanger Institute. Genotype imputation was performed using MACH 1.0[13] and HapMap II CEU (release 22, NCBI build 36, dbSNP 126) samples as a reference. After imputation, 2 543 887 SNPs were available. SNPs with a squared correlation ($r^2$) of <0.30 between imputed and true genotypes were eliminated from the analysis.

In the FINCAVAS, ANGES, and LURIC cohorts, genotyping was performed by using the Metabochip, which is a custom Illumina iSelect genotyping array designed to test ≈200 000 SNPs identified through genome-wide meta-analyses for metabolic and atherosclerotic/cardiovascular diseases and traits. In *KORA*, genotyping was accomplished by using the IBC 50K array, which is an Illumina iSelect genotyping array designed to test ≈50 000 SNPs identified through genome-wide meta-analyses associated with a range of cardiovascular, metabolic, and inflammatory syndromes.[14] The studied top SNP (rs676210) passed quality check (call rate >0.95, minor allele frequency >0.01, Hardy-Weinbert equilibrium $P>10^{-6}$) in all 3 cohorts (online-only Data Supplement, section A4).

### Statistical Analyses

For the GWAS analysis, oxLDL was Box-Cox transformed. Residuals were obtained using a linear regression model in which the variables were adjusted for sex, age, and body mass index, as well as principal components (to control population stratification)[15] and apoB. Tests for additive genetic effects were carried out on a linear scale by means of linear regression. Genotypes were coded as 0, 1, or 2 when the SNP was genotyped and by dosage (scale 0–2) when imputed. In true genotyped SNPs, the minor allele was the effect allele. The imputation software (MACH 1.0) used HapMap II as reference to assign the alleles for imputed SNPs. Tests were performed to assess the association of SNPs with the standardized residuals using PLINK[16] for the genotyped data. ProbABEL[17] was employed to fit the linear regression model, taking into account the genotype uncertainty in imputed SNPs. *P* values were combined from the analysis by favoring genotyped SNPs over imputed ones. Quartile-quartile and Manhattan plots were drawn for the analysis of the results. The *P* value for genome-wide significance was set at $P<5\times10^{-8}$, corresponding to a target $\alpha$ of 0.05 with a Bonferroni correction for 1 million independent tests.

The severity (functionality) of mutations was assessed by PolyPhen-2 version 2.1.0 software.[18] Further statistical analyses were

performed using the R Statistical package v. 2.11.1 (http://www.r-project.org). To define associations nonredundantly associated with oxLDL, we applied a forward-selection algorithm.[19] We associated the SNPs with genome-wide significance (top SNPs) with cardiovascular-disease–related endpoints (angioraphically verified CAD, severity of CAD, and MI) in FINCAVAS, ANGES, and LURIC. The associations were assessed using the appropriate statistical models ($\chi^2$ test, ANOVA, linear regression, or Cox Proportional-Hazards regression) in R. Meta-analyses were performed using a fixed effects model when the *P* for cohort heterogeneity was >0.05. KORA did not have angiographic data and was only used for the replication of the oxLDL association. The YFS participants were young (<39 years of age, average age 31.7 years in 2001), still without major clinical endpoints, and it was therefore not possible to include them in these analyses. $P<0.05$ were considered significant (online-only Data Supplement, section A5).

## Results

### General Characteristics of the Study Populations

There was a predominance of women in the YFS population and a predominance of men in the FINCAVAS, LURIC, and KORA populations (online-only Data Supplement Table I). Furthermore, age differences existed between the cohorts, with YFS as the youngest and LURIC as the oldest population. The mean oxLDL levels also varied, being highest in KORA and lowest in LURIC (online-only Data Supplement Tables I and II).

### GWAS Analysis of Oxidized LDL Pinpoints a Novel Variant Affecting LDL Oxidation

In the YFS discovery GWAS, 328 SNPs were associated with oxLDL with genome-wide statistical significance ($P<5\times10^{-8}$). All statistical adjustments applied produced identical top SNP associations. All of these SNPs were within 210 kb from the apoB-100 precursor coding region (OMIM 107730) on chromosome 2 (see the quartile-quartile plot [online-only Data Supplement Figure I], Manhattan plot [online-only Data Supplement Figure II], and regional plot [Figure 1]).

Using a forward-selection algorithm with a *P* value cut-off of $5\times10^{-8}$, only 1 SNP (rs676210) was independently associated with oxLDL, implicating a role as the proxy for all of the associations. Eleven SNPs had $r^2>0.5$ with rs676210 (rs1042034, rs6728178, rs6754295, rs673548, rs6711016, rs11902417, rs10184054, rs6544366, rs4564803, rs7557067, and rs2678379), and forward-selection algorithm cannot separate, which is the true proxy. We chose to include rs676210 in the subsequent analyses because it is biologically most plausible from its linkage disequilibrium-block and also because it causes a missense mutation. We also ran the GWAS by adjusting for rs676210, in addition to other covariates. No other independent associations were found (online-only Data Supplement Figure III). Moreover, in a haplotype analysis of all associated missense mutations, only rs676210 (or rs1042034, which is in perfect linkage disequilibrium with rs676210) showed an independent effect on oxLDL.

### Structural Testing of the Pro2739Leu (rs676210) and Other Related Missense Variants

The variant rs676210 causes a missense mutation (change of proline to leucine at position 2739) in apoB. This mutation was predicted to be damaging in the analysis performed with the PolyPhen-2 software. This supports the notion of the biological functionality of the amino acid change caused by the SNP. The SNP in perfect linkage disequilibrium with rs676210 (rs1042034 [Ser4338Asn]) was predicted to be benign, which further supports the role of rs676210 as the functional variant. Two other nearby missense mutations (rs533617 [His1923Arg] and rs679899 [Ala618Val]) were also predicted to be probably damaging to apoB but showed no independent effect on oxLDL in the haplotype analyses and were, therefore, not studied further. The other significantly associated missense mutations (rs1801695 [Ala4481Thr], rs1367117 [Thr98Ile], and rs1042031 [Glu4181Lys]) were predicted to be benign and did not show independent effect on oxLDL in the haplotype analyses. As rs676210 was the most probable SNP behind all the found associations, further analyses were conducted with this SNP only.

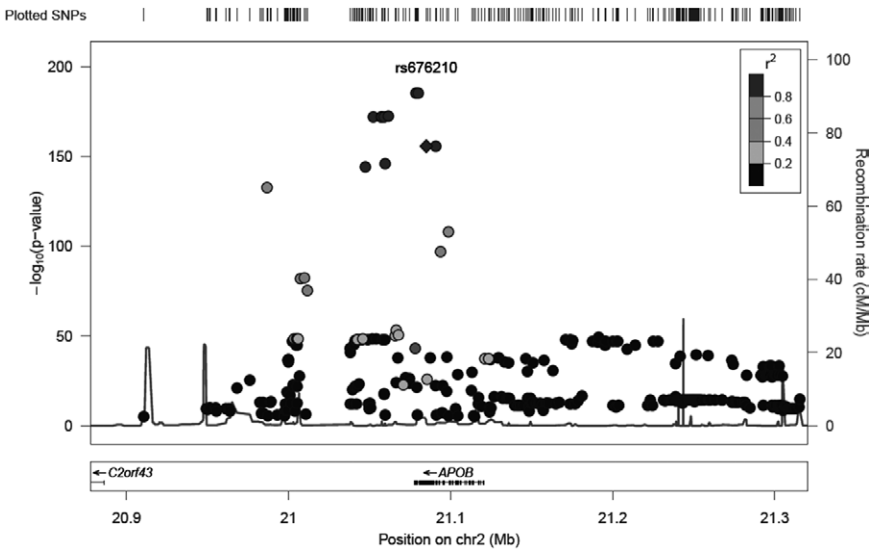



**Figure 1.** Regional plot of the single-nucleotide polymorphisms (SNPs) associated significantly with oxidized low-density lipoprotein. Dots represent –2log10 (*P* values) of SNPs; the color represents the $r^2$ value of the most significantly associated SNP. The gray line shows recombination rates yielded by the HapMap database. The lower part indicates RefSeq genes in the locus. The plot was drawn using LocusZoom version 1.1.[20]

### Replication of the Association of rs676210 With oxLDL in 2 Independent cohorts

In YFS, rs676210 was associated with oxLDL with a *P* value of $4.3\times10^{-136}$ and an effect size of 13.2 U/L oxLDL per allele. The major (risk) allele carriers had significantly higher levels of oxLDL (Table 1). In addition to conventional risk factors, rs676210 explained 11% of the variation in oxLDL ($r^2$=0.11). The association was replicated in 2 independent cohorts, in LURIC with a *P* value of $2.5\times10^{-47}$ and an effect size of 10.5 U/L, and in KORA with a *P* value of $1.1\times10^{-11}$ and an effect size of 7.8 U/L. The association was also strong with the oxLDL/LDL, oxLDL/apoB, and oxLDL/LDL-apoB ratios (Table 1).

In a linear regression model adjusted for age, sex, and body mass index in the YFS, rs676210 also associated significantly with the apoB-epitope-structure-independent measurement of oxLDL, LDL diene conjugation, with a *P* value of 0.028 and an effect size of 0.73 µmol/L, confirming the effect of the studied SNP on LDL oxidation.

Furthermore, rs676210 was significantly associated with apoB in YFS: with triglyceride concentrations in YFS and LURIC; with very-low-density lipoprotein (VLDL) cholesterol concentration in LURIC; with high-density lipoprotein (HDL) cholesterol concentrations in YFS, LURIC, and KORA; and with total cholesterol/HDL cholesterol concentrations in YFS and LURIC (Tables 1).

### Meta-Analyses of the Association Between the apoB Pro2739Leu (rs676210) Mutation and Cardiovascular Outcomes

We found statistically significant differences in traits between genotypes in the angiographic cohorts (FINCAVAS, ANGES, and LURIC; see Tables 1 and 2). Therefore, and because of the pleiotropic effect of rs676210 with lipids, the meta-analyses were performed adjusting for age, sex, body mass index, LDL, HDL, and triglycerides (Figure 2). We also performed the analyses without adjusting for LDL, HDL, and triglycerides to see what proportion of the effects is caused by the pleiotropic effects of rs676210 (online-only Data Supplement Figure IV).

The apoB Pro2739Leu (rs676210 allele G, also the allele causing high oxLDL levels) was not associated with CAD, 3-vessel CAD, or MI after adjustment for age, sex, body mass index, statin use, LDL, HDL, and triglycerides (Figure 2). Results without adjustment for LDL, HDL, and triglycerides were also not significant in the meta-analysis (online-only Data Supplement Figure IV).

## Discussion

Of the 328 SNPs associated with circulating oxLDL levels in a healthy white adult population (YFS), only 1 missense mutation leading to a proline-to-leucine interchange in apoB (SNP rs676210, Pro2739Leu) on chromosome 2 remained significant in further analysis. The association of apoB rs676210 with oxLDL was convincingly replicated in the LURIC and KORA cohorts. We also tested the association of rs676210 with cardiovascular endpoints in a meta-analysis of 3 independent clinical cohorts but did not find significant associations.

The YFS subjects had higher mean oxLDL levels than the LURIC subjects, although the LURIC subjects were older and had more comorbidities. This could simply be due to different storage conditions or differences in the use of statin medication[21] or the levels of oxidants or antioxidants. In LURIC, there was a higher proportion of statin users in comparison with the healthier YFS population (mean oxLDL levels were lower among statin users than among nonusers; data not shown).

There are few previous reports about the genetics of LDL oxidation. To our knowledge, ours is the first GWAS on circulating oxLDL. Employing a forward-selection algorithm, performing haplotype analyses, assessing the damage-producing probability of the SNPs, and running the GWAS with top-SNP adjustment strongly suggest that apoB rs676210 is the most probable functional variant and the proxy for all of the found 328 associations.

Each LDL particle contains 1 apoB moiety. In the current study, the missense mutation of proline to leucine (rs676210, Pro2739Leu) in apoB increased plasma oxLDL levels in a step-wise manner in the genotype order of AA (Leu/Leu, the minor allele), GA (Pro/Leu), and GG (Pro/Pro). Our results are supported by the fact that the variation in apoB probably changes the 3-dimensional structure of apoB[22] in a way that makes LDL less prone to oxidation in homozygous apoB (Leu/Leu) minor allele carriers.

There seem to be some pleiotropic effects for this variation as rs676210 was also associated with HDL cholesterol in all 3 cohorts and with triglyceride levels in YFS and LURIC. The oxLDL association, however, was adjusted for these, so the main effect seems to be on the oxLDL levels.

There are a few studies related to the genetic variation rs676210. In a previous study, the rs676210 minor allele (A) was associated with lower triglyceride, total cholesterol, and LDL cholesterol levels and with higher HDL cholesterol levels, with a $P<5\times10^{-8}$, in comparison with major allele (G) carriers.[23] These results are in accordance with our study. In another earlier report, rs676210 was found to associate with VLDL-related fractions, triglycerides, and mean VLDL/LDL size.[24] In that study, minor allele carriers had larger VLDL/LDL particles and lower VLDL cholesterol and triglyceride concentrations. These findings are also in accordance with our results, showing a linear trend for the minor allele carriers in LURIC to have lower VLDL cholesterol levels (*P*=0.0034). VLDL cholesterol was not measured in YFS with this method.

Interestingly, the minor allele of rs676210 (A) has been linked to an improved response to fenofibrate treatment,[25] the treatment was reported to lower triglyceride levels by 24.7%, 28.3%, and 34.5% according to the rs676210 genotypes GG, GA, and AA, respectively. The allele found to decrease LDL oxidation also seems to improve the response to fenofibrate. Furthermore, minor allele carriers had lower triglyceride levels in YFS and LURIC when compared with major risk allele carriers. We do not have data on fenofibrate use in our cohorts.

It seems that the oxLDL levels associated apoB Pro2739Leu mutation is not associated with cardiovascular endpoints. This was observed across the cohorts studied in the combined meta-analysis. The association of serum oxLDL levels measured with Mercodia ELISA assay with CAD is controversial.

**Table 1. Clinical and Biochemical Trait Profiles of Oxidized LDL Study Cohorts According to rs676210 (Pro2739Leu) Genotype**

| | ApoB rs676210 (Pro2739Leu) Genotype | | | |
|---|---|---|---|---|
| Cohort/Trait | AA | GA | GG | *P* Value |
| YFS | (n=149) | (n=806) | (n=1125) | |
| OxLDL (U/L) | 61.7 (20) | 76.2 (22) | 91.3 (24.6) | 4.3E-136 |
| OxLDL/LDL (U/L per mmol per L) | 19.8 (5.98) | 23.7 (5.12) | 28 (5.22) | 2E-113 |
| OxLDL/apoB (U/L per g per L) | 60.9 (17.1) | 73.4 (13.9) | 85.7 (13.8) | 4.9E-124 |
| Sex (% male) | 67 (45%) | 357 (44.3%) | 516 (45.9%) | 0.789 |
| Hypertension (%) | 5 (3.36%) | 20 (2.48%) | 30 (2.67%) | 0.828 |
| Diabetes mellitus (%) | 2 (1.34%) | 3 (0.372%) | 7 (0.622%) | 0.341 |
| Statin use (%) | 0 (0%) | 2 (0.248%) | 5 (0.444%) | 0.582 |
| Age (y) | 31.7 (5.09) | 31.8 (4.9) | 31.7 (5.04) | 0.694 |
| BMI (kg/m$^2$) | 25.3 (3.97) | 25 (4.45) | 25.2 (4.5) | 0.636 |
| Triglycerides (mmol/L) | 1.33 (1.01) | 1.27 (0.747) | 1.37 (0.884) | 0.0483 |
| Total cholesterol (mmol/L) | 5.03 (0.919) | 5.13 (0.964) | 5.17 (0.979) | 0.109 |
| LDL cholesterol (mmol/L) | 3.15 (0.828) | 3.25 (0.844) | 3.28 (0.827) | 0.0801 |
| HDL cholesterol (mmol/L) | 1.32 (0.337) | 1.31 (0.321) | 1.27 (0.308) | 0.00605 |
| Total/HDL cholesterol | 4.05 (1.25) | 4.16 (1.39) | 4.3 (1.37) | 0.00671 |
| apoB (g/L) | 1.02 (0.261) | 1.04 (0.261) | 1.07 (0.266) | 0.00501 |
| LURIC | (n=165) | (n=986) | (n=1761) | |
| OxLDL (U/L) | 58.9 (52.6) | 68.6 (26.2) | 80.1 (24.5) | 2.5E-47 |
| OxLDL/LDL (U/L per mg per dL) | 0.507 (0.393) | 0.614 (0.263) | 0.74 (0.312) | 9.7E-36 |
| OxLDL/apoB (U/L per mg per dL) | 0.554 (0.414) | 0.659 (0.223) | 0.768 (0.201) | 4.3E-49 |
| OxLDL/LDL-apoB (U/L per mg per dL) | 0.69 (0.53) | 0.825 (0.309) | 0.981 (0.357) | 7.2E-38 |
| Sex (% male) | 119 (72.1%) | 676 (68.6%) | 1224 (69.5%) | 0.636 |
| Hypertension (%) | 122 (73.9%) | 706 (71.6%) | 1285 (73%) | 0.684 |
| Diabetes mellitus (%) | 26 (15.8%) | 175 (17.7%) | 305 (17.3%) | 0.819 |
| Statin use (%) | 80 (48.5%) | 447 (45.3%) | 819 (46.5%) | 0.701 |
| Age (y) | 62.6 (10.8) | 62.9 (10.6) | 62.5 (10.8) | 0.447 |
| BMI (kg/m$^2$) | 26.9 (3.44) | 27.3 (4.12) | 27.5 (4.1) | 0.0501 |
| Triglycerides (mg/dL) | 154 (84) | 169 (115) | 178 (125) | 0.00487 |
| Total cholesterol (mg/dL) | 194 (35.8) | 192 (39.3) | 192 (38.6) | 0.933 |
| LDL cholesterol (mg/dL) | 119 (33.8) | 116 (34.9) | 115 (33.5) | 0.111 |
| VLDL cholesterol (mg/dL) | 34.3 (25.8) | 36.3 (25.4) | 38.8 (27.5) | 0.0034 |
| HDL cholesterol (mg/dL) | 40.6 (12.5) | 39.1 (10.5) | 38.2 (10.7) | 0.00168 |
| Total/HDL cholesterol | 5.12 (1.63) | 5.31 (3.53) | 5.37 (1.77) | 0.245 |
| apoB (mg/dL) | 104 (22.4) | 104 (25.2) | 105 (24.2) | 0.236 |
| LDL-apoB (mg/dL) | 85.1 (21.3) | 84.7 (22.9) | 84.5 (21.9) | 0.697 |
| KORA | (n=61) | (n=462) | (n=803) | |
| OxLDL (U/L) | 79.9 (22.7) | 88.0 (24.2) | 96.8 (26.7) | 1.05E-11 |
| OxLDL/LDL (U/L per mg per dL) | 0.583 (0.126) | 0.627 (0.154) | 0.674 (0.167) | 0.000000235 |
| Sex (% male) | 33 (54.1%) | 230 (49.8%) | 441 (54.9%) | 0.209 |
| Hypertension (%) | 23 (37.7%) | 194 (42.0%) | 355 (44.2%) | 0.507 |
| Diabetes mellitus (%) | 2 (3.3%) | 28 (6.1%) | 48 (6.0%) | 0.675 |
| Statin use (%) | 0 (0%) | 2 (0.43%) | 4 (0.50%) | 0.853 |
| Age (y) | 50.9 (11.0) | 52.5 (10.6) | 52.9 (10.5) | 0.341 |
| BMI (kg/m$^2$) | 26.7 (4.1) | 27.1 (3.9) | 27.3 (4.2) | 0.513 |
| Triglycerides (mg/dL) | 145.6 (83.2) | 167.8 (107.5) | 180.4 (127.0) | 0.334 |
| Total cholesterol (mg/dL) | 232.7 (40.7) | 237.1 (42.2) | 238.8 (46.4) | 0.524 |
| LDL cholesterol (mg/dL) | 142.8 (40.2) | 147.7 (41.7) | 150.0 (43.9) | 0.455 |
| HDL cholesterol (mg/dL) | 55.5 (15.0) | 57.7 (17.3) | 55.1 (16.5) | 0.03 |
| Total/HDL cholesterol | 4.52 (1.59) | 4.50 (1.86) | 4.73 (1.84) | 0.079 |

Statistics: values are numbers (percentages) in cases of categorical data and means (SDs) in cases of continuous data; *P* values (difference between rs676210 genotype groups) calculated with $\chi^2$ test for categorical data, with a linear regression model for oxLDL, oxLDL/LDL, and oxLDL/apoB, and with ANOVA for other continuous data. The OxLDL and oxLDL/LDL models have been adjusted for age, sex, BMI, HDL, apoB, triglycerides, and current smoking in YFS and LURIC. The OxLDL and oxLDL/LDL models have been adjusted for age, sex, survey number, and BMI in KORA. The OxLDL/apoB models have been adjusted for age, sex, BMI, HDL, and triglycerides. Conversion factors: total cholesterol, LDL cholesterol, and HDL cholesterol 1 mg/dL=0.0259 mmol/L; triglycerides: 1 mg/dL=0.0113 mmol/L; apoB: 1 mg/dL=0.01 g/L. apoB indicates apolipoprotein B; BMI, body mass index; HDL, high-density lipoprotein; KORA, Kooperative Gesundheitsforschung in der Region Augsburg Study; LURIC, the Ludwigshafen Risk and Cardiovascular Health Study; OxLDL, oxidized low-density lipoprotein (LDL); VLDL, very-low-density lipoprotein; and YFS, the Cardiovascular Risk in Young Finns Study.

**Table 2. Risk Profiles of the Additional Meta-Analysis Study Cohorts According to apoB rs676210 (Pro2739Leu) Genotype**

| | ApoB rs676210 (Pro2739Leu) Genotype | | | |
|---|---|---|---|---|
| Cohort/Trait | AA | AG | GG | P Value |
| FINCAVAS | (n=99) | (n=408) | (n=608) | |
| Age, (y) | 60.2 (10.4) | 60 (10.7) | 59.9 (10.2) | 0.858 |
| Sex (% female) | 26 (26.3%) | 118 (28.9%) | 178 (29.3%) | 0.828 |
| BMI (kg/m2) | 28.2 (4.14) | 27.6 (4.36) | 27.8 (4.56) | 0.814 |
| Hypertension (%) | 88 (88.9%) | 357 (87.5%) | 540 (88.8%) | 0.856 |
| Diabetes mellitus (%) | 17 (17.2%) | 50 (12.3%) | 69 (11.3%) | 0.285 |
| Statin use (%) | 53 (53.5%) | 222 (54.4%) | 348 (57.2%) | 0.597 |
| LDL cholesterol (mmol/L) | 2.63 (0.791) | 2.67 (0.934) | 2.75 (0.931) | 0.112 |
| HDL cholesterol (mmol/L) | 1.32 (0.444) | 1.28 (0.409) | 1.27 (0.379) | 0.337 |
| Triglycerides (mmol/L) | 1.42 (0.837) | 1.42 (0.742) | 1.61 (1.97) | 0.0813 |
| Total cholesterol mmol/L | 4.58 (0.879) | 4.59 (1.04) | 4.73 (1.15) | 0.0475 |
| ANGES | (n=54) | (n=309) | (n=441) | |
| Age, (y) | 65.1 (9.09) | 62.8 (10.6) | 62.6 (9.61) | 0.217 |
| Sex (% female) | 18 (33.3%) | 119 (38.5%) | 155 (35.1%) | 0.573 |
| BMI (kg/m2) | 28.2 (4.07) | 27.6 (4.1) | 28.4 (4.46) | 0.0908 |
| Hypertension (%) | 51 (94.4%) | 286 (92.6%) | 408 (92.5%) | 0.873 |
| Diabetes mellitus (%) | 14 (25.9%) | 89 (28.8%) | 126 (28.6%) | 0.894 |
| Statin use (%) | 37 (68.5%) | 198 (64.1%) | 304 (68.9%) | 0.368 |
| LDL cholesterol (mmol/L) | 2.57 (0.656) | 2.79 (0.783) | 2.77 (0.826) | 0.434 |
| HDL cholesterol (mmol/L) | 1.19 (0.373) | 1.19 (0.351) | 1.16 (0.311) | 0.297 |
| Triglycerides (mmol/L) | 1.5 (0.923) | 1.52 (0.806) | 1.58 (0.972) | 0.314 |
| Total cholesterol (mmol/L) | 4.42 (0.993) | 4.5 (1.02) | 4.5 (0.938) | 0.719 |

Statistics: Values are numbers (percentages) in cases of categorical data and means (standard deviations) in cases of continuous data; p values are calculated with chi-squared for categorical variables and with analysis of variance (ANOVA) for continuous variables. Conversion factors: total cholesterol, LDL cholesterol, and HDL cholesterol 1 mg/dl = 0.0259 mmol/l; triglycerides: 1 mg/dl = 0.0113 mmol/l. ANGES indicates The Angiography and Genes Study; BMI, body mass index; FINCAVAS, The Finnish Cardiovascular Study; HDL, high-density lipoprotein; and LDL, low-density lipoprotein.

Many smaller studies report oxLDL as a predictor of CAD; however, larger studies have been negative after adjusting for standard lipid variables.[26] Our results are in line with these larger nongenetic studies by showing that serum oxLDL levels associated gene variant is not associated with the cardiovascular endpoints.

## Limitations and Strengths of the Study

A strength of the study is the replication of the major results in multiple independent cohorts. The association of rs676210 with LDL oxidation is convincing. The oxidation of LDL was, however, measured by structural changes in the apoB protein and does not fully account for the lipid component of these apoB-containing particles. To confirm that our results are not due to the possibility of the Pro2739Leu substitution altering the binding of the used monoclonal antibody to apoB, we also applied a second, apoB-epitope-structure-independent assay to assess the effect of the SNP on the LDL oxidation, with parallel results. In this independent assay, the SNP rs676210 was significantly associated with LDL diene conjugation. These results put together show that our findings are not due to the oxLDL assay employed and imply a true biological function for rs676210 in LDL oxidation.

The step-wise algorithm and top-SNP–adjusted GWAS strongly suggest that rs676210 is the proxy SNP. We also carried out manual haplotype analyses, arriving at the same conclusion of rs676210 as the implied proxy. This information, together with the predicted damaging change caused in apoB (proline-to-leucine interchange), indicates that out of the available SNPs, this is the most likely functional top-SNP candidate. The SNP in perfect linkage disequilibrium with rs676210 was predicted to be benign. Two other nearby missense mutations were also predicted to be damaging but showed no independent effect in the haplotype analysis nor with the forward-selection algorithm; other associated missense mutations were predicted not to be damaging. These results strongly suggest that rs676210 is the true functional variant. However, it is important to note that the true proxy cannot be found with 100% certainty by means of bioinformatic methods, and, therefore, we performed the analyses with the most probable one.

The effect was studied only in whites and might not be generalizable to other populations. There were significant

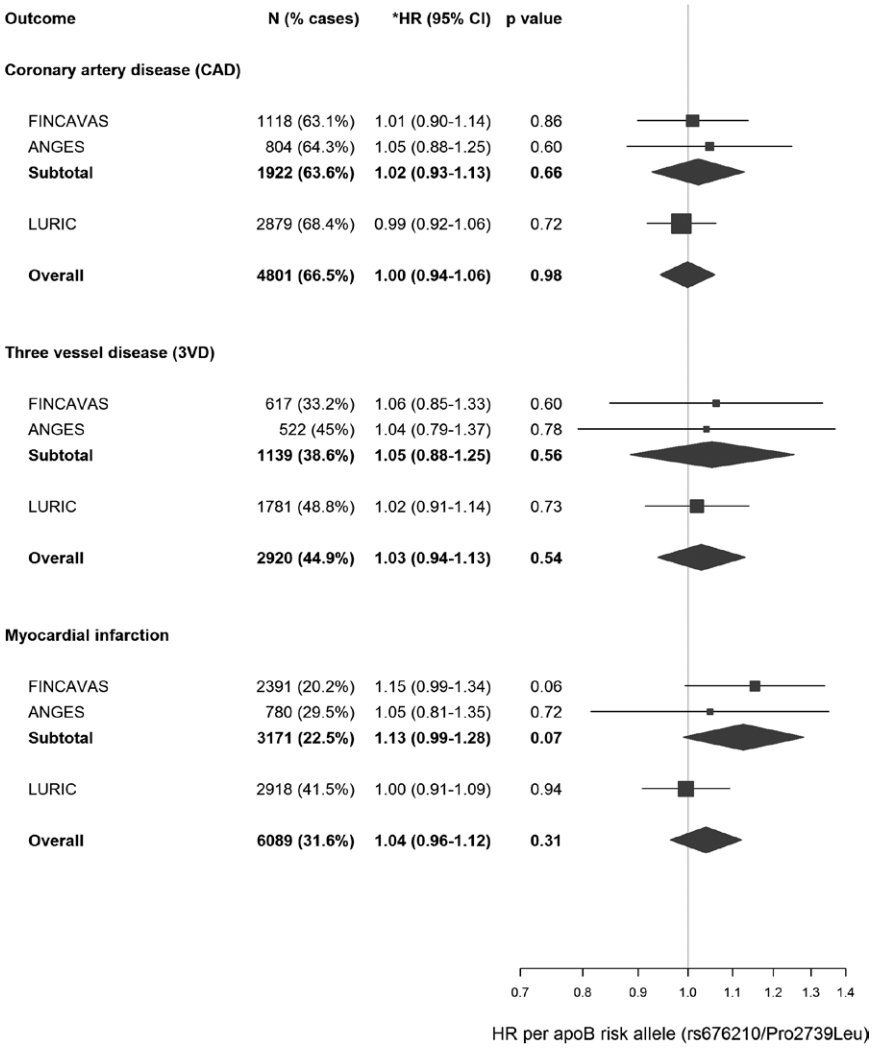



**Figure 2.** Meta-analyses of the association between apolipoprotein B alleles (rs676210/Pro2739Leu) and the age of onset of cardiovascular outcomes in 3 independent clinical study cohorts. Statistics: *Hazard ratios (HR; gray squares, sizes indicate the relative weight of each study) and 95% confidence intervals (CI; horizontal lines) were estimated by a per-risk-allele (G) additive Cox Proportional-Hazards regression using age as time variable adjusted for sex, body mass index (BMI), low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides, and statin use. The fixed effect meta-analysis (*P* value for heterogeneity >0.05 for all) estimates are shown as gray diamonds. ANGES indicates the Angiography and Genes Study; CAD, coronary artery disease, with or without angiographically verified 50% stenosis; FINCAVAS, Finnish Cardiovascular Study; LURIC, Ludwigshafen Risk and Cardiovascular Health Study; MI, myocardial infarction, with or without MI; and 3VD, 3-vessel disease, with or without angiographically verified 50% stenosis in three vessels.

differences in sex distributions and mean ages between the cohorts. However, the association of rs676210 with oxLDL concentration remained significant despite these and other differences in the background populations.

In comparison, the cohorts used for the clinical endpoint association assessments were quite similar. The current analyses related to oxLDL did not include the FINCAVAS or ANGES study populations because oxLDL was not measured in them.

The clinical implications of rs676210 leading to Pro2739Leu missense mutation require further investigation. However, we did not find significant associations in a meta-analysis of 3 independent studies.

## Conclusions

We found an SNP that is important in regulating oxLDL levels but not a major genetic factor for studied cardiovascular endpoints.

## Acknowledgments

For the YFS, the expert technical assistance in data management and statistical analyses done by Irina Lisinen and Ville Aalto are gratefully acknowledged. For the LURIC study, we extend our appreciation to the participants of the LURIC study; without their collaboration, this article would not have been written. We thank the LURIC study team that was either temporarily or permanently involved in patient recruitment as well as sample and data handling, in addition to the laboratory staff at the Ludwigshafen General Hospital and the Universities of Freiburg, Ulm, and Graz, Germany. For the KORA study, we would like to thank Gerlinde Trischler (University of Ulm) for her expert technical assistance.

## Sources of Funding

This work was supported by the European Union 7th Framework Program funding for the AtheroRemo project (201668). The Cardiovascular Risk in Young Finns Study has been financially supported by the Academy of Finland (grants 126925, 121584, 124282, 129378 [Salve], 117787 [Gendi], and 41071 [Skidi]); the Social Insurance Institution of Finland; the Kuopio, Tampere, and Turku University Hospital Medical Funds (grant 9M048 and 9N035 for TeLeht); the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Finnish Foundation of Cardiovascular Research; the Finnish Cultural Foundation; as well as the Tampere Tuberculosis Foundation and the Emil Aaltonen Foundation (Dr Lehtimäki). The Monitoring Trends and Determinants in Cardiovascular Disease/KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, and supported by grants from the German Federal Ministry of Education and Research. A part of the work was financed by the German National Genome Research Network (NGFNPlus, project

number 01GS0834), the German Research Foundation (TH-784/2–2), and additional funds from the University of Ulm. Furthermore, the research was supported within the Munich Center of Health Sciences (MC Health) as part of the Ludwig-Maximilians University innovative. The LURIC study received funding through the 6th Framework Program (integrated project Bloodomics, grant LSHM-CT-2004–503485) and the 7th of Framework Program (integrated project AtheroRemo, Grant Agreement number 201668) of the European Union. Mercodia AB, Uppsala, provided the reagents for the determination of oxLDL free of charge but did not assume any other role in the conducting of this study. FINCAVAS and ANGES have been financially supported by the Tampere University Hospital Medical Fund (grants 9M048 and 9N035 for TeLeht), the Tampere Tuberculosis Foundation, and the Emil Aaltonen Foundation (Dr Lehtimäki). This work has been financially supported by a grant awarded to Dr Mäkelä by the Urho Känkänen foundation.

## Disclosures

None.

## Appendix

From the Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital and School of Medicine, University of Tampere, Tampere, Finland (K.-M.M., I.S., J.A.H., L.-P.L., N.O., R.L., T. Lehtimäki); Department of Internal Medicine, North Karelia Central Hospital, Joensuu, Finland (J.A.H.); Division of Vascular Surgery, Department of Surgery, Tampere University Hospital, Tampere, Finland (N.O.); LURIC nonprofit LLC, Freiburg im Breisgau and Mannheim Institute of Public Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany (M.E.K.); Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria (H.S.); Mannheim Institute of Public Health, Social and Preventive Medicine, Medical Faculty Mannheim, Rupprecht Karls University, Heidelberg, Germany (T.B.G.); Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany (J.B., B.T.); Population Studies Unit, National Institute of Health and Welfare, Turku, Finland (A.J.); Department of Pediatrics, University of Tampere and Tampere University Hospital, Tampere, Finland (N.H.-K.); Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Internal Medicine, Turku University Hospital, Turku, Finland (M.J.); Department of Clinical Physiology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland (T. Laitinen); Department of Forensic Medicine, School of Medicine at the University of Tampere and Centre for Laboratory Medicine, Tampere University Hospital, Tampere, Finland (P.J.K.); Department of Cardiology, Heart Center, Tampere University Hospital and University of Tampere, Tampere, Finland (K.C.N.); Division of Cardiology, Helsinki University Central Hospital, Helsinki, Finland (T.N.); Department of Internal Medicine, Päijät-Häme Central Hospital, Lahti, Finland (T.N.); Department of Cardio-Thoracic Surgery, Heart Center, Tampere University Hospital and University of Tampere, Tampere, Finland (J.L., P.K., M.T.); Department of Biomedical Engineering, Tampere University of Technology and BioMediTech, Tampere, Finland (J.V.); Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Munich, Germany (N.K., T.I.); Hannover Unified Biobank, Hannover Medical School, Hannover, Germany (N.K., T.I.); Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland (J.K.); Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland (J.K.); MCA Research Laboratory, Department of Physiology (M.A.) and Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland (O.T.R.); Department of Medicine, University of Turku and Turku University Hospital, Turku, Finland (J.S.A.V.); Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland (O.T.R.); Department of Clinical Physiology, Tampere University Hospital and University of Tampere, Tampere, Finland (M. Kähönen); Department of Internal Medicine II–Cardiology, University of Ulm Medical Center, Ulm, Germany (M. Karakas, W.K.); Division of Endocrinology, Diabetes and Metabolism, Graduate School of Molecular Diabetology and Endocrinology, Ulm University, Ulm, Baden-Württemberg, Germany (B.O.B.); ClinPhenomics Study Center and University of Frankfurt, Frankfurt, Germany (B.R.W.); Synlab Services GmbH, Mannheim and Mannheim Institute of Public Health, Medical Faculty Mannheim, University of Heidelberg, Germany and Clinical Institute of Medical and Chemical Laboratory Diagnostics Medical University of Graz, Graz, Austria (W.M.).

## References

1. Packard RR, Libby P. Inflammation in atherosclerosis: from vascular biology to biomarker discovery and risk prediction. *Clin Chem*. 2008;54:24–38.
2. Steinberg D. Low density lipoprotein oxidation and its pathobiological significance. *J Biol Chem*. 1997;272:20963–20966.
3. Ishigaki Y, Katagiri H, Gao J, Yamada T, Imai J, Uno K, et al. Impact of plasma oxidized low-density lipoprotein removal on atherosclerosis. *Circulation*. 2008;118:75–83.
4. Raitakari OT, Juonala M, Rönnemaa T, Keltikangas-Järvinen L, Räsänen L, Pietikäinen M, et al. Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol*. 2008;37:1220–1226.
5. Winkelmann BR, März W, Boehm BO, Zotz R, Hager J, Hellstern P, et al; LURIC Study Group (Ludwigshafen RIsk and Cardiovascular Health). Rationale and design of the LURIC study–a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics*. 2001;2(Suppl 1):S1–S73.
6. Thorand B, Schneider A, Baumert J, Doring A, Marowsky-Koppl M, Heier M, et al. Case-cohort studies: an effective design for the investigation of biomarkers as risk factors for chronic diseases--demonstrated by the example of the MONICA/KORA Augsburg Case-Cohort Study 1984–2002. *Gesundheitswesen*. 2005;67(Suppl 1):S98–S102.
7. Löwel H, Döring A, Schneider A, Heier M, Thorand B, Meisinger C; MONICA/KORA Study Group. The MONICA Augsburg surveys–basis for prospective cohort studies. *Gesundheitswesen*. 2005;67(Suppl 1):S13–S18.
8. Nieminen T, Lehtinen R, Viik J, Lehtimäki T, Niemelä K, Nikus K, et al. The Finnish Cardiovascular Study (FINCAVAS): characterising patients with high risk of cardiovascular morbidity and mortality. *BMC Cardiovasc Disord*. 2006;6:9.
9. Raitoharju E, Seppälä I, Levula M, Kuukasjärvi P, Laurikka J, Nikus K, et al. Common variation in the ADAM8 gene affects serum sADAM8 concentrations and the risk of myocardial infarction in two independent cohorts. *Atherosclerosis*. 2011;218:127–133.
10. Mennander A, Kuukasjärvi P, Laurikka J, Nikus K, Karhunen PJ, Tarkka M, et al. Diagnostic performance of plasma high sensitive C-reactive protein in detecting three-vessel coronary artery disease: modification by apolipoprotein E genotype. *Scand J Clin Lab Invest*. 2008;68:714–719.
11. Holvoet P, Stassen JM, Van Cleemput J, Collen D, Vanhaecke J. Oxidized low density lipoproteins in patients with transplant-associated coronary artery disease. *Arterioscler Thromb Vasc Biol*. 1998;18:100–107.
12. Ahotupa M, Marniemi J, Lehtimäki T, Talvinen K, Raitakari OT, Vasankari T, et al. Baseline diene conjugation in LDL lipids as a direct measure of *in vivo* LDL oxidation. *Clin Biochem*. 1998;31:257–261.
13. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34:816–834.
14. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, et al. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS ONE*. 2008;3:e3583.
15. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–909.
16. 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;81:559–575.
17. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*. 2010;11:134.
18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248–249.
19. Paré G, Chasman DI, Parker AN, Zee RR, Mälarstig A, Seedorf U, et al. Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of

13 974 participants in the Women's Genome Health Study. *Circ Cardiovasc Genet*. 2009;2:142–150.
20. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–2337.
21. Homma Y, Michishita I, Hayashi H, Shigematsu H; Kanagawa Lipid Research Group. Effects of low-dose simvastatin on the distribution of plasma cholesterol and oxidized low-density lipoprotein in three ultracentrifugally separated low-density lipoprotein subfractions: 12-month, open-label trial. *J Atheroscler Thromb*. 2010;17:1049–1053.
22. Kumar V, Butcher SJ, Öörni K, Engelhardt P, Heikkonen J, Kaski K, et al. Three-dimensional cryoEM reconstruction of native LDL particles to 16Å resolution at physiological body temperature. *PLoS ONE*. 2011;6:e18841.
23. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707–713.
24. Chasman DI, Paré G, Mora S, Hopewell JC, Peloso G, Clarke R, et al. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet*. 2009;5:e1000730.
25. Wojczynski MK, Gao G, Borecki I, Hopkins PN, Parnell L, Lai CQ, et al. Apolipoprotein B genetic variants modify the response to fenofibrate: a GOLDN study. *J Lipid Res*. 2010;51:3316–3323.
26. Wu T, Willett WC, Rifai N, Shai I, Manson JE, Rimm EB. Is plasma oxidized low-density lipoprotein, measured with the widely used antibody 4E6, an independent predictor of coronary heart disease among U.S. men and women? *J Am Coll Cardiol*. 2006;48:973–979.

## CLINICAL PERSPECTIVE

Atherogenic oxidized low-density lipoprotein (LDL) is regarded as a key element in atherosclerosis. By performing a genome-wide association analysis, we identified the most important single-nucleotide polymorphism associating with circulating levels of oxidized LDL. The discovered single-nucleotide polymorphism (rs676210) leads to a Pro2739Leu missense mutation in the apolipoprotein B gene, likely rendering LDL less prone to oxidation and resulting in significantly lower levels of circulating oxidized LDL. Oxidized LDL levels have previously been reported to predict coronary artery disease but often not independently after adjusting for standard lipid variables. Correspondingly, we did not observe statistically significant associations between rs676210 and the age of onset of cardiovascular endpoints in a meta-analysis of 3 independent patient populations undergoing angiography (N=4801). However, population stratification by genetic and environmental factors regulating the oxidation of LDL could potentially lead to improved accuracy in patient selection for interventions such as statin therapy, novel antioxidant-molecule-based therapies, or diet-based prevention strategies. Supporting this, the discovered single-nucleotide polymorphism has previously been found to associate with an improved response to fenofibrate treatment, with the minor (protective) allele associating with greater reductions in triglyceride levels. Our study provides a useful framework for future investigations to elucidate the oxidative processes underlying coronary artery disease and to investigate how genes and environmental exposures (such as smoking) interact in causing the condition.

# Supplemental Material

## Supplementary introduction

Atherosclerosis, a major cause of disability and death, is a multifactorial disease with a strong hereditary and environmental background [1-3].

The conversion of low-density lipoprotein (LDL) to oxidized LDL (oxLDL) increases the atherogenic potential of LDL and is regarded as a key event in the development of fatty streaks—the early atherosclerotic lesions [4-8]. OxLDL can be found in the macrophages in atherosclerotic lesions, but not in healthy arteries [9]. Macrophages do not take up oxLDL via the LDL receptor [10], but instead via scavenger receptors [4-8, 11]. The binding of oxLDL to the scavenger receptors, and their subsequent internalization, is a key step in the accumulation of cholesterol in macrophages, which transforms them into lipid-laden foam cells [11].

The mentioned studies were engaged in cooperation with the ongoing AtheroRemo Consortium [12] and have been widely involved with other international genetic consortia [13-19].

## A. Supplementary Materials and Methods

### 1. Detailed cohort descriptions

*The Cardiovascular Risk in Young Finns Study (YFS)*

The first cross-sectional study was conducted in 1980 at five different centers. It included 3,596 participants in the age groups of 3, 6, 9, 12, 15, and 18. The participants were randomly selected from the national population register. Following recruitment in 1980, these subjects have been re-examined in 1983 and 1986 as young individuals and in 2001 and 2007 as

adults. During the follow-up in 2001, a total of 2,283 participants aged 24–39 years were examined for numerous study variables, including serum lipoproteins, glucose, insulin, obesity indices, blood pressure, lifestyle factors, smoking status, alcohol use, and general health status. [20]

*The Finnish Cardiovascular study (FINCAVAS)*

The exercise test indications were a diagnosis of CAD, a post-MI assessment, evaluation of drug therapy, arrhythmia, assessment of performance (working capacity), or an evaluation prior to surgery. The purpose of FINCAVAS is to construct a risk profile of individuals at high risk of cardiovascular diseases, events, and deaths. FINCAVAS has an extensive set of data on patient history, genetic variation (the Metabochip), cardiovascular parameters, ECG markers, and follow-up data on clinical events, hospitalizations, and deaths. Of the patients included, 43.6% also underwent coronary angiography. [21]

*The Angiography and Genes Study (ANGES)*

The population studied consists of 1,000 Finnish individuals participating in the ongoing ANGES study. Angiographic, genetic, and covariate data was available for 808 individuals (516 men and 292 women; mean age 62 ± 10). The data was collected between September 2002 and July 2005. All patients underwent coronary angiography at Tampere University Hospital due to clinically suspected coronary artery disease [22]. The study is a cross-sectional study, and after the angiography, patients were treated according to the Finnish Current Care Guidelines. Patients were also interviewed by a study nurse, and a questionnaire was used to collect general information—age, sex, body mass index, alcohol consumption, smoking, medication as well as traditional risk factors of atherosclerosis and myocardial infarction

(MI). The whole study has been approved by the Ethics Committee of Pirkanmaa Hospital District, and written informed consent was obtained from each patient. [22]

**2. Other clinical and biochemical characteristics in more detail**

*Oxidized LDL, lipid, and apolipoprotein B (apoB) measurements*

*In the YFS, LURIC, and KORA,* the circulating oxidized LDL levels were measured using the same immunoassay. In brief, circulating serum oxLDL levels were assayed with a competitive enzyme-linked immunosorbent assay (ELISA) utilizing a specific murine monoclonal antibody, mAB 4E6[23] (Mercodia, Uppsala, Sweden; detection limit < 0.3 U/l). The monoclonal antibody is directed against a conformational epitope in the apoB-100 moiety of LDL that is generated by substituting at least 60 lysine residues of apolipoprotein B-100 with aldehydes. The substitute aldehydes can be produced by the peroxidation of LDL lipids, resulting in the generation of oxLDL. Aldehydes released by endothelial cells under oxidative stress or by activated platelets may also induce the oxidative modification of apolipoprotein B-100 in the absence of the peroxidation of LDL lipids [24].

*In YFS,* LDL baseline diene conjugation was measured by determining the level of baseline diene conjugation in lipids extracted from LDL [25]. In brief, serum LDL was isolated by means of precipitation with buffered heparin. Lipids were extracted from LDL samples with chloroform-methanol, dried under nitrogen, then redissolved in cyclohexane and analyzed spectrophotometrically at 234 nm.

*In YFS, LURIC, ANGES, and FINCAVAS,* venous blood samples were drawn after an overnight fast. *In KORA,* non-fasting venous blood samples were drawn. Standard enzymatic methods were used for serum total cholesterol, triglycerides, and high-density lipoprotein (HDL) cholesterol. LDL cholesterol was calculated in YFS, KORA, ANGES, and FINCAVAS by the Friedewald formula: LDL cholesterol = total cholesterol – HDL

cholesterol – triglyceride-concentration / 2.2 [26]. In LURIC, lipoproteins were separated by means of a combined ultracentrifugation–precipitation method (ß-quantification) [27].

*In YFS,* apoB was analyzed immunoturbidometrically (Orion Diagnostica, Espoo, Finland); the interassay CV was 2.8% for apoB [28]. *In LURIC*, apoB was analyzed with a photometric assay using an antihuman apoB antibody (apoB Test, Rolf Greiner Bio-chemica, Flacht, Germany).

*In YFS, ANGES, and FINCAVAS,* participants were classified as having type 2 diabetes mellitus (T2DM) if they *(i)* had a fasting plasma glucose level of ≥7.0 mmol/L (≥125 mg/dL); or *(ii)* reported receiving oral hypoglycemic agents and/or insulin injections and did not have type 1 diabetes mellitus; or *(iii)* reported a history of physician-diagnosed T2DM, which is consistent with the World Health Organization definition. Women who reported having physician-diagnosed diabetes mellitus only during the term of their pregnancy were considered to have had gestational diabetes and were classified as not currently having T2DM, provided that their plasma glucose levels were not ≥7.0 mmol/L (≥125 mg/dL).

*In LURIC*, Individuals were classified as having diabetes mellitus if their plasma glucose level was > 125 mg/dL (7 mmol/l) in the fasting state or >200 mg/dL (11 mmol/l) 2 h after the oral glucose load (performed on individuals with no previous diabetes mellitus diagnosis), or if individuals were receiving oral antidiabetics or insulin.

*In all studies*, hypertension was diagnosed if the systolic and/or diastolic blood pressure exceeded 140 and/or 90 mmHg, respectively, or if there was a significant history of hypertension (use of antihypertensive medication).

*In YFS, FINCAVAS, ANGES, and LURIC*, dyslipidemia was defined as usage of lipid-lowering medication or having a triglyceride concentration of over 2 mmol/l, an LDL cholesterol concentration of over 3 mmol/l (LURIC 4 mmol/l), an HDL cholesterol

concentration below 1 mmol/l, or an HDL/total cholesterol concentration of over 4 mmol/l. In

In YFS, CRP was analyzed by an automated analyzer (Olympus AU400) with a latex turbidimetric immunoassay kit (CRP-UL assay, Wako Chemicals, Neuss, Germany). The detection limit reported by the manufacturer for the assay was 0.06 mg/l.

*In LURIC*, 'sensitive' CRP was measured by means of immunonephelometry on a Behring Nephelometer II (N High Sensitivity CRP, Dade Behring, Marburg, Germany) after the completion of the patient recruitment in 2001 on samples stored at -80 °C. In the CRP assay used, the limit of detection for CRP is >0.17 mg/L; it is linear up to 500 mg/L. The lowest and the highest CRP concentrations encountered in this study were 0.17 mg/L and 269 mg/L, respectively. *In KORA,* CRP concentrations were measured using a high-sensitivity immunoradiometric assay (IRMA; range, 0.05–10 mg/l; Survey 1: men aged 45–64; Survey 3) or a high-sensitivity latex enhanced nephelometric assay on a BN II analyzer (Dade Behring, Marburg, Germany; Survey 1: men aged 35–44 and all women; Survey 2). Both methods gave similar results when the same samples were analyzed [29]. The respective intra- and inter-assay coefficients of variation for quality control test sera were 4.0% and 12.0% for the IRMA assay, and 2.5% and 5.1% for the nephelometric assay.

*Assessments of risk factors and medications*

*In all studies*, body mass index (BMI) was calculated with the formula BMI = weight (kg) / height (m) $^2$.

*In all studies*, subjects were asked to fill out questionnaires that included questions concerning the use of medications (including lipid-lowering medication [statins]).

**3. Cardiovascular endpoint determinations and definitions of different cohorts**

*Cardiovascular endpoints studied in the Ludwigshafen Risk and Cardiovascular Health*

(*LURIC) study.*

Coronary artery disease (CAD) was assessed by angiography using the maximum luminal narrowing estimated by visual analysis [27]. Clinically relevant CAD was defined as the occurrence of at least one 50% stenosis in at least 1 of 15 coronary segments.

*Cardiovascular endpoints studied in the Finnish Cardiovascular Study (FINCAVAS)*

The presence of coronary artery disease (CAD) in the patients prior to the exercise test was determined based on patient interviews, resting ECG recordings, and existing hospital records [21]. Of the patients, 43.6% had undergone coronary angiography that unambiguously revealed the status of the coronary arteries. The data on the presence of hypertension, valvular conditions, cardiomyopathies, other heart diseases, and diabetes (types 1 and 2) were based on interviews and hospital records.

*The Angiography and Genes Study (ANGES)*

A clinical diagnosis of MI was based on symptoms, electrocardiographic findings, and biochemical marker tests measuring troponin I and creatinekinase. Information concerning previous cardiovascular diseases, surgical procedures, and MIs was collected from patient records at Tampere University Hospital. Coronary angiography was performed using the standard Judkins technique. Transluminal narrowing of at least 50% in any major coronary artery (left anterior descending, left circumflex, or right coronary artery) was the criterion for the diagnosis of coronary artery disease. [30]

**4. Genotyping and quality control (QC) in the different cohorts**

In YFS, genomic DNA was extracted from peripheral blood leukocytes using a commercially

available kit and the Qiagen BioRobot M48 Workstation according to the manufacturer's instructions (Qiagen, Hilden, Germany). Genotyping was performed on 2,556 samples using a custom-built Illumina Human 670k BeadChip at the Welcome Trust Sanger Institute. Genotypes were called using the Illuminus clustering algorithm. Fifty-six samples failed to meet the Sanger genotyping pipeline QC criteria (i.e. duplicated samples, heterozygosity, low call rate, or Sequenom fingerprint discrepancy). Out of the remaining 2,500 samples, one failed the gender check, three were removed due to low genotyping call rate (< 0.95), and 54 were excluded for possible relatedness (pi-hat > 0.2). Based on the Hardy–Weinberg equilibrium (HWE) test, 11,766 SNPs were excluded ($p \leq 10^{-6}$), and 7,746 SNPs failed the missingness test (call rate < 0.95 ) and another 34,596 failed the frequency test (MAF < 0.01). After quality control, 2,442 samples and 546,677 genotyped SNPs were available for further analysis. [31]

Genotype imputation was performed using MACH[32, 33] 1.0 and HapMap II CEU (release 22, NCBI build 36, dbSNP 126) samples as a reference. Palindromic A/T and C/G SNPs were removed before imputation. After imputation, 2,543,887 SNPs were available. SNPs with a squared correlation ($r^2$) of $\geq 0.30$ between imputed and true genotypes were considered well imputed.

For the DNA extraction in FINCAVAS and ANGES, 9.0 ml EDTA whole blood was drawn from the participants and stored at -20 °C. Genomic DNA was extracted from peripheral blood leukocytes by using the QIAamp DNA Blood Midi Kit and automated biorobot M48 extraction (QIAGEN GmbH, Hilden, Germany). Genotyping was completed for 2,824 samples using the Illumina Cardio-Metabo Chip (Illumina Inc., San Diego, CA, USA) at the Helmholtz Zentrum, München, Germany. The chips were scanned with the Illumina iScan system and genotypes called with Illumina GenomeStudio software. The following quality control filters were applied: GenCall score < 0.2, sample and SNP call rate

< 0.95, HWE p value < 10-6, MAF < 0.01, cryptic relatedness (pi-hat > 0.2), and gender check. After quality control, 2,620 samples and 120,721 SNPs were available. Both genotype and clinical data were available for 2,390 samples in FINCAVAS and for 808 in ANGES.

In LURIC, genomic DNA was prepared from EDTA anticoagulated peripheral blood by using a common salting-out procedure. Genotyping was accomplished for 2,966 samples using a custom-built Illumina 200k BeadChip (Cardio-Metabo Chip) at the Institute of Human Genetics at the Department of Genomics, Life & Brain Center, University of Bonn, Germany. Forty-seven samples failed to meet the QC criteria (i.e. duplicated samples, possible relatedness, low call rate [< 0.95], or gender discrepancy). Out of the SNPs, 5,384 failed to meet the QC criteria (Hardy–Weinberg equilibrium test [p ≤ 10–6], call rate < 0.95 ). After quality control, 2,919 samples and 191,341 genotyped SNPs were available for further analysis. Both genotype and clinical data were available for 2,912 samples in this study.

**5. Statistical analyses in more detail**

For the GWAS analysis, oxLDL was Box-Cox transformed. Residuals were obtained using a linear regression model in which the variables were adjusted for sex, age, and BMI, as well as principal components (to control population stratification [34]) and apoB. The GWAS was adjusted for apoB to identify SNPs affecting the oxidation process only (each LDL particle has one apoB molecule and the measured oxLDL strongly correlates with apoB). A GWAS was also performed on oxLDL without adjusting for apoB as well as by adjusting for LDL concentrations. Residuals were standardized (mean 0, s.d. 1) and their distributions confirmed to be very close to normal by means of visual Q-Q plot analysis. We also verified that the estimates for the beta coefficients from the GWAS were not driven by a few outliers by plotting leverage versus standardized residuals plots for the residuals.

Tests for additive genetic effects were carried out on a linear scale by means of linear

regression. Genotypes were coded as 0, 1, or 2 when the SNP was genotyped and by dosage (scale 0–2) when imputed. Tests were performed to assess the association of SNPs with the standardized residuals using PLINK[35] for the genotyped data. ProbABEL [36] was employed to fit the linear regression model, taking into account the genotype uncertainty in imputed SNPs. P values were combined from the analysis by favoring genotyped SNPs over imputed ones. Q-Q and Manhattan plots were drawn for the analysis of the results. The p value for genome-wide significance was set at $p < 5 \times 10^{-8}$, corresponding to a target $\alpha$ of 0.05 with a Bonferroni correction for one million independent tests.

The severity (functionality) of mutations was assessed by PolyPhen-2 version 2.1.0 software [37].

Further statistical analyses were performed using the R Statistical package v. 2.11.1 (http://www.r-project.org). In order to define associations non-redundantly associated with oxLDL, we applied a forward selection algorithm (as described in [38]). In brief, all the top SNPs with a p value below $5x10^{-8}$ and the covariates were inserted in the same linear model, and a stepwise model selection (Akaike Information Criterion, AIC) algorithm in the R package Modern Applied Statistics with S (MASS) was used with the Bayesian IC (BIC) criterion to leave only the individually associated SNPs and covariates in the model. Linkage disequilibrium (LD) was also analyzed visually with the Haploview software with an $r^2$ threshold of 0.8 using HapMap (phase II, release 22 CEU) haplotypes [39]. Moreover, the possible haplotypic effect of the associated SNPs was studied by using the haplo.stats package in R. To assess the proportion of oxLDL explained by the top SNP, $r^2$ was calculated twice—first by using a linear regression model explaining oxLDL with the SNP and all covariates, and secondly only with the covariates. The remainder of these two was considered as $r^2$ for the SNP.

We associated the SNPs with genome-wide significance (top SNPs) with

cardiovascular-disease-related endpoints (angioraphically verified CAD, three vessel CAD, and MI) in FINCAVAS, ANGES, and LURIC. The associations were assessed using the appropriate statistical models (chi-squared test, analysis of variance [ANOVA], linear regression, or Cox Proportional-Hazards regression) in R. Meta-analyses were performed using a fixed effects model when the p for cohort heterogeneity was higher than 0.05. KORA was only used for the replication of the oxLDL association. The YFS participants were young (< 39 yrs, average age 31.7 yrs in 2001), still without major clinical endpoints, and it was therefore not possible to include them in these analyses. P values below 0.05 were considered significant.

## B. Supplementary References


1. Packard RR, Libby P. Inflammation in atherosclerosis: from vascular biology to biomarker discovery and risk prediction. *Clin Chem.* 2008; 54: 24-38.

2. Ross R. Atherosclerosis is an inflammatory disease. *Am Heart J.* 1999; 138: S419-20.

3. Lusis AJ, Mar R, Pajukanta P. Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet.* 2004; 5: 189-218.

4. Steinberg D. Low density lipoprotein oxidation and its pathobiological significance. *J Biol Chem.* 1997; 272: 20963-20966.

5. Berliner JA, Navab M, Fogelman AM, Frank JS, Demer LL, Edwards PA, et al. Atherosclerosis: basic mechanisms. Oxidation, inflammation, and genetics. *Circulation.* 1995; 91: 2488-2496.

6. Steinberg D. Lewis A. Conner Memorial Lecture. Oxidative modification of LDL and atherogenesis. *Circulation.* 1997; 95: 1062-1071.

7. Heinecke JW. Oxidants and antioxidants in the pathogenesis of atherosclerosis: implications for the oxidized low density lipoprotein hypothesis. *Atherosclerosis.* 1998; 141: 1-15.

8. Witztum JL, Horkko S. The role of oxidized LDL in atherogenesis: immunological response and anti-phospholipid antibodies. *Ann N Y Acad Sci.* 1997; 811: 88-96; discussion 96-9.

9. Ylä-Herttuala S. Is oxidized low-density lipoprotein present in vivo? *Curr Opin Lipidol.* 1998; 9: 337-344.

10. Brown MS, Goldstein JL. Lipoprotein metabolism in the macrophage: implications for cholesterol deposition in atherosclerosis. *Annu Rev Biochem.* 1983; 52: 223-261.

11. Chisolm GM,3rd, Hazen SL, Fox PL, Cathcart MK. The oxidation of lipoproteins by monocytes-macrophages. Biochemical and biological mechanisms. *J Biol Chem.* 1999; 274: 25959-25962.

12. Conde L, Bevan S, Sitzer M, Klopp N, Illig T, Thiery J, et al. Novel associations for coronary artery disease derived from genome wide association studies are not associated with increased carotid intima-media thickness, suggesting they do not act via early atherosclerosis or vessel remodeling. *Atherosclerosis.* 2011; .

13. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel

pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011; 478: 103-109.

14. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature.* 2010; 467: 832-838.

15. 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; 43: 940-947.

16. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet.* 2011; 43: 1005-1011.

17. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature.* 2011; 477: 54-60.

18. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010; 42: 937-948.

19. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet.* 2011; 43: 1164-1164a.

20. Raitakari OT, Juonala M, Rönnemaa T, Keltikangas-Järvinen L, Räsänen L, Pietikäinen M, et al. Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol.* 2008; 37: 1220-1226.

21. Nieminen T, Lehtinen R, Viik J, Lehtimäki T, Niemelä K, Nikus K, et al. The Finnish Cardiovascular Study (FINCAVAS): characterising patients with high risk of cardiovascular morbidity and mortality. *BMC Cardiovasc Disord.* 2006; 6: 9.

22. Mennander A, Kuukasjarvi P, Laurikka J, Nikus K, Karhunen PJ, Tarkka M, et al. Diagnostic performance of plasma high sensitive C-reactive protein in detecting three-vessel coronary artery disease: modification by apolipoprotein E genotype. *Scand J Clin Lab Invest.* 2008; 68: 714-719.

23. Holvoet P, Stassen JM, Van Cleemput J, Collen D, Vanhaecke J. Oxidized low density lipoproteins in patients with transplant-associated coronary artery disease. *Arterioscler Thromb Vasc Biol.* 1998; 18: 100-107.

24. Paniagua JA, Lopez-Miranda J, Perez-Martinez P, Marin C, Vida JM, Fuentes F, et al. Oxidized-LDL levels are changed during short-term serum glucose variations and lowered with statin treatment in early Type 2 diabetes: a study of endothelial function and microalbuminuria. *Diabet Med.* 2005; 22: 1647-1656.

25. Ahotupa M, Marniemi J, Lehtimäki T, Talvinen K, Raitakari OT, Vasankari T, et al. Baseline diene conjugation in LDL lipids as a direct measure of in vivo LDL oxidation. *Clin Biochem.* 1998; 31: 257-261.

26. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem.* 1972; 18: 499-502.

27. Winkelmann BR, März W, Boehm BO, Zotz R, Hager J, Hellstern P, et al. Rationale and design of the LURIC study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics.* 2001; 2: S1-73.

28. Juonala M, Viikari JS, Kähönen M, Solakivi T, Helenius H, Jula A, et al. Childhood levels of serum apolipoproteins B and A-I predict carotid intima-media thickness and brachial endothelial function in adulthood: the cardiovascular risk in young Finns study. *J Am Coll Cardiol.* 2008; 52: 293-299.

29. Khuseyinova N, Imhof A, Trischler G, Rothenbacher D, Hutchinson WL, Pepys MB, et al. Determination of C-reactive protein: comparison of three high-sensitivity immunoassays. *Clin Chem.* 2003; 49: 1691-1695.

30. Raitoharju E, Seppälä I, Levula M, Kuukasjarvi P, Laurikka J, Nikus K, et al. Common variation in the ADAM8 gene affects serum sADAM8 concentrations and the risk of myocardial infarction in two independent cohorts. *Atherosclerosis.* 2011; .

31. Smith EN, Chen W, Kähönen M, Kettunen J, Lehtimäki T, Peltonen L, et al. Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet.* 2010; 6: e1001094.

32. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* 2010; 34: 816-834.

33. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet.* 2009; 10: 387-406.

34. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38: 904-909.

35. 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; 81: 559-575.

36. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics.* 2010; 11: 134.

37. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7: 248-249.

38. Pare G, Chasman DI, Parker AN, Zee RR, Malarstig A, Seedorf U, et al. Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. *Circ Cardiovasc Genet.* 2009; 2: 142-150.

39. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21: 263-265.

## C. Supplementary figure legends

**Figure S1. Q-Q plot of the genome-wide association study of oxidized low-density lipoprotein showing a clear deviation from normal distribution (lambda = 0.997).**

**Figure S2. Manhattan plot of the genome-wide association study of oxidized low-density lipoprotein (oxLDL) showing the association of multiple single-nucleotide polymorphisms with oxLDL in chromosome two.**

**Figure S3. Manhattan plot of the genome-wide association study of oxidized low-density lipoprotein adjusted for the top single-nucleotide polymorphism (rs676210) showing no further independent associations, with a $p < 5 \times 10^{-8}$.**

**Figure S4. Meta-analyses of the association between apoB alleles (rs676210/Pro2739Leu) and the age of onset of cardiovascular outcomes in three independent clinical study cohorts.**

**Statistics:** *Hazard ratios (HR; blue squares, sizes indicate the relative weight of each study) and 95% confidence intervals (CI; horizontal lines) were estimated by a per-risk-allele (G) additive Cox Proportional-Hazards regression using age as time variable adjusted for sex, body mass index (BMI), and statin use. The fixed effect meta-analysis (p value for heterogeneity > 0.05 for all) estimates are shown as blue diamonds.

**Definitions:** a) coronary artery disease (CAD), with or without angiographically verified 50% stenosis; b) three vessel disease (3VD), with or without angiographically verified 50% stenosis in three vessels; c) myocardial infarction (MI), with or without MI

**Abbreviations:** N, number; FINCAVAS, The Finnish Cardiovascular Study; ANGES, The Angiography and Genes Study; LURIC, The Ludwigshafen Risk and Cardiovascular Health Study.

## D. Supplementary Tables

**Table S1.** General characteristics of the study populations.

| Study | Acronym | n* | Age, years† | % Male |
|---|---|---|---|---|
| The Cardiovascular Risk in Young Finns Study | **YFS** | 2080 | 31.7 (5.0) | 45.2% |
| The Ludwigshafen Risk and Cardiovascular Health Study | **LURIC** | 2912 | 62.6 (10.7) | 69.3% |
| Kooperative Gesundheitsforschung in der Region Augsburg (Cooperative Health Research in the Region of Augsburg) Study | **KORA** | 1326 | 52.7 (10.6) | 53.1% |
| The Finnish Cardiovascular Study | **FINCAVAS** | 1118 | 59.9 (10.4) | 71.1% |
| The Angiography and Genes Study | **ANGES** | 808 | 62.9 (10.0) | 63.9% |

**Abbreviations:** n, number. *n for subjects having data required for the present study; †values expressed as means (standard deviations).

**Table S2.** Clinical and biochemical trait profiles of oxidized LDL study cohorts according to rs676210 (Pro2739Leu) genotype.

| | ApoB rs676210 (Pro2739Leu) genotype | | | |
|---|---|---|---|---|
| Cohort / trait | GG | GA | AA | p value* |
| **YFS** | (n = 1125) | (n = 806) | (n = 149) | |
| OxLDL (U/l) | 91.3 (24.6) | 76.2 (22) | 61.7 (20) | 4.3E-136 |
| OxLDL / LDL (U/l/mmol/l) | 28 (5.22) | 23.7 (5.12) | 19.8 (5.98) | 2E-113 |
| OxLDL / apoB (U/l/g/l) | 85.7 (13.8) | 73.4 (13.9) | 60.9 (17.1) | 4.9E-124 |
| Male sex (yes) | 516 (45.9%) | 357 (44.3%) | 67 (45%) | 0.789 |
| Hypertension (yes) | 30 (2.67%) | 20 (2.48%) | 5 (3.36%) | 0.828 |
| Type I DM (yes) | 7 (0.622%) | 3 (0.372%) | 2 (1.34%) | 0.341 |
| Type II DM (yes) | 1 (0.0889%) | 0 (0%) | 0 (0%) | 0.654 |
| Current smoking (yes) | 269 (23.9%) | 182 (22.6%) | 36 (24.2%) | 0.771 |
| Dyslipidemia (yes) | 782 (69.5%) | 539 (66.9%) | 100 (67.1%) | 0.446 |
| Statin use (yes) | 5 (0.444%) | 2 (0.248%) | 0 (0%) | 0.582 |
| Age (years) | 31.7 (5.04) | 31.8 (4.9) | 31.7 (5.09) | 0.694 |
| BMI (kg/m$^2$) | 25.2 (4.5) | 25 (4.45) | 25.3 (3.97) | 0.636 |
| C-reactive protein (mg/l) | 2.02 (4.33) | 1.95 (3.74) | 1.32 (1.53) | 0.116 |
| Triglycerides (mmol/l) | 1.37 (0.884) | 1.27 (0.747) | 1.33 (1.01) | 0.0483 |
| Total cholesterol (mmol/l) | 5.17 (0.979) | 5.13 (0.964) | 5.03 (0.919) | 0.109 |
| LDL cholesterol (mmol/l) | 3.28 (0.827) | 3.25 (0.844) | 3.15 (0.828) | 0.0801 |
| VLDL cholesterol | NA | NA | NA | - |
| HDL cholesterol (mmol/l) | 1.27 (0.308) | 1.31 (0.321) | 1.32 (0.337) | 0.00605 |
| Total / HDL cholesterol | 4.3 (1.37) | 4.16 (1.39) | 4.05 (1.25) | 0.00671 |
| ApoB (g/l) | 1.07 (0.266) | 1.04 (0.261) | 1.02 (0.261) | 0.00501 |
| | | | | |
| **LURIC** | (n = 1761) | (n = 986) | (n = 165) | |
| OxLDL (U/l) | 80.1 (24.5) | 68.6 (26.2) | 58.9 (52.6) | 2.5E-47 |
| OxLDL / LDL (U/l/mg/dl) | 0.74 (0.312) | 0.614 (0.263) | 0.507 (0.393) | 9.7E-36 |
| OxLDL / apoB (U/l/mg/dl) | 0.768 (0.201) | 0.659 (0.223) | 0.554 (0.414) | 4.3E-49 |
| OxLDL / LDL-apoB (U/l / mg/dl) | 0.981 (0.357) | 0.825 (0.309) | 0.69 (0.53) | 7.2E-38 |
| Male sex (yes) | 1224 (69.5%) | 676 (68.6%) | 119 (72.1%) | 0.636 |
| Hypertension (yes) | 1285 (73%) | 706 (71.6%) | 122 (73.9%) | 0.684 |
| Type I DM (yes) | 4 (0.227%) | 2 (0.203%) | 2 (1.21%) | 0.0601 |
| Type II DM (yes) | 305 (17.3%) | 175 (17.7%) | 26 (15.8%) | 0.819 |
| Current smoking (yes) | 1146 (65.1%) | 608 (61.7%) | 116 (70.3%) | 0.0492 |
| Dyslipidemia (yes) | 1278 (72.6%) | 646 (65.5%) | 103 (62.4%) | 0.0000698 |
| Statin use (yes) | 819 (46.5%) | 447 (45.3%) | 80 (48.5%) | 0.701 |
| Age (years) | 62.5 (10.8) | 62.9 (10.6) | 62.6 (10.8) | 0.447 |

| | | | | |
|---|---|---|---|---|
| BMI (kg/m$^2$) | 27.5 (4.1) | 27.3 (4.12) | 26.9 (3.44) | 0.0501 |
| C-reactive protein (mg/l) | 9.29 (19.2) | 8.48 (15.8) | 8.85 (15.6) | 0.346 |
| Triglycerides (mg/dl) | 178 (125) | 169 (115) | 154 (84) | 0.00487 |
| Total cholesterol (mg/dl) | 192 (38.6) | 192 (39.3) | 194 (35.8) | 0.933 |
| LDL cholesterol (mg/dl) | 115 (33.5) | 116 (34.9) | 119 (33.8) | 0.111 |
| VLDL cholesterol (mg/dl) | 38.8 (27.5) | 36.3 (25.4) | 34.3 (25.8) | 0.0034 |
| HDL cholesterol (mg/dl) | 38.2 (10.7) | 39.1 (10.5) | 40.6 (12.5) | 0.00168 |
| Total / HDL cholesterol | 5.37 (1.77) | 5.31 (3.53) | 5.12 (1.63) | 0.245 |
| apoB (mg/dl) | 105 (24.2) | 104 (25.2) | 104 (22.4) | 0.236 |
| LDL-apoB (mg/dl) | 84.5 (21.9) | 84.7 (22.9) | 85.1 (21.3) | 0.697 |
| **KORA** | (n = 803) | (n = 462) | (n = 61) | |
| OxLDL (U/l) | 96.8 (26.7) | 88.0 (24.2) | 79.9 (22.7) | 1.05E-11 |
| OxLDL / LDL (U/l / mg/dl) | 0.674 (0.167) | 0.627 (0.154) | 0.583 (0.126) | 0.000000235 |
| Male sex (yes) | 441 (54.9%) | 230 (49.8%) | 33 (54.1%) | 0.209 |
| Hypertension (yes) | 355 (44.2%) | 194 (42.0%) | 23 (37.7%) | 0.507 |
| DM (yes) | 48 (6.0%) | 28 (6.1%) | 2 (3.3%) | 0.675 |
| Dyslipidemia (yes) | 297 (40.0%) | 172 (37.2%) | 16 (26.2%) | 0.228 |
| Current smoking (yes) | 193 (24.0%) | 121 (26.2%) | 18 (29.5%) | 0.495 |
| Statin use (yes) | 4 (0.50%) | 2 (0.43%) | 0 (0%) | 0.853 |
| Age (years) | 52.9 (10.5) | 52.5 (10.6) | 50.9 (11.0) | 0.341 |
| BMI (kg/m$^2$) | 27.3 (4.2) | 27.1 (3.9) | 26.7 (4.1) | 0.513 |
| C-reactive protein (mg/l) | 3.20 ( 5.67) | 2.89 (5.95) | 3.03 (4.56) | 0.646 |
| Triglycerides (mg/dl) | 180.4 (127.0) | 167.8 (107.5) | 145.6 (83.2) | 0.334 |
| Total cholesterol (mg/dl) | 238.8 (46.4) | 237.1 (42.2) | 232.7 (40.7) | 0.524 |
| LDL cholesterol (mg/dl) | 150.0 (43.9) | 147.7 (41.7) | 142.8 (40.2) | 0.455 |
| HDL cholesterol (mg/dl) | 55.1 (16.5) | 57.7 (17.3) | 55.5 (15.0) | 0.03 |
| Total / HDL cholesterol | 4.73 (1.84) | 4.50 (1.86) | 4.52 (1.59) | 0.079 |

**Abbreviations:** YFS, The Cardiovascular Risk in Young Finns Study; LURIC, The Ludwigshafen Risk and Cardiovascular Health Study; KORA, Kooperative Gesundheitsforschung in der Region Augsburg Study; OxLDL, oxidized low-density lipoprotein (LDL); DM, Diabetes Mellitus; apoB, apolipoprotein-B; BMI, body mass index; HDL, high-density lipoprotein; VLDL, very-low-density lipoprotein; NA, not available.

**Statistics:** Values are numbers (percentages) in cases of categorical data and means (standard deviations) in cases of continuous data; * p values (difference between rs676210 genotype groups) calculated with chi-square test for categorical data, with a linear regression model for

oxLDL, oxLDL/LDL, and oxLDL/apoB, and with analysis of variance (ANOVA) for other continuous data. The OxLDL and oxLDL/LDL models have been adjusted for age, sex, BMI, HDL, APOB, triglycerides, and current smoking in YFS and LURIC. The OxLDL and oxLDL/LDL models have been adjusted for age, sex, survey number, and BMI in KORA. The OxLDL/apoB models have been adjusted for age, sex, BMI, HDL, and triglycerides.

**Conversion factors**: total cholesterol, LDL cholesterol, and HDL cholesterol 1 mg/dl = 0.0259 mmol/l; triglycerides: 1 mg/dl = 0.0113 mmol/l; apoB: 1mg/dl = 0.01 g/l.

**Figure S1**

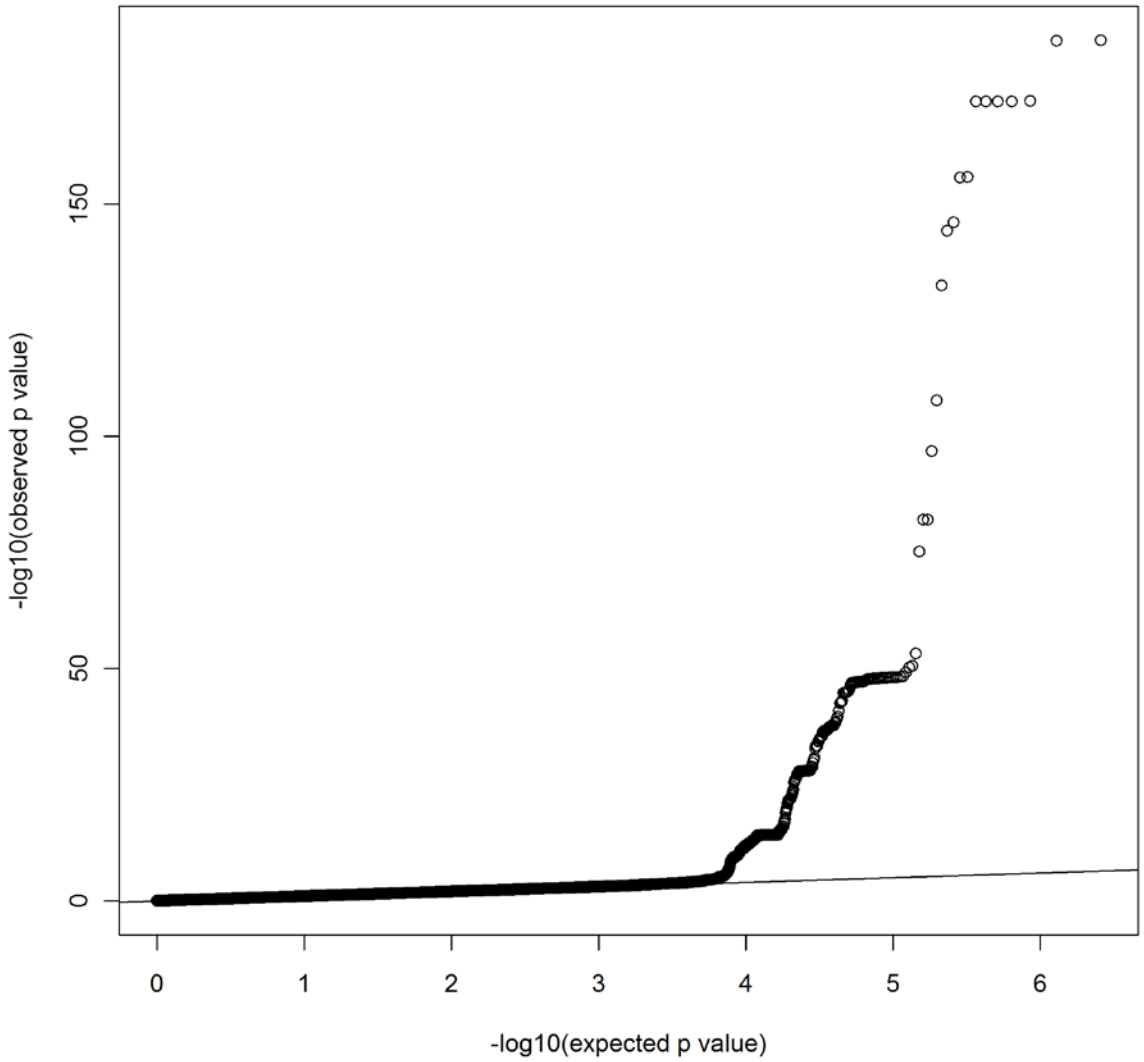


**Figure S2**

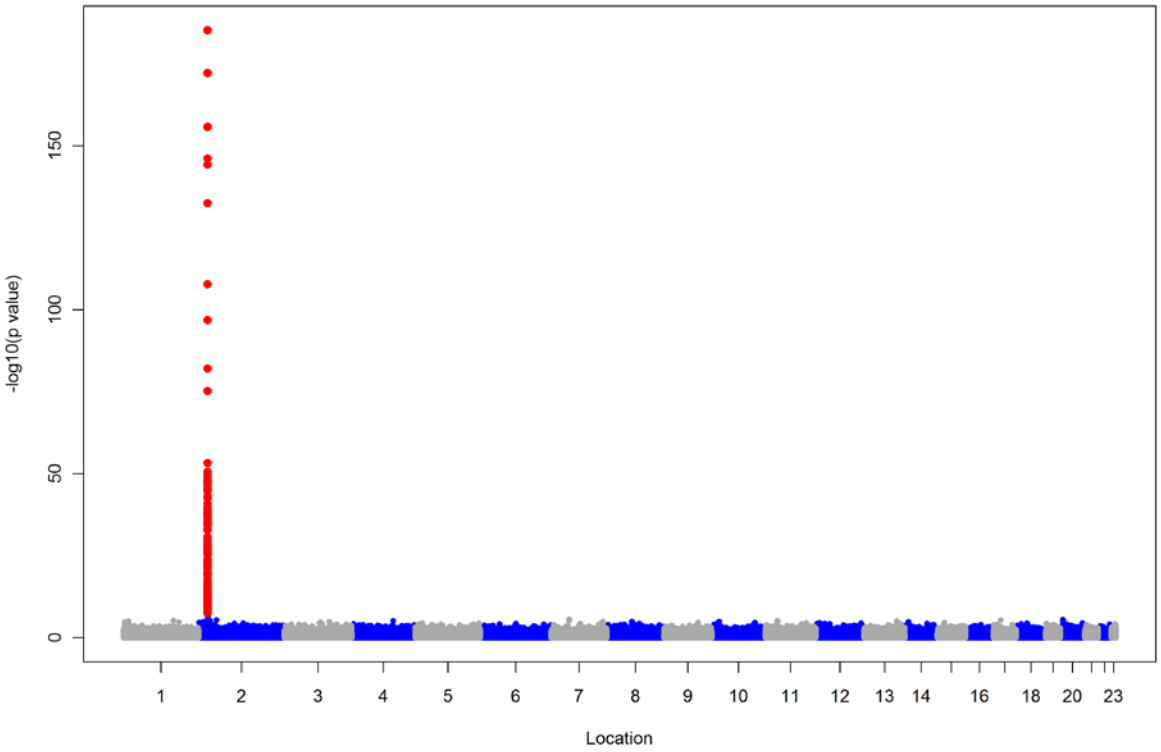


**Figure S3**

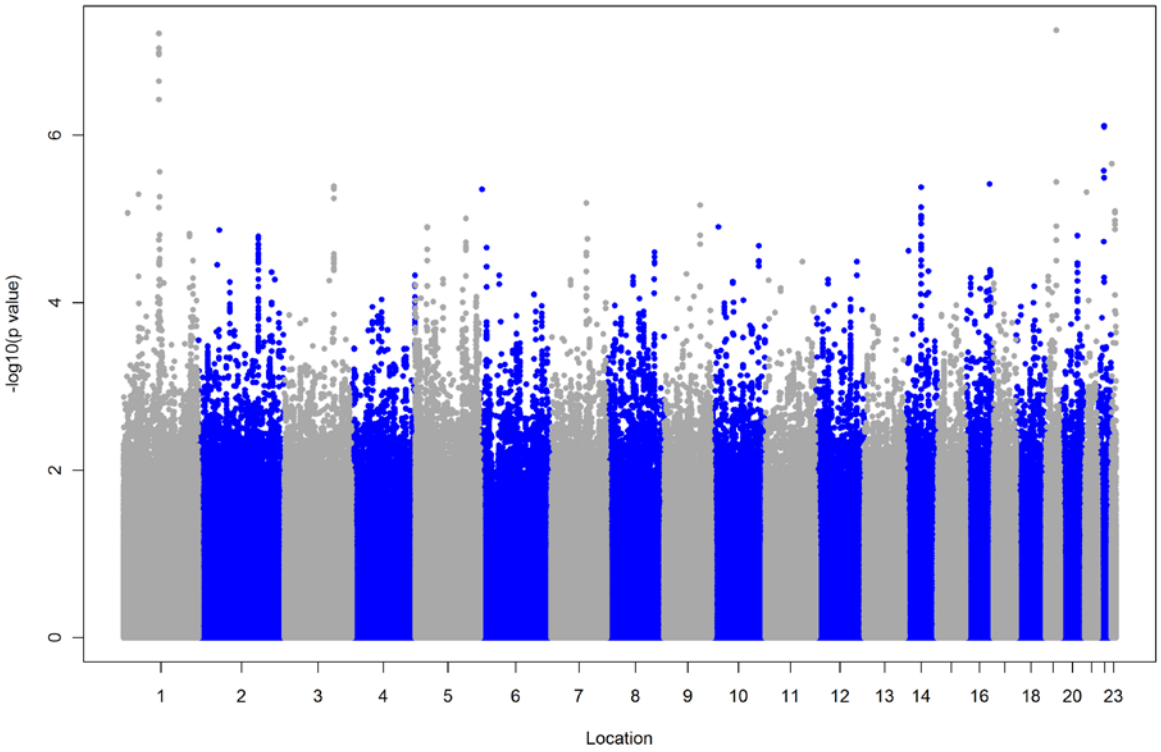
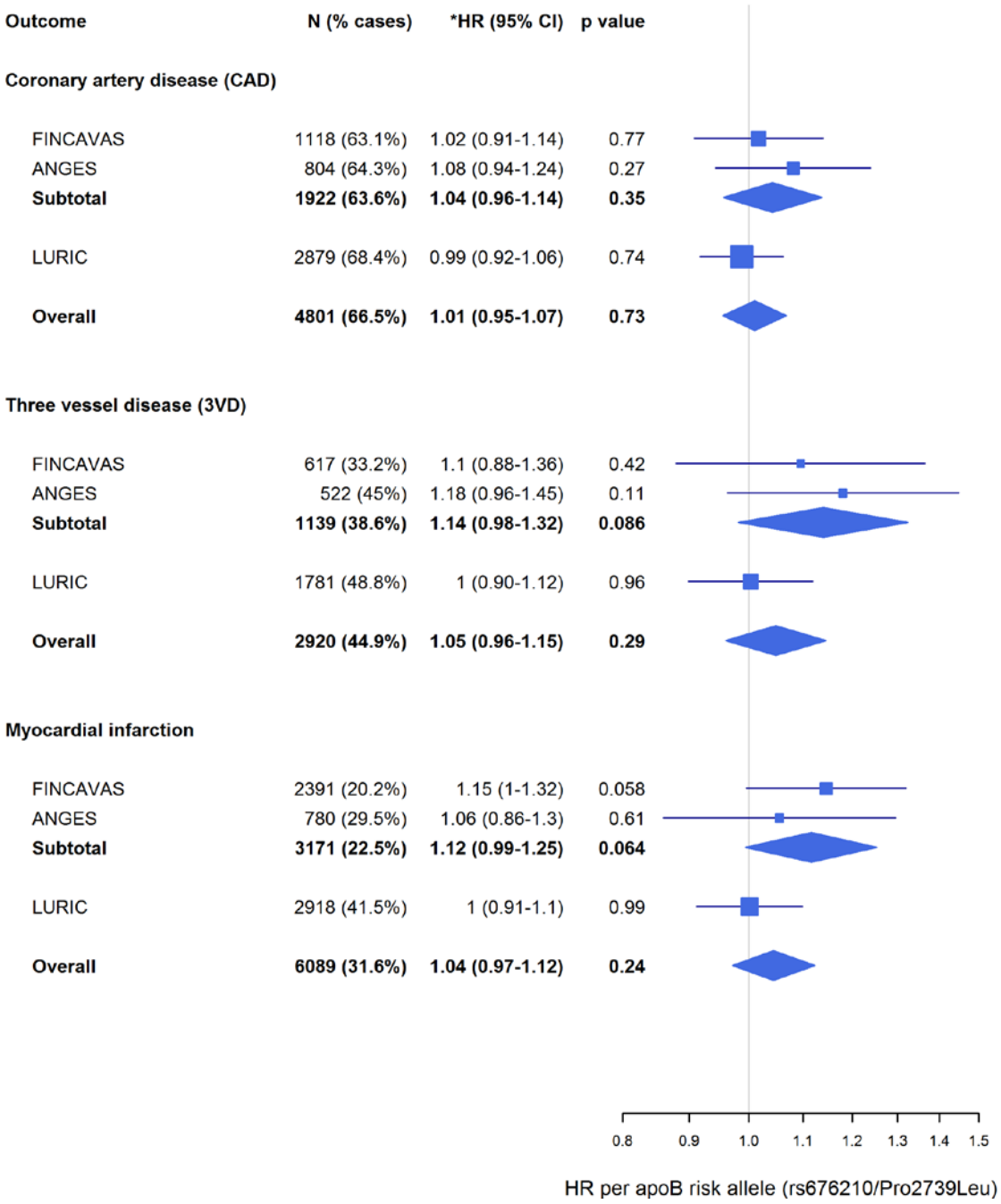


**Figure S4**



Atherosclerosis 234 (2014) 214–217

Contents lists available at ScienceDirect

Atherosclerosis

journal homepage: www.elsevier.com/locate/atherosclerosis

# Association of the novel single-nucleotide polymorphism which increases oxidized low-density lipoprotein levels with cerebrovascular disease events

Kari-Matti Mäkelä [a,b,*], Matthew Traylor [c], Niku Oksala [a,b,d], Marcus E. Kleber [e], Ilkka Seppälä [a,b], Leo-Pekka Lyytikäinen [a,b], Jussi A. Hernesniemi [a,b,f], Mika Kähönen [g], Steve Bevan [m], Peter M. Rothwell [h], Cathie Sudlow [i], Martin Dichgans [j,k], Wellcome Trust Case Control Consortium 2 (WTCCC2)[1], Graciela Delgado [e], Tanja B. Grammer [e], Hubert Scharnagl [l], Hugh S. Markus [m], Winfried März [n,e,l,2], Terho Lehtimäki [a,b,2]

[a] School of Medicine, University of Tampere, Tampere, Finland
[b] Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
[c] Stroke and Dementia Research Centre, St George's University of London, London, United Kingdom
[d] Division of Vascular Surgery, Department of Surgery, Tampere University Hospital, Tampere, Finland
[e] Medical Clinic V, Mannheim Medical Faculty, University of Heidelberg, Germany
[f] Department of Internal Medicine, North Karelia Central Hospital, Joensuu, Finland
[g] Department of Clinical Physiology, Tampere University Hospital and University of Tampere, Tampere, Finland
[h] Stroke Prevention Research Unit, Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, UK
[i] Division of Clinical Neurosciences and Insititute of Genetics and Molecular Medicine, University of Edinburgh, UK
[j] Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany
[k] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
[l] Clinical Institute of Medical and Chemical Laboratory Diagnostics Medical University of Graz, Graz, Austria
[m] Clinical Neurosciences, University of Cambridge, Cambridge, UK
[n] Synlab Services GmbH, Mannheim, Germany

ARTICLE INFO

*Article history:*
Received 3 December 2013
Received in revised form
6 February 2014
Accepted 3 March 2014
Available online 14 March 2014

*Keywords:*
Oxidized low-density lipoprotein
Stroke
Transient ischaemic attack
Genetics
Single nucleotide polymorphism

ABSTRACT

*Background and purpose:* Patients with genetic background for high circulating oxidized low-density lipoprotein (oxLDL) levels might be at an increased risk of cerebrovascular disease (CVD).
*Methods:* The association of oxLDL-variant rs676210 with CVD events was studied in patients undergoing coronary angiography (study A; $N = 2913$ [271 cases]). We sought to replicate the results in a large genome-wide association study meta-analysis of ischaemic stroke (study B; $N = 3548$ cases, 5972 controls).
*Results:* In study A, the prevalence of hypertension, diabetes and >50% carotid stenosis as well as the levels of LDL cholesterol differed significantly between cases and controls. In a logistic regression model adjusted for the significant covariates, rs676210 associated with CVD events ($p = 0.030$; odds ratio = 1.29 [95% confidence interval 1.03–1.63] for risk allele G). In study B, rs676210 did not associate with the history of ischaemic stroke.
*Conclusions:* The oxLDL levels increasing variant rs676210 associates with CVD events in patients undergoing coronary angiography.



## 1. Introduction

In our earlier genome-wide association study (GWAS) on oxidized low-density lipoprotein (oxLDL), we found an apolipoprotein-B (apoB) Pro2739Leu missense mutation (rs676210) associating with oxLDL but not with history of coronary artery disease (CAD) or myocardial infarction (MI) [1]. Since there is some

* Corresponding author. Department of Clinical Chemistry, Finn-Medi 2, P.O. Box 66, FI-33101 Tampere, Finland. Tel.: +358 3 3117 4051; fax: +358 3 3117 4168.
*E-mail address:* kari-matti.makela@uta.fi (K.-M. Mäkelä).
[1] WTCCC2 membership is listed in the Supplementary material.
[2] Equally contributed.

http://dx.doi.org/10.1016/j.atherosclerosis.2014.03.002
0021-9150/

evidence that oxLDL associates with ischaemic stroke [2–4], rs676210 may also associate with ischaemic stroke.

Ischaemic stroke can be classified into five subtypes according to Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria [5]; 1) large-artery atherosclerosis (LAA), 2) small-vessel disease (SVD), 3) cardioembolic stroke (CE), 4) other aetiology, or 5) unknown aetiology. One would expect oxLDL to associate especially with LAA since oxLDL is considered to be essential in the atherosclerotic process [6]. There is also some evidence that oxLDL could bind to thrombocytes and induce platelet adhesion to the vascular wall in acute coronary syndromes [7], which could also increase the risk of LAA subtype of stroke. However, only few studies indicate an association of circulating oxLDL with LAA [4], and, furthermore, we did not find an association between rs676210 and the history of CAD in the previous study [1].

Since rs676210 associates strongly with oxLDL in both the young healthy population and elderly angiography patients [1], it could be used to substitute the effect of the lifetime risk of increased oxLDL levels on ischaemic stroke. Therefore, we first studied the association of rs676210 with the history of cerebrovascular disease (CVD) events (transient ischaemic attack [TIA], or stroke) in patients undergoing coronary angiography (the Ludwigshafen Risk and Cardiovascular Health [LURIC] study [8]; $N = 2918$ [271 cases]). In the second phase, we sought to replicate the results in a large GWAS meta-analysis of history of ischaemic stroke and its pathophysiological subtypes (Wellcome Trust Case Control Consortium 2 [WTCCC2] ischaemic stroke GWAS [9]; total $N = 3548$ cases, 5972.controls).

## 2. Materials and methods

For all cohorts, the recruitment of patients was approved by the relevant local ethics committees, and studies were conducted in accordance with the Declaration of Helsinki.

*The Ludwigshafen Risk and Cardiovascular Health (LURIC) study* consists of 3316 Caucasian patients who were referred for coronary angiography because of chest pain at a tertiary care centre in Southwest Germany between 1997 and 2000 [8]. For the 2913 LURIC participants, all necessary covariate and endpoint data were available and they were included in the present study. Previous CVD events were defined as a documented history of a foregoing transient ischaemic attack (TIA), prolonged ischaemic deficit, or cerebral infarction with or without a remaining neurologic deficit [8].

*Discovery stroke cohorts in WTCCC2* ischaemic stroke GWAS included samples from the UK (a–c) and Germany (d), with a total of 3548 cases and 5972 controls [9]. Cases were phenotyped and classified into mutually exclusive aetiologic subtypes according to the TOAST classification [5]. The control data set for the British discovery samples was the WTCCC2 common control set, which includes healthy blood donors from the United Kingdom Blood Service's (UKBS) collection and individuals from the 1958 Birth Cohort dataset (58C). The control data set for the German cases was taken from the MONICA/KORA Augsburg Study's population-based controls from the same region in Germany.

### 2.1. Genotyping and quality control (QC) in the different cohorts

In LURIC, Metabochip and in WTCCC2 Illumina chips were used for genotyping the studied single-nucleotide polymorphism (SNP), rs676210. The SNP was directly genotyped in all studies and met all quality control criteria.

Statistics for the LURIC study were performed using logistic regression in the R Statistical package v. 2.15.2 (http://www.r-project.org). In the WTCCC2, analysis was performed with logistic regression using PLINK [10] on the separate groups; meta-analysis using an inverse-variance-weighted approach was performed using METAL [11].

## 3. Results

The general characteristics of the LURIC cohort and the difference between CVD event cases and controls are displayed in Supplementary Table 1. More than 60% of the LURIC population was diagnosed with CAD, and CAD and MI were more prevalent among CVD event cases. Out of the known ischaemic stroke risk factors, hypertension, diabetes, carotid stenosis, atrial fibrillation and LDL cholesterol were associated with history of CVD events. OxLDL did not associate with CVD events in LURIC ($p = 0.955$).

In the logistic regression model with no covariates, rs676210 associated with CVD events ($p = 0.030$, odds ratio [OR] = 1.28 [95% confidence interval, CI = 1.03–1.60] for risk allele G). In the logistic regression model adjusted for the significant risk factors (hypertension, diabetes, carotid stenosis, atrial fibrillation and LDL cholesterol), rs676210 remained significantly associated with CVD events ($p = 0.030$, OR = 1.29 [1.03–1.63] for risk allele G). The studied LURIC variables according to rs676210 genotype are displayed in Table 1. In addition to oxLDL and CVD events, only BMI and smoking were borderline-significantly associated with rs676210.

We attempted to replicate the association of rs676210 with CVD events in LURIC in the meta-analysis of WTCCC2 ischaemic stroke cohorts. The general characteristics as well as number of cases and controls in WTCCC2 are displayed in Supplementary Table 2. Logistic regression without adjustments was used as in LURIC. There was no significant association of rs676210 with all types of ischaemic stroke ($p = 0.81$, OR = 1.00 [0.93–1.09]), LAA ($p = 0.85$, OR = 0.99 [0.87–1.12]), CE ($p = 0.66$, OR = 1.03 [0.90–1.18]), or SVD ($p = 0.65$, OR = 0.97 [0.83–1.13]).

**Table 1**
The association of the apolipoprotein-B Pro2739Leu missense mutation rs676210 with cerebrovascular disease (CVD) events and its risk factors in LURIC.

| Variables | rs676210 genotype | | | *p*-value |
|---|---|---|---|---|
| | AA | AG | GG | |
| | ($N = 165$) | ($N = 987$) | ($N = 1761$) | |
| CVD event (yes) | 11 (6.67%) | 80 (8.11%) | 180 (10.2%) | 0.030 |
| oxLDL (U/l) | 58.9 (52.6) | 68.7 (26.2) | 80.1 (24.5) | <<0.001 |
| Sex (female) | 46 (27.9%) | 310 (31.4%) | 537 (30.5%) | 0.643 |
| Hypertension (yes) | 95 (57.6%) | 562 (56.9%) | 1049 (59.6%) | 0.392 |
| Diabetes (yes) | 52 (31.5%) | 327 (33.1%) | 562 (31.9%) | 0.787 |
| Carotid stenosis (yes) | 12 (7.27%) | 42 (4.26%) | 80 (4.54%) | 0.229 |
| Atrial fibrillation (yes) | 17 (10.3%) | 127 (12.9%) | 218 (12.4%) | 0.633 |
| CAD (yes) | 119 (72.1%) | 662 (67.1%) | 1183 (67.2%) | 0.469 |
| MI (yes) | 62 (37.6%) | 423 (42.9%) | 722 (41%) | 0.373 |
| Smoking (yes, ever) | 116 (70.3%) | 607 (61.5%) | 1133 (64.3%) | 0.0643 |
| BMI (kg/m$^2$) | 26.9 (3.44) | 27.3 (4.12) | 27.5 (4.1) | 0.0501 |
| LDL cholesterol (mg/dl) | 119 (33.8) | 116 (34.9) | 115 (33.5) | 0.111 |

**Abbreviations**: LURIC, Ludwigshafen Risk and Cardiovascular Health study; CAD, coronary artery disease (over 50% stenosis); MI, myocardial infarction; BMI, body-mass index; LDL, low-density lipoprotein; oxLDL, oxidized LDL.
**Statistics**: Values are numbers (percentages) in cases of categorical data and means (standard deviations) in cases of continuous data; *p* values (difference between rs676210 genotype groups) calculated with chi-square test for categorical data, with logistic regression for CVD events and with analysis of variance (ANOVA) for other continuous data. The CVD event model is adjusted for hypertension, diabetes, carotid stenosis, atrial fibrillation and low-density lipoprotein cholesterol.

## 4. Discussion

We found an association between the oxLDL-levels-increasing apolipoprotein-B missense mutation Pro2739Leu (rs676210) and history of CVD events in a cohort of patients referred for coronary angiography. The result was not, however, replicated in a large meta-analysis of ischaemic stroke cohorts.

The evident difference between the LURIC and WTCCC2 cohorts could explain the distinct association of rs676210 with CVD events in LURIC and ischaemic stroke in WTCCC2. The WTCCC2 cohorts were a collected as series of stroke patients enrolled to hospital, whereas LURIC is a series of patients referred for coronary angiography, most of whom were diagnosed with CAD. The expected association between oxLDL and ischaemic stroke would be for the LAA subtype of stroke due to the known involvement of oxLDL in the atherosclerotic process [6]. However, the absence of an association between rs676210 and CAD or MI [1] and the controversial evidence of circulating oxLDL associating with CAD [12] suggest that this would not explain the observed association of the SNP with CVD events. One explanation could be found in the evidence of oxLDL binding to thrombocytes and increasing their adhesion to vessel walls [7]. There is also evidence of higher serum oxLDL levels especially in CAD patients [13]. This higher oxidative stress burden in CAD patients could be the key to explaining the increase in the risk of at least the LAA subtype of stroke in the cases of CAD among the LURIC patients. Unfortunately, we found no association for this subtype in WTCCC2. The CVD event definition in LURIC includes TIAs which also could be one explanation for the distinct association between the studies, as increased oxLDL levels in LURIC could also increase the risk for TIA.

The predictive value of cardiovascular disease by oxLDL biomarkers is considered controversial [12] and, accordingly, there is no clinical use for them at present. A genetic variant such as rs676210 which has a high impact on oxLDL levels would be an interesting alternative for oxLDL biomarkers, as it describes the lifetime risk for high oxLDL levels in comparison to measurements of oxLDL at one point in time. Unfortunately, the evidence of circulating oxLDL associating with ischaemic stroke is scarce [2–4] and our current results are inconsistent. In our results, the cross-sectional circulating oxLDL levels did not associate with history of CVD events even though rs676210 did. Therefore, rs67210 could have predictive value for ischaemic stroke in a specific population of CAD risk patients. This needs to be verified in another similar sample.

### 4.1. Limitations and strengths of the study

The obvious limitation of this study is the lack of replication of the association of rs676210 with ischaemic stroke. To verify the association, it needs to be replicated in a sample similar to LURIC. Moreover, the definitions for CVD differ between LURIC and WTCCC2. The CVD event definition in LURIC includes all TIAs and strokes. The absence of replication of this association in a general ischaemic stroke setting means that there is not clear clinical use for the studied SNP per se in, for example, the prediction of stroke in the general population. However, if the result could be replicated in another cohort of angiography patients, there might be some predictive value for this SNP in patients with suspected CAD. We also do not have stroke subtypes available from LURIC to verify whether the association is with a specific subtype of ischaemic stroke. Furthermore, the effect was only studied in white European individuals and might not be generalisable to other ethnicities.

## 5. Conclusions

We have found evidence that the common apoB missense mutation Pro2739Leu (rs676210) that increases oxLDL levels associates with CVD events in patients with high probability of CAD undergoing coronary angiography.

## Sources of funding

This work received funding from the European Union 7th Framework Programme funding for the AtheroRemo project (201668), the Tampere University Hospital Medical Fund grant 9N035, the Emil Aaltonen Foundation, the Finnish Foundation of Cardiovascular Research, and the Tampere Tuberculosis Foundation (T.L.); furthermore, the work was funded with grants awarded to K-M.M. by the Kliinisen Kemian Tutkimussäätiö, Orion-Farmos Research Foundation, Yrjö Jahnsson, and Laboratoriolääketieteen edistämissäätiö foundations.

WTCCC2 was principally funded by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). The Edinburgh Stroke Study was supported by the Wellcome Trust (clinician scientist award to Dr. Sudlow) and the Binks Trust (research fellowship to Dr. Jackson). The collection of the Munich cases and data analysis were supported by the Vascular Dementia Research Foundation. The Oxford Vascular Study is funded by the Medical Research Council, the Stroke Association, the Dunhill Medical Trust, the National Institute of Health Research (NIHR) and the NIHR Biomedical Research Centre, Oxford. Segal, Burgess and Poole are funded by the NIHR Biomedical Research Centre, Oxford. The collection of some of the St George's stroke cohort was supported by project grant support from the Stroke Association.

LURIC received funding through the 6th Framework Programme (integrated project Bloodomics, grant LSHM-CT-2004-503485) and the 7th Framework Programme (integrated project AtheroRemo, grant agreement number 201668 and RiskyCAD, grant agreement number 305739) of the European Union.

Mercodia AB, Uppsala, provided the reagents for the determination of oxLDL free of charge, but did not assume any other role in the conducting of this study.

## Disclosures

None.

## Appendix A. Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.atherosclerosis.2014.03.002.

## References

[1] Mäkelä KM, Seppälä I, Hernesniemi JA, et al. Genome-wide association study pinpoints a new functional apolipoprotein B variant influencing oxidized low-density lipoprotein levels but not cardiovascular events: AtheroRemo consortium. Circ Cardiovasc Genet 2013;6:73–81.
[2] Uno M, Kitazato KT, Nishi K, Itabe H, Nagahiro S. Raised plasma oxidised LDL in acute cerebral infarction. J Neurol Neurosurg Psychiatry 2003;74:312–6.
[3] Serebruany V, Sani Y, Eisert C, et al. Effects of aggrenox and aspirin on plasma endothelial nitric oxide synthase and oxidised low-density lipoproteins in patients after ischaemic stroke. the AGgrenox versus aspirin therapy evaluation (AGATE) biomarker substudy. Thromb Haemost 2011;105:81–7.
[4] Vibo R, Korv J, Roose M, et al. Acute phase proteins and oxidised low-density lipoprotein in association with ischemic stroke subtype, severity and outcome. Free Radic Res 2007;41:282–7.
[5] Adams Jr HP, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. definitions for use in a multicenter clinical trial. TOAST. trial of org 10172 in acute stroke treatment. Stroke 1993;24:35–41.
[6] Ylä-Herttuala S. Is oxidized low-density lipoprotein present in vivo? Curr Opin Lipidol 1998;9:337–44.
[7] Stellos K, Sauter R, Fahrleitner M, et al. Binding of oxidized low-density lipoprotein on circulating platelets is increased in patients with acute coronary syndromes and induces platelet adhesion to vascular wall in vivo–brief report. Arterioscler Thromb Vasc Biol 2012;32:2017–20.

[8] Winkelmann BR, März W, Boehm BO, et al. Rationale and design of the LURIC study–a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. Pharmacogenomics 2001;2: S1–73.
[9] International Stroke Genetics Consortium (ISGC), Wellcome Trust Case Control Consortium 2 (WTCCC2), Bellenguez C, Bevan S, Gschwendtner A, et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. Nat Genet 2012;44: 328–33.
[10] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81: 559–75.
[11] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–1.
[12] Tsimikas S. Oxidized low-density lipoprotein biomarkers in atherosclerosis. Curr Atheroscler Rep 2006;8:55–61.
[13] Weinbrenner T, Cladellas M, Isabel Covas M, et al. High oxidative stress in patients with stable coronary heart disease. Atherosclerosis 2003;168:99–106.

# Evidence HDAC9 Genetic Variant Associated With Ischemic Stroke Increases Risk via Promoting Carotid Atherosclerosis

Hugh S. Markus, DM, FRCP; Kari-Matti Mäkelä, BM, BSc; Steve Bevan, PhD; Emma Raitoharju, MSc; Niku Oksala, MD, PhD; Joshua C. Bis, PhD; Chris O'Donnell, MD, MPH*; Atticus Hainsworth, PhD; Terho Lehtimäki, MD, PhD

***Background and Purpose***—A novel association between a single nucleotide polymorphism on chromosome 7p21.1 and large-vessel ischemic stroke was recently identified. The most likely underlying gene is histone deacetylase 9 (HDAC9). The mechanism by which HDAC9 increases stroke risk is not clear; both vascular and neuronal mechanisms have been proposed.

***Methods***—We determined whether the lead single nucleotide polymorphisms were associated with asymptomatic carotid plaque (N=25 179) and carotid intima-media thickness (N=31 210) detected by carotid ultrasound in a meta-analysis of population-based and community cohorts. Immunohistochemistry was used to determine whether HDAC9 was expressed in healthy human cerebral and systemic arteries. In the Tampere Vascular Study, we determined whether HDAC9 mRNA expression was altered in carotid (N=29), abdominal aortic (N=15), and femoral (N=24) atherosclerotic plaques compared with control (left internal thoracic, N=28) arteries.

***Results***—Both single nucleotide polymorphisms (rs11984041 and rs2107595) were associated with common carotid intima-media thickness (rs2107595; $P$=0.0018) and with presence of carotid plaque (rs2107595; $P$=0.0022). In both cerebral and systemic arteries, HDAC9 labeling was seen in nuclei and cytoplasm of vascular smooth muscle cells, and in endothelial cells. HDAC9 expression was upregulated in carotid plaques compared with left internal thoracic controls ($P$=0.00000103). It was also upregulated in aortic and femoral plaques compared with controls, with mRNA expression increased in carotid compared with femoral plaques ($P$=0.0038).

***Conclusions***—Our results are consistent with the 7p21.1 association acting via promoting atherosclerosis, and consistent with alterations in HDAC9 expression mediating this increased risk. Further studies in experimental models are required to confirm this link. (***Stroke.*** **2013;44:1220-1225.)**

**Key Words:** atherosclerosis ■ carotid stenosis ■ expression experiments ■ genetics ■ intima-media thickness

The Wellcome Trust Case Control Consortium 2 (WTCCC2) Ischemic Stroke Study identified a novel association between a single nucleotide polymorphism (SNP) on chromosome 7p21.1 and ischemic stroke.[1] This association was present with large artery atherosclerotic stroke but not with cardioembolic or small vessel stroke.

The most likely underlying gene is histone deacetylase 9 (HDAC9). All SNPs showing an association resided within a peak between 2 recombination hotspots and encompassed the tail end of HDAC9.[1] The mechanisms underlying this association remain uncertain. HDAC9 is a member of a large family of genes that encode proteins responsible for deacetylation of histones, and therefore regulation of chromatin structure and gene transcription.[2] Although known as histone deacetylases, these proteins also act on other substrates and lead to both upregulation and downregulation of genes. How HDAC9 might increase stroke risk is not clear. The specific association with large artery stroke would be consistent with a mechanism via accelerating atherosclerosis or promoting plaque instability. The HDAC9 protein inhibits myogenesis and is involved in heart development, but no specific mechanism that could predispose to pathological processes in large systemic arteries has yet been reported.[3] Alternatively, it could increase risk by altering brain ischemic responses, and therefore have effects on neuronal survival; although how this would only predispose to large artery

Received November 19, 2012; accepted January 29, 2013.

From the Stroke and Dementia Research Centre, St George's University of London, London, United Kingdom (H.S.M., S.B., A.H.); Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital, and School of Medicine, University of Tampere, Finland (K.-M.M., E.R., N.O., T.L.); Division of Vascular Surgery, Department of Surgery, Tampere University and Tampere University Hospital, Finland (N.O.); Department of Medicine, Cardiovascular Health Research Unit, University of Washington, Seattle, WA (J.C.B.); and National Heart, Lung and Blood Institute (NHLBI) and the NHLBI Framingham Heart Study, Framingham, MA (C.O.D.).

*Membership of Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium is listed in online-only Data Supplement.

**The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.111.000217/-/DC1.**

Correspondence to Hugh S. Markus, DM, FRCP, Stroke and Dementia Research Centre, St George's University of London, London SW19 3QZ, United Kingdom. E-mail hmarkus@sgul.ac.uk



***Stroke*** **is available at http://stroke.ahajournals.org** **DOI: 10.1161/STROKEAHA.111.000217**

stroke is unclear.[4] HDAC inhibitors have been suggested as a treatment for ischemic stroke.[4]

We performed a series of experiments to explore the mechanisms underlying the 7p21.1 association with large artery stroke. First, we determined whether the SNP associated with large artery stroke was also associated with asymptomatic carotid plaque and carotid intima-media thickness (IMT) measured in community populations. Duplex ultrasound imaging can noninvasively visualize atherosclerotic plaques themselves and diffuse thickening of the arterial wall (thickened IMT), which is an independent predictor of stroke.[5] Second, we determined whether HDAC9 was expressed in cerebral and systemic large arteries. Finally, we determined whether mRNA expression of HDAC9 was altered in atherosclerotic plaque using data from the Tampere Vascular Study.[6]

## Methods

### Associations With Carotid Plaque and IMT

Associations with carotid plaque and common carotid artery (CCA) IMT were examined in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium that brings together 5 population-based studies, and 4 additional community-based studies that had collaborated with the CHARGE consortium in a previous genome-wide association study (GWAS) of these phenotypes.[7] All individuals have GWAS data; these were used to perform a look-up of the SNPs.

Measurements of CCA-IMT were available on 31 210 participants from 9 studies, and of carotid artery plaque on 25 179 participants from 7 studies. The individual studies were as follows: Aging Gene-Environment Susceptibility-Reykjavik Study (AGES), Atherosclerosis Risk in Communities (ARIC) Study, Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), the Rotterdam Study I (RS-I), Old Order Amish (Amish) Study, Erasmus Ruchpen Family (ERF) Study, SardiNIA Study, and Study of Health in Pomerania (SHIP). For all studies in the meta-analyses, each participant provided written informed consent, and the local institutional review board approved the study. Studies contributing to this meta-analysis have been described in detail previously[7]; details are summarized in Table 1.

Each study evaluated the carotid arteries using B-mode ultrasonography and previously described reading protocols.[7] Data were used from the baseline examination, or the first examination in which carotid ultrasonography was obtained. CCA-IMT was typically summarized as the mean of the maximum of several measurements. For most studies, this was an average of multiple measurements of both the left and right arteries. All studies measured the far wall, and in addition, several included the near wall. We also examined atherosclerotic thickening of the carotid artery wall, defined in 7 of the 9 studies by either the presence of plaque (ARIC, AGES, ERF, CHS, RS-I, SHIP) or the proxy measure of stenosis >25% (FHS).

### Genotyping and Imputation

The 9 studies used commercial genotyping platforms available from Illumina and Affymetrix. Each study performed genotyping quality control checks and imputed ≈2.5 million polymorphic autosomal SNPs described in the HapMap Utah residents with ancestry from northern and western Europe population for each participant using available imputation methods. Details of individual study genotyping, imputation, and quality control procedures have been previously published.

### Statistical Analysis Within Studies

Each study independently implemented a predefined GWAS analysis plan. For the continuous measures of CCA-IMT, we evaluated cross-sectional associations of log(IMT)and genomewide variation using linear regression models (or linear mixed effects models, in Amish, FHS, and ERF to account for family relatedness). For each of the 2.5 million SNPs, each study fit additive genetic models relating genotype dosage (0–2 copies of the variant allele) with the study trait. For the dichotomous outcome of plaque, each study used logistic regression models (or general estimating equations clustering on family to account for familial correlations in FHS and ERF). In our primary analyses, all studies adjusted for age and sex. Some studies made additional adjustments, including study site (ARIC and CHS), familial structure (Amish, FHS, and ERF), or for whether the DNA had been whole genome amplified (FHS). Full details have been previously published.[7]

**Table 1. Details of the Individual Cohorts in the CHARGE Collaboration**

| Characteristics | AGES (N=3073) | Amish (N=1054) | ARIC (N=7767) | CHS (N=3261) | ERF (N=1809) | FHS (N=3004) | RS-I (N=4699) | SardiNIA (N=4235) | SHIP (N=2309) |
|---|---|---|---|---|---|---|---|---|---|
| Age, y | 76.4 (5.4) | 48.1 (15.9) | 54.3 (5.7) | 72.3 (5.4) | 48.5 (14.5) | 58.5 (9.7) | 68.9 (8.70) | 43.5 (17.5) | 61.8 (9.5) |
| Women, % | 57.7% | 49.4% | 53% | 61% | 56.5 | 53.3% | 59.3% | 56.2% | 48.6% |
| Hypertension, % | 80.6% | 9.3% | 27% | 51% | 51.4% | 40.5% | 55.9% | 29.1% | 72.4% |
| Diabetes mellitus, % | 11.6% | 2.1% | 8% | 12% | 6.1% | 8.6% | 10% | 4.8% | 10.1% |
| Current smoker, % | 12.6% | 9.4% | 25% | 11% | 39.4% | 15.6% | 23.4% | 20.2% | 19.2% |
| Total cholesterol, mg/dL | 217.9 (44.5) | 211.3 (48.1) | 214.7 (40.5) | 213.0 (38.9) | 214.4 (42.6) | 205.9 (39.7) | 256.0 (46.8) | 208.6 (42.1) | 234.3 (47.9) |
| HDL cholesterol, mg/dL | 61.0 (17.1) | 55.7 (14.8) | 50.7 (16.8) | 55.3 (15.8) | 49.5 (14.1) | 51.1 (16.1) | 51.8 (13.9) | 64.4 (14.9) | 55.3 (17.8) |
| Triglyceride, mg/dL | 107.0 (59.0) | 74.9 (47.1) | 136.0 (89.5) | 140.4 (76.4) | 118.6 (68.1) | 142.3 (138.6) | N/A | 87.2 (61.4) | 177.6 (134.8) |
| BMI, kg/m2 | 27.1 (4.5) | 26.9 (4.7) | 26.9 (4.7) | 26.3 (4.5) | 26.8 (4.7) | 27.9 (5.1) | 26.3 (3.7) | 25.3 (4.7) | 28.5 (4.6) |
| Prevalent CVD | 21.9% | 6.9% | 5% | 0% | 3.1% | 10.4% | 30.8% | 1.7% | 8.4% |
| IMT common carotid | 0.97 (0.1) | 0.74 (0.2) | 0.77 (0.2) | 1.03 (0.2) | 0.82 (0.2) | 0.74 (0.2) | 1.02 (0.2) | 0.54 (0.1) | 0.93 (0.2) |

Numbers in table are mean (SD) or percentage. N in the column headers indicates number of participants with common carotid IMT available. Diabetes mellitus was defined as fasting blood glucose >125 mg/dL, a random blood glucose of >200 mg/dL, or use of insulin or oral hypoglycemic agents; hypertension was defined as blood pressure >140/90 mm Hg or on antihypertensive medication; current cigarette smoking was defined as self-reported cigarette smoking of ≥1 cigarette per day for a year at any attended examination; cardiovascular disease was defined as coronary heart disease, stroke or transient ischemic attack, or congestive heart failure.

AGES indicates Aging Gene-Environment Susceptibility-Reykjavik Study; ARIC, Atherosclerosis Risk in Communities; BMI, body mass index; CHS, Cardiovascular Health Study; CVD, cardiovascular disease; ERF, Erasmus Ruchpen Family; FHS, Framingham Heart Study; HDL, high-density lipoprotein; IMT, intima-media thickness; RS-I, Rotterdam Study I; and SHIP, Study of Health in Pomerania.

**Table 2. CHARGE Results of the Association Analyses for the 2 Lead SNPs Tested for CCA-IMT**

| CCA-IMT SNP | Allele 1 | Allele 2 | Freq1.Z | Effect | SE | *P* Value | Direction | N | N_Eff | *P* Value.Z |
|---|---|---|---|---|---|---|---|---|---|---|
| rs11984041 | t | c | 0.088 | 0.0077 | 0.0027 | 0.00391 | ++++++++−+ | 31 210 | 26 187.95 | 0.006243 |
| rs2107595 | a | g | 0.146 | 0.0065 | 0.0021 | 0.00184 | +−+++++++ | 31 210 | 27 035.76 | 0.001833 |

CCA-IMT indicates common carotid artery intima-media thickness; and SNP, single nucleotide polymorphism.

## Meta-analysis

We conducted a meta-analysis of beta estimates and SEs from the 9 studies using an inverse-variance weighting approach as implemented in METAL.[8] Before meta-analysis, we calculated a genomic inflation factor ($\lambda_{gc}$) for each study to screen for cryptic population substructure or undiagnosed irregularities that might have inflated the test statistics. Inflation was low, with $\lambda_{gc}$ <1.09 in all studies. We applied genomic control to each study whose genomic inflation factor was >1.00 by multiplying all of the SEs by the square root of the study-specific $\lambda_{gc}$. For IMT, we express the association of each SNP and log(IMT) as the regression slope(β), its SE(β), and a corresponding *P* value. For the presence of plaque, we calculated a meta-analysis odds ratio, which represents the increase or decrease in the odds of plaque for each additional copy of the coded allele of the SNP.

We performed a look-up of 2 SNPs at 7p21.1. The WTCCC2 study found the strongest association with rs11984041, whereas recent GWAS meta-analysis in 12 389 ischemic stroke individuals and 62 004 controls found the strongest association with rs2107595, which is in linkage disequilibrium with rs11984041.[9] To account for the 2 SNPs, we applied a Bonferroni correction and predefined a significance level of 0.025.

## Immunohistochemistry Studies of HDAC9 in Normal Arteries

Expression of HDAC9 was examined by immunohistochemistry in human large arteries derived from surgical or post-mortem material: aorta (n=7), internal carotid (n=5), middle cerebral (n=5), and coronary arteries (n=5). Tissues were used with ethical approval via the UK National Research Ethics Service.

Anti-HDAC9 antibodies were rabbit polyclonal, 18970, and 59718 (both Abcam, Cambridge, United Kingdom); 18970 is raised against the peptide EVPVGLEPISPLDLRT (corresponding to residues 12–27 of human HDAC9 isoform 1) present in human HDAC9 isoforms 1,3,5,6,7,CRA_g, CRA_i, CRA_j, and 59718 is raised against a peptide corresponding to amino acids 541 to 590 at the C-terminal of human HDAC9 isoform 6 that is found in human HDAC9 isoforms 3,6,7,8,9,10, CRA_h, CRA_i, CRA_j.

Other antibodies used for immunohistochemistry were as follows: CD31(PECAM1), CD45 (leukocyte common antigen; clones 2B11 and PD7/26), and CD68 (clone PG-M1); all mouse monoclonals from Dako, Ely, United Kingdom; and smooth muscle α-actin (mouse monoclonal, clone 1A4) and smooth muscle myosin (mouse monoclonal, clone h-SMV) from Sigma-Aldrich, Poole, United Kingdom.

Paraffin wax–embedded sections (6 μm) were processed for standard immunohistochemical labeling. Endogenous peroxidase activity was quenched by $H_2O_2$ (3% v/v, aqueous solution) for 8 minutes. After high-pressure heat-induced antigen retrieval (30 s, 125°C, in pH 7.8 Tris-citrate buffer), sections were exposed to primary antibodies. HDAC9 primary antibodies ab18970 and ab59718 were applied to human tissues (1:300) and to pig tissue (1:500) overnight at 4°C.

Antibody labeling was visualized using a peroxidase-conjugated secondary reagent (Envision kit, K-5007, Dako, Ely, United Kingdom) and diaminobenzidine chromogen, then counterstained with Mayer's hematoxylin. Sections were examined on a Zeiss Axioplan-2 microscope driven by Axiovision software (version 4.7).

## Messenger RNA Expression Studies

Carotid, femoral, and aortic atherosclerotic plaques constituting the intima and inner media were prospectively obtained between 2005 and 2009 from patients fulfilling the following inclusion criteria: (1) carotid endarterectomy attributable to asymptomatic or symptomatic >70% carotid stenosis, or (2) femoral or (3) aortic endarterectomy with aortoiliac or aortobifemoral bypass attributable to symptomatic peripheral arterial disease. Whole thickness left internal thoracic artery samples were used as controls and obtained during coronary artery bypass surgery. All open vascular surgical procedures were performed at the Division of Vascular Surgery and Heart Center, Tampere University Hospital. The study was approved by the local ethics committee; all patients gave informed consent.

Fresh tissue samples were immediately soaked in RNALater solution (Ambion Inc) and homogenized using an Ultra-Turrax T80 homogenizer (IKA). RNA was extracted with the Trizol reagent (Invitrogen) and miRNEasy Mini-Kit (Qiagen) with the RNase-Free DNase Set (Qiagen) according to manufacturer instructions. The RNA isolation protocol was validated by analyzing the integrity of the RNA with the RNA 6000 Nano Chip Kit (Agilent).

The expression levels were analyzed with an Illumina HumanHT-12 v3 Expression BeadChip (Illumina). In brief, 300–500 ng of RNA was reverse transcribed in cRNA and biotin-UTP labeled using the IlluminaTotalPrep RNA Amplification Kit (Ambion), and 1500 ng of cRNA was then hybridized to the Illumina HumanHT-12 v3 Expression BeadChip.

The BeadChips were scanned with the Illuminai Scan system. After background subtraction, raw intensity data were exported using the Illumina Genome Studio software. Further data processing was conducted by means of R language and appropriate Bioconductor modules. Data were log2-transformed, and robust multichip average and robust spline normalization (rma_rsn) were used.

# Results

## Associations With Carotid Plaque and IMT

Both SNPs (rs11984041 and rs 2107595) were associated with both CCA-IMT (Table 2) and with presence of carotid plaque (Table 3). The strongest associations were seen for SNPs rs 2107595; CCA-IMT *P*=0.0018; carotid plaque *P*=0.0022.

## Immunohistochemistry Studies of HDAC9 in Normal Arteries

In all arterial beds, strong HDAC9 labeling was seen in nuclei and cytoplasm of VSMC, and in endothelial cells, where intact endothelia were visible (example in Figure 1A). In the medial layer, a high fraction of cells were labeled (80% to 90%) with

**Table 3. CHARGE Results of the Association Analyses for the 2 Lead SNPs Tested for Carotid Plaque**

| Carotid Plaque SNP | Allele 1 | Allele 2 | Freq1.Z | Effect | SE | *P* Value | Direction | N | N_eff | *P* Value.Z |
|---|---|---|---|---|---|---|---|---|---|---|
| rs11984041 | t | c | 0.097 | 0.1069 | 0.0374 | 0.00425 | +++−+++ | 25 179 | 21 616.84 | 0.002554 |
| rs2107595 | a | g | 0.159 | 0.0911 | 0.0298 | 0.00222 | +++++++ | 25 179 | 22 257.60 | 0.001395 |

SNP indicates single nucleotide polymorphism.

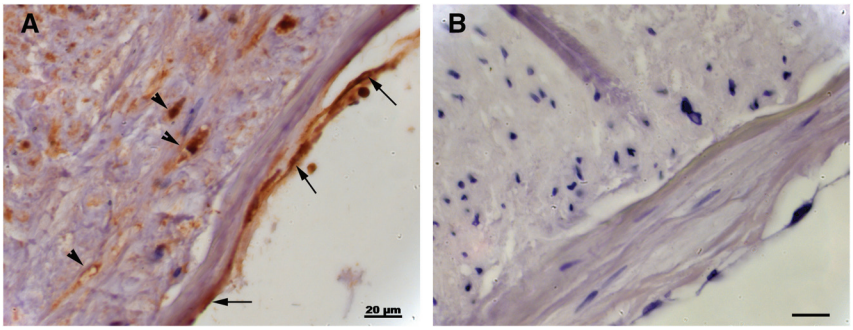



**Figure 1.** Immunohistochemical labeling of histone deacetylase 9 (HDAC9) in human middle cerebral artery. Post-mortem sample of human middle cerebral artery labeled for HDAC9 (**A**) or with no primary antibody (**B**). **A**, Cellular HDAC9 immunoreactivity (brown) is evident in endothelial cells (examples marked with **arrows**) and in myoctes within the medial layer (examples marked with **arrowheads**). **B**, Labeling was absent from neighboring sections treated identically but with no primary antibody. Scale bars=20 μm.

a similar distribution to the VSMC-specific marker smooth muscle myosin (Figure 1). A minority of medial cells, with fibroblast-like morphology, were negative for HDAC9 and smooth muscle myosin, consistent with normal incidence of structural fibroblasts. Similar results were obtained with the 2 different anti-HDAC9 antibodies used.

Distinct patterns of labeling were observed with primary antibodies to leukocyte common antigen (CD45) and a lysosomal marker for macrophage/monocytic cells (CD68) at similar titer. Immunolabeling was absent in adjacent negative control sections treated without primary antibody (Figures 1 and 2).

### Messenger RNA Expression Studies

Gene expression was analyzed from 29 carotid, 15 abdominal aorta, 24 femoral plaques, and 28 atherosclerosis free left internal thoracic artery controls. Demographics and American Heart Association plaque grading[10] for the different plaques are shown in Table 4.

HDAC9 expression was upregulated in carotid plaques compared with left internal thoracic artery controls. ($P$=0.00000103; fold change [FC]=3.06). It was also upregulated in aortic plaques ($P$=0.0038; FC=1.76) and femoral plaques ($P$=0.038; FC=1.57) compared with controls. HDAC9 mRNA expression was greater in carotid compared with femoral plaques ($P$=0.0038; FC=1.76), although there was no significant difference between carotid and aortic plaques ($P$=0.90; FC=1.19).

## Discussion

Our results show the 7p21.1 locus, previously associated with large artery stroke, is associated with asymptomatic carotid plaque and carotid IMT in community populations. This is consistent with a mechanism related to acceleration of the progression of atherosclerosis. HDAC9 is the most likely gene underlying this association. Consistent with this, we demonstrated that HDAC9 is expressed in VSMC and endothelium of healthy human adult large arteries, including cerebral and systemic arteries. A similar pattern was obtained with 2 antibodies raised against 2 nonoverlapping HDAC9-specific sequences. Consistent with a role in atherosclerosis, we found increased expression of HDAC9 mRNA in carotid atherosclerotic plaques.

Although canonical HDACs are ubiquitously expressed, Class IIa HDACs (including HDAC9) have more restricted expression. Expression in heart, pancreatic islets, spinal cord, and brain of mouse embryos has been demonstrated, and

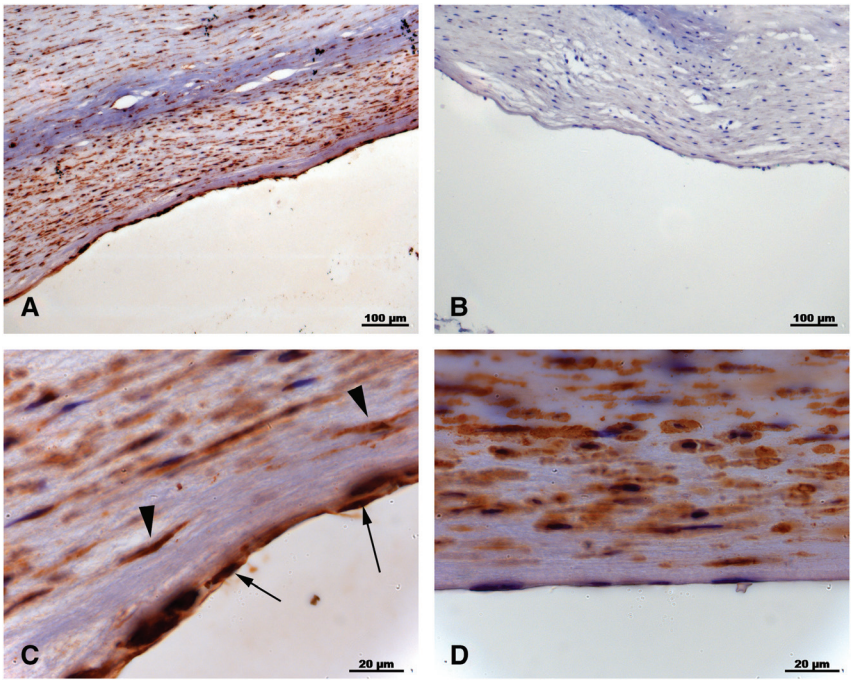



**Figure 2.** Immunohistochemical labeling of histone deacetylase 9 (HDAC9) in human aorta. Cellular HDAC9 immunoreactivity (brown) in adult human aorta (**A**) was absent from a neighboring negative control section (**B**). **C**, Higher magnification shows HDAC9 in vascular smooth muscle cells (VSMC; **arrowheads**) and endothelial cells (**arrows**). In VSMC, a similar pattern of labeling is seen with smooth muscle myosin (SMM; **D**). Hematoxylin counterstain (blue) labels nuclear chromatin. Scale bars, 100 μm (**A** and **B**); 20 μm (**C** and **D**).

**Table 4. Demographics, Risk Factors, and AHA Plaque Class of Plaques From Different Vascular Beds**

| | Carotid Plaque | Aortic Plaque | Femoral Plaque | Control Arteries | All |
|---|---|---|---|---|---|
| N | 29 | 15 | 24 | 28 | 96 |
| Age (median), y (SD) | 70 (9.5) | 61.0 (10.8) | 76.0 (9.4) | 69.0 (9.6) | 69.0 (10.2) |
| Men, % | 62.1 | 73.3 | 70.8 | 82.1 | 71.9 |
| Body mass index (median), kg/m$^2$ (SD) | 25.6 (3.4) | 25.9 (4.2) | 26.7 (4.3) | 28.2 (5.1) | 27.0 (4.4) |
| Dyslipidemia, % | 75.9 | 46.7 | 70.8 | 85.7 | 72.9 |
| Statins, % | 100.0 | 40.0 | 62.5 | 82.1 | 76 |
| Hypertension, % | 79.3 | 80.0 | 87.5 | 100.0 | 87.5 |
| Blood pressure medication, % | 82.8 | 80.0 | 79.2 | 92.9 | 84.4 |
| History of smoking % | 65.5 | 100.0 | 70.8 | 64.3 | 71.9 |
| AHA Class V–VI, %, of the atherosclerotic arteries | 82.8 | 73.3 | 62.5 | NA | 74.6 |

AHA indicates American Heart Association.

human tissue lysates for HDAC9 mRNA show high expression in skeletal muscle and brain. There are reports of HDAC9 protein expression using immunohistochemical labeling in cerebral medulloblastoma tumors (using one of the antibodies we used, ab59718)[11] and in teeth, using a different antibody.[12] We have been unable to find published data on HDAC9 expression in human blood vessels.

Since its discovery as a risk factor for stroke, a recent very large GWAS meta-analysis in 63 746 coronary artery disease cases and 130 681 controls has found an association of the 7p21.1 locus with coronary artery disease but with a much smaller effect size[13]; the odds ratio was 1.09 compared with 1.42 with large artery stroke in WTCCC2.[1] This suggests this locus predisposes to large artery disease in the carotid arteries to a much greater extent than to coronary artery disease. Interestingly, we found HDAC9 mRNA expression was greater in carotid compared with femoral plaques. How such a risk factor would preferentially increase risk of carotid plaque is uncertain. One possible factor is flow-dependent mechanisms dependent on local anatomy; local hemodynamic factors, and the anatomy of the carotid bifurcation, are known to be related to early atherosclerotic changes.[14]

Taken together, our results are consistent with the 7p21.1 locus acting as a risk factor for atherosclerosis. Such an association with large artery stroke could be via increasing plaque development, or by mechanisms that result in plaque instability and increase the risk of subsequent thromboembolism, the major cause of stroke in large artery disease. The association with asymptomatic carotid plaque, plaques that have not yet become unstable and symptomatic, would support the former mechanism. We also found an association with carotid IMT, consistent with increased risk occurring at the earlier stages of plaque formation. Increased carotid IMT is believed to occur with both early atherosclerosis and also vascular remodeling.[15]

HDACs catalyze removal of acetyl groups from ε-amino groups of lysine residues in a variety of proteins. HDACs have been studied mainly in the context of chromatin, where they serve an epigenetic function by deacetylating nucleosomal histones and altering the electrostatic properties of chromatin leading to gene repression. However, it is now recognized that HDACs deacetylate many nonhistone proteins, and are therefore also referred to as lysine deacetylases.[16] There are 18 HDACs that are encoded by different genes and grouped into 4 classes on the basis of similarity to yeast transcriptional repressors.[17] HDAC9 is a member of the class IIa HDACs. The class IIa HDACs interact with members of the myocyte enhancer factor-2 transcription factor family,[18] which are regulators of VSMC proliferation. Given the VSMC expression of HDAC9, increased risk of large-vessel disease could be via promotion of atherosclerosis as a consequence of MDAC9-mediated increased VSMC proliferation, an action impeded by HDAC9 inhibition in vitro.[19] HDAC inhibitors also have been shown to reduce proinflammatory cytokine expression, which has been implicated in atherosclerosis.[17] The antiepileptic drug sodium valproate has nonspecific HDAC inhibitory properties and has been shown to inhibit atherosclerosis in animal models.[20] Intriguingly, sodium valproate therapy in man has been associated with lower stroke and myocardial infarction rates compared with other antiepileptic drugs.[21] Specific inhibitors to a variety of HDACs are currently being developed and might offer potential in stroke and cardiovascular prevention.[17]

The 2 HDAC9 SNPs we assessed for association with IMT were those most strongly associated previously with large artery stroke.[1,9] They are in close linkage disequilibrium; in the 1000 genomes European ancestry individuals, linkage disequilibrium measures between the 2 SNPs are $R^2$=0.568, D-prime=0.936. Both are likely to be markers for an as-yet-unknown functional variant.

There are potential limitations to this study. Not all patients in whom IMT was measured had carotid plaque measured also. However, this would have tended to reduce power to detect association with plaque, and we found such an association. The CHARGE consortium includes several different populations, which introduces heterogeneity; therefore, we analyzed using a meta-analysis approach, and the associations we found were consistent across almost all populations (Table 2). In the mRNA expression studies, we used relatively small sample sizes, although we were still able to detect upregulation of HDAC9 in atherosclerotic plaque.

In conclusion, our results are consistent with the 7p21.1 association, which has previously been associated with symptomatic large artery stroke, acting via alterations in HDAC9 expression promoting atherosclerosis. Further studies in experimental models are now required to prove this association is indeed mediated via accelerated carotid atherosclerosis.

## Acknowledgments

We thank the staff of Oxford Brain Collection and St George's Hospital Cellular Pathology for supplying human tissue.

## Sources of Funding

This study was supported in part by European Union 7th Framework Program funding for the AtheroRemo project [201668]. Tampere Vascular study has been supported by the Academy of Finland, Tampere, University Hospital Medical Fund (grant 9M048 and 9N035 for TeLeht), the Finnish Foundation of Cardiovascular Research, the Tampere Tuberculosis Foundation, and the Emil Aaltonen Foundation. Dr Hainsworth is supported by Action Medical Research and Neuroscience Research Foundation.

## Disclosures

None.

## References

1. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet*. 2012;44:328–33.
2. Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet*. 2009;10:32–42.
3. Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA, Olson EN. Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell*. 2002;110:479–488.
4. Langley B, Brochier C, Rivieccio MA. Targeting histone deacetylases as a multifaceted approach to treat the diverse outcomes of stroke. *Stroke*. 2009;40:2899–2905.
5. Lorenz MW, Markus HS, Bots ML, Rosvall M, Sitzer M. Prediction of clinical cardiovascular events with carotid intima-media thickness: a systematic review and meta-analysis. *Circulation*. 2007;115:459–467.
6. Levula M, Oksala N, Airla N, Zeitlin R, Salenius JP, Järvinen O, et al. Genes involved in systemic and arterial bed dependent atherosclerosis–Tampere Vascular study. *PLoS ONE*. 2012;7:e33787.
7. Bis JC, Kavousi M, Franceschini N, Isaacs A, Abecasis GR, Schminke U, et al; CARDIoGRAM Consortium. 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;43:940–947.
8. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191.
9. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell, JC, Cheng Y, et al. Genome-wide associations with ischaemic stroke and its subtypes- the METASTROKE collaboration. *Lancet Neurol*. 2012;11:951–962
10. Stary HC, Chandler AB, Dinsmore RE, Fuster V, Glagov S, Insull W Jr, et al. A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Arterioscler Thromb Vasc Biol*. 1995;15:1512–1531.
11. Milde T, Oehme I, Korshunov A, Kopp-Schneider A, Remke M, Northcott P, et al. HDAC5 and HDAC9 in medulloblastoma: novel markers for risk stratification and role in tumor cell growth. *Clin Cancer Res*. 2010;16:3240–3252.
12. Klinz FJ, Korkmaz Y, Bloch W, Raab WH, Addicks K. Histone deacetylases 2 and 9 are coexpressed and nuclear localized in human molar odontoblasts in vivo. *Histochem Cell Biol*. 2012;137:697–702.
13. The CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25–33.
14. Sitzer M, Puac D, Buehler A, Steckel DA, von Kegler S, Markus HS, et al. Internal carotid artery angle of origin: a novel risk factor for early carotid atherosclerosis. *Stroke*. 2003;34:950–955.
15. Mathiesen EB, Johnsen SH. Ultrasonographic measurements of subclinical carotid atherosclerosis in prediction of ischemic stroke. *ActaNeurolScand Suppl*. 2009;189:68–72.
16. Yang XJ, Seto E. Lysine acetylation: codified crosstalk with other post-translational modifications. *Mol Cell*. 2008;31:449–461.
17. McKinsey TA. Targeting inflammation in heart failure with histone deacetylase inhibitors. *Mol Med*. 2011;17:434–441.
18. McKinsey TA, Zhang CL, Olson EN. MEF2: a calcium-dependent regulator of cell division, differentiation and death. *Trends Biochem Sci*. 2002;27:40–47.
19. Okamoto H, Fujioka Y, Takahashi A, Takahashi T, Taniguchi T, Ishikawa Y, et al. Trichostatin A, an inhibitor of histone deacetylase, inhibits smooth muscle cell proliferation via induction of p21(WAF1). *J Atheroscler Thromb*. 2006;13:183–191.
20. Bowes AJ, Khan MI, Shi Y, Robertson L, Werstuck GH. Valproate attenuates accelerated atherosclerosis in hyperglycemic apoE-deficient mice: evidence in support of a role for endoplasmic reticulum stress and glycogen synthase kinase-3 in lesion development and hepatic steatosis. *Am J Pathol*. 2009;174:330–342.
21. Olesen JB, Abildstrøm SZ, Erdal J, Gislason GH, Weeke P, Andersson C, et al. Effects of epilepsy and selected antiepileptic drugs on risk of myocardial infarction, stroke, and death in patients with or without previous stroke: a nationwide cohort study. *Pharmacoepidemiol Drug Saf*. 2011;20:964–971.

**SUPPLEMENTAL MATERIAL**

# Evidence HDAC9 genetic variant associated with ischaemic stroke increases risk via promoting carotid atherosclerosis.

## CHARGE consortium membership

Joshua C Bis, Maryam Kavousi, Nora Franceschini, Aaron Isaacs, Gonçalo R Abecasis, Ulf Schminke, Wendy S Post, Albert V Smith, L Adrienne Cupples, Hugh S Markus, Reinhold Schmidt, Jennifer E Huffman, Terho Lehtimäki, Jens Baumert, Thomas Münzel, Susan R Heckbert, Abbas Dehghan, Kari North, Ben Oostra, Steve Bevan, Eva-Maria Stoegerer, Caroline Hayward, Olli Raitakari, Christa Meisinger, Arne Schillert, Serena Sanna, Henry Völzke, Yu-Ching Cheng, Bolli Thorsson, Caroline S Fox, Kenneth Rice, Fernando Rivadeneira, Vijay Nambi, Eran Halperin, Katja E Petrovic, Leena Peltonen, H Erich Wichmann, Renate B Schnabel, Marcus Dörr, Afshin Parsa, Thor Aspelund, Serkalem Demissie, Sekar Kathiresan, Muredach P Reilly, the CARDIoGRAM Consortium, Kent Taylor, Andre Uitterlinden, David J Couper, Matthias Sitzer, Mika Kähönen, Thomas Illig, Philipp S Wild, Marco Orru, Jan Lüdemann, Alan R Shuldiner, Gudny Eiriksdottir, Charles C White, Jerome I Rotter, Albert Hofman, Jochen Seissler, Tanja Zeller, Gianluca Usala, Florian Ernst, Lenore J Launer, Ralph B D'Agostino Sr, Daniel H O'Leary, Christie Ballantyne, Joachim Thiery, Andreas Ziegler, Edward G Lakatta, Ravi Kumar Chilukoti, Tamara B Harris, Philip A Wolf, Bruce M Psaty, Joseph F Polak, Xia Li, Wolfgang Rathmann, Manuela Uda, Eric Boerwinkle, Norman Klopp, Helena Schmidt, James F Wilson, Jorma Viikari, Wolfgang Koenig, Stefan Blankenberg, Anne B Newman, Jacqueline Witteman, Gerardo Heiss, Cornelia van Duijn, Angelo Scuteri, Georg Homuth,

Braxton D Mitchell, Vilmundur Gudnason & Christopher J O'Donnell

The following studies contributed to the CHARGE Consortium GWAS discovery stage:

The Age, Gene/Environment Susceptibility (AGES) Reykjavik Study was funded by US National Institutes of Health (NIH) contract N01-AG-12100, the National Institute on Aging (NIA) Intramural Research Program, Hjartavernd (the Icelandic Heart Association) and the Althingi (the Icelandic Parliament).

The Old Order Amish Studies were supported by grants and contracts from the NIH including R01 AG18728 (Amish Longevity Study), R01 HL088119 (Amish Calcification Study), U01 GM074518-04 (The Amish Pharmacogenomics of Anti-Platelet Intervention (PAPI) study) and U01 HL072515-06 (the Heredity and Phenotype Intervention (HAPI) heart study), the University of Maryland General Clinical Research Center grant M01 RR 16500, the Baltimore Veterans Administration Medical Center Geriatrics Research and Education Clinical Center and the Paul Beeson Physician Faculty Scholars in Aging Program. We thank our Amish research volunteers for their long-standing partnership in research and the research staff at the Amish Research Clinic for their hard work and dedication.

The Atherosclerosis Risk in Communities Study (ARIC) was carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and US National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for

their important contributions. Study infrastructure was partly supported by grant number UL1RR025005, a component of the US National Institutes of Health and NIH Roadmap for Medical Research.

The Erasmus Rucphen Family Study was supported by grants from The Netherlands Organisation for Scientific Research, Erasmus Medical Center and the Centre for Medical Systems Biology (CMSB). We are grateful to all study participants and their relatives, the general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.

Cardiovascular Health Study: This CHS research was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, HHSN268201200036C and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from NINDS. Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the NIA. See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by the National Center for Research Resources, grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The Framingham Heart Study of the National Heart, Lung, and Blood Institute of the US National Institutes of Health and Boston University School of Medicine was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (Contract No.

N01-HC-25195) and its contract with Affymetrix, Inc. for genotyping services (Contract No. N02-HL-6-4278) and by grants from the National Institute of Neurological Disorders and Stroke (NS17950, P.A.W.) and the National Institute of Aging (AG08122, AG16495, P.A.W. and AG033193, S.S.). A portion of this research used the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project.

Rotterdam Study I and II (RS I and RS II): The Rotterdam GWAS was funded by the Netherlands Organisation of Scientific Research (NWO, De Nederlandse Organisatie voor Wetenschappelijk Onderzoek) Investments (number 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Aging (NCHA) project number 050-060-810. This study was further supported by an NWO grant (vici, 918-76-619). The Rotterdam Study was funded by the Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are very grateful to the participants and staff from the Rotterdam Study, the participating general practioners and the pharmacists. We thank P. Arp, M. Jhamai, M. Moorhouse, M. Verkerk and S. Bervoets for their help in creating the GWAS database. We would like to thank T.A. Knoch, L.V. de Zeeuw, A. Abuseiris and R. de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the National German MediGRID and Services@MediGRID part of the German D-Grid, both funded by

the German Bundesministeriumfuer Forschung und Technology under grants #01 AK 803 A-H and #01 IG 07015 G, for access to their grid resources.

The SardiNIA Study: This work was supported by the Intramural Research Program of the National Institute on Aging, NIH. The SardiNIA ('Progenia') team was supported by Contract NO1-AG-1–2109 from the National Institute on Aging. The efforts of G.R.A. were supported in part by contract 263-MA-410953 from the National Institute on Aging to the University of Michigan and by research grants HG005581 and HL084729 from the National Institutes of Health (to G.R.A.).

The Study of Health in Pomerania (SHIP) is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants number 01ZZ9603, 01ZZ0103 and 01ZZ0403), and the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany, and the Federal State of Mecklenburg-West Pomerania. The SHIP authors are grateful to the contribution of A. Teumer, A. Hoffmann and A. Petersmann in generating the SNP data. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG.

OPEN ACCESS Freely available online

PLOS | GENETICS

# A Novel *MMP12* Locus Is Associated with Large Artery Atherosclerotic Stroke Using a Genome-Wide Age-at-Onset Informed Approach

**Matthew Traylor[1]*, Kari-Matti Mäkelä[2,3], Laura L. Kilarski[1], Elizabeth G. Holliday[4,5], William J. Devan[6,7], Mike A. Nalls[8], Kerri L. Wiggins[9], Wei Zhao[10], Yu-Ching Cheng[11,12], Sefanja Achterberg[13], Rainer Malik[14], Cathie Sudlow[15], Steve Bevan[16], Emma Raitoharju[2,3], METASTROKE, International Stroke Genetics Consortium, Wellcome Trust Case Consortium 2 (WTCCC2)[¶], Niku Oksala[2,3,17], Vincent Thijs[18,19,20], Robin Lemmens[18,19,20], Arne Lindgren[21,22], Agnieszka Slowik[23], Jane M. Maguire[4,5,24,25], Matthew Walters[26], Ale Algra[13,27], Pankaj Sharma[28], John R. Attia[4,5,25], Giorgio B. Boncoraglio[29], Peter M. Rothwell[30], Paul I. W. de Bakker[7,27,31,32], Joshua C. Bis[9], Danish Saleheen[33,34], Steven J. Kittner[12], Braxton D. Mitchell[11], Jonathan Rosand[6,7], James F. Meschia[35], Christopher Levi[5,25], Martin Dichgans[14,36], Terho Lehtimäki[2,3], Cathryn M. Lewis[37,38], Hugh S. Markus[16]**

**1** Stroke and Dementia Research Centre, St George's University of London, London, United Kingdom, **2** Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland, **3** School of Medicine, University of Tampere, Tampere, Finland, **4** Center for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, Callaghan, New South Wales, Australia, **5** Hunter Medical Research Institute, New Lambton Heights, New South Wales, Australia, **6** Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **7** Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, **8** Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland, United States of America, **9** Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, United States of America, **10** Perelman School of Medicine, Division of Translational Medicine and Human Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **11** Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **12** Research and Development Program, Veterans Affairs Maryland Health Care System, Baltimore, Maryland, United States of America, **13** Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands, **14** Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany, **15** Division of Clinical Neurosciences and Insititute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, **16** Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom, **17** Department of Surgery, Tampere University Hospital, Tampere, Finland, **18** KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology - Laboratory of Neurobiology, Leuven, Belgium, **19** VIB - Vesalius Research Center, Leuven, Belgium, **20** University Hospitals Leuven, Department of Neurology, Leuven, Belgium, **21** Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden, **22** Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden, **23** Department of Neurology, Jagiellonian University, Krakow, Poland, **24** School of Nursing and Midwifery, University of Newcastle, Callaghan, New South Wales, Australia, **25** Centre for Translational Neuroscience and Mental Health, University of Newcastle, Callaghan, New South Wales, Australia, **26** Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, United Kingdom, **27** Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, **28** Imperial College Cerebrovascular Research Unit (ICCRU), Imperial College London, London, United Kingdom, **29** Department of Cereberovascular Disease, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Neurologico Carlo Besta, Milan, Italy, **30** Stroke Prevention Research Unit, Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, United Kingdom, **31** Department of Medical Genetics, University Medical Centre, Utrecht, The Netherlands, **32** Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **33** Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **34** Center for Non-Communicable Diseases, Karachi, Pakistan, **35** Department of Neurology, Mayo Clinic, Jacksonville, Florida, United States of America, **36** Munich Cluster for Systems Neurology (SyNergy), Ludwig-Maximilians-Universität, Munich, Germany, **37** Department of Medical & Molecular Genetics, King's College London, London, United Kingdom, **38** Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom

## Abstract

Genome-wide association studies (GWAS) have begun to identify the common genetic component to ischaemic stroke (IS). However, IS has considerable phenotypic heterogeneity. Where clinical covariates explain a large fraction of disease risk, covariate informed designs can increase power to detect associations. As prevalence rates in IS are markedly affected by age, and younger onset cases may have higher genetic predisposition, we investigated whether an age-at-onset informed approach could detect novel associations with IS and its subtypes; cardioembolic (CE), large artery atherosclerosis (LAA) and small vessel disease (SVD) in 6,778 cases of European ancestry and 12,095 ancestry-matched controls. Regression analysis to identify SNP associations was performed on posterior liabilities after conditioning on age-at-onset and affection status. We sought further evidence of an association with LAA in 1,881 cases and 50,817 controls, and examined mRNA expression levels of the nearby genes in atherosclerotic carotid artery plaques. Secondly, we performed permutation analyses to evaluate the extent to which age-at-onset informed analysis improves significance for novel loci. We identified a novel association with an *MMP12* locus in LAA (rs660599; $p = 2.5 \times 10^{-7}$), with independent replication in a second population ($p = 0.0048$, OR(95% CI) = 1.18(1.05–1.32); meta-analysis $p = 2.6 \times 10^{-8}$). The nearby gene, *MMP12*, was significantly overexpressed in carotid plaques compared to atherosclerosis-free control arteries ($p = 1.2 \times 10^{-15}$; fold change = 335.6). Permutation analyses demonstrated improved significance for associations when accounting for age-at-onset in all four stroke phenotypes ($p < 0.001$). Our results show that a covariate-informed design, by adjusting for age-at-onset of stroke, can detect variants not identified by conventional GWAS.

**Citation:** Traylor M, Mäkelä K-M, Kilarski LL, Holliday EG, Devan WJ, et al. (2014) A Novel *MMP12* Locus Is Associated with Large Artery Atherosclerotic Stroke Using a Genome-Wide Age-at-Onset Informed Approach. PLoS Genet 10(7): e1004469. doi:10.1371/journal.pgen.1004469

**Editor:** Timothy M. Frayling, University of Exeter Medical School, United Kingdom

**Received** December 21, 2013; **Accepted** May 14, 2014; **Published** July 31, 2014



**Funding:** The principal funding for this study was provided by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). HSM is supported by an NIHR Senior Investigator award and the NIHR Biomedical Research Centre at Cambridge. We acknowledge support from the National Institutes of Health Research Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. The PROMISe Study (The Netherlands) was made possible, in part, by a Complementation Grant to PIWdB from the Biobanking and Biomolecular Resources Research Infrastructure in the Netherlands (BBMRI-NL). SA is supported in part by a grant from the Netherlands Heart Foundation (grant no. 2005B031) and a grant from the Dutch Brain Foundation (project 2008(1).10). The Australian Stroke Genetics Collaboration (ASGC) Australian population control data was derived from the Hunter Community Study. We also thank the University of Newcastle for funding and the men and women of the Hunter region who participated in this study. This research was funded by grants from the Australian National and Medical Health Research Council (NHMRC Project Grant ID: 569257), the Australian National Heart Foundation (NHF Project Grant ID: G 04S 1623), the University of Newcastle, the Gladys M Brawn Fellowship scheme and the Vincent Fairfax Family Foundation in Australia. EGH is supported by the Australian NHMRC Fellowship scheme. The Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA was supported by the National Institutes of Health Genes, Environment and Health Initiative (GEI) Grant U01 HG004436, as part of the GENEVA consortium under GEI, with additional support provided by the Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488); and the Office of Research and Development, Medical Research Service, and the Baltimore Geriatrics Research, Education, and Clinical Center of the Department of Veterans Affairs. Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the National Institutes of Health to the Johns Hopkins University (contract number HHSN268200782096C). Assistance with data cleaning was provided by the GENEVA Coordinating Center (U01 HG 004446; PI Bruce S Weir). Study recruitment and assembly of datasets were supported by a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and by grants from the National Institute of Neurological Disorders and Stroke (NINDS) and the NIH Office of Research on Women's Health (R01 NS45012, U01 NS069208-01). The Heart and Vascular Health Study (HVH) research reported in this article was funded by NHLBI grants R01 HL085251 and R01 HL073410 The Ischemic Stroke Genetics Study (ISGS)/Siblings With Ischemic Stroke Study (SWISS) study was supported in part by the Intramural Research Program of the National Institute on Aging, NIH project Z01 AG-000954-06. ISGS/SWISS used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds), human subjects protocol numbers 2003-081 and 2004-147. ISGS/SWISS used stroke-free participants from the Baltimore Longitudinal Study of Aging (BLSA) as controls. The inclusion of BLSA samples was supported in part by the Intramural Research Program of the National Institute on Aging, NIH project Z01 AG-000015-50, human subjects protocol number 2003-078.The ISGS study was funded by NIH-NINDS Grant R01 NS-42733 (JFM, P.I.). The SWISS study was funded by NIH-NINDS Grant R01 NS-39987 (JFM, P.I.). This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (http://biowulf.nih.gov). The MGH Genes Affecting Stroke Risk and Outcome Study (MGH-GASROS) GASROS was supported by The National Institute of Neurological Disorders and Stroke (U01 NS069208), the American Heart Association/Bugher Foundation Centers for Stroke Prevention Research 0775010N, the National Institutes of Health and National Heart, Lung, and Blood Institute's STAMPEED genomics research program (R01 HL087676) and a grant from the National Center for Research Resources. The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278 from the National Center for Research resources. Milano - Besta Stroke Register Collection and genotyping of the Milan cases within CEDIR were supported by Annual Research Funding of the Italian Ministry of Health (Grant Numbers: RC 2007/LR6, RC 2008/LR6; RC 2009/LR8; RC 2010/LR8). FP6 LSHM-CT-2007-037273 for the PROCARDIS control samples. The Wellcome Trust Case-Control Consortium 2 (WTCCC2) The principal funding for the WTCCC2 stroke study was provided by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). The Stroke Association provided additional support for collection of some of the St George's, London cases. The Oxford cases were collected as part of the Oxford Vascular Study which is funded by the Medical Research Council, Stroke Association, Dunhill Medical Trust, National Institute of Health Research (NIHR) and the NIHR Biomedical Research Centre, Oxford. PMR has a Wellcome Trust Senior Investigator Award and an NIHR Senior Investigator Award. The Edinburgh Stroke Study was supported by the Wellcome Trust (clinician scientist award to CS), and the Binks Trust. Sample processing occurred in the Genetics Core Laboratory of the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. This work was supported by grants received from the German Federal Ministry of Education and Research (BMBF) in the context of the e:Med program (e:AtheroSysMed) and the FP7 European Union project CVgenes@target (261123) to Martin Dichgans. Much of the neuroimaging occurred in the Scottish Funding Council Brain Imaging Research Centre (www.sbirc.ed.ac.uk), Division of Clinical Neurosciences, University of Edinburgh, a core area of the Wellcome Trust Clinical Research Facility and part of the SINAPSE (Scottish Imaging Network – A Platform for Scientific Excellence) collaboration (www.sinapse.ac.uk), funded by the Scottish Funding Council and the Chief Scientist Office. Collection of the Munich cases and data analysis was supported by the Vascular Dementia Research Foundation. This work made use of data and samples generated by the 1958 Birth Cohort (NCDS). Access to these resources was enabled via the 58READIE Project funded by Wellcome Trust and Medical Research Council (grant numbers WT095219MA and G1001799). A full list of the financial, institutional and personal contributions to the development of the 1958 Birth Cohort Biomedical resource is available at http://www2.le.ac.uk/projects/birthcohort. Tampere Vascular Study (TVS) was supported by the European Union 7th Framework Programme funding for the AtheroRemo project [201668], the Finnish Foundation of Cardiovascular Research (TL), the Finnish Cultural Foundation, the Tampere Tuberculosis Foundation, the Emil Aaltonen Foundation (NO & TL), the Tampere University Hospital Medical Fund (grant 9M048 and 9N035) and the the Foundation of Clinical Chemistry. Immunochip case data for UK, Polish, Belgian, German and Swedish cohorts was genotyped by the Sanger Centre, Cambridge, UK as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA), as were the UK and Polish control cohorts. German Control Genotypes was provided through the POPGEN Consortium. Swedish Immunochip Control Samples wereprovided by the Swedish SLE network and the Uppsala Bioresource. Belgian Immunochip Control Sample Genotypes were provided through the efforts of the International Multiple Sclerosis Genetics Consortium (IMSGC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* Email: mtraylor@sgul.ac.uk

◔ These authors contributed equally to this work

¶ Members of Wellcome Trust Case Consortium 2 are listed in the Text S1.

## Introduction

Genome-wide association studies (GWAS) in ischaemic stroke have begun to identify the common genetic variants that confer risk of the disease. However, there is considerable heterogeneity present in stroke phenotypes: GWAS analyses have primarily looked at the three main subtypes; cardioembolic (CE), large artery atherosclerosis (LAA) and small vessel disease stroke (SVD). Within these subtype analyses, numbers of cases are smaller, but the expectation is that the effects of SNPs identified within the subtypes will be considerably larger. Indeed, all validated GWAS SNPs for ischaemic stroke to date have been stroke subtype-specific [1,2,3,4,5], indicating the importance of subtyping of cases.

Clinical risk factors are important in stroke; as many as 77% of first-ever stroke patients are hypertensive [6], and other factors such as diabetes mellitus and elevated serum cholesterol confer a considerable proportion of disease risk [7]. These risk factors increase in prevalence in older age groups, suggesting older stroke patients may have a reduced stroke-specific genetic contribution.

## Author Summary

Ischaemic stroke places an enormous burden on global healthcare. However, the disease processes that lead to stroke are not fully understood. Genome-wide association studies have recently established that common genetic variants can increase risk of ischaemic stroke and its subtypes. In this study, we aimed to identify novel genetic associations with ischaemic stroke and its subtypes by addressing the fact that younger onset cases may have a stronger genetic component, and using this information in our analyses. We identify a novel genetic variant on chromosome 11 (rs660599), which is associated with increased risk of large artery stroke. We also show that mRNA expression of the nearest gene (*MMP12*) is higher in arteries with the disease process underlying large artery stroke (atherosclerosis). Finally, we evaluate our novel analysis approach, and show that our method is likely to identify further associations with ischaemic stroke.


Indeed, IS is uncommon in individuals below middle age, but increases greatly in prevalence beyond the age of 65 [8], with a lifetime risk of 1 in 5 for women and 1 in 6 for men [9].

Under the assumptions of the liability threshold model, the low prevalence of IS in younger age ranges suggests that individuals who do suffer strokes in this age group are likely to have an increased genetic predisposition. This is supported by family history data; with stronger family history seen in younger onset cases [10,11,12], and twin studies [13], which suggest that early onset cases may have higher heritability. We recently showed stronger effects for all stroke-associated SNPs in younger age groups, found evidence genome-wide that a significant number of SNPs show stronger association p-values when the oldest cases are removed, and showed increased pseudoheritability estimates for younger onset cases in certain stroke subtypes, thereby supporting this hypothesis [14]. However, the question of how best to integrate this information into GWAS analyses of ischaemic stroke remains unanswered. Previous GWAS have analysed younger subsets of ischaemic stroke cases [1,15], but this approach may not be optimal for existing GWAS datasets if the increase in odds ratios for SNPs in younger cases are not sufficient to justify discarding a large proportion of the ascertained cases. All previous young onset analyses have been restricted to all ischaemic stroke cases versus controls; this may be particularly relevant given that all known loci for ischaemic stroke to date are for stroke subtypes [16].

A recent publication [17], outlined a novel method of informing genetic association analyses on important clinical covariates. Using the liability threshold model in conjunction with estimates of disease prevalence for individuals with specific clinical covariates, the method estimates posterior disease liabilities for each individual in a GWAS, and uses these liabilities in regression analyses to test for association with genome-wide SNPs. This approach avoids issues due to multiple testing across age-at-onset thresholds, and provides a simple solution that is rooted is previous epidemiological research. In the present study, we extend the clinical covariate informed analysis approach to imputed genotypes, informing our analyses on the age-at-onset to identify novel variants associated with IS. We perform a genome-wide analysis with four stroke phenotypes (IS, CE, LAA, SVD), and then determine the utility of the approach in ischaemic stroke GWAS, testing whether SNPs increase in significance.

## Results

### Association analysis

We performed age-at-onset informed association analysis for a total of 6,778 ischaemic stroke cases and 12,095 controls across four ischaemic stroke phenotypes; all IS and the three major subtypes: CE, LAA, and SVD (Table 1); with 1,637, 1,316, and 1,108 cases in the CE, LAA and SVD analyses respectively. With the exception of the young Milanese cohort, the age-at-onset distributions were similar in all cohorts (Table S3).

We identified a group of twenty SNPs proximal to *MMP3* and *MMP12* on chromosome 11 in the LAA subtype that met our criteria for replication. The strongest associated of these was rs662558 ($p = 1.4\times10^{-7}$), a SNP that is in 1000 Genomes, but not HapMap II. Therefore, to enable replication in existing METASTROKE datasets, which were imputed to HapMap II, we selected the most strongly associated SNP from the HapMap II panel, which was in perfect LD with the lead SNP in our discovery meta-analysis (rs660599: uninformed, $p = 1.6\times10^{-6}$; informed, $p = 2.5\times10^{-7}$; Figure 1) [16]. We found no evidence of between-study heterogeneity at either SNP (Cochran's Q $p = 0.22$ and $p = 0.19$ for rs662558 and rs660599, respectively). The evidence of an age-at-onset effect at rs660599 was $p = 0.011$ (from permutations). We calculated age-at-onset quartiles for all large artery stroke cases from the discovery cohorts, and used these to evaluate this region at different age-at-onset thresholds. The median age-at-onset was 71 years, and the interquartile range was between 61 and 78 years. Post-hoc analyses of rs660599 in the discovery cohorts using logistic regression (full details in Text S2) showed considerably stronger associations in younger age-at-onset quantiles (Q1; OR(95% CI) = 1.83 (1.46–2.30), Q1–Q2; 1.56 (1.33–1.83), Q1–Q3; 1.30 (1.14–1.49), Q1–Q4; 1.30 (1.15–1.46)). No other regions met our criteria for replication.

### Replication analysis

The associated locus was evaluated in a further 1,881 large artery stroke cases and ancestry matched controls in 9 cohorts from METASTROKE (Table 2). We found evidence for replication of the SNP (rs660599) in all large artery stroke cases of European Ancestry ($p = 0.0048$, OR(95% CI) = 1.18(1.05–1.32)). Combining this result with the discovery p-value gave a genome-wide significant p-value of $2.6\times10^{-8}$ (Table 3). Secondly, we used the Han and Eskin random effects meta-analysis approach to evaluate the association [18] after including a further 355 cases and 1,390 controls of Pakistani ancestry. The evidence for replication in this sample was $p = 0.0063$, giving an overall p-value of $3.4\times10^{-8}$. Age-at-onset information was available across all age-at-onset quantiles for a subset of the replication studies (1,240 cases, 9,238 controls; ASGC, HVH, ISGS/SWISS, MGH-GASROS, Utrecht). We evaluated the SNP (rs660599) in these studies at different age-at-onset quantiles using logistic regression, meta-analysing as previously. We again found the strongest effects in the youngest age quantile, consistent with a stronger effect in younger onset cases (Q1; OR(95% CI) = 1.27(1.02–1.57), Q1–Q2; 1.18(1.00–1.39), Q1–Q3; 1.22(1.05–1.40), Q1–Q4; 1.22(1.07–1.41)).

### mRNA expression in carotid plaques

mRNA expression of the two proximal genes, *MMP3* and *MMP12* was analysed from 29 carotid, 15 abdominal aorta, 24 femoral plaques, and 28 atherosclerosis free left internal thoracic artery controls. *MMP12* expression was upregulated in carotid plaques compared with left internal thoracic artery controls ($P = 1.2\times10^{-15}$; fold change [FC] = 335.6). It was also upregulated in femoral plaques ($P = 3.2\times10^{-14}$; FC = 306.0) and abdominal

**Table 1.** Sample size of discovery populations.

| Study Population | IS | CE | LAA | SVD | Controls |
|---|---|---|---|---|---|
| Belgium – Immunochip | 396 | 147 | 57 | 49 | 319 |
| Germany-Immunochip | 421 | 127 | 101 | - | 2,355 |
| Krakow – Immunochip | 384 | 119 | 33 | 28 | 255 |
| Sweden – Immunochip | 796 | 246 | 56 | 183 | 997 |
| UK – Immunochip | 867 | 130 | 152 | 257 | 1,790 |
| Germany – WTCCC2 | 1,174 | 330 | 346 | 106 | 797 |
| UK – WTCCC2 | 2,374 | 474 | 498 | 460 | 5,175 |
| Milano | 366 | 64 | 73 | 25 | 407 |
| Total (Discovery) | 6,778 | 1,637 | 1,316 | 1,108 | 12,095 |

IS, all ischaemic stroke; CE, cardioembolic stroke; LAA, large artery stroke; SVD, small vessel disease.
doi:10.1371/journal.pgen.1004469.t001

plaques ($P = 5.0 \times 10^{-11}$; FC = 399.3) compared with controls. Conversely, *MMP3* was not significantly overexpressed in carotid, femoral or abdominal plaques versus controls ($p > 0.05$).

## Regulatory information from ENCODE

Eight SNPs were identified that were perfect proxies ($r^2 = 1$) with the associated SNP (rs660599) in the region. Seven of the SNPs were in an intergenic region between *MMP3* and *MMP12*, while one fell within an intron of *MMP12*. We investigated the evidence that any of these SNPs are functional variants using RegulomeDB [19]. Of the eight SNPs, we found strong evidence that one of these SNPs (rs586701) affects binding. The SNP overlaps both CHIP-seq and DNA-seq peaks from ENCODE analyses, indicating that there is open chromatin in the region, and therefore that the SNP is likely to be functional. There is also evidence from a separate CHIP-seq analysis that the SNP affects protein binding [20], and evidence from multiple sources that the SNP overlaps a predicted motif [21,22,23]. Histone modifications were observed in CHIP-seq experiments from ENCODE in a number of cells types, including Human umbilical vein endothelial (Huvec) cells. Two other SNPs (rs17368582, rs2276109) in moderate LD with the associated SNP ($r^2 = 0.64$) have been previously shown to directly influence *MMP12* expression by affecting the affinity of an AP-1 binding site in the *MMP12* promoter region [24,25]. Using RegulomeDB, we found further evidence from ENCODE that one of these SNPs (rs2276109) is indeed functional, giving evidence that the associated locus in this analysis is likely to affect *MMP12* expression through altered

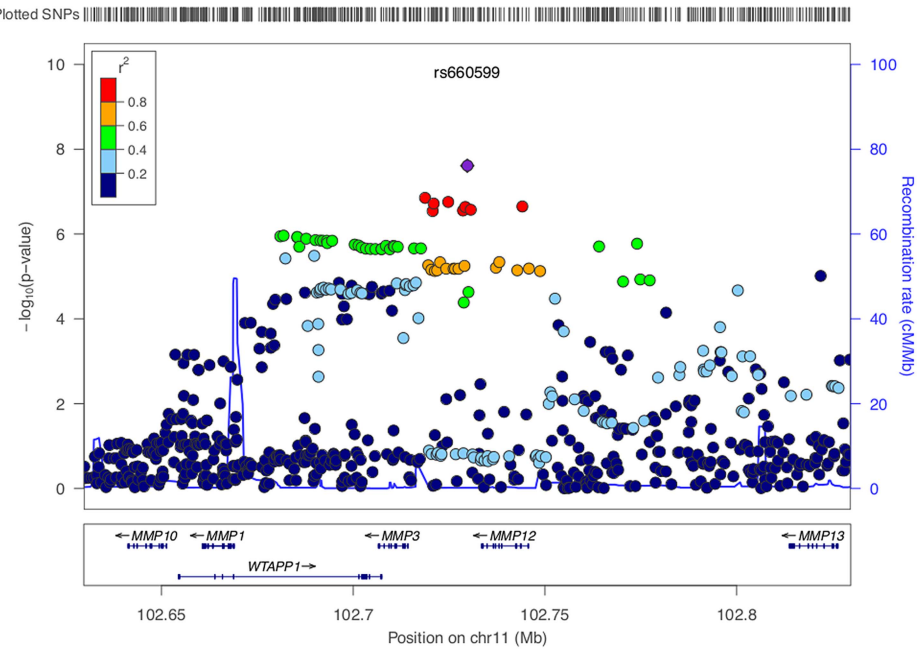



**Figure 1. LocusZoom plot of *MMP12* association using age-at-onset informed approach.** SNPs are colored based on their correlation ($r^2$) with the labeled top SNP, which has the smallest P value in the region. The fine-scale recombination rates estimated from 1000 Genomes (EUR) data are marked in light blue, with genes marked below by horizontal blue lines. Arrows on the horizontal blue lines show the direction of transcription, and rectangles are exons. SNP p-values are from the discovery meta-analysis only with the exception of rs660599, for which the given p-value indicates the overall evidence for association from the discovery and replication cohorts.
doi:10.1371/journal.pgen.1004469.g001

**Table 2.** Sample size of replication populations.

| Study Population | LAA (age<61) | LAA (age<71) | LAA (age<78) | LAA (all ages) | IS | Controls |
|---|---|---|---|---|---|---|
| ARIC | - | - | - | 31 | 385 | 8,803 |
| ASGC | 81 | 179 | 277 | 421 | 1,162 | 1,244 |
| deCODE | - | - | - | 255 | 2,391 | 26,970 |
| GEOS | - | - | - | 37 | 448 | 498 |
| HVH | 18 | 39 | 63 | 71 | 566 | 2,072 |
| ISGS/SWISS | 84 | 130 | 179 | 217 | 1,070 | 1,370 |
| MGH-GASROS (Affymetrix) | 31 | 60 | 79 | 102 | 485 | 3,030 |
| MGH-GASROS (Illumina) | 22 | 47 | 59 | 68 | 296 | 377 |
| PROMISe | 134 | 230 | 301 | 324 | 556 | 1,145 |
| RACE | - | - | - | 355 | 1,390 | 5,308 |
| Total (Replication) | 370 | 685 | 958 | 1,881 | 8,749 | 50,817 |

LAA, large artery stroke; IS, all ischaemic stroke; ARIC, the Atherosclerosis Risk in communities study; ASGC, the Australian Stroke Genetics collaboration; deCODE, deCODE genetics; GEOS, the Genetics of early onset stroke study; HVH, the heart and vascular health study; ISGS/SWISS, the Ischaemic stroke genetics study/Siblings with Ischaemic stroke study; MGH-GASROS, Massachusetts General Hospital – Genetics affecting stroke risk and outcome; PROMISe, Prognostic modeling in ischaemic stroke study [55]; RACE, Risk Assessment of Cerebrovascular Events study. For further details of these populations please see the original METASTROKE publication [16].
doi:10.1371/journal.pgen.1004469.t002

transcription. Detailed results for all analysed SNPs are given in Table S1. Additionally, we investigated if these SNPs (rs17368582, rs2276109, rs586701) were associated with *MMP12* expression in tissues from the GTEx project [26]. However, we could not confirm an association with *MMP12* expression in any relevant tissues ($p>0.4$ in whole blood, tibial artery, aortic artery).

## Evaluation of age-at-onset informed approach

Finally, we evaluated the overall utility of the age-at-onset informed approach in permutation analyses for SNPs that met p-value thresholds in the case control discovery data set. We generated 1000 permutations of age-at-onset within each centre, and performed age-at-onset informed analysis and subsequent meta-analysis for these SNPs, in the relevant stroke subtype.

We compared the sum of the meta-analysis Z scores from all SNPs with $p<0.05$ in the observed age at onset informed meta-analysis with those from permutations. At this p-value selection threshold, we found strong evidence ($p<0.001$) for genome-wide age-at-onset effects in each of the stroke phenotypes, with consistently increased summed Z scores in the observed age-at-onset informed meta-analysis compared to the permutations (Figure 2, red points, right hand axis). These results suggest that many of the risk variants for each stroke subphenotype have a higher frequency in younger onset cases. As the p-value selection threshold decreased, the summed Z score statistic became less significant in each stroke type, possibly reflecting lower overall power when fewer SNPs are included, even as these SNPs may have larger average effects. Further details are seen from the median proportion of SNPs more significant in the age-at-onset informed analysis than in the permutations (Figure 2, blue points, left hand axis). For CE and LAA stroke, the proportions increased with more stringent p-value thresholds (from 52.1% to 56.3% for $p<0.05$ and $p<0.00005$ thresholds in CE, and from 51.4% to 56.0% for $p<0.05$ and $p<0.00005$ thresholds in LAA). Interestingly, in the all ischaemic stroke analysis the median proportion of SNPs more significant in the observed results than permutations dropped from 55.1% for SNPs with $p<0.05$ to 49.2% for only SNPs with $p<0.00005$. This result may indicate a reduced proportion of true associations at stricter p-value thresholds for all ischaemic stroke compared to the subtypes, which is consistent with the observation that all common variants associated with stroke are for stroke subtypes, rather than for the phenotype of all ischaemic stroke [16].

The previously reported GWAS associations from a recent ischaemic stroke meta-analysis (9p21, *HDAC9*, *PITX2*, *ZFHX3*) were all found to be more significant using the age-at-onset informed approach than the uninformed analysis (Figure 3). The increase in significance ranged from over half an order of

**Table 3.** Evidence for association of A allele of rs660599 (chromosome 11; Base position 102,234,967) with large artery atherosclerotic stroke and all ischaemic stroke.

| Subtype | SNP | RAF | p-value (discovery) | OR (95% CI) (EUR replication) | p-value (EUR replication, overall) | p-value (ALL replication, overall) |
|---|---|---|---|---|---|---|
| LAA | rs660599 | 0.19 | $2.5.\times10^{-7}$ | 1.18 (1.05–1.32) | 0.0048, $2.6\times10^{-8}$ | 0.0063, $3.4\times10^{-8}$ |
| IS | " | " | $3.2\times10^{-4}$ | 1.05 (1.00–1.11) | 0.050, $1.9\times10^{-4}$ | 0.098, $3.6\times10^{-4}$ |
| CE | " | " | 0.13 | - | - | - |
| SVD | " | " | 0.30 | - | - | - |

LAA, large artery stroke; IS, all ischaemic stroke; SNP, single nucleotide polymorphism; RAF, risk allele frequency; OR, odds ratio; 95% CI, 95% confidence interval; EUR, meta-analysis in individuals of European ancestry alone; ALL, trans-ethnic meta-analysis of all individuals. Forest plots of effect sizes and standard errors for each replication centre are given in Figures S3, S4.
doi:10.1371/journal.pgen.1004469.t003

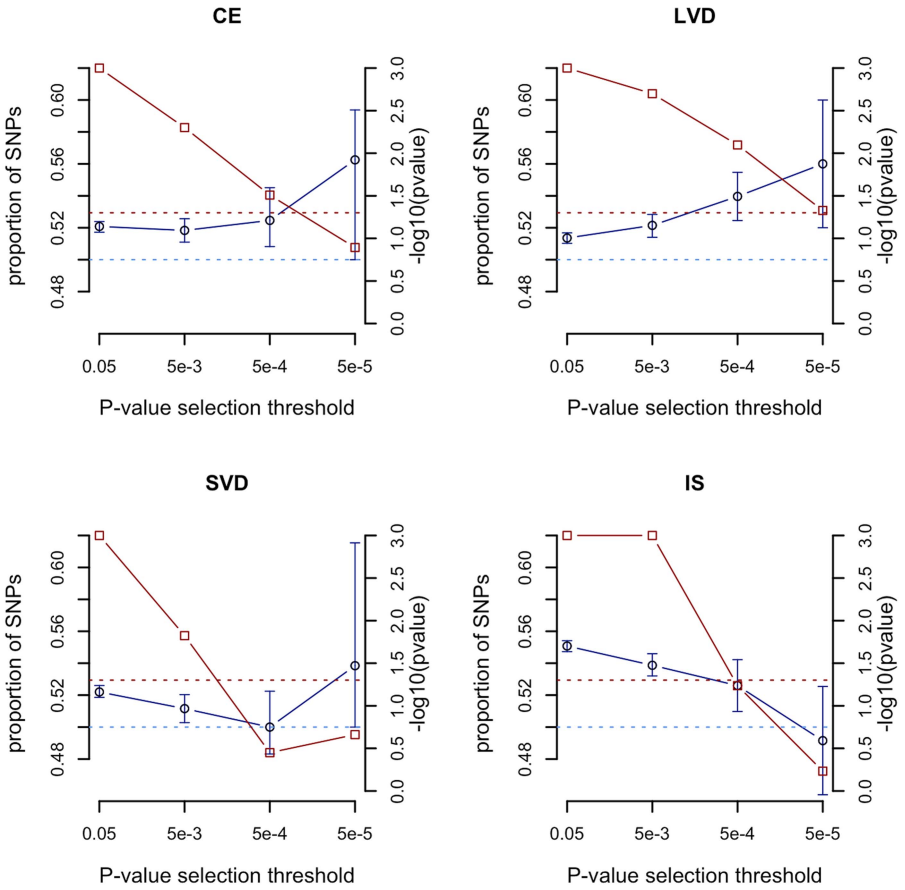



**Figure 2. Evaluation of evidence genome-wide for SNPs exhibiting greater significance using the age-at-onset informed approach compared to permutations.** -log10(p value) from permutations for evidence of age-at-onset effect at given SNP p-value selection threshold shown in red; median proportion of SNPs (with IQR) more significant in observed age-at-onset informed meta-analysis compared to permutations shown in blue; horizontal line at p = 0.05 in red; horizontal line at median proportion of SNP = 0.5 in blue; IS, all ischaemic stroke; CE, cardioembolic stroke; LAA, large artery atherosclerotic stroke; SVD, small vessel disease. See Table S5 for number of SNPs included at each p-value selection threshold.
doi:10.1371/journal.pgen.1004469.g002

magnitude ($7.9\times10^{-9}$ to $1.5\times10^{-9}$ for rs879324 in *ZFHX3*, CE), to under half an order of magnitude ($5.7\times10^{-9}$ to $2.5\times10^{-9}$ for rs2107595 in *HDAC9*, LVD). To ensure these analysis methods were comparable, we calculated genomic inflation factors and plotted QQ-plots. These were similar in the standard and the age-at-onset informed approach (Table S4, Figure S1, S2). For these four associated SNPs, we further used the permuted data sets to assess the observation of increased significance in the age-at-onset informed analysis. We compared the observed meta-analysis p-value to those from the permutations, generating an empirical p-value by dividing the number of permutations more significant than the observed results by the number of permutations. In LAA stroke, we observed a significant age-at-onset effect (p = 0.018, 0.011 and 0.002 for the *HDAC9*, *MMP12* and 9p21-associated SNPs in Figure 3, respectively). Similarly, for CE, we observed a significant age-at-onset effect for rs879324 (*ZFHX3*, p = 0.026), and a near-significant effect in rs6843082 (*PITX2*, p = 0.081). This result provides further evidence that risk variants associated with ischaemic stroke subtypes have a stronger role in younger onset cases, and suggests that the age-at-onset informed approach will produce improved significance when the magnitude of genetic effects are stronger in younger onset cases.

## Discussion

We used a large GWAS dataset to evaluate the utility of an age-at-onset informed analysis approach to ischaemic stroke, and to identify novel variants associated with ischaemic stroke phenotypes. We identified a novel *MMP12* locus that is associated with large artery atherosclerotic stroke, and verified that the age-at-onset informed approach produces improved significance for loci associated with each of the stroke phenotypes studied, as well as demonstrating that it increased the significance of four previous GWAS associations with ischemic stroke, all without systematic inflation of the test statistic. Importantly, the novel associated SNP would not have been identified using a standard logistic regression framework.

We identified a group of SNPs proximal to Matrix Metalloproteinase 12 (*MMP12*) that showed increased significance when using the age-at-onset informed approach. The increase in significance from the equivalent uninformed analysis was of almost an order of magnitude (from $p = 1.6\times10^{-6}$ to $p = 2.5\times10^{-7}$ for rs660599). We took a single SNP from this region forward for replication in an independent dataset, finding further evidence that the region is associated with large artery stroke. Two SNPs (rs17368582, rs2276109) in this LD-block have previously been shown to directly influence *MMP12* expression by

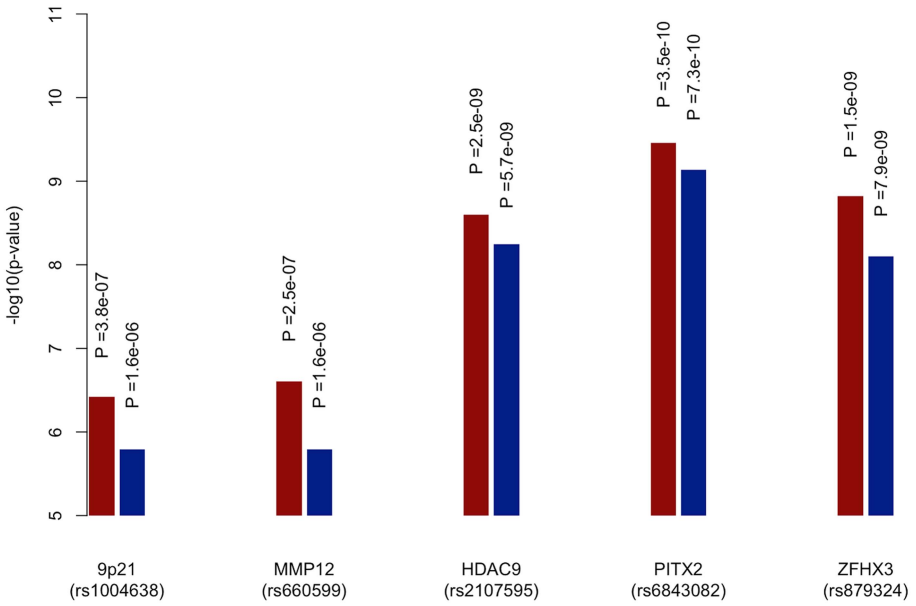



**Figure 3. Meta-analysis p-values of known loci for ischaemic stroke subtypes using age-at-onset informed approach compared to uninformed approach.** -log10 of p-values derived from meta-analysis of all discovery cohorts using age-at-onset informed approach (red) and uninformed approach (blue). 9p21 (rs1004638), *MMP12* (rs660599) and *HDAC9* (rs2107595) p-values calculated within large artery atherosclerosis subtype of stroke, *PITX2* (rs6843082) and *ZFHX3* (rs879324) p-values calculated with cardioembolic stroke subtype.
doi:10.1371/journal.pgen.1004469.g003

affecting the affinity of an AP-1 binding site in the *MMP12* promoter region [24,25], and another variant in this block (rs17361668) is associated with increased fibrinogen levels, leading to an increased risk of developing advanced carotid atherosclerotic lesions, and an increased risk of myocardial infarction. We identified a second functional candidate (rs586701), which falls within both CHIP-seq and DNA-seq peaks from ENCODE, and is in complete LD with the associated SNP in our analysis.

We investigated mRNA expression of *MMP12* and *MMP3* in carotid atherosclerotic plaques in individuals from the Tampere Vascular Study. *MMP12* was overexpressed in diseased tissue compared to healthy controls, while no significant difference was found for the other nearby gene, *MMP3*. *MMP12* is a member of the Matrix Metalloproteinase (MMP) family of proteases, which are capable of degrading extracellular matrix proteins, and have a prominent role in atherosclerosis. They are thought to promote macrophage invasion [27,28,29], promote angiogenesis [30], and show increased activity in atheromatous plaques [31]. *MMP12* deletions are associated with smaller, more stable lesions in the brachiocephalic artery of rabbits [32], and reduced elastin degradation in the aortic arch [33], indicating that *MMP12* may have a role in destabilising plaques. Studies in humans have found *MMP12* is localized to the core of advanced plaques, in macrophages with decreased arginase-I expression [34], that *MMP12* localizes selectively to macrophages at the borders of the lipid core [35], and that *MMP12* is significantly overexpressed in ruptured plaques when compared with thick or thin cap plaques, or with plaques with pathological intimal thickening [36]. This indicates that *MMP12* is likely be involved in late-stage plaque instability: our study suggests that genetic variation impacts on this process.

Secondly, we performed extensive permutation analyses to assess the utility of the age-at-onset informed approach genome-wide. In each phenotype studied we found evidence that SNPs were more strongly associated using the approach than would be expected by chance, indicating that multiple risk variants are likely to be more common in younger onset cases. The significance was strongest when more SNPs were included in the analysis, which likely reflects the cumulative impact of age-at-onset effects on many SNPs. An alternative explanation might be that the increased significance for lower p-value thresholds is the result of the cumulative effects of subtle confounding. However, this is unlikely because any subtle biases will also be present in the permutations, and should therefore not affect the significance of the results. This result supports observations from family history and prospective cohort studies, which have observed stronger effects in younger onset cases [6,11]. Furthermore, all known associations with stroke were more significant using the age-at-onset informed approach. The increase in significance was around half an order of magnitude (e.g from $p = 7.9\times10^{-9}$ to $1.5\times10^{-9}$ for *ZFHX3*, Figure 2), and was significant in all but one locus, as assessed by permutation. Taken together, these results indicate that age-at-onset is an important measure to stratify stroke cases, and show that, as expected by theory [17], integrating this information into association studies is likely to increase power to identify novel loci when the relative contribution of genetic is dependent on age-at-onset.

Our study has limitations. We used imputed data from the Immunochip platform, meaning we only had access to ~40% of the genome across all centres. Secondly, cases were drawn from a number of international centres, meaning that despite efforts to standardize phenotyping, we cannot rule out differences in screening and clinical ascertainment.

Of complex diseases, IS has a particularly large degree of heterogeneity, exemplified by the fact that all validated associations identified to date have been within subtypes defined by clinical and radiological information. Further heterogeneity by risk factor and clinical covariate profiles is likely to exist, but the optimal method of incorporating this information into analyses remains an unanswered question. Our results indicate that a

covariate-informed design, conditioning on age-at-onset of stroke, can unearth further associated variants. We provide evidence for this by identifying an association with a novel *MMP12* locus in large artery stroke, supported by increased mRNA expression of the implicated gene in carotid plaques. GWAS in ischaemic stroke have begun to identify the genetic component of the disease, but these results are not yet clinically useful. Our study suggests that a more refined approach to analysis of genetic data, incorporating covariate information, is an important step in this process, and will help to ensure success in future GWAS.

## Materials and Methods

### Ethics statement

All studies were approved by their local ethics committees; all patients gave informed consent.

### Description of datasets

The initial dataset consisted of 6,778 ischaemic stroke cases of European ancestry and 12,095 ancestry-matched controls from the Wellcome Trust Case-Control Consortium II project in ischaemic stroke [1], as well as a cohort from Milan, Italy [16]. These included 2,858 cases and 5,716 matched controls genotyped using the Immunochip platform; and 3,940 cases genotyped using either the Illumina 610 k or 660 k platforms matched with 6,379 controls genotyped on the Illumina Human 1.2M Duo (UK), Illumina Human 550 k (German) and Illumina 610 k platforms (Italian) (Table 1). The Immunochip cases were described in the previous WTCCC2 ischaemic study, where they formed the replication effort [1], as well as in a recent paper [37]. Genotyping of the five Immunochip case cohorts on the commercially available Immunochip array (Illumina, San Diego, CA, USA) was performed at the Sanger Centre, Hinxton, Cambridge UK. Swedish controls were provided and genotyped by the Swedish SLE network, Uppsala, Sweden. Belgian control samples were provided through the efforts of the International Multiple Sclerosis Genetics Consortium (IMSGC). German controls were derived from the PopGen biobank, [38]. UK controls were derived from the 1958 Birth cohort. Any of the 1958 Birth controls overlapping with those from the WTCCC2 datasets, as assessed by IBD estimates, were removed prior to analysis. Standard quality control procedures were undertaken on all centres, before centre-wise imputation to the 1000 Genomes phase 1 integrated variant set (March 2012), using IMPUTE v2.2.0 [39,40]. SNPs with poor imputation quality (info$<$0.3) or low minor allele frequency (MAF$<$0.01) were discarded.

Ischemic stroke was defined as a typical clinical syndrome with radiological confirmation; ascertained cases were classified into individual stroke subtypes using the Trial of Org 10172 in acute stroke (TOAST) criteria in all centres [41]. Age-at-onset was defined as age at first hospital admission for stroke; where this information was unavailable, age at blood draw was used (7.3% of cases). The age-at-onset and gender distributions of the populations are given in Table S3. Age-at-onset quantiles were calculated from all the cases from the discovery datasets in the four stroke phenotypes (all IS and the three stroke subtypes: CE, LAA, SVD) and these were used to evaluate associated loci at different age-at-onset thresholds.

### Association analysis

The prevalence of ischaemic stroke by age was obtained from a recent publication [9]; gender-specific estimates were averaged, and prevalences within each of the stroke subtypes were assumed to be approximately 20% of the overall total, similar to proportions seen in population-based studies [42]. We modeled phenotype data using a continuous unobserved quantitative trait called the disease liability, which we used to approximate the effect of age-at-onset on the liability scale, based on estimates of ischaemic stroke prevalence by age from epidemiological data (full details in Text S2). We developed two models for our analysis; one based on the prevalence rates for all ischaemic stroke cases, and secondly for the three stroke subtypes. We used these models to calculate posterior mean liabilities after conditioning on age-at-onset for the four stroke phenotypes separately. Controls were modeled in the same way, but were assumed to take the posterior mean from the lower (unaffected) portion of the distribution in the liability threshold model. Where age data was missing, individuals were assigned the median age value. Full descriptions of the models used and the formulae used to calculate posterior mean liabilities are given in Text S2. Regression was then performed on posterior liabilities by multiplying the number of samples by the squared correlation between the expected genotype dosage and posterior mean liabilities for each of the discovery cohorts in the four ischaemic stroke phenotypes (CE, LAA, SVD, IS), following a previous approach [17]. Ancestry-informative principal components were included where appropriate (6 of 8 centres), using the EIGENSTRAT procedure [43]. All analysis was performed using the R statistical software.

The results from each centre were meta-analysed for each of the four phenotypes using Stouffer's Z-score weighted approach, as implemented in METAL [44]. Genomic control was used to correct for any residual inflation due to population stratification [45]. Between-study heterogeneity was assessed using Cochran's Q statistic. We considered only SNPs present in at least 75% of the cases, and with no evidence of heterogeneity (Cochran's Q p-value$>$0.001). All SNPs analysed were either genotyped or imputed in both the Immunochip and the genome-wide datasets. After meta-analysis, the resulting p-values were compared with the equivalent values from an unconditioned analysis. For SNPs more significant in the age-at-onset informed analysis and with $p< 5\times10^{-6}$, we determined the evidence of a true age-at-onset effect by generating 1000 permutations of age-at-onset and rerunning the age-at-onset informed analysis, meta-analysing as previously. We calculated an empirical p-value by dividing the number of permuted observations showing greater significance in the meta-analysis than the observed results by the number of permutations. Any novel SNP with a meta-analysis $p<5\times10^{-6}$ and evidence of an age-at-onset effect at $p<0.05$ were taken forward for replication. We set the experiment-wide significance threshold at $p<5\times10^{-8}$.

### Replication analysis

Replication of an associated variant was performed in a further 10 cohorts from METASTROKE. Nine of the centres used a cross-sectional design, while one was a large prospective, population based cohort (ARIC). Nine of the centres were of European ancestry, while one consisted of individuals of Pakistani ancestry (RACE) (Table 2). All centres used a case-control methodology; centres with a cross sectional design used logistic regression to model the association of genotype dosages from imputation with the dichotomous outcome of ischaemic stroke and prospective cohorts used Cox proportional-hazards models to evaluate time to first stroke, fitting an additive model relating genotype dose to the stroke outcome. European ancestry replication centres were meta-analysed using a fixed effects inverse-variance weighted method. To assess the evidence for association of the SNP for replication samples of all ancestries, we

performed a trans-ethnic meta-analysis using a random-effects model to control for any resulting heterogeneity [18]. To evaluate the overall evidence for association, the results of the discovery and replication analyses were combined using Fisher's Method.

### mRNA expression in carotid atherosclerotic plaques

Expression of the two genes proximal to the associated variant was tested in atherosclerotic plaques from the Tampere Vascular study [27,46,47,48,49]. Carotid, femoral, and aortic atherosclerotic plaques constituting the intima and inner media were prospectively obtained between 2005 and 2009 from patients fulfilling the following inclusion criteria: (1) carotid endarterectomy attributable to asymptomatic or symptomatic >70% carotid stenosis, or (2) femoral or (3) aortic endarterectomy with aortoiliac or aortobifemoral bypass attributable to symptomatic peripheral arterial disease. Whole thickness left internal thoracic artery samples obtained during coronary artery bypass surgery and identified as being microscopically atherosclerosis free were used as controls. The patients were consecutively recruited and stratified according to indication for surgery. All open vascular surgical procedures were performed at the Division of Vascular Surgery and Heart Center, Tampere University Hospital.

Fresh tissue samples were immediately soaked in RNALater solution (Ambion Inc) and homogenized using an Ultra-Turrax T80 homogenizer (IKA). RNA was extracted with the Trizol reagent (Invitrogen) and miRNEasy Mini-Kit (Qiagen) with the RNase-Free DNase Set (Qiagen) according to manufacturer instructions. The RNA isolation protocol was validated by analyzing the integrity of the RNA with the RNA 6000 Nano Chip Kit (Agilent). The expression levels were analyzed with an Illumina HumanHT-12 v3 Expression BeadChip (Illumina). In brief, 300–500 ng of RNA was reverse transcribed in cRNA and biotin-UTP labeled using the IlluminaTotalPrep RNA Amplification Kit (Ambion), and 1500 ng of cRNA was then hybridized to the Illumina HumanHT-12 v3 Expression BeadChip.

The BeadChips were scanned with the Illumina iScan system. After background subtraction, raw intensity data were exported using the Illumina Genome Studio software. Further data processing was conducted by means of R language and appropriate Bioconductor modules. Data were log2-transformed, and robust multichip average and robust spline normalization (rma_rsn) were used. Accuracy of the expression array was validated with qRT-PCR [50]. mRNA Expression levels in the tissues were determined; a fold change statistic was estimated between the two tissues, and significance was calculated using a t test.

### Regulatory information using RegulomeDB

Recent evidence indicates that a significant proportion of GWAS SNPs fall within regions that are likely to affect binding of nearby proteins, such as transcription factor binding sites [51,52]. We used the RegulomeDB database to access regulatory information from ENCODE and other existing publications [19], investigating the evidence that the SNPs in the associated locus have a regulatory function. First, the linkage-disequilibrium (LD) patterns amongst the most strongly associated SNPs were determined. We then used PLINK to determine the LD structure of the associated region, using LD-patterns from the 85 Utah residents from the 1000 Genomes project [53,54]. All SNPs with $r^2>0.6$ were identified within a 2,000 kb window from the index SNP. All of the SNPs identified were then investigated using RegulomeDB to determine the evidence that any of the SNPs have a regulatory function.

### Evaluation of age-at-onset informed approach

Permutation analysis was performed to evaluate the age-at-onset informed approach, to show that including age at onset information directly led to the increased significance, due solely to inclusion of age-at-onset information at tested SNPs. First, we identified a set of SNPs enriched for true association in the case control analysis of ischaemic stroke and subtypes. An expanded set of discovery and METASTROKE studies were analysed using standard case control methods and subsequent meta-analysis (see Table S2). SNPs with $p<0.05$ and no evidence of heterogeneity ($p>0.0001$) were extracted and pruned for LD (300 kb window, $r^2<0.25$), leaving a set of almost independent SNPs for further analysis. Each retained SNP represented the most significant association in each LD block, as determined by the ''clump'' procedure in PLINK, based on LD patterns from the CEU individuals from 1000 Genomes. The number of SNPs used in each analysis is given in Table S5. These SNP subsets were derived for ischaemic stroke, and for each stroke subset and then used in the age-at-onset informed analysis. Analysis was performed as previously for each stroke subtype using the age-at-onset informed method within studies and meta-analysis across studies (giving *observed* results, as obtained above). We then performed a permutation study to obtain the expected distribution of p-values at these SNPs. Age at onset for cases was permuted within stroke subtypes within each study, and then the data were re-analysed, for 1000 permutations. Two summary statistics were constructed: (1) within permutations, we compared p-values from analysis of permuted age at onset with p-values from the observed data, and tabulated the proportion of SNPs with increased significance in the observed data set than in the permuted data set; across permutations, we calculated the median proportion of SNPs with increased significance in the observed data; (2) Within permutations, we converted each SNP p-value to a Z score and summed the absolute value of the Z score across SNPs (sumZ). An empirical p-value for the age-informed analysis was calculated from the proportion of simulated data sets where sumZ exceeded the value in the observed analysis. This analysis was performed at SNP subsets defined from four SNP p-value thresholds in the discovery and METASTROKE studies: $p<0.05$, $p<0.005$, $p<0.0005$, and $p<0.00005$.

Finally, we assessed the evidence of an age-at-onset effect at the four stroke loci identified in the METASTROKE ischaemic stroke collaboration (9p21, *HDAC9*, *PITX2*, *ZFHX3*) [16]. For each SNP, we generated an empirical p-value from the proportion of permutations showing stronger association than in the observed age-at-onset informed analysis.

## Supporting Information

**Figure S1** QQ-plots for cardioembolic stroke and all ischaemic stroke analyses. QQ-plots of expected p-values (x-axis) against observed p-values (y-axis) for analyses of (clockwise from top left) cardioembolic stroke (age-at-onset informed), cardioembolic stroke (uninformed), all ischaemic stroke (uninformed), all ischaemic stroke (age-at-onset informed). Lambda values for each plot are given in Table S4.
(DOCX)

**Figure S2** QQ-plots for large artery atherosclerotic stroke and small vessel disease stroke analyses. QQ-plots of expected p-values (x-axis) against observed p-values (y-axis) for analyses of (clockwise from top left) large artery stroke (age-at-onset informed), large artery stroke (uninformed), small vessel stroke (uninformed), small

vessel stroke (age-at-onset informed). Lambda values for each plot are given in Table S4.
(DOCX)

**Figure S3** Forest plot of SNP effects for rs660599 in the large artery atherosclerotic stroke replication populations. ASGC, the Australian Stroke Genetics collaboration; deCODE, deCODE genetics; GEOS, the Genetics of early onset stroke study; HVH, the heart and vascular health study; ISGS/SWISS, the Ischaemic stroke genetics study/Siblings with Ischaemic stroke study; MGH-GASROS, Massachusetts General Hospital – Genetics affecting stroke risk and outcome. PROMISe, Prognostic modeling in ischaemic stroke study; RACE, Risk Assessment of Cerebrovascular Events study.
(DOCX)

**Figure S4** Forest plot of SNP effects for rs660599 in the large artery atherosclerotic stroke replication populations for cases with age <61 years. ASGC, the Australian Stroke Genetics collaboration; HVH, the heart and vascular health study; ISGS/SWISS, the Ischaemic stroke genetics study/Siblings with Ischaemic stroke study; MGH-GASROS, Massachusetts General Hospital – Genetics affecting stroke risk and outcome. PROMISe, Prognostic modeling in ischaemic stroke study.
(DOCX)

**Table S1** Results from RegulomeDB, showing the evidence that SNPs in the associated *MMP12* region have a regulatory function. Scores indicate the following degrees of evidence: Score 2b, TF binding + any motif + DNase Footprint + DNase peak; Score 4, TF binding + DNase peak; Score 5, TF binding or DNase peak; Score 6, other; "No data" indicates that RegulomeDB holds no information about the given SNP, meaning there currently exists no evidence to suggest that the SNP has a regulatory function. In some cases this may indicate that the SNP falls within a protein-coding region. SNP, single nucleotide polymorphism.
(DOCX)

**Table S2** Expanded set of populations used to generate SNPs with $p<0.05$ to evaluate the age-at-onset informed approach. ARIC, The Atherosclerosis Risk in Communities study; ASGC, Australian Stroke Genetics Collaborative; CHS, Cardiovascular Health Study; FHS, Framingham Heart Study; HPS, Heart Protection Study; HVH, The Heart and Vascular Health Study; ISGS/SWISS, The Ischemic Stroke Genetics Study/Sibling with Ischaemic Stroke Study; MGH-GASROS, The MGH Genes Affecting Stroke Risk and Outcome Study; WTCCC2-Germany, The Wellcome Trust Case-Consortium II Munich; WTCCC2-UK, The Wellcome Trust Case-Consortium II UK; RACE, Risk Assessment of Cerebrovascular Events Study, Pakistan.
(DOCX)

**Table S3** Age and gender distributions of populations. ARIC, The Atherosclerosis Risk in Communities study; ASGC, Australian Stroke Genetics Collaborative; CHS, Cardiovascular Health Study; FHS, Framingham Heart Study; HPS, Heart Protection Study; HVH, The Heart and Vascular Health Study; ISGS/SWISS, The Ischemic Stroke Genetics Study/Sibling with Ischaemic Stroke Study; MGH-GASROS, The MGH Genes Affecting Stroke Risk and Outcome Study; WTCCC2-Germany, The Wellcome Trust Case-Consortium II Munich; WTCCC2-UK, The Wellcome Trust Case-Consortium II UK; RACE, Risk Assessment of Cerebrovascular Events Study, Pakistan. IS, all ischaemic stroke; CE, cardioembolic stroke; LAA, large artery stroke; SVD, small vessel disease.
(DOCX)

**Table S4** Genomic inflation ($\lambda$) rates for discovery populations for age-at-onset informed and uninformed approaches. IS, all ischaemic stroke; CE, cardioembolic stroke; LAA, large artery stroke; SVD, small vessel disease.
(DOCX)

**Table S5** Number of SNPs used in evaluation of age-at-onset informed approach. IS, all ischaemic stroke; CE, cardioembolic stroke; LAA, large artery stroke; SVD, small vessel disease.
(DOCX)

**Text S1** Membership of Wellcome Trust Case Control Consortium 2 (WTCCC2).
(DOCX)

**Text S2** Liability threshold models.
(DOCX)

## Acknowledgments

The authors thank all study staff and participants for their important contributions, and METASTROKE for granting access to study data.

## Author Contributions

Conceived and designed the experiments: MT CML HSM. Performed the experiments: MT KMM ER NO EGH WJD MAN KLW WZ YCC SA. Analyzed the data: MT KMM. Contributed reagents/materials/analysis tools: LLK RM CS SB VT RL AL AS JMM MW AA PS JRA GBB PMR PIWdB JCB DS SJK BDM JR JFM CL MD TL. Wrote the paper: MT CML HSM.

## References

1. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, et al. (2012) Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. Nat Genet 44: 328–333.
2. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, et al. (2012) Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. Nat Genet 44: 1147–1151.
3. Gschwendtner A, Bevan S, Cole JW, Plourde A, Matarin M, et al. (2009) Sequence variants on chromosome 9p21.3 confer risk for atherosclerotic stroke. Ann Neurol 65: 531–539.
4. Gudbjartsson DF, Holm H, Gretarsdottir S, Thorleifsson G, Walters GB, et al. (2009) A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. Nat Genet 41: 876–878.
5. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, et al. (2008) Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. Ann Neurol 64: 402–409.
6. Seshadri S, Beiser A, Kelly-Hayes M, Kase CS, Au R, et al. (2006) The lifetime risk of stroke: estimates from the Framingham Study. Stroke 37: 345–350.
7. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, et al. (2010) Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet 376: 112–123.
8. Rothwell PM, Coull AJ, Silver LE, Fairhead JF, Giles MF, et al. (2005) Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). Lancet 366: 1773–1783.
9. Seshadri S, Wolf PA (2007) Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. Lancet Neurol 6: 1106–1114.
10. Seshadri S, Beiser A, Pikula A, Himali JJ, Kelly-Hayes M, et al. (2010) Parental occurrence of stroke and risk of stroke in their children: the Framingham study. Circulation 121: 1304–1312.
11. Jerrard-Dunne P, Cloud G, Hassan A, Markus HS (2003) Evaluating the genetic component of ischemic stroke subtypes: a family history study. Stroke 34: 1364–1369.
12. MacClellan LR, Mitchell BD, Cole JW, Wozniak MA, Stern BJ, et al. (2006) Familial aggregation of ischemic stroke in young women: the Stroke Prevention in Young Women Study. Genet Epidemiol 30: 602–608.

13. Brass LM, Isaacsohn JL, Merikangas KR, Robinette CD (1992) A study of twins and stroke. Stroke 23: 221–223.
14. Traylor M, Bevan S, Rothwell PM, Sudlow C, Dichgans M, et al. (2013) Using phenotypic heterogeneity to increase the power of genome-wide association studies: application to age at onset of ischaemic stroke subphenotypes. Genetic Epidemiology 37: 495–503.
15. Cheng YC, O'Connell JR, Cole JW, Stine OC, Dueker N, et al. (2011) Genome-wide association analysis of ischemic stroke in young adults. G3 (Bethesda) 1: 505–514.
16. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, et al. (2012) Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. Lancet Neurol 11: 951–962.
17. Zaitlen N, Lindstrom S, Pasaniuc B, Cornelis M, Genovese G, et al. (2012) Informed conditioning on clinical covariates increases power in case-control association studies. PLoS Genet 8: e1003032.
18. Han B, Eskin E (2011) Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet 88: 586–598.
19. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, et al. (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22: 1790–1797.
20. Verzi MP, Shin H, He HH, Sulahian R, Meyer CA, et al. (2010) Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev Cell 19: 713–726.
21. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. Science 324: 1720–1723.
22. Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, et al. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 36: D102–106.
23. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, et al. (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res 21: 447–455.
24. Jormsjo S, Ye S, Moritz J, Walter DH, Dimmeler S, et al. (2000) Allele-specific regulation of matrix metalloproteinase-12 gene activity is associated with coronary artery luminal dimensions in diabetic patients with manifest coronary artery disease. Circ Res 86: 998–1003.
25. Motterle A, Xiao Q, Kiechl S, Pender SL, Morris GE, et al. (2012) Influence of matrix metalloproteinase-12 on fibrinogen level. Atherosclerosis 220: 351–354.
26. (2013) The Genotype-Tissue Expression (GTEx) project. Nature Genetics 45: 580–585.
27. Oksala N, Levula M, Airla N, Pelto-Huikko M, Ortiz RM, et al. (2009) ADAM-9, ADAM-15, and ADAM-17 are upregulated in macrophages in advanced human atherosclerotic plaques in aorta and carotid and femoral arteries–Tampere vascular study. Ann Med 41: 279–290.
28. Shipley JM, Wesselschmidt RL, Kobayashi DK, Ley TJ, Shapiro SD (1996) Metalloelastase is required for macrophage-mediated proteolysis and matrix invasion in mice. Proc Natl Acad Sci U S A 93: 3942–3946.
29. Johnson JL, Baker AH, Oka K, Chan L, Newby AC, et al. (2006) Suppression of atherosclerotic plaque progression and instability by tissue inhibitor of metalloproteinase-2: involvement of macrophage migration and apoptosis. Circulation 113: 2435–2444.
30. Pepper MS (2001) Role of the matrix metalloproteinase and plasminogen activator-plasmin systems in angiogenesis. Arterioscler Thromb Vasc Biol 21: 1104–1117.
31. Choudhary S, Higgins CL, Chen IY, Reardon M, Lawrie G, et al. (2006) Quantitation and localization of matrix metalloproteinases and their inhibitors in human carotid endarterectomy tissues. Arterioscler Thromb Vasc Biol 26: 2351–2358.
32. Johnson JL, George SJ, Newby AC, Jackson CL (2005) Divergent effects of matrix metalloproteinases 3, 7, 9, and 12 on atherosclerotic plaque stability in mouse brachiocephalic arteries. Proc Natl Acad Sci U S A 102: 15575–15580.
33. Luttun A, Lutgens E, Manderveld A, Maris K, Collen D, et al. (2004) Loss of matrix metalloproteinase-9 or matrix metalloproteinase-12 protects apolipoprotein E-deficient mice against atherosclerotic media destruction but differentially affects plaque growth. Circulation 109: 1408–1414.
34. Thomas AC, Sala-Newby GB, Ismail Y, Johnson JL, Pasterkamp G, et al. (2007) Genomics of foam cells and nonfoamy macrophages from rabbits identifies arginase-I as a differential regulator of nitric oxide production. Arterioscler Thromb Vasc Biol 27: 571–577.
35. Halpert I, Sires UI, Roby JD, Potter-Perigo S, Wight TN, et al. (1996) Matrilysin is expressed by lipid-laden macrophages at sites of potential rupture in atherosclerotic lesions and localizes to areas of versican deposition, a proteoglycan substrate for the enzyme. Proc Natl Acad Sci U S A 93: 9748–9753.
36. Morgan AR, Rerkasem K, Gallagher PJ, Zhang B, Morris GE, et al. (2004) Differences in matrix metalloproteinase-1 and matrix metalloproteinase-12 transcript levels among carotid atherosclerotic plaques with different histopathological characteristics. Stroke 35: 1310–1315.
37. Kilarski L, Achterberg S, Devan WJ, Traylor M, Malik R, et al. (2014) Meta-analysis in over 17,900 cases of ischaemic stroke reveals a novel association at 12q24.12. Neurology (in press).
38. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, et al. (2006) PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet 9: 55–61.
39. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44: 955–959.
40. Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet 11: 499–511.
41. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, et al. (1993) Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke 24: 35–41.
42. Lovett JK, Coull AJ, Rothwell PM (2004) Early risk of recurrence by subtype of ischemic stroke in population-based incidence studies. Neurology 62: 569–573.
43. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.
44. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190–2191.
45. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004.
46. Levula M, Oksala N, Airla N, Zeitlin R, Salenius JP, et al. (2012) Genes involved in systemic and arterial bed dependent atherosclerosis–Tampere Vascular study. PLoS One 7: e33787.
47. Niinisalo P, Oksala N, Levula M, Pelto-Huikko M, Jarvinen O, et al. (2010) Activation of indoleamine 2,3-dioxygenase-induced tryptophan degradation in advanced atherosclerotic plaques: Tampere vascular study. Ann Med 42: 55–63.
48. Oksala N, Levula M, Pelto-Huikko M, Kytomaki L, Soini JT, et al. (2010) Carbonic anhydrases II and XII are up-regulated in osteoclast-like cells in advanced human atherosclerotic plaques-Tampere Vascular Study. Ann Med 42: 360–370.
49. Oksala N, Parssinen J, Seppala I, Raitoharju E, Kholova I, et al. (2013) Association of Neuroimmune Guidance Cue Netrin-1 and its Chemorepulsive Receptor UNC5B with Atherosclerotic Plaque Expression Signatures and Stability in Human(s) - Tampere Vascular Study. Circ Cardiovasc Genet 6:579–87.
50. Raitoharju E, Seppala I, Lyytikainen LP, Levula M, Oksala N, et al. (2013) A comparison of the accuracy of Illumina HumanHT-12 v3 Expression BeadChip and TaqMan qRT-PCR gene expression results in patient samples from the Tampere Vascular Study. Atherosclerosis 226: 149–152.
51. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, et al. (2013) All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. PLoS Genet 9: e1003449.
52. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.
53. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56–65.
54. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
55. Achterberg S, Kappelle LJ, Algra A (2008) Prognostic modelling in ischaemic stroke study, additional value of genetic characteristics. Rationale and design. Eur Neurol 59: 243–252.