Genetics and genomics of cholesterol and polyunsaturated fatty acid metabolism in relation to coronary heart disease risk

Yingchang (Kevin) Lu

Thesis committee

Thesis supervisor

Prof. dr. E.J.M. Feskens

Personal chair at the Division of Human Nutrition

Wageningen University

Prof. dr. M.R. Müller

Professor of Nutrition, Metabolism & Genomics

Wageningen University

Thesis co-supervisor

Dr. J.M.A. Boer

Senior researcher, Centre for Nutrition and Health

National Institute for Public Health and the Environment (RIVM), De Bilt

Other members

Prof. dr. M.A.M. Groenen, Wageningen University

Prof. dr. J.M. Ordovas, Jean Mayer USDA HNRCA at Tufts University, Boston, USA

Prof. dr. ir. R.P. Mensink, Maastricht University Medical Center

Dr. ir. K. Willems van Dijk, Leiden University Medical Center (LUMC)

This research was conducted under the auspices of the Graduate School VLAG

Genetics and genomics of cholesterol and polyunsaturated fatty acid metabolism in relation to coronary heart disease risk

Yingchang (Kevin) Lu

Thesis

submitted in fulfilment of the requirements

for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. dr. M.J. Kropff

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday November 8 2011

at 1.30 p.m. in the Aula.

Yingchang (Kevin) Lu

Genetics and genomics of cholesterol and polyunsaturated fatty acid metabolism in relation to coronary heart disease risk, 216 pages.

Thesis, Wageningen University, Wageningen, NL (2011)

With references, with summaries in English and Dutch

ISBN 978-94-6173-035-0

To my mother

Abstract

Background

Coronary heart disease (CHD) continues to be a leading cause of morbidity and mortality among adults worldwide. Deregulated lipid metabolism (dyslipidemia) that manifests as hypercholesterolemia, hypertriglyceridemia, low high-density-lipoprotein (HDL) cholesterol levels or a combination of those, is an established risk factor for CHD among other established risk factors. Linoleic acid (LA, C18:2n-6) and alpha-linolenic acid (ALA, C18:3n-3) are polyunsaturated fatty acids (PUFAs) that cannot be synthesized *de novo* by human or animal cells, and therefore must be obtained from the diet. From these two PUFAs, two series of long-chain PUFAs are formed; the omega-6 series that are synthesized from LA, and the omega-3 series that are from ALA. Formation of these long-chain PUFAs involves a series of alternate desaturation and elongation processes. These PUFAs, especially, omega-3 PUFAs, have long been observed to reduce CHD risk. In contrast to the consistently observed cardiovascular protective effects of omega-3 PUFAs, accumulating evidence suggests a potential pro-atherogenic effects of omega-6 PUFAs, which is now still under debate.

It has been estimated that genetic factors account for 26%-69% of inter-individual variation in CHD risk. These genetic factors are thought to influence CHD risk both directly and through effects on known CHD risk factors such as plasma lipid levels. The heritability of plasma lipid levels (total cholesterol, LDL cholesterol, HDL cholesterol, and triglycerides (TG)) is estimated to be about 50% (ranging from 28%-78%). With the success of recent genome-wide association studies (GWAS), many genetic variants underlying intermediate risk factors of CHD (including plasma lipid levels) and CHD itself have been identified. Whether this new genetic information could be used to improve CHD risk prediction is still marginally explored, and for some variants, the underlying mechanisms for their mediated effects on CHD risk are still unknown. The aim of this research is to investigate common genetic determinants of plasma lipid levels (cholesterol and polyunsaturated fatty acid levels) using a pathway-driven approach, and to explore whether such common genetic variants could be used to improve CHD prediction using a population based genetic approach. An additional aim was to explore the underlying mechanisms of cardiovascular protective effects of PUFAs using a genomic approach.

Methods

In order to explore whether common genetic variants are involved in determining plasma cholesterol levels, we used data from 3575 men and women from the Doetinchem cohort, which was examined thrice over 11 years. They were genotyped on 384 single nucleotide polymorphisms (SNPs) across 251 genes in regulatory pathways that control fatty acid, glucose, cholesterol and bile salt homeostasis.

In order to explore whether common genetic variants could be used to predict future CHD risk, we used the data from CAREMA cohort that involved 15,236 middle-aged subjects and was followed up for a median of 12.1 years. 179 SNPs associated with CHD or its risk factors in GWAS published up to May 2, 2011 were genotyped in the 2221 subcohort members and 742 incident CHD cases. In addition, fatty acids from plasma cholesteryl esters were quantified in 1323 subcohort members and 537 CHD cases. They were used to explore whether δ-5 and δ-6 desaturase activities were associated with CHD risk.

In order to perform a comparative analysis of the effects of fenofibrate and fish oil at transcriptome and metabolome level, 34 mice were randomized by weight-matching into three groups (n = 10 in control group, and n = 12 in fenofibrate or fish oil intervention group), and fed a research diet supplemented with sunflower oil (containing 81.3% oleic acid, 7% energy intake) in control group, sunflower oil (7% energy intake) and fenofibrate (0.03% w/w) in fenofibrate group, and fish oil (Marinol C-38 fish oil: 23.1% EPA and 21.1% DHA, 7% energy intake) in fish oil group for 2 weeks. At the end of treatment, mice were fasted with drinking water available, and were subsequently sacrificed by cervical dislocation under isoflurane anesthesia. Blood was collected via orbital puncture. Livers were dissected, directly frozen in liquid nitrogen and stored at−80°C until further analysis. Microarray analysis was performed on individual mouse livers. The LC-MS method was used for measuring plasma lipids and non-esterified free fatty acids, and the GC-MS method was used for measuring a broad range of metabolites.

Results

In chapter 2, 3, and 4, common genetic variants in the genes along known cholesterol metabolic pathways, such as bile acid and bile metabolic pathways, the HDL cholesterol metabolic pathway, and the plasma total cholesterol metabolic pathway, are involved in determining plasma cholesterol levels. The modest effect associated with each individual variant, however, caused the amount of heritability explained by them in aggregate to be relatively small: 13 single nucleotide polymorphisms (SNPs) explained 4% of inter-individual variation in HDL cholesterol levels (Chapter 3), whereas 12 SNPs explained 6.9% of inter-individual variation in total cholesterol levels (Chapter 4).

In chapter 5, we found that genetic variants in the FADS1 gene potentially interact with dietary PUFA intake to affect plasma cholesterol levels. A high intake of omega-3 PUFAs was associated with increased plasma non-HDL cholesterol levels, consistent with increased plasma LDL cholesterol levels observed in fish oil intervention studies. Increased LDL cholesterol levels could be due to hepatic downregulation of the LDL receptor gene (LDLR) in subjects with high omega-3 PUFA intakes. This is further confirmed by the findings described in Chapter 6 that the hepatic LDLR gene was significantly downregulated in fish oil treated mice. This study also confirmed PUFAs to be weak PPAR ligands. The increased plasma HDL cholesterol levels in the subjects with high PUFA intakes in Chapter 5 could be due to PPARs-mediated genes that are directly involved in HDL lipoprotein metabolism. All these may explain the changes in blood cholesterol levels upon PUFA intake observed in human studies.

In Chapter 6, we found that not only downregulation in the hepatic lipogenic pathway but also upregulation in hepatic fatty acid oxidation pathways are involved in lowering plasma TG levels upon fish oil treatment. The striking parallel between fenofibrate and fish oil in hepatic downregulation of blood coagulation and fibrinolysis pathways suggest that hepatic activation of PPARα is potentially one of the mechanisms responsible for anticoagulation effects of fish oil treatment observed in humans.

In Chapter 7, with confirmed effects of rs174547 in FADS1 on PUFA levels and δ-5 desaturase activities and also protective effects of DHA on CHD, we observed a reduced CHD risk of increased δ-5 desaturase activity. Increased δ-5 desaturase activity could contribute to the intracellular increase of EPA and especially arachidonic acid (C20:4n-6) levels. Despite the potential pro-coagulant and pro-inflammatory effects of increased exposures of arachidonic acid and its derived eicosanoid metabolites, there is no evidence of increased CHD risk with increased habitual arachidonic acid intake so far. Some of the oxygenated metabolites of arachidonic acid were found to have anti-inflammatory and pro-resolving actions. High dietary n-6 PUFA intakes or high plasma n-6 PUFA levels are associated with increased blood HDL cholesterol levels and reduced TG (or VLDL particle) levels. All these point to a potential cardiovascular protective effect of n-6 PUFAs. The fact that increased EPA and/or DHA levels associated with increased δ-5 desaturase activity protect against CHD is consistent with the current established cardiovascular protective effect of increased n-3 PUFA exposure, especially EPA and DHA.

In Chapter 8, the current known common genetic variants associated with CHD risk factors (blood pressure, obesity, blood lipid levels, and type 2 diabetes) and CHD itself from published GWAS are examined to see whether they provide additional value in CHD risk prediction beyond established traditional CHD risk factors. We constructed several gene risk scores (GRS) for CHD that consisted of SNPs directly associated with CHD or intermediate CHD risk factors in GWAS, and tested their relationship to incident CHD and their potential to improve risk prediction. The weighted GRS based on 29 CHD SNPs predicted future CHD independently from established traditional risk factors. However, the GRS based on 153 SNPs associated with intermediate risk factors and the GRS based on the total 179 SNPs did not. None of them improved risk discrimination. Risk classification of CHD, measured by the net reclassification index, improved only when the GRS based on the 29 CHD SNPs was used. These results are generally consistent with the results from other recent studies that took a similar approach as ours. However, the final conclusions on GRS application could not be drawn at this early stage. With a great understanding of the genetic architecture of CHD in the future, more research should be done on this topic.

Conclusions

Our studies in this thesis demonstrated that common genetic variants along the known candidate cholesterol metabolic pathways are involved in determining the plasma cholesterol levels. PUFAs are not only weak PPARα ligands, but also inhibit SREBPs' activities. All these could explain part of the cardiovascular protective effects (increased HDL cholesterol levels and reduced TG levels) of PUFAs, increased LDL cholesterol levels upon fish oil treatment in humans, and potentially reduced CHD risk of high δ-5

desaturase activities. At present, many questions remain about the feasibility of genetic risk prediction of CHD. Clinicians should continue to inquire about family history of CHD for risk prediction, because this represents a simple, cheap, and useful risk factor for CHD that likely represents the net integrated effects from hundreds of genetic risk variants.

Contents

Chapter 1

General introduction, rationale and outline of this thesis

General introduction

Coronary heart disease (CHD) continues to be a leading cause of morbidity and mortality among adults worldwide. It is caused by atherosclerosis in one or more of the coronary arteries. The development of atherosclerosis begins early in life with the deposition of lipids to form fatty streaks in the arterial wall occurring in childhood and adolescence (1, 2). Rapid progression of the early lesions to fibrous plaques occurs in the third and fourth decades of life, with the rate of progression directly associated with the number of cardiovascular risk factors (1, 2). The presence of atherosclerosis early in life, its relationship to major cardiovascular risk factors at a young age, and its steady progression toward cardiovascular events later in life have indicated that the optional age to begin prevention of atherosclerosis and CHD is as young as possible (1-3).

Deregulated lipid metabolism (dyslipidemia) that manifests as hypercholesterolemia, hypertriglyceridemia, low high-density-lipoprotein (HDL) cholesterol levels or a combination of those is an established risk factor for CHD among other established risk factors. The liver is of major importance in maintaining whole-body lipid metabolic homeostasis and integrates exogenous (dietary) source of lipids with endogenous *de novo* synthesis. Dietary lipids are taken up by intestinal cells and packed into chylomicron particles. Upon secretion into the systemic circulation, most of triglycerides (TG) in chylomicrons are hydrolyzed by lipoprotein lipase, releasing free fatty acids (FFAs) for uptake by peripheral tissues such as muscle and adipose tissue. The remaining chylomicron remnants are delivered to the liver. During fasting, plasma levels of insulin fall, whereas levels of glucagon and epinephrine increase, stimulating TG hydrolysis in adipose tissue. The released FFAs are transported to the liver. FFAs in liver can be oxidized in mitochondria to produce energy and ketone bodies, re-esterified to TG and stored in lipid droplets, or coupled to apolipoproteins and secreted as a constituent of very-low-density-lipoprotein (VLDL) particles. Upon secretion, the TG in VLDL are hydrolyzed by the lipases in peripheral tissues resulting in cholesterol-dense low-density-lipoprotein (LDL) particles that are mainly taken up by the liver.

Linoleic acid (LA, C18:2n-6) and alpha-linolenic acid (ALA, C18:3n-3) are polyunsaturated fatty acids (PUFAs) that cannot be synthesized *de novo* by human or animal cells. They are indispensable for normal development and function, and therefore must be obtained from the diet. From these two PUFAs, two series of long-chain PUFAs are formed; the omega-6 series that are synthesized from LA, and the omega-3 series that are from ALA. Formation of these long-chain PUFAs involves a series of alternate desaturation (insertion of a double bond by fatty acid desaturases) and elongation (addition of two carbon atoms by fatty acid elongases) processes that occur predominantly in the endoplasmic reticulum of the liver. These PUFAs, especially omega-3 PUFAs, have long been observed to reduce CHD risk (4-8). In addition to the established blood TG-lowering and HDL cholesterol increasing effects of PUFAs (6, 9), omega-3 PUFAs have also been observed to have anti-thrombotic, anti-inflammatory, and anti-arrhythmic effects in humans (5, 7, 8). However, the underlying mechanisms for most of these effects are still incompletely understood. In contrast to the consistently observed cardiovascular protective

effects of omega-3 PUFAs, accumulating evidence suggests a potential pro-atherogenic effect of omega-6 PUFAs (10-14), which is still under debate (6, 15, 16).

Family history of premature CHD is an independent risk factor of CHD, suggesting that inherited genetic factors contribute to CHD risk (17-19). It has been estimated that genetic factors account for 26% to 69% of inter-individual variation in CHD risk (18, 20, 21). The influence of genetic factors on CHD risk decreases as age increases (17, 19). These genetic factors are thought likely to influence CHD risk both directly and through effects on known CHD risk factors such as plasma lipid levels. The heritability of plasma lipid levels (total cholesterol, LDL cholesterol, HDL cholesterol, and TG) is estimated to be about 50% (ranging from 28%-78%), significantly greater among young subjects than among old subjects (22). These observations have motivated investigators to undertake a variety of studies to identify the genes responsible for the heritability of CHD and its risk factors. Early candidate gene studies of CHD are however often plagued with a lack of consistency and reproducibility, most of which stems from studying one single nucleotide polymorphism, lack of power and/or population stratification. More recently, the completion of the Human Genome Project (23, 24) and the International Haplotype Map Project (25, 26) has made it possible to perform genome-wide screens for common DNA sequence variants that are associated with phenotypes of interest, including CHD and its risk factors. This approach benefits from being “hypothesis free”, and is therefore not subject to constraints and potential biases seen in candidate gene studies. This approach has substantially expanded our knowledge of the genetic basis of CHD and plasma lipid levels, with 24 and 95 unequivocal genetic loci recently identified to be associated with CHD and plasma lipid levels, respectively (27, 28).

For the past 5 decades, the major CHD risk factors, namely, male sex, hypertension, dyslipidemia, smoking, and diabetes mellitus, have been well established (29, 30). On the basis of these factors, a number of risk prediction models have been developed, including the Framingham risk score (31) and the European cardiovascular score (SCORE) (32). These prediction models, however, cannot predict many cases of incident CHD, motivating the identification of new risk factors and the refinement of existing prediction models. Haunted by the notion of family history being one of the strong independent CHD risk factors, individual genetic variants identified from the genome-wide association studies on the association with CHD and its risk factors, have been tested to improve CHD risk prediction. Like for many other novel CHD risk biomarkers, however, many fundamental and specific questions remain to be answered: How many variants and what allelic spectrum underlie the genetic architecture of CHD? How many genetic variants are required to reliably improve risk prediction? Can genetic variants or gene risk scores improve CHD risk prediction beyond the established traditional CHD risk factors?

Rationale

With the success of recent genome-wide association studies, many genetic variants underlying intermediate risk factors of CHD (including plasma lipid levels, type 2 diabetes, and hypertension, etc.) and CHD itself have been identified. Whether this new genetic information could be used to improve CHD risk prediction is still marginally explored, and

for some variants, the underlying mechanisms for their mediated effects on CHD risk are still unknown. The aim of the research described in this thesis is to investigate common genetic determinants of plasma lipid levels (cholesterol and polyunsaturated fatty acid levels) using a pathway-driven approach and to explore whether such common genetic variants could be used to improve CHD prediction using a population based genetic approach. An additional aim was to explore the underlying mechanisms of cardiovascular protective effects of PUFAs using a genomic approach.

Outline of this thesis

In Chapter 2, a detailed review is presented on the potential influence of genetic variants in genes along bile acid and bile metabolic pathways on blood cholesterol levels. Chapter 3 and 4 aim to explore genetic determinants of plasma HDL and total cholesterol levels using a candidate pathway-driven approach. In addition, in Chapter 4, it is examined whether the genetic information from these variants could add value to predicting future risk of hypercholesterolemia. In Chapter 5, the potential mechanism underlying the influence of dietary PUFAs on plasma cholesterol levels is explored using a population-based genetic approach. In Chapter 6, novel mechanistic insights on the TG-lowering and anti-thrombotic effects of fish oil (omega-3 PUFAs) treatment is explored in comparison with fenofibrate treatment in mice using a genomic approach. In Chapter 7, the genetic determinants of plasma PUFA levels, and potential effects of omega-6 and omega-3 PUFAs and desaturase activities on CHD risk are explored using a population-based genetic approach. In Chapter 8, the current known common genetic variants associated with CHD risk factors (blood pressure, obesity, blood lipid levels, and type 2 diabetes) and CHD itself, derived from published genome-wide association studies, are examined to see whether they provide additional value in CHD risk prediction beyond established traditional CHD risk factors. Finally, in Chapter 9, the general discussion and conclusions are presented.

References

1. McGill, H.C., Jr., McMahan, C.A., and Gidding, S.S. 2008. Preventing heart disease in the 21st century: implications of the Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study. Circulation 117:1216-1227.
2. Berenson, G.S., Srinivasan, S.R., Bao, W., Newman, W.P., 3rd, Tracy, R.E., and Wattigney, W.A. 1998. Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study. *New Engl J Medicine* 338:1650-1656.
3. Gidding, S.S. 2010. Assembling evidence to justify prevention of atherosclerosis beginning in youth. *Circulation* 122:2493-2494.
4. Hu, F.B., Stampfer, M.J., Manson, J.E., Rimm, E., Colditz, G.A., Rosner, B.A., Hennekens, C.H., and Willett, W.C. 1997. Dietary fat intake and the risk of coronary heart disease in women. *New Engl J Med* 337:1491-1499.
5. Harris, W.S., Miller, M., Tighe, A.P., Davidson, M.H., and Schaefer, E.J. 2008. Omega-3 fatty acids and coronary heart disease risk: clinical and mechanistic perspectives. *Atherosclerosis* 197:12-24.
6. Harris, W.S., Mozaffarian, D., Rimm, E., Kris-Etherton, P., Rudel, L.L., Appel, L.J., Engler, M.M., Engler, M.B., and Sacks, F. 2009. Omega-6 fatty acids and risk for cardiovascular disease: a science advisory from the American Heart Association Nutrition Subcommittee of the Council on Nutrition, Physical Activity, and Metabolism; Council on Cardiovascular Nursing; and Council on Epidemiology and Prevention. *Circulation* 119:902-907.

7. Mozaffarian, D. 2008. Fish and n-3 fatty acids for the prevention of fatal coronary heart disease and sudden cardiac death. *Am J Clin Nutr* 87:1991S-1996S.
8. De Caterina, R. 2011. n-3 fatty acids in cardiovascular disease. New Engl J Med 364:2439-2450.
9. Mensink, R.P., Zock, P.L., Kester, A.D., and Katan, M.B. 2003. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am J Clin Nutr* 77:1146-1155.
10. Kark, J.D., Kaufmann, N.A., Binka, F., Goldberger, N., and Berry, E.M. 2003. Adipose tissue n-6 fatty acids and acute myocardial infarction in a population consuming a diet high in polyunsaturated fatty acids. *Am J Clin Nutr* 77:796-802.
11. Baylin, A., and Campos, H. 2004. Arachidonic acid in adipose tissue is associated with nonfatal acute myocardial infarction in the central valley of Costa Rica. *J Nutr* 134:3095-3099.
12. Dwyer, J.H., Allayee, H., Dwyer, K.M., Fan, J., Wu, H., Mar, R., Lusis, A.J., and Mehrabian, M. 2004. Arachidonate 5-lipoxygenase promoter genotype, dietary arachidonic acid, and atherosclerosis. *New Engl J Med* 350:29-37.
13. Martinelli, N., Girelli, D., Malerba, G., Guarini, P., Illig, T., Trabetti, E., Sandri, M., Friso, S., Pizzolo, F., Schaeffer, L., et al. 2008. FADS genotypes and desaturase activity estimated by the ratio of arachidonic acid to linoleic acid are associated with inflammation and coronary artery disease. *Am J Clin Nutr* 88:941-949.
14. Ramsden, C.E., Hibbeln, J.R., Majchrzak, S.F., and Davis, J.M. 2010. n-6 fatty acid-specific and mixed polyunsaturate dietary interventions have different effects on CHD risk: a meta-analysis of randomised controlled trials. *Brit J Nutr* 104:1586-1600.
15. Calder, P.C. 2010. The American Heart Association advisory on n-6 fatty acids: evidence based or biased evidence? *Brit J Nutr* 104:1575-1576.
16. Martinelli, N., Consoli, L., and Olivieri, O. 2009. A 'desaturase hypothesis' for atherosclerosis: Janus-faced enzymes in omega-6 and omega-3 polyunsaturated fatty acid metabolism. *J Nutrigenet Nutrigenom* 2:129-139.
17. Marenberg, M.E., Risch, N., Berkman, L.F., Floderus, B., and de Faire, U. 1994. Genetic susceptibility to death from coronary heart disease in a study of twins. *New Engl J Med* 330:1041-1046.
18. Zdravkovic, S., Wienke, A., Pedersen, N.L., Marenberg, M.E., Yashin, A.I., and De Faire, U. 2002. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* 252:247-254.
19. Lloyd-Jones, D.M., Nam, B.H., D'Agostino, R.B., Sr., Levy, D., Murabito, J.M., Wang, T.J., Wilson, P.W., and O'Donnell, C.J. 2004. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* 291:2204-2211.
20. Fischer, M., Broeckel, U., Holmer, S., Baessler, A., Hengstenberg, C., Mayer, B., Erdmann, J., Klein, G., Riegger, G., Jacob, H.J., et al. 2005. Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction. *Circulation* 111:855-862.
21. Nora, J.J., Lortscher, R.H., Spangler, R.D., Nora, A.H., and Kimberling, W.J. 1980. Genetic--epidemiologic study of early-onset ischemic heart disease. *Circulation* 61:503-508.
22. Heller, D.A., de Faire, U., Pedersen, N.L., Dahlen, G., and McClearn, G.E. 1993. Genetic and environmental influences on serum lipid levels in twins. *New Engl J Med* 328:1150-1156.
23. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
24. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* 291:1304-1351.

25. 2005. A haplotype map of the human genome. *Nature* 437:1299-1320.
26. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851-861.
27. Schunkert, H., Konig, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C., et al. 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 43:333-338.
28. 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., et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707-713.
29. Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., and Stokes, J., 3rd. 1961. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. *Ann Intern Med* 55:33-50.
30. Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J., et al. 2004. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 364:937-952.
31. Wilson, P.W., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., and Kannel, W.B. 1998. Prediction of coronary heart disease using risk factor categories. *Circulation* 97:1837-1847.
32. Conroy, R.M., Pyorala, K., Fitzgerald, A.P., Sans, S., Menotti, A., De Backer, G., De Bacquer, D., Ducimetiere, P., Jousilahti, P., Keil, U., et al. 2003. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 24:987-1003.

Chapter 2

Review

The potential influence of genetic variants in genes along bile acid and bile metabolic pathway on blood cholesterol levels in the population

Yingchang Lu, Edith J.M. Feskens,
Jolanda M.A. Boer and Michael Müller

Atherosclerosis. 2010;210(1):14-27.

Abstract

The liver is currently known to be the major organ to eliminate excess cholesterol from our body. It accomplishes this function in two ways: conversion of cholesterol molecules into bile acids (BAs) and secretion of unesterified cholesterol molecules into bile. BAs are synthesized in the hepatocytes, secreted into bile and delivered to the lumen of the small intestine where they act as detergents to facilitate absorption of fats and fat-soluble vitamins. About 95% of BAs are recovered in the ileum during each cycle of the enterohepatic circulation. Five percent are lost and replaced by newly synthesized BAs, which amounts to ~500 mg/day in adult humans. In contrast to the efficiency of the BAs' enterohepatic circulation, 50% of the 1000 mg of cholesterol secreted daily into bile is lost in feces. It is known that rare human mutations in certain genes in bile acid and bile metabolic pathway influence blood cholesterol levels. With the recent success of genome-wide association studies, we are convinced that common genetic variants also play a role in the genetic architecture of plasma lipid traits. In this review, we summarized the current state of knowledge about genetic variations in bile acid and bile metabolic pathway, and assessed their impact on blood cholesterol levels and cholesterol metabolic kinetics in the population.

Key Words: cholesterol catabolism; bile acid synthesis; enterohepatic circulation; genetics

Introduction

The liver is currently known to be the major organ to eliminate excess cholesterol from our body, despite the emerging role of the small intestine (1, 2). The liver accomplishes this function in two ways; it converts cholesterol molecules into bile acids (BAs), and secretes unesterified (free) cholesterol molecules into bile. BAs are exclusively synthesized in hepatocytes, secreted into bile and delivered to the lumen of the small intestine, where they act as detergents to facilitate absorption of dietary lipids, cholesterol, and fat-soluble vitamins. BAs are reabsorbed from the ileum by the action of specific transporter systems, transported from the intestine to the liver via the portal circulation and then resecreted into bile. During each cycle of this enterohepatic circulation, about 95% of BAs are recovered, and 5 percent are lost and replaced by newly synthesized BAs. Although the fractional loss of BAs per cycle is relatively small, total daily BA synthesis in adult humans amounts to ~500 mg, accounting for 90% of the cholesterol that is actively metabolized in the body. In contrast to the efficiency of the BAs' enterohepatic circulation, about 50% of the 1000 mg of cholesterol secreted daily into bile is lost in feces as neutral sterols, thereby, almost equal to the loss of cholesterol as BAs (3). The remaining 50%, together with half of the dietary cholesterol, is absorbed into enterocytes in the proximal small intestine, packed into chylomicrons, secreted into the mesenteric lymph and then enters the bloodstream.

Liver is one of the two major organs to synthesize apoA-I. After initial lipidation of lipid-poor apoA-I in the liver, the nascent HDL particle is secreted. With its intravascular maturation and remodelling, the HDL particle facilitates the uptake of peripheral cholesterol and its return to the liver directly or by transferring cholesterol to VLDL and LDL in plasma, which eventually deliver much of their cholesterol to the liver (4, 5). The liver can then channel the excess cholesterol for excretion into the bile. As the major physiological process for our body to clear excess cholesterol, the bile acid and bile metabolic pathway plays a critical role in the maintenance of whole-body cholesterol homeostasis. It has already been shown that rare human mutations in some genes in this pathway influence blood cholesterol levels dramatically: rare mutations in cholesterol 7a-hydroxylase gene (CYP7A1) (6) and ATP binding cassette transporter G5 and G8 genes (ABCG5 and ABCG8) (7-9) cause elevated plasma cholesterol levels, while rare mutations in apical sodium bile acid transporter gene (ASBT) (10) cause reduced plasma cholesterol levels. With the recent success of genome-wide association (GWA) studies, we are convinced that common genetic variants in bile metabolic pathway also play a role in determining plasma cholesterol levels, but with a modest effect, individually (11-13). So far, the effects of genetic variants in bile acid and bile metabolic pathway on blood cholesterol levels haven't systematically been reviewed. Therefore, in this review, we summarized the current state of knowledge about genetic variants in CYP7A1, ABCG5, ABCG8, NPC1L1, LXRa, HNF4α and other potentially relevant genes in this pathway (Figure 1), and assessed their impact on blood cholesterol levels and cholesterol metabolic kinetics in the population. The literature search for variations in genes in the bile acid and bile metabolic pathway was performed in PubMed using search terms: genetic variant or polymorphism or mutation and cholesterol. We also explored associations between common genetic variants in those genes important in this pathway and plasma LDL and HDL cholesterol levels in the publicly available meta-analysis of seven GWA studies (12).

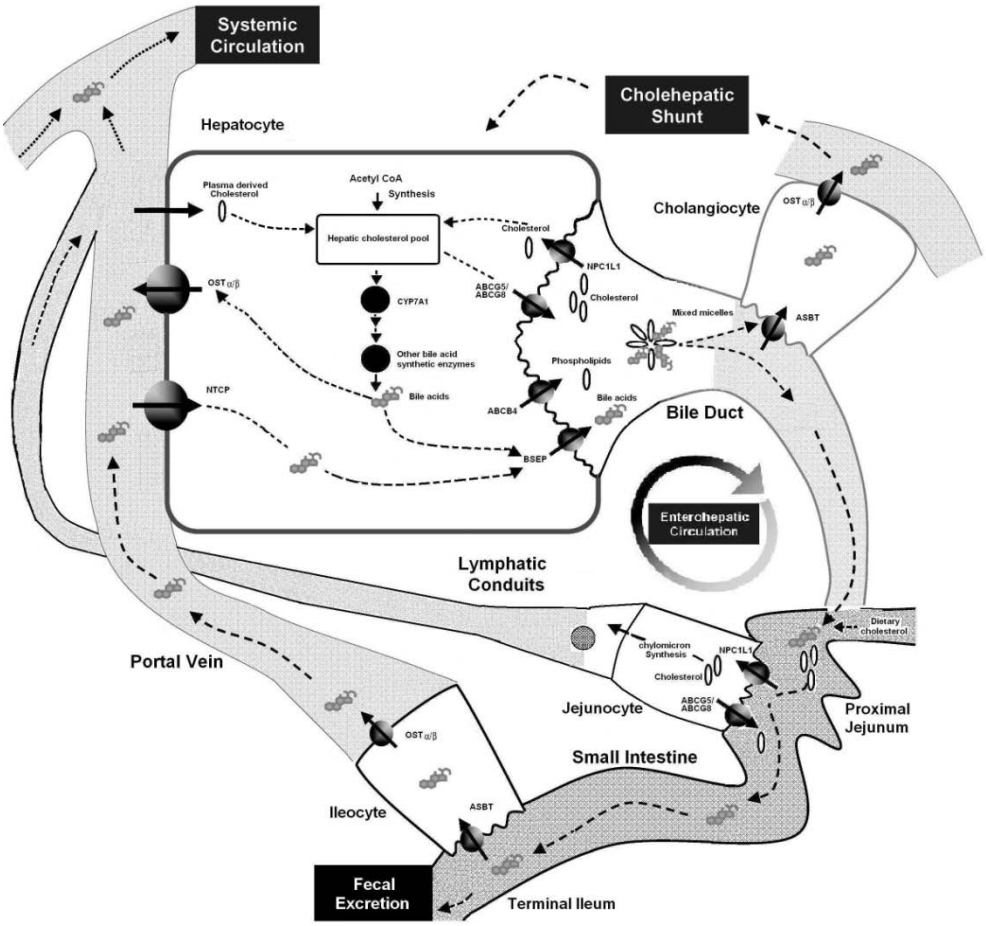

Figure 1: Bile acid and cholesterol transport in enterohepatic circulation. Bile acids are taken up from the intestinal lumen mainly by the apical sodium-dependent bile acid transporter (ASBT) in the terminal ileum. After intracellular transport in the ileocyte mediated by ileal bile acid binding protein (I-BABP, not shown), bile acids are excreted by the heterodimeric organic solute transporter α/β (OST α/β) into the portal circulation. Upon reaching the liver sinusoids, bile acids are taken up by hepatocytes via Na^+-dependent taurocholate cotransport peptide (NTCP) (conjugated form) or Na^+-independent organic anion transporting polypeptides (conjugated or unconjugated form, not shown). The newly synthesized bile acids by a complex series of enzymes including cholesterol 7a hydroxylase (CYP7A1) and other enzymes in the hepatocytes, together with the ones taken up from the portal circulation, are secreted across the canalicular membrane into bile by the bile sale export pump (BSEP). After secretion into the bile, the majority of bile acids travel through the bile ducts to reach the intestinal lumen. A small proportion of bile acids can be taken up by cholangiocytes lining the bile ducts via ASBT and secreted across the basolateral membrane back into circulation, destined for either reuptake by the liver (cholehepatic shunt) or the systemic circulation. Hepatobiliary excretion of cholesterol is mediated by the heterodimer of ATP-binding cassette protein G5 and G8 (ABCG5/ABCG8). With bile acids and phospholipids (its secretion is medicated by ABCB4 in canalicular membrane of hepatocytes) together, they form micelles in bile and are

secreted into small intestine. Niemann-Pick C1-like1 (NPC1L1) along the canalicular membrane of hepatocytes facilitates the reuptake of cholesterol from bile. Half of biliarily secreted and diet-derived cholesterol is taken up through NPC1L1 on the apical surface of absorptive enterocytes in the proximal small intestine. Together, with other absorbed lipids, they are packaged into chylomicrons and secreted into the lymph, and then enter the bloodstream. Part of the intracellular cholesterol of the enterocyte in the small intestine is excreted into lumen through ABCG5/ABCG8. Nuclear receptors important in regulating bile acids and cholesterol metabolism, such as farnesoid X receptor (FXR), liver X receptor alpha (LXRα), pregnane X receptor (PXR), retinoid X receptor alpha (RXRα), small heterodimer partner (SHP), liver receptor homologue-1 (LRH-1) and hepatocyte nuclear factor 4α (HNF4α), are also reviewed but not shown (for regulating mechanisms see reference 31 and 77). This figure was adapted from Kosters et al (85).

1. CYP7A1

The hydroxylation step catalyzed by the microsomal cytochrome P-450 enzyme cholesterol 7α-hydroxylase (CYP7A1), yielding 7α-hydroxy-cholesterol, is the first step of the so-called “classical pathway” of BA biosynthesis. CYP7A1 is considered to be the rate-limiting enzyme for cholesterol catabolism into BAs. CYP7A1 induction stimulates the conversion of cholesterol to BAs, resulting in a relative deprivation of hepatic microsomal cholesterol content, followed by upregulation of LDL receptor expression and activity, which consequently reduces plasma LDL cholesterol levels (14). A frameshift mutation in the CYP7A1 gene (L413fsX414) has been identified in a family of patients with elevated plasma cholesterol levels, decreased BA excretion, and accumulation of cholesterol in the liver (6). Also, several common genetic variants in this gene have been found. Based on re-sequencing data in a limited number (n = 20) of Caucasian subjects (15, 16), and HapMap CEU data (17), it is now known that a linkage disequilibrium (LD) block exists in Caucasians, spanning a 14-kb region from the proximal promoter (rs3824260) to the 3′-downstream (rs10504255) of the CYP7A1 gene. Two common SNPs (rs3808607, r.-203A>C; rs3824260, r.-469C>T) in the proximal promoter region of CYP7A1 have been studied in relation to lipids. Since rs3824260 is in complete LD with rs3808607, rs3808607 was most studied. Initially, Wang et al. (15) reported that CC homozygotes of rs3808607 had higher plasma LDL cholesterol levels compared to common AA homozygotes in both family based and unrelated American Caucasian populations. This association was replicated in male subjects from the Framingham offspring study (18). In contrast, carriers of the C allele had lower plasma total cholesterol levels compared to AA homozygotes in 139 Dutch hypertriglyceridaemic patients (19). No association was observed in Swedish male populations (16), which is perhaps due to the relative small sample sizes with deviated allele frequencies. Also, no association was observed in 715 male Dutch patients with coronary atherosclerosis (20). This may have been due to a less pronounced effect of this polymorphism in these preselected patients who had elevated total and LDL cholesterol levels.

In other ethnic populations, genetic polymorphisms in the promoter region of CYP7A1 were associated with blood cholesterol or apolipoprotein levels. Hegele et al. (21) studied three distinct Canadian populations. In 594 Hutterites, the C allele carriers of r.-203A>C had high HDL cholesterol and apoA-I levels compared to AA homozygotes; in 190 Keewatin Inuit members, the C allele was the major allele and associated with lower plasma total and LDL cholesterol levels compared to AA homozygotes; and no association

was observed in an OjiCree population. In 1102 Micronesians, Han et al. (22) observed the rare CC homozygotes of r.-203A>C had a higher apoA-I levels compared to homozygote AA. In blacks (n = 1939), Klos et al. (23) found two SNPs (rs1023649, rs1023651) in the promoter region of CYP7A1 to be associated with plasma LDL cholesterol and total cholesterol levels with increased levels in carriers of the rare alleles. Moreover, a 5′region CYP7A1 haplotype (consisting of rs1023649, rs8192871 and rs11786580) was associated with plasma apoA-I levels.

In a population with hypercholesterolemia, mainly composed of Caucasians, Kajinami et al. (24) demonstrated that the rare C allele was independently and possibly gene-dose dependently associated with a poor response to atorvastatin in terms of LDL cholesterol lowering. This effect was more striking in men, and was enhanced by the coexistence of common variants of APOE (ε2 or ε4). This poor response was also only observed in subjects carrying ABCG8 19H, 54CC, 400TT and 632AA (for details see ABCG8 part) by the same investigators (25). Similarly, in 363 male Dutch patients with coronary atherosclerosis, Hofman et al. (20) observed a poor response to pravastatin among -203C allele carriers regarding serum total cholesterol levels. The poor reduction in LDL cholesterol levels among C allele carriers did not reach statistical significance compared to AA homozygotes (p = 0.09). Recently, in a GWA study of patients with coronary heart disease, C allele carriers responded poorly to atorvastatin regarding LDL cholesterol lowering, but it did not reach statistical significance (p = 0.06) (26). The poor response to statins among C allele carriers is reminiscent of the refractory hypercholesterolemia in response to HMG-CoA reductase inhibitors in 2 patients carrying homozygous rare frameshift mutation in CYP7A1 (6). In agreement with the difference in response to statins between CYP7A1 genotypes, the C allele was found to be associated with a higher response of plasma total cholesterol and HDL cholesterol levels to an increase in dietary cholesterol intake in Dutch intervention studies (27). Additionally, Kovar et al. (28) reported that CC homozygotes demonstrated a significant increase in serum total cholesterol and LDL cholesterol levels after a high-fat diet challenge. This increase was not observed in AA homozygotes. Interestingly, the same group earlier reported a larger decrease in plasma total cholesterol levels in C allele carriers after a pronounced change in dietary composition (greatly decreased meat, eggs, butter and animal fat consumption, and increased vegetables, fruits and vegetable oils consumption) over an 8-year-follow-up (29).

It is speculated that r.-203A>C modulates transcriptional activity of the CYP7A1 gene. The -203C variant allele could be associated with decreased CYP7A1 gene expression and, consequently, decreased cholesterol catabolism into BAs. Studies on the transcriptional regulation of CYP7A1 revealed that the promoter region around position -203 contains several liver-specific elements (30, 31). According to GenBank data, the nucleotide sequence from -206 to -199 of the human CYP7A1 promoter is completely conserved in chimpanzee, orangutan, rhesus monkey, dog, mouse, rat and opossum. However, no association was observed between r.-203A>C and hepatic cholesterol 7a-hydroxylase activity in 21 liver biopsies obtained from patients who had undergone cholecystectomy (16). Additionally, no association was observed with plasma parameters of BA synthesis rate (7α-hydoxy-4-cholesten-3-one concentration) in 30 subject with asymptomatic

gallstone disease, and also with the synthesis rates of total BAs, cholic acid and chenodeoxycholic acid in another 30 subjects (16); however, the sample sizes in these studies were relatively small. In a recent study conducted among 65 Czech patients with ileal resection, serum cholesterol and non-HDL cholesterol adjusted 7α-hydoxy-4-cholesten-3-one concentrations were significantly higher in -203 AA than CC homozygotes, and the differences were more pronounced in patients with extensive resection (32). Considering the high LD among common genetic variants in this gene in Caucasians, it cannot be excluded that another yet unidentified functional variant that is in high linkage disequilibrium with r.-203A>C is responsible for this observed influence on plasma total and LDL cholesterol levels.

Overall, based on the current information described above, it can be concluded that genetic variation in the promoter region of CYP7A1 (r.-203A>C in Caucasian populations) is associated with blood cholesterol levels, particularly with LDL cholesterol. Moreover, in Caucasian populations, r.-203A>C may modulate the cholesterol response to drug and diet intervention. The rare C allele carriers may respond poorly to statin-based treatment, but strongly to a high fat diet regarding blood LDL levels. However, it should be noted that these gene (r.-203A>C) -drug or diet interactions were only investigated in very few populations, most with small sample size. Therefore, more research among multiple large cohorts is needed in the future before a final conclusion can be drawn.

Regarding other important BA biosynthetic enzymes (such as sterol 27-hydroxylase (CYP27A1), important for alternative BA biosynthesis and sterol side chain oxidation; oxysterol 7a-hydroxylase (CYP7B1), converting oxysterol to BA intermediates in alternative BA biosynthesis; and sterol 12a-hydroxylase (CYP8B1), channelling BA intermediates ultimately into cholic acid) (3, 33, 34), rare mutations are reported in the CYP27A1 gene (35) and CYP7B1 gene (36, 37), but these mutations do not influence blood cholesterol levels. Based on the recent meta-analysis of seven GWA studies (Supplement table 1 and 2), common genetic variations in these genes may not have a large influence on plasma LDL and HDL cholesterol levels.

2. ABCG5 and ABCG8

Hepatobiliary secretion of BAs drives excretion of cholesterol from liver cells into bile, which is mediated by the heterodimer of ATP-binding cassette protein G5 and G8 (ABCG5 and ABCG8) that are expressed in the canalicular membrane of the liver. Rare mutations in ABCG5 or G8 in human cause sitosterolemia that is characterized by an accumulation of sterols in blood and tissues, consequent to the enhanced intestinal absorption and decreased biliary removal of cholesterol and plant sterol (7-9). Since ABCG5 and G8 are also expressed in the apical membrane of the brush border of the small intestine, the effect of the augmented excretion of cholesterol from the liver into bile cannot be distinguished from that of the enhanced efflux of cholesterol from enterocytes to the small intestinal lumen with respect to influence on blood cholesterol levels. However, based on the observation that a patient with sitosterolemia who underwent a liver transplantation had a normalized plant sterol levels despite his genetic defect in the intestine (38), the effect on biliary cholesterol excretion in the liver may be more significant.

Given the important function of ABCG5 and G8 in the excretion of cholesterol, common sequence variants in the ABCG5 or ABCG8 gene may have subtle effects on sterol metabolism, and may contribute to inter-individual variation in blood cholesterol levels. Several polymorphisms and haplotype structures for ABCG5 and ABCG8 have been reported (39, 40). Several polymorphisms in ABCG5 (Q604E, rs6720173) and ABCG8 (T400K, rs4148217; D19H, rs11887534; A632V, rs6544718; and Y54C, rs4148211) have been found to be associated with several facets of cholesterol metabolism, including cholesterol level, cholesterol kinetics, and individual responsiveness of blood cholesterol to dietary and pharmaceutical intervention (41).

Regarding Q604E of ABCG5, Weggemans et al. (42) demonstrated that EE homozygotes had higher plasma total cholesterol levels than carriers of the wild-type allele. Viturro et al. (43) found that heterozygote boys had higher plasma total cholesterol, LDL cholesterol and apoB levels compared to homozygotes of wide-type allele, but only in those within the lowest tertile of saturated fat intake. However, other studies failed to observe any of these associations (44-48). In contrast, Acalovschi et al. (49) observed the opposite effect. The 604E allele carriers had lower plasma total cholesterol levels and higher HDL cholesterol levels compared with QQ homozygotes in 68 siblings with gallstone disease. Also Plat et al. (50) observed lower serum LDL cholesterol levels among E allele carriers taking a low-erucic rapeseed oil-based margarine and shortening diet in 112 healthy Dutch volunteers. In line with these findings of low blood cholesterol levels associated with the E allele, Gylling et al. (45) demonstrated that the E allele was associated with low cholesterol absorption from intestine (low serum cholesterol adjusted campesterol and sitosterol levels) and high cholesterol synthesis (high serum cholesterol adjusted cholestenol levels) in 262 Finish subjects with mild to moderate hypercholesterolemia. Along the same line, Santosa et al. (48) reported that homozygous 604E subjects had a larger reduction in cholesterol absorption and greater increase in synthesis during weight loss compared to heterozygotes and homozygous wild-type carriers in 35 hypercholesterolemic Canadian women. In dietary cholesterol intervention studies, Herron et al. (51) found that, after one additional egg consumption per day over 30 days, 640E allele carriers responded less compared to homozygote QQ with respect to plasma total and LDL cholesterol levels. However, an opposite trend was observed earlier by Weggemans et al. (42) regarding serum total cholesterol levels change after dietary cholesterol challenge, but without reaching statistical significance. No modulating effect was observed from Q604E on cholesterol lowering response to atorvastatin (25, 26). Overall, the various studies reviewed on effects of Q604E of ABCG5 gene on blood cholesterol levels and cholesterol metabolic kinetics suggest that the rare E allele are consistently associated with lower cholesterol absorption and higher cholesterol synthesis. No consistent result on association with total and LDL cholesterol levels could be drawn now.

Several studies consistently demonstrated that subjects with the 19H allele of ABCG8 had lower blood total cholesterol levels (45, 47-49, 52) and lower LDL cholesterol levels (45, 47, 48) compared to the subjects without. The association with LDL cholesterol levels was recently replicated in a GWA study ($p = 1\times10^{-11}$) (12). However, in another relatively big

study that consisted of 2012 patients with heterozygous familial hypercholesterolemia, Koeijvoets et al. (53) did not observe any association between this polymorphism and lipid levels. This may be due to the attenuation of the moderate effects from19H allele by high plasma cholesterol levels in these preselected subjects. The 19H allele was also observed to be more common in subjects with relatively low cholesterol absorption (45), and was associated with lower cholesterol absorption marker sterols (serum cholesterol adjusted campesterol, sitosterol, and cholestanol levels) (44, 45). Furthermore, it was associated with higher cholesterol synthesis marker sterols (serum cholesterol adjusted serum cholestenol and lathosterol levels) (45).

The finding of a consistent association between D19H and cholesterol metabolic kinetics (baseline cholesterol levels, cholesterol absorption and synthesis) suggests that the substitution of histidine for aspartic acid at amino acid 19 alters the function of ABCG8. Since the plasma sterol (both cholesterol and phytosterol) levels were lower in individuals with histidine at this residual, one would expect an increased transporter function. This is supported by observations that the presence of the 19H mutant allele was associated with cholesterol gallstones, suggesting that the mutated allele might confer a more efficient transport of cholesterol into bile lumen, causing cholesterol hypersaturation of the bile and eventually promoting the formation of cholesterol gallstones (54). In the 19H allele carriers, this increased transporter function causes more efflux of cholesterol into intestinal lumen and as a result, inefficiency in intestinal cholesterol absorption, which consequently induces an upregulated endogenous cholesterol synthesis. This may explain the increased LDL cholesterol lowering response to atorvastatin in the 19H allele carriers (25, 26, 52).

No association was observed between T400K of the ABCG8 gene and total and LDL cholesterol levels (44, 45, 47-50, 53, 55). In 120 male Czech participants, Hubacek et al. (46) reported that K allele carriers exhibited a smaller decrease in plasma total and LDL cholesterol levels than TT homozygotes after a pronounced change in dietary composition (considerably decreased red meat, eggs, and animal fat consumption, and increased vegetables, fruits, cereals, and vegetable oils consumption) over an 8-year-follow-up. However, in 143 healthy American Caucasians, Berge et al. (44) reported that the K allele was associated with lower levels of cholesterol absorption marker sterol (serum cholesterol adjusted sitosterol levels) and high cholesterol synthesis marker sterol (serum cholesterol adjusted desmosterol and lathosterol levels). In line with the lower absorption effect with K allele, in 120 healthy Dutch volunteers, Plat et al. (50) observed a higher campesterol and sitosterol absorption, and a stronger inhibitor effect of plant stanol ester on campesterol and sitosterol absorption in TT homozygotes. However, these associations were not replicated by Gylling et al. (45) in 262 Finish participants with mild to moderate hypercholesterolemia.

Regarding A632V in ABCG8, Berge et al. (44) observed that the V allele was associated with a high plasma total cholesterol levels. In 380 Spanish children, Viturro et al. (43) found that the heterozygotes had higher plasma total cholesterol and apoB levels than AA homozygotes, but only in the group with low cholesterol intake. No association of A632V and total cholesterol levels was observed in other studies (46, 49, 50). No association was

observed between Y54C of ABCG8 and blood total and LDL cholesterol levels (44, 45, 47-49). In 35 hypercholesterolemic Canadian women, Santosa et al. (48) showed that the heterozygous Y54C carriers had a smaller decline in cholesterol synthesis compared with homozygous YY carriers during weight loss through decreasing dietary energy intake and increasing energy expenditure. Additionally, in 139 female Czech subjects, Hubacek et al. (46) reported that the Y allele carriers had a larger plasma total and LDL cholesterol decrease compared to the homozygous CC allele carriers after a change in dietary composition (considerably decreased red meat, eggs, and animal fat consumption, and increased vegetables, fruits, cereals, and vegetable oils consumption) over 8 years' follow-up. No such association was observed in males (46).

Overall, no consistent results on effects of T400K, A632V and Y54C in ABCG8 gene on blood cholesterol levels and cholesterol metabolic kinetics were reported so far. Failure to identify a consistent association may be due to variations in populations examined, including healthy, hypercholesterolemic, and overweight/obese subjects; modulating environmental factors such as diet or pharmaceutical treatments; or simply, a lack of power to allow for any robust conclusion to be drawn. No modulating effect was observed from T400K, A632V or Y54C on cholesterol lowering response to atorvastatin (25, 26).

Recently, a series of GWA studies was conducted in Caucasian populations. The minor T allele of rs6756629 (R50C, C>T) in ABCG5 was reported to be associated with lower levels of total cholesterol ($p = 1.5\times10^{-11}$) and LDL cholesterol ($p = 2.6\times10^{-10}$) compared to the major allele (11). The minor A allele of rs4953023 (G>A) in ABCG8 was associated with lower levels of LDL cholesterol ($p = 4\times10^{-8}$) compared to the major allele (12). Based on the HapMap CEU data, these SNPs are in complete LD with each other and with D19H that showed similar association result. The minor T allele of rs6544713 (C>T) in ABCG8 was associated with higher LDL cholesterol levels ($p = 2\times10^{-29}$) compared to the major allele (12). Rs6544713 is not in LD with the above three SNPs ($r2 < 0.03$), suggesting that it contributed an independent association.

Common polymorphisms in ABCG5/G8 were also studied in non-Caucasian populations. In 100 hypercholesterolaemic Japanese subjects, Miwa et al. (56) reported that carriers of the M429V variant of ABCG8 or a specific haplotype (wild-type allele of Q604E ABCG5, and wild-type allele of C54Y, wild-type allele of T400K, mutant allele of M429V ABCG8) had higher cholesterol absorption efficiency than non-carriers. However, no difference was observed in serum lipid profiles in relation to common polymorphisms studied previously in Caucasian populations [ABCG5 (Q604E) and ABCG8 (A632V, T400K, D19H and C54Y)]. This might be explained by the fact that carriers of ABCG8 D19H and A632V polymorphisms are rare among Japanese compared to Caucasian populations. Interestingly, in 1046 Chinese, Chen et al. (57) observed that the heterozygote D19H of ABCG8 had a higher serum total and LDL cholesterol levels than homozygote DD, which is opposite to the effect observed in Caucasian populations. The author speculated that this opposite effect may be due to the specific Chinese dietary pattern with high intake of plant sterols. No association with C54Y and T400K of ABCG8 regarding total and LDL cholesterol levels was observed. A632V of ABCG8 was monomorphic in this Chinese population.

Recently, in 845 self-identified Puerto Ricans from Boston, Junyent et al. (47) reported that ABCG5/G8 (i7892T>C, rs4131229; 5U145A>C, rs3806471; Y54C; T400K) SNPs were significantly associated with HDL-C concentrations. Carriers of the minor alleles at these loci and homozygotes for the T400 allele displayed lower HDL cholesterol levels. A significant gene-smoking interaction was also found. Carriers of the minor alleles at ABCG5/G8 (Q604E; D19H; i14222A>G, rs6709904) SNPs displayed lower levels of HDL cholesterol only if they were smokers. Also, for ABCG8 T400K, smokers, but not nonsmokers, homozygous for the T allele displayed lower HDL cholesterol levels. The result further supported a significant haplotype global effect on lowering HDL cholesterols among smokers. The association between polymorphisms in ABCG5/G8 and plasma HDL cholesterol levels possibly reconciles with an old concept: hepatobililarily excreted cholesterol mainly originates from HDL-derived cholesterol (58). However, these associations were not observed in some earlier studies conducted in Caucasian populations with relatively large sample size (43, 45, 46) and also in recent GWA studies (Supplement table 2) (11, 12). Therefore, this issue needs further exploration.

3. NPC1L1

Niemann-Pick C1-like 1 (NPC1L1) is essential to the body's ability to absorb cholesterol (59). The main function of NPC1L1 is to mediate the absorption of cholesterol, especially free cholesterol in the small intestine. Although there are conflicting reports about the relative expression level of NPC1L1 in human liver compared to the small intestine (59-61), NPC1L1 on the canalicular membrane of the hepatocyte is postulated to have a role in conserving cholesterol from what would be otherwise unopposed efflux by ABCG5 and ABCG8 transporters in the liver (62). Both rare mutations and common polymorphisms have been reported in the NPC1L1 gene and associated with blood cholesterol levels, cholesterol kinetics and LDL cholesterol lowering response to ezetimibe.

In the Dallas Heart Study, Cohen et al. (63) demonstrated that in African-Americans, rare non-synonymous genetic variations in NPC1L1 cumulatively contribute to the variability in cholesterol uptake and plasma LDL cholesterol levels. Nineteen nonsynonymous variants and one nonsense mutation were only present in low absorbers of cholesterol and not in high absorbers. These variants were found in 6.2% of African-Americans and were less frequent in Whites (1.8%) or Hispanics (1.7%). Average plasma LDL cholesterol levels were reduced with about 10% among African-Americans with one of the variant NPC1L1 alleles. The variations were mostly found among sequences that are highly conserved across the phylogeny (63), most of which were found to interfere with the formation of a stable protein in a follow-up functional study (64). Interestingly, one of the rare variants that was only found in the high absorbers in the Dallas Heart Study, I1233N (63), was also identified as one of the compound mutations in a male Caucasian with hyperlipidemia who did not respond to ezetimibe treatment (65). This could indicate that the residual at I1233N may be one of the critical residuals upon which ezetimibe acts.

Besides these rare mutations, polymorphisms in the NPC1L1 gene have been associated with variation in LDL cholesterol lowering response to ezetimibe (65-68). Simon et al. (66) identified a 3 SNP haplotype (consisting of g.-133A>G, g.-18C>A and g.1679C>G

(rs2072183, L272L)) that was associated with responsiveness to ezetimibe in two independent studies: carriers of the A-A-G haplotype had a significantly better response to ezetimibe compared to non-carriers. The minor allele of one of the SNPs in this haplotype (rs2072183) was subsequently reported to be associated with increased response to ezetimibe in two small cohorts of hyperlipidemic individuals (68). This result is also supported by an early study. Hegele et al. (67) reported that subjects carrying the haplotype consisting of all the major alleles of rs2072183, g.25342A>C (rs217428), and g.27677T>C (rs217434) had a smaller reduction in plasma LDL cholesterol in response to ezetimibe compared to non-carriers.

However, common variants in NPC1L1 were not associated with cholesterol levels before treatment (66, 67), or in a general population (63). Recently, Chan et al. (69) demonstrated that the NPC1L1 2/2 haplotype (consisting of all the majors alleles of rs2072183, rs217428 and rs217434) was associated with increased baseline plasma total and LDL cholesterol levels, apoB levels, and LDL-apoB pool size, and decreased LDL-apoB fractional catabolic rate (FCR). In addition, 2/2 haplotype carriers showed a greater reduction in plasma levels of total cholesterol and apoB levels and in LDL-apoB pool size, as well as a greater increase in LDL-apoB FCR after atorvastatin treatment over 6 weeks compared to non-2/2 subjects. However, rs217428 and rs217434 were not associated with LDL cholesterol response to atorvastatin in a GWA study (26), and also, none of the SNPs (g.-133A>G, g.-18C>A and rs2072183) showed an association with the LDL cholesterol response to statin treatment in a study by Simon et al. (66). In one study conducted in Chinese subjects, Chen et al. (70) reported that two common SNPs (g.-762T>C; g.1679C>G, rs2072183) were found among 50 Chinese through re-sequencing and they were in high LD. The C allele of -762T>C had a higher promoter activity based on luciferase assay. In 224 Chinese subjects, the -762C allele carriers had higher serum total and LDL cholesterol levels than TT homozygotes. It should be noted that regarding association with baseline blood cholesterol levels, the sample size of the two studies described above (69, 70) were still small. Although no SNP in NPC1L1 has reached the GWA significance-level in relation to blood LDL cholesterol levels in recent GWA studies (supplement table 1) (11, 12), considering the consistent gene-drug (ezetimibe) interaction regarding LDL cholesterol lowering response, the effect from these common genetic variants may only become manifest under stressed conditions such as ezetimibe treatment.

4. LXRα

Liver X receptor (LXR) is the major transcription factor that acts as a sensor of cholesterol levels via its interaction with oxysterols, which in turn, drives the disposal of excess cholesterol. Although there is a clear role of LXRα in inducing CYP7A1 in mice after a challenge with diets high in cholesterol (33), LXRα induction of CYP7A1 does not seem to occur in humans (71, 72). Nevertheless, LXRα is highly expressed in the liver and also abundant in the intestine (3, 73), and it potentially regulates the expression of ABCG5 and ABCG8 in the liver and intestine (73, 74).

No phenotypic effect due to rare mutation in LXRα has been reported so far. One study conducted in 732 French-Canadians studied common polymorphisms in the LXRα gene

(75). Plasma total cholesterol levels were higher in carriers of the minor -115A, -840A and -1830C allele, compared with the -115G/G, -840C/C, and -1830T/T homozygotes at c.-115G>A (rs12221497), c.-840C>A (rs61896015) and c.-1830T>C (rs3758674). No additional haplotype analysis was pursued due to the complete LD among the three SNPs. When continuous dietary intake of cholesterol was put into a model to check the its potential interaction with the three genotypes, LXRα -115G>A explained 1.8% and 2.1% of the variance in total cholesterol and LDL cholesterol levels, respectively, whereas the interaction term explained 2.9% and 2.8%, respectively. When subjects were divided into two groups according to the median intake of cholesterol (290.8mg) and -115G>A genotype, there was a positive association between dietary cholesterol intake and plasma total cholesterol and LDL cholesterol levels in -115A carriers, but not in -115G/G homozygotes. These results suggest that dietary cholesterol intake interacts with LXRα variants to modulate plasma total cholesterol and LDL cholesterol levels. In another study conducted in 2290 French subjects, carriers of the minor allele (-6A) of rs11039115 was associated with increased plasma HDL cholesterol levels in a dominant model (GG vs GA+AA) (76). This SNP is also located in the LXRα promoter region, but isn't in LD with the three SNPs mentioned above.

Recently, in a GWA study conducted in a Finnish birth cohort derived from a genetic isolate (n = 4763) (13), two SNPs (rs2167079 and rs7120118 in complete LD) in LXRα were associated with plasma HDL cholesterol levels ($p = 2.23\times10^{-8}$ and 1.53×10^{-8}, respectively). The association between genetic variants in LXRα and plasma HDL cholesterol levels are understandable since LXRα is also highly expressed in adipose tissue and macrophages (3, 73), and it directly regulates genes involved in HDL lipoprotein metabolism, such as ABCA1, ABCG1, LPL, PLTP and CETP (73). To what extent this association is mediated through ABCG5/G8 is still unclear.

5. HNF4α

Hepatocyte nuclear factor 4α (HNF4α) plays a crucial role in the development and function of vital organs, such as the liver, colon, and pancreas. It is a major regulator of genes involved in the control of lipid homeostasis (77). HNF4α is involved in regulating expressions of BA biosynthetic enzymes (CYP7A1, CYP7B1, CYP8B1 and CYP27A1) (33, 34, 77-79), and also BA transporter of NTCP (77, 80). Mutations in HNF4α in humans are directly linked to the onset of MODY1 (maturity onset diabetes of the young 1) (81, 82), and associated with variations in HDL cholesterol levels (81, 83). Common genetic variant in HNF4α (rs1800961) is associated with plasma HDL cholesterol levels in both candidate association study (84) and GWA study ($p = 8\times10^{-10}$) (12). Since HNF4α also regulates hepatic expression of a number of genes directly involved in HDL lipoprotein metabolism, including apoA-I, apoA-II and SR-BI (77), it is not clear to what extent the HDL cholesterol modulating effect of HNF4α is mediated through BA metabolic pathway.

6. Other genes

Farnesoid X receptor (FXR) is the transcription factor that senses the intracellular levels of BAs. It inhibits BA synthesis via downregulation of CYP7A1, CYP8B1 and CYP27A1 (33, 34, 79). The expression of BA transporters (NTCP, BSEP, ASBT, OSTα and OSTβ; for

their exact functions, see Figure 1) is, to a considerable extent, regulated by FXR (34, 79, 85). Fibroblast growth factor 19 (FGF19), the recent discovered putative intestine-derived molecule regulating BA synthesis is also regulated by FXR (79, 86). Because of the central role of FXR in the maintenance of BA homeostasis, genetic variations in its sequence have been related to several dysregulations of BA metabolism, such as intrahepatic cholestasis of pregnancy [-1G>T, +1A>G (M1V), and 518T>C (M173T)] (87) and cholelithiasis (-20647T>G, -1G>T and IVS7 -31A>T) (88). Recently, the role of FXR in LDL and HDL cholesterol metabolisms has been reviewed (34). However, common genetic variants in FXR seem not to have a big influence on blood cholesterol levels based on the recent meta-analysis of seven GWA studies (Supplement table 1 and 2).

LXR and FXR form obligate heterodimers with RXR proteins. The activities of the heterodimeric LXR and FXR receptors with respect to the regulation of cholesterol and BA metabolism are modulated by ligands that interact with the RXR subunit (89). Since RXRα is the predominant heterodimeration partner for nuclear receptors in the liver, its genetic variants have also been reviewed. Only one study on plasma cholesterol levels was reported (90). Three SNPs (rs3132293, rs3118570 and rs1536475) and their haplotypes were analyzed in 405 Alzheimer's patients and 347 controls. None of the SNPs or haplotypes were associated with plasma cholesterol levels. However, based on the recent meta-analysis of seven GWA studies (Supplement table 2), it seems that common genetic variants in RXRα could potentially affect plasma HDL cholesterol levels.

The pregnane X receptor (PXR), a nuclear receptor, inhibits CYP7A1 expression (33, 78, 79, 91), and also CYP39A1 expression (92), one of the enzymes involved in alternative BA biosynthesis (33). Clinical activation of PXR is associated with hyperlipidemia and increased blood cholesterol levels (92). Based on the recent meta-analysis of seven GWA studies (Supplement table 1), it seems that common genetic variants in PXR affect plasma LDL cholesterol levels, relatively strongly.

Short heterodimer partner (SHP) and liver receptor homologue-1 (LRH-1), two nuclear receptors, are involved in BA feedback inhibition of BA synthesis mediated by FXR (33, 79). After binding BAs, FXR activates the transcription of SHP that is a common transcriptional repressor of nuclear receptor. SHP subsequently suppresses BA synthesis by directly participating in the inhibition of CYP7A1 expression through chromatin remodelling or binds to LRH-1, which is required for expression of CYP7A1 and CYP8B1 (3, 33, 79). Based on the recent meta-analysis of seven GWA studies (Supplement table 1 and 2), it seems that common genetic variants in SHP and LRH-1 could potentially affect plasma cholesterol levels.

ABCB4 acts as a flippase and translocates phosphatidylcholine across the canalicular membrane (34). Genetic mutations and polymorphism in ABCB4 have been associated with progressive familial intrahepatic cholestasis type 3, gallstone and cholestasis of pregnancy (93). However, no association between ABCB4 genetic variants and blood cholesterol levels has been reported so far.

BA transporters are now understood to play central roles in driving bile flow. The pathophysiological consequences of genetic mutations and polymorphisms in these genes encoding BA transporters have been reviewed recently (80, 85). The effect of genetic variation in SLCO1B1 on statin-mediated cholesterol lowering efficacy has recently been highlighted (94-97). SLCO1B1 encodes OATP1B1, one of the organic anion transporting polypeptides (OATP) expressed on the hepatic sinusoidal membrane and facilitating the uptake of BAs and also statin-class drugs (34, 80, 85). A number of polymorphisms and haplotypes have been identified in SLCO1B1 to affect the pharmacokinetic profiles of the statin-class drugs (98). Two small candidate gene studies conducted in eastern Asian subjects and one GWA study conducted in Caucasian subjects consistently reported that carriers of the C allele for rs4149056 (c.521T>C, V174A) had less reduction of total or LDL cholesterol levels with statin treatment compared to TT homozygotes (94, 95, 97). This could be explained by less efficient uptake of pravastatin, atorvastatin or simvastatin into hepatocytes in the C allele carriers compared to the TT homozygotes (97, 98). This reduced uptake of statins also explained the high risk of myopathy in the C allele carriers due to elevated systemic exposure to simvastatin in high dose users (80 mg simvastatin/day) (97). The V174A variation is predicted to be functionally 'damaging' in both SIFT and PloyPhen (99, 100), explaining the limited hepatic uptake of statins. The C allele of another SNP in SLCO1B1, rs11045819 (c.463C>A, P155T), is also reported to be associated with less total and LDL cholesterol reduction compared to AA genotypes in 420 elderly hypercholesterolemic French subjects after extended-release fluvastatin treatment (96). As reviewed by Lefebvre et al. (34), BAs affect HDL cholesterol levels in both human and animal studies. Given the pivotal role of BSEP in generating bile flow (34, 85), and also the results of recent meta-analysis of seven GWA studies (Supplement table 1 and 2), common genetic variants in BSEP could affect plasma HDL cholesterol levels relatively strongly.

Conclusion

The studies reviewed above dealing with potential effects of genetic variations along bile acid and bile metabolic pathway on blood cholesterol levels and cholesterol metabolic kinetics are summarized in table 1. Together, with the latest results from GWA studies, we have sufficient evidence that genetic variants at CYP7A1, ABCG5, ABCG8, NPC1L1, LXRα and HNF4α loci affect blood cholesterol levels. Some of them also appear to modulate the association between blood cholesterol levels and environmental factors including dietary fat and some drugs (statins or ezetimibe).

We note that conflicting results were reported for some genetic variants. This may be due to either different populations studied, different modulating environmental factors, or modest effect of each studied variant, especially when combined with small sample sizes. Large studies in the future or meta-analysis of all available similar studies (published or not) on these genetic variants may guide interpretation (101). Due to poor coverage of rare and structural variants in current GWA genotyping platforms and the stringent statistic criteria adopted in GWA studies (102), some important genetic variants in the reviewed genes may not have been captured so far. With more detailed cataloguing of these rare and structural variants and more advanced sequencing and genotyping technologies, they could

be well studied in the near future (103). Based on the results from recent meta-analysis of seven GWA studies, some common variants in genes involving regulating cholesterol and BA metabolism, such as RXRα, PXR, SHP and LRH-1 and also one of BA transporter gene, BSEP, could potentially modulate blood cholesterol levels. With more genetic variants in these genes explored in the future, new information will be obtained. Since the variants in the genes along the bile metabolic pathway exert their effect on cholesterol metabolism over a lifetime, all together, these that are repeatedly confirmed to be associated with lipid levels or response to lipid-lowering treatment may eventually be used in genetic risk scores to predict lipid levels or to individualize therapeutic options. The ultimate goal will be to optimize the risk-benefit ratio for therapies that decrease the cardiovascular risk.

Acknowledgments:

The authors thank Mark Boekschoten for critical reading of this manuscript. This work is supported by grant 2006B195 of the Netherlands Heart Foundation.

Supplementary data:

http://www.sciencedirect.com/science/article/pii/S0021915009009095

Reference:

1. Kruit, J.K., Plosch, T., Havinga, R., Boverhof, R., Groot, P.H., Groen, A.K., and Kuipers, F. 2005. Increased fecal neutral sterol loss upon liver X receptor activation is independent of biliary sterol secretion in mice. Gastroenterology 128:147-156.
2. van der Velde, A.E., Vrins, C.L., van den Oever, K., Kunne, C., Oude Elferink, R.P., Kuipers, F., and Groen, A.K. 2007. Direct intestinal cholesterol secretion contributes significantly to total fecal neutral sterol excretion in mice. Gastroenterology 133:967-975.
3. Repa, J.J., and Mangelsdorf, D.J. 2000. The role of orphan nuclear receptors in the regulation of cholesterol homeostasis. *Annu Rev Cell Dev Biol* 16:459-481.
4. Rader, D.J. 2006. Molecular regulation of HDL metabolism and function: implications for novel therapies. *J Clin Invest* 116:3090-3100.
5. Lewis, G.F. 2006. Determinants of plasma HDL concentrations and reverse cholesterol transport. *Curr Opin Cardiol* 21:345-352.
6. Pullinger, C.R., Eng, C., Salen, G., Shefer, S., Batta, A.K., Erickson, S.K., Verhagen, A., Rivera, C.R., Mulvihill, S.J., Malloy, M.J., et al. 2002. Human cholesterol 7alpha-hydroxylase (CYP7A1) deficiency has a hypercholesterolemic phenotype. *J Clin Invest* 110:109-117.
7. Bhattacharyya, A.K., and Connor, W.E. 1974. Beta-sitosterolemia and xanthomatosis. A newly described lipid storage disease in two sisters. *J Clin Invest* 53:1033-1043.
8. Bjorkhem, K., Boberg, K., and Leitersdorf, E. 2001. Inborn errors in bile acid biosynthesis and storage of sterols other than cholesterol. In: Scriver CR, Beaudet AL, Valle DS, Sly W, editors. The Metabolic and Molecular Bases of Inherited Disease. New York:McGraw-Hill p2961-2988.
9. Hubacek, J.A., Berge, K.E., Cohen, J.C., and Hobbs, H.H. 2001. Mutations in ATP-cassette binding proteins G5 (ABCG5) and G8 (ABCG8) causing sitosterolemia. *Hum Mutat* 18:359-360.
10. Oelkers, P., Kirby, L.C., Heubi, J.E., and Dawson, P.A. 1997. Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). *J Clin Invest* 99:1880-1887.

11. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., et al. 2009. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 41:47-55.
12. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., et al. 2009. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41:56-65.
13. Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M., et al. 2009. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 41:35-46.
14. Cohen, J.C. 1999. Contribution of cholesterol 7alpha-hydroxylase to the regulation of lipoprotein metabolism. *Curr Opin Lipidol* 10:303-307.
15. Wang, J., Freeman, D.J., Grundy, S.M., Levine, D.M., Guerra, R., and Cohen, J.C. 1998. Linkage between cholesterol 7alpha-hydroxylase and high plasma low-density lipoprotein cholesterol concentrations. *J Clin Invest* 101:1283-1291.
16. Abrahamsson, A., Krapivner, S., Gustafsson, U., Muhrbeck, O., Eggertsen, G., Johansson, I., Persson, I., Angelin, B., Ingelman-Sundberg, M., Bjorkhem, I., et al. 2005. Common polymorphisms in the CYP7A1 gene do not contribute to variation in rates of bile acid synthesis and plasma LDL cholesterol concentration. *Atherosclerosis* 182:37-45.
17. Nakamoto, K., Wang, S., Jenison, R.D., Guo, G.L., Klaassen, C.D., Wan, Y.J., and Zhong, X.B. 2006. Linkage disequilibrium blocks, haplotype structure, and htSNPs of human CYP7A1 gene. *BMC Genet* 7:29.
18. Couture, P., Otvos, J.D., Cupples, L.A., Wilson, P.W., Schaefer, E.J., and Ordovas, J.M. 1999. Association of the A-204C polymorphism in the cholesterol 7alpha-hydroxylase gene with variations in plasma low density lipoprotein cholesterol levels in the Framingham Offspring Study. *J Lipid Res* 40:1883-1889.
19. Hofman, M.K., Groenendijk, M., Verkuijlen, P.J., Jonkers, I.J., Mohrschladt, M.F., Smelt, A.H., and Princen, H.M. 2004. Modulating effect of the A-278C promoter polymorphism in the cholesterol 7alpha-hydroxylase gene on serum lipid levels in normolipidaemic and hypertriglyceridaemic individuals. *Eur J Hum Genet* 12:935-941.
20. Hofman, M.K., Princen, H.M., Zwinderman, A.H., and Jukema, J.W. 2005. Genetic variation in the rate-limiting enzyme in cholesterol catabolism (cholesterol 7alpha-hydroxylase) influences the progression of atherosclerosis and risk of new clinical events. *Clin Sci (Lond)* 108:539-545.
21. Hegele, R.A., Wang, J., Harris, S.B., Brunt, J.H., Young, T.K., Hanley, A.J., Zinman, B., Connelly, P.W., and Anderson, C.M. 2001. Variable association between genetic variation in the CYP7 gene promoter and plasma lipoproteins in three Canadian populations. *Atherosclerosis* 154:579-587.
22. Han, Z., Heath, S.C., Shmulewitz, D., Li, W., Auerbach, S.B., Blundell, M.L., Lehner, T., Ott, J., Stoffel, M., Friedman, J.M., et al. 2002. Candidate genes involved in cardiovascular risk factors by a family-based association study on the island of Kosrae, Federated States of Micronesia. *Am J Med Genet* 110:234-242.
23. Klos, K.L., Sing, C.F., Boerwinkle, E., Hamon, S.C., Rea, T.J., Clark, A., Fornage, M., and Hixson, J.E. 2006. Consistent effects of genes involved in reverse cholesterol transport on plasma lipid and apolipoprotein levels in CARDIA participants. *Arterioscler Thromb Vasc Biol* 26:1828-1836.
24. Kajinami, K., Brousseau, M.E., Ordovas, J.M., and Schaefer, E.J. 2005. A promoter polymorphism in cholesterol 7alpha-hydroxylase interacts with apolipoprotein E genotype in the LDL-lowering response to atorvastatin. *Atherosclerosis* 180:407-415.

25. Kajinami, K., Brousseau, M.E., Ordovas, J.M., and Schaefer, E.J. 2004. Interactions between common genetic polymorphisms in ABCG5/G8 and CYP7A1 on LDL cholesterol-lowering response to atorvastatin. *Atherosclerosis* 175:287-293.
26. Thompson, J.F., Hyde, C.L., Wood, L.S., Paciga, S.A., Hinds, D.A., Cox, D.R., Hovingh, G.K., and Kastelein, J.J.P. 2009. Comprehensive Whole-Genome and Candidate Gene Analysis for Response to Statin Therapy in the Treating to New Targets (TNT) Cohort *Circ Cardiovasc Genet* 2:173-181.
27. Hofman, M.K., Weggemans, R.M., Zock, P.L., Schouten, E.G., Katan, M.B., and Princen, H.M. 2004. CYP7A1 A-278C polymorphism affects the response of plasma lipids after dietary cholesterol or cafestol interventions in humans. *J Nutr* 134:2200-2204.
28. Kovar, J., Suchanek, P., Hubacek, J.A., and Poledne, R. 2004. The A-204C polymorphism in the cholesterol 7alpha-hydroxylase (CYP7A1) gene determines the cholesterolemia responsiveness to a high-fat diet. *Physiol Res* 53:565-568.
29. Hubacek, J.A., Pitha, J., Skodova, Z., Poledne, R., Lanska, V., Waterworth, D.M., Humphries, S.E., and Talmud, P.J. 2003. Polymorphisms in CYP-7A1, not APOE, influence the change in plasma lipids in response to population dietary change in an 8 year follow-up; results from the Czech MONICA study. *Clin Biochem* 36:263-267.
30. Molowa, D.T., Chen, W.S., Cimis, G.M., and Tan, C.P. 1992. Transcriptional regulation of the human cholesterol 7 alpha-hydroxylase gene. *Biochemistry* 31:2539-2544.
31. Wang, D.P., Stroup, D., Marrapodi, M., Crestani, M., Galli, G., and Chiang, J.Y. 1996. Transcriptional regulation of the human cholesterol 7 alpha-hydroxylase gene (CYP7A) in HepG2 cells. *J Lipid Res* 37:1831-1841.
32. Lenicek, M., Komarek, V., Zimolova, M., Kovar, J., Jirsa, M., Lukas, M., and Vitek, L. 2008. CYP7A1 promoter polymorphism -203A>C affects bile salt synthesis rate in patients after ileal resection. *J Lipid Res* 49:2664-2667.
33. Russell, D.W. 2003. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem* 72:137-174.
34. Lefebvre, P., Cariou, B., Lien, F., Kuipers, F., and Staels, B. 2009. Role of bile acids and bile acid receptors in metabolic regulation. *Physiol Rev* 89:147-191.
35. Verrips, A., Hoefsloot, L.H., Steenbergen, G.C., Theelen, J.P., Wevers, R.A., Gabreels, F.J., van Engelen, B.G., and van den Heuvel, L.P. 2000. Clinical and molecular genetic characteristics of patients with cerebrotendinous xanthomatosis. *Brain* 123 (Pt 5):908-919.
36. Setchell, K.D., Schwarz, M., O'Connell, N.C., Lund, E.G., Davis, D.L., Lathe, R., Thompson, H.R., Weslie Tyson, R., Sokol, R.J., and Russell, D.W. 1998. Identification of a new inborn error in bile acid synthesis: mutation of the oxysterol 7alpha-hydroxylase gene causes severe neonatal liver disease. *J Clin Invest* 102:1690-1703.
37. Goizet, C., Boukhris, A., Durr, A., Beetz, C., Truchetto, J., Tesson, C., Tsaousidou, M., Forlani, S., Guyant-Marechal, L., Fontaine, B., et al. 2009. CYP7B1 mutations in pure and complex forms of hereditary spastic paraplegia type 5. *Brain* 132:1589-1600.
38. Miettinen, T.A., Klett, E.L., Gylling, H., Isoniemi, H., and Patel, S.B. 2006. Liver transplantation in a patient with sitosterolemia and cirrhosis. *Gastroenterology* 130:542-547.
39. Hazard, S.E., and Patel, S.B. 2007. Sterolins ABCG5 and ABCG8: regulators of whole body dietary sterols. *Pflugers Arch* 453:745-752.
40. Pandit, B., Ahn, G.S., Hazard, S.E., Gordon, D., and Patel, S.B. 2006. A detailed Hapmap of the Sitosterolemia locus spanning 69 kb; differences between Caucasians and African-Americans. *BMC Med Genet* 7:13.
41. Rudkowska, I., and Jones, P.J. 2008. Polymorphisms in ABCG5/G8 transporters linked to hypercholesterolemia and gallstone disease. *Nutr Rev* 66:343-348.

42. Weggemans, R.M., Zock, P.L., Tai, E.S., Ordovas, J.M., Molhuizen, H.O., and Katan, M.B. 2002. ATP binding cassette G5 C1950G polymorphism may affect blood cholesterol concentrations in humans. *Clin Genet* 62:226-229.
43. Viturro, E., de Oya, M., Lasuncion, M.A., Gorgojo, L., Moreno, J.M., Benavente, M., Cano, B., and Garces, C. 2006. Cholesterol and saturated fat intake determine the effect of polymorphisms at ABCG5/ABCG8 genes on lipid levels in children. *Genet Med* 8:594-599.
44. Berge, K.E., von Bergmann, K., Lutjohann, D., Guerra, R., Grundy, S.M., Hobbs, H.H., and Cohen, J.C. 2002. Heritability of plasma noncholesterol sterols and relationship to DNA sequence polymorphism in ABCG5 and ABCG8. *J Lipid Res* 43:486-494.
45. Gylling, H., Hallikainen, M., Pihlajamaki, J., Agren, J., Laakso, M., Rajaratnam, R.A., Rauramaa, R., and Miettinen, T.A. 2004. Polymorphisms in the ABCG5 and ABCG8 genes associate with cholesterol absorption and insulin sensitivity. *J Lipid Res* 45:1660-1665.
46. Hubacek, J.A., Berge, K.E., Stefkova, J., Pitha, J., Skodova, Z., Lanska, V., and Poledne, R. 2004. Polymorphisms in ABCG5 and ABCG8 transporters and plasma cholesterol levels. *Physiol Res* 53:395-401.
47. Junyent, M., Tucker, K.L., Smith, C.E., Garcia-Rios, A., Mattei, J., Lai, C.Q., Parnell, L.D., and Ordovas, J.M. 2009. The effects of ABCG5/G8 polymorphisms on plasma HDL cholesterol concentrations depend on smoking habit in the Boston Puerto Rican Health Study. *J Lipid Res* 50:565-573.
48. Santosa, S., Demonty, I., Lichtenstein, A.H., Ordovas, J.M., and Jones, P.J. 2007. Single nucleotide polymorphisms in ABCG5 and ABCG8 are associated with changes in cholesterol metabolism during weight loss. *J Lipid Res* 48:2607-2613.
49. Acalovschi, M., Ciocan, A., Mostean, O., Tirziu, S., Chiorean, E., Keppeler, H., Schirin-Sokhan, R., and Lammert, F. 2006. Are plasma lipid levels related to ABCG5/ABCG8 polymorphisms? A preliminary study in siblings with gallstones. *Eur J Intern Med* 17:490-494.
50. Plat, J., Bragt, M.C., and Mensink, R.P. 2005. Common sequence variations in ABCG8 are related to plant sterol metabolism in healthy volunteers. *J Lipid Res* 46:68-75.
51. Herron, K.L., McGrane, M.M., Waters, D., Lofgren, I.E., Clark, R.M., Ordovas, J.M., and Fernandez, M.L. 2006. The ABCG5 polymorphism contributes to individual responses to dietary cholesterol and carotenoids in eggs. *J Nutr* 136:1161-1165.
52. Kajinami, K., Brousseau, M.E., Nartsupha, C., Ordovas, J.M., and Schaefer, E.J. 2004. ATP binding cassette transporter G5 and G8 genotypes and plasma lipoprotein levels before and after treatment with atorvastatin. *J Lipid Res* 45:653-656.
53. Koeijvoets, K.C., van der Net, J.B., Dallinga-Thie, G.M., Steyerberg, E.W., Mensink, R.P., Kastelein, J.J., Sijbrands, E.J., and Plat, J. 2008. ABCG8 gene polymorphisms, plasma cholesterol concentrations, and risk of cardiovascular disease in familial hypercholesterolemia. *Atherosclerosis*.
54. Buch, S., Schafmayer, C., Volzke, H., Becker, C., Franke, A., von Eller-Eberstein, H., Kluck, C., Bassmann, I., Brosch, M., Lammert, F., et al. 2007. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet* 39:995-999.
55. Chan, D.C., Watts, G.F., Barrett, P.H., Whitfield, A.J., and van Bockxmeer, F.M. 2004. ATP-binding cassette transporter G8 gene as a determinant of apolipoprotein B-100 kinetics in overweight men. *Arterioscler Thromb Vasc Biol* 24:2188-2191.
56. Miwa, K., Inazu, A., Kobayashi, J., Higashikata, T., Nohara, A., Kawashiri, M., Katsuda, S., Takata, M., Koizumi, J., and Mabuchi, H. 2005. ATP-binding cassette transporter G8 M429V polymorphism as a novel genetic marker of higher cholesterol absorption in hypercholesterolaemic Japanese subjects. *Clin Sci (Lond)* 109:183-188.

57. Chen, Z.C., Shin, S.J., Kuo, K.K., Lin, K.D., Yu, M.L., and Hsiao, P.J. 2008. Significant association of ABCG8:D19H gene polymorphism with hypercholesterolemia and insulin resistance. *J Hum Genet* 53:757-763.
58. Carey, M.C. 1997. Homing-in on the origin of biliary steroids. *Gut* 41:721-722.
59. Altmann, S.W., Davis, H.R., Jr., Zhu, L.J., Yao, X., Hoos, L.M., Tetzloff, G., Iyer, S.P., Maguire, M., Golovko, A., Zeng, M., et al. 2004. Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption. *Science* 303:1201-1204.
60. Davies, J.P., Scott, C., Oishi, K., Liapis, A., and Ioannou, Y.A. 2005. Inactivation of NPC1L1 causes multiple lipid transport defects and protects against diet-induced hypercholesterolemia. *J Biol Chem* 280:12710-12720.
61. Zuniga, S., Molina, H., Azocar, L., Amigo, L., Nervi, F., Pimentel, F., Jarufe, N., Arrese, M., Lammert, F., and Miquel, J.F. 2008. Ezetimibe prevents cholesterol gallstone formation in mice. *Liver Int* 28:935-947.
62. Temel, R.E., Tang, W., Ma, Y., Rudel, L.L., Willingham, M.C., Ioannou, Y.A., Davies, J.P., Nilsson, L.M., and Yu, L. 2007. Hepatic Niemann-Pick C1-like 1 regulates biliary cholesterol concentration and is a target of ezetimibe. *J Clin Invest* 117:1968-1978.
63. Cohen, J.C., Pertsemlidis, A., Fahmi, S., Esmail, S., Vega, G.L., Grundy, S.M., and Hobbs, H.H. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A* 103:1810-1815.
64. Fahmi, S., Yang, C., Esmail, S., Hobbs, H.H., and Cohen, J.C. 2008. Functional characterization of genetic variants in NPC1L1 supports the sequencing extremes strategy to identify complex trait genes. *Hum Mol Genet* 17:2101-2107.
65. Wang, J., Williams, C.M., and Hegele, R.A. 2005. Compound heterozygosity for two non-synonymous polymorphisms in NPC1L1 in a non-responder to ezetimibe. *Clin Genet* 67:175-177.
66. Simon, J.S., Karnoub, M.C., Devlin, D.J., Arreaza, M.G., Qiu, P., Monks, S.A., Severino, M.E., Deutsch, P., Palmisano, J., Sachs, A.B., et al. 2005. Sequence variation in NPC1L1 and association with improved LDL-cholesterol lowering in response to ezetimibe treatment. *Genomics* 86:648-656.
67. Hegele, R.A., Guy, J., Ban, M.R., and Wang, J. 2005. NPC1L1 haplotype is associated with inter-individual variation in plasma low-density lipoprotein response to ezetimibe. *Lipids Health Dis* 4:16.
68. Pisciotta, L., Fasano, T., Bellocchio, A., Bocchi, L., Sallo, R., Fresa, R., Colangeli, I., Cantafora, A., Calandra, S., and Bertolini, S. 2007. Effect of ezetimibe coadministered with statins in genotype-confirmed heterozygous FH patients. *Atherosclerosis* 194:e116-122.
69. Chan, D.C., Watts, G.F., Wang, J., Hegele, R.A., van Bockxmeer, F.M., and Barrett, P.H. 2008. Variation in Niemann-Pick C1-like 1 gene as a determinant of apolipoprotein B-100 kinetics and response to statin therapy in centrally obese men. *Clin Endocrinol (Oxf)* 69:45-51.
70. Chen, C.W., Hwang, J.J., Tsai, C.T., Su, Y.N., Hsueh, C.H., Shen, M.J., and Lai, L.P. 2009. The g.-762T>C polymorphism of the NPC1L1 gene is common in Chinese and contributes to a higher promoter activity and higher serum cholesterol levels. *J Hum Genet*.
71. Chiang, J.Y., Kimmel, R., and Stroup, D. 2001. Regulation of cholesterol 7alpha-hydroxylase gene (CYP7A1) transcription by the liver orphan receptor (LXRalpha). *Gene* 262:257-265.
72. Goodwin, B., Watson, M.A., Kim, H., Miao, J., Kemper, J.K., and Kliewer, S.A. 2003. Differential regulation of rat and human CYP7A1 by the nuclear oxysterol receptor liver X receptor-alpha. *Mol Endocrinol* 17:386-394.
73. Repa, J.J., and Mangelsdorf, D.J. 2002. The liver X receptor gene team: potential new players in atherosclerosis. *Nat Med* 8:1243-1248.

74. Yu, L., York, J., von Bergmann, K., Lutjohann, D., Cohen, J.C., and Hobbs, H.H. 2003. Stimulation of cholesterol excretion by the liver X receptor agonist requires ATP-binding cassette transporters G5 and G8. *J Biol Chem* 278:15565-15570.
75. Robitaille, J., Houde, A., Lemieux, S., Gaudet, D., Perusse, L., and Vohl, M.C. 2007. The lipoprotein/lipid profile is modulated by a gene-diet interaction effect between polymorphisms in the liver X receptor-alpha and dietary cholesterol intake in French-Canadians. *Br J Nutr* 97:11-18.
76. Legry, V., Cottel, D., Ferrieres, J., Chinetti, G., Deroide, T., Staels, B., Amouyel, P., and Meirhaeghe, A. 2008. Association between liver X receptor alpha gene polymorphisms and risk of metabolic syndrome in French populations. *Int J Obes (Lond)* 32:421-428.
77. Hayhurst, G.P., Lee, Y.H., Lambert, G., Ward, J.M., and Gonzalez, F.J. 2001. Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. *Mol Cell Biol* 21:1393-1403.
78. Li, T., and Chiang, J.Y. 2005. Mechanism of rifampicin and pregnane X receptor inhibition of human cholesterol 7 alpha-hydroxylase gene transcription. *Am J Physiol Gastrointest Liver Physiol* 288:G74-84.
79. Chiang, J.Y. 2009. Bile acids: regulation of synthesis. *J Lipid Res*.
80. Dawson, P.A., Lan, T., and Rao, A. 2009. Bile acid transporters. *J Lipid Res*.
81. Pearson, E.R., Pruhova, S., Tack, C.J., Johansen, A., Castleden, H.A., Lumb, P.J., Wierzbicki, A.S., Clark, P.M., Lebl, J., Pedersen, O., et al. 2005. Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large European collection. *Diabetologia* 48:878-885.
82. Lu, P., Rha, G.B., Melikishvili, M., Wu, G., Adkins, B.C., Fried, M.G., and Chi, Y.I. 2008. Structural basis of natural promoter recognition by a unique nuclear receptor, HNF4alpha. Diabetes gene product. *J Biol Chem* 283:33685-33697.
83. Zhu, Q., Yamagata, K., Miura, A., Shihara, N., Horikawa, Y., Takeda, J., Miyagawa, J., and Matsuzawa, Y. 2003. T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. *Diabetologia* 46:567-573.
84. Lu, Y., Dolle, M.E., Imholz, S., van 't Slot, R., Verschuren, W.M., Wijmenga, C., Feskens, E.J., and Boer, J.M. 2008. Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations. *J Lipid Res* 49:2582-2589.
85. Kosters, A., and Karpen, S.J. 2008. Bile acid transporters in health and disease. *Xenobiotica* 38:1043-1071.
86. Holt, J.A., Luo, G., Billin, A.N., Bisi, J., McNeill, Y.Y., Kozarsky, K.F., Donahee, M., Wang, D.Y., Mansfield, T.A., Kliewer, S.A., et al. 2003. Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis. *Genes Dev* 17:1581-1591.
87. Van Mil, S.W., Milona, A., Dixon, P.H., Mullenbach, R., Geenes, V.L., Chambers, J., Shevchuk, V., Moore, G.E., Lammert, F., Glantz, A.G., et al. 2007. Functional variants of the central bile acid sensor FXR identified in intrahepatic cholestasis of pregnancy. *Gastroenterology* 133:507-516.
88. Kovacs, P., Kress, R., Rocha, J., Kurtz, U., Miquel, J.F., Nervi, F., Mendez-Sanchez, N., Uribe, M., Bock, H.H., Schirin-Sokhan, R., et al. 2008. Variation of the gene encoding the nuclear bile salt receptor FXR and gallstone susceptibility in mice and humans. *J Hepatol* 48:116-124.
89. Repa, J.J., Turley, S.D., Lobaccaro, J.A., Medina, J., Li, L., Lustig, K., Shan, B., Heyman, R.A., Dietschy, J.M., and Mangelsdorf, D.J. 2000. Regulation of absorption and ABC1-mediated efflux of cholesterol by RXR heterodimers. *Science* 289:1524-1529.

90. Kolsch, H., Lutjohann, D., Jessen, F., Popp, J., Hentschel, F., Kelemen, P., Friedrichs, S., Maier, W., and Heun, R. 2008. RXRA gene variations influence Alzheimer's disease risk and cholesterol metabolism. *J Cell Mol Med*.
91. Staudinger, J.L., Goodwin, B., Jones, S.A., Hawkins-Brown, D., MacKenzie, K.I., LaTour, A., Liu, Y., Klaassen, C.D., Brown, K.K., Reinhard, J., et al. 2001. The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. *Proc Natl Acad Sci U S A* 98:3369-3374.
92. Zhou, C., King, N., Chen, K.Y., and Breslow, J.L. 2009. Activation of pregnane X receptor induces hypercholesterolemia in wild-type and accelerates atherosclerosis in apolipoprotein E deficient mice. *J Lipid Res*.
93. Oude Elferink, R.P., and Paulusma, C.C. 2007. Function and pathophysiological importance of ABCB4 (MDR3 P-glycoprotein). *Pflugers Arch* 453:601-610.
94. Tachibana-Iimori, R., Tabara, Y., Kusuhara, H., Kohara, K., Kawamoto, R., Nakura, J., Tokunaga, K., Kondo, I., Sugiyama, Y., and Miki, T. 2004. Effect of genetic polymorphism of OATP-C (SLCO1B1) on lipid-lowering response to HMG-CoA reductase inhibitors. *Drug Metab Pharmacokinet* 19:375-380.
95. Zhang, W., Chen, B.L., Ozdemir, V., He, Y.J., Zhou, G., Peng, D.D., Deng, S., Xie, Q.Y., Xie, W., Xu, L.Y., et al. 2007. SLCO1B1 521T-->C functional genetic polymorphism and lipid-lowering efficacy of multiple-dose pravastatin in Chinese coronary heart disease patients. *Br J Clin Pharmacol* 64:346-352.
96. Couvert, P., Giral, P., Dejager, S., Gu, J., Huby, T., Chapman, M.J., Bruckert, E., and Carrie, A. 2008. Association between a frequent allele of the gene encoding OATP1B1 and enhanced LDL-lowering response to fluvastatin therapy. *Pharmacogenomics* 9:1217-1227.
97. Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M., and Collins, R. 2008. SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med* 359:789-799.
98. Meyer Zu Schwabedissen, H.E., and Kim, R.B. 2009. Hepatic OATP1B Transporters and Nuclear Receptors PXR and CAR: Interplay, Regulation of Drug Disposition Genes, and Single Nucleotide Polymorphisms. *Mol Pharm*.
99. Ramensky, V., Bork, P., and Sunyaev, S. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30:3894-3900.
100. Ng, P.C., and Henikoff, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31:3812-3814.
101. Hirschhorn, J.N., Lohmueller, K., Byrne, E., and Hirschhorn, K. 2002. A comprehensive review of genetic association studies. *Genet Med* 4:45-61.
102. Sethupathy, P., and Collins, F.S. 2008. MicroRNA target site polymorphisms and human disease. *Trends Genet* 24:489-497.
103. 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., et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747-753.
104. Pearson, T.A., Denke, M.A., McBride, P.E., Battisti, W.P., Brady, W.E., and Palmisano, J. 2005. A community-based, randomized trial of ezetimibe added to statin therapy to attain NCEP ATP III goals for LDL cholesterol in hypercholesterolemic patients: the ezetimibe add-on to statin for effectiveness (EASE) trial. *Mayo Clin Proc* 80:587-595.

Table 1 Summary of studies that have linked polymorphisms in genes along bile acid and bile metabolic pathways to variation in blood cholesterol levels, cholesterol kinetics and their responsiveness to interventions. [a]

Genes	Genetic variant	Reference	Study population	Observed association [b]
CYP7A1	r.-203A> C (rs3808607)	Wang et al. (1998) (15)	24 Caucasian nuclear families and 295 unrelated Caucasians	plasma LDL-C higher in CC than AA
		Couture et al. (1999) (18)	2330 Caucasians from Framingham offspring study	plasma TC and LDL-C higher in AC than AA in men, no association was observed in women
		Hofman et al. (2004) (19)	139 Dutch hypertriglyceridaemic patients	TC lower in C allele carriers than AA
		Abrahamsson et al. (2005)(16)	491 healthy male Caucasians; 179 MI survivors and 186 healthy controls.	no difference in LDL-C among the genotypes
		Hofman et al. (2005) (20)	715 male Dutch patients with coronary atherosclerosis	no difference in TC, LDL-C and HDL-C among the genotypes, but lower reduction in serum TC in C allele carriers than AA in response to pravastatin
		Hegele et al. (2001) (21)	594 Hutterites	higher HDL-C and apoA-I levels in C allele carriers than AA
		Hegele et al. (2001) (21)	190 Keewatin Inuit subjects	lower TC and LDL-C in major C allele carriers than AA
		Hegele et al. (2001) (21)	325 OjiCree subjects	no association was observed
		Han et al. (2002) (22)	1102 Micronesians	higher apoA-I levels in rare CC than AA
		Kajinami et al. (2005) (24)	324 hypercholesterolemic subjects, mainly Caucasians	less LDL-C reduction in C allele carriers after atorvastatin treatment, more striking in men

				and in ε2 or ε4 carriers of APOE
		Kajinami et al.(2004) (25)	337 hypercholesterolemic subjects, mainly Caucasians	less LDL-C reduction in C allele carriers after atorvastatin treatment only in ABCG8 D19H variant carriers or ABCG8 homozygote (54CC, 400TT, and 632AA)
		Hofman et al. (2004) (27)	209 Caucasians for TC response and 179 Caucasians for HDL-C response	higher response of plasma TC and HDL-C in C allele carriers after an increase in dietary cholesterol intake
		Kovar et al. (2004) (28)	11 healthy Czech men (6 CC homozygotes and 5 AA homozygotes)	increased serum TC and LDL-C in CC after a high-fat diet challenge, which was not observed in AA
		Hubacek et al. (2003) (29)	114 Czech males	more reductions in TC and LDL-C in C allele carriers ($p < 0.01$ and $p = 0.07$, respectively)
CYP7A1	rs1023649, rs1023651	Klos et al. (2006) (23)	2054 whites and 1939 blacks	two SNPs associated with TC and LDL-C in black with increased levels in carriers of the rare alleles, no association observed in white
ABCG5	Q604E (rs6720173, C>G)	Weggemans et al. (2002) (42)	486 Dutch subjects	higher TC in EE than wide type allele carriers
		Viturro et al. (2006) (43)	1227 healthy Spanish school children	heterozygotes (CG) higher in TC, LDL-C and apoB levels than CC, but only observed in the 70 boys of lowest tertile of saturated fat intake
		Berge et al. (2002) (44)	142 healthy American Caucasians	No difference in TC
		Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	No association between E allele carriers and wide type homozygotes in TC, LDL-C and HDL-C

		Hubacek et al. (2004) (46)	285 Czech participants	No difference in TC, LDL-C and HDL-C
		Junyent et al. (2009) (47)	845 self-identified Puerto Ricans	No difference in TC, LDL-C and HDL-C
		Santosa et al. (2007) (48)	42 overweight/obese Canadian women	No difference in TC and LDL-C
		Acalovschi et al. (2006) (49)	68 Romanian siblings with gallstone disease	lower TC and higher HDL-C in E allele carriers than QQ
		Plat et al. (2005) (50)	112 healthy Dutch volunteers	lower LDL-C in E allele carriers than QQ, no difference in HDL-C
		Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	lower cholesterol absorption (lower level of serum cholesterol adjusted campesterol and sitosterol) and higher cholesterol synthesis (high levels of serum cholesterol adjusted cholestenol) in E allele carriers than QQ
		Santosa et al. (2007) (48)	35 hypercholesterolemic Canadian women.	larger reduction in cholesterol absorption and greater increase in synthesis in EE than Q carriers during weight loss
		Herron et al. (2006) (51)	91 Caucasian subjects (40 men and 51 premenopausal women)	E allele carriers responded less compared to QQ in TC and LDL-C after 1 more egg consumption/day over 30 days, no difference in HDL-C
		Kajinami et al. (2004) (25)	337 hypercholesterolemic subjects, mainly Caucasians	No modulating effect from Q604E on cholesterol lowering response to atorvastatin
ABCG8	D19H (rs11887534, G>C)	Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	lower TC and LDL-C in H allele carriers than DD, no difference in HDL-C

	Junyent et al. (2009) (47)	845 self-identified Puerto Ricans	lower TC and LDL-C in H allele carriers than DD, no difference in HDL-C
	Santosa et al. (2007) (48)	42 overweight/obese Canadian women	lower TC and LDL-C in H allele carriers than DD
	Acalovschi et al. (2006) (49)	68 Romanian siblings with gallstone disease	lower TC in H allele carriers than DD, no difference in HDL-C
	Kajinami et al. (2004) (52)	338 hypercholesterolemic subjects, mainly Caucasians	lower TC in H allele carriers than DD, no difference in LDL-C and HDL-C
	Koeijvoets et al. (2008) (53)	2012 Dutch patients with heterozygous familial hypercholesteroemia	no difference in TC, LDL-C and HDL-C
	Kajinami et al. (2004) (25)	337 hypercholesterolemic subjects, mainly Caucasians	larger reduction in LDL-C in H allele carriers than DD in response to atorvastatin
	Kajinami et al. (2004) (52)	338 hypercholesterolemic subjects, mainly Caucasians	larger reduction in LDL-C in H allele carriers than DD in response to atorvastatin, no difference in change of TC and HDL-C
	Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	H allele more common in the low cholesterol absorption tertile compared with intermediate and high absorption group
	Berge et al. (2002) (44)	142 healthy American Caucasians	lower cholesterol absorption marker sterols (serum cholesterol adjusted campesterol, sitosterol, and cholestanol levels) in H allele carriers than DD
	Chen et al. (2008) (57)	1046 Chinese recruited from the general population	higher TC and LDL-C in heterozygote D19H than homozygote DD, no difference in HDL-C

		Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	lower cholesterol absorption marker sterols (serum cholesterol adjusted campesterol, sitosterol, and cholestanol levels) and higher cholesterol synthesis marker sterols (serum cholesterol adjusted cholestenol and lathosterol levels) in H allele carriers than DD
ABCG8	T400K(rs4148217, C>A)	Berge et al. (2002) (44)	143 healthy American Caucasians	no difference in TC
		Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	no difference in TC, LDL-C and HDL-C
		Junyent et al. (2009) (47)	845 self-identified Puerto Ricans	no difference in TC and LDL-C, but K allele carriers had higher HDL-C than TT
		Santosa et al. (2007) (48)	42 overweight/obese Canadian women	no difference in TC, LDL-C and HDL-C
		Acalovschi et al. (2006) (49)	68 Romanian siblings with gallstone disease	no difference in TC and HDL-C
		Plat et al. (2005) (50)	112 healthy Dutch volunteers	no difference in LDL-C and HDL-C
		Koeijvoets et al. (2008) (53)	2012 Dutch patients with heterozygous familial hypercholesteroemia	no difference in TC, LDL-C and HDL-C
		Chan et al. (2004) (55)	47 nonsmoking overweight/obese Australian men	no difference in TC, LDL-C and HDL-C
		Hubacek et al. (2004) (46)	285 Czech participants	a smaller decrease in TC and LDL-C in K allele carriers than TT after changing dietary habits (less meat and more vegetable)
		Berge et al. (2002) (44)	143 healthy American Caucasians	lower cholesterol absorption marker sterol (serum cholesterol adjusted sitosterol levels)

				and higher cholesterol synthesis marker sterol (serum cholesterol adjusted desmosterol and lathosterol levels) in K allele carriers than TT
		Plat et al. (2005) (50)	112 healthy Dutch volunteers	higher absorption in campesterol and sitosterol and a stronger inhibitor effect of plant stanol ester on campesterol and sitosterol absorption in TT than K allele carriers
		Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	no difference in plasma cholesterol absorption and synthesis markers
		Kajinami et al. (2004) (25)	337 hypercholesterolemic subjects, mainly Caucasians	no modulating effect from T400K on cholesterol lowering response to atorvastatin
ABCG8	A632V (rs6544718, C>T)	Berge et al. (2002) (44)	143 healthy American Caucasians	higher TC in V allele carriers than AA
		Viturro et al. (2006) (43)	1227 healthy Spanish school children	higher TC, LDL-C and apoB levels in heterozygotes than homozygotes of wide-type allele, but only in the low cholesterol intake group (380 children)
		Hubacek et al. (2004) (46)	285 Czech participants	no difference in TC, LDL-C and HDL-C
		Acalovschi et al. (2006) (49)	68 Romanian siblings with gallstone disease	no difference in TC and HDL-C
		Plat et al. (2005) (50)	112 healthy Dutch volunteers	no difference in LDL-C and HDL-C
		Kajinami et al. (2004) (25)	337 hypercholesterolemic subjects, mainly Caucasians	no modulating effect from A632V on cholesterol lowering response to atorvastatin
ABCG8	Y54C (rs4148211, A>G)	Berge et al. (2002) (44)	139 healthy American Caucasians	no difference in TC

		Gylling et al. (2004) (45)	262 mildly to moderately hypercholesterolemic Finnish subjects	no difference in TC, LDL-C and HDL-C
		Junyent et al. (2009) (47)	845 self-identified Puerto Ricans	no difference in TC and LDL-C, but C allele carriers had a lower HDL-C than YY.
		Santosa et al. (2007) (48)	42 overweight/obese Canadian women	no difference in TC and LDL-C
		Acalovschi et al. (2006) (49)	68 Romanian siblings with gallstone disease	no difference in TC and HDL-C
		Santosa et al. (2007) (48)	42 overweight/obese Canadian women	smaller decline in heterozygous Y54C carriers in cholesterol synthesis than YY during weight loss
		Hubacek et al. (2004) (46)	285 Czech participants	Y allele carriers had larger reduction in TC and LDL-C than the CC, but only in female subjects.
		Kajinami et al. (2004) (25)	337 hypercholesterolemic subjects, mainly Caucasians	no modulating effect from Y54C on cholesterol lowering response to atorvastatin
ABCG8	M429V (A>G)	Miwa et al. (2005) (56)	100 hypercholesterolaemic Japanese subjects	M429V variant associated with higher cholesterol absorption
ABCG8	rs4131229 (T>C), rs3806471 (A>C)	Junyent et al. (2009) (47)	845 self-identified Puerto Ricans	lower HDL-C in rare allele carriers than wild-type homozygotes, no difference in TC and LDL-C
ABCG8	rs6709904 (A>G)	Junyent et al. (2009) (47)	845 self-identified Puerto Ricans	lower LDL-C in rare allele carriers than wild-type homozygotes, no difference in TC and HDL-C
NPC1L1	g.-113A>G-g.-18C>A-g.1679C>G	Simon et al. (2005) (66)	1208 hypercholesterolemic individuals participating in the ezetimibe+statin	carriers of the haplotype consisting of A-A-G had a significantly improved LDL-C response

	(L272L, rs2072183)		treatment arm of the EASE trial (104) and 1132 hypercholesterolemic individuals participating in Vytorin vs. Atorvastatin clinical trial.	to ezetimibe compared to non-carriers.
NPC1L1	L272L (rs2072183,C>G)	Pisciotta et al. (2007) (68)	50 primary hypercholesterolemia intolerant to statin and 65 heterozygous familial hypercholesterolemia	higher prevalence of the G allele in hyper-responders, higher LDL-C reduction in G allele carriers than CC in response to ezetimibe
NPC1L1	g. 1735C>G (rs2072183)-g.25342A>C (rs217428)-g.27677T>C (rs217434)	Hegele et al. (2005) (67)	101 Canadian dyslipidemic subjects	lower reduction in LDL-C in subjects carrying the haplotype consisting of all the major alleles than non-carriers in response to ezetimibe
NPC1L1	rs2072183-rs217428-rs217434	Chan et al (2008) (69)	37 obese Australian	subjects carrying the homozygous haplotype (consisting of all the majors alleles) had increased baseline TC, LDL-C, apoB levels, and LDL-apoB pool size, and decreased LDL-apoB fractional catabolic rate (FCR) compared to non-homozygous haplotype carriers.
	rs2072183-rs217428-rs217434	Chan et al (2008) (69)	37 obese Australian	subjects carrying homozygous haplotype (consisting of all the majors alleles) had a greater reduction in TC and apoB and in LDL-apoB pool size, as well as a greater increase in LDL-apoB FCR after atorvastatin treatment over 6 weeks compared to non-homozygous

				haplotype carriers.
NPC1L1	g.-762T>C(ss119336593) c.-115G>A	Chen et al. (2009) (70)	224 Chinese recruited from the hospital	higher TC and LDL-C in C allele carriers than TT
LXRα	(rs12221497), c.-840C>A (rs61896015), and c.-1830T>C (rs3758674)	Robitaille et al. (75)	732 French-Canadians	higher TC in carriers of the minor -115A, -840A and -1830C allele, compared with the -115G/G, -840C/C, and -1830T/T
LXRα	rs11039115 (c.-6G>A)	Legry et al. (2005) (76)	2290 French subjects	higher HDL-C in carriers of the A allele than GG
HNF4α	rs1800961 (C>T)	Lu et al. (2008) (84)	3575 Dutch subjects	higher HDL-C in CC than carriers of the T allele
SLCO1B1	V174A (rs4149056, T>C)	Tachibana-Iimori et al. (2004) (94)	66 Japanese hyperlipidemic patients	lower TC lowering effect in TC heterozygotes than TT homozygotes with pravastatin, atorvastatin or simvastatin treatment for 2 months
		Zhang et al. (2007) (95)	45 Chinese patients with coronary heart disease	lower TC lowering effect in TC heterozygotes than TT homozygotes with pravastatin treatment for 1 month
	P155T (rs11045819, C>A)	Couvert et al. (2008) (96)	420 French hypercholesterolemic patients	lower TC and LDL-C lowering effect in CC homozygotes than CA heterozygotes or AA homozygotes with extended-release fluvastatin treatment for 2 months

[a] The results from GWA studies not included.
[b] TC: total cholesterol levels, LDL-C: low-density lipoprotein cholesterol levels, HDL-C: high-density lipoprotein cholesterol levels; AA, TT, CC and GG mean the homozygous carriers for the corresponding nucleotides, and it is also true for the letters representing the amino acids.

Chapter 3

Research article

Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations

Yingchang Lu, Martijn E.T. Dollé, Sandra Imholz, Ruben van 't Slot, W.M.M. Verschuren, Cisca Wijmenga, Edith J.M. Feskens, and Jolanda M.A. Boer

J Lipid Res. 2008;49(12):2582-9.

Abstract

The known genetic variants determining plasma HDL cholesterol (HDL-C) levels explain only part of its variation. 384 single nucleotide polymorphisms (SNPs) across 251 genes based on pathways potentially relevant to HDL-C metabolism were selected and genotyped in 3575 subjects from the Doetinchem cohort, which was examined thrice over 11 years. 353 SNPs in 239 genes passed the quality control criteria. Seven SNPs (rs1800777 and rs5882 in CETP; rs3208305, rs328 and rs268 in LPL; rs1800588 in LIPC; rs2229741 in NRIP1) were associated with plasma HDL-C levels with false discovery rate adjusted q-values (FDR_q) < 0.05. Five other SNPs (rs17585739 in SC4MOL, rs11066322 in PTPN11, rs4961 in ADD1, rs6060717 near SCAND1 and rs3213451 in MBTPS2 in women) were associated with plasma HDL-C levels with FDR_q between 0.05 and 0.2. Two less well replicated associations (rs3135506 in APOA5 and rs1800961 in HNF4A) known from the literature were also observed but their significance disappeared after adjustment for multiple testing ($p = 0.008$, FDR_q = 0.221 for rs3135506; $p = 0.018$, FDR_q = 0.338 for rs1800961, respectively). In addition to replication of previous results for candidate genes (CETP, LPL, LIPC, HNF4A and APOA5), we found interesting new candidate SNPs (rs2229741 in NRIP1, rs3213451 in MBTPS2, rs17585739 in SC4MOL, rs11066322 in PTPN11, rs4961 in ADD1 and rs6060717 near SCAND1) for plasma HDL-C levels that should be evaluated further.

Supplementary key words: high density lipoprotein, cholesterol, single nucleotide polymorphism, pathway-driven approach, random coefficient model.

INTRODUCTION

Numerous clinical and epidemiological studies have demonstrated an inverse and independent association between plasma concentrations of HDL cholesterol (HDL-C) and the risk of coronary heart disease (1). The most popular mechanistic explanation has been that HDL functions in reverse cholesterol transport, removing cholesterol from peripheral tissues and delivering it to the liver for biliary excretion and to steroidogenic organs for steroid hormones synthesis (2). Although efflux of cholesterol from macrophages represents only a tiny fraction of overall cellular cholesterol efflux, it is in fact, the most important with regard to antiatherogenic effects (3). More recently, a variety of other functions of HDL have been described, primarily based on in vitro assays, including anti-inflammatory, antioxidant, antithrombotic, and nitric oxide-inducing mechanisms that could also contribute to its antiatherogenic effects (4, 5).

Current evidence suggests that blood lipids are complex phenotypes, influenced by both environmental and genetic factors. It has been well established that body weight (6), current smoking habits (7), exercise (8), alcohol use (9) and dietary fat intake (10) influence plasma HDL-C levels. Several twin and family studies indicate that heritability estimates for blood levels of HDL-C range from 24% to 83%, with most studies in the 40% to 60% range (11). Mutations in genes including ABC transporter A1 (ABCA1), apolipoprotein A1, and lecithin cholesterol transferase (LCAT) are implicated in rare mendelian forms of HDL deficiency and familial hypoalphalipoproteinemia (12, 13). A mutation in the cholesteryl ester transfer protein gene (CETP) found in two Japanese siblings causes CETP-deficiency and extremely elevated levels of HDL-C (14). Findings from candidate gene studies suggest that genetic polymorphisms, including single nucleotide polymorphisms (SNPs) located in CETP, lipoprotein lipase (LPL), hepatic lipase (LIPC) (15, 16), and apolipoprotein A1/C3 (17) genes are important sources of genetically determined variation in plasma HDL-C. However, they explain only a small part of the variation. Most of the DNA sequence variants that contribute to variation in plasma HDL-C levels in the general population are still largely unknown.

Since low plasma HDL-C levels always cluster with other kinds of dyslipidemia and also insulin resistance, in this study, we extended the candidate gene scope and emphasized on the intricate links within lipid metabolic pathways and also between glucose and lipid metabolic pathways (18). We postulated that more genuine signals from the genes that are involved in the metabolism of HDL-C could be captured. The purpose of this study was twofold: first, to survey genetic variants in a large number of candidate genes in relation to plasma HDL-C levels, and second, to investigate whether these associations could be modified by traditional environmental factors, such as drinking, smoking and dietary fat intake.

METHODS

Study population

Our study was conducted in Doetinchem, a town in a rural area in the east of the Netherlands. Random samples were taken from the municipal population register between 1987 and 1991. A total of 12404 inhabitants aged 20-59 years were willing to participate

and underwent a first measurement (response 62%). A sub-sample was re-invited for a second measurement between 1993 and 1997 and 6100 participants were re-examined (response 79%). Between 1998 and 2002, a third measurement took place and 4917 participants were re-assessed (response 75%). For each survey approval had been obtained from the Medical Ethics Committee. Informed consent was obtained from all participants. Overall, the Doetinchem Cohort consisted of 4662 persons for which three measurements were available. Participants who changed their smoking habits (n = 872), who had missing data on smoking status (n = 11) or who were pregnant at the time of measurement (n = 122) were excluded. Finally, 3779 participants met the inclusion criteria of this study. The design and detailed methods have been reported earlier (19). The subjects were surveyed on demographic, anthropometric and lifestyle information (smoking, alcohol use, physical activity and dietary habits), disease history and medications by questionnaires. A non-fasting blood sample was taken from all participants, fractionated into blood plasma, white blood cells and erythrocytes and subsequently stored. A validated semi-quantitative food frequency questionnaire was used in the second and third surveys to assess the habitual consumption of 178 food items during the previous year (20). Nutrient and energy intake were quantified for each individual using a updated computerized Dutch food composition table.

Laboratory assessment of HDL-C.
Plasma HDL-C levels were assayed in the Lipid Reference Laboratory (LRL) of the University Hospital Dijkzigt in Rotterdam using standardized enzymatic methods within three weeks after storage. The LRL Rotterdam is a permanent member of the International Cholesterol Reference Method Laboratory Network. It has been standardized to the Centers for Diseases Control and Prevention (CDC) through participating in the CDC/National Heart Lung and Blood Institute Lipid Standardization Program. HDL-C was determined in the supernatant after precipitation of apoB-containing lipoproteins with phophotungstic acid/$MgCl_2$ (Boehringer). The accuracy of HDL-C determination fulfilled National Cholesterol Education Program (NCEP) recommendations throughout the entire period. From December 1999, a homogen liquid method was used. From March 2002, the homogen liquid (2^{nd} generation) was used. Whenever a new method was introduced, a careful calibration was performed to make sure there was no difference between different assaying methods.

Gene and SNP selection and genotyping
Candidate genes were selected by a pathway-driven approach based on literature, with emphasis on regulatory pathways that control fatty acid, glucose, cholesterol and bile salt homeostasis; for a recent review see reference 18. The selection procedure started from the master regulator genes encoding nuclear receptors (PPARs, LXR, and FXR), transcription factors (SREBPs) and continued by selecting their co-activators, co-repressors and target genes. In addition, hormonal receptors (insulin receptor) and their downstream signaling proteins were selected. Furthermore, several candidate genes, described in literature to be associated with blood lipids or blood pressure were added. The selection resulted in 251 candidate genes.

Where possible eligible SNPs for these genes were selected based on published associations with any disorder or functional parameter, using databases from NCBI (PubMed, Gene and SNP), the Genetic Association Database (21) from CDC, and SNPper (22). In part limited by various constrains of the Golden Gate genotyping assay, a total of 153 SNPs across 91 genes were finally selected from the literature. Subsequently, we used the web based program SNPselector (23) to query all 251 genes for potential candidate SNPs. We performed a "SNPs by gene" search, including 5kb 5' and 1kb 3' flanking sequences, with slightly modified default ranking settings. Firstly, SNPs located in repeat regions were excluded (Repeat_score>0), to avoid potential genotyping difficulties. Secondly, Caucasian minor allele frequencies (MAF_Caucasian>0) had to be available. Thirdly, SNPs in predicted transcription factor binding sites (Regulatory=*TFBS*) were preferred. Finally, subsequent ranking was based on the highest function score (Function_score), followed by the highest regulatory score (Regulatory_score), discriminating between SNPs that might affect gene transcript structure or protein product, and the regulatory potential of the SNP, respectively. A total of 226 SNPs across 178 genes, selected with SNPselector, were included on the Illumina array. Four of the 251 candidate genes remained without eligible SNPs in the SNP search described above. For each of these genes (SCAP, ACSL1, CEBPA, E2F4) a single SNP was handpicked based on Caucasian allele frequency, SNP location and validation information in NCBI SNP. Together a final set of 383 SNPs across 251 candidate genes, with one to seven SNPs per gene, passed the Illumina Assay Design Tool and were included. Besides the 383 selected SNPs, one Y chromosome marker was determined to serve as gender control. The complete list of SNPs and additional data for all SNPs are provided in supplementary table 1.

Genomic DNA was extracted from the buffy coat fraction with a salting out method. A total of 139 subjects were not eligible for genotyping, mainly because DNA extraction did not succeed or DNA was not available. For 3639 subjects high throughput SNP genotyping was performed with the Illumina Golden Gate assay using the Sentrix Array Matrix platform (Illumina Inc, San Diego, California). Illumina GenCall software (version 6.1.3.28) was used for automated genotype clustering and calling. Genotyping failed for 43 subjects due to an overall absence of any signal. In addition, 21 subjects were excluded because of discordance for the gender control. For 28 SNPs the genotype calling did not succeed due to low signal (n = 11), overlap between the genotype clusters (n = 13), multiple genotype clusters (n = 3) or scattering of clusters (n = 1). For the eight SNPs most out of Hardy Weinberg Equilibrium (HWE), genotyping of the particular SNP was verified in a random sample (n = 96) using Taqman, pyrosequencing or sequencing. Two of these failed the verification and were excluded. Furthermore, genotype calling was not completely convincing for 42 SNPs. The latter were included in the analysis, but when an association was found with body mass index, waist circumference or lipids levels, genotyping was verified. All SNPs passed the verification. Finally, for 3575 participants data was available of 353 SNPs in 239 genes.

Statistical analysis

All analyses were performed with SAS version 9.1 software (SAS Institute, Cary, NC). Paired Student's *t* tests and chi-square tests were used for comparisons of means and

proportions between measurement rounds. If data were not normally distributed, the Wilcoxon Signed-Rank test was used to make the comparison. Distributions of genotypes were tested for deviation from HWE by chi-square analysis (PROC ALLELE). Random coefficient models (multi-level modeling) were used to study the relationship between SNPs and repeated measurements of plasma HDL-C levels (PROC MIXED). Men and women were analyzed separately for five X-linked SNPs (rs2073115, rs3213451, rs5969919, rs1403543 and rs3048). When the overall difference was statistically significant, the Tukey-Kramer method was used to identify significant differences between the genotype groups. The exact follow-up time in years was put into the model as a continuous variable. To adjust for the potential confounding effects and to improve model fitting, age, age^2, sex, current smoking (yes or no), alcohol use and body mass index (BMI) were added to the model as covariates. The intercept and time were treated as random effects allowing unique baseline levels and unique changes of plasma HDL-C concentration over time for each individual. Potential gene-environment interactions in relation to plasma HDL-C level were also explored using MIXED models by including interaction terms into the model. All reported p values were two-tailed, and statistical significance without adjustment for multiple testing was defined at $\alpha = 0.05$ level. The false discovery rate (FDR) was applied to take multiple testing into account (PROC MULTTEST). To date, there is no conventional q value threshold to categorize a discovery as significant. As in similar research, a q value threshold of 0.20 was used to define significance (24). Certain SNPs in previously confirmed candidate genes affecting plasma HDL-C levels were also reported ($p < 0.05$) although their q values exceeded 0.2.

RESULTS

The mean age of the subjects at the first survey was 40.8 years, ranging from 20 to 60 years (table 1). More people took lipid-lowering medication in the second and the third round compared to the first examination. The average plasma HDL-C level increased from the first survey (1.26 ± 0.31 mmol/L) to the second survey (1.38 ± 0.38 mmol/L), but decreased slightly thereafter in the third survey (1.36 ± 0.39 mmol/L). Correlation coefficients between the three measurements of plasma HDL-C levels ranged from 0.76 to 0.82.

Twenty-eight SNPs were found to be significantly associated with variations in plasma HDL-C levels ($p < 0.05$, table 2) after adjustment for age, age^2, sex, current smoking habit, alcohol use and BMI. Eleven of them had a FDR_q-value < 0.2 after adjustment for multiple testing; seven SNPs (rs1800777 and rs5882 in CETP; rs3208305, rs328 and rs268 in LPL; rs1800588 in LIPC and rs2229741 in NRIP1) had a FDR_q value < 0.05, while four more SNPs (rs17585739 in SCMOL, rs11066322 in PTPN11, rs4961 in ADD1 and rs6060717 near SCAND1) had a FDR_q-value between 0.05 and 0.2. Two of the SNPs for which the significance disappeared after adjustment for multiple testing concerned less well replicated associations in the literature (rs3135506 in APOA5, $p = 0.008$ and FDR_q = 0.221; rs1800961 in HNF4A, $p = 0.018$ and FDR_q = 0.338, respectively). With respect to the SNPs located on the X chromosome, a SNP in MBTPS2 (rs3213451) was associated with plasma HDL-C levels, but only in female subjects ($p = 0.02$ and FDR_q = 0.12). The p values for the above top SNPs (FDR_q-value < 0.2) did not change too much after we

excluded the subjects who had started lipid-lowering medication in each round of survey (data not shown).

TABLE 1. General characteristics of the study population (n=3575, 1710 men and 1865 women)

	First Survey	Second Survey	Third Survey
Time since baseline (yr) [a]	-	6.0(5.9-6.1)	11.0(10.9-11.1)
Age(yrs)	40.8 ± 9.8	46.7 ± 9.8	51.7 ± 9.8
BMI(kg/m^2) [b]	24.5 ± 3.3	25.4 ± 3.6	26.1 ± 3.8
HDL cholesterol (mmol/L) [b]	1.26 ± 0.31	1.38 ± 0.38	1.36 ± 0.39
Lipid-lowering medication (%) [b]	6(0.17)	52(1.45)	149(4.17)
Current smokers (%)	915(25.6)	916(25.6)	916(25.6)
Alcohol consumption (glass/day) [a, b]	0.57(0-1.43)	0.57(0-1.57)	0.71(0-1.71)
Dietary intake	Na		
Energy intake (kj/day) [b]	-	9403 ± 2570	9081 ±2415
Total fat (% energy) [b]	-	35.0 ± 4.7	34.8 ± 4.8
Saturated fat (% energy) [b]	-	14.6 ± 2.3	14.4 ± 2.3
Monounsaturated fat (% energy)	-	13.3 ± 2.1	13.2 ± 2.2
Polyunsaturated fat (% energy)	-	6.8 ± 1.6	6.9 ± 1.6
Cholesterol (mg/day) [b]	-	240.1 ± 82.8	229.8 ± 79.6

BMI, body mass index; Na: not available.
Data are expressed as mean ± standard deviation or N(%) unless otherwise indicated.
[a] Median (Q1-Q3).
[b] Significant difference between the surveys, $p<0.05$.

Compared to the ancestral allele, rs1800777 in CETP, rs3208305 in LPL, rs11066322 in PTPN11, rs4961 in ADD1, rs3135506 in APOA5, and rs1800961 in HNF4A were associated with decreased plasma HDL-C levels, while rs268 and rs328 in LPL, rs5882 in CETP, rs2229741 in NRIP1, rs1800588 in LIPC, rs17585739 in SC4MOL, rs6060717 near SCAND1 and rs3213451 in MBTPS2 were associated with increased plasma HDL-C levels (table 3 and supplementary table 2). Rs328 and rs3208305 in LPL, rs1800588 in LIPC, and rs17585739 in SC4MOL had relatively big effects on plasma HDL-C levels compared to other SNPs ($\geq$ 0.1 mmol/L between the two homozygote genotypes).

Considering that variation in plasma HDL-C level also relates to lifestyle factors, including smoking, body weight, alcohol use and dietary fat intake, we further examined whether these factors could modulate the observed associations between the 14 SNPs identified above (SNPs with FDR_q < 0.2, rs3135506 and rs1800961) and plasma HDL-C level. For dietary fat intake (saturated fat, monounsaturated fat and polyunsaturated fat), only information in the second and third surveys was included in the analysis, as information in the first survey is less valid. Interactions with the other lifestyle factors were based on three rounds of measurements. Before adjustment for multiple testing, there were possible interactions between NRIP1 genotype (rs2229741) and BMI , APOA5 genotype (rs3135506) and alcohol use (low vs. high), and APOA5 genotype (rs3135506) and saturated fat intake (low vs. high) (p for interaction = 0.017, 0.023 and 0.042, respectively).

However, the significance of the interactions disappeared after adjustment for multiple testing (FDR_q > 0.2).

TABLE 2. Association between SNPs and plasma HDL cholesterol levels

Nearest Gene	SNP	MAF	p [a]	FDR_q [b]
LPL	rs3208305	0.30	<0.0001	<0.0001
LPL	rs328	0.11	<0.0001	<0.0001
CETP	rs1800777	0.03	<0.0001	<0.0001
CETP	rs5882	0.31	<0.0001	0.0002
NRIP1	rs2229741	0.41	<0.0001	0.0013
LIPC	rs1800588	0.20	<0.0001	0.0025
LPL	rs268	0.02	0.0003	0.0164
MBTPS2 [c]	rs3213451	0.35	0.0231	0.1153
SC4MOL	rs17585739	0.06	0.0027	0.1190
PTPN11	rs11066322	0.19	0.0031	0.1196
ADD1	rs4961	0.21	0.0036	0.1264
SCAND1	rs6060717	0.20	0.0061	0.1924
PIAS1	rs1489599	0.43	0.0082	0.2211
APOA5	rs3135506	0.08	0.0083	0.2211
NCOR2	rs2229840	0.17	0.0099	0.2451
LPL	rs1059507	0.16	0.0114	0.2645
MAP2K1 [d]	rs17586159	0.02	0.0132	0.2862
PRKCA	rs7210446	0.40	0.0140	0.2862
HNF4A	rs1800961	0.04	0.0175	0.3383
NRIP1	rs2229742	0.12	0.0196	0.3532
ACADM	rs11549022	0.30	0.0203	0.3532
PRKCA	rs17633437	0.37	0.0234	0.3886
ILK	rs2288283	0.13	0.0254	0.4014
ADIPOQ	rs17300539	0.07	0.0312	0.4722
MYBBP1A	rs751670	0.16	0.0383	0.4978
MLYCD	rs11649200	0.17	0.0384	0.4978
NDN	rs850791	0.09	0.0386	0.4978
SAH	rs5716	0.08	0.0386	0.4978
NCOA2	rs10112498	0.39	0.0444	0.5518

SNP, single nucleotide polymorphism; MAF, minor allele frequency; FDR, false discovery rate.
[a] Adjusted for age, age^2, sex, current smoking habits, alcohol use and body mass index.
[b] Adjusted for multiple testing (348 SNPs) with the FDR method.
[c] Statistical analysis was conducted in women only and was adjusted for multiple testing (5 SNPs) with FDR method.
[d] There was only one subject with the genotype AA. This subject was incorporated into the genotype GA group during the statistical analysis.

DISCUSSION

In this longitudinal genetic association study, with a pathway-driven approach, we confirmed several genetic variants in genes along known HDL metabolic pathway to be associated with plasma HDL-C levels. At the same time, we found some new additional

candidate genes whose genetic variants may be associated with plasma HDL-C levels. For these genes the results need to be replicated in other population studies.

TABLE 3. Distribution of plasma HDL cholesterol levels according to the genotypes of 14 SNPs significantly associated with plasma HDL cholesterol levels.

Nearest Gene	SNP	Nucleotide substitution [a]	p(HWE)	Frequency	HDL-C (mmol/ L)
LPL	rs3208305	26696 T→A	0.97		
	TT			315	1.39a
	AT			1490	1.32b
	AA			1767	1.27c
	rs328 (Ser447Ter)	22772 C→G	0.90		
	GG			40	1.43a
	CG			683	1.36a
	CC			2851	1.28b
CETP	rs268 (Asn291Ser)	16577 A→G	0.76		
	GG			2	1.39
	AG			148	1.21a
	AA			3423	1.30b
	rs1800777(Arg468Gln)	21427 G→A [b]	0.60		
	AA			3	1.22
	AG			233	1.20a
	GG			3332	1.31b
NRIP1	rs5882 (Ile422Val)	20200 A→G	0.20		
	GG			328	1.35a
	AG			1563	1.31a
	AA			1682	1.28b
	rs2229741	224 A→G	0.05		
	AA			634	1.26a
	AG			1674	1.30b
	GG			1263	1.32b
LIPC	rs1800588	-557 C→T	0.63		
	TT			149	1.39a
	CT			1134	1.30b
	CC			2267	1.29b
SC4MOL	rs17585739	3968 G→A	0.55		
	AA			15	1.52a
	AG			401	1.32b
	GG			3159	1.30b
PTPN11	rs11066322	65613 A→G	0.02		
	AA			150	1.36a
	AG			1049	1.28b
	GG			2342	1.30b
ADD1	rs4961 (Gly460Trp)	29064 G→T	0.67		
	TT			146	1.27
	TG			1172	1.28a
	GG			2252	1.31b
SCAND1	rs6060717	-2441 C→T	0.81		
	CC			134	1.23a
	CT			1127	1.31b
	TT			2309	1.30b
APOA5	rs3135506 (Ser19Trp)	169 G→C	0.66		
	CC			19	1.24
	CG			508	1.27a

	GG			3042	1.31b
	rs1800961(Thr117/139Ile)	12351 C→T	0.44		
HNF4A	TT			3	1.17
	CT			250	1.26a
	CC			3321	1.30b
	rs3213451	3581 A→G	0.36		
MBTPS2 [c]	GG			214	1.49a
	AG			860	1.43b
	AA			787	1.45

SNP, single nucleotide polymorphism; p(HWE), p value from x^2 test of Hardy-Weinberg Equilibrium; HDL-C, high density lipoprotein cholesterol.
[a] Human mutation nomenclature(50).
[b] Information on ancestral allele is not available from chimpanzee, but from macaque.
[c] statistical analysis and calculation were only conducted in women.
Genotypes with different letters marked with (a, b and c) differ significantly from each other in plasma HDL cholesterol levels.

Genetic variants among CETP, LPL and LIPC have previously been found to be associated with plasma HDL-C levels in candidate gene studies (15, 16). Recently, in a series of genome-wide association (GWA) analyses (25-27), the associations between genetic variants from these genes and plasma HDL-C levels were also confirmed (rs711752, rs7205804, rs5880 and rs3764261 in CETP; rs326 and rs12678919 in LPL; rs11858164 and rs10468017 in LIPC). The influence from these genetic variants (rs1800777 and rs5882 in CETP, rs3208305, rs328 and rs268 in LPL and rs1800588 in LIPC) on plasma HDL-C levels was once again observed in our study. These above mentioned significant loci in LPL or LIPC are in fact located in one haplotype block (28), which makes them more frequently inherited together. All the loci in CETP are located in two connected haplotype blocks (28). The roles these genes play in the metabolic process of HDL-C have well been elucidated either from experimentation in cell culture and in animal models or from observational and interventional studies in humans (3).

Two less well replicated associations (rs1800961 in HNF4A and rs3135506 in APOA5) with plasma HDL-C levels were also observed in our study, although their statistical significance disappeared after adjustment for multiple testing. In line with our findings serum HDL-C concentration was lower in subjects with the T130I (rs1800961) mutation compared with those without this mutation in a study among Japanese patients with late-onset type 2 diabetes (29). A large European Caucasian collection of MODY (maturity onset diabetes of the young) patients due to HNF4A mutations (30) also presented with lower fasting apolipoprotein A-II, A-I and HDL-C levels than control subjects. Hepatic nuclear factor 4 alpha (HNF4A) regulates hepatic expression of a number of genes associated with lipoprotein metabolism, including genes encoding apoA-I, A-II, A-IV, B, C-II, C-III, E, microsomal triglyceride transfer protein, cholesterol 7 α-hydroxylase, SR-BI, and PPARα (31). Mice with targeted mutations of HNF4A have a dramatic decrease in LDL and HDL cholesterol levels, and the HDL particles in these mice are small and lipid-poor (31). These results altogether show that HNF4A is important in HDL-C metabolism.

The APOA5 gene is located ~ 27 kb upstream of the well characterized APOA1/C3/A4 gene cluster. It is suggested that HDL is a reservoir for apoA-V based on the observation

that there was a greater than 20-fold higher level of apoA-V in HDL than in VLDL of APOA5 transgenic mice (32). There are numerous reports on the association of genetic variants of APOA5 with low HDL-C levels, such as rs662799, rs651821, rs2075291 in Japanese populations (33) and Chinese populations (34). The SNP of rs3135506 is significantly associated with blood HDL-C levels in northern American (35) and Austrian Caucasian subjects (36). In our Dutch population, rs3135506 was associated with plasma HDL-C levels before adjustment for multiple testing. This SNP, substituting Trp for Ser at residual 19 within the predicted signal peptide, is the only common variant with known influence on APOA5 expression (37). Furthermore, another SNP (rs28927680) in linkage disequilibrium with rs3135506 ($r^2 = 0.98$) is recently reported to be associated with HDL-C concentrations in a genome wide association analysis (27). Therefore, we think rs3135506 is an important determinant of HDL-C levels.

A SNP (rs2229741) in nuclear receptor interacting protein 1 (NRIP1) was found to be associated with plasma HDL-C levels in our study. Another SNP (rs2229742), ~ 1kb downstream of rs2229741 was also found to be associated with plasma HDL-C levels, although the significance disappeared after adjustment for multiple testing (table 2). These two SNPs are correlated with each other ($r^2 = 0.44$). Two other SNPS (rs1297214 and rs2142450) in this gene are shown to be associated with age-sex/multivariable adjusted NMR intermediate HDL-C concentration in the Framingham Offspring study (Exam 4) using a generalized estimating equation regression (GEE) approach (p = 0.02/0.04 for rs1297214; p = 0.02/0.04 for rs2142450) (38). NRIP1 is reported to be involved in the estrogen regulation of apoA1 enhancer activity. At low ratios of NRIP1 to estrogen receptor α, estradiol repressed apoA1 enhancer activity, whereas at high ratios this repression was reversed (39). Recent studies have identified a hepatocyte specific role for NRIP1 as a cofactor for LXR in different ways, namely serving as a coactivator in lipogenesis and as a corepressor in gluconeogenesis (40). Since NRIP1 is a widely expressed cofactor for nuclear receptors involved in the regulation of metabolic gene expression (41), its exact role in HDL-C metabolism warrants further research.

The SNP of rs17585739 in Sterol C4-methyloxidase-like (SC4MOL), that catalyzes sterol-4α-methyl oxidation process in cholesterol biosynthesis (42), was associated with plasma HDL-C levels in our population. Another SNP (rs1550270) in this gene was also associated with multivariable adjusted NMR large HDL-C concentration (p = 0.03) and HDL size (p = 0.04) in the Offspring Exam 4 of the Framingham SNP Health Association Resource (38). Interestingly, also along the cholesterol biosynthetic pathway, MVK (mevalonate kinase), encoding mevalonate kinase to convert mevalonate into 5-phosphomevalonate (2), is reported to possibly affect HDL-C concentrations in a recent genome wide association study (25). However, the SNP (rs7957619) in MVK we studied was not associated with plasma HDL-C concentrations (p = 0.45). Another SNP in a gene regulating this pathway, rs3213451 in MBTPS2, a metalloprotease required for intramembrane proteolysis of sterol regulatory element-binding proteins (2), was associated with plasma HDL-C concentrations in women in our study. Although this finding needs replication in other studies, it is tempting to speculate that de novo biosynthesis of cholesterol is an important contributor to plasma HDL-C. However, this remains to be confirmed.

Also for other SNPs, our results are more speculative. The results are inconsistent with those found in other studies, as in those studies no associations with HDL-C levels have been reported but with other (lipid) parameters. Also no association has been found between these genes and HDL-C in recent GWA studies. Therefore, also these findings need replication. The genetic variant of PTPN11 (rs11066322) is reported to be associated with serum ApoB levels and LDL cholesterol levels in a study conducted in Caucasian female twins (43), but in our study this genetic variant was associated with plasma HDL-C levels. Furthermore, the genetic variant of ADD1 (rs4961) has been associated with susceptibility to hypertension, but not with plasma HDL-C levels (44). However, we did find a significant association between rs4961 and plasma HDL-C levels in our study. SCAN-domain-containing protein 1 (SCAND1) is the nearest gene to the locus of rs6060717 that was found to be associated with plasma HDL-C levels in our study. SCAND1 is a widely expressed nuclear protein that may function as a key regulator of zinc finger transcription factor function (45). The capability of SCAND1 to interact with ZNF202 and PPARγ2 as two crucial transcription factors involved in lipid metabolism suggests that SCAND1 could function as important coregulator in lipid metabolism (46). In view of the above, a role of these genes in lipid metabolism is possible, but the association with HDL-C needs to be studied further.

There are some limitations to acknowledge in our study, which could have lead to either type I (false positive) or type II (false negative) errors. We adopted the false discovery rate method to adjust for multiple testing, as we took a candidate gene approach (and not random markers on the genome), and Bonferroni correction would have been too stringent. We think it is justified to take a more liberal threshold of FDR_q-value < 0.2. Although we took this approach to avoid false-positive results, we cannot exclude the possibility that the novel SNPs are spuriously associated with plasma HDL-C levels; therefore, the results for novel SNPs need further replication in other populations. A number of factors could have resulted in a type II error, leading to the inability to detect a true underlying association. Firstly, only a limited number of SNPs within a candidate gene has been studied. Failure to find an association with these SNPs does not exclude the possibility that other SNPs in the gene are related to plasma HDL-C levels. Secondly, less common variants with low minor allele frequencies (between 0.001 and 0.01) have been proven to affect plasma HDL-C levels (47). Although our sample size is large enough to allow us to detect effects as low as 0.4% variation in plasma HDL-C levels with 80% power ($p = 0.005$, FDR_q = 0.2) for such less common genetic variants, we did not take these less common variants into consideration in our study. Recently, some authors argued that with the findings from the previous linkage studies, candidate gene studies and also current genome wide association studies, we still cannot explain the high heritability of complex phenotypes. In fact, all the significant 13 SNPs (except rs3213451 in MBTPS2) identified in our study only explain 4% of the inter-individual variation in average HDL-C concentrations over the three measurements in an additive model in our study. It is suggested that deep resequencing of some candidate genes to find these culprit (rare) variants may be a solution (48). Finally, our blood samples were taken from the subjects in a non-fasting condition. However, there was no clear association between the time to last meal and plasma HDL-C levels in the

second and third surveys (data not shown). Also others found that fasting does not influence HDL-C levels to a large extent (49). Therefore, we do not think the non-fasting condition has influenced our results.

In conclusion, in addition to replication of previous results for candidate genes (SNPs among CETP, LPL, LIPC, HNF4A and APOA5), we found interesting new candidate SNPs for plasma HDL-C levels (rs2229741 in NRIP1, rs3213451 in MBTPS2, rs17585739 in SC4MOL, rs11066322 in PTPN11, rs4961 in ADD1 and rs6060717 near SCAND1). The validity and relevance of these novel associations with plasma HDL-C levels require further validation in other study populations.

ACKNOWLEDGEMENTS

The Doetinchem Cohort Study was financially supported by the Ministry of Health, Welfare and Sport of The Netherlands and the National Institute for Public Health and the Environment. The authors thank the epidemiologists and fieldworkers of the Municipal Health Service in Doetinchem for their contribution to the data collection for this study. Logistic management was provided by J Steenbrink and P Vissink, and administrative support by EP van der Wolf. Data management was provided by A Blokstra, AWD van Kessel and PE Steinberger. The authors would also like to thank B Hoebee for her role in the conception of the study, C Strien for DNA preparation, K Siezen for her assistance in SNP selection, SW van den Berg and AG Heidema for their statistical advice. Y Lu is funded by grant 2006B195 of the Netherlands Heart Foundation.

Supplementary data:

http://www.jlr.org/content/49/12/2582/suppl/DC1

REFERENCES

1. Gordon, D. J., and B. M. Rifkind. 1989. High-density lipoprotein--the clinical implications of recent studies. *N Engl J Med* **321**: 1311-1316.
2. Lodish, H., A. Berk, P. Matsudaira, C. A. Kaiser, M. Krieger, M. P. Scott, S. Zipursky, and J. Darnell. 2004. Molecular cell biology. 5th ed. W. H. Freeman and Company, New York. 750-767.
3. Lewis, G. F., and D. J. Rader. 2005. New insights into the regulation of HDL metabolism and reverse cholesterol transport. *Circ Res* **96**: 1221-1232.
4. Barter, P. J., S. Nicholls, K. A. Rye, G. M. Anantharamaiah, M. Navab, and A. M. Fogelman. 2004. Antiinflammatory properties of HDL. *Circ Res* **95**: 764-772.
5. Mineo, C., H. Deguchi, J. H. Griffin, and P. W. Shaul. 2006. Endothelial and antithrombotic actions of HDL. *Circ Res* **98**: 1352-1364.
6. Dattilo, A. M., and P. M. Kris-Etherton. 1992. Effects of weight reduction on blood lipids and lipoproteins: a meta-analysis. *Am J Clin Nutr* **56**: 320-328.
7. Moffatt, R. J. 1988. Effects of cessation of smoking on serum lipids and high density lipoprotein-cholesterol. *Atherosclerosis* **74**: 85-89.
8. Kokkinos, P. F., J. C. Holland, P. Narayan, J. A. Colleran, C. O. Dotson, and V. Papademetriou. 1995. Miles run per week and high-density lipoprotein cholesterol levels in healthy, middle-aged men. A dose-response relationship. *Arch Intern Med* **155**: 415-420.
9. McConnell, M. V., I. Vavouranakis, L. L. Wu, D. E. Vaughan, and P. M. Ridker. 1997. Effects of a single, daily alcoholic beverage on lipid and hemostatic markers of cardiovascular risk. *Am J Cardiol* **80**: 1226-1228.

10. Mooradian, A. D., M. J. Haas, and N. C. Wong. 2006. The effect of select nutrients on serum high-density lipoprotein cholesterol and apolipoprotein A-I levels. *Endocr Rev* **27**: 2-16.
11. Peacock, J. M., D. K. Arnett, L. D. Atwood, R. H. Myers, H. Coon, S. S. Rich, M. A. Province, and G. Heiss. 2001. Genome scan for quantitative trait loci linked to high-density lipoprotein cholesterol: The NHLBI Family Heart Study. *Arterioscler Thromb Vasc Biol* **21**: 1823-1828.
12. Brooks-Wilson, A., M. Marcil, S. M. Clee, L. H. Zhang, K. Roomp, M. van Dam, L. Yu, C. Brewer, J. A. Collins, H. O. Molhuizen, O. Loubser, B. F. Ouelette, K. Fichter, K. J. Ashbourne-Excoffon, C. W. Sensen, S. Scherer, S. Mott, M. Denis, D. Martindale, J. Frohlich, K. Morgan, B. Koop, S. Pimstone, J. J. Kastelein, J. Genest, Jr., and M. R. Hayden. 1999. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet* **22**: 336-345.
13. Hovingh, G. K., E. de Groot, W. van der Steeg, S. M. Boekholdt, B. A. Hutten, J. A. Kuivenhoven, and J. J. Kastelein. 2005. Inherited disorders of HDL metabolism and atherosclerosis. *Curr Opin Lipidol* **16**: 139-145.
14. Brown, M. L., A. Inazu, C. B. Hesler, L. B. Agellon, C. Mann, M. E. Whitlock, Y. L. Marcel, R. W. Milne, J. Koizumi, H. Mabuchi, and et al. 1989. Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins. *Nature* **342**: 448-451.
15. Knoblauch, H., A. Bauerfeind, M. R. Toliat, C. Becker, T. Luganskaja, U. P. Gunther, K. Rohde, H. Schuster, C. Junghans, F. C. Luft, P. Nurnberg, and J. G. Reich. 2004. Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet* **13**: 993-1004.
16. Isaacs, A., F. A. Sayed-Tabatabaei, O. T. Njajou, J. C. Witteman, and C. M. van Duijn. 2004. The -514 C->T hepatic lipase promoter region polymorphism and plasma lipids: a meta-analysis. *J Clin Endocrinol Metab* **89**: 3858-3863.
17. Brown, C. M., T. J. Rea, S. C. Hamon, J. E. Hixson, E. Boerwinkle, A. G. Clark, and C. F. Sing. 2006. The contribution of individual and pairwise combinations of SNPs in the APOA1 and APOC3 genes to interindividual HDL-C variability. *J Mol Med* **84**: 561-572.
18. Desvergne, B., L. Michalik, and W. Wahli. 2006. Transcriptional regulation of metabolism. *Physiol Rev* **86**: 465-514.
19. Verschuren, W., A. Blokstra, H. Picavet, and H. Smit. 2008. Cohort Profile: The Doetinchem Cohort Study. *Int J Epidemiol*. doi:10.1093/ije/dym292.
20. Ocke, M. C., H. B. Bueno-de-Mesquita, M. A. Pols, H. A. Smit, W. A. van Staveren, and D. Kromhout. 1997. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. *Int J Epidemiol* **26 Suppl 1**: S49-58.
21. Becker, K. G., K. C. Barnes, T. J. Bright, and S. A. Wang. 2004. The genetic association database. *Nat Genet* **36**: 431-432.
22. Riva, A., and I. S. Kohane. 2004. A SNP-centric database for the investigation of the human genome. *BMC Bioinformatics* **5**: 33.
23. Xu, H., S. G. Gregory, E. R. Hauser, J. E. Stenger, M. A. Pericak-Vance, J. M. Vance, S. Zuchner, and M. A. Hauser. 2005. SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatics* **21**: 4181-4186.
24. Smith, N. L., L. A. Hindorff, S. R. Heckbert, R. N. Lemaitre, K. D. Marciante, K. Rice, T. Lumley, J. C. Bis, K. L. Wiggins, F. R. Rosendaal, and B. M. Psaty. 2007. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *Jama* **297**: 489-498.
25. Willer, C. J., S. Sanna, A. U. Jackson, A. Scuteri, L. L. Bonnycastle, R. Clarke, S. C. Heath, N. J. Timpson, S. S. Najjar, H. M. Stringham, J. Strait, W. L. Duren, A. Maschio, F. Busonero, A. Mulas, G. Albai, A. J. Swift, M. A. Morken, N. Narisu, D. Bennett, S. Parish, H. Shen, P. Galan, P. Meneton, S. Hercberg, D. Zelenika, W. M. Chen, Y. Li, L. J. Scott, P. A. Scheet, J. Sundvall, R. M. Watanabe, R. Nagaraja, S. Ebrahim, D. A. Lawlor, Y. Ben-Shlomo, G. Davey-Smith, A. R. Shuldiner, R. Collins, R. N. Bergman, M. Uda, J. Tuomilehto, A. Cao, F. S. Collins, E. Lakatta, G.

M. Lathrop, M. Boehnke, D. Schlessinger, K. L. Mohlke, and G. R. Abecasis. 2008. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* **40**: 161-169.
26. Kooner, J. S., J. C. Chambers, C. A. Aguilar-Salinas, D. A. Hinds, C. L. Hyde, G. R. Warnes, F. J. Gomez Perez, K. A. Frazer, P. Elliott, J. Scott, P. M. Milos, D. R. Cox, and J. F. Thompson. 2008. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet* **40**: 149-151.
27. Kathiresan, S., O. Melander, C. Guiducci, A. Surti, N. P. Burtt, M. J. Rieder, G. M. Cooper, C. Roos, B. F. Voight, A. S. Havulinna, B. Wahlstrand, T. Hedner, D. Corella, E. S. Tai, J. M. Ordovas, G. Berglund, E. Vartiainen, P. Jousilahti, B. Hedblad, M. R. Taskinen, C. Newton-Cheh, V. Salomaa, L. Peltonen, L. Groop, D. M. Altshuler, and M. Orho-Melander. 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**: 189-197.
28. Frazer, K. A., D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**: 851-861.
29. Zhu, Q., K. Yamagata, A. Miura, N. Shihara, Y. Horikawa, J. Takeda, J. Miyagawa, and Y. Matsuzawa. 2003. T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. *Diabetologia* **46**: 567-573.
30. Pearson, E. R., S. Pruhova, C. J. Tack, A. Johansen, H. A. Castleden, P. J. Lumb, A. S. Wierzbicki, P. M. Clark, J. Lebl, O. Pedersen, S. Ellard, T. Hansen, and A. T. Hattersley. 2005. Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large European collection. *Diabetologia* **48**: 878-885.
31. Hayhurst, G. P., Y. H. Lee, G. Lambert, J. M. Ward, and F. J. Gonzalez. 2001. Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. *Mol Cell Biol* **21**: 1393-1403.
32. Nelbach, L., X. Shu, R. J. Konrad, R. O. Ryan, and T. M. Forte. 2008. Effect of apolipoprotein A-V on plasma triglyceride, lipoprotein size and composition in genetically engineered mice. *J Lipid Res* **49**:572-580.
33. Yamada, Y., H. Matsuo, S. Warita, S. Watanabe, K. Kato, M. Oguri, K. Yokoi, N. Metoki, H. Yoshida, K. Satoh, S. Ichihara, Y. Aoyagi, A. Yasunaga, H. Park, M. Tanaka, and Y. Nozawa. 2007. Prediction of genetic risk for dyslipidemia. *Genomics* **90**: 551-558.
34. Chien, K. L., M. F. Chen, H. C. Hsu, T. C. Su, W. T. Chang, C. M. Lee, and Y. T. Lee. 2008. Genetic association study of APOA1/C3/A4/A5 gene cluster and haplotypes on triglyceride and HDL cholesterol in a community-based population. *Clin Chim Acta* **388**: 78-83.
35. Lai, C. Q., S. Demissie, L. A. Cupples, Y. Zhu, X. Adiconis, L. D. Parnell, D. Corella, and J. M. Ordovas. 2004. Influence of the APOA5 locus on plasma triglyceride, lipoprotein subclasses, and CVD risk in the Framingham Heart Study. *J Lipid Res* **45**: 2096-2105.
36. Grallert, H., E. M. Sedlmeier, C. Huth, M. Kolz, I. M. Heid, C. Meisinger, C. Herder, K. Strassburger, A. Gehringer, M. Haak, G. Giani, F. Kronenberg, H. E. Wichmann, J. Adamski, B. Paulweber, T. Illig, and W. Rathmann. 2007. APOA5 variants and metabolic syndrome in Caucasians. *J Lipid Res* **48**: 2614-2621.
37. Talmud, P. J., J. Palmen, W. Putt, L. Lins, and S. E. Humphries. 2005. Determination of the functionality of common APOA5 polymorphisms. *J Biol Chem* **280**: 28215-28220.
38. Resource, Framingham SNP Health Association. *http://www.ncbi.nlm.nih.gov/projects/gap/cgi-*

bin/analysis.cgi?study_id=phs000007.v2.p1&phv=&phd=&pha=547&phsf=0&phvf=&phdf=&phaf=747.

39. Harnish, D. C., M. J. Evans, M. S. Scicchitano, R. A. Bhat, and S. K. Karathanasis. 1998. Estrogen regulation of the apolipoprotein AI gene promoter through transcription cofactor sharing. *J Biol Chem* **273**: 9270-9278.

40. Herzog, B., M. Hallberg, A. Seth, A. Woods, R. White, and M. G. Parker. 2007. The nuclear receptor cofactor, receptor-interacting protein 140, is required for the regulation of hepatic lipid and glucose metabolism by liver X receptor. *Mol Endocrinol* **21**: 2687-2697.

41. Parker, M. G., M. Christian, and R. White. 2006. The nuclear receptor co-repressor RIP140 controls the expression of metabolic gene networks. *Biochem Soc Trans* **34**: 1103-1106.

42. Rodriguez, C., B. Raposo, J. Martinez-Gonzalez, V. Llorente-Cortes, G. Vilahur, and L. Badimon. 2003. Modulation of ERG25 expression by LDL in vascular cells. *Cardiovasc Res* **58**: 178-185.

43. Jamshidi, Y., S. B. Gooljar, H. Snieder, X. Wang, D. Ge, R. Swaminathan, T. D. Spector, and S. D. O'Dell. 2007. SHP-2 and PI3-kinase genes PTPN11 and PIK3R1 may influence serum apoB and LDL cholesterol levels in normal women. *Atherosclerosis* **194**: e26-33.

44. Yazdanpanah, M., F. A. Sayed-Tabatabaei, A. Hofman, Y. S. Aulchenko, B. A. Oostra, B. H. Stricker, H. A. Pols, S. W. Lamberts, J. C. Witteman, J. A. Janssen, and C. M. van Duijn. 2006. The alpha-adducin gene is associated with macrovascular complications and mortality in patients with type 2 diabetes. *Diabetes* **55**: 2922-2927.

45. Sander, T. L., and J. F. Morris. 2002. Characterization of the SCAN box encoding RAZ1 gene: analysis of cDNA transcripts, expression, and cellular localization. *Gene* **296**: 53-64.

46. Schmitz, G., S. Heimerl, and T. Langmann. 2004. Zinc finger protein ZNF202 structure and function in transcriptional control of HDL metabolism. *Curr Opin Lipidol* **15**: 199-208.

47. Cohen, J. C., R. S. Kiss, A. Pertsemlidis, Y. L. Marcel, R. McPherson, and H. H. Hobbs. 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**: 869-872.

48. Petretto, E., E. T. Liu, and T. J. Aitman. 2007. A gene harvest revealing the archeology and complexity of human disease. *Nat Genet* **39**: 1299-1301.

49. Craig, S. R., R. V. Amin, D. W. Russell, and N. F. Paradise. 2000. Blood cholesterol screening influence of fasting state on cholesterol results and management decisions. *J Gen Intern Med* **15**: 395-399.

50. den Dunnen, J. T., and S. E. Antonarakis. 2001. Nomenclature for the description of human sequence variations. *Hum Genet* **109**: 121-124.

Chapter 4

Research article

Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study

Yingchang Lu, Edith J.M. Feskens, Jolanda M.A. Boer, Sandra Imholz, W.M.Monique Verschuren, Cisca Wijmenga, Anika Vaarhorst, Eline Slagboom, Michael Müller, Martijn E.T. Dollé

Atherosclerosis. 2010;213(1):200-5.

ABSTRACT

Background-Plasma total cholesterol (TC) levels are highly genetically determined. Although ample evidence of genetic determination of separate lipoprotein cholesterol levels has been reported, using TC level directly as a phenotype in a relatively large broad-gene based association study has not been reported to date.

Methods and Results-We genotyped 361 single nucleotide polymorphisms (SNPs) across 243 genes based on pathways potentially relevant to cholesterol metabolism in 3575 subjects that were examined thrice over 11 years. Twenty-three SNPs were associated with TC levels after adjustment for multiple testing. We used 12 of them (rs7412 and rs429358 in APOE, rs646776 in CELSR2, rs1367117 in APOB, rs6756629 in ABCG5, rs662799 in APOA5, rs688 in LDLR, rs10889353 in DOCK7, rs2304130 in NCAN, rs3846662 in HMGCR, rs2275543 in ABCA1, rs7275 in SMARCA4) that were confirmed in previous candidate association or genome-wide-association studies to define a gene risk score (GRS). Average TC levels increased from 5.23 ± 0.82 mmol/L for those with 11 or less cholesterol raising alleles to 6.03 ± 1.11 mmol/L for those with 18 or more (P for trend < 0.0001). The association with TC levels was slightly stronger when the weighted GRS that weighted the magnitude of allelic effects was used.

Conclusion-A panel of common genetic variants in the genes pivotal in cholesterol metabolism could possibly help identify those people who are at risk of high cholesterol levels.

Key Words: cholesterol; single nucleotide polymorphism; pathway-driven approach; random coefficient model.

Introduction

Mammalian cells require cholesterol for maintenance of membrane integrity and multiple cellular functions. Cells obtain cholesterol by either de novo synthesis in the endoplasmic reticulum or receptor-mediated uptake of lipoproteins, processes that are tightly controlled by feedback regulation to prevent the toxicity of excess unesterified cholesterol in membranes. Only 9% of the body cholesterol pool is easily accessible in plasma, the remaining 91% being found in tissues. Plasma total cholesterol (TC) levels represent a combination of cholesterol in structurally and metabolically heterogeneous groups of lipoprotein, such as very low density lipoprotein (VLDL), low density lipoprotein (LDL), high density lipoprotein (HDL) and cholesterol in other lipoproteins, among which cholesterol (and other lipids) and certain apolipoproteins are constantly exchanged (1-3). It is known that about two-thirds of the blood TC is attributable to LDL and the remaining mainly to HDL and VLDL. Ample evidence for a genetic basis of separate lipoprotein cholesterol levels has been reported so far (4-7). The heritability of TC levels has been estimated to be around 40-60% (8). However, plasma TC levels, as a comprehensive index of our body cholesterol pool, have rarely been explored directly as a phenotype in a relatively large, broad-gene based association study despite its being more sensitive in predicting cardiovascular disease risk compared with LDL cholesterol levels (9).

In this study, we followed 3575 subjects over 11 years during which their plasma TC levels were measured three times, which enabled us to better account for intra-individual variation. In order to get a broad list of genes potentially relevant to cholesterol metabolism, we adopted a pathway driven approach to select genes known to be involved in the regulatory pathways that control fatty acid, glucose, cholesterol and bile acid homeostasis. We emphasized on the intricate links within lipid metabolism and between glucose and lipid metabolism through common transcriptional factors (10). By surveying genetic variants in this broad list of candidate genes, we want to provide some new insights into the genetic determinants of plasma TC levels, and investigate whether common genetic variants in genes involved in cholesterol metabolism could predict the plasma cholesterol levels.

Materials and methods

Study population

Our study was conducted in Doetinchem, a town in a rural area in the east of the Netherlands. The design and detailed methods have been reported earlier (11). At baseline (round 1, 1987-1991), an age-and-sex stratified random sample of the population was drawn. A total of 12,405 subjects aged 20-59 yrs were willing to participate and underwent a first measurement. A random two-third of those measured in round 1 were invited for round 2 (1993-97), and 6,118 subjects participated. They were invited for a third measurement round (1998-2002) and 4,917 subjects participated. For 4662 participants, all three measurements were available. Participants who changed their smoking habits, had missing data on smoking status or were pregnant at the time of measurement were excluded. Finally, 3779 participants met the inclusion criteria for this study. Informed consent was obtained from all participants. In brief, the subjects were surveyed on demographic, anthropometric and lifestyle information, disease history and medication by

questionnaires. A validated semi-quantitative food frequency questionnaire was used in the second and third surveys to assess the habitual consumption of 178 food items during the previous year (12). A non-fasting blood sample was taken from all participants, fractionated into blood plasma, white blood cells and erythrocytes and subsequently stored at -20°C. Plasma TC was measured using a CHOD-PAP method (Boehringer) (13) after being sent to the Lipid Reference Laboratory of the University Hospital Dijkzigt (LRL) in Rotterdam once every week or three weeks. The LRL in Rotterdam is a permanent member of the International Cholesterol Reference Method Laboratory Network (14). For replication of associations between APOC3 SNPs (rs2854116 and rs4520) and TC levels, we used a random sample of 1984 subjects with a single measurement of TC from a population-based cohort study in Maastricht (15, 16), The Netherlands, which used the same methodology as the Doetinchem cohort.

Gene and SNP selection and genotyping

Genomic DNA was extracted from the buffy coat fraction with a salting out method. For 3639 subjects high throughput SNP genotyping was performed with the Illumina Golden Gate assay using the Sentrix Array Matrix platform (Illumina Inc, San Diego, California). Gene selection started from the master regulator genes (encoding the nuclear receptors [PPARs, LXR, and FXR] or transcription factors [SREBPs]), and continued by selecting their associated co-activators, co-repressors and target genes. Several candidate genes, described in literature to be associated with blood lipids or blood pressure were also added. The SNPs in each gene were selected based on either published associations or the web-based program SNPselector (17). Detailed information on the gene and SNP selecting procedures, quality control and the full gene and SNP list have been reported before (18). Two SNPs that failed in the Illumia assay (rs7412 and rs429358 in APOE) were successfully re-genotyped with a Taqman assay. In addition, 6 SNPs associated with blood total cholesterol levels from the latest genome wide association (GWA) study that was published after our initial gene selection (rs646776 in CELSR2, rs3846662 in HMGCR, rs2304130 in NCAN, rs10889353 in DOCK7, rs6756629 in ABCG5, and rs174570 in FADS2) (4) were genotyped by KBioscience (Hoddesdon, Hertfordshire, UK) using the KASPar chemistry, a competitive allele specific PCR SNP genotyping system using FRET quencher cassette oligonucleotides (http://www.kbioscience.co.uk). After exclusion of subjects for whom all genotypes failed, exclusion of failed genotypes and quality control, data were available of 361 SNPs in 243 genes for a total of 3575 participants. APOC3 SNPs (rs2854116 and rs4520) were genotyped in the replication sample at the Leiden University Medical Center using the Sequenom iPLEX platform.

Statistical analysis

Statistical analyses were performed with SAS version 9.1 software (SAS Institute, Cary, NC), unless indicated otherwise. The paired Student's *t* test and Chi-square test were used for comparisons of means and proportions between measurement rounds, respectively. The Wilcoxon Signed-Rank test was used for comparisons of alcohol consumption between two rounds of survey. Testing of deviation from Hardy Weinberg Equilibrium (HWE) and calculation of linkage disequilibrium (LD) were done with PROC ALLELE. For testing of deviation from HWE, the chi-square goodness-of-fit test was used ($p < 0.05$). The TC

levels of participants who took lipid-lowering medication at the time of survey were treated as missing values (Table 1). The intake of total fat and cholesterol was averaged over the second and third surveys and the medians (34.9 energy% and 224.9 mg/day, respectively) were used to categorize the subjects into low or high intake of fat or cholesterol. Since inheritance patterns of the causal alleles are unknown, an additive model was used, which gives good overall performance in any of the three potential modes of inheritance (19). Random coefficient models (multi-level modeling, PROC MIXED) were used to study the relationship between SNPs and repeated measurements of TC levels, which accounts for correlation between repeated measurements within subjects (20). Men and women were analyzed separately for the five X-linked SNPs (rs2073115, rs3213451, rs5969919, rs1403543, and rs3048). When the overall difference was statistically significant, the Tukey-Kramer method was used to identify significant differences between the genotype groups. To adjust for potential confounding effects and to improve model fitting, age, sex, current smoking status (yes or no), alcohol consumption, and body mass index (BMI) were added to the model as covariates. The exact follow-up time in years was put into the model as a continuous variable. Intercept and time were treated as random effects allowing unique baseline levels and unique changes of TC level over time for each individual. Gene*time interactions were explored by including interaction terms into the MIXED models. The odds ratio (OR) of having an average cholesterol level over the three rounds $\geq$ 6.5 mmol/L (hypercholesterolemia) in the subjects carrying $\geq$ 18 total cholesterol raising alleles was calculated using a logistic regression model adjusting for age, sex, current smoking habits, alcohol use and BMI. Subjects with < 18 total cholesterol raising alleles were used as reference. All reported *p* values were two-tailed, and statistical significance before adjustment for multiple testing was defined at the $\alpha = 0.05$ level. The false discovery rate (FDR, PROC MULTTEST) was applied to take multiple testing into account. To date, there is no conventional q-value threshold to categorize a discovery as significant. As in similar research, a q-value threshold of 0.20 was used to define significance (21).

Genetic risk score computation and analysis

A genetic risk score (GRS) was calculated on the basis of those SNPs associated with TC levels after adjustment for multiple testing that were replicated in other studies. If SNPs were in LD ($r^2 > 0.9$), only the SNP with the most significant main effect in our study was included in the score. As reported before (22), two methods were used to create the GRS: a simple count method (count GRS) and a weighted method (weighted GRS). Both methods assume each SNP to be independently associated with TC levels. We assumed an additive genetic model for each SNP, applying a linear weighting of 0, 1 and 2 to genotypes containing 0, 1, or 2 risk alleles (raising total cholesterol levels), respectively. The count method assumes that each SNP in the panel contributes equally to the cholesterol increasing effect and was calculated by summing the total number of risk alleles, producing a maximal score of 24. For the weighted GRS, each SNP was weighted by the β-coefficients (Supplementary Table 4) obtained from the linear regression of individual SNP on average TC levels over the three rounds of measurements by using an additive model in a randomly selected half of the total subjects with no missing value in any of the SNPs included in GRS calculation (n = 1668). The weighted GRS was calculated by multiplying each β-coefficient by the number of corresponding risk alleles (0, 1 or 2) and then summing

the products. This produces a score out of 3.4 (twice the sum of the β-coefficients), which was then divided by 3.4 and multiplied by 24 to facilitate interpretation and comparison with the count GRS. The weighted GRS was associated with plasma TC levels in the other half of the subjects (n = 1669). In doing so, we reduced the bias of yielding overly optimistic associations between the weighted GRS and plasma TC levels. To improve the comparability, the association between count GRS and TC levels was also assessed in the same subjects (Supplementary Table 5). To assess possible discriminative improvement for hypercholesterolemia attributable to the GRS, we calculated the areas under the receiver-operating characteristic curves (AUCs or c-index) from a logistic regression model including conventional risk factors only (age, sex, current smoking habit, alcohol use, and BMI) and a model which additionally included the GRS. This analysis was done in Stata 11 (StataCorp, Texas).

Results

Mean age of the subjects at the first survey was 40.8 years, ranging from 20 to 60 years (Table 1). The average TC levels increased from the first survey (5.47±1.03 mmol/L) to the second survey (5.52±1.02 mmol/L), and the third survey (5.74±1.03 mmol/L). More people took lipid-lowering medication in the second (1993-1997) and third survey (1998-2002) compared to the first survey (1987-1991). Correlation coefficients between the measurements of TC levels ranged from 0.74 to 0.76.

Table 1. General characteristics of the study population (n=3575, 1710 men and 1865 women)

	First Survey 1987-1991	Second Survey 1993-1997	Third Survey 1998-2002
Age (yrs)	40.8 ± 9.8	46.7 ± 9.8	51.7 ± 9.8
BMI (kg/m^2)	24.5 ± 3.3	25.4 ± 3.6	26.1 ± 3.8
Current smokers (%)	915 (25.6)	916 (25.6)	916 (25.6)
Alcohol consumption (glass/day) [a,b]	0.57 (0-1.43)	0.57 (0-1.57)	0.71 (0-1.71)
Lipid-lowering medication (%) [b]	6 (0.17)	52 (1.45)	149 (4.17)
Total cholesterol (mmol/L) [b]	5.47 ± 1.03	5.52 ± 1.02	5.74 ± 1.03
Dietary intake	Na		
Energy intake (kj/day) [b]	-	9403 ± 2570	9081 ± 2415
Total fat (% energy) [b]	-	35.0 ± 4.7	34.8 ± 4.8
Cholesterol (mg/day) [b]	-	240.1 ± 82.8	229.8 ± 79.6

BMI, body mass index; Na, not available.
Data expressed as mean ± standard deviation or N (%) unless otherwise indicated.
[a] Median (interquartile range).
[b] $P<0.05$.

Thirty-five SNPs were found to be significantly associated with TC levels ($p < 0.05$, Table 2) after adjustment for age, sex, current smoking status, alcohol consumption and BMI. Twenty-three of them remained statistically significant after adjustment for multiple testing ($FDR\_q < 0.2$). Eleven of them (rs7412 and rs429358 in APOE (23), rs646776 in CELSR2 (4), rs1367117 in APOB (6, 7), rs6756629 in ABCG5 (4, 24), rs662799 in APOA5 (6), rs688 and rs5925 in LDLR (25, 26), rs10889353 in DOCK7 (4, 24), rs2304130 in NCAN

(4, 24), rs3846662 in HMGCR (4)) were reported to be associated with blood TC or LDL levels in candidate genetic association studies or GWA studies. For the other 12 SNPs, we explored the association with plasma cholesterol levels in the publicly available results of a meta-analysis of seven GWA studies on LDL and HDL cholesterol levels (5). In this meta-analysis, rs2275543 in ABCA1 was associated with both LDL (p = 0.004) and HDL (p = 1.437E-7) cholesterol levels. Also, rs7275 in SMARCA4 was associated with LDL cholesterol levels (p = 1.921E-6, Supplementary Table 1). For the two SNPs in APOC3 (rs2854116 and rs4520) that were associated with TC levels in our study, no corresponding SNPs or tagging SNPs were genotyped or imputed in the meta-analysis dataset. Therefore, we carried out a replication study in 2221 subjects randomly selected from another population study using the same methodology as the Doetinchem cohort (details in Supplementary Table 2). Rs2854116 was not associated with TC levels (p > 0.05). Despite a larger difference in TC levels between genotypes of rs4520 in the replication samples than in the Doetinchem cohort, the association did not reach statistical significance (p > 0.05).

As an explorative analysis, we investigated whether TC levels changed differentially over the 11 years of follow-up for the genotypes of the above identified 13 significant SNPs (11 replicated in literature, 2 replicated in the meta-analysis of GWA studies). Four SNPs (rs7412 in APOE, rs646776 in CELSR2, rs662799 in APOA5, and rs7275 in SMARCA4) interacted with follow-up time on TC levels (p for interaction<0.05, Supplementary Table 3). No interaction between the above identified significant SNPs and gender was detected (data not shown).

In order to evaluate to what extent the 13 identified SNPs in aggregate explained the variation in TC levels, we constructed a gene risk score for each subject by using a simple count (count GRS) or a weighted (weighted GRS) approach. One SNP (rs5925) was excluded from the GRS calculation because of its LD with rs688 in LDLR (r2 = 0.99). Both the count and weighted GRSs ranged from 7 to 21. The median count and weighted GRSs were 14 and 15.5, respectively. Demographic and life style characteristics (age, gender, current smoking status, alcohol consumption and BMI) did not differ significantly across quartiles of the count or weighted GRS (data now shown). Based on the count GRS, the average increase of TC levels per risk allele was 0.11 ± 0.01 mmol/L after adjustment of age, sex, current smoking status, alcohol consumption and BMI.

Average TC levels increased from 5.23 ± 0.82 mmol/L for subjects with a score of 11 or less to 6.03 ± 1.11 mmol/L for subjects with a score of 18 or more (*p* for trend<0.0001, Table 3). Estimates were slightly higher (0.14 ± 0.01 mmol/L per 1 unit increase in weighted GRS) and the association between GRS and TC levels was slightly stronger when using the weighted GRS (Table 3 and Supplementary Table 5). In total, the variance in average TC levels explained by these 12 SNPs was 6.9% (7.2% and 6.7% in the low and high-fat intake group, respectively; 7.6% and 6.1% in the low and high-cholesterol intake group, respectively). Subjects in the top 5% of the distribution of the number of TC raising alleles (≥ 18 risk alleles) had an increased odds of having hypercholesterolemia using 6.5 mmol/L as a cut-off point (OR: 2.4; 95% CI: 1.7-3.5) as compared to the subjects with < 18

Table 2. Association between SNPs and repeated measures of plasma total cholesterol levels among 3575 men and women from the Doetinchem cohort.[a]

Gene	SNP	Minor allele	MAF	P [b]	FDR_q [c]
APOE	rs7412	T	0.08	<0.0001	<0.0001
APOE	rs429358	C	0.16	<0.0001	<0.0001
CELSR2/PSRC1/SORT1	rs646776	G	0.23	<0.0001	<0.0001
APOB	rs1367117	A	0.31	<0.0001	<0.0001
ABCA1	rs2275543	C	0.09	<0.0001	0.0031
ABCG5	rs6756629	G	0.07	0.0001	0.0066
APOA5	rs662799	G	0.07	0.0001	0.0066
GNB3	rs5443	T	0.31	0.0005	0.0234
LDLR	rs688	T	0.43	0.0012	0.0493
LDLR	rs5925	C	0.43	0.0015	0.0541
SMARCA4/LDLR	rs7275	C	0.31	0.0022	0.0686
ANGPTL3/DOCK7	rs10889353	C	0.35	0.0023	0.0686
SREBF1	rs4925119	A	0.13	0.0029	0.0801
CEBPA	rs12691	T	0.15	0.0037	0.0935
NCAN	rs2304130	G	0.09	0.0042	0.0962
PPARG	rs3856806	T	0.12	0.0045	0.0962
SREBF1	rs8066560	A	0.35	0.0046	0.0962
APOC3	rs2854116	G	0.36	0.0049	0.0966
APOC3	rs4520	T	0.27	0.0053	0.0995
HMGCR	rs3846662	C	0.45	0.0065	0.1166
PLA2G7	rs1051931	A	0.19	0.0079	0.1344
NDN	rs850791	G	0.09	0.0119	0.1929
MBTPS2d	rs3213451	G	0.34	0.0341	0.1963
ABCA1	rs2230806	A	0.26	0.0132	0.2047
ESRRA	rs2276014	T	0.16	0.0140	0.2078
PTPN11	rs11066322	A	0.19	0.0161	0.2291
FADS1	rs174546	T	0.33	0.0169	0.2314
HRAS	rs4963176	C	0.34	0.0176	0.2325
CETP	rs1800777	A	0.03	0.0196	0.2490
SMARCA2	rs17712152	A	0.05	0.0237	0.2912
PPARG	rs709158	G	0.33	0.0291	0.3345
LPL	rs3208305	T	0.30	0.0291	0.3345
PIAS1	rs1489599	G	0.43	0.0421	0.4681
INSIG1	rs9769506	C	0.43	0.0437	0.4713
SREBF2	rs222814	C	0.24	0.0476	0.4988

SNP, single nucleotide polymorphism; MAF, minor allele frequency; FDR_q, false discovery rate adjusted q value.

[a] Plasma total cholesterol levels of subjects who took lipid-lowering medication at the time of survey were treated as missing during the analysis.

[b] *P* values were derived from random coefficient (MIXED) models after adjustment for age, sex, current smoking status, alcohol consumption and body mass index.

[c] Adjusted for multiple testing (356 SNPs on the autosome) with the FDR method.

[d] Association was significant in men only and was adjusted for multiple testing with the FDR method (for 5 SNPs on the X chromosome).

risk alleles. The GRS improved the discriminative accuracy of hypercholesterolemia, measured by the AUC, from 0.705 (95%CI: 0.684-0.727) for conventional risk factors (age, sex, current smoking habit, alcohol use and BMI) to 0.734 (95%CI: 0.713-0.755) when the count GRS was added ($p < 0.0001$) (Fig. 1). Similar improvement was obtained when the weighted GRS was included (data not shown).

Discussion

In this longitudinal study, out of 361 SNPs in 243 genes, 23 SNPs were associated with plasma total cholesterol levels. Eleven of them in 9 genes were reported in previous genetic association or GWA studies on associations with blood total or LDL cholesterol levels. Two (rs2275543 in ABCA1 and rs7275 in SMARCA4) were found to be associated with cholesterol levels in a meta-analysis of seven GWA studies. A gene risk score based on these significant SNPs was strongly associated with TC levels and the prevalence of hypercholesterolemia. Four of the SNPs (rs7412 in APOE, rs646776 in CELSR2, rs662799 in APOA5, and rs7275 in SMARCA4) interacted with follow-up time on TC levels.

Rs2275543 in ABCA1 was in high LD with the recently reported rs3905000 ($r2 = 0.89$) known to be strongly associated with total and HDL cholesterol levels (4). Compared to the major T allele of rs2275543, the C allele was associated with lower TC levels (Supplementary Table 3). This cholesterol decreasing effect (both HDL and LDL cholesterol levels) associated with the C allele was also observed in the results of a meta-analysis of seven GWA studies (Supplementary Table 1) (5). Rare mutations in ABCA1 in humans, causing Tangier disease, were reported to affect not only plasma HDL cholesterol levels (6% of normal), but also total cholesterol levels (32% of normal) and LDL cholesterol levels (37% of normal) (27). SMARCA4 genetic variation (rs1529729) was recently reported to be associated with serum LDL cholesterol levels ($r2=0.38$ between rs1529729 and rs7275 in SMARCA4) (6). However, we think that rs7275 might tag certain functional SNPs in the LDLR gene due to its proximity to this gene (29kb upstream). This may explain the association between rs7275 and TC levels observed in our study.

Among the 4 SNPs that interacted with follow-up time on TC levels, rs662799 in APOA5 was recently reported to interact with dietary fat intake. G allele carriers had higher TC levels compared to AA homozygotes only in the high fat intake group (28). Therefore, we explored whether fat intake explained the strong TC increase over time in GG homozygotes observed in our study. However, adjustment for fat intake did not alter our results and no statistically significant interaction between dietary fat intake and rs662799 was detected (data not shown). Higher TC levels with alcohol consumption (29) and stronger positive associations between HDL cholesterol levels and alcohol consumption (30) have been reported in APOE ε2 carriers compared to ε3 or ε4 carriers. However, although alcohol consumption increased over the follow-up in our study, no statistically significant

Table 3. Plasma total cholesterol levels over three rounds of measurement according to gene risk score.[a, c]

	Count Gene Risk Score (N)								*p* for trend
	≤11 (230)	12 (316)	13 (509)	14 (661)	15 (672)	16 (495)	17 (280)	≥18 (174)	
TC (1st measurement)	5.11±0.89	5.17±1.04	5.28±0.94	5.41±1.02	5.60±1.06	5.62±0.99	5.71±1.01	5.92±1.18	<0.0001
TC (2nd measurement)	5.13±0.89	5.24±1.03	5.33±0.97	5.46±1.03	5.62±1.01	5.69±0.95	5.82±0.98	5.87±1.11	<0.0001
TC (3rd measurement)	5.42±0.97	5.53±0.99	5.60±1.01	5.71±1.04	5.83±1.01	5.95±0.99	6.07±0.96	6.10±1.18	<0.0001
Average TC over three measurements	5.23±0.82	5.32±0.93	5.41±0.88	5.55±0.95	5.72±0.96	5.78±0.90	5.90±0.92	6.03±1.11	<0.0001
	Weighted Gene Risk Score (N)								
	<12 (86)	12~13 (95)	13~14 (177)	14~15 (313)	15~16 (364)	16~17 (356)	17~18 (177)	≥18 (101)	
TC (1st measurement) [b]	4.74±0.95	5.25±0.91	5.25±0.94	5.38±1.04	5.45±1.06	5.64±0.97	5.70±0.99	5.95±1.09	<0.0001
TC (2nd measurement) [b]	4.97±1.22	5.11±0.89	5.33±0.89	5.38±1.01	5.49±1.01	5.73±0.98	5.76±0.91	6.04±1.16	<0.0001
TC (3rd measurement) [b]	5.20±1.31	5.53±0.90	5.56±0.88	5.61±0.95	5.79±1.05	5.92±0.99	6.10±1.02	6.22±1.18	<0.0001
Average TC over three measurements	4.97±1.05	5.31±0.81	5.39±0.83	5.47±0.91	5.59±0.95	5.79±0.90	5.89±0.91	6.11±1.06	<0.0001

TC: plasma total cholesterol levels in mmol/L; values presented as mean ± SD.

[a] The count gene risk score represents the number of unfavorable alleles (raising total cholesterol levels) at 12 SNPs. These 12 SNPs were rs7412 and rs429358 in APOE, rs646776 in CELSR2, rs1367117 in APOB, rs6756629 in ABCG5, rs662799 in APOA5, rs688 in LDLR, rs10889353 in DOCK7, rs2304130 in NCAN, rs3846662 in HMGCR, rs2275543 in ABCA1, and rs7275 in SMARCA4. The SNPs included in the weighted gene risk score computation were the same as in the count gene risk score. See the Methods section for detailed weighted gene risk score computation. The subjects who took lipid-lowering medication at the time of survey were excluded from the analysis.

[b] The analyses of association between weighted gene risk score and TC levels were conducted in a random half of the subjects (n = 1669, see the Method section for the details).

[c] The explained variation in average TC levels: 5.3% for the count gene risk score and 7.2% for the weighted gene risk score.

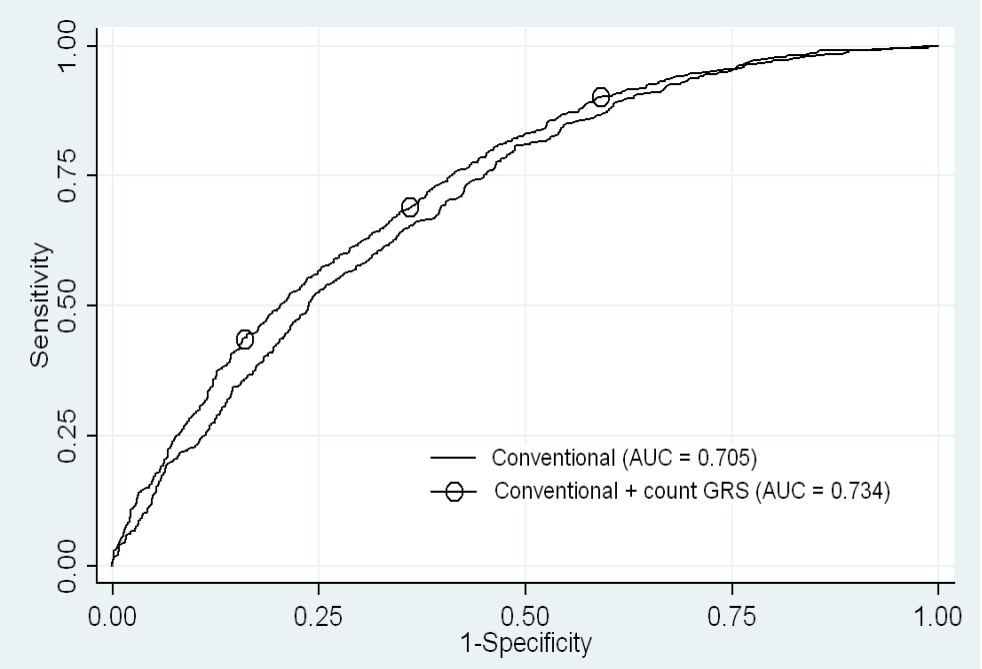

Fig. 1. Receiver-operating characteristic curves for hypercholesterolemia (total cholesterol levels 6.5 mmol/L). The curves are based on logistic regression models incorporating conventional risk factors (age, sex, current smoking habit, alcohol use and BMI) with and without the count GRS. AUC = area under the curve; GRS = gene risk score.

interaction between APOE genotype and alcohol consumption was observed (data not shown). Therefore these environmental factors could not explain our gene*time interactions. They may represent a differential effect of aging on TC levels according to genotype, but we can not rule out the possibility that our results are due to chance findings. This should be investigated further in other studies.

We investigated whether the replicated SNPs in aggregate could predict plasma cholesterol levels. A statistically significant positive trend was consistently observed in each round of survey with each count GRS point associated with an increase of 0.11 mmol/L in average plasma TC levels. This effect is similar in size to the effect per count GRS point based on 21 independent SNPs in 9 genes or loci on LDL cholesterol levels recently observed by Talmud et al.(6). These 21 SNPs could explain 14.6% of the LDL cholesterol variance. In our study, the variance in TC levels explained by the 12 SNPs was smaller (6.9%) and slightly higher in the low-fat (cholesterol) intake group compared to the high-fat (cholesterol) intake group. Additional independent variants (including common or rare SNPs and structural variants) in the genes included in our study or other genes, such as PCSK9 (5, 7), or context-dependent genetic effects (gene-gene or gene-environment

interactions) are likely to contribute in part to the remaining unexplained variance (31). Additionally, with fine-mapping to identify the causal variants underlying most of the reported associations, more variance will be explained in the future (31, 32). The weighted GRS that accounts for the strength of allelic effects performed slightly better than the count GRS on association with TC levels in our study. However, the improvement was not very substantial as also shown by other studies (6, 22). This may be because the effect for each allele tends to be normally distributed in most populations, and alleles with large effects are counterbalanced by those with smaller effects. When summing these effects, the weighted mean approximates that of the unweighted mean (33).

It is well understood that for a particular disease or trait, most individuals will have inherited some sequence variants that confer increased risk and some variants that provide protection, resulting in an overall risk around the average. However, a small proportion of people will have inherited mainly variants that confer risk of developing disease (32). In our study, carrying $\geq$ 18 TC raising alleles (5% of the studied population) confers an OR of 2.4 of the carrier having hypercholesterolemia. This indicates that occupancy of extreme distributions of certain traits (hypercholesterolemia in this case) is achieved in some individuals by carriage of a large repertoire of common alleles of modest effect (32, 34). In others, it will result perhaps, from carrying a small number of rare alleles of large effect, e.g. in LDLR. Addition of the GRS significantly improved the discriminative accuracy of hypercholesterolemia beyond that afforded by conventional risk factors with a 3% increase of the AUC. However, the actual predictive ability might have increased even more, as the increase in AUC may be an insensitive measure of the improvement in risk prediction when a novel risk factor is considered (22). Unfortunately, our current design prohibited us from estimating the predictive power of GRS more precisely by using other methods, such as net reclassification improvement (35). Nevertheless, our results suggest that a gene risk score based on a panel of comprehensive and independent risk alleles could be used to help identify those people who are at risk of high plasma cholesterol levels and enable early preventive strategies. Such scores may be more relevant to cardiovascular disease risk prediction because it may reflect life-time exposure better compared to a single-time-point measurement of cholesterol levels (26, 34). A recent study demonstrated that a GRS composed of 11 SNPs (including most of our top SNPs, such as rs10889353, rs646776, rs3846662, rs2304130, rs6756629, etc) was significantly associated with coronary heart disease and intima media thickness (even after adjusting for blood TC levels for intima media thickness) (4).

We adopted the false discovery rate method to adjust for multiple testing. As we took a candidate gene approach (and not random markers on the genome), Bonferroni correction would have been too stringent and we consider it justified to take a more liberal threshold of FDR_q-value <0.2. A number of factors could have resulted in a type II error, leading to the inability to detect a true underlying association. Firstly, only a limited number of SNPs within a candidate gene have been studied. Failure to find an association with a SNP does not exclude the possibility that other SNPs in the gene are related to TC levels. Secondly, our candidate gene list may still not be broad enough, as some genes that are recently found to be involved in the cholesterol metabolism were not taken into account, such as PCSK9

(5, 7). Our blood samples were taken from the subjects in a non-fasting condition. However, TC levels at most change minimally in response to normal food intake in individuals in the general population (36); therefore, we think the non-fasting state has not influenced our results.

In summary, in this relatively large, broad-gene based association study, we found that the common variants in genes in regulating cholesterol biosynthesis (HMGCR), VLDL metabolism (APOE, APOA5), LDL metabolism (APOB, LDLR), HDL metabolism (ABCA1, APOA5), intestinal or hepatic cholesterol efflux (ABCG5) affect plasma TC levels. We also demonstrated that a panel of SNPs in genes pivotal in cholesterol metabolism could possibly help identify those people who are likely to have high plasma cholesterol levels.

Acknowledgements

The Doetinchem Cohort Study was financially supported by the Ministry of Health, Welfare and Sport of The Netherlands and the National Institute for Public Health and the Environment. The authors thank the epidemiologists and fieldworkers of the Municipal Health Service in Doetinchem for their contribution to the data collection for this study. Logistic management was provided by J Steenbrink and P Vissink, and administrative support by EP van der Wolf. Data management was provided by A Blokstra, AWD van Kessel and PE Steinberger. The authors would also like to thank B Hoebee for her role in the conception of the study, C Strien for DNA preparation, Ruben vant ‘t Slot and HM Hodemaekers for their role in genotyping, and K Siezen for her assistance in SNP selection, and also E Reiling for critically reading the manuscript. Y Lu and A Vaarhorst are funded by grant 2006B195 of the Netherlands Heart Foundation.

Supplementary data:

http://www.sciencedirect.com/science/article/pii/S0021915010006957

References:

1. Havel, R.J. Triglyceride-rich lipoproteins and plasma lipid transport. *Arterioscler Thromb Vasc Biol* 30:9-19.
2. Schwartz, C.C., VandenBroek, J.M., and Cooper, P.S. 2004. Lipoprotein cholesteryl ester production, transfer, and output in vivo in humans. *J Lipid Res* 45:1594-1607.
3. Davidsson, P., Hulthe, J., Fagerberg, B., and Camejo, G. Proteomics of apolipoproteins and associated proteins from plasma high-density lipoproteins. *Arterioscler Thromb Vasc Biol* 30:156-163.
4. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., et al. 2009. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 41:47-55.
5. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., et al. 2009. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41:56-65.
6. Talmud, P.J., Drenos, F., Shah, S., Shah, T., Palmen, J., Verzilli, C., Gaunt, T.R., Pallas, J., Lovering, R., Li, K., et al. 2009. Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. *Am J Hum Genet* 85:628-642.

7. Chasman, D.I., Pare, G., Mora, S., Hopewell, J.C., Peloso, G., Clarke, R., Cupples, L.A., Hamsten, A., Kathiresan, S., Malarstig, A., et al. 2009. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet* 5:e1000730.
8. Lusis, A.J., Mar, R., and Pajukanta, P. 2004. Genetics of atherosclerosis. *Annu Rev Genomics Hum Genet* 5:189-218.
9. Mora, S., Otvos, J.D., Rifai, N., Rosenson, R.S., Buring, J.E., and Ridker, P.M. 2009. Lipoprotein particle profiles by nuclear magnetic resonance compared with standard lipids and apolipoproteins in predicting incident cardiovascular disease in women. *Circulation* 119:931-939.
10. Desvergne, B., Michalik, L., and Wahli, W. 2006. Transcriptional regulation of metabolism. *Physiol Rev* 86:465-514.
11. Verschuren, W.M., Blokstra, A., Picavet, H.S., and Smit, H.A. 2008. Cohort profile: the Doetinchem Cohort Study. *Int J Epidemiol* 37:1236-1241.
12. Ocke, M.C., Bueno-de-Mesquita, H.B., Pols, M.A., Smit, H.A., van Staveren, W.A., and Kromhout, D. 1997. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. *Int J Epidemiol* 26 Suppl 1:S49-58.
13. Kattermann, R., Jaworek, D., Moller, G., Assmann, G., Bjorkhem, I., Svensson, L., Borner, K., Boerma, G., Leijnse, B., Desager, J.P., et al. 1984. Multicentre study of a new enzymatic method of cholesterol determination. *J Clin Chem Clin Biochem* 22:245-251.
14. Myers, G.L., Cooper, G.R., Henderson, O., Hassemer, D.J., and Kimberly, M. 1997. Standardization of Lipid and Lipoprotein Measurements. In: Rifai N, Warnick GR, Dominiczak MH (eds). Handbook of Lipoprotein Testing. *Washington DC:AACC Press*.
15. Verschuren, W., Leer, E., Blokstra, A., Seidell, J., Smit, H.A., BuenodeMesquita, H., Obermann-de Boer, G., and Kromhout, D. 1993. Cardiovascular disease risk factors in The Netherlands. *Neth J Cardiol* 6:205-210.
16. Houterman, S., Verschuren, W.M., Oomen, C.M., Boersma-Cobbaert, C.M., and Kromhout, D. 2001. Trends in total and high density lipoprotein cholesterol and their determinants in The Netherlands between 1993 and 1997. *Int J Epidemiol* 30:1063-1070.
17. Xu, H., Gregory, S.G., Hauser, E.R., Stenger, J.E., Pericak-Vance, M.A., Vance, J.M., Zuchner, S., and Hauser, M.A. 2005. SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatics* 21:4181-4186.
18. Lu, Y., Dolle, M.E., Imholz, S., van 't Slot, R., Verschuren, W.M., Wijmenga, C., Feskens, E.J., and Boer, J.M. 2008. Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations. *J Lipid Res* 49:2582-2589.
19. Lettre, G., Lange, C., and Hirschhorn, J.N. 2007. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet Epidemiol* 31:358-362.
20. Twisk, J.W. 2003. Applied Longitudinal Data Analysis for Epidemiology, A Practical Guide.
21. Smith, N.L., Hindorff, L.A., Heckbert, S.R., Lemaitre, R.N., Marciante, K.D., Rice, K., Lumley, T., Bis, J.C., Wiggins, K.L., Rosendaal, F.R., et al. 2007. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *Jama* 297:489-498.
22. Cornelis, M.C., Qi, L., Zhang, C., Kraft, P., Manson, J., Cai, T., Hunter, D.J., and Hu, F.B. 2009. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Ann Intern Med* 150:541-550.
23. Klos, K.L., Sing, C.F., Boerwinkle, E., Hamon, S.C., Rea, T.J., Clark, A., Fornage, M., and Hixson, J.E. 2006. Consistent effects of genes involved in reverse cholesterol transport on plasma lipid and apolipoprotein levels in CARDIA participants. *Arterioscler Thromb Vasc Biol* 26:1828-1836.

24. Paynter, N.P., Chasman, D.I., Pare, G., Buring, J.E., Cook, N.R., Miletich, J.P., and Ridker, P.M. Association between a literature-based genetic risk score and cardiovascular events in women. *Jama* 303:631-637.
25. Zhu, H., Tucker, H.M., Grear, K.E., Simpson, J.F., Manning, A.K., Cupples, L.A., and Estus, S. 2007. A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol. *Hum Mol Genet* 16:1765-1772.
26. Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burtt, N.P., Roos, C., Hirschhorn, J.N., Berglund, G., Hedblad, B., Groop, L., et al. 2008. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 358:1240-1249.
27. Serfaty-Lacrosniere, C., Civeira, F., Lanzberg, A., Isaia, P., Berg, J., Janus, E.D., Smith, M.P., Jr., Pritchard, P.H., Frohlich, J., Lees, R.S., et al. 1994. Homozygous Tangier disease and cardiovascular disease. *Atherosclerosis* 107:85-98.
28. Mattei, J., Demissie, S., Tucker, K.L., and Ordovas, J.M. 2009. Apolipoprotein A5 polymorphisms interact with total dietary fat intake in association with markers of metabolic syndrome in Puerto Rican older adults. *J Nutr* 139:2301-2308.
29. Bednarska-Makaruk, M., Rodo, M., Markuszewski, C., Rozenfeld, A., Swiderska, M., Habrat, B., and Wehr, H. 2005. Polymorphisms of apolipoprotein E and angiotensin-converting enzyme genes and carotid atherosclerosis in heavy drinkers. *Alcohol Alcohol* 40:274-282.
30. Loktionov, A., Scollen, S., McKeown, N., and Bingham, S.A. 2000. Gene-nutrient interactions: dietary behaviour associated with high coronary heart disease risk particularly affects serum LDL cholesterol in apolipoprotein E epsilon4-carrying free-living individuals. *Br J Nutr* 84:885-890.
31. 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., et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747-753.
32. Donnelly, P. 2008. Progress and challenges in genome-wide association studies in humans. *Nature* 456:728-731.
33. He, M., Cornelis, M.C., Franks, P.W., Zhang, C., Hu, F.B., and Qi, L. Obesity genotype score and cardiovascular risk in women with type 2 diabetes mellitus. *Arterioscler Thromb Vasc Biol* 30:327-332.
34. Hegele, R.A. 2009. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet* 10:109-121.
35. Pencina, M.J., D'Agostino, R.B., Sr., D'Agostino, R.B., Jr., and Vasan, R.S. 2008. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 27:157-172; discussion 207-112.
36. Langsted, A., Freiberg, J.J., and Nordestgaard, B.G. 2008. Fasting and nonfasting lipid levels: influence of normal food intake on lipids, lipoproteins, apolipoproteins, and cardiovascular risk prediction. *Circulation* 118:2047-2056.

Chapter 5

Research article

Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL cholesterol levels in the Doetinchem Cohort Study

Yingchang Lu, Edith JM Feskens, Martijn ET Dollé, Sandra Imholz, W.M.Monique Verschuren, Michael Müller, and Jolanda MA Boer

Am J Clin Nutr. 2010;92(1):258-65.

Abstract

Background: The δ-5 and δ-6 desaturases, encoded by the FADS1 and FADS2 genes, are rate-limiting enzymes in polyunsaturated fatty acid (PUFA) biosynthesis. Single-nucleotide polymorphisms in the FADS gene cluster region have been associated with both PUFA levels in plasma or erythrocyte membrane phospholipids and cholesterol levels in recent genome wide association studies.

Objective: We examined whether genetic variations in the FADS gene cluster region interact with dietary intakes of n-3 and n-6 PUFAs to affect plasma total, HDL and non-HDL cholesterol levels.

Design: Rs174546, rs482548, and rs174570 in the FADS gene cluster region, dietary intakes of n-3 and n-6 PUFAs, and plasma levels of total and HDL cholesterol were measured in 3575 subjects of the second survey of the Doetinchem Cohort Study.

Results: Significant associations between rs174546 genotypes and total and non-HDL cholesterol levels were observed in the group with a high intake of n-3 PUFAs (≥0.51% of total energy; P = 0.006 and 0.047, respectively), but not in the low-intake group (P for interaction = 0.32 and 0.51, respectively). The C allele was associated with high total and non-HDL cholesterol levels. Furthermore, the C allele was significantly associated with high HDL cholesterol levels in the group with a high intake of n-6 PUFAs (≥5.26% of total energy, P = 0.004), but not in the group with a low intake (P for interaction = 0.02).

Conclusion: Genetic variation in the FADS1 gene potentially interacts with dietary PUFA intakes to affect plasma cholesterol levels, which should be investigated further in other studies.

INTRODUCTION

The δ-5 and δ-6 desaturases, encoded by *FADS1* and *FADS2* genes, are rate-limiting enzymes in the biosynthesis of long-chain polyunsaturated n-3 (omega-3) and n-6 (omega-6) fatty acids (n-3 PUFAs and n-6 PUFAs). They introduce *cis* double bonds at specific positions in a fatty acid chain (1, 2). *FADS3* shares 62% and 70% nucleotide sequence homology with *FADS1* and *FADS2* respectively and encodes for an as-yet-uncharacterized protein (3, 4). These three genes are located on chromosome 11 (11q12-13.1) and form the *FADS* gene cluster (**Figure 1**) (3). Several single-nucleotide polymorphisms (SNPs) and haplotypes in this region have been shown to be associated with blood or erythrocyte membrane phospholipid PUFA concentrations in either candidate or genome-wide association (GWA) studies (5-11). Some of these SNPs have also been associated with blood cholesterol levels in recent GWA studies (12-14). Recent emerging evidence suggests that the observed association between genetic variation in the *FADS* gene cluster region with blood cholesterol levels is functionally related to the availability of PUFAs with four and more double bonds and its impact on the homeostasis of different glycerophospholipids (10). It has long been known that blood cholesterol levels are influenced by dietary PUFA intakes (15-21); however, the underlying molecular mechanisms of these associations are still unclear. We want to investigate whether the SNPs in the *FADS* gene cluster region participate in the influence of dietary PUFAs on blood cholesterol levels (22, 23). In the present study, we evaluated the potential interaction between three candidate SNPs in the *FADS* gene cluster region and dietary n-3 and n-6 PUFA intakes on plasma total, HDL and non-HDL cholesterol levels.

SUBJECTS AND METHODS

Study population

This study was performed within the framework of the Doetinchem Cohort Study, a regional survey aimed at monitoring risk factors for chronic diseases, conducted in a rural area in the east of the Netherlands. Institutional review boards approved the Doetinchem Cohort Study. A detailed description of the design and methods was published elsewhere (24). Briefly, subjects were surveyed between 1987 and 1991 for baseline information and a sub-sample of the subjects was followed up about every 5 or 6 years. The second survey of the Doetinchem Cohort Study was approved by the ethical review board of TNO and all participants provided their written informed consent. The subjects were surveyed on demographic, anthropometric and lifestyle information (smoking, alcohol use, physical activity and dietary habits), disease history and medications by questionnaires. A non-fasting blood sample was taken from all participants, fractionated into plasma, white blood cells and erythrocytes and subsequently stored. A validated semi-quantitative food frequency questionnaire was used in the second and third surveys to assess the habitual consumption of 178 food items during the previous year (25). Nutrient and energy intake were quantified for each individual using the updated computerized Dutch food composition table (26). Detailed dietary n-3 and n-6 PUFA intakes were only available in the second survey; therefore, the data of the second survey were used in the present study.

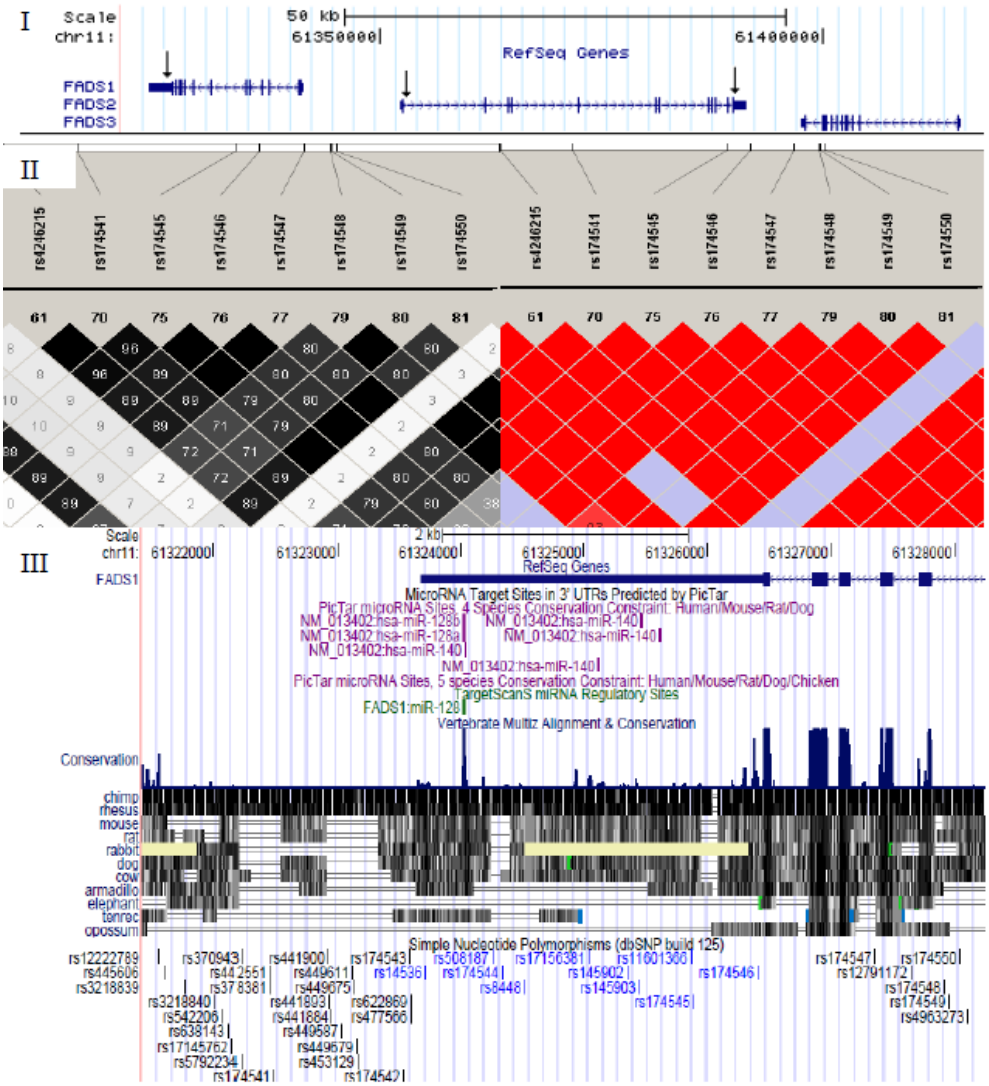

Figure 1. Genomic context of rs174546, rs174570 and rs482548 and surrounding SNPs in the *FADS* gene cluster region. The genomic locations of genes in the *FADS* gene cluster region on chromosome 11 (61.32-61.42 Mb) are shown. Arrows indicate rs174546, rs174570 and rs482548 according to their position on the genomic sequence. The data are from the UCSC Genome Browser 'RefSeq Gene' track (I). Pairwise disequilibrium coefficients for SNPs with MAF > 0.1% from HapMap CEU data (Phase II, release 22) in a 6.5kb interval (between rs4246215 and rs174550) in the 3'UTR of *FADS1* in which rs174546 is located are generated using Haploview software. The strength of the linkage disequilibrium between SNPs increases from white to black (r^2 = 1, depicted by black diamonds) or from purple to red′(Đ 1, depicted by red diamonds) using Haploview's standard color scheme (II). The predicted or experimentally verified miRNA binding sites relative to rs174546 from 'PicTar miRNA and T-ScanS miRNA' tracks at UCSC Genome Browser are shown as small vertical lines,

pink and green, respectively. MULTIZ vertebrate alignment of 11 mammal species shows evolutionary conservation (66). The lowest row shows the position of known SNPs according to dbSNP build 125 (III).

Characteristics of the study population have been reported before (27, 28). In brief, 48% of them were male. The mean age at the second survey was 46.7 ± 9.8 [SD] years. They had consistent smoking habits, and women were not pregnant at any of the three surveys. Intake of n-3 PUFAs (a-linolenic acid [ALA; 18:3n-3], eicosapentaenoic acid [EPA; 20:5n-3], all-cis-7,10,13,16,19-docosapentaenoic acid [DPA; 22:5n-3], and docosahexaenoic acid [DHA; 22:6n-3]) and n-6 PUFAs (linoleic acid [LA; 18:2n-6], eicosadienoic acid [EDA; 20:2n-6], dihomo-γ-linolenic acid [20:3n-6], arachidonic acid [AA; 20:4n-6], adrenic acid [22:4n-6], and all-cis-4,7,10,13,16-docosapentaenoic acid [22:5n-6]) were calculated in grams per day. N-3 PUFAs and n-6 PUFAs were summed up separately and converted to the percentage of the energy.

Laboratory assessment of total and HDL cholesterol levels.
Plasma total and HDL cholesterol levels were assayed in the Lipid Reference Laboratory (LRL) of the University Hospital Dijkzigt in Rotterdam using standardized enzymatic methods within three weeks after storage. Total cholesterol was measured using a CHOD-PAP method (Boehringer) (29). HDL cholesterol was determined in the supernatant after precipitation of apoB-containing lipoproteins with phophotungstic acid/$MgCl_2$ (Boehringer) (30). Non-HDL cholesterol includes all lipoproteins that contain apolipoprotein B (apo B) and were calculated as total cholesterol minus HDL cholesterol (31). The LRL Rotterdam is a permanent member of the International Cholesterol Reference Method Laboratory Network (32). It has been standardized to the Centers for Diseases Control and Prevention (CDC) through participating in the CDC/National Heart Lung and Blood Institute Lipid Standardization Program. The accuracy of total and HDL cholesterol determination fulfilled National Cholesterol Education Program (NCEP) recommendations throughout the entire period (32-34).

SNP Selection and Genotype determination
SNPs were determined in 3575 subjects who participated in three surveys. Two SNPs (rs174546 in *FADS1* 3′ UTR and rs482548 in *FADS2* 3′ UTR) in the *FADS* gene cluster region (targets genes [*FADS1* and *FADS2*] regulated by transcription factor SREBP1α (35)) were selected using the web-based program SNPselector that can prioritize the SNPs based on their potential functional impact in each gene, minor allele frequency, and haplotype block information (36). At the time of SNP selection, no study on associations of them with metabolic endpoints was available. They were included in a high throughput SNP genotyping platform-Illumina Golden Gate assay (Illumina Inc, San Diego, California). Illumina GenCall software (version 6.1.3.28) was used for automated genotype clustering and calling. The detailed quality control has been reported before (27). Another SNP (rs174570 in *FADS2* intron_1) in *FADS* gene cluster region was selected later based on its latest published associations with blood cholesterol levels (12) and genotyped by KBioscience (Hoddesdon, Hertfordshire, UK) using the KASPar chemistry, which is a competitive allele specific PCR SNP genotyping system using FRET quencher cassette oligonucleotides (http://www.kbioscience.co.uk).

Statistical Analysis

Deviation from Hardy-Weinberg equilibrium (HWE) and pairwise linkage disequilibrium (LD) were tested using R genetics package. The relationships between SNPs and plasma total, HDL and non-HDL cholesterol levels were explored with General Linear Models (GLM) using SAS version 9.1 software (SAS Institute, Cary, NC). To adjust for potential confounding effects and to improve model fitting, age, sex, current smoking (yes or no), alcohol use (glasses of alcohol per day), body mass index (BMI in kg/m^2), fasting status (blood taken after≥8 hours fasting: yes or no), physical activity (meeting the guideline of physical activity: yes or no), and cholesterol lowering medication were added to the models as covariates. Multiple linear regression models were used to assess the relation between n-3 PUFAs or n-6 PUFAs intake as percentage of energy and plasma HDL cholesterol levels with additional adjustment for n-6 PUFAs or n-3 PUFAs intakes (respectively) as percentage of energy if necessary. Dietary n-3 and n-6 PUFA intakes were also classified into two groups according to the PUFA-specific median value of the percentage of energy: n-3 PUFAs (low, <0.51 en%; high, ≥ 0.51 en%) and n -6 PUFAs (low, <5.26 en%; high, ≥5.26 en%). The interactions between genotypes and dichotomized n-3 PUFAs (n-6 PUFAs) intakes on total, HDL and non-HDL cholesterol levels were explored using GLM models by including additional interaction terms into the models. Distributions of continuous variables in groups were expressed as means ± SEMs. All reported p values were two-tailed, and statistical significance was defined at $\alpha = 0.05$ level based on our candidate SNP approach.

RESULTS

The genotype distributions of the three SNPs studied were consistent with HWE expectations (Supplementary Table 1). The three SNPs were not in strong pairwise LD (r^2 = 0.01-0.36). Mean n-3 PUFA intake was 0.53% of the total energy intake, while mean n-6 PUFA intake was 5.47% of the total energy intake. PUFA intake did not significantly differ according to *FADS* genotypes (*P* = 0.12-0.99, Supplementary Table 2).

Subjects with high n-3 PUFA intakes had significantly higher HDL cholesterol levels than subjects with low n-3 PUFA intake (*P* = 0.02, Supplementary Table 3). No significant associations between n-6 PUFA intake and any of the lipid parameters were observed. Mean total, HDL and non-HDL cholesterol levels according to the genotypes of three SNPs are presented in **Table 1**. No associations were found between any of the SNPs and HDL or non-HDL cholesterol levels, but total cholesterol levels differed significantly according to the rs174546 genotype (*P* = 0.02). The C allele was associated with higher total cholesterol levels. However, this association was more pronounced and only statistically significant (*P* = 0.006) in subjects with a high n-3 PUFA intake (**Table 2**). In the high n-3 PUFA intake group, the C allele was also associated with higher non-HDL cholesterol levels (*P* = 0.047). However, the *P* values for interaction between rs174546 genotypes and n-3 PUFA intake on total and non-HDL cholesterol levels were not statistically significant (*P* for interaction = 0.32 and 0.51, respectively). Also, n-6 PUFA intake modified the association of the rs174546 genotype with lipid levels **(Table 3)**. The C allele was associated with a statistically significant higher HDL cholesterol levels in subjects with a high n-6 PUFA intake (*P* = 0.004), while no difference was observed in group with a low n-6 PUFA intake

(*P* = 0.59; *P* for interaction = 0.024). PUFA intake did not modify associations between the other two SNPs (rs174570 and rs482548) and total, HDL or non-HDL cholesterol levels. A modest correlation between n-3 and n-6 PUFA intakes was observed in this population (ρ = 0.23, $p < 0.0001$). Further adjustments of the above associations between PUFA intakes or SNPs and plasma total, HDL and non-HDL cholesterol levels with corresponding PUFA intake did not substantially change the results (data now shown). Results from haplotype analysis either including all three SNPs or only rs174546 and rs174570 showed that the haplotype associations seemed to be mainly derived from and were similar to those observed with rs174546 alone (data not shown).

Table 1. Plasma total cholesterol, HDL cholesterol and non-HDL cholesterol levels according to rs174546, rs174570 and rs482548 in the *FADS* gene cluster region

	Total cholesterol (mmol/L)	*P*[1]	HDL cholesterol (mmol/L)	*P*[1]	non-HDL cholesterol (mmol/L)	*P*[1]
rs174546						
TT (n = 362)	5.48 ± 0.09[2]		1.27 ± 0.03		4.21 ± 0.09	
TC (n = 1503)	5.56 ± 0.07		1.30 ± 0.03		4.26 ± 0.08	
CC (n = 1505)	5.62 ± 0.08	0.02	1.31 ± 0.03	0.09	4.31 ± 0.08	0.11
rs174570						
TT (n = 95)	5.46 ± 0.12		1.28 ± 0.04		4.18 ± 0.12	
TC (n = 874)	5.52 ± 0.08		1.31 ± 0.03		4.21 ± 0.08	
CC (n = 2559)	5.57 ± 0.07	0.25	1.30 ± 0.02	0.52	4.27 ± 0.07	0.22
rs482548						
TT (n = 40)	5.67 ± 0.16		1.30 ± 0.06		4.37 ± 0.17	
TC (n = 590)	5.56 ± 0.08		1.31 ± 0.03		4.26 ± 0.08	
CC (n = 2914)	5.55 ± 0.07	0.72	1.30 ± 0.02	0.71	4.26 ± 0.07	0.75

[1] P values for differences between genotypes were obtained by using general linear models (ANOVA) adjusted for age, sex, BMI, fasting status, current smoking status, alcohol consumption, physical activity, and cholesterol lowering medication.
[2] All values are presented as adjusted mean ± SEM.

DISCUSSION

In the present study, the previously reported association between rs174546 in *FADS1* and cholesterol levels was confirmed, while the association between rs174570 in *FADS2* and cholesterol levels was not replicated. We observed that rs174546 genotypes potentially interacted with dietary n-3 and n-6 PUFA intakes to affect total, HDL and non-HDL cholesterol levels.

Table 2. Plasma total cholesterol, HDL cholesterol and non-HDL cholesterol levels according to n-3 PUFA intake and genotypes of rs174546, rs174570 and rs482548 in the *FADS* gene cluster region[1,4]

	Total cholesterol (mmol/L)		HDL cholesterol (mmol/L)		non-HDL cholesterol (mmol/L)	
	Low intake (<0.51 en%)	High intake (≥ 0.51 en%)	Low intake (<0.51 en%)	High intake (≥ 0.51 en%)	Low intake (<0.51 en%)	High intake (≥ 0.51 en%)
rs174546						
TT (n = 362)	5.52 ± 0.113	5.46 ± 0.14	1.29 ± 0.04	1.26 ± 0.05	4.23 ± 0.12	4.20 ± 0.14
TC (n = 1503)	5.58 ± 0.10	5.57 ± 0.12	1.30 ± 0.03	1.32 ± 0.04	4.28 ± 0.10	4.26 ± 0.13
CC (n = 1505)	5.61 ± 0.10	5.68 ± 0.12	1.30 ± 0.03	1.32 ± 0.04	4.30 ± 0.10	4.36 ± 0.13
P[2]	0.55	0.006	0.82	0.08	0.67	0.047
rs174570						
TT (n = 95)	5.53 ± 0.16	5.42 ± 0.18	1.32 ± 0.05	1.24 ± 0.06	5.53 ± 0.16	5.42 ± 0.18
TC (n = 874)	5.54 ± 0.10	5.54 ± 0.12	1.32 ± 0.03	1.30 ± 0.04	5.54 ± 0.10	5.54 ± 0.12
CC (n = 2559)	5.55 ± 0.09	5.64 ± 0.11	1.30 ± 0.03	1.30 ± 0.04	5.55 ± 0.09	5.64 ± 0.11
P[2]	0.98	0.064	0.52	0.40	0.98	0.064
rs482548						
TT (n = 40)	5.78 ± 0.22	5.55 ± 0.25	1.29 ± 0.07	1.33 ± 0.09	4.49 ± 0.22	4.22 ± 0.26
TC (n = 590)	5.55 ± 0.10	5.62 ± 0.12	1.33 ± 0.03	1.30 ± 0.04	4.22 ± 0.11	4.32 ± 0.13
CC (n = 2914)	5.54 ± 0.09	5.61 ± 0.11	1.30 ± 0.03	1.30 ± 0.04	4.24 ± 0.09	4.31 ± 0.12
P[2]	0.49	0.948	0.34	0.92	0.44	0.904

[1] n-3 PUFA, n-3 polyunsaturated fatty acids, includes a-linolenic acid (18:3n-3), eicosapentanoic acid (20:5n-3), docosapentaenoic acid (22:5n-3), and docosahexaenoic acid (22:6n-3); n-3 PUFA intake was classified into low and high intake according to the median intake expressed as the percentage of energy.

[2] P values for differences between genotypes were obtained by using general linear models (ANOVA) adjusted for age, sex, BMI, fasting status, current smoking status, alcohol consumption, physical activity, and cholesterol lowering medication.
[3] All values are presented as adjusted mean ± SEM.
[4] P values for interaction between genotypes of rs174546, rs174570, and rs482548 and n-3 PUFA intake (low and high) were obtained by including the interaction terms in the general linear models and were 0.32, 0.31, 0.59 for total cholesterol levels, 0.45, 0.36, 0.51 for HDL cholesterol levels, and 0.51, 0.60, 0.49 for non-HDL cholesterol levels, respectively.

Table 3. Plasma total cholesterol, HDL cholesterol and non-HDL cholesterol levels according to n-6 PUFA intake and genotypes of rs174546, rs174570 and rs482548 in the *FADS* gene cluster region[1,4]

	Total cholesterol (mmol/L)		HDL cholesterol (mmol/L)		non-HDL cholesterol (mmol/L)	
	Low intake (<5.26 en%)	High intake (≥5.26 en%)	Low intake (<5.26 en%)	High intake (≥5.26 en%)	Low intake (<5.26 en%)	High intake (≥5.26 en%)
rs174546						
TT (n = 362)	5.55 ± 0.133	5.42 ± 0.12	1.30 ± 0.04	1.23 ± 0.04	4.25 ± 0.13	4.18 ± 0.12
TC (n = 1503)	5.53 ± 0.11	5.60 ± 0.10	1.29 ± 0.04	1.32 ± 0.03	4.25 ± 0.12	4.27 ± 0.11
CC (n = 1505)	5.64 ± 0.11	5.61 ± 0.10	1.30 ± 0.04	1.31 ± 0.03	4.34 ± 0.12	4.29 ± 0.11
P[2]	0.06	0.07	0.59	0.004	0.15	0.43
rs174570						
TT (n = 95)	5.63 ± 0.16	5.25 ± 0.18	1.30 ± 0.06	1.24 ± 0.06	4.33 ± 0.17	3.99 ± 0.19
TC (n = 874)	5.52 ± 0.11	5.53 ± 0.11	1.29 ± 0.04	1.33 ± 0.03	4.23 ± 0.11	4.20 ± 0.11
CC (n = 2559)	5.57 ± 0.10	5.57 ± 0.10	1.29 ± 0.04	1.31 ± 0.03	4.29 ± 0.11	4.25 ± 0.10
P[2]	0.49	0.11	0.87	0.257	0.45	0.19
rs482548						
TT (n = 40)	5.45 ± 0.29	5.74 ± 0.20	1.37 ± 0.10	1.27 ± 0.07	4.08 ± 0.30	4.47 ± 0.21
TC (n = 590)	5.52 ± 0.11	5.62 ± 0.11	1.29 ± 0.04	1.33 ± 0.04	4.23 ± 0.12	4.29 ± 0.11
CC (n = 2914)	5.57 ± 0.10	5.54 ± 0.10	1.29 ± 0.04	1.30 ± 0.03	4.29 ± 0.11	4.24 ± 0.10
P[2]	0.60	0.30	0.65	0.443	0.49	0.34

[1] n-6 PUFA, n-6 polyunsaturated fatty acids, includes linoleic acid (18:2n-6), eicosadienoic acid (20:2n-6), dihomo-γ-linolenic acid (20:3n-6), arachidonic acid (20:4n-6), adrenic acid (22:4n-6), and docosapentaenoic acid (22:5n-6); n-6 PUFA intake was classified into low and high intake according to the median intake expressed as the percentage of energy.

[2] P values for differences between genotypes were obtained by using general linear models (ANOVA) adjusted for age, sex, BMI, fasting status, current smoking status, alcohol consumption, physical activity, and cholesterol lowering medication.
[3] All values are presented as adjusted mean ± SEM.
[4] P values for interaction between genotypes of rs174546, rs174570, and rs482548 and n-6 PUFA intake (low and high) were obtained by including the interaction terms in the general linear models and were 0.13, 0.16, 0.21 for total cholesterol levels, 0.02, 0.42, 0.47 for HDL cholesterol levels, and 0.49, 0.31, 0.21 for non-HDL cholesterol levels, respectively.

The association between genetic variants in the *FADS* gene cluster region and blood cholesterol levels has been highlighted in three recent GWA studies (12-14). The minor T allele of rs174546 was associated with decreased LDL (10, 11, 14), HDL (13, 37) and total cholesterol levels (11) based on high LD among SNPs in the *FADS1* 3′UTR region (Figure 1). Several miRNA target sites have been predicted in this region (Figure 1) (38, 39). Based on PolymiRTS and Patrocles databases, rs174546 is within one of the miRNA target sites, and it is also in complete LD with another miRNA target site polymorphism (rs174545) in this region, which results in disruption of one of the conserved miRNA target sites. The potential impact on *FADS1* transcript abundance of rs174546 (or other SNPs in high LD with rs174546 in this region) is further corroborated by recently published human transcription data. The major C allele of rs174546 was positively associated with *FADS1* expression in lymphoblastoid cells (40), and also in the liver (13, 41), and adipose tissue (42) based on complete LD between rs174546 and rs174547. Rs174546 is in high LD with recently reported SNPs that are associated with plasma or erythrocyte membrane PUFA levels (5-11). It has been observed that the haplotype including the major C allele is associated with an increase in the levels of desaturase products, while associated with a decrease of its substrates (5-11). Despite the lack of a study directly assaying desaturase activity so far, all the above described evidence indicates that the major C allele of rs174546 may be associated with an increased efficiency of the fatty acid δ-5 desaturase reaction. The potential mechanism underlying the association between high desaturase activity and high total or LDL cholesterol levels (10, 11, 14) and high HDL cholesterol levels (13, 37) is still unclear. Since δ-5 and δ-6 desaturases are highest expressed in the liver (1, 2, 4), a major contribution of the liver desaturase–elongase enzyme system to the observed association between genetic variation in *FADS1* and *FADS2* genes and plasma cholesterol metabolism seems likely.

In our population, significant associations between rs174546 and total and non-HDL cholesterol levels were only found in the group with a high n-3 PUFA intake, not in the group with a low intake. Non-HDL cholesterol is highly correlated with total apo B levels (31, 43, 44) and could indirectly indicate the total number of apo B containing particles (VLDL, IDL and LDL) (43-45). The major C allele of rs174546 is associated with increased levels of n-3 PUFA with four and more double bonds (5-11), whereas n-3 PUFA intervention decreases hepatic triglyceride synthesis and VLDL secretion (15, 18, 21). We would therefore expect that C allele-associated high desaturase activity would be associated with low non-HDL cholesterol levels in a high dietary n-3 PUFA intake population. However, we observed a non-HDL cholesterol increasing effect. Considering the kinetics of the whole apo B pool, this may suggest that with a high dietary n-3 PUFA intake C allele-associated high desaturase activity suppresses the uptake of apo B containing particles, such as VLDL remnants or LDL. This hypothesis is supported by both positive associations between serum n-3 PUFA levels and LDL cholesterol levels in a Japanese observational study (46) and increased LDL cholesterol levels in EPA and DHA or fish oil intervention studies (15-18, 20). It is also mechanistically supported by in vitro and in vivo observations of reduced rat hepatic precursor and nuclear (much stronger) form of *SREBP-1* with fish oil (or DHA) treatment (47-49). One of the isoforms encoded by *SREBP-1* (*SREBP-1α*) is a potent transcription factor regulating *LDLR* gene expression (35, 50). Our

lab recently demonstrated that feeding fish oil to mice downregulated hepatic *LDLR* gene expression (1.9 fold, *Q*-value = 4.23E-6; M. Boekschoten, personal communication), consistent with the previous observation in rabbits (51).

A statistically significant interaction between dietary n-6 PUFA intakes and rs174546 genotypes on HDL cholesterol levels was observed in our study. There was a significant difference in HDL cholesterol levels among rs174546 genotypes in the high n-6 PUFA intake group. In controlled trials, substituting carbohydrates isoenergetically with PUFAs (largely n-6) is associated with increased HDL cholesterol levels (19). Additionally, serum n-6 PUFA levels were positively associated with HDL cholesterol levels in both Caucasian and Japanese middle-aged men (46). The more efficient desaturase reaction associated with the major C allele (see above), together with a high dietary intake of n-6 PUFAs, could lead to much larger differences in the availability of n-6 PUFAs with four and more double bonds between C allele carriers and TT homozygotes of rs174546 (10) as compared to a situation when n-6 PUFA intake is low. These PUFAs, and especially their metabolic derivatives are potent activators of PPARs (52-55), resulting in the increased expression of genes directly involved in HDL production (56, 57). Interestingly, however, in TT homozygotes, HDL cholesterol levels were even lower in the high n-6 PUFA intake group than in the low intake group (P = 0.014). From a quantitative point of view, linoleic acid (precursor of n-6 PUFAs) is 10-100 times more abundant in Western diets than α-linolenic acid (precursor of n-3 PUFAs)(8). N-3 and n-6 PUFAs competitively use the same desaturase–elongase enzyme system (58), and n-3 PUFAs are relatively potent PPARs ligands compared to n-6 PUFAs (53-55, 58). With a high n-6 PUFA intake, linoleic acid accumulates, especially in TT homozygotes. This potentially poses a strong inhibiting effect on n-3 PUFAs' metabolism (58-61) and their associated HDL cholesterol increasing effects (17, 18, 21, 46, 62). This hypothesis is supported by our own observation that the association between dietary n-3 PUFA intake and HDL cholesterol levels became stronger after further adjustment for dietary n-6 PUFA intakes (Supplementary Table 4).

The present study has some limitations. Firstly, we did not measure the internal levels (plasma or tissue membrane) of different PUFAs that are directly involved in the underlying biological processes. However, relatively strong correlations between dietary intakes of PUFAs and internal PUFA levels have been observed (15, 17, 59-61). Furthermore, the detailed questionnaire used in this study covered all food items habitually consumed in the Netherlands and was well validated for micronutrient intake (25). Secondly, despite the relatively large sample size, the interactions between n-3 PUFA intakes and rs174546 genotypes on total and non-HDL cholesterol levels did not reach statistical significance (P > 0.05), as larger sample size is needed to obtain statistical significance. Thirdly, considering our candidate SNP approach, the associations were not adjusted for multiple testing. However, our findings should be replicated in other studies. Fourthly, although non-HDL cholesterol and apo B levels do not change while other lipid profiles at most change minimally in response to normal food intake in the general population (63-65), our results based on the non-fasting condition should be investigated in other conditions in independent populations.

In conclusion, *FADS1* genetic polymorphisms may play an important modulating role in the influence of dietary n-3 and n-6 PUFAs on plasma cholesterol levels. The health benefit of specific PUFAs may depend on our individual *FADS1* genotypes or δ-5 desaturase activity.

ACKNOWLEDGEMENTS

We thank the epidemiologists and fieldworkers of the Municipal Health Service in Doetinchem for their contribution to the data collection for this study. Logistic management was provided by J Steenbrink and P Vissink, administrative support by EP van der Wolf, data management by A Blokstra, AWD van Kessel and PE Steinberger. We would also like to thank B Hoebee for her role in the conception of the study, C Strien for DNA preparation, K Siezen for her assistance in SNP selection, and HM Hodemaekers for her assistance with the genotyping, and M Boekschoten for communicating unpublished data. We are further indebted to Cisca Wijmenga and Ruben vant 't Slot for providing the genotyping facilities.

The authors' responsibilities were as follows study concept and design: YL, JMAB, and EJMF; Acquisition of data: METD, SI, JMAB, EJMF, WMMV, and MM; Analysis and interpretation of data: YL, JMAB, EJMF, and MM; writing of the manuscript: YL; Critical revision of the manuscript: JMAB, EJMF, MM, METD, SI, and WMMV. None of the authors had any conflict of interest.

Supplementary data:

http://www.ajcn.org/content/92/1/258/suppl/DC1

REFERENCES

1. Cho HP, Nakamura MT, Clarke SD. Cloning, expression, and nutritional regulation of the mammalian Delta-6 desaturase. J Biol Chem 1999;274:471-7.
2. Cho HP, Nakamura M, Clarke SD. Cloning, expression, and fatty acid regulation of the human delta-5 desaturase. J Biol Chem 1999;274:37335-9.
3. Marquardt A, Stohr H, White K, Weber BH. cDNA cloning, genomic structure, and chromosomal localization of three members of the human fatty acid desaturase family. Genomics 2000;66:175-83.
4. Pedrono F, Blanchard H, Kloereg M, et al. The fatty acid desaturase 3 (Fads3) gene encodes for different FADS3 protein isoforms in mammalian tissues. J Lipid Res 2009.
5. Schaeffer L, Gohlke H, Muller M, et al. Common genetic variants of the FADS1 FADS2 gene cluster and their reconstructed haplotypes are associated with the fatty acid composition in phospholipids. Hum Mol Genet 2006;15:1745-56.
6. Malerba G, Schaeffer L, Xumerle L, et al. SNPs of the FADS gene cluster are associated with polyunsaturated fatty acids in a cohort of patients with cardiovascular disease. Lipids 2008;43:289-99.
7. Xie L, Innis SM. Genetic variants of the FADS1 FADS2 gene cluster are associated with altered (n-6) and (n-3) essential fatty acids in plasma and erythrocyte phospholipids in women during pregnancy and in breast milk during lactation. J Nutr 2008;138:2222-8.
8. Martinelli N, Girelli D, Malerba G, et al. FADS genotypes and desaturase activity estimated by the ratio of arachidonic acid to linoleic acid are associated with inflammation and coronary artery disease. Am J Clin Nutr 2008;88:941-9.

9. Rzehak P, Heinrich J, Klopp N, et al. Evidence for an association between genetic variants of the fatty acid desaturase 1 fatty acid desaturase 2 (FADS1 FADS2) gene cluster and the fatty acid composition of erythrocyte membranes. Br J Nutr 2009;101:20-6.
10. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet 2008;4:e1000282.
11. Tanaka T, Shen J, Abecasis GR, et al. Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. PLoS Genet 2009;5:e1000338.
12. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 2009;41:47-55.
13. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 2009;41:56-65.
14. Sabatti C, Service SK, Hartikainen AL, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet 2009;41:35-46.
15. Harris WS. Fish oils and plasma lipid and lipoprotein metabolism in humans: a critical review. J Lipid Res 1989;30:785-807.
16. Harris WS. n-3 fatty acids and serum lipoproteins: human studies. Am J Clin Nutr 1997;65:1645S-1654S.
17. Harris WS, Ginsberg HN, Arunakul N, et al. Safety and efficacy of Omacor in severe hypertriglyceridemia. J Cardiovasc Risk 1997;4:385-91.
18. Harris WS, Miller M, Tighe AP, Davidson MH, Schaefer EJ. Omega-3 fatty acids and coronary heart disease risk: clinical and mechanistic perspectives. Atherosclerosis 2008;197:12-24.
19. Mensink RP, Zock PL, Kester AD, Katan MB. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. Am J Clin Nutr 2003;77:1146-55.
20. Theobald HE, Chowienczyk PJ, Whittall R, Humphries SE, Sanders TA. LDL cholesterol-raising effect of low-dose docosahexaenoic acid in middle-aged men and women. Am J Clin Nutr 2004;79:558-63.
21. Chan DC, Watts GF, Barrett PH, Beilin LJ, Redgrave TG, Mori TA. Regulatory effects of HMG CoA reductase inhibitor and fish oils on apolipoprotein B-100 kinetics in insulin-resistant obese male subjects with dyslipidemia. Diabetes 2002;51:2377-86.
22. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 2001;358:1356-60.
23. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet 2005;6:287-98.
24. Verschuren WM, Blokstra A, Picavet HS, Smit HA. Cohort profile: the Doetinchem Cohort Study. Int J Epidemiol 2008;37:1236-41.
25. Ocke MC, Bueno-de-Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol 1997;26 Suppl 1:S49-58.
26. NEVO. S. Nederlands Voedingsstoffenbestand 1996. Den Haag: Voorlichtingsbureau voor de Voeding, 1996 (in Dutch). 1996.
27. Lu Y, Dolle ME, Imholz S, et al. Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations. J Lipid Res 2008;49:2582-9.
28. van den Berg SW, Dolle ME, Imholz S, et al. Genetic variations in regulatory pathways of fatty acid and glucose metabolism are associated with obesity phenotypes: a population-based cohort study. Int J Obes (Lond) 2009;33:1143-52.
29. Kattermann R, Jaworek D, Moller G, et al. Multicentre study of a new enzymatic method of cholesterol determination. J Clin Chem Clin Biochem 1984;22:245-51.
30. Lopes-Virella MF, Stone P, Ellis S, Colwell JA. Cholesterol determination in high-density lipoproteins separated by three different methods. Clin Chem 1977;23:882-4.

31. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002;106:3143-421.
32. Myers GL, Cooper GR, Henderson O, Hassemer DJ, Kimberly M. Standardization of Lipid and Lipoprotein Measurements. In: Rifai N, Warnick GR, Dominiczak MH(eds). Handbook of Lipoprotein Testing. Washington DC: AACC Press. 1997.
33. Bachorik PS, Ross JW. National Cholesterol Education Program recommendations for measurement of low-density lipoprotein cholesterol: executive summary. The National Cholesterol Education Program Working Group on Lipoprotein Measurement. Clin Chem 1995;41:1414-20.
34. Warnick GR, Wood PD. National Cholesterol Education Program recommendations for measurement of high-density lipoprotein cholesterol: executive summary. The National Cholesterol Education Program Working Group on Lipoprotein Measurement. Clin Chem 1995;41:1427-33.
35. Horton JD, Shah NA, Warrington JA, et al. Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. Proc Natl Acad Sci U S A 2003;100:12027-32.
36. Xu H, Gregory SG, Hauser ER, et al. SNPselector: a web tool for selecting SNPs for genetic association studies. Bioinformatics 2005;21:4181-6.
37. Hicks AA, Pramstaller PP, Johansson A, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet 2009;5:e1000672.
38. Lall S, Grun D, Krek A, et al. A genome-wide map of conserved microRNA targets in C. elegans. Curr Biol 2006;16:460-71.
39. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15-20.
40. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. Nat Genet 2007;39:1202-7.
41. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol 2008;6:e107.
42. Plaisier CL, Horvath S, Huertas-Vazquez A, et al. A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. PLoS Genet 2009;5:e1000642.
43. Vega GL, Grundy SM. Does measurement of apolipoprotein B have a place in cholesterol management? Arteriosclerosis 1990;10:668-71.
44. Abate N, Vega GL, Grundy SM. Variability in cholesterol content and physical properties of lipoproteins containing apolipoprotein B-100. Atherosclerosis 1993;104:159-71.
45. Elovson J, Chatterton JE, Bell GT, et al. Plasma very low density lipoproteins contain a single molecule of apolipoprotein B. J Lipid Res 1988;29:1461-73.
46. Motoyama KR, Curb JD, Kadowaki T, et al. Association of serum n-6 and n-3 polyunsaturated fatty acids with lipids in 3 populations of middle-aged men. Am J Clin Nutr 2009;90:49-55.
47. Botolin D, Wang Y, Christian B, Jump DB. Docosahexaneoic acid (22:6,n-3) regulates rat hepatocyte SREBP-1 nuclear abundance by Erk- and 26S proteasome-dependent pathways. J Lipid Res 2006;47:181-92.
48. Jump DB. N-3 polyunsaturated fatty acid regulation of hepatic gene transcription. Curr Opin Lipidol 2008;19:242-7.
49. Xu J, Nakamura MT, Cho HP, Clarke SD. Sterol regulatory element binding protein-1 expression is suppressed by dietary polyunsaturated fatty acids. A mechanism for the coordinate suppression of lipogenic genes by polyunsaturated fats. J Biol Chem 1999;274:23577-83.
50. Wang X, Sato R, Brown MS, Hua X, Goldstein JL. SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis. Cell 1994;77:53-62.

51. Wilkinson J, Higgins JA, Fitzsimmons C, Bowyer DE. Dietary fish oils modify the assembly of VLDL and expression of the LDL receptor in rabbit liver. Arterioscler Thromb Vasc Biol 1998;18:1490-7.
52. Kliewer SA, Sundseth SS, Jones SA, et al. Fatty acids and eicosanoids regulate gene expression through direct interactions with peroxisome proliferator-activated receptors alpha and gamma. Proc Natl Acad Sci U S A 1997;94:4318-23.
53. Forman BM, Chen J, Evans RM. Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. Proc Natl Acad Sci U S A 1997;94:4312-7.
54. Xu HE, Lambert MH, Montana VG, et al. Molecular recognition of fatty acids by peroxisome proliferator-activated receptors. Mol Cell 1999;3:397-403.
55. Pawar A, Jump DB. Unsaturated fatty acid regulation of peroxisome proliferator-activated receptor alpha activity in rat primary hepatocytes. J Biol Chem 2003;278:35931-9.
56. Duval C, Muller M, Kersten S. PPARalpha and dyslipidemia. Biochim Biophys Acta 2007;1771:961-71.
57. Akiyama TE, Sakai S, Lambert G, et al. Conditional disruption of the peroxisome proliferator-activated receptor gamma gene in mice results in lowered expression of ABCA1, ABCG1, and apoE in macrophages and reduced cholesterol efflux. Mol Cell Biol 2002;22:2607-19.
58. Schmitz G, Ecker J. The opposing effects of n-3 and n-6 fatty acids. Prog Lipid Res 2008;47:147-55.
59. Hussein N, Ah-Sing E, Wilkinson P, Leach C, Griffin BA, Millward DJ. Long-chain conversion of [13C]linoleic acid and alpha-linolenic acid in response to marked changes in their dietary intake in men. J Lipid Res 2005;46:269-80.
60. Goyens PL, Spilker ME, Zock PL, Katan MB, Mensink RP. Conversion of alpha-linolenic acid in humans is influenced by the absolute amounts of alpha-linolenic acid and linoleic acid in the diet and not by their ratio. Am J Clin Nutr 2006;84:44-53.
61. Liou YA, King DJ, Zibrik D, Innis SM. Decreasing linoleic acid with constant alpha-linolenic acid in dietary fats increases (n-3) eicosapentaenoic acid in plasma phospholipids in healthy men. J Nutr 2007;137:945-52.
62. Volcik KA, Nettleton JA, Ballantyne CM, Boerwinkle E. Peroxisome proliferator-activated receptor [alpha] genetic variation interacts with n-6 and long-chain n-3 fatty acid intake to affect total cholesterol and LDL-cholesterol concentrations in the Atherosclerosis Risk in Communities Study. Am J Clin Nutr 2008;87:1926-31.
63. Warnick GR, Nakajima K. Fasting versus nonfasting triglycerides: implications for laboratory measurements. Clin Chem 2008;54:14-6.
64. Langsted A, Freiberg JJ, Nordestgaard BG. Fasting and nonfasting lipid levels: influence of normal food intake on lipids, lipoproteins, apolipoproteins, and cardiovascular risk prediction. Circulation 2008;118:2047-56.
65. Craig SR, Amin RV, Russell DW, Paradise NF. Blood cholesterol screening influence of fasting state on cholesterol results and management decisions. J Gen Intern Med 2000;15:395-9.
66. Blanchette M, Kent WJ, Riemer C, et al. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 2004;14:708-15.

Chapter 6

Research article

Comparative Transcriptomic and Metabolomic Analysis of Fenofibrate and Fish Oil Treatments in Mice

Yingchang Lu*, Mark V. Boekschoten*,
Suzan Wopereis, Michael Müller[#], Sander Kersten[#]
*equal contribution
#joint corresponding authors

Physiological Genomics (In press)

Summary:

Elevated circulating triglycerides, which are considered a risk factor for cardiovascular disease, can be targeted by treatment with fenofibrate or fish oil. To gain insight into underlying mechanisms, we carried out a comparative transcriptomics and metabolomics analysis of the effect of 2 week treatment with fenofibrate and fish oil in mice. Plasma triglycerides were significantly decreased by fenofibrate (-49.1%) and fish oil (-21.8%), whereas plasma cholesterol was increased by fenofibrate (+29.9%) and decreased by fish oil (-32.8%). Levels of various phospholipid species were specifically decreased by fish oil, while levels of Krebs cycle intermediates were increased specifically by fenofibrate. Plasma levels of many amino acids were altered by fenofibrate and to a lesser extent by fish oil. Both fenofibrate and fish oil upregulated genes involved in fatty acid metabolism, and downregulated genes involved in blood coagulation and fibrinolysis. Significant overlap in gene regulation by fenofibrate and fish oil was observed, reflecting their property as high or low affinity agonist for PPARα, respectively. Fenofibrate specifically downregulated genes involved in complement cascade and inflammatory response. Fish oil specifically downregulated genes involved in cholesterol and fatty acid biosynthesis, and upregulated genes involved in amino acid and arachidonic acid metabolism. Taken together, the data indicate that despite being similarly potent towards modulating plasma free fatty acids, cholesterol and triglyceride levels, fish oil causes modest changes in gene expression likely via activation of multiple mechanistic pathways, whereas fenofibrate causes pronounced gene expression changes via a single pathway, reflecting the key difference between nutritional and pharmacological intervention.

Introduction

One of the components of the metabolic syndrome and frequent complication of insulin resistance is hypertriglyceridemia (7, 14). Similar to elevated plasma cholesterol levels, hypertriglyceridemia is considered an independent risk factor for atherosclerosis (7, 14, 52). Elevation of circulating triglycerides (TG) in insulin resistance is related to the diverse actions of insulin on hepatic secretion of very low density lipoprotein (VLDL) particles (1).

Apart from weight loss and increased physical activity, both of which improve insulin sensitivity, a limited number of therapeutic options exist to lower circulating TG (15, 29, 56). A class of drugs that effectively lowers plasma TG are the fibrates. Reduction in plasma TG by fibrate treatment depends on baseline TG levels but can reach 40% (15). Fibrates also raise plasma levels of HDL-cholesterol, making it an attractive drug for the treatment of diabetic dyslipidemia, which is characterized by elevated TG and low HDL levels. Overall results of clinical trials, however, are mixed and not overwhelmingly in support of a broad application of fibrates to correct dyslipidemia and lower risk of coronary heart disease. The decrease in plasma TG by fibrates has been attributed to inhibition of synthesis and secretion of TG by the liver and stimulation of degradation of TG-rich lipoproteins (29, 56). Fibrates stimulate a large panel of genes involved in hepatic mitochondrial and peroxisomal fatty acid oxidation and lipoprotein metabolism by direct high-affinity binding and consequent activation of peroxisome proliferator-activated receptor alpha (PPARα) (13, 27, 29, 44, 56), thus serving as a direct pharmacological ligand of PPARα. PPARα is a transcription factor abundant in liver that mediates the adaptive response to fasting.

Besides via pharmacological approach, plasma TG can be effectively lowered via nutritional intervention in the form of fatty fish or fish oil. Fish oil, rich in eicosapentaenoic acid (EPA, C20:5n-3) and docosahexaenoic acid (DHA, C22:6n-3), has long been considered as potential treatment to lower risk of coronary heart disease (12, 18, 19, 38). Similar to fibrates, fatty acids present in fish oil induce hepatic expression of numerous genes via activation of PPARα (51). In addition, EPA and DHA suppress activity of the pro-lipogenic transcription factor sterol regulatory element binding protein-1 (SREBP-1) by inhibiting proteolytic processing of SREBP-1, a process required to generate the active mature SREBP-1 protein (24, 25). While fenofibrate and fish oil thus both lower plasma TG and can activate the same transcription factor, a comparative analysis of the effects of fenofibrate and fish oil at transcriptome and metabolome level has yet to be performed. Therefore, to gain further insight into mechanisms underlying the effects of fenofibrate and fish oil on cardiovascular risk factors and to investigate whether these mechanisms are shared between fenofibrate and fish oil, we performed hepatic transcriptional profiling and plasma metabolite profiling in mice treated with fenofibrate or fish oil for two weeks. Inasmuch as fenofibrate and fish oil are already known to both stimulate PPARα-dependent gene regulation, the focus of the analysis was on genes and pathways uniquely regulated by either fenofibrate or fish oil.

Materials and Methods

Animals and Experimental Design.
12-week-old male C57BL/6 mice (Charles River, L'Arbresle Cedex, France) were housed 2 per cage in a light-and-temperature controlled facility (lights on 6:30 to 18:30, 21°C) and acclimated for 3 weeks. The mice were randomized by weight-matching into three groups (n = 10 in control group, and n = 12 in fenofibrate or fish oil intervention group), and fed a research diet (No. D10012M, Research Diets, NJ, Supplementary Table 1) supplemented with sunflower oil (containing 81.3% oleic acid, 7% energy intake) in control group, sunflower oil (7% energy intake) and fenofibrate (0.03% w/w) in fenofibrate group, and fish oil (Marinol C-38 fish oil: 23.1% EPA and 21.1% DHA, 7% energy intake) in fish oil group for 2 weeks. Mice received fresh diet every 3rd day, and food consumption rate and body weight gain were monitored. At the end of treatment, mice were fasted from 7:00 to 13:00 with drinking water available and were subsequently sacrificed by cervical dislocation under isoflurane anesthesia. Blood was collected via orbital puncture. Livers were dissected, directly frozen in liquid nitrogen and stored at−80°C until further analysis. Blood was centrifuged (4000 × g for 10 min at 4°C), and plasma was stored at−80°C. The animal experiments were approved by the animal ethics committee of Wageningen University.

Affymetrix GeneChip microarray analysis.
Total RNA was prepared from mouse livers using Trizol reagent (Invitrogen, Breda, The Netherlands), treated with DNase and purified on columns using the RNeasy Mini Kit (Qiagen, Venlo, The Netherlands) following the supplier's protocol. RNA concentrations were measured by absorbency at 260 nm, and the quality and integrity were verified with the RNA 6000 Nano assay on the Agilent 2100 Bioanalyzer (Agilent Technologies, Amsterdam, the Netherlands) according to the manufacturer's instructions. Microarray analysis was performed on individual mouse livers. Five micrograms of RNA were labelled using the Affymetrix One –Cycle Target Labelling Assay kit (Affymetrix, Santa Clara, CA). Hybridization, washing and scanning of Affymetrix Mouse Genome 430 2.0 Arrays were done according to standard Affymetrix protocols. Scans of the Affymetrix arrays were processed using packages from the Bioconductor project (16). Raw signal intensities were normalized by using the GCRMA algorithm (69). Probesets were defined according to Dai et al. using remapped CDF version 11.0.2 based on the Entrez gene database (10). The Affymetrix Mouse Genome 430 2.0 Arrays target 16,331 unique genes based on this CDF. Genes were filtered on expression value >20 in 5 samples, resulting in a set of 7,400 expressed genes. The Bioconductor R package Linear models for microarray data (LIMMA) was used to identify differentially expressed genes. In order to balance between unspecific responses from the two treatments and relative weak transcriptional effects by fish oil, the genes that met the cut-off of mean absolute fold change > 1.2 and false discovery rate corrected q-value < 0.05 (55) were considered to be significantly regulated. Among the significantly regulated genes, only those that are associated with a canonical pathway in the Ingenuity Pathway Knowledge Base were considered for Ingenuity Pathway Analysis (IPA, Ingenuity Systems, Redwood City, CA). In addition, all genes represented on the array were also considered for the unbiased Gene Set Enrichment Analysis (GSEA) (58). This analysis was run using 1000 permutations per gene set. All microarray data are

MIAME compliant and have been submitted to the Gene Expression Omnibus (accession number GSE pending).

Lipidomic and metabolic profiling.
The LC-MS methods for measuring plasma lipids and non-esterified free fatty acids (NEFAs) and the GC-MS method for measuring a broad range of metabolites were identical to the methods reported by Wopereis et al (68). The samples were analyzed in randomized order. Data for each sample were corrected for the recovery of the internal standard for injection. The performance of the methods was carefully monitored by using multiple internal standards (5–10 depending on the method, including analogs and 2H- and 13C-labeled metabolites) as described previously (68). Furthermore, a quality control sample prepared by pooling of plasma from all samples was analyzed after every 10th study sample. Batches were only accepted if the relative standard deviation (RSD) of the peak area ratio for all internal standards was < 20%. Metabolites were only accepted if the RSD was < 20%, unless large differences between treatment groups were observed. Batch to batch differences in data were removed by synchronizing medians of quality control samples per batch. Metabolites were annotated by using an in-house metabolite database containing retention time information, MS spectra (electron impact ionization for GC-MS data), MS/MS spectra (LC-MS), and accurate mass data (LC-MS) of reference substances. The confidence of identification was 100% unless indicated otherwise. Accurate MS and MS/MS data of reference substances and metabolites in the study samples were acquired by using Thermo LTQ-FT and Thermo LTQ-Orbitrap instruments (Thermo Fisher Scientific, Waltham, MA). Finally, the LC-MS NEFA dataset contained 22 free fatty acids; the LC-MS lipid dataset contained 184 lipid metabolites; and the GC-MS dataset contained 137 metabolites. Also, 41 different metabolite ratios and sums were calculated. Detailed information on these metabolites could be acquired upon request.

Statistical analysis.
Results are reported as mean ± SEM. The comparison of different groups was carried out using analysis of variance (ANOVA) and unpaired 2-tailed Student's t test. The Kruskal-Wallis test or Mann-Whitney U test was used if groups did not show equal variance. Spearman rank correlation was used to correlate hepatic gene expression signals with plasma metabolite levels. Differences were considered statistically significant when $p < 0.05$.

Results

Changes in liver and body weight and selected plasma metabolites
Mice received fenofibrate (0.03% w/w) or fish oil (3% w/w) in their feed for 2 weeks. Neither fenofibrate nor fish oil influenced food intake (Table 1). Fish oil but not fenofibrate significantly increased bodyweight (+6.2% vs +2.7%), while fenofibrate but not fish oil increased liver weight (+1.8% vs -0.1%, respectively). Fenofibrate raised plasma total cholesterol (+29.9%), whereas fish oil had the opposite effect (-32.8%). Both fenofibrate and fish oil reduced plasma triglycerides although the effect of fenofibrate was more pronounced (-49.1% vs -21.8%). Consistent with stimulation of hepatic fatty acid oxidation, fenofibrate markedly raised plasma ketone bodies (+316%). A similar effect was

observed for ribose. Interestingly, fish oil significantly reduced plasma levels of phospholipids (-42.3% for phosphatidylcholine and phosphatidylethanolamine), lysophospholipids (-28.0% for lysophosphatidylcholine and lysophosphatidylethanolamine) and sphingomyelins (-19.4%), which was not observed for fenofibrate.

Fenofibrate reduced plasma total NEFA levels (-18.8%), with its effect equally distributed among the various fatty acid classes (SFA, MUFA, PUFA). Fish oil lowered total plasma NEFA to a similar extent as fenofibrate but its effect across the various fatty acid classes was different. Fish oil decreased levels of all n-6 and n-9 long-chain fatty acids and, as expected, increased plasma levels of n-3 polyunsaturated fatty acid levels (Table 2). The stronger reduction in monounsaturated fatty acids by fish oil compared to fenofibrate was mainly attributable to reduced levels of C18:1 fatty acids and presumably due to lower content of oleic acid in the feed of fish oil treated mice (Table 2 and Supplementary Table 1).

Table 1. General characteristics and selected plasma metabolite levels

Parameter	Control (n = 10)	Fenofibrate (n = 12)	Fish oil (n = 12)
Food intake (g/day)	3.3 ± 0.1	3.5 ± 0.1	3.3 ± 0.1
Body weight change (%)	1.6 ± 0.6	2.7 ± 0.8	6.2 ± 0.8***
Liver weight (% body weight)	3.9 ± 0.1	5.7 ± 0.1***	3.8 ± 0.1
Cholesterol levels‡	0.67 ± 0.05	0.87 ± 0.05*	0.45 ± 0.04**
Triacylglycerol (mmol/L)§	0.55 ± 0.05	0.28 ± 0.02***	0.43 ± 0.03*
Triacylglycerol†	75.6 ± 6.2	38.9 ± 3.1***	42.2 ± 2.9**
Total ketone bodies‡	0.30 ± 0.09	0.95 ± 0.08***	0.26 ± 0.02
β-hydroxybutyrate‡	0.29 ± 0.09	0.91 ± 0.08***	0.25 ± 0.02
Glucose (mmol/L)§	9.91 ± 0.49	10.63 ± 0.30	10.14 ± 0.55
Glucose‡	7.2 ± 0.7	8.5 ± 0.5	7.5 ± 0.5
Ribose‡	0.012 ± 0.003	0.030 ± 0.006*	0.015 ± 0.002
Pyruvic acid‡	0.93 ± 0.10	0.92 ± 0.07	0.69 ± 0.07
Lactic acid‡	12.30 ± 1.10	11.55 ± 0.74	10.05 ± 0.39
Glycerol‡	0.25 ± 0.03	0.19 ± 0.01	0.24 ± 0.02
PC†	21.57 ± 0.90	21.21 ± 0.69	12.27 ± 0.54***
LPC†	23.95 ± 0.88	24.94 ± 0.72	17.20 ± 0.34***
LPE†	0.42 ± 0.03	0.40 ± 0.02	0.35 ± 0.02
PE†	0.87 ± 0.07	0.73 ± 0.03	0.67 ± 0.03**
SM†	1.55 ± 0.05	1.53 ± 0.04	1.25 ± 0.05**

PC, phosphatidylcholine; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; PE, phosphatidylethanolamine; SM, sphingomyelin.
Data are expressed as mean ± SEM.
†LCMS platform, n = 10, 12 and 12 for control, fenofibrate and fish oil group, respectively; arbitrary units used.
‡GCMS platform, n = 9, 12 and 11 for control, fenofibrate and fish oil group, respectively; arbitrary units used.
§Triacylglycerol assayed with the glycerol based kit; and glucose assayed with glucose oxidization based kit.
*P < 0.05, **P < 0.01, ***P < 0.001 versus control.

Fenofibrate and fish oil treatments lead to PPARα activation

Since fenofibrate and EPA/DHA are known ligands of PPARα, we determined to what extent PPARα was activated by the two treatments. Fenofibrate and fish oil caused

upregulation of many genes involved in PPARα-dependent pathways such as hepatic fatty acid uptake (Fig. 1A), mitochondrial fatty acid β-oxidation (Fig. 1B), peroxisomal fatty acid β-oxidation (Fig. 1C), microsomal fatty acid ω-hydroxylation (Fig. 1D) and ketogenesis, including many classical PPARα target genes (Fig. 1E) (50). Consistent with fenofibrate being a higher affinity PPARα agonist compared to EPA/DHA, more pronounced inductions were observed in fenofibrate treated mice. These data demonstrate that fenofibrate and fish oil treatment led to activation of PPARα, in line with published data (36).

Figure 1

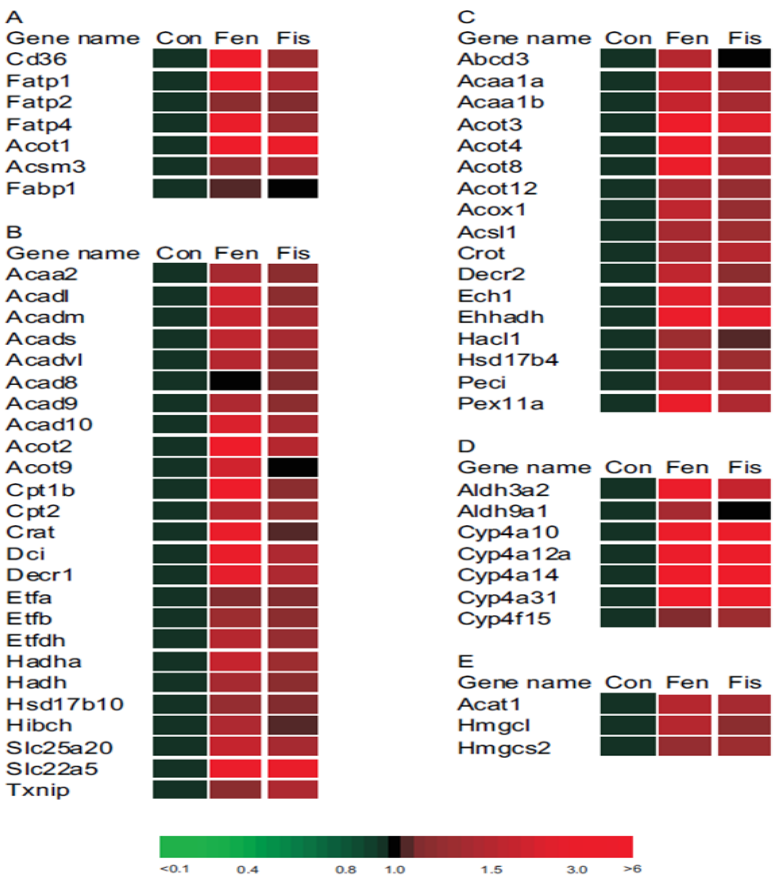

Figure 1. Heat map showing the parallel induction of specific PPARα-dependent gene sets in liver of fenofibrate and fish oil treated mice. (A) Genes involved in fatty acid uptake, activation and binding. (B) Genes involved in mitochondrial fatty acid β-oxidation/oxidative phosphorylation. (C) Genes involved in peroxisomal fatty acid β-oxidation. (D) Genes involved in microsomal fatty acid ω-hydroxylation. (E) Genes involved in ketogenesis. The expression levels in the control (Con) mice were set at 1 (black) and expression levels in fenofibrate (Fen) and fish oil (Fis) treated groups were calculated relative to the control group. The definition of

the gene sets were based on reference (50).

Comparative analysis of gene regulation by fenofibrate and fish oil

Dendrogram of hierarchical clustering of the various mice based on the complete liver transcriptome showed that fenofibrate treated mice formed a highly distinct group from fish oil treated and control mice, illustrating the much more pronounced effects of fenofibrate on hepatic gene expression compared to fish oil (Fig. 2). Furthermore, since the three groups clustered separately, the analysis indicated that the variability in gene expression within the three groups was much less compared to the variability between the groups. In order to compare the whole hepatic genome effects of fenofibrate and fish oil, scatter plot analysis was carried out (Fig. 3). The similarity in gene regulation between fenofibrate and fish oil treated mice was relatively small, and was mainly observed with respect to upregulation of gene expression. Fold inductions of gene expression were generally higher in fenofibrate compared to fish oil treated mice. Interestingly, some genes strongly upregulated by fenofibrate were downregulated by fish oil, including Ly6d, Cidec, Pdk4, Defb1, Ucp2, Cidea and Pltp. Other genes were upregulated by fish oil but not by fenofibrate, including Mt2, Derl3 and Agxt2l1, or even strongly downregulated by fenofibrate, such as Clec2h, Aox3, Cyp2c37 and Hsd3b5 (Fig. 3). As fenofibrate and fish oil both activate PPARα, differential hepatic gene regulation by fish oil compared to fenofibrate suggest another type of regulatory mechanism by fish oil treatment not involving PPARα.

Figure 2

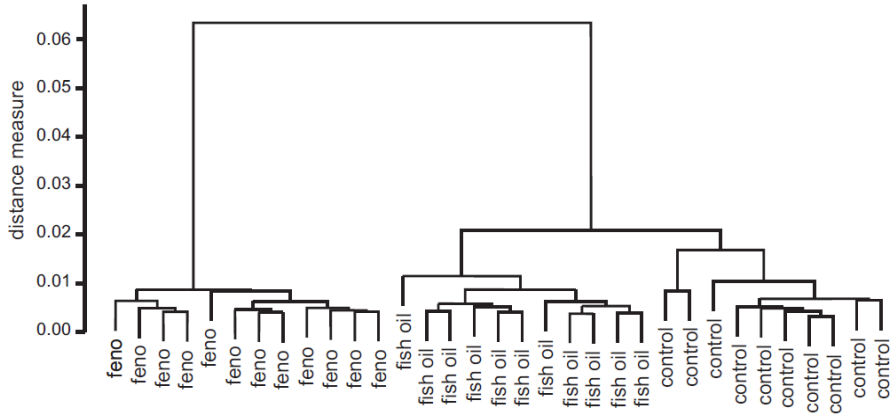

Figure 2. Hierarchical clustering of microarray data from mice treated with fenofibrate, fish oil, or receiving control treatment. Dendrogram shows separate clustering of mice treated with fenofibrate compared to fish oil or control treated mice, illustrating the much more pronounced effects of fenofibrate on hepatic gene expression compared to fish oil.

Pathway analysis of microarray data

To gain insight into pathways uniquely regulated by fenofibrate and fish oil, gene set enrichment analysis (GSEA) was performed. Partial overlap in the upregulated gene sets was observed between fenofibrate and fish oil, mostly covering various aspects of fatty acid metabolism (Fig. 4). Many gene sets were specifically induced by fenofibrate, including electron transport chain and TCA cycle (Fig. 4A). Also, a small number of gene sets was upregulated by fenofibrate but decreased by fish oil, including lipogenesis, glycolysis and gluconeogenesis, and the pentose phosphate pathway. Interestingly, autophagy, ABC transporters and arachidonic acid metabolism were specifically induced by fish oil (Fig. 4B). In terms of downregulation, blood clotting cascades, complement cascades, and antigen processing and presentation were commonly regulated by fenofibrate and fish oil. Gene sets that were exclusively downregulated by fish oil included lipoprotein metabolism, cholesterol synthesis and esterification, and prostaglandin synthesis regulation. Taken together, several pathways specifically regulated by either fenofibrate or fish oil could be identified. Expectedly, most of the common regulation by fenofibrate and fish oil relates to PPARα-dependent gene sets connected to fatty acid catabolism.

Table 2. Plasma non-esterified fatty acid levels

Free Fatty acid	Control (n = 10)	Fenofibrate (n = 12)	Fish oil (n = 12)
Total NEFA	57.3 ± 3.0	46.5 ± 1.6**	46.1 ± 2.2**
Total SFA	23.07 ± 1.08	19.45 ± 0.60**	20.99 ± 1.33
C12:0	0.37 ± 0.02	0.38 ± 0.03	0.39 ± 0.05
C14:0	1.16 ± 0.11	1.03 ± 0.08	1.30 ± 0.14
C16:0	14.7 ± 0.8	13.0 ± 0.4	13.8 ± 0.7
C17:0	0.77 ± 0.05	0.67 ± 0.03	0.74 ± 0.04
C18:0	5.92 ± 0.22	4.31 ± 0.13***	4.60 ± 0.47*
C20:0	0.11 ± 0.01	0.06 ± 0.00***	0.13 ± 0.02
Total MUFA	22.69 ± 1.71	18.76 ± 0.80*	13.36 ± 0.54***
C16:1n-7	3.85 ± 0.46	3.21 ± 0.21	3.28 ± 0.13
C18:1n-7/9	18.24 ± 1.24	15.14 ± 0.59*	9.76 ± 0.40***
C20:1n-9	0.60 ± 0.05	0.41 ± 0.02**	0.32 ± 0.02***
Total PUFA	11.06 ± 0.51	7.50 ± 0.32***	11.58 ± 0.49
Total PUFA (n-6)†	9.18 ± 0.43	6.25 ± 0.26***	4.70 ± 0.24***
Total PUFA (n-3)	1.87 ± 0.10	1.25 ± 0.06***	6.88 ± 0.32***
C18:2n-6/7	5.32 ± 0.30	3.93 ± 0.17***	3.61 ± 0.21***
C18:3n-3	0.14 ± 0.01	0.09 ± 0.01**	0.22 ± 0.01***
C18:3n-6	0.13 ± 0.02	0.07 ± 0.00***	0.07 ± 0.00**
C20:2n-6	0.10 ± 0.01	0.07 ± 0.00**	0.04 ± 0.00***
C20:3n-6	0.24 ± 0.01	0.28 ± 0.01*	0.02 ± 0.00***
C20:3n-9	0.51 ± 0.04	0.76 ± 0.04***	0.12 ± 0.01***
C20:4n-3	0.16 ± 0.01	0.09 ± 0.00***	0.12 ± 0.01**
C20:4n-6	3.12 ± 0.18	1.82 ± 0.10***	0.90 ± 0.05***
C20:5n-3	0.04 ± 0.00	0.03 ± 0.00	1.90 ± 0.10***
C22:4n-6	0.11 ± 0.00	0.06 ± 0.00***	0.02 ± 0.00***
C22:5n-3	0.07 ± 0.00	0.05 ± 0.00***	0.62 ± 0.03***

C22:5n-6	0.17 ± 0.01	0.04 ± 0.00***	0.05 ± 0.00***
C22:6n-3	1.46 ± 0.09	0.99 ± 0.05***	4.02 ± 0.19***

NEFA, non-esterified fatty acid; SFA, saturated fatty acids; MUFA, monounsaturated fatty acid; PUFA, polyunsaturated fatty acids.
All metabolites were measured in the LCMS free fatty acid platform, arbitrary units used. Data are expressed as mean ± SEM.
†C18:2n-6/7 is included in total n-6 PUFA calculation.
*P < 0.05, **P < 0.01, ***P < 0.001 versus control.

Figure 3

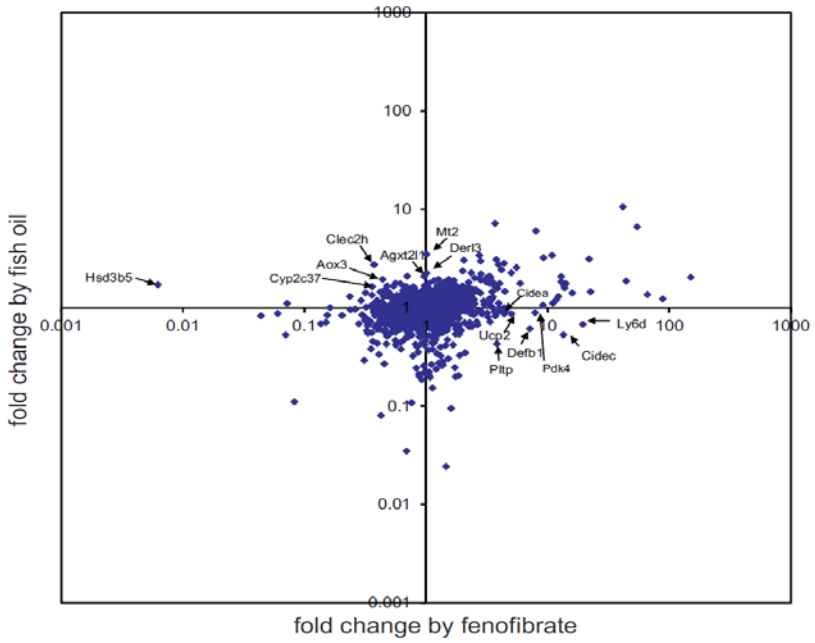

Figure 3. Similarities and differences in gene expression in livers between fenofibrate and fish oil treated mice. Scatter plot shows fold-change in gene expression after treatment with fenofibrate (x-axis) plotted against fold-change in gene expression after treatment with fish oil (y-axis). Selected genes that were upregulated or downregulated specifically by fenofibrate or fish oil are indicated.

One of the gene sets specifically induced by fenofibrate corresponded to TCA cycle. In agreement with this finding, fenofibrate but not fish oil significantly raised plasma levels of several TCA cycle intermediates, including isocitrate, α-ketoglutarate, succinate, fumarate and malate (Table 3). In accordance with altered hepatic amino acid metabolic pathways upon fenofibrate and fish oil treatment, significant changes in plasma amino acids levels were observed in both treatment groups (Supplementary Table 3). Most assayed amino acids were increased by fenofibrate, including glutamate, glycine, isoleucine, leucine, phenylalanine, serine, tryptophan, tyrosine and valine, whereas only glycine was increased by fish oil (Supplementary Table 3). These data correspond well with downregulation of

genes involved in urea cycle and metabolism of amino groups in fenofibrate treated mice (Fig. 4A) (Supplementary Table 2).

Genes specifically regulated by fenofibrate or fish oil treatment

Venn diagrams were created to identify genes that were specifically regulated by fenofibrate or fish oil treatment (Fig. 5A and B). Most genes upregulated or downregulated by fish oil were similarly regulated by fenofibrate. To better characterize fenofibrate and fish oil-specific gene regulation, we compiled the top 40 of genes specifically upregulated (Fig. 5C and D) and downregulated (Fig. 5E and F) by fenofibrate and fish oil, respectively, and ranked according to fold change. Changes in gene expression by fish oil and fenofibrate are shown in parallel.

Table 3. Plasma metabolite (organic acids) involved in TCA cycle

Parameter	Control (n = 9)	Fenofibrate (n = 12)	Fish oil (n = 11)
Citrate	1.13 ± 0.08	1.19 ± 0.07	0.98 ± 0.06
Fumaric acid	0.056 ± 0.07	0.098 ± 0.006***	0.050 ± 0.004
Isocitric acid	0.078 ± 0.003	0.092 ± 0.004*	0.075 ± 0.003
Malic acid	0.38 ± 0.05	0.63 ± 0.05**	0.36 ± 0.03
Succinic acid	0.0018 ± 0.0002	0.0030 ± 0.0002**	0.0024 ± 0.0002*
α-Ketoglutaric acid	0.027 ± 0.003	0.047 ± 0.004**	0.031 ± 0.002

All organic acids were measured in the GCMS platform, arbitrary units used. Data are expressed as mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001 versus control.

A relatively large proportion of top 40 genes specifically upregulated by fenofibrate were known PPARα target genes (50). In contrast to the classical PPARα target genes involved in fatty acid catabolism shown in Fig. 1, these genes, which included Cidec, Pdk4, Ucp2, Cidea, Pltp, Abcd2, Me1 and Fabp4, were downregulated by fish oil (Fig. 5C). Among the top 40 genes specifically downregulated by fenofibrate treatment, several genes are involved in complement cascade (C8b, C6 and C9), coagulation cascade (Serpine2 and F11) and general inflammatory regulation and response (Orm2, Bcl6, Saa1, Saa2, Ifit1, Il6ra, Il1r1 and Irf5) (Fig. 5E), which is consistent with the strongly downregulated acute phase response signaling (Ingenuity pathway analysis, data not shown), complement cascade and coagulation cascade in fenofibrate treated mice (Fig. 4A and Supplementary Table 2). Also, a number of fenofibrate-specific genes were related to steroid metabolism, which fits with reduction in androgen and estrogen metabolism as indicated by GSEA (Fig. 4A). With respect to fish oil treatment, several of the top 40 genes specifically upregulated by fish oil are involved in amino acid metabolism (Agxt2l1, Aox3, Clpx, Cyp7b1 and Aadat) and arachidonic acid metabolism (Cyp2c37, Cyp2c44 and Cyp2j9) (Fig. 5D). Most of them were downregulated by fenofibrate, suggesting that the underlying mechanism of regulation is fish oil specific and not PPARα related. Among the top 40 genes specifically downregulated by fish oil, a marked enrichment of SREBP target genes was apparent (21, 43) (Fig. 5F and Supplementary Table 2), which is consistent with the observed downregulation in both cholesterol and fatty acid biosynthetic pathways (Fig. 4B). Strikingly, expression of several SREBP1 targets involved in lipogenesis (Fasn, Elovl6, Fads2 and Fads1) was upregulated by fenofibrate, but downregulated by fish oil (Fig. 5F). The hepatic down- and upregulation of SREBP1 target stearoyl-CoA desaturase (Scd1,

Supplementary Table 2) by fish oil and fenofibrate was substantiated by corresponding changes in plasma Δ9-desaturation indices (stearoyl-CoA desaturase activity) (Table 4).

The precise molecular mechanism behind the plasma TG lowering effect of fish oil remains controversial. To gain insight into potential mechanisms, we determined which genes showed the highest correlation with plasma TG levels in the combined control and fish-oil treated mice (Fig. 6A). Remarkably, many genes showing a highly significant negative correlation with plasma TG were PPARα target genes involved in fatty acid metabolism. In comparison, lipogenic and cholesterol biosynthetic genes, despite being strongly downregulated by fish oil, showed weaker correlation with plasma TG. Similar observations were made for correlations of gene expressions with plasma cholesterol levels and PC (Fig. 6A). Ingenuity pathways analysis on genes showing the highest positive or negative correlation with plasma TG, cholesterol and PC identified "fatty acid metabolism", a PPARα regulated pathway, as the most significant pathway (data not shown). Pathways related to fatty acid or cholesterol biosynthesis were much less significant. Specific examples of genes showing a highly significant correlation with plasma TG, cholesterol or lysophosphatidylcholine are shown in Fig. 6B. These results may imply that the effects of fish oil on various plasma lipids may occur via changes in fatty acid oxidation via activation of PPARα, and to a lesser extent via suppression of fatty acid and cholesterol biosynthesis, which are under control of SREBP.

Table 4. Liver Δ9-desaturation indices (Stearoyl-CoA desaturase activity)

Δ9-desaturation indices	Control (n = 10)	Fenofibrate (n = 12)	Fish oil (n = 12)
LPC_C16:1/LPC_C16:0	0.036 ± 0.002	0.047 ± 0.001***	0.02 ± 0.001***
LPC_C18:1/LPC_C18:0	1.23 ± 0.08	2.55 ± 0.12***	0.79 ± 0.02**
LPC_C20:1/ LPC_C20:0	2.78 ± 0.20	5.55 ± 0.19***	0.86 ± 0.03***

Δ9-Desaturation indices were calculated from the ratios between C16:1(n-7) and C16:0, C18:1(n-7/n-9) and C18:0, and C20:1(n-9) and C20:0, respectively, in the plasma lysophosphatidylcholines. All lysophosphatidylcholine (LPC) were measured in the LCMS platform, arbitrary units used. Data are expressed as mean ± SEM.

*P < 0.05, **P < 0.01, ***P < 0.001 versus control

Discussion

Using a combination of transcriptional and plasma metabolite profiling, we carried out a comprehensive comparison of the effects of fenofibrate and fish oil treatment in mice. Due to the treatment duration, some of the observed changes in gene expression changes will be secondary to metabolic perturbations elicited by the treatment, although a major portion of gene expression changes likely reflects direct regulation. Both fenofibrate and fish oil induced numerous genes involved in hepatic fatty acid catabolism and other PPARα-dependent pathways. Fenofibrate consistently caused higher fold-inductions, in agreement with fenofibrate being a better PPARα agonist compared to the EPA and DHA (51). In contrast to fish oil, fenofibrate caused hepatomegaly and raised plasma ketone bodies, which are known PPARα-dependent effects. The data thus indicate that compared to fenofibrate, fish oil treatment leads to modest activation of PPARα in liver. However, while effects of fenofibrate are almost entirely mediated by PPARα, fish oil additionally acts upon other regulatory pathways to exert multiple effects, reflecting a property characteristic

of nutrients.

Consumption of fish oil lowers circulating triglycerides in humans (19), which could be reproduced in mice. Fish oil fatty acids may stimulate post-ER presecretory proteolysis of ApoB via a mechanism dependent on fatty acid peroxidation, leading to reduced triglyceride secretion by hepatocytes (47). Alternatively, lowering of circulating TG may occur via the observed downregulation of fatty acids synthesis. In addition, a number of other mechanisms has been proposed, including activation of PPARγ in adipose tissue (12, 18). Conflicting data exist on whether lowering of plasma TG and cholesterol by fish oil is dependent on PPARα (11, 59). We observed pronounced enrichment for PPARα targets and genes involved in fatty acid metabolism among genes showing the most significant negative correlation with plasma TG, cholesterol, and phospholipids. In comparison, lipogenic and cholesterol biosynthetic genes showed good correlation with plasma phospholipids but weaker correlation with plasma TG and cholesterol. Our results may suggest that the effect of fish oil on plasma TG and cholesterol primarily occurs via activation of PPARα, whereas the effect on plasma phospholipids seems to rely proportionally more on suppression of SREBP-dependent regulation of lipogenesis and cholesterol metabolism.

In contrast to earlier studies in rats but in line with the effect in humans (13), fenofibrate increased plasma cholesterol, which in mice is almost exclusively carried in HDL. The reason for the discrepancy with previous rat studies is unclear (42, 57). Besides via changes in apoAI expression, which was slightly but significantly reduced by both fenofibrate and fish oil, the HDL-raising effect of fenofibrate in humans may be mediated via changes in Pltp expression and activity (33, 66). Alternatively, fenofibrate may raise plasma cholesterol levels in mice by downregulating Scarb1 (SR-BI), which was confirmed in our study (35).

In line with numerous papers, fish oil downregulated expression of numerous SREBP target genes involved in fatty acid and cholesterol synthesis (24, 25, 59, 60). Although SREBP1 and SREBP2 have both been suggested to be inhibited by PUFAs, data implicating SREBP1 in downregulation of gene expression by PUFAs are much more plentiful. Recently, the target of PUFAs was identified as Ubxd8, a ER membrane-bound protein that facilitates the degradation of Insig-1, thereby promoting proteolytic processing and activation of SREBP-1 (30). It was shown that PUFAs inhibit the activity of Ubxd8 (30). Hence, downregulation of hepatic fatty acid and cholesterol biosynthetic genes by fish oil may occur via inhibition of Ubxd8.

While several classical PPARα targets involved in fatty acid oxidation were upregulated by both fenofibrate and fish oil, other established PPARα targets were induced only by fenofibrate, including Cidec, Cidea, Pdk4, Ucp2, Pltp, Abcd2, Me1 and Fabp4. With the exception of Abcd2, none of these genes are involved in fatty acid oxidation but instead participate in other (lipid) metabolic pathways. Interestingly, malic enzyme (Me1) is controlled by both PPARα and SREBP, as are Ucp2 (37, 50), Fads1, Fads2, Scd1, Acsl3, Acsl4 and Acsl5 (21, 50), Abcd2 (50, 67) and Fabp4 (26, 50). Similarly, the PPARα targets

Pltp and lipid droplet -associated protein Cidea were shown to be regulated by SREBP1 (43, 64). Thus, it is possible that the above gene set is induced by fish oil via PPARα but that the effect is counterbalanced by suppression of SREBP-mediated transcriptional regulation. These data indicate that responsiveness to synthetic PPARα agonist may not always properly predict regulation by dietary PPARα agonists, as the latter act via multiple mechanistic pathways that may converge on a single gene.

Data abound indicating that fish oil fatty acid and fibrates stimulate hepatic fatty acid uptake and catabolism (5), which likely explains the reduced plasma NEFA levels in fenofibrate and fish oil treated mice. Although fish oil caused much weaker induction of genes involved in fatty acid catabolic pathways compared to fenofibrate, the reduction in plasma NEFA was identical between the treatments. Fish oil may reduce plasma NEFA levels and increase weight gain by attenuating fatty acid release from adipose tissues (18, 23, 45, 48), perhaps via activation of PPARγ. EPA and DHA present in fish oil are endogenous PPARγ ligands (9, 12, 61) and fish oil has been shown to upregulate PPARγ and its responsive genes in epididymal adipose tissue (8, 9, 17, 22, 39, 61). Fish oil may thus mimic the stimulatory effect of synthetic PPARγ agonists on fatty acid trapping (9, 12, 61), and weight gain (3). Since we did not collect adipose tissue, it is impossible to determine whether fish oil induced PPARγ target genes in adipose tissue and whether it increased adipose tissue mass.

Remarkably, pathways relevant to blood coagulation and fibrinolysis were strongly downregulated by fenofibrate and fish oil. Individual genes within these pathways were consistently downregulated by fenofibrate and fish oil, although the effect of fenofibrate was much more pronounced (Supplementary Table 2). Suppression of hepatic expression of these genes upon pharmacological PPARα activation has been reported in rats and monkeys (6, 28) and was shown to be PPARα dependent for fibrinogen α, β and γ chain (28). In humans, fibrates reduce plasma fibrinogen levels by 12-25% (65, 70). Recently, several genes in this pathway were found to be downregulated in human hepatocytes by synthetic PPARα agonist (49).

A weak hypocoagulant effect of fish oil has been observed in studies in humans and is likely mediated by suppression of clotting factors. Fish oil was shown to decrease hepatic transcription of kallikrein B, fibrinogen β chain, antithrombin III, and protein C genes (60, 63) and lower blood fibrinogen, factors II, V, VII, and X, antithrombin III, and protein C levels in rodents (2, 32, 40, 41, 63). Reduced activity of factor V, VII, VIII, IX, X, XI, and XII, and protein C (20, 34) and reduced plasma levels of fibrinogen, factor V, VII and X, protein C, antithrombin III, plasminogen activator inhibitor, and α2 antiplasmin were also observed in human fish oil intervention studies (4, 20, 34, 46, 54, 62). The striking parallel downregulation of blood coagulation and fibrinolysis pathways by fenofibrate and fish oil suggest they occur via a common mediator, e.g. PPARα.

Interestingly, consistent with previous data showing enhanced TCA cycle flux (53) and increased TCA cycle enzymes (31) by synthetic PPARα ligand, plasma levels of all TCA cycle intermediates was induced by fenofibrate, perhaps as a result of enhanced amino acid

degradation by fenofibrate, as revealed by GSEA.

In conclusion, the transcriptomic and metabolomic effects of fish oil and fenofibrate reflect the highly specific activity of fenofibrate towards PPARα, whereas fish oil engages additional regulatory pathways to impact numerous biological processes. Our data provide better insight into how fish oil modulates circulating levels of lipids and other metabolites in humans. Moreover, the data may provide clues towards additional potential health benefits of fish oil.

Acknowledgements

This study was financed by the Netherlands Nutrigenomics Centre and the Netherlands Heart Foundation (2006B195). We would like to thank Rob Vreeken of the Netherlands Metabolomics Centre for the lipidomics measurements and Ivana Bobeldijk for GCMS and free fatty acids measurements. Co-authors are members of NuGO association.

Supplementary data:

Physiol. Genomics; doi:10.1152/physiolgenomics.00100.2011.

References

1. Adiels M, Olofsson SO, Taskinen MR, and Boren J. Overproduction of very low-density lipoproteins is the hallmark of the dyslipidemia in the metabolic syndrome. Arteriosclerosis, thrombosis, and vascular biology 28: 1225-1236, 2008.
2. Andriamampandry M, Freund M, Wiesel ML, Rhinn S, Ravanat C, Cazenave JP, Leray C, and Gachet C. Diets enriched in (n-3) fatty acids affect rat coagulation factors dependent on vitamin K. Comptes rendus de l'Academie des sciences 321: 415-421, 1998.
3. Balas B, Belfort R, Harrison SA, Darland C, Finch J, Schenker S, Gastaldelli A, and Cusi K. Pioglitazone treatment increases whole body fat but not total body water in patients with non-alcoholic steatohepatitis. Journal of hepatology 47: 565-570, 2007.
4. Barcelli U, Glas-Greenwalt P, and Pollak VE. Enhancing effect of dietary supplementation with omega-3 fatty acids on plasma fibrinolysis in normal subjects. Thrombosis research 39: 307-312, 1985.
5. Bremer J. The biochemistry of hypo- and hyperlipidemic fatty acid derivatives: metabolism and metabolic effects. Progress in lipid research 40: 231-268, 2001.
6. Cariello NF, Romach EH, Colton HM, Ni H, Yoon L, Falls JG, Casey W, Creech D, Anderson SP, Benavides GR, Hoivik DJ, Brown R, and Miller RT. Gene expression profiling of the PPAR-alpha agonist ciprofibrate in the cynomolgus monkey liver. Toxicol Sci 88: 250-264, 2005.
7. Chahil TJ, and Ginsberg HN. Diabetic dyslipidemia. Endocrinology and metabolism clinics of North America 35: 491-510, vii-viii, 2006.
8. Chambrier C, Bastard JP, Rieusset J, Chevillotte E, Bonnefont-Rousselot D, Therond P, Hainque B, Riou JP, Laville M, and Vidal H. Eicosapentaenoic acid induces mRNA expression of peroxisome proliferator-activated receptor gamma. Obesity research 10: 518-525, 2002.
9. Christodoulides C, and Vidal-Puig A. PPARs and adipocyte function. Molecular and cellular endocrinology 318: 61-68, 2010.
10. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, and Meng F. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic acids research 33: e175, 2005.
11. Dallongeville J, Bauge E, Tailleux A, Peters JM, Gonzalez FJ, Fruchart JC, and Staels B. Peroxisome proliferator-activated receptor alpha is not rate-limiting for the lipoprotein-lowering

action of fish oil. The Journal of biological chemistry 276: 4634-4639, 2001.
12. Deckelbaum RJ, Worgall TS, and Seo T. n-3 fatty acids and gene expression. The American journal of clinical nutrition 83: 1520S-1525S, 2006.
13. Duval C, Muller M, and Kersten S. PPARalpha and dyslipidemia. Biochimica et biophysica acta 1771: 961-971, 2007.
14. Eckel RH, Grundy SM, and Zimmet PZ. The metabolic syndrome. Lancet 365: 1415-1428, 2005.
15. Farnier M. Update on the clinical utility of fenofibrate in mixed dyslipidemias: mechanisms of action and rational prescribing. Vasc Health Risk Manag 4: 991-1000, 2008.
16. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, and Zhang J. Bioconductor: open software development for computational biology and bioinformatics. Genome biology 5: R80, 2004.
17. Gonzalez-Periz A, Horrillo R, Ferre N, Gronert K, Dong B, Moran-Salvador E, Titos E, Martinez-Clemente M, Lopez-Parra M, Arroyo V, and Claria J. Obesity-induced insulin resistance and hepatic steatosis are alleviated by omega-3 fatty acids: a role for resolvins and protectins. Faseb J 23: 1946-1957, 2009.
18. Harris WS, and Bulchandani D. Why do omega-3 fatty acids lower serum triglycerides? Current opinion in lipidology 17: 387-393, 2006.
19. Harris WS, Miller M, Tighe AP, Davidson MH, and Schaefer EJ. Omega-3 fatty acids and coronary heart disease risk: clinical and mechanistic perspectives. Atherosclerosis 197: 12-24, 2008.
20. Hayashi K, Ohtani H, Kurushima H, Nomura S, Koide K, Kunita T, Tanaka K, Yasunobu Y, Kambe M, and Kajiyama G. Decreases in plasma lipid content and thrombotic activity by ethyl eicosapentate purified from fish oils. Curr Ther Res 56: 24–31, 1995.
21. Horton JD, Shah NA, Warrington JA, Anderson NN, Park SW, Brown MS, and Goldstein JL. Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. Proceedings of the National Academy of Sciences of the United States of America 100: 12027-12032, 2003.
22. Ide T. Interaction of fish oil and conjugated linoleic acid in affecting hepatic activity of lipogenic enzymes and gene expression in liver and adipose tissue. Diabetes 54: 412-423, 2005.
23. Itoh M, Suganami T, Satoh N, Tanimoto-Koyama K, Yuan X, Tanaka M, Kawano H, Yano T, Aoe S, Takeya M, Shimatsu A, Kuzuya H, Kamei Y, and Ogawa Y. Increased adiponectin secretion by highly purified eicosapentaenoic acid in rodent models of obesity and human obese subjects. Arteriosclerosis, thrombosis, and vascular biology 27: 1918-1925, 2007.
24. Jump DB, Botolin D, Wang Y, Xu J, Demeure O, and Christian B. Docosahexaenoic acid (DHA) and hepatic gene transcription. Chemistry and physics of lipids 153: 3-13, 2008.
25. Kim HJ, Takahashi M, and Ezaki O. Fish oil feeding decreases mature sterol regulatory element-binding protein 1 (SREBP-1) by down-regulation of SREBP-1c mRNA in mouse liver. A possible mechanism for down-regulation of lipogenic enzyme mRNAs. The Journal of biological chemistry 274: 25892-25898, 1999.
26. Kim JB, and Spiegelman BM. ADD1/SREBP1 promotes adipocyte differentiation and gene expression linked to fatty acid metabolism. Genes & development 10: 1096-1107, 1996.
27. Knight BL, Hebbachi A, Hauton D, Brown AM, Wiggins D, Patel DD, and Gibbons GF. A role for PPARalpha in the control of SREBP activity and lipid synthesis in the liver. The Biochemical journal 389: 413-421, 2005.
28. Kockx M, Gervois PP, Poulain P, Derudas B, Peters JM, Gonzalez FJ, Princen HM, Kooistra T, and Staels B. Fibrates suppress fibrinogen gene expression in rodents via activation of the peroxisome proliferator-activated receptor-alpha. Blood 93: 2991-2998, 1999.
29. Lalloyer F, and Staels B. Fibrates, glitazones, and peroxisome proliferator-activated receptors. Arteriosclerosis, thrombosis, and vascular biology 30: 894-899, 2010.

30. Lee JN, Kim H, Yao H, Chen Y, Weng K, and Ye J. Identification of Ubxd8 protein as a sensor for unsaturated fatty acids and regulator of triglyceride synthesis. Proceedings of the National Academy of Sciences of the United States of America 107: 21424-21429, 2010.
31. Leonard JF, Courcol M, Mariet C, Charbonnier A, Boitier E, Duchesne M, Parker F, Genet B, Supatto F, Roberts R, and Gautier JC. Proteomic characterization of the effects of clofibrate on protein expression in rat liver. Proteomics 6: 1915-1933, 2006.
32. Leray C, Wiesel ML, Freund M, Cazenave JP, and Gachet C. Long-chain n-3 fatty acids specifically affect rat coagulation factors dependent on vitamin K: relation to peroxidative stress. Arteriosclerosis, thrombosis, and vascular biology 21: 459-465, 2001.
33. Lie J, Lankhuizen IM, Gross B, van Gent T, van Haperen R, Scheek L, Staels B, de Crom R, and van Tol A. Fenofibrate reverses the decline in HDL cholesterol in mice overexpressing human phospholipid transfer protein. Biochimica et biophysica acta 1738: 48-53, 2005.
34. Lox CD. The effects of dietary marine fish oils (omega-3 fatty acids) on coagulation profiles in men. General pharmacology 21: 241-246, 1990.
35. Mardones P, Pilon A, Bouly M, Duran D, Nishimoto T, Arai H, Kozarsky KF, Altayo M, Miquel JF, Luc G, Clavey V, Staels B, and Rigotti A. Fibrates down-regulate hepatic scavenger receptor class B type I protein expression in mice. The Journal of biological chemistry 278: 7884-7890, 2003.
36. Martin PG, Guillou H, Lasserre F, Dejean S, Lan A, Pascussi JM, Sancristobal M, Legrand P, Besse P, and Pineau T. Novel aspects of PPARalpha-mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. Hepatology (Baltimore, Md 45: 767-777, 2007.
37. Medvedev AV, Robidoux J, Bai X, Cao W, Floering LM, Daniel KW, and Collins S. Regulation of the uncoupling protein-2 gene in INS-1 beta-cells by oleic acid. The Journal of biological chemistry 277: 42639-42644, 2002.
38. Mozaffarian D. Fish and n-3 fatty acids for the prevention of fatal coronary heart disease and sudden cardiac death. The American journal of clinical nutrition 87: 1991S-1996S, 2008.
39. Neschen S, Morino K, Rossbacher JC, Pongratz RL, Cline GW, Sono S, Gillum M, and Shulman GI. Fish oil regulates adiponectin secretion by a peroxisome proliferator-activated receptor-gamma-dependent mechanism in mice. Diabetes 55: 924-928, 2006.
40. Nieuwenhuys CM, Beguin S, Offermans RF, Emeis JJ, Hornstra G, and Heemskerk JW. Hypocoagulant and lipid-lowering effects of dietary n-3 polyunsaturated fatty acids with unchanged platelet activation in rats. Arteriosclerosis, thrombosis, and vascular biology 18: 1480-1489, 1998.
41. Nieuwenhuys CM, Feijge MA, Vermeer C, Hennissen AH, Beguin S, and Heemskerk JW. Vitamin K-dependent and vitamin K-independent hypocoagulant effects of dietary fish oil in rats. Thrombosis research 104: 137-147, 2001.
42. Ohta T, Masutomi N, Tsutsui N, Sakairi T, Mitchell M, Milburn MV, Ryals JA, Beebe KD, and Guo L. Untargeted metabolomic profiling as an evaluative tool of fenofibrate-induced toxicology in Fischer 344 male rats. Toxicologic pathology 37: 521-535, 2009.
43. Okazaki H, Goldstein JL, Brown MS, and Liang G. LXR-SREBP-1c-phospholipid transfer protein axis controls very low density lipoprotein (VLDL) particle size. The Journal of biological chemistry 285: 6801-6810.
44. Oosterveer MH, Grefhorst A, van Dijk TH, Havinga R, Staels B, Kuipers F, Groen AK, and Reijngoud DJ. Fenofibrate simultaneously induces hepatic fatty acid oxidation, synthesis, and elongation in mice. The Journal of biological chemistry 284: 34036-34044, 2009.
45. Oosterveer MH, van Dijk TH, Tietge UJ, Boer T, Havinga R, Stellaard F, Groen AK, Kuipers F, and Reijngoud DJ. High fat feeding induces hepatic fatty acid elongation in mice. PloS one 4: e6066, 2009.
46. Oosthuizen W, Vorster HH, Jerling JC, Barnard HC, Smuts CM, Silvis N, Kruger A, and Venter CS. Both fish oil and olive oil lowered plasma fibrinogen in women with high baseline fibrinogen levels. Thrombosis and haemostasis 72: 557-562, 1994.

47. Pan M, Cederbaum AI, Zhang YL, Ginsberg HN, Williams KJ, and Fisher EA. Lipid peroxidation and oxidant stress regulate hepatic apolipoprotein B degradation and VLDL production. The Journal of clinical investigation 113: 1277-1287, 2004.
48. Raclot T, Groscolas R, Langin D, and Ferre P. Site-specific regulation of gene expression by n-3 polyunsaturated fatty acids in rat white adipose tissues. Journal of lipid research 38: 1963-1972, 1997.
49. Rakhshandehroo M, Hooiveld G, Muller M, and Kersten S. Comparative analysis of gene regulation by the transcription factor PPARalpha between mouse and human. PloS one 4: e6796, 2009.
50. Rakhshandehroo M, Knoch B, Muller M, and Kersten S. Peroxisome proliferator-activated receptor alpha target genes. PPAR research 2010: 2010.
51. Sanderson LM, de Groot PJ, Hooiveld GJ, Koppen A, Kalkhoven E, Muller M, and Kersten S. Effect of synthetic dietary triglycerides: a novel research paradigm for nutrigenomics. PloS one 3: e1681, 2008.
52. Sarwar N, Sandhu MS, Ricketts SL, Butterworth AS, Di Angelantonio E, Boekholdt SM, Ouwehand W, Watkins H, Samani NJ, Saleheen D, Lawlor D, Reilly MP, Hingorani AD, Talmud PJ, and Danesh J. Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. Lancet 375: 1634-1639, 2010.
53. Satapati S, He T, Inagaki T, Potthoff M, Merritt ME, Esser V, Mangelsdorf DJ, Kliewer SA, Browning JD, and Burgess SC. Partial resistance to peroxisome proliferator-activated receptor-alpha agonists in ZDF rats is associated with defective hepatic mitochondrial metabolism. Diabetes 57: 2012-2021, 2008.
54. Schmidt EB, Varming K, Ernst E, Madsen P, and Dyerberg J. Dose-response studies on the effect of n-3 polyunsaturated fatty acids on lipids and haemostasis. Thrombosis and haemostasis 63: 1-5, 1990.
55. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology 3: Article3, 2004.
56. Staels B, Dallongeville J, Auwerx J, Schoonjans K, Leitersdorf E, and Fruchart JC. Mechanism of action of fibrates on lipid and lipoprotein metabolism. Circulation 98: 2088-2093, 1998.
57. Staels B, van Tol A, Andreu T, and Auwerx J. Fibrates influence the expression of genes involved in lipoprotein metabolism in a tissue-selective manner in the rat. Arterioscler Thromb 12: 286-294, 1992.
58. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102: 15545-15550, 2005.
59. Sugiyama E, Ishikawa Y, Li Y, Kagai T, Nobayashi M, Tanaka N, Kamijo Y, Yokoyama S, Hara A, and Aoyama T. Eicosapentaenoic acid lowers plasma and liver cholesterol levels in the presence of peroxisome proliferators-activated receptor alpha. Life sciences 83: 19-28, 2008.
60. Takahashi M, Tsuboyama-Kasaoka N, Nakatani T, Ishii M, Tsutsumi S, Aburatani H, and Ezaki O. Fish oil feeding alters liver gene expressions to defend against PPARalpha activation and ROS production. American journal of physiology 282: G338-348, 2002.
61. Tontonoz P, and Spiegelman BM. Fat and beyond: the diverse biology of PPARgamma. Annual review of biochemistry 77: 289-312, 2008.
62. Vanschoonbeek K, Feijge MA, Paquay M, Rosing J, Saris W, Kluft C, Giesen PL, de Maat MP, and Heemskerk JW. Variable hypocoagulant effect of fish oil intake in humans: modulation of fibrinogen level and thrombin generation. Arteriosclerosis, thrombosis, and vascular biology 24: 1734-1740, 2004.

63. Vanschoonbeek K, Wouters K, van der Meijden PE, van Gorp PJ, Feijge MA, Herfs M, Schurgers LJ, Hofker MH, de Maat MP, and Heemskerk JW. Anticoagulant effect of dietary fish oil in hyperlipidemia: a study of hepatic gene expression in APOE2 knock-in mice. Arteriosclerosis, thrombosis, and vascular biology 28: 2023-2029, 2008.
64. Wang R, Kong X, Cui A, Liu X, Xiang R, Yang Y, Guan Y, Fang F, and Chang Y. Sterol-regulatory-element-binding protein 1c mediates the effect of insulin on the expression of Cidea in mouse hepatocytes. The Biochemical journal 430: 245-254, 2010.
65. Watts GF, and Dimmitt SB. Fibrates, dyslipoproteinaemia and cardiovascular disease. Current opinion in lipidology 10: 561-574, 1999.
66. Watts GF, Ji J, Chan DC, Ooi EM, Johnson AG, Rye KA, and Barrett PH. Relationships between changes in plasma lipid transfer proteins and apolipoprotein B-100 kinetics during fenofibrate treatment in the metabolic syndrome. Clin Sci (Lond) 111: 193-199, 2006.
67. Weinhofer I, Kunze M, Rampler H, Bookout AL, Forss-Petter S, and Berger J. Liver X receptor alpha interferes with SREBP1c-mediated Abcd2 expression. Novel cross-talk in gene regulation. The Journal of biological chemistry 280: 41243-41251, 2005.
68. Wopereis S, Rubingh CM, van Erk MJ, Verheij ER, van Vliet T, Cnubben NH, Smilde AK, van der Greef J, van Ommen B, and Hendriks HF. Metabolic profiling of the response to an oral glucose tolerance test detects subtle metabolic changes. PloS one 4: e4525, 2009.
69. Wu Z, Irizarry R, Gentleman R, Martinez-Murillo F, and Spencer F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association 99: 909–917, 2004.
70. Zambon A, Gervois P, Pauletto P, Fruchart JC, and Staels B. Modulation of hepatic inflammatory risk markers of cardiovascular diseases by PPAR-alpha activators: clinical and experimental evidence. Arteriosclerosis, thrombosis, and vascular biology 26: 977-986, 2006.

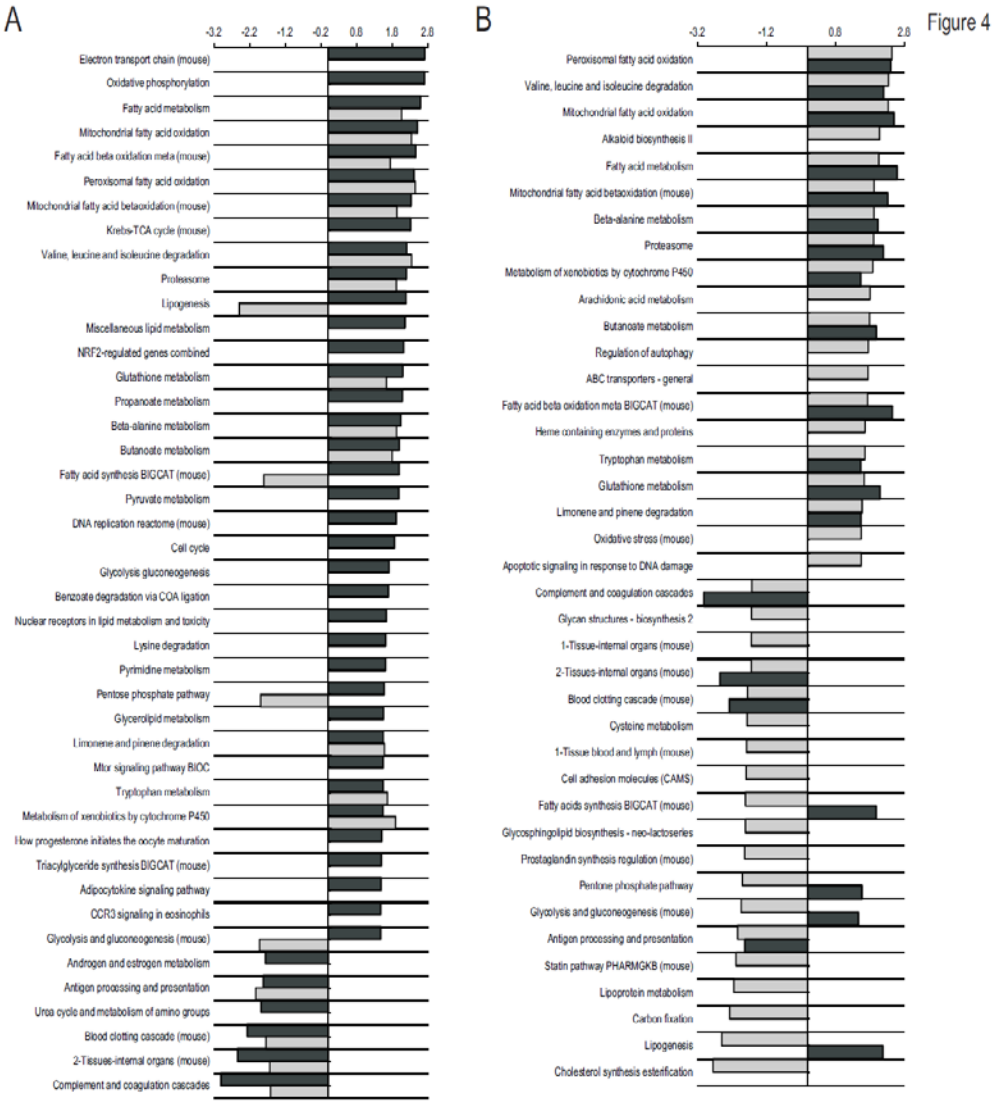
A
-3.2 -2.2 -1.2 -0.2 0.8 1.8 2.8
Electron transport chain (mouse)
Oxidative phosphorylation
Fatty acid metabolism
Mitochondrial fatty acid oxidation
Fatty acid beta oxidation meta (mouse)
Peroxisomal fatty acid oxidation
Mitochondrial fatty acid betaoxidation (mouse)
Krebs-TCA cycle (mouse)
Valine, leucine and isoleucine degradation
Proteasome
Lipogenesis
Miscellaneous lipid metabolism
NRF2-regulated genes combined
Glutathione metabolism
Propanoate metabolism
Beta-alanine metabolism
Butanoate metabolism
Fatty acid synthesis BIOCAT (mouse)
Pyruvate metabolism
DNA replication reactome (mouse)
Cell cycle
Glycolysis gluconeogenesis
Benzoate degradation via COA ligation
Nuclear receptors in lipid metabolism and toxicity
Lysine degradation
Pyrimidine metabolism
Pentose phosphate pathway
Glycerolipid metabolism
Limonene and pinene degradation
Mtor signaling pathway BIOC
Tryptophan metabolism
Metabolism of xenobiotics by cytochrome P450
How progesterone initiates the oocyte maturation
Triacylglyceride synthesis BIOCAT (mouse)
Adipocytokine signaling pathway
CCR3 signaling in eosinophils
Glycolysis and gluconeogenesis (mouse)
Androgen and estrogen metabolism
Antigen processing and presentation
Urea cycle and metabolism of amino groups
Blood clotting cascade (mouse)
2-Tissues-internal organs (mouse)
Complement and coagulation cascades
B
-3.2 -1.2 0.8 2.8
Peroxisomal fatty acid oxidation
Valine, leucine and isoleucine degradation
Mitochondrial fatty acid oxidation
Alkaloid biosynthesis II
Fatty acid metabolism
Mitochondrial fatty acid betaoxidation (mouse)
Beta-alanine metabolism
Proteasome
Metabolism of xenobiotics by cytochrome P450
Arachidonic acid metabolism
Butanoate metabolism
Regulation of autophagy
ABC transporters - general
Fatty acid beta oxidation meta BIOCAT (mouse)
Heme containing enzymes and proteins
Tryptophan metabolism
Glutathione metabolism
Limonene and pinene degradation
Oxidative stress (mouse)
Apoptotic signaling in response to DNA damage
Complement and coagulation cascades
Glycan structures - biosynthesis 2
1-Tissue-internal organs (mouse)
2-Tissues-internal organs (mouse)
Blood clotting cascade (mouse)
Cysteine metabolism
1-Tissue blood and lymph (mouse)
Cell adhesion molecules (CAMS)
Fatty acids synthesis BIOCAT (mouse)
Glycosphingolipid biosynthesis - neo-lactoseries
Prostaglandin synthesis regulation (mouse)
Pentose phosphate pathway
Glycolysis and gluconeogenesis (mouse)
Antigen processing and presentation
Statin pathway PHARMGKB (mouse)
Lipoprotein metabolism
Carbon fixation
Lipogenesis
Cholesterol synthesis esterification

Figure 4

Figure 4. Overrepresented pathways in the liver after fenofibrate and fish oil treatments identified by Gene Set Enrichment Analysis. (A) Ranking based on the normalized enrichment score (NES) of pathways regulated by fenofibrate (black bars), with pathways regulated by fish oil shown in parallel (grey bars). Pathways with FDR q-value < 0.2 are shown. The NES reflects the degree to which a gene set in certain pathway is overrepresented at the top (upregulated) or bottom (downregulated) of the ranked gene list and is corrected for gene set size. Sources of the gene sets consist of BioCarta, GenMAPP, KEGG, Sigma-Aldrich pathway, and Signal Transduction Knowledge Environment. (B) Ranking based on the normalized enrichment score of pathways regulated by fish oil (grey bars), with pathways regulated by fenofibrate shown in parallel (grey bars). For additional information, see supplementary note.

Comparative transcriptomics and metabolomics of fenofibrate and fish oil

Figure 5. N**on-overlapping genes between fenofibrate and fish oil treatment.** Venn diagrams show the overlap in significantly upregulated (A) and downregulated (B) genes after treatment with fenofibrate or fish oil for two weeks. Genes were included if mean fold-change (MFC) by fenofibrate or fish oil exceeded the value of 1.2 and Bayesian corrected q was lower than 0.05. (C) Top 40 genes most strongly upregulated (based on MFC) by fenofibrate but not fish oil. Known PPARα target genes upregulated by fish oil by >1.2 fold yet for which the effect failed to meet statistical significance due to large variation were removed in order to exclusively highlight fenofibrate-specific genes (Mogat1, Acot3, Cd36, Cpt1b, Slc16a13, Acot4, Crat and Slc27a4). (D) Top 40 genes most strongly upregulated by fish oil but not fenofibrate. (E) Top 40 genes most strongly downregulated by fenofibrate but not fish oil. (F) Top 40 genes most strongly downregulated by fish oil but not fenofibrate. Expression levels in the control mice (Con) were set at 1 (black) and expression levels in two treated groups were calculated relative to the control group. The list was sorted based on fold-change of fenofibrate (Fen: C and E) or fish oil (Fis: D and F) treated mice.

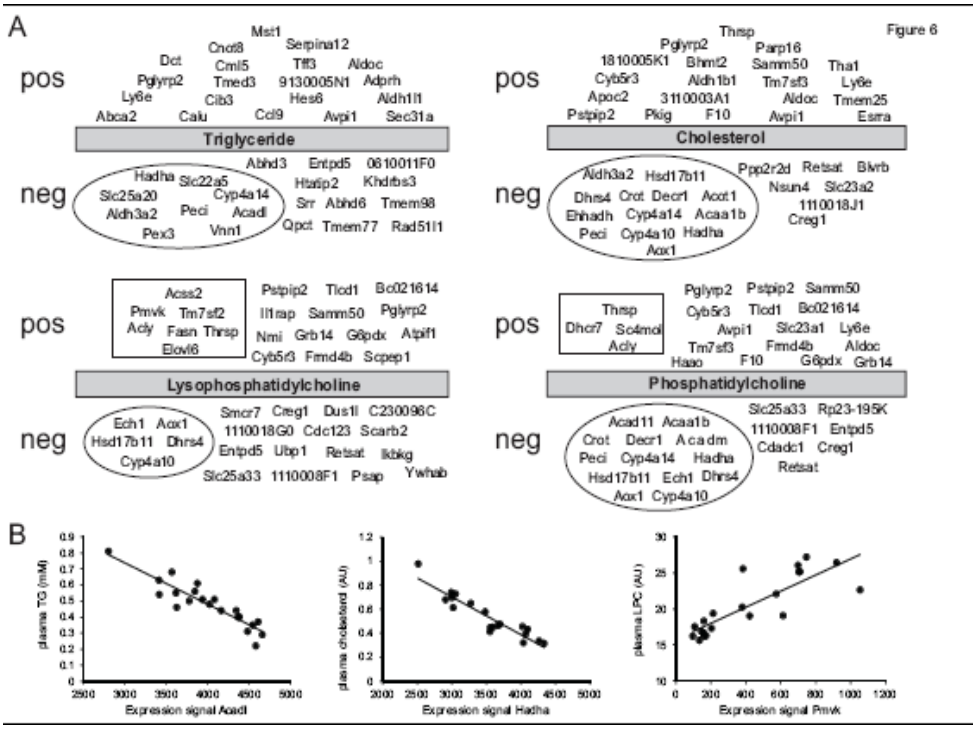

Figure 6. Correlation analysis between plasma metabolites and hepatic gene expression signals in control and fish oil treated mice. Spearman rank correlation analyses were conducted between plasma metabolites and hepatic gene expression signals for those genes that were statistically significant regulated after fish oil treatment ($q < 0.05$). (A) The respective top 20 positive (pos, $r > 0.77$) and negative (neg, $r < -0.81$) correlated genes with plasma triglyceride (TG), cholesterol, lysophosphatidylcholine (LPC), and phosphatidylcholine levels. The genes are either PPARα regulated and involved in fatty acid metabolism (neg side, in circle), or SREBP regulated and involved in lipogenesis and cholesterol biosynthesis (pos side, in square). (B) Scatterplots between plasma triglyceride (TG), cholesterol and lysophosphatidylcholine (LPC) levels and hepatic expression level of top correlated genes. AU: arbitrary units.

Chapter 7

Research article

Endogenous desaturase activity and risk of coronary heart disease in the CAREMA cohort study

Yingchang Lu, Anika Vaarhorst, Audrey H.H. Merry, Martijn E.T. Dollé, Robert Hovenier, Sandra Imholz, Leo J. Schouten, Bastiaan T. Heijmans, Michael Müller, P. Eline Slagboom, Piet A. van den Brandt, Anton P.M. Gorgels, Jolanda M.A. Boer, and Edith J.M. Feskens

Yingchang Lu, Robert Hovenier, Michael Müller, Edith J.M. Feskens
Division of Human Nutrition, Wageningen University and Research Center, Wageningen, The Netherlands
Anika Vaarhorst, Bastiaan T. Heijmans, P. Eline Slagboom
Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
Audrey H.H. Merry, Piet A. van den Brandt
Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands
Yingchang Lu, Martijn E.T. Dollé, Sandra Imholz, Jolanda M.A. Boer
National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
Leo J. Schouten, Piet A. van den Brandt
Department of Epidemiology, GROW School of Oncology and Developmental Biology, Maastricht University,
Maastricht, The Netherlands
Anton P.M. Gorgels
Department of Cardiology, University Hospital Maastricht,
Maastricht, The Netherlands

(Submitted)

ABSTRACT:

BACKGROUND: The long-term role of fatty acid desaturases in the etiology of coronary heart disease (CHD) remains unexplained. C20:4n-6/C20:3n-6 and C18:3n-6/C18:2n-6 ratios are markers of endogenous desaturase activity, but have never been studied in relation to incident CHD. Therefore, the aim of this study was to investigate the relation between these ratios as well as genotypes of FADS1 rs174547 and CHD incidence.

METHODS: We applied a case-cohort design within the CAREMA cohort, a large prospective study among the general Dutch population followed up for a median of 12.1 years. Fatty acid profile in plasma cholesteryl esters at baseline was measured in a random subcohort (n = 1323) and incident CHD cases (n = 537). Main outcome measures were hazard ratios (HRs) of incident CHD adjusted for major CHD risk factors.

RESULTS: FADS1 genotypes were associated with plasma polyunsaturated fatty acid (PUFA) levels and δ-5 and δ-6 desaturase' activity, but not with CHD risk. In multivariable adjusted models, high baseline δ-5 desaturase activity was associated with reduced CHD risk (P for trend = 0.02), especially among those carrying the high desaturase activity genotype (AA): HR (95% CI) = 0.35 (0.15-0.81) for comparing the extreme quintiles. High plasma C22:6n-3 (DHA) levels were also associated with reduced CHD risk.

CONCLUSIONS: In this prospective cohort study, we showed evidence for a reduced CHD risk with increased δ-5 desaturase activity. This effect may be partly mediated by increased DHA levels. Furthermore, we confirmed effects of rs174547 in the FADS1 gene on PUFA levels and markers of δ-5 desaturase activities.

INTRODUCTION

Polyunsaturated fatty acids (PUFAs) are generally believed to reduce coronary heart disease (CHD) risk.[1-4] Intakes of n-3 PUFAs, especially eicosapentaenoic acid (EPA, C20:5n-3) and docosahexaenoic acid (DHA, C22:6n-3) present in fish oil, are confirmed to prevent fatal CHD and sudden cardiac death in both observational studies and large-scale randomized controlled trials (RCTs).[1,3] However, direct evidence for the preventive effect of n-3 PUFAs on non-fatal CHD was only recently found in some, but not all, large-scale RCTs.[5-7] The replacement of saturated fatty acids by n-6 PUFAs protected against incident CHD based on a recent meta-analysis including 8 RCTs.[8] As some of these RCTs also included n-3 PUFAs in addition to n-6 PUFAs,[2,8] the effects specific to n-6 PUFAs, however, remain unclear.

PUFA profile of various biological tissues is often used as a biomarker of dietary PUFA intake. Adipose tissue reflects the intake of past months to years, while erythrocyte membranes, and plasma or serum phospholipids or cholesteryl esters reflect the intake of several weeks.[9,10] However, the PUFA profile in biological tissues does not only reflect dietary intake, but is also strongly dependent on the endogenous metabolism of PUFAs.[10,11] Therefore, PUFA biomarkers in biological tissues mirror the internal PUFA exposure that may be biologically more relevant. Several PUFAs can be endogenously synthesized by a series of alternate desaturation and elongation processes.[11,12] The δ-5 desaturase and δ-6 desaturase are rate-limiting enzymes for synthesizing long-chain n-3 and n-6 PUFAs.[11,13-15] They are encoded by the *FADS1* and *FADS2* genes on chromosome 11 (11q12-13.1), respectively.[11,16] The potential functional genetic variants in these genes have been identified,[17] and confirmed in recent genome-wide association studies.[18,19] They have an impact on mRNA abundance of *FADS1*,[20-24] and, as a result, on desaturase activity, plasma PUFA levels, and endogenous PUFA pools.[17-19,24-27] Since it is difficult to directly assay the enzyme activities of δ-5 and δ-6 desaturase,[11,13,14] especially in large-scale epidemiological studies, their activities have been traditionally estimated by using PUFA product-to-precursor ratios.[25,26]

Although few prospective cohort studies have investigated PUFA biomarkers in relation to the incidence of CHD,[28] the relation with desaturase activities estimated from PUFA product-to-precursor ratios has, to the best of our knowledge, never been evaluated. In this prospective cohort study, we therefore aim to investigate whether δ-5 and δ-6 desaturase activity, that affect our life-time exposure to PUFAs, influence CHD risk.

METHODS

STUDY POPULATION

We conducted a case-cohort study within the Monitoring Project on Cardiovascular Disease Risk Factors 1987-1991,[29] one of the two monitoring studies that were included in the Cardiovascular Registry Maastricht (CAREMA) study. The CAREMA study was described in detail before.[30] In total, 12,486 men and women, born between 1927 and 1967 and living in the Maastricht area, participated in the Monitoring Project on Cardiovascular Disease Risk Factors and had given informed consent to retrieve information from the municipal population registries and from the general practitioner and specialist.

CARDIOLOGICAL FOLLOW-UP

The cardiologic follow-up has been described before.[30] In brief, 97.6% of the CAREMA members could be found by linking the cohort to the hospital information system of University Hospital Maastricht (UHM). They were linked to the cardiology information system of the Department of Cardiology to obtain information about the occurrence of myocardial infarction (MI), unstable angina pectoris (UAP), coronary artery bypass grafting (CABG), or percutanous transluminal coronary angioplasty surgery (PTCA). For participants who died, the cause of death was obtained from Statistics Netherlands. In addition, the CAREMA cohort was linked to the hospital discharge registry of the UHM to increase the completeness of the cardiologic follow-up. Follow-up ended on 31 December 2003 with a median follow-up of 12.1 yrs (range: 0.0-16.9 yrs).

SUBCOHORT AND INCIDENT CHD SELECTION FOR CASE-COHORT DESIGN

For the present study, participants who were younger than 30 years at baseline (n = 2204), had a history of MI, UAP, CABG, or PTCA before baseline (n = 118), or were lost to follow-up (n = 2) were excluded. Thus, the eligible cohort consisted of 10,164 participants. All 620 participants who developed incident CHD during follow-up (315 MIs, 244 UAPs and 61 CHD deaths) were included in the case-cohort study. From the eligible cohort, 1483 participants were randomly drawn as a subcohort.[31] By randomly selecting a subcohort and using the appropriate statistics for this type of research design, the results are expected to be generalizable to the entire cohort without the need of biomarker measurements in the entire cohort.[31-33]

RISK FACTOR DETERMINATION

At baseline, all participants filled in a questionnaire on life-style characteristics, medical history, and parental history of MI. During a medical examination, information was collected on blood pressure, height, and weight. In addition, non-fasting blood samples were collected using EDTA tubes. The blood was centrifuged for 10 minutes at 1000 rpm and fractioned into blood plasma, white blood cells and erythrocytes and subsequently stored at -20°C. Within three weeks, the plasma samples were transported to the Lipid Reference Laboratory of the University Hospital Dijkzigt (LRL) in Rotterdam where the total and HDL-cholesterol levels were determined using a CHOD-PAP method.[34] The LRL in Rotterdam is a permanent member of the International Cholesterol Reference Method Laboratory Network.

FATTY ACID DETERMINATION

Fatty acids from plasma cholesteryl esters were quantified by gas-liquid chromatography between 2010 and 2011 at the Department of Human Nutrition of Wageningen University. The case and non-case samples were randomly distributed over the batches. Peak retention times and area percentages of total fatty acids were identified by using known cholesteryl ester standards (mixture of FAME components from Sigma (MO) and NuChek (MN)) and analyzed with the Agilent Technologies ChemStation software (Agilent, Amstelveen, The Netherlands). For certain fatty acids, the values were too low to be reliably detected in some subjects, and “0” was assigned to their values. Interassay coefficients of variance in fatty acids in plasma cholesteryl esters were 1.68% for C16:0, 1.01% for C18:2n-6, 1.88%

for C20:4n-6, and 5.02% for C22:6n-3, respectively. Desaturase activities were estimated by calculating fatty acid product-to-precursor ratios, C20:4n-6/C20:3n-6 to reflect δ-5 desaturase activity, and C18:3n-6/C18:2n-6 to reflect δ-6 desaturase activity. The subjects with "0" value for C20:3n-6 and C18:2n-6 were not included in the δ-5 and δ-6 desaturase activity analyses, respectively. Information on plasma fatty acids was available on 1323 subcohort members and 537 CHD cases.

DNA EXTRACTION AND GENOTYPING

DNA was extracted from the white blood cell fraction (buffy coats), using a standard procedure.[35] The resulting DNA pellet was dissolved in TE buffer and DNA concentrations were determined using the Nanodrop ND1000 Spectrophotometer. The single nucleotide polymorphism (SNP) of rs174547 was selected based on its association with blood cholesterol and triglyceride levels in a genome-wide association study.[21] This SNP is in high linkage disequilibrium with several other SNPs around *FADS1* and *FADS2* gene region, which have an impact on mRNA abundance of *FADS1*,[20-23] desaturase activity, plasma PUFA levels, and endogenous PUFA pools.[17-19,24-27] Rs174547 was genotyped using the iPLEX Gold chemistry of Sequenom's MassARRAY platform (San Diego, CA) at the Leiden University Medical Center. Sequenom's MassARRAY® Assay Design 3.1 Software was used for SNP assay design, and Sequenom's SpectroTyper 4.0 software was used to call genotypes automatically, followed by manual review. The total genotyping success rate was 93%. Among the subjects who were measured for plasma fatty acid levels, information on rs174547 genotype was available for 1246 subcohort members and 492 CHD cases. The genotype distribution was consistent with Hardy-Weinberg equilibrium expectations.

STATISTICAL ANALYSIS

Generalized linear models adjusted for age and sex were used to study the relations between rs174547 genotypes and PUFAs, δ-5 desaturase activity, and δ-6 desaturase activity. Cox proportional hazards models adapted for the case-cohort design according to the Prentice's method[32] were used to calculate hazard ratios as measures for relative risk.[33] We estimated hazard ratios for quintiles of fatty acids (expressed as the percentage of total fatty acids present in the chromatogram) and desaturase activities based on subcohort distributions, and the respective lowest quintile was used as reference. The base models included age and sex. Additional models were further adjusted for current smoking, systolic blood pressure, hypertensive medication use, total and HDL cholesterol, history of diabetes and parental history of MI. The models were also further adjusted for the total percentage of n-3 PUFAs or n-6 PUFAs in plasma cholesteryl esters where necessary. Additional covariates studied were alcohol use and physical activity. The significance of a linear trend across quintiles was examined by including the exposure as a continuous variable in the model. Potential interactions between continuous desaturase activity and dichotomized rs174547 genotype (homozygous major allele carriers vs. minor allele carriers) were tested by including interaction terms into the model. Statistical significance was considered to be met with a P value < 0.05 and all testing was 2-sided. All statistical analyses were performed with SAS version 9.1 software (SAS Institute, Cary, NC).

RESULTS

The general characteristics of the study population by subcohort-case status are shown in Table 1. As expected, cases were older, more frequently male, had higher blood pressure and total cholesterol levels, lower HDL cholesterol levels, smoked more often, and more often reported to have diabetes and a parental history of MI.

Carrying the minor G allele of rs174547 was associated with higher levels of substrates for desaturases (C18:2n-6, C20:3n-6, and C18:3n-3) and lower levels of products from desaturases (C18:3n-6, C20:4n-6, C20:5n-3, and C22:6n-3) in the plasma cholesteryl esters. Consequently, lower δ-5 and δ-6 desaturase activity, assessed by the ratio of C20:4n-6 to C20:3n-6 and C18:3n-6 to C18:2n-6, were observed in carriers of the G allele as compared to those with the AA genotype (Table 2).

High baseline δ-5 desaturase activity was associated with reduced CHD risk (Table 3). A 30% reduction in CHD risk was observed among the participants in the second, third, fourth and fifth quintile of δ-5 desaturase activity compared with those in the first quintile after adjustment for age, sex, systolic blood pressure, hypertensive medication use, current smoking, diabetes, total cholesterol, and high-density lipoprotein cholesterol (*P* for trend = 0.02). Although the statistical interaction between rs174547 and δ-5 desaturase activity was not significant (*P* = 0.56), the protective effect of high δ-5 desaturase activity was mainly confined to subjects with the AA genotype (Table 4). In this group, the effect was stronger with a 65% risk reduction for the subjects in the fifth quintile compared with the first quintile (*P* for trend = 0.02). Rs174547 itself was not associated with CHD risk, the age- and sex-adjusted HR per G-allele being 0.99 (95% CI 0.84-1.16, Supplementary Table 1).

No association was observed between δ-6 desaturase activity and CHD risk (Table 5), also not after stratification by rs174547 genotype (data not shown).

The results for the four n-6 PUFAs that determine δ-5 and δ-6 desaturase activity are shown in Supplementary Table 2. No associations with CHD were observed for the C20 precursor (C20:3n-6) and product (C20:4n-6, arachidonic acid) of δ-5 desaturase, or for the C18 precursor (C18:2n-6, linoleic acid) and product (C18:3n-6) of δ-6 desaturase after adjustment for age, sex, systolic blood pressure, hypertensive medication use, current smoking, diabetes, total cholesterol, and high-density lipoprotein cholesterol (*P* for trend > 0.16).

Regarding the n-3 PUFAs affected by desaturases, a significant inverse association was observed for C22:6n-3 (DHA). This association became stronger after adjustment for plasma total and HDL cholesterol levels, and percentages of n-6 PUFA in plasma cholesteryl esters (*P* for trend = 0.027, Supplementary Table 3). The proportion of plasma C20:5n-3 (EPA) was not associated with incident CHD (*P* for trend = 0.724, Supplementary Table 3). No association was observed between C18:3n-3 (α-linolenic acid) and CHD risk (data not shown). To explore whether there is any independent effect of δ-5 desaturase activity on CHD beyond DHA, we additionally adjusted the models in Table 3 and 4 for percentages of DHA. The association between δ-5 desaturase activity and CHD

Table 1. Baseline characteristics of sub-cohort subjects and cases of incident coronary heart disease in the CAREMA cohort study.*

	Subcohort (n = 1323)†	Cases (n = 537)	Crude HR (95% CI)‡	Adjusted HR (95% CI)§
Age (y)	45.2 ± 8.5	49.7 ± 7.3	1.07 (1.06-1.09)	1.05 (1.04-1.07)
Male sex	608 (46.0%)	392 (73.0%)	3.34 (2.69-4.15)	2.22 (1.66-2.99)
Total cholesterol (mmol/L)	5.7 ± 1.1	6.4 ± 1.2	1.71 (1.56-1.87)	1.42 (1.26-1.60)
HDL cholesterol (mmol/L)	1.2 ± 0.3	1.0 ± 0.2	0.04 (0.03-0.06)	0.09 (0.05-0.16)
Systolic blood pressure (mmHg)	119.2 ± 14.9	128.0 ± 16.9	1.03 (1.02-1.04)	1.02 (1.01-1.03)
Hypertensive medication use	67 (5.1%)	58 (10.8%)	2.34 (1.63-3.35)	1.27 (0.79-2.05)
Diabetes mellitus	13 (1.0%)	20 (3.7%)	5.33 (2.74-10.36)	2.83 (1.39-5.78)
Current smoking	551 (41.8%)	304 (56.7%)	1.81 (1.49-2.21)	1.72 (1.33-2.22)
Parental history of MI	452 (34.3%)	228 (42.5%)	1.40 (1.14-1.71)	1.51 (1.16-1.95)

*Data are expressed as mean ± SD or n (%) unless otherwise indicated. HDL: high-density lipoprotein; MI: myocardial infarction; and HR (95% CI): hazard ratio and 95% confidence interval.
†Including 84 cases.
‡Hazard ratios were calculated per unit increase in total cholesterol, HDL cholesterol, and systolic blood pressure, and for the presence of the categorical traits.
§All variables were added into one multivariable Cox proportional hazards model.

Table 2. Association of rs174547 in *FADS1* with baseline PUFAs in plasma cholesteryl esters and desaturase activities in the sub-cohort (n = 1246).*

PUFA	*Rs174547*			*P* value†
	AA (545)	AG (569)	GG (132)	
n-6 PUFA				
C18:2n-6 (%)	44.30 ± 0.27†	44.88 ± 0.26	46.60 ± 0.54	7.48×10^{-4}
C18:3n-6 (%)	0.60 ± 0.009	0.48 ± 0.009	0.40 ± 0.019	6.87×10^{-28}
C20:3n-6 (%)	0.42 ± 0.005	0.43 ± 0.005	0.44 ± 0.010	0.051
C20:4n-6 (%)	4.29 ± 0.05	3.56 ± 0.05	2.89 ± 0.09	3.92×10^{-46}
n-3 PUFA				
C18:3n-3 (%)	0.40 ± 0.005	0.41 ± 0.005	0.45 ± 0.010	3.28×10^{-4}
C18:4n-3 (%)‡	0.18 ± 0.007	0.18 ± 0.007	0.17 ± 0.014	0.708
C20:5n-3 (%)	0.56 ± 0.01	0.46 ± 0.01	0.40 ± 0.03	8.71×10^{-8}
C22:6n-3 (%)	0.34 ± 0.006	0.31 ± 0.006	0.30 ± 0.013	0.005
δ-5§	10.65 ± 0.09	8.59 ± 0.09	6.86 ± 0.19	6.40×10^{-85}
δ-6§	0.014 ± 0.0002	0.011 ± 0.0002	0.009 ± 0.0005	2.51×10^{-27}

* 77 subjects in the subcohort had missing values for rs174547. PUFAs: polyunsaturated fatty acids.
† General linear models were used, and all values are mean ± SEM, adjusted for age and sex.
‡ Only few subjects were successfully measured (AA=161, AG=185, and GG=42).
§ δ-5 and δ-6 desaturase activities were assessed by the ratio of C20:4n-6 to C20:3n-6 and C18:3n-6 to C18:2n-6 in plasma cholesteryl esters, respectively.

Table 3. Association between baseline δ-5 desaturase activity and incident coronary heart disease (CHD).

	Quintile of δ-5 desaturase activity*					*P* value for trend†
	First (6.45)	Second (7.93)	Third (9.07)	Fourth (10.32)	Fifth (12.52)	
Incident CHD, n	155	117	94	93	67	
Model 1‡	1	0.70 (0.51-0.97)	0.60 (0.42-0.83)	0.60 (0.43-0.83)	0.49 (0.34-0.70)	<0.0001
Model 2§	1	0.75 (0.54-1.06)	0.66 (0.46-0.94)	0.57 (0.39-0.82)	0.51 (0.35-0.75)	<0.0001
Model 3¶	1	0.68 (0.47-0.98)	0.66 (0.45-0.96)	0.69 (0.46-1.01)	0.68 (0.45-1.02)	0.0249
Model 4$	1	0.71 (0.49-1.03)	0.70 (0.48-1.04)	0.74 (0.50-1.09)	0.77 (0.50-1.18)	0.1114

* δ-5 desaturase activity was assessed by the ratio of C20:4n-6 to C20:3n-6 in plasma cholesteryl esters, and median ratios in each quintile are listed between brackets.
† From models with desaturase activity included as a continuous variable.
‡ Model 1 is adjusted for age and sex.
§ Model 2 is adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, and diabetes.
¶ Model 3 is adjusted for all covariates in model 2, total cholesterol, and high-density lipoprotein cholesterol.
$ Model 4 is adjusted for all covariates in model 3 and percentages of C22:6n-3 (DHA) in plasma cholesteryl esters.

Table 4. Association between baseline δ-5 desaturase activity and incident coronary heart disease according to rs174547 genotypes.

rs174547 (# cases)		Quintile of δ-5 desaturase activity*				*P* value for trend†
	First	Second	Third	Fourth	Fifth	
AA (n = 205)	1‡	0.61 (0.27-1.41)	0.35 (0.15-0.79)	0.36 (0.16-0.80)	0.25 (0.11-0.54)	<0.0001
	1§	0.60 (0.25-1.43)	0.35 (0.15-0.83)	0.48 (0.21-1.11)	0.35 (0.15-0.81)	0.022
	1¶	0.63 (0.26-1.53)	0.40 (0.17-0.97)	0.55 (0.24-1.27)	0.44 (0.19-1.04)	0.087
AG/GG (n = 276)	1‡	0.69 (0.46-1.03)	0.62 (0.38-1.00)	0.54 (0.31-0.93)	1.15 (0.57-2.33)	0.027$
	1§	0.62 (0.39-0.98)	0.71 (0.42-1.21)	0.72 (0.40-1.30)	1.64 (0.76-3.53)	0.463
	1¶	0.63 (0.40-1.02)	0.75 (0.44-1.27)	0.78 (0.43-1.41)	1.86 (0.84-4.12)	0.649

* δ-5 desaturase activity was assessed by the ratio of C20:4n-6 to C20:3n-6 in plasma cholesteryl esters.
† From models with desaturase activity included as continuous variable.
‡ Model was adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, and diabetes.
§ Model was adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, diabetes, total cholesterol, and high-density lipoprotein cholesterol.
¶ Model was adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, diabetes, total cholesterol, high-density lipoprotein cholesterol, and percentages of C22:6n-3 (DHA) in plasma cholesteryl esters.
$ Hazard ratio per unit desaturase activity (95% confidence interval) = 0.89 (0.80-0.99).

Table 5. Association between baseline δ-6 desaturase activity and incident coronary heart disease (CHD).

	Quintile of δ-6 desaturase activity*					*P* value for trend†
	First (0.0055)	Second (0.0084)	Third (0.0104)	Fourth (0.0132)	Fifth (0.019)	
Incident CHD, n	92	99	93	122	131	
Model 1‡	1	0.99 (0.69-1.42)	0.87 (0.60-1.25)	1.09 (0.76-1.55)	1.03 (0.73-1.45)	0.606
Model 2§	1	1.03 (0.70-1.51)	0.89 (0.61-1.31)	1.07 (0.73-1.58)	0.93 (0.63-1.36)	0.627
Model 3¶	1	1.07 (0.71-1.63)	0.86 (0.55-1.33)	1.11 (0.73-1.69)	0.96 (0.63-1.47)	0.897

* δ-6 desaturase activity was assessed by ratio of C18:3n-6 to C18:2n-6 in plasma cholesteryl esters, and median ratios in each quintile are listed between brackets.
† From desaturase activity as continuous variable in the Cox proportional hazards model.
‡ Model 1 was adjusted for age and sex.
§ Model 2 was adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, and diabetes.
¶ Model 3 was adjusted for all covariates in model 2, total cholesterol, and high-density lipoprotein cholesterol.

risk attenuated, but there still existed a strong protective effect, especially among the AA carriers of rs174547 (HR: 95% CI = 0.44: 0.19-1.04 for comparing the extreme quintiles, Table 4).

Additional adjustment for parental history of MI, alcohol use or physical activity did not materially change any of the aforementioned associations (data not shown).

DISCUSSION

In this prospective cohort study, we observed an inverse association between δ-5 desaturase activity and incident CHD risk, but no association with δ-6 desaturase activity. This association was partly mediated by DHA. Furthermore we confirmed associations of rs174547 in the *FADS1* gene with plasma PUFA levels and δ-5 desaturase activity.[17-19,25,26] Consistent with the established cardiovascular protective effects of n-3 PUFAs,[1,3] and especially tissue DHA,[4,28] high DHA in plasma cholesteryl esters was associated with a reduced CHD risk. However, no association was observed between arachidonic acid or other n-6 PUFAs related to δ-5 or δ-6 desaturase activity in plasma cholesteryl esters and CHD risk.

Common genetic variants (including rs174547) in the *FADS* gene region have been associated with plasma lipid levels (total, LDL and HDL cholesterol, and triglycerides),[18,21,36] glycemic traits (fasting glucose and beta-cell function),[24] and resting heart rate[37] in recent genome-wide association studies. However, none of them have been associated with CHD risk directly.[36,38] This was also the case in our relatively large prospective study. In contrast, when using the estimated δ-5 desaturase activity based on the fatty acid proportion in plasma cholesteryl esters, we found a significant inverse association with incident CHD. This seems contradictory, as a strong association between rs174547 genotypes and estimated δ-5 desaturase activities was observed. However, the reduced risk was already observed with relatively low δ-5 desaturase activities (the second quintile) and remained constant over the following quintiles. Therefore, the majority of the participants with the GG genotype of rs174547 might have sufficient δ-5 desaturase activity to protect them from CHD. This might explain why no association between rs174547 genotypes and CHD risk was found. Both rs174547 genotypes and C20:4n-6/C20:3n-6 ratio reflect endogenous δ-5 desaturase activity, but from two different perspectives. The former can be regarded as the desaturase effect conferred by a single common genetic variant in the *FADS1* gene,[19,24-27] and the latter as an approximate estimation of full desaturase activity.[25,26] Their combination might provide the most accurate estimate of δ-5 desaturase activity. This might explain the stronger CHD risk reduction with high δ-5 desaturase activity in the subjects who inherited the AA genotype.

The exact biological mechanisms that link δ-5 desaturase activity with CHD risk are still not well understood. Arachidonic acid, EPA, and DHA are currently considered to be potentially involved directly in the pathogenesis of CHD through thrombotic, inflammatory, arrhythmic and/or lipid regulatory pathways.[1,3,11,12,39-41] δ-5 Desaturase is the key enzyme synthesizing these PUFAs, while δ-6 desaturase is important at the beginning of the n-3 and n-6 PUFA synthetic pathways.[13,14] Therefore, it is biologically plausible that

CHD risk could be influenced by δ-5 desaturase activity, but not by δ-6 desaturase activity[11,12] as was shown in our data. Increased δ-5 desaturase activity might contribute to the intracellular increase of EPA and especially arachidonic acid levels.[15] In non-fish eating populations, arachidonic acid is the predominant tissue very-long-chain PUFA, reaching 80% of the total.[28,39] Despite the potential pro-coagulant and pro-inflammatory effects of increased exposures to arachidonic acid and its derived eicosanoid metabolites,[2,12,39-43] there is no evidence of increased CHD risk with 5 -7 times habitual arachidonic acid intake based on short-term small-scale controlled feeding studies.[2,44-48] Tissue arachidonic acid levels are generally not associated with CHD risk.[28] This was supported by our finding based on the fatty acid profile in plasma cholesteryl esters, which suggests that arachidonic acic does not mediate the observed association between δ-5 desaturase activity and CHD risk.

Our results showed that the protective effect of increased δ-5 desaturase activity on CHD could be partly mediated by increased endogenous exposure to DHA. The fact that increased DHA levels associated with increased δ-5 desaturase activity protect against CHD is consistent with the established cardiovascular protective effect of increased n-3 PUFA exposure (EPA and/or DHA).[1,3] Accumulating evidence from observational studies suggests DHA might be more protective than EPA on CHD,[4,28] which is consistent with our findings. However, EPA and DHA are usually correlated with each other in tissues, and their potential effects cannot be easily discerned. In addition to blood triglyceride lowering and HDL cholesterol increasing effects of EPA and DHA, n-3 PUFAs have long been observed to have anti-thrombotic, anti-inflammatory, anti-arrhythmic, and blood pressure-lowering effects in humans even though the underlying mechanisms for these effects are incompletely understood.[1,3,11,12,41] Interestingly, the protective effects on fatal CHD and sudden cardiac death have been shown to level off with a modest intake of EPA and/or DHA (250mg/day), and little additional benefit was observed with higher intakes.[1] This is also consistent with our results for δ-5 desaturase activity. Nevertheless, there might be other unidentified pleiotropic cardiovascular protective effects of increased δ-5 desaturase activity. For example, these desaturases are also significantly expressed in immune cells[49] that play important roles in atherosclerotic CHD progression.

Our results should be interpreted in the context of several limitations. First, our analyses were based on a single baseline measurement of fatty acid levels that may not accurately reflect long-term fatty acid exposures. However, we did detect the established protective effect of DHA against CHD.[1,3,4,11,12,28] Second, we estimated δ-5 and δ-6 desaturase activities based on n-6 PUFAs, while δ-5 and δ-6 desaturases are not only involved in n-6 PUFA conversion, but also in n-3 PUFA conversion. However, in comparison to n-6 PUFA conversion, the amount of n-3 PUFA conversion is relatively small,[15] which should not have affected our results. Third, other potential unmeasured environmental or physiological factors could have confounded the observed associations. However, the relatively large magnitude of the protective effect of increased δ-5 desaturase activity relative to the effect of other risk factors for CHD makes confounding with other unknown risk factors unlikely. Finally, our models that included total and HDL cholesterol may have been over-adjusted,

as these are probably intermediates in the metabolic pathway between desaturase and CHD risk.

In conclusion, in this prospective cohort study, we showed evidence for a reduced CHD risk with increased δ-5 desaturase activity that was partly mediated by DHA. δ-5 Desaturase might play a fundamental role in protecting us from CHD.

Acknowledgements

The monitoring project on cardiovascular disease risk factors was financially supported by the Ministry of Public Health, Welfare and Sports of the Netherlands. The authors wish to thank D Kromhout for supervision of the project and the project steering committee consisting of (not mentioning co-authors of this article) H. B. Bueno de Mesquita, H. A. Smit, and J. C. Seidell (project leader). Furthermore, the authors thank epidemiologists and field-workers of the Municipal Health Services in Maastricht for their contribution to baseline data collection and those involved in the logistic management (A. Jansen and J. Steenbrink) and the data management (A. Blokstra, A. van Kessel, P. Steinberger, E. den Hoedt, I. Toxopeus, J. van der Laan). The authors further wish to thank D. Jaspers, A. Hilton, V. Visser, P. Erkens, S. Philippens, J. Bremen, B. Bleijlevens, T. van Moergastel, S. van de Crommert for assistance in clinical data collection, and E. Vijge and A. Engelen for preparing the plasma samples. Statistics Netherlands is acknowledged for providing data on causes of death.

Source of funding

This study was supported by grant 2006B195 of the Netherlands Heart Foundation.

Disclosure

None

REFERENCES

1. Mozaffarian D. Fish and n-3 fatty acids for the prevention of fatal coronary heart disease and sudden cardiac death. *Am J Clin Nutr.* Jun 2008;87(6):1991S-1996S.

2. Harris WS, Mozaffarian D, Rimm E, et al. Omega-6 fatty acids and risk for cardiovascular disease: a science advisory from the American Heart Association Nutrition Subcommittee of the Council on Nutrition, Physical Activity, and Metabolism; Council on Cardiovascular Nursing; and Council on Epidemiology and Prevention. *Circulation.* Feb 17 2009;119(6):902-907.

3. De Caterina R. n-3 fatty acids in cardiovascular disease. *N Engl J Med.* Jun 23 2011;364(25):2439-2450.

4. Joensen AM, Overvad K, Dethlefsen C, et al. Marine n-3 Polyunsaturated Fatty Acids in Adipose Tissue and the Risk of Acute Coronary Syndrome. *Circulation.* Sep 13 2011;124(11):1232-1238.

5. Yokoyama M, Origasa H, Matsuzaki M, et al. Effects of eicosapentaenoic acid on major coronary events in hypercholesterolaemic patients (JELIS): a randomised open-label, blinded endpoint analysis. *Lancet.* Mar 31 2007;369(9567):1090-1098.

6. Kromhout D, Giltay EJ, Geleijnse JM. n-3 fatty acids and cardiovascular events after myocardial infarction. *N Engl J Med.* Nov 18 2010;363(21):2015-2026.

7. Rauch B, Schiele R, Schneider S, et al. OMEGA, a randomized, placebo-controlled trial to test the effect of highly purified omega-3 fatty acids on top of modern guideline-adjusted therapy after myocardial infarction. *Circulation.* Nov 23 2010;122(21):2152-2159.

8. Mozaffarian D, Micha R, Wallace S. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. *PLoS medicine.* Mar 2010;7(3):e1000252.

9. Baylin A, Campos H. The use of fatty acid biomarkers to reflect dietary intake. *Curr Opin Lipidol.* Feb 2006;17(1):22-27.

10. Raatz SK, Bibus D, Thomas W, Kris-Etherton P. Total fat intake modifies plasma fatty acid composition in humans. *J Nutr.* Feb 2001;131(2):231-234.

11. Nakamura MT, Nara TY. Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. *Annu Rev Nutr.* 2004;24:345-376.

12. Schmitz G, Ecker J. The opposing effects of n-3 and n-6 fatty acids. *Prog Lipid Res.* Mar 2008;47(2):147-155.

13. Cho HP, Nakamura MT, Clarke SD. Cloning, expression, and nutritional regulation of the mammalian Delta-6 desaturase. *J Biol Chem.* Jan 1 1999;274(1):471-477.

14. Cho HP, Nakamura M, Clarke SD. Cloning, expression, and fatty acid regulation of the human delta-5 desaturase. *J Biol Chem.* Dec 24 1999;274(52):37335-37339.

15. Burdge G. Alpha-linolenic acid metabolism in men and women: nutritional and biological implications. *Curr Opin Clin Nutr Metab Care.* Mar 2004;7(2):137-144.

16. Marquardt A, Stohr H, White K, Weber BH. cDNA cloning, genomic structure, and chromosomal localization of three members of the human fatty acid desaturase family. *Genomics.* Jun 1 2000;66(2):175-183.

17. Schaeffer L, Gohlke H, Muller M, et al. Common genetic variants of the FADS1 FADS2 gene cluster and their reconstructed haplotypes are associated with the fatty acid composition in phospholipids. *Hum Mol Genet.* Jun 1 2006;15(11):1745-1756.

18. Tanaka T, Shen J, Abecasis GR, et al. Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. *PLoS Genet.* Jan 2009;5(1):e1000338.

19. Lemaitre RN, Tanaka T, Tang W, et al. Genetic Loci Associated with Plasma Phospholipid n-3 Fatty Acids: A Meta-Analysis of Genome-Wide Association Studies from the CHARGE Consortium. *PLoS genetics.* Jul 2011;7(7):e1002193.

20. Dixon AL, Liang L, Moffatt MF, et al. A genome-wide association study of global gene expression. *Nat Genet.* Oct 2007;39(10):1202-1207.

21. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* Jan 2009;41(1):56-65.

22. Schadt EE, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* May 6 2008;6(5):e107.

23. Plaisier CL, Horvath S, Huertas-Vazquez A, et al. A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genet.* Sep 2009;5(9):e1000642.

24. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet.* Feb 2010;42(2):105-116.

25. Bokor S, Dumont J, Spinneker A, et al. Single nucleotide polymorphisms in the FADS gene cluster are associated with delta-5 and delta-6 desaturase activities estimated by serum fatty acid ratios. *J Lipid Res.* Aug 2010;51(8):2325-2333.

26. Mathias RA, Vergara C, Gao L, et al. FADS genetic variants and omega-6 polyunsaturated fatty acid metabolism in a homogeneous island population. *J Lipid Res.* Sep 2010;51(9):2766-2774.

27. Lu Y, Feskens EJ, Dolle ME, et al. Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem Cohort Study. *Am J Clin Nutr.* Jul 2010;92(1):258-265.

28. Harris WS, Poston WC, Haddock CK. Tissue n-3 and n-6 fatty acids and risk for coronary heart disease events. *Atherosclerosis.* Jul 2007;193(1):1-10.

29. Verschuren W, Leer E, Blokstra A, et al. Cardiovascular disease risk factors in The Netherlands. *Neth J Cardiol.* 1993;6:205-210.

30. Merry AH, Boer JM, Schouten LJ, et al. Validity of coronary heart diseases and heart failure based on hospital discharge and mortality data in the Netherlands using the cardiovascular registry Maastricht cohort study. *Eur J Epidemiol.* 2009;24(5):237-247.

31. Cai J, Zeng D. Sample size/power calculation for case-cohort studies. *Biometrics.* Dec 2004;60(4):1015-1024.

32. Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. *J Clin Epidemiol.* Dec 1999;52(12):1165-1172.

33. Langholz B, Jiao J. Computational methods for case-cohort studies. *Computational Statistics & Data Analysis.* 2007;51:3737-3748.

34. Kattermann R, Jaworek D, Moller G, et al. Multicentre study of a new enzymatic method of cholesterol determination. *Journal of clinical chemistry and clinical biochemistry. Zeitschrift fur klinische Chemie und klinische Biochemie.* Mar 1984;22(3):245-251.

35. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* Feb 11 1988;16(3):1215.

36. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* Aug 5 2010;466(7307):707-713.

37. Eijgelsheim M, Newton-Cheh C, Sotoodehnia N, et al. Genome-wide association analysis identifies multiple loci related to resting heart rate. *Human molecular genetics.* Oct 1 2010;19(19):3885-3894.

38. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011;43(4):333-338.

39. Calder PC. Dietary arachidonic acid: harmful, harmless or helpful? *Br J Nutr.* Sep 2007;98(3):451-453.

40. Calder PC. The American Heart Association advisory on n-6 fatty acids: evidence based or biased evidence? *Br J Nutr.* Dec 2010;104(11):1575-1576.

41. Serhan CN, Chiang N, Van Dyke TE. Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. *Nature reviews. Immunology.* May 2008;8(5):349-361.

42. Seyberth HW, Oelz O, Kennedy T, et al. Increased arachidonate in lipids after administration to man: effects on prostaglandin biosynthesis. *Clinical pharmacology and therapeutics.* Nov 1975;18(5 Pt 1):521-529.

43. Simopoulos AP. Evolutionary aspects of diet, the omega-6/omega-3 ratio and genetic variation: nutritional implications for chronic diseases. *Biomed Pharmacother.* Nov 2006;60(9):502-507.

44. Kusumoto A, Ishikura Y, Kawashima H, Kiso Y, Takai S, Miyazaki M. Effects of arachidonate-enriched triacylglycerol supplementation on serum fatty acids and platelet aggregation in healthy male subjects with a fish diet. *Br J Nutr.* Sep 2007;98(3):626-635.

45. Nelson GJ, Schmidt PC, Bartolini G, Kelley DS, Kyle D. The effect of dietary arachidonic acid on platelet function, platelet fatty acid composition, and blood coagulation in humans. *Lipids.* Apr 1997;32(4):421-425.

46. Nelson GJ, Schmidt PC, Bartolini G, et al. The effect of dietary arachidonic acid on plasma lipoprotein distributions, apoproteins, blood lipid levels, and tissue fatty acid composition in humans. *Lipids.* Apr 1997;32(4):427-433.

47. Ferretti A, Nelson GJ, Schmidt PC, Kelley DS, Bartolini G, Flanagan VP. Increased dietary arachidonic acid enhances the synthesis of vasoactive eicosanoids in humans. *Lipids.* Apr 1997;32(4):435-439.
48. Kelley DS, Taylor PC, Nelson GJ, Schmidt PC, Mackey BE, Kyle D. Effects of dietary arachidonic acid on human immune response. *Lipids.* Apr 1997;32(4):449-456.
49. http://biogps.org/#goto=genereport&id=3992. Accessed on August 15 2011.

Supplemental Appendix

Supplementary Table 1. Association of rs174547 with incident coronary heart disease (CHD) risk.

	rs174547 genotype			rs174547 G allele*
	AA	*AG*	*GG*	
Incident CHD, n	234	259	56	-
Model 1†	1.0	1.00 (0.80-1.26)	0.96 (0.67-1.37)	0.99 (0.84-1.16)
Model 2‡	1.0	1.15 (0.88-1.49)	0.91 (0.60-1.39)	1.02 (0.85-1.22)

* The model assumes that each copy of the G allele contributes equally to coronary heart disease risk.
† Model 1 was adjusted for age and sex.
‡ Model 2 was adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, diabetes, total cholesterol, and high-density lipoprotein cholesterol.

Supplementary Table 2. Association between baseline n-6 PUFA in plasma cholesteryl esters (precursors and products of δ5- or δ6-desaturase) and incident coronary heart disease (CHD).

	Quintile of C18:2n-6 (linoleic acid)					*P* value for trend*
	First	Second	Third	Fourth	Fifth	
Incident CHD, n	127	125	102	97	86	
Model 1†	1	1.11 (0.80-1.54)	1.02 (0.72-1.44)	1.10 (0.78-1.56)	0.80 (0.57-1.13)	0.247
Model 2‡	1	1.16 (0.81-1.66)	1.27 (0.88-1.83)	1.30 (0.89-1.89)	0.91 (0.62-1.32)	0.861
Model 3§	1	1.09 (0.74-1.60)	1.34 (0.91-1.96)	1.38 (0.92-2.09)	1.01 (0.68-1.48)	0.395
Model 4¶	1	1.09 (0.74-1.61)	1.34 (0.91-1.96)	1.39 (0.92-2.09)	1.01 (0.69-1.49)	0.389
	Quintile of C18:3n-6					*P* value for trend*
	First	Second	Third	Fourth	Fifth	
Incident CHD, n	93	109	100	107	128	
Model 1†	1	1.05 (0.74-1.50)	0.89 (0.62-1.28)	0.92 (0.64-1.32)	1.00 (0.71-1.41)	0.706
Model 2‡	1	1.12 (0.76-1.64)	0.90 (0.61-1.33)	1.02 (0.69-1.50)	0.93 (0.64-1.37)	0.867
Model 3§	1	1.13 (0.75-1.72)	0.90 (0.58-1.38)	1.01 (0.66-1.53)	1.05 (0.70-1.57)	0.605
Model 4¶	1	1.14 (0.75-1.73)	0.90 (0.58-1.39)	1.02 (0.67-1.56)	1.06 (0.70-1.61)	0.536
	Quintile of C20:3n-6					*P* value for trend*
	First	Second	Third	Fourth	Fifth	
Incident CHD, n	88	80	107	125	137	
Model 1†	1	0.98 (0.68-1.43)	1.14 (0.80-1.64)	1.29 (0.91-1.82)	1.39 (0.99-1.96)	0.024
Model 2‡	1	0.91 (0.61-1.34)	1.06 (0.73-1.55)	1.14 (0.78-1.66)	1.43 (1.00-2.05)	0.011
Model 3§	1	0.94 (0.61-1.45)	1.13 (0.75-1.69)	1.00 (0.67-1.50)	1.11 (0.75-1.67)	0.494

Model 4¶	1	0.95 (0.62-1.47)	1.15 (0.76-1.73)	1.03 (0.67-1.57)	1.14 (0.75-1.74)	0.420

	Quintile of C20:4n-6 (arachidonic acid)					*P* value for trend*
	First	Second	Third	Fourth	Fifth	
Incident CHD, n	127	107	120	79	104	
Model 1†	1	0.89 (0.63-1.24)	0.96 (0.70-1.34)	0.65 (0.45-0.92)	0.76 (0.54-1.06)	0.031
Model 2‡	1	0.86 (0.60-1.22)	0.92 (0.65-1.31)	0.64 (0.43-0.94)	0.77 (0.53-1.10)	0.088
Model 3§	1	1.06 (0.72-1.56)	1.09 (0.74-1.60)	0.79 (0.51-1.21)	0.86 (0.58-1.28)	0.163
Model 4¶	1	1.04 (0.70-1.55)	1.07 (0.73-1.58)	0.76 (0.48-1.20)	0.83 (0.53-1.28)	0.150

* From models with fatty acids included as continuous variables.
† Model 1 is adjusted for age and sex.
‡ Model 2 is adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, and diabetes.
§ Model 3 is adjusted for all covariates in model 2, total cholesterol, and high-density lipoprotein cholesterol.
¶ Model 4 is adjusted for all covariates in model 3 and baseline n-3 PUFA in plasma cholesteryl esters.

Supplementary Table 3. Association of baseline C20:5n-3 (EPA) and C22:6n-3 (DHA) in plasma cholesteryl esters with incident coronary heart disease (CHD).

	Quintile of C20:5n-3 (EPA)					*P* value for trend†
	First (0.00)*	Second (0.28)	Third (0.38)	Fourth (0.52)	Fifth (0.83)	
Incident CHD, n	96	100	104	114	123	
Model 1‡	1	0.81 (0.57-1.16)	0.88 (0.62-1.26)	0.82 (0.57-1.16)	0.82 (0.58-1.17)	0.348
Model 2§	1	0.86 (0.59-1.24)	0.91 (0.62-1.33)	0.88 (0.61-1.27)	0.76 (0.52-1.11)	0.243
Model 3¶	1	0.90 (0.60-1.34)	0.90 (0.59-1.39)	0.79 (0.52-1.20)	0.89 (0.58-1.35)	0.733
Model 4$	1	0.90 (0.60-1.35)	0.90 (0.59-1.39)	0.80 (0.52-1.22)	0.89 (0.58-1.35)	0.724
	Quintile of C22:6n-3 (DHA)					*P* value for trend†
	First (0.00)*	Second (0.20)	Third (0.27)	Fourth (0.34)	Fifth (0.46)	
Incident CHD, n	89	139	93	112	104	
Model 1‡	1	1.45 (1.03-2.05)	0.87 (0.60-1.25)	1.07 (0.74-1.53)	0.93 (0.65-1.34)	0.286
Model 2§	1	1.49 (1.05-2.13)	0.84 (0.58-1.24)	1.13 (0.77-1.67)	0.86 (0.58-1.29)	0.268
Model 3¶	1	0.96 (0.65-1.43)	0.61 (0.41-0.92)	0.89 (0.59-1.34)	0.65 (0.42-0.99)	0.049
Model 4$	1	0.95 (0.64-1.40)	0.57 (0.38-0.87)	0.82 (0.53-1.26)	0.59 (0.37-0.93)	0.027

* Median percentages of EPA and DHA in each quintile are listed between brackets.
† From models with fatty acids included as continuous variables.
‡ Model 1 is adjusted for age and sex.
§ Model 2 is adjusted for age, sex, systolic blood pressure, hypertensive medication use, current smoking, and diabetes.
¶ Model 3 is adjusted for all covariates in model 2, total cholesterol, and high-density lipoprotein cholesterol.
$ Model 4 is adjusted for all covariates in model 3 and baseline n-6 PUFA in plasma cholesteryl esters.

Chapter 8

Research article

Literature-based genetic risk scores for coronary heart disease; the CAREMA prospective-cohort study

Anika A.M. Vaarhorst[a], Yingchang Lu[b, c], Bastiaan T. Heijmans[a], Martijn E.T. Dollé[c], Stefan Böhringer[d], Hein Putter[d], Sandra Imholz[c], Audrey H.H. Merry[e] , Marleen M. van Greevenbroek[f], J. Wouter Jukema[g,h,i], Anton P.M. Gorgels[j], Piet A. van den Brandt[e,k], Michael R. Müller[b], Leo J. Schouten[k], Edith J.M. Feskens[b], Jolanda M.A. Boer[c], P. Eline Slagboom[a, l]

[a] Molecular Epidemiology section, Leiden University Medical Center, Leiden, The Netherlands
[b] Division of Human Nutrition, Wageningen University and Research Center, Wageningen, The Netherlands
[c] National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
[d] Department of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands
[e] Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands
[f] Cardiovascular Research Institute Maastricht (CARIM), Maastricht University Medical Centre, Maastricht, The Netherlands
[g] Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands,
[h] Interuniversity Cardiology Institute of the Netherlands (ICIN), Utrecht, The Netherlands
[i] The Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands
[j] Department of Cardiology, University Hospital Maastricht, Maastricht, The Netherlands
[k] Department of Epidemiology, GROW School of Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands
[l] Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands

(Submitted)

Abstract

Background: Genome-wide association studies (GWAS) have identified many single nucleotide polymorphisms (SNPs) associated with coronary heart disease (CHD) or its risk factors. Using a case-cohort study within the prospective CAREMA cohort, we tested if genetic risk scores (GRS) based on those SNPs are associated with and predictive for future CHD.

Methods and Results: 742 incident cases, i.e. participants who developed CHD during a median follow up of 12.1 years (range 0.0-16.9 years) were compared with a randomly selected sub-cohort of 2221 participants selected from the total cohort (N = 21,148). 179 SNPs associated with CHD or its factors in GWAS published up to May 2, 2011 were genotyped. Five GRS were investigated; an Overall GRS of all SNPs, a Risk Factor GRS of 153 SNPs associated with CHD risk factors, a non-weighted and a weighted CHD GRS of 29 CHD associated SNPs and a LASSO GRS of the 14 most discriminating SNPs identified using LASSO regression. The weighted CHD GRS (HR = 1.12 per weighted risk allele; 95%-CI=1.04-1.21) and the LASSO GRS (HR = 1.39 per weighted risk allele; 95%-CI=1.26-1.53) were associated with CHD independent of traditional risk factors. The net reclassification index (NRI) improved when the CHD or the LASSO GRS were added to a model including traditional CHD risk factors (NRI=3.3%, *P*=0.017; NRI=8.1%, *P*=0.001, respectively).

Conclusions The weighted CHD GRS and the LASSO GRS are associated with CHD independent of traditional risk factors and particularly the LASSO GRSs may improve risk classification.

***Keywords**:* case-cohort study, genetic risk score, coronary heart disease, risk prediction

Introduction

Coronary heart disease (CHD) is a complex disease influenced by both life-style and genetic factors. During the last years, genome-wide association studies (GWAS) have identified multiple common single nucleotide polymorphisms (SNPs) that are robustly associated with the risk of CHD[1-6] and its risk factors including blood pressure[7, 8] and plasma lipid levels.[9, 10] It was hoped that GWAS would identify SNPs that were useful in predicting CHD risk. Individually, however, these SNPs have a relatively small effect size. Greater effect sizes may be expected for genetic risk scores (GRS) that comprise the cumulative effect of individual SNPs.[11, 12]

Two previous prospective studies investigated the association of a GRS with incident CHD.[13, 14] Paynter and co-workers constructed a GRS from 101 published GWAS SNPs known to be associated with cardiovascular disease and its risk factors.[13] The GRS was tested for an association with a combined end-point of CHD and stroke in US women. An association was observed, but this disappeared after adjustment for traditional cardiovascular risk factors. The association was somewhat stronger for a GRS based on 12 of these SNPs which were previously shown to be associated with cardiovascular disease only.[13] In a recent Finnish cohort a GRS based on 13 SNPs was independently associated with 10 year incidence of CHD.[14] However, this GRS did not improve risk discrimination over traditional risk factors and a family history of MI.[14]

An important observation from these studies was that SNPs identified in GWAS, which commonly compare cases with a history of CHD to control groups of various origins, were not always detected in population-based prospective studies.[13, 14] Case-control studies may overestimate effect sizes of SNPs associated to disease,[15] thereby potentially diluting a GRS based on these SNPs applied to a study with a prospective design, resulting in a poor association. To construct a meaningful GRS, it may therefore be important to select a subset of SNPs that are actually associated with CHD in a population-based setting. The complication is that it is not possible to select SNPs based on their association with CHD, construct a GRS and test this GRS in the same population, because this would result in over-fitting the data and over optimistic results. A combination of cross-validation and LASSO regression can avoid this problem.[16-18]

In our study, we constructed five different GRSs on the basis of 179 SNPs that were associated with CHD or its traditional risk factors in GWAS published up to May 2, 2011, and investigated their association with the occurrence of CHD during a median follow-up of 12.1 years in a population-based setting.

Material and methods

Study population

The Cardiovascular Registry Maastricht (CAREMA) study population has been described before.[19] In short, the study participants, living in the Maastricht region, were selected from two large monitoring projects in the Netherlands: the monitoring project on cardiovascular disease risk factors (PPHVZ) 1987-1991 and the monitoring project on chronic disease risk factors (MORGEN Project) 1993-1997, including the transition year (1992) between these

two projects. Each year, a random sample of people aged between 20 and 59 years were selected from municipal registries of Maastricht and surrounding communities (i.e. Eijsden, Margraten, Meerssen, and Valkenburg aan de Geul). Between 1987 and 1997, 21,662 men and women, born between 1927 and 1997, were included in this study, of whom 21,148 participants (97.6%) had given informed consent to retrieve information from the municipal population registries and from their general practitioner and specialist.

Cardiologic follow-up
The cardiologic follow-up has been described in detail earlier.[19] In short by linking the cohort to the hospital information system of the University Hospital Maastricht (UHM), 97.6% of the cohort members could be found (N=20,632). Next, these subjects were linked to the cardiology information system of the department of cardiology. For participants who died, the cause of death was obtained from Statistics Netherlands. In addition, the CAREMA cohort was linked to the hospital discharge registry of the UHM to enlarge the completeness of the cardiologic follow-up.

All participants, who were younger than 30 years of age at baseline (N=3505), who had an acute myocardial infarction (MI), unstable angina pectoris (UAP), a coronary artery bypass graft (CABG) or a percutaneous transluminal coronary angioplasty (PTCA) before baseline (N=187), were lost to follow-up (N=17), or were included in the transition year 1992 (N=2203), were excluded from further analysis.

Case-cohort design
During the follow-up period (median, 12.1 yrs; range, 0.0-16.9 yrs), which ended on 31 December 2003, 742 participants developed incident CHD, consisting of 368 patients with incident MI, 294 patients with UAP and 80 patients who have died because of CHD. No blood for DNA extraction was available for 59 patients, and in 15 cases the DNA extraction was unsuccessful. Twenty-two cases were excluded because no complete information on traditional risk factors (i.e. total cholesterol, HDL-cholesterol, systolic blood pressure, current smoking, body mass index (BMI), current diabetes status and a family history of MI) was available. Thus the available case group consists of 646 patients.

From the full eligible CAREMA cohort, randomly a sub-cohort of 2221 participants was drawn. No blood was available for 188 participants, DNA extraction was unsuccessful in 49 participants and no complete information on traditional risk factors was available for 71 participants. Thus the available sub-cohort group consisted of 1913 participants.

Determination of risk factors
At baseline, all participants filled in a questionnaire on medical history, parental history of MI, and lifestyle factors such as smoking. During a medical examination at baseline, information was collected on blood pressure, height and weight. In addition, non-fasting blood samples were collected using EDTA tubes. The blood was centrifuged for 10 minutes at 1000 rpm and fractioned into blood plasma, white blood cells fractions (buffy coats) and erythrocytes and subsequently stored at -20°C. Within three weeks, the blood plasma samples were transported to the Lipid Reference Laboratory of the University Hospital

Dijkzigt (LRL) in Rotterdam where the total and HDL-cholesterol levels were determined using a CHOD-PAP method.[20] The LRL in Rotterdam is a permanent member of the International Cholesterol Reference Method Laboratory Network.

SNP selection
Using the GWAS Catalog,[21] available at: www.genome.gov/gwastudies and accessed at May 2, 2011, 248 SNPs were identified that showed an association with coronary heart disease or with its intermediate risk factors (blood pressure, anthropomorphic traits (BMI, waist circumference and waist to hip ratio), blood lipid levels and type 2 diabetes) in at least two GWA studies or in a meta-analysis at a genome-wide significance levels ($P < 5*10^{-8}$). For the SNPs that were in perfect linkage disequilibrium (LD) (D'= 1; R^2=1) with each other, one of them was excluded (N=22). For 29 SNPs it was not possible to design primers or they did not fit in the SNP assay designs, leaving 197 SNPs for genotyping. For a detailed overview see supplementary table s1.

DNA extraction & Genotyping
DNA was extracted from the white blood cell fractions (buffy coats), using a standard procedure.[22] The resulting DNA pellet was dissolved in TE buffer and DNA concentrations were determined using the Nanodrop ND1000 Spectrophotometer. All SNPs were genotyped using the iPLEX Gold chemistry of Sequenom's MassARRAY platform (San Diego, CA, USA). Sequenom's MassARRAY® Assay Design 3.1 Software was used for the SNP assay designs, and Sequenom's SpectroTyper 4.0 software was used to call genotypes automatically, followed by manual review. Nine SNPs failed genotyping (e.g no PCR product or an abnormal cluster plot), two SNP were removed due to a success rate lower than 90% and seven SNPs were removed because they were out of Hardy Weinberg equilibrium (P-value<0.001) calculated in the sub-cohort, leaving 179 SNPs for analysis. For details see supplementary table s1. Individuals (four cases and eight sub-cohort members) were removed if their success rate was lower than 80%, leaving 642 cases and 1905 sub-cohort members.

Statistical analysis
Not every SNP was successfully genotyped in every person (for success rate per SNP see supplementary table s1). Therefore, we applied a multiple imputation method (R packages *mi* (version 0.08-08)) to impute these missing values. First, missing values for the SNPs were randomly imputed, where the chance of being a major homozygote, heterozygote or minor homozygote depends on the distribution of the particular SNPs. This process was repeated 20 times, thus 20 new datasets with non-missing data were created. Next, in every imputed dataset the appropriate analysis was performed. Finally the results from every imputed dataset were combined using the R package *mitools* (version 2.0).

We composed different GRS. The first genetic risk score (GRS) indicated as Overall GRS was constructed by counting the number of all risk alleles previously associated with CHD and its intermediate risk factors (179 SNPs). Risk alleles associated with more than one risk factor were included only once. Using the same method, three additional GRS were constructed. The first one was based on SNPs previously associated with traditional risk

factors including blood pressure, anthropomorphic traits (i.e, BMI, waist and waist/hip ratio), type 2 diabetes and blood lipid levels (153 SNPs, indicated as Risk Factor GRS). The next two were based on SNPs previously associated with CHD, namely a non-weighted CHD GRS composed of 29 SNPs previously associated with CHD and a weighted CHD GRS. A weighted CHD GRS was based on all the 29 risk alleles for CHD multiplied by their log odds ratio divided by the log odds ratio of the SNP with the highest effect size. As a result the SNP with the highest odds ratio has a weight of one. For estimating the weights of the individual SNPs, the odds ratios of the relevant SNPs were obtained from two large GWAS on coronary heart disease.[23, 24] For more details see supplementary table s2.

Because it is not likely that all GWAS-identified SNPs would contribute to the risk of future CHD in our population-based prospective study,[25] we also constructed a GRS based on least absolute shrinkage and selection operator (LASSO) regression (indicated as LASSO GRS). With LASSO regression it is possible to perform variable selection and shrinkage at the same time, which makes it an effective method for finding prediction rules based in high dimensional data.[16-18] Since we did not have an independent population which could be used as a discovery population for selecting SNPs and their appropriate weights, a cross-validation approach in combination with LASSO regression was used. We divided our population in ten equally sized subgroups and nine of these groups were combined as a training dataset. In this training dataset relevant SNPs were selected using LASSO regression,[16, 17] and their coefficients were calculated. Next, these SNPs and their coefficients were tested in the independent tenth group. This process was repeated for each of the ten testing sets. The resulting SNPs and their weights, based on the coefficients resulting from the LASSO regression, were then used to construct the LASSO GRS. To perform this analysis we used the R package *penalized* (version 0.9-32).[18]

We tested whether the five different GRS were associated with future CHD, using Cox-proportional hazards models with robust variance estimation to adjust for the case-cohort design according to the method of Prentice.[26] These models were used to estimate the 10-year risk of CHD using a base model with and without each GRS. The base model consists of the covariates sex, current smoking, history of diabetes, parental history of MI, HDL cholesterol, total cholesterol, systolic blood pressure and BMI. Age in years was used as the time-scale variable and we adjusted for delayed entry. The same analyses were also done for every SNP separately.

Due to a change in scaling, the non-weighted CHD GRS theoretically ranges from 0-58, whereas the weighted CHD GRS ranges from 0-22.14, the hazard ratio of a weighted GRS cannot be directly compared to a non-weighted GRS. Thus, for comparing the effects of the different GRS, we calculated the z-scores for all five GRS in all individuals and performed the above analysis on these standardized GRS.

To investigate whether using a GRS in addition to traditional risk factors improved risk discrimination, we compared the estimates of the 10-year risk of CHD between the base model based on the traditional risk factors alone and the base model plus the GRS using the

c-statistic. To assess the differences in c-statistic between the base-model and the base-model with the GRS, we used the non-parametric method of DeLong et.al.[27] Next, to see whether participants were categorized into more appropriate risk categories when using the GRS in addition to the traditional risk factors, these 10-year risk estimates were compared with the 10-year risks from the base model, using the net reclassification improvement index (NRI) as described by Pencina et al.[28] However, this NRI is not directly applicable to our study because of non-cases (N=550) who left the study prematurely (i.e. before 10 years of follow-up) and cases (N=208) that got an event after 10 years of follow-up.[29] Therefore cases who developed an event after 10 years of follow-up were treated as controls that left the study at 10 years of follow-up (N=208), which resulted in 434 cases and 2012 non-cases. Next, we calculated a Kaplan-Meier curve for the sub-cohort. Based on this Kaplan-Meier curve, weights were assigned to every individual. Weights of zero were assigned to non-cases who left the study before 10 years of follow-up (N=550). Using the 10-year risk estimates, subjects were divided in the following risk categories: 0-<2%, 2-<5%, 5-<10%, and≥10%. The choice of the risk categories were based on the SCORE risk.[30] For all analyses we used the software R (version 2.10.1, www.r-project.org) in combination with the package survival (version 2.35-8) for calculating the Cox-proportional hazards models according to Prentice and the Kaplan-Meier curve. For calculating the c-statistics and the corresponding p-values, we used the R package pROC (version 1.4.1).

Results

Baseline characteristics

In Table 1, it is shown that all traditional risk factors for CHD are associated with future CHD in the CAREMA study population. When all traditional risk factors were entered in a Cox proportional hazards model, with age as the time-scale variable, BMI was no longer significantly associated with future CHD.

Table 1. Baseline characteristics of the CAREMA case-cohort study and their association with future CHD

Variables	Cases (N=646)	Sub-cohort (N=1913)*	HR (95% CI)†	Adjusted HR (95% CI)‡
Age (years)	50.2 ± 7.1	45.2 ± 8.4	1.08 (1.07-1.09)§	-
Tot Chol (mmol/L)	6.3 ± 1.2	5.6 ± 1.1	1.47 (1.35-1.61)	1.33 (1.20-1.48)
HDL-C (mmol/L)	1.0 ± 0.3	1.3 ± 0.3	0.05 (0.03-0.07)	0.14 (0.08-0.22)
SBP (mmHg)	128.9 ± 17.1	120.2 ± 15.2	1.02 (1.02-1.03)	1.02 (1.01-1.03)
BMI (kg/m^2)	26.6 ± 4.0	25.0 ± 3.7	1.08 (1.05-1.10)	1.02 (0.99-1.05)
Men	483 (74.8%)	870 (45.5%)	1.41 (3.31-50.3)	2.52 (1.93-3.29)
Current smoking‖	373 (57.7%)	790 (41.3%)	2.03 (1.68-2.46)	1.95 (1.54-2.46)
Type 2 diabetes	24 (3.7%)	17 (0.9%)	3.60 (1.83-7.08)	3.03 (1.46-6.26)
Parental history of MI	270 (41.8%)	670 (35.0%)	1.31 (1.12-1.54)	1.41 (1.16-1.72)

Results are presented as means ± SD or N (%)

* Includes 101 cases

† Univariate analysis; age in years was used as the time-scale variable

‡ All the variables were added into one multivariable Cox proportional hazards models model and age was used as the time scale-variable.

[§] Follow-up time in years was used as the time-scale variable
[‖] Current smokers compared to non-current smokers
HR, hazard ratio; *95% CI*, 95% confidence interval; *Tot Chol*, total cholesterol; *HDL-C*, high density lipoprotein cholesterol; *SBP*, systolic blood pressure; *BMI*, body mass index; *MI*, myocardial infarction

Genetic Risk Scores
The Overall GRS constructed of all the 179 SNPs had a mean value (SD) of 182.9 (9.2) for all participants, with a range from 142 to 217 risk alleles. As shown in table 2, this Overall GRS was associated with CHD with a hazard ratio (HR) of 1.02 per risk allele (95% CI = 1.01-1.03); after adjustment for the traditional risk factors the effect size further attenuated (HR/risk allele = 1.01; 95% CI = 1.00-1.02). For a detailed overview of all the separate SNPs from which the GRS was constructed, and their association with incident CHD before and after adjustment for traditional risk factors, see supplementary table s3.

As shown in Table 2, the Risk Factor GRS constructed from the 153 SNPs selected for their previous association with CHD risk factors (blood pressure, anthropomorphic traits, blood lipid levels and type 2 diabetes) was associated with future CHD (HR/risk allele=1.02; 95% CI =1.01-1.03), but not after adjustment for all traditional risk factors (HR/risk allele=1.00; 95% CI =0.99-1.02).

Next, the CHD GRS constructed from the 29 SNPs selected for their previous association with CHD was investigated. The CHD GRS was associated with future CHD with a HR of 1.04 per risk allele (95% CI = 1.01-1.07). After adjustment for the traditional risk factors this GRS was no longer associated with future CHD (HR = 1.03; 95% CI = 0.95-1.12). The CHD GRS that was weighted for previously reported effect sizes of the included SNPs, however, was associated with future CHD both prior (HR = 1.15; 95% CI = 1.07-1.23) and after the adjustment for the traditional risk factors (HR = 1.12; 95% CI = 1.04-1.21).

Cross-validation and a LASSO GRS
Using LASSO regression,[18] a subset of 14 SNPs was selected as the optimal predictor of CHD in the current data-set: rs10757278 (locus = 9p21.3, nearby genes = *CDKN2A, CDKN2B*), rs2925979 (locus = 16q23.2, nearby genes = *CMIP*), rs6882076 (locus = 5q33.3, nearby genes = *TIMD4*), rs2954029 and rs6987702 (locus = 8q24.13, nearby genes = *TRIB1*), rs10889352 (locus = 1p31.3, nearby gene = *DOCK7*), rs2972146 (locus = 2q36.3, nearby genes = *IRS1*), rs11556924 (locus = 7q32.2, nearby genes = *ZC3HC1*), rs514230 (locus = 1q42.3, nearby genes = *IRF2BP2*), rs8050136 (locus = 16q12.2, nearby gene = *FTO*), rs181362 (locus = 22q11.21, nearby genes = *UBE2L3)*, rs646776 (locus = 1p13.3, nearby genes = *CELSR2, PSRC1, SORT1*), rs925946 (locus = 11p14.1, nearby genes = *BDNF*) and rs2000999 (locus = 16q22.2, nearby genes = *HPR*). These 14 SNPs were used to construct a LASSO GRS with weighting factors (see supplementary table s4). This weighted LASSO GRS was associated with future CHD (HR = 1.48; 95% CI = 1.31-1.29) and remained significantly associated with CHD after adjustment for traditional risk factors (HR = 1.39; 95% CI = 1.26-1.53). To be able to compare the effect sizes of the various GRSs tested, we computed the HR for each GRS per standard deviation (sd) increase in number of risk alleles (see table 2). The standardized HR of the LASSO GRS

(HR/sd increase = 1.49; 95% CI =1.31-1.69) was higher than that of the weighted CHD GRS (HR/ sd increase = 1.21; 95% = 1.10-1.33).

Table 2. Associations between the different genetic risk scores and future CHD.

Genetic Risk Scores	N SNPs	Cases (N=642)	Sub-cohort (N=1905)	Coronary heart disease		
				HR (95% CI)	HR std[†] (95% CI)	*P*
Overall GRS	179	184.1 ± 8.8	182.5 ± 9.2	1.02 (1.01-1.03)	1.20 (1.09-1.32)	<0.001
Adjusted for TRF[*]				1.01 (1.00-1.02)	1.07 (0.99-1.16)	0.084
Risk Factor GRS	153	156.5 ± 8.3	155.3 ± 8.5	1.02 (1.01-1.03)	1.16 (1.06-1.27)	0.002
Adjusted for TRF[*]				1.00 (0.99-1.02)	1.03 (0.95-1.12)	0.452
CHD GRS	29	31.2 ± 3.3	30.7 ± 3.5	1.04 (1.01-1.07)	1.16 (1.04-1.26)	0.004
Adjusted for TRF[*]				1.03 (0.95-1.12)	1.12 (0.85-1.49)	0.413
Weighted CHD GRS	29	11.9 ± 1.3	11.6 ± 1.4	1.15 (1.07-1.23)	1.21 (1.10-1.33)	<0.001
Adjusted for TRF[*]				1.12 (1.04-1.21)	1.17 (1.06-1.30)	0.001
LASSO GRS	14	4.0 ± 1.0	3.7 ± 1.0	1.48 (1.31-1.67)	1.49 (1.31-1.69)	<0.001
Adjusted for TRF[*]				1.39 (1.26-1.53)	1.40 (1.27-1.53)	<0.001

[*] Traditional risk factors include sex, current smoking (yes/no), systolic blood pressure, total cholesterol, HDL cholesterol, self reported diabetes (yes/no), body mass index and parental history of MI parents with MI, one parent with MI, both parents with MI). Age was used as the time scale variable

[†] Increase in hazard ratio per standard deviation

HR, hazard ratio; *std*, standardized; *SNPs*, single nucleotide polymorphisms; *CHD*, coronary heart disease; TRF, traditional risk factors; *GRS*, genetic risk score; *LASSO*, least absolute shrinkage and selection operator

C-indexes

The discriminatory capabilities of the five GRSs without including other variables were low (c-index = 0.547 for the overall GRS, 0.538 for the risk factors GRS, 0.527 for the CHD GRS, 0.550 for the weighted CHD GRS, and 0.599 for the LASSO GRS; Table 3), but all statistically significant (see table 3). When using a prediction model based on all the CHD risk factors (c-index = 0.816), only the LASSO GRS improved the fit of the model (c-index = 0.824, P = 0.011), whereas the other GRSs did not.

Table 3. Improvement in discrimination, as assessed with the C index, when the GRSs (Overall GRS, Risk Factor GRS CHD GRS, weighted GRS or LASSO GRS) are added to the base model (composed age, sex, current smoking and the traditional risk factors[*]).

Model	**C-index**	**P-value**
Overall GRS	0.547	<0.001
Risk Factor GRS	0.538	0.002
CHD GRS	0.527	0.021
Weighted CHD GRS	0.550	<0.001
LASSO GRS	0.599	<0.001
Base model age, sex, current smoking plus traditional risk factors[*]	0.816	-
Base model and overall GRS	0.816	0.966
Base model and risk factor GRS	0.816	0.858
Base model and CHD GRS	0.817	0.417
Base model and weighted CHD GRS	0.818	0.134
Base model and LASSO GRS	0.824	0.011

[*] total cholesterol, HDL cholesterol, systolic blood pressure, self reported diabetes and family history of myocardial infarction

GRS, genetic risk score; *CHD*, coronary heart disease; *LASSO*, least absolute shrinkage and selection operator

Net reclassification index (NRI)

In addition, we calculated the NRI, and found no improvement when the Overall GRS or the Risk Factor GRS were added to the base model (Table 4, for more details see supplementary table 5a-e). Reclassification improved for the non-weighted CHD GRS, but only with 2.3% (P = 0.035). When we used the weighted CHD GRS, 10.0% of the participants were classified into a more appropriate risk category, but for 6.7% of the population the classification got worse, thus the NRI improved with 3.3% (P = 0.017), which could be attributed to an improvement in reclassification for events (i.e. individuals who did have a CHD event; NRI = 2.5%, P = 0.044). An improved reclassification was observed for 23.4% of the population and a worsening for 15.3% of the population when we used the LASSO GRS (NRI = 8.1%, P = 0.001). In this case, the net improvement in NRI could be accredited to an improvement for events (NRI = 5.0%, P = 0.017) as well as non-events (NRI = 3.1%, P < 0.001).

Discussion

Over the past five years, the pace of identification of genetic variants underlying susceptibility to CHD has rapidly increased, leading to an interest in investigating if and how this information might be used in improving CHD risk prediction. Therefore, we constructed a GRS composed of 179 SNPs previously associated with CHD and its intermediate risk factors. This Overall GRS was associated with CHD, but not independently of traditional risk factors and did not improve risk prediction above a model using traditional risk factors. We separated this GRS into a weighted CHD GRS based on SNPs selected for previous association with CHD, and a Risk Factor GRS based on SNPs previously associated with intermediate CHD risk factors. The Risk Factor GRS was not associated with incident CHD. However, the weighted CHD GRS based on 29 CHD SNPs was associated with incident CHD, the effect being independent from the traditional risk factors. This weighted CHD GRS also improved risk reclassification and but not discrimination.

Table 4. Improvement in reclassification of future CHD using a model with different genetic risk scores in addition to traditional risk factors[*] compared to a model with only traditional risk factors.

Genetic risk score		Reclassification				
		P Up	**P Down**	**NRI**	**Z-score**	**P-value**
Overall GRS	Event	0.015	0.013	0.001	0.185	0.427
	Non-event	0.013	0.017	0.003	0.912	0.181
	Total	-	-	0.005	0.568	0.285
Risk Factor GRS	Event	0.011	0.004	0.007	1.217	0.112
	Non-event	0.006	0.006	0.001	0.288	0.387
	Total	-	-	0.008	1.234	0.109
CHD GRS	Event	0.041	0.023	0.019	1.611	0.054
	Non-event	0.023	0.027	0.004	0.845	0.199
	Total	-	-	0.023	1.813	0.035
Weighted CHD GRS	Event	0.065	0.040	0.025	1.706	0.044
	Non-event	0.027	0.035	0.008	1.512	0.065
	Total	-	-	0.033	2.125	0.017
LASSO GRS	Event	0.156	0.106	0.050	2.121	0.017
	Non-event	0.047	0.078	0.031	3.955	<0.001
	Total	-	-	0.081	3.264	0.001

[*] sex, age, total cholesterol, HDL cholesterol, systolic blood pressure, body mass index, current smoking, diabetes and a parental history of MI.

NRI, net reclassification index; SNPs, single nucleotide polymorphisms; *P*, proportion; *GRS*, genetic risk score; *CHD*, coronary heart disease; *LASSO*, least absolute shrinkage and selection operator

The effect of the Overall GRS can be diluted. GWAS-identified SNPs may confer very different and difficult to estimate risks for CHD (e.g. SNPs directly associated with CHD and those associated with risk factors) and in addition not all GWAS-identified SNPs may be associated with CHD in a population-based prospective study.[15] Therefore, we used a new strategy to construct a GRS to overcome these limitations, namely a cross-validation approach in combination with LASSO regression, which can be used to achieve data selection and shrinkage at the same time.[18] The association of the 14 SNP LASSO GRS with future CHD risk was stronger than the other GRSs, and modestly but significantly improved both risk discrimination and reclassification. Of the 14 SNPs selected, three (rs10757278, rs646776 and rs11556924)[2, 4, 33] were previously reported to be associated with CHD in GWAs. Also, in both the LASSO GRS and the weighted GRS, rs10757278 was the SNP with the highest weight. Of the remaining 11 SNPs, one was previously associated with BMI (rs925946)[31], two with diabetes (rs8050136 and rs181362)[32-34] and eight with lipid levels (rs2925979, rs6882076, rs2954029, rs6897702, rs10889352, rs2972146, rs514230 and rs2000999).[9, 35] For locus and nearby genes see supplementary table s4.

Our findings for the Overall GRS based on all 179 SNPs are in line with that of a recently published large prospective cohort study.[13] In this study, a GRS composed of 101 SNPs previously associated with CVD or an intermediate phenotype was not associated with CVD after adjustment for the intermediate CHD risk factors systolic blood pressure, total cholesterol and HDL cholesterol. This GRS also did not improve discrimination or risk reclassification.[13] The non-weighted CHD GRS was not associated with CHD after adjustment for traditional risk factors in our population, which confirms findings in a recently published large prospective cohort study of women, where a GRS composed of 12 SNPs associated with CHD was no longer associated with CHD after adjusting for traditional risk factors.[13] A similar GRS, composed of 13 SNPs associated with CHD and sharing eight genetic variants with our CHD GRS, was tested in a large Finnish cohort study.[14] In contrast to our findings, this GRS was associated with CVD after adjustment for traditional risk factors. In a recently published meta-analysis of 14 case-control studies,[23] a weighted GRS of 23 SNPs associated with CHD was constructed. This GRS was highly associated with CHD, although the outcome was not adjusted for traditional risk factors. The currently tested weighted CHD GRS encompassing 29 SNPs was associated to future CHD in our prospective case-cohort also after adjustment for traditional risk factors.

From our study we can conclude the following. First, a GRS composed of common SNPs previously associated with CHD and its traditional risk factors, currently has no value in predicting incident CHD compared to measuring classical biochemical parameters and systolic blood pressure measurement at one single point in time. Second, it might be important to account for the different effect sizes of SNPs in a GRS, given the fact that the weighted CHD GRS was still associated with CHD after adjustment for traditional risk factors and the non-weighted GRS was not. These results are in line with Davies et al.[25] Third, when using the

weighted CHD GRS in addition to traditional risk factors and when accounting for the fact that using a new predictor also can deteriorate reclassification, 3.3% of the participants were placed in a more appropriate risk category, for the LASSO GRS this was 8.1%.

The SNPs we used for constructing the GRS were detected in GWA studies with large sample sizes. Our study has lower power to detect the small effect sizes, which will have contributed to our observation that most of SNPs were not significantly associated with CHD when analyzed individually. We did remove SNPs in complete LD, but not the ones in high LD, which could bias the results of the overall GRS and the risk factor GRS, depending on the regions that were over represented. For the CHD GRSs this is not an issue, since none of the included SNPs in the CHD GRSs were in very high LD with each other. Also, for constructing the LASSO GRS, highly correlated SNPs were not a problem, since penalized logistic regression methods, such as LASSO regression, only select one SNP from many correlated SNPs.[36] The LASSO GRS was not tested in an independent population and therefore we cannot rule out that the LASSO GRS performs less if tested in an independent population. On the other hand, we used a cross-validation in combination with LASSO regression approach to prevent over-fitting of the data.

In conclusion, GRSs based on CHD SNPs and the LASSO GRS are associated with future CHD independent of traditional risk factors. The weighted CHD GRS and especially the LASSO GRS improve risk reclassification.

Acknowledgements

The Monitoring Project on Cardiovascular Disease Risk Factors (PPHVZ) and the MORGEN study were financially supported by the Ministry of Public Health, Welfare and Sports of the Netherlands. The authors wish to thank D Kromhout for supervision of the PPHVZ project and the MORGEN-project steering committee consisting of (not mentioning co-authors of this article) H. B. Bueno de Mesquita, H. A. Smit, and J. C. Seidell (project leader). Furthermore the authors thank epidemiologists and field-workers of the Municipal Health Services in Maastricht, Amsterdam, and Doetinchem for their contribution to baseline data collection and those involved in the logistic management (A. Jansen and J. Steenbrink) and the data management (A. Blokstra, A. van Kessel, P. Steinberger, E. den Hoedt, I. Toxopeus, J. van der Laan). The authors further wish to thank D. Jaspers, A. Hilton, V. Visser, P. Erkens, S. Philippens, J. Bremen, B. Bleijlevens, T. van Moergastel, S. van de Crommert for assistance in clinical data collection. Statistics Netherlands is acknowledged for providing data on causes of death.

Source of funding

This study was funded by a grant from the Netherlands Heart foundation (2006B083) and the Netherlands Consortium for Healthy Ageing (050 60 810). The funding source had no involvement in the design and conduct of the study; or with the collection, management,

analysis or interpretation of the data and reporting or the decision to submit the manuscript for publication.

Disclosure: None.

References

1. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-678

2. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet DA, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, El Mokhtari NE, Schafer A, Marz W, Renner W, Bugert P, Kluter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat.Genet.* 2009;41:280-282

3. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-1493

4. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, 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, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Daly MJ, Nemesh J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burtt N, Gabriel SB, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall AS, Linsel-Nitschke P, Lieb W, Ziegler A, Konig I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein SE, Scheffold T, Berger K, Huge A, Martinelli N, Olivieri O, Corrocher R, McKeown P, Erdmann E, Konig IR, Holm H, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Do R, Xie C, Siscovick D. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat.Genet.* 2009;41:334-341

5. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443-453

6. Tregouet DA, Konig IR, Erdmann J, Munteanu A, Braund PS, Hall AS, Grosshennig A, Linsel-Nitschke P, Perret C, DeSuremain M, Meitinger T, Wright BJ, Preuss M, Balmforth AJ, Ball SG, Meisinger C, Germain C, Evans A, Arveiler D, Luc G, Ruidavets JB, Morrison C, van der Harst P, Schreiber S, Neureuther K, Schafer A, Bugert P, El Mokhtari NE, Schrezenmeir J, Stark K, Rubin D, Wichmann HE, Hengstenberg C, Ouwehand W, Ziegler A, Tiret L, Thompson JR, Cambien F, Schunkert H, Samani NJ. Genome-wide haplotype association study identifies the slc22a3-lpal2-lpa gene cluster as a risk locus for coronary artery disease. *Nat Genet*. 2009;41:283-285

7. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Kottgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FU, Smith AV, Taylor K, Scharpf RB, Hwang SJ, Sijbrands EJ, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JC, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, van Duijn CM. Genome-wide association study of blood pressure and hypertension. *Nat.Genet.* 2009;41:677-687

8. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, Papadakis K, Voight BF, Scott LJ, Zhang F, Farrall M, Tanaka T, Wallace C, Chambers JC, Khaw KT, Nilsson P, van der HP, Polidoro S, Grobbee DE, Onland-Moret NC, Bots ML, Wain LV, Elliott KS, Teumer A, Luan J, Lucas G, Kuusisto J, Burton PR, Hadley D, McArdle WL, Brown M, Dominiczak A, Newhouse SJ, Samani NJ, Webster J, Zeggini E, Beckmann JS, Bergmann S, Lim N, Song K, Vollenweider P, Waeber G, Waterworth DM, Yuan X, Groop L, Orho-Melander M, Allione A, Di GA, Guarrera S, Panico S, Ricceri F, Romanazzi V, Sacerdote C, Vineis P, Barroso I, Sandhu MS, Luben RN, Crawford GJ, Jousilahti P, Perola M, Boehnke M, Bonnycastle LL, Collins FS, Jackson AU, Mohlke KL, Stringham HM, Valle TT, Willer CJ, Bergman RN, Morken MA, Doring A, Gieger C, Illig T, Meitinger T, Org E, Pfeufer A, Wichmann HE, Kathiresan S, Marrugat J, O'Donnell CJ, Schwartz SM, Siscovick DS, Subirana I, Freimer NB, Hartikainen AL, McCarthy MI, O'Reilly PF, Peltonen L, Pouta A, de Jong PE, Snieder H, van Gilst WH, Clarke R, Goel A, Hamsten A, Peden JF, Seedorf U, Syvanen AC, Tognoni G, Lakatta EG, Sanna S, Scheet P, Schlessinger D, Scuteri A, Dorr M, Ernst F, Felix SB, Homuth G, Lorbeer R, Reffelmann T, Rettig R, Volker U, Galan P, Gut IG, Hercberg S, Lathrop GM, Zelenika D, Deloukas P, Soranzo N, Williams FM, Zhai G, Salomaa V, Laakso M, Elosua R, Forouhi NG, Volzke H, Uiterwaal CS, van der Schouw YT, Numans ME, Matullo G, Navis G, Berglund G, Bingham SA, Kooner JS, Connell JM, Bandinelli S, Ferrucci L, Watkins H, Spector TD, Tuomilehto J, Altshuler D, Strachan DP, Laan M, Meneton P, Wareham NJ, Uda M, Jarvelin MR, Mooser V, Melander O, Loos RJ, Elliott P, Abecasis GR, Caulfield M, Munroe PB. Genome-wide association study identifies eight loci associated with blood pressure. *Nat.Genet.* 2009;41:666-676

9. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T, Martin NG, Pedersen NL, Kyvik KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllensten U, Campbell H, Rudan I, Johansson A, Marroni F, Hayward C, Vitart V, Jonasson I, Pattaro C, Wright A, Hastie N, Pichler I, Hicks AA, Falchi M, Willemsen G, Hottenga JJ, de Geus EJ, Montgomery GW, Whitfield J, Magnusson P, Saharinen J, Perola M, Silander K, Isaacs A, Sijbrands EJ, Uitterlinden AG, Witteman JC, Oostra BA, Elliott P, Ruokonen A, Sabatti C, Gieger C, Meitinger T, Kronenberg F, Doring A, Wichmann HE, Smit JH, McCarthy MI, van Duijn CM, Peltonen L. Loci influencing lipid levels and coronary heart disease risk in 16 european population cohorts. *Nat Genet*. 2009;41:47-55

10. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ,

Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat.Genet.* 2009;41:56-65

11. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, Hunter DJ, Hu FB. Joint effects of common genetic variants on the risk for type 2 diabetes in u.S. Men and women of european ancestry. *Ann Intern Med*. 2009;150:541-550

12. van Heemst D, Beekman M, Mooijaart SP, Heijmans BT, Brandt BW, Zwaan BJ, Slagboom PE, Westendorp RG. Reduced insulin/igf-1 signalling and human longevity. *Aging Cell*. 2005;4:79-85

13. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, Ridker PM. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA*. 2010;303:631-637

14. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki ML, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S. A multilocus genetic risk score for coronary heart disease: Case-control and prospective cohort analyses. *Lancet*. 2010;376:1393-1400

15. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: Methodological considerations for future research. *Genome Med*. 2009;1:20

16. Tibshirani R. Regression shrinkage and selection via the lasso. *J.R.Statist.Soc.B.* 1996;58:267-288

17. Tibshirani R. The lasso method for variable selection in the cox model. *Stat.Med.* 1997;16:385-395

18. Goeman JJ. L1 penalized estimation in the cox proportional hazards model. *Biom.J.* 2010;52:70-84

19. Merry AH, Boer JM, Schouten LJ, Feskens EJ, Verschuren WM, Gorgels AP, van den Brandt PA. Validity of coronary heart diseases and heart failure based on hospital discharge and mortality data in the netherlands using the cardiovascular registry maastricht cohort study. *Eur J Epidemiol*. 2009;24:237-247

20. Kattermann R, Jaworek D, Moller G, Assmann G, Bjorkhem I, Svensson L, Borner K, Boerma G, Leijnse B, Desager JP. Multicentre study of a new enzymatic method of cholesterol determination. *J.Clin.Chem.Clin.Biochem.* 1984;22:245-251

21. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106:9362-9367

22. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16:1215

23. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA,

Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoep JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333-338

24. A genome-wide association study in europeans and south asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43:339-344

25. Davies RW, Dandona S, Stewart AF, Chen L, Ellis SG, Tang WH, Hazen SL, Roberts R, McPherson R, Wells GA. Improved prediction of cardiovascular disease based on a panel of snps identified through genome wide association studies. *Circ.Cardiovasc.Genet.* 2010;3:468-474

26. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. 1986;73:1-11

27. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*. 1988;44:837-845

28. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: From area under the roc curve to reclassification and beyond. *Stat Med*. 2008;27:157-172; discussion 207-112

29. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30:11-21

30. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De BG, De BD, Ducimetiere P, Jousilahti P, Keil U, Njolstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM. Estimation of ten-year risk of fatal cardiovascular disease in europe: The score project. *Eur.Heart J.* 2003;24:987-1003

31. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, Styrkarsdottir U, Gretarsdottir S, Thorlacius S, Jonsdottir I, Jonsdottir T, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Jonsson F, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Lauritzen T, Aben KK, Verbeek AL, Roeleveld N, Kampman E, Yanek LR, Becker LC, Tryggvadottir L, Rafnar T, Becker DM, Gulcher J, Kiemeney LA, Pedersen O, Kong A, Thorsteinsdottir U, Stefansson K. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat.Genet.* 2009;41:18-24

32. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-

wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*. 2007;316:1341-1345

33. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. Replication of genome-wide association signals in uk samples reveals risk loci for type 2 diabetes. *Science*. 2007;316:1336-1341

34. Timpson NJ, Lindgren CM, Weedon MN, Randall J, Ouwehand WH, Strachan DP, Rayner NW, Walker M, Hitman GA, Doney AS, Palmer CN, Morris AD, Hattersley AT, Zeggini E, Frayling TM, McCarthy MI. Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes*. 2009;58:505-510

35. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, 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 JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, Konig IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA, Jr., Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713

36. Ayers KL, Cordell HJ. Snp selection in genome-wide and candidate gene studies via penalized logistic regression. *Genet Epidemiol*. 2010;34:879-891

Table s1. Quality control for the selected SNPs (based on association between CHD and intermediate risk factors from previously published GWAS) genotyped in the CAREMA cohort study.

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
SNPs selected for their association with coronary heart disease								
RS46522	17q21.32	46988597	UBE2Z, GIP, ATP5G1, SNF8	CAD[1]		1.00	0.160	Yes
RS216172	17p13.3	2126504	SMG6, SRR	CAD[1]		1.00	0.843	Yes
RS579459	9q34.2	136154168	ABO	CAD[1]		0.99	0.412	Yes
RS599839	1p13.3	109822166	CELSR2,PSRC1	CD[2], CAD[1]	Abnormal clusterplot	0.00		No
RS646776	1p13.3	109818530	CELSR2,PSRC1,SORT1	MI(early onset)[3], CHD[4]		0.99	0.687	Yes
RS964184	11q23.3	116648917	APOA1	CAD[1], TG[5]		0.99	0.708	Yes
RS974819	11q22.3	103660567	PDGFD	CAD[1]		1.00	0.638	Yes
RS1122608	19p13.2	11163601	LDLR	MI(early onset)[3]		0.99	0.352	Yes
RS1333049	9p21.3	22125503	CDKN2A,CDKN2B	CD	In LD with rs10757278			No
RS1412444	10q23.31	91002927	LIPA	CAD[6]		0.99	0.130	Yes
RS1746048	10q11.21	44775824	CXCL12	MI(early onset)[3]		0.96	0.328	Yes
RS2505083	10p11.23	30335122	KIAA1462	CAD[6]		0.99	0.852	Yes
RS2895811	14q32.2	100133942	HHIPL1	CAD[1]		1.00	0.889	Yes
RS3825807	15q25.1	79089111	ADAMTS7	CAD[1]		1.00	0.405	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS4380028	15q25.1	79111093	ADAMTS7, MORF4L1	CAD[6]		1.00	0.156	Yes
RS4773144	13q34	110960712	COL4A1, COL4A2	CAD[1]		1.00	0.555	Yes
RS4977574	9p21.3	22098574	CDKN2A, CDKN2B	MI(early onset)[3], CAD[1]	In LD with rs10757278			No
RS6725887	2q33.2	203745885	WDR12	MI(early onset)[3], CAD[1]		0.95	0.290	Yes
RS6922269	6q25.1	151252985	MTHFD1L	CD[6]		0.97	0.347	Yes
RS9818870	3q22.3	138122122	MRAS	CAD[7]		0.98	0.740	Yes
RS9982601	21q22.11	35599128	SLC5A3, MRPS6, KCNE2	MI(early onset)[3], CAD[1]		0.97	0.238	Yes
RS10757278	9p21.3	22124477	CDKN2A,CDKN2B	MI[8]		0.96	0.614	Yes
RS10953541	7q22.3	107244545	-	CAD[6]		1.00	0.503	Yes
RS11206510	1p32.3	55496039	PCSK9	MI(early onset)[3], LDL-C[9, 10]		0.99	0.155	Yes
RS11556924	7q32.2	129663496	ZC3HC1	CAD[1]		1.00	0.672	Yes
RS12190287	6q23.2	134214525	TCF21	CAD[1]		1.00	0.144	Yes
RS12413409	10q24.32	104719096	CYP17A1, CNNM2, NT5C2	CAD[1]		1.00	0.377	Yes
RS12526453	6p24.1	12927544	PHACTR1	MI(early onset)[3], CAD[1]		0.98	0.142	Yes
RS12936587	17p11.2	17543722	RASD1, SMCR3, PEMT	CAD[1]		1.00	0.681	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS17114036	1p32.2	56962821	PPAP2B	CAD[1]		0.99	0.119	Yes
RS17465637	1q41	222823529	MIA3	MI(early onset)[3], CAD[1]		0.99	0.210	Yes
RS17609940	6p21.31	35034800	ANKS1A	CAD[1]		0.99	0.298	Yes
SNPs selected for their association with blood pressure								
RS381815	11p15.1	16902268	PLEKHA7	SBP[11]		0.99	0.267	Yes
RS653178	12q24.12	112007756	ATXN2, SH2B3	DBP[12]	In LD with rs3184504			
RS1004467	10q24.32	104594507	CYP17A1	SBP[11]		0.99	0.680	Yes
RS1378942	15q24.1	75077367	CYP1A1, CYP1A2, CSK, LMAN1L, CPLX3, ARID3	DBP[12]		1.00	0.643	Yes
RS1530440	10q21.2	63524591	c10orf107, TMEM26, RTKN2, RHOBTB1, ARID5B	DBP[12]		1.00	0.705	Yes
RS2384550	12q24.21	115352731	TBX3, TBX5	DBP[11]	Did not fit in i-plex			No
RS2681472	12q21.33	90008959	ATP2B1	DBP[11]		1.00	0.670	Yes
RS2681492	12q21.33	90013089	ATP2B1	SBP[11]	Did not fit in i-plex			No
RS3184504	12q24.12	111884608	SH2B3	DBP[11]		1.00	0.653	Yes
RS6495122	15q24.1	75125645	CSK, ULK3	DBP[11]	Did not fit in			No

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
					i-plex			
RS9815354	3p22.1	41912651	ULK4	DBP[11]		1.00	1.000	Yes
RS11014166	10p12.31	18708798	CACNB2	DBP[11]		0.99	0.687	Yes
RS11191548	10q24.32	104846178	CYP17A1, AS3MT, CNNM2, NT5C2	SBP[12]		1.00	0.304	Yes
RS12946454	17q21.31	43208121	PLCD3, ACBD4, HEXIM1, HEXIM2	SBP[12]		1.00	0.861	Yes
RS16948048	17q21.33	47440466	ZNF652, PHB	DBP[12]		1.00	0.568	Yes
RS16998073	4q21.21	81184341	FGF5, PRDM8, c4orf22	DBP[12]		1.00	0.704	Yes
RS17367504	1p36.22	11862778	MTHFR, NPPA, CLCN6, NPPB, AGTRAP	SBP[12]		1.00	0.590	Yes
SNPs selected for their association with diabetes type II								
RS5215	11p15.1	17408630	KCNJ11	Type 2 diabetes[13]		0.97	0.807	Yes
RS5219	11p15.1	17409572	KCNJ11	Type 2 diabetes[14, 15]		0.94	0.616	Yes
RS864745	7p15.1	28180556	JAZF1	Type 2 diabetes[16]		0.97	0.201	Yes
RS1111875	10q23.33	94462882	HHEX	Type 2 diabetes[14, 17]		0.97	0.157	Yes
RS2237892	11p15.4	2839751	KCNQ1	Type 2 diabetes[18]		0.96	0.415	Yes
RS2237897	11p15.4	2858546	KCNQ1	Type 2 diabetes[19]		0.97	0.013	Yes
RS2383208	9p21.3	22132076	CDKN2A, CDKN2B	Type 2 diabetes[20]	In ld with			No

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
					rs10811661			
RS2943641	2q36.3	227093745	LOC64673, IRS1	Type 2 diabetes[21]	Did not fit in i-plex			No
RS4402960	3q27.2	185511687	IGF2BP2	Type 2 diabetes[13, 14, 17]		0.99	0.917	Yes
RS4506565	10q25.2	114756041	TCF7L2	Type 2 diabetes[22]		0.97	0.032	Yes
RS4607103	3p14.1	64711904	ADAMTS9	Type 2 diabetes[16]		0.98	0.334	Yes
RS4689388	4p16.1	6270056	WFS1, PPP2R2C	Type 2 diabetes[21]		1.00	0.710	Yes
RS4712523	6p22.3	20657564	CDKAL1	Type 2 diabetes[20, 21]	In ld with rs10946398			No
RS4712524	6p22.3	20657865	CDKAL1	Type 2 diabetes[19]		0.96	0.313	Yes
RS6769511	3q27.2	185530290	IGF2BP2	Type 2 diabetes[19]				
RS6931514	6p22.3	20703952	CDKAL1	Type 2 diabetes[16]		0.96	0.908	Yes
RS7578597	2p21	43732823	THADA	Type 2 diabetes[16]		0.98	0.820	Yes
RS7754840	6p22.3	20661034	CDKAL1	Type 2 diabetes[14, 17]	In ld with rs10946398			No
RS7756992	6p22.3	20679709	CDKAL1	Type 2 diabetes[23]	In ld with rs6931514			No
RS7901695	10q25.2	114754088	TCF7L2	Type 2 diabetes[13]	Did not fit in i-plex			No
RS7903146	10q25.2	114758349	TCF7L2	Type 2 diabetes[14-17, 20, 21,	Did not fit in i-plex			No

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
				[23-25]				
RS7961581	12q21.1	71663102	TSPAN8,LGR5	Type 2 diabetes[16]		0.96	0.274	Yes
RS8050136	16q12.2	53816275	FTO	Type 2 diabetes[13-15]		0.96	0.002	Yes
RS10811661	9p21.3	22134094	CDKN2A,CDKN2B	Type 2 diabetes[13, 14, 17]		0.98	0.826	Yes
RS10923931	1p12	120517959	NOTCH2, ADAM30	Type 2 diabetes[16]		0.96	0.217	Yes
RS10946398	6p22.3	20661034	CDKAL1	Type 2 diabetes[13]		0.98	0.598	Yes
RS12779790	10p13	12328010	CDC123,CAMK1D	Type 2 diabetes[16]		0.95	1.000	Yes
RS13266634	8q24.11	118184783	SLC30A8	Type 2 diabetes[13, 14, 17]		0.98	0.457	Yes
SNPs selected for their association with blood lipid levels								
RS328	8p21.3	19819724	LPL	TG[26], HDL-C[26]		0.96	0.027	Yes
RS693	2p24.1	21232195	APOB	TC[4], LDL-C[4, 26, 27]		0.98	0.468	Yes
RS7679	20q13.12	44576502	PLTP	TG[10]		0.95	0.04	Yes
RS12916	5q13.3	74692295	HMGCR	TC[5]		0.99	0.927	Yes
RS157580	19q13.32	45395266	APO cluster	LDL-C[4]		0.97	0.884	Yes
RS171457	7q11.23	142827855	BCL7B, TBL2, MLXIPL	TG[9, 26]	In ld with RS2240466			No
RS173539	16q13	56988044	CETP	HDL-C[10]	No PCR product	0.00		No

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS174546	11q12.2	61326406	FADS1-2-3	TG[5]	In ld with RS174547			No
RS174547	11q12.2	61570783	FADS1, FADS2, FADS3	TG[10]		0.98	0.608	Yes
RS174570	11q12.2	61597212	FADS2, FADS3	LDL-C[4], TC[4]		0.94	0.003	Yes
RS181362	22q11.21	20262068	UBE2L3	HDL-C[27]		0.99	0.14	Yes
RS255049	16q22.1	68013471	LCAT	HDL-C[27]	Did not fit in i-plex			No
RS386000	19q13.42	59484573	LILRA3	HDL-C[5]	In ld with rs3905000			No
RS439401	19q13.32	50106291	APOE	TG[4]		0.96	0.076	Yes
RS442177	4q22.1	88249285	KLHL8	TG[5]		0.99	0.225	Yes
RS471364	9p22.3	15289578	TTC39B	HDL-C[10]		0.96	0.375	Yes
RS492602	19q13.33	53898229	FLJ36070	TC[5]		0.99	0.437	Yes
RS514230	1q42.3	232925220	IRF2BP2	TC[5]		0.99	0.752	Yes
RS515135	2p24.1	21286057	APOB	LDL-C[10]	Abnormal clusterplot	0.00		No
RS562338	2p24.1	21288321	APOB	LDL-C[9, 28]	In LD with rs515135			No
RS581080	9p22.3	15295378	TTC39B	HDL[5]		0.99	0.461	Yes
RS599839	1p13.3	109822166	CELSR2.PSRC1	LDL-C[2, 9, 28]	No PCR product	0.00		No

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS605066	6q24.1	139871359	CITED2	HDL-C[5]	Out of HWE	0.96	<0.001	No
RS629301	1p13.3	109619829	SORT1	LDL-C[5]		0.99	0.787	Yes
RS645040	3q22.3	137409312	MSL2L1	TG[5]		0.99	0.786	Yes
RS646776	1p13.3	109818530	CELSR2.PSRC1.SORT1	LDL-C[4, 26, 27]		0.99	0.687	Yes
RS714052	7q11.23	72864869	MLXIPL	TG[10]	In ld with RS2240466			No
RS737337	19p13.2	11208493	LOC55908	HDL-C[5]		0.96	0.456	Yes
RS780094	2p23.3	27741237	GCKR	TG[4, 9, 26]		0.97	0.451	Yes
RS838880	12q24.31	123827546	SCARB1	HDL-C[5]		1.00	0.367	Yes
RS964184	11q23.3	116154127	APOA1	TG[10]		0.99	0.708	Yes
RS1042034	2p24.1	21078786	APOB	TG[5]		1.00	0.697	Yes
RS1084651 (RS1652507)	6q26	161009807	LPA	HDL-C[5]		1.00	0.787	Yes
RS1167998	1p31.3	62931632	DOCK7	TG[4]		0.98	0.174	Yes
RS1169288	12q24.31	119901033	HNF1A	TC[5]	Did not fit in i-plex			No
RS1260326	2p23.3	27584444	GCKR	TG[10]		0.97	0.457	Yes
RS1367117	2p24.1	21117405	APOB	LDL-C[5]	Out of HWE	0.80	<0.001	No
RS1495741	8p22	18317161	NAT2	TG[5]		1.00	0.080	Yes
RS1501908	5q33.3	156398169	TIMD4, HAVCR1	LDL-C[10]		0.97	0.961	Yes
RS1532085	15q21.3	56470658	LIPC	HDL-C[4]		0.99	1.000	Yes
RS1532624	16q13	57005479	CETP	HDL-C[4]		0.97	0.710	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS1564348	6q25.3	160498850	LPA	LDL-C[5]		0.98	0.086	Yes
RS1689800	1q25.3	180435508	ZNF648	HDL-C[5]		0.99	0.076	Yes
RS1748195	1p31.3	63049593	ANGPTL3	TG[9]	In ld with RS12970134			No
RS1800562	6p22.2	26201120	HFE	LDL-C[5]		1.00	0.571	Yes
RS1800588	15q21.3	58723675	LIPC	HDL-C[26]		0.99	0.848	Yes
RS1800775	16q13	55552737	CETP. NUP93. SLC12A3. HERPUD1	TG[17], HDL-C[26, 29]	Did not fit in i-plex			No
RS1800961	20q13.12	42475778	HNF4A	HDL-C[5, 10]		0.97	0.388	Yes
RS1864163	16q13	56997233	CETP	HDL-C[9]	Did not fit in i-plex			No
RS1883025	9q31.1	106704122	ABCA1	HDL-C[5, 10]		0.96	0.216	Yes
RS2000999	16q22.2	70665594	HPR	TC[5]		1.00	1.000	Yes
RS2068888	10q23.33	94829632	CYP26A1	TG[5]		0.99	0.388	Yes
RS2072183	7p13	44545705	NPC1L1	TC[5]		0.99	0.569	Yes
RS2075650	19q13.32	45395619	TOMM40. APOE	TC[4]		0.97	0.072	Yes
RS2081687	8q12.1	59551119	CYP7A1	TC[5]		1.00	0.374	Yes
RS2083637	8p21.3	19865175	LPL	HDL-C[4]		0.98	0.200	Yes
RS2131925	1p31.3	62798530	ANGPTL3	TG[5]	In ld with RS1167998			No
RS2144300	1q42.13	230294916	GALNT2	HDL-C[9]	Did not fit in i-plex			No

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS2156552	18q21.1	47181668	LIPG	HDL-C[9]	Did not fit in i-plex			No
RS2228671	19p13.2	11210912	LDLR	TC[4], LDL-C[4]		0.99	0.383	Yes
RS2240466	7q11.23	72856269	MLXIPL	TG[4]		0.98	1.000	Yes
RS2247056	6p21.33	31373469	HLA	TG[5]	No PCR product	0.00		No
RS2254287	6p21.32	33143948	B3GALT4	LDL-C[9]	Did not fit in i-plex			No
RS2255141	10q25.2	113923876	GPAM	TC[5]		1.00	0.045	Yes
RS2271293	16q22.1	67902070	CTCF. PRMT8	HDL-C[4, 10]		0.95	0.766	Yes
RS2277862	20q11.22	33616196	ERGIC3	TC[5]		1.00	0.013	Yes
RS2290159	3p25.2	12603920	RAF1	TC[5]		0.99	0.397	Yes
RS2293889	8q23.3	116668374	TRPS1	HDL-C[5]		0.99	0.708	Yes
RS2304130	19p13.11	19789528	NCAN	TC[4]		0.99	0.393	Yes
RS2338104	12q24.11	109895168	MMAB.MVK	HDL-C[9, 10]		0.96	0.130	Yes
RS2412710	15q15.1	40471079	CAPN3	TG[5]		1.00	0.624	Yes
RS2479409	1p32.3	55277238	PCSK9	LDL-C[5]		0.99	0.056	Yes
RS2642442	1q41	219040186	MOSC1	TC[5]	Out of HWE	0.93	<0.001	No
RS2650000	12q24.31	121388962	LEF1	LDL-C[10, 27]		0.97	0.032	Yes
RS2652834	15q22.2	61183920	LACTB	HDL[5]		0.99	0.362	Yes
RS2737229	8q23.3	116717740	TRPS1	TC[5]		1.00	0.759	Yes
RS2814944	6p21.31	34660775	C6orf106	HDL-C[5]		0.99	0.930	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS2814982	6p21.31	34654538	C6orf106	TC[5]		1.00	0.394	Yes
RS2902940	20q12	38524901	MAFB	TC[5]		0.99	0.069	Yes
RS2923084	11p15.4	10345358	AMPD3	HDL-C[5]		0.99	0.610	Yes
RS2925979	16q23.2	80092291	CMIP	HDL-C[5]		0.99	0.045	Yes
RS2929282	15q15.3	42033223	FRMD5	TG[5]		1.00	0.241	Yes
RS2954029	8q24.13	126560154	TRIB1	TG[10]		0.98	0.159	Yes
RS2967605	19p13.2	8469738	ANGPTL4	HDL-C[10]	Out of HWE	0.93	<0.001	No
RS2972146	2q36.3	226808942	IRS1	HDL-C[5]		1.00	0.155	Yes
RS3136441	11p11.2	46699823	LRP4	HDL-C[5]	Did not fit in i-plex			No
RS3177928	6p21.32	32520413	HLA	TC[5]	Out of HWE	0.99	<0.001	No
RS3757354	6p22.3	16235386	MYLIP	LDL-C[5]		1.00	0.839	Yes
RS3764261	16q13	55550825	CETP	HDL-C[5]		0.99	1.000	Yes
RS3846662	5q13.3	74651084	HMGCR	TC[4]	Success-rate to low	0.86		No
RS3846663	5q13.3	74655726	HMGCR	LDL-C[10] HDL-C[26]	in LD with rs3846662			No
RS3890182	9q31.1	107647655	ABCA1		in LD with rs3905000			No
RS3905000	9q31.1	107657070	ABCA1	HDL-C[4]		0.95	0.007	Yes
RS4129767	17q25.3	73915579	PGS1	HDL-C[5]		0.99	0.499	Yes
RS4148008	17q24.2	64386889	ABCA8	HDL-C[5]		0.99	0.310	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
				HDL-C[9]	Did not fit in			
RS4149268	9q31.1	107647220	ABCA1		i-plex			No
RS4299376	2p21	43926080	ABCG5/8	LDL-C[5]		0.97	0.870	Yes
RS4420638	19q13.32	50114786	APOE	LDL-C[10]	Out of HWE	0.95	<0.001	No
RS4660293	1p34.3	39800767	PABPC4	HDL-C[5]		0.98	0.903	Yes
RS4731702	7q32.3	130083924	KLF14	HDL-C[5]		0.98	0.084	Yes
				HDL-C[5]	No primer design			
RS4759375	12q24.31	122362191	SBNO1		possible			No
RS4765127	12q24.31	123026120	ZNF664	HDL-C[5]		0.99	0.077	Yes
				HDL-C[9], TG[9]	Did not fit in			
RS4775041	15q22.1	58674695	LIPC		i-plex			No
RS4846914	1q42.13	228362314	GALNT2	HDL-C[10]		0.95	0.777	Yes
RS4939883	18q21.1	47167214	LIPG	HDL-C[4, 10]		0.93	0.495	Yes
					Success-rate			
RS5756931	22q13.1	36875979	PLA2G6	TG[5]	to low	0.89		No
RS6029526	20q12	39106032	TOP1	LDL-C[5]		0.99	0.238	Yes
				HDL-C[5]	No PCR			
RS6065906	20q13.12	43987422	PLTP		product	0.00		No
					No primer design			
RS6102059	20q12	39228784	MAFB	LDL-C[10]	possible			No

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS6450176	5q11.2	53333782	ARL15	HDL-C[5]		0.99	0.081	Yes
RS6511720	19p13.2	11063306	LDLR	LDL-C[10]		0.98	0.017	Yes
RS6544713	2p21	44073881	ABCG8	LDL-C[10]		0.97	0.870	Yes
RS6754295	2p24.1	21206183	APOB	TG[4]		0.97	0.621	Yes
RS6756629	2p21	44065090	ABCG5	TC[4]		0.99	0.345	Yes
RS6882076	5q33.3	156322875	TIMD4	TC[5]		1.00	0.846	Yes
RS6987702	8q24.13	126504726	TRIB1	TC[4]		0.99	1.000	Yes
RS7134375	12p12.2	20365025	PDE3A	HDL-C[5]		1.00	0.116	Yes
RS7134594	12q24.11	108484576	MVK	HDL-C[5]	Did not fit in i-plex			No
RS7206971	17q21.32	42780114	OSBPL7	LDL-C[5]	Did not fit in i-plex			No
RS7240405	18q21.1	45413088	LIPG	HDL-C[30]	In LD with RS4939883			No
RS7241918	18q21.1	45414951	LIPG	HDL-C[5]	In LD with RS4939883			No
RS7255436	19p13.2	8339196	ANGPTL4	HDL-C[5]		0.99	0.022	Yes
RS7395662	11p11.2	48518893	MADD. FOLH1	HDL-C[4]		0.98	0.061	Yes
RS7515577	1p22.1	92782026	EVI5	TC[5]		1.00	0.750	Yes
RS7557067	2p24.1	21208211	APOB	TG[10]		0.97	0.667	Yes
RS7570971	2q21.3	135554376	RAB3GAP1	TC[5]		1.00	0.011	Yes
RS7819412	8p23.1	11045161	XKR6. AMAC1L2	TG[10]		0.97	0.964	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS7941030	11q24.1	122027585	UBASH3B	TC[5]		1.00	0.474	Yes
RS8017377	14q12	23953727	NYNRIN	LDL-C[5]		1.00	0.892	Yes
RS9411489	9q34.2	133184804	ABO	LDL-C[5]	Did not fit in i-plex			No
RS9488822	6q22.1	116419586	FRK	TC[5]		1.00	0.722	Yes
RS9686661	5q11.2	55897543	MAP3K1	TG[5]		1.00	0.828	Yes
RS9987289	8p23.1	9220768	PPP1R3B	HDL-C[5]		0.99	0.337	Yes
RS9989419	16q13	56985139	CETP	HDL-C[9, 30]	Did not fit in i-plex			No
RS10096633	8p21.3	19830921	LPL	TG[4, 27]	Did not fit in i-plex			No
RS10128711	11p15.1	18589560	SPTY2D1	TC[5]	Did not fit in i-plex			No
RS10195252	2q24.3	165221337	COBLL1	TG[5]	Abnormal cluster plot			No
RS10401969	19p13.11	19268718	CILP2	TC[10], LDL-C[10]		0.99	1.000	Yes
RS10468017	15q21.3	58678512	LIPC	HDL-C[10]		0.98	0.139	Yes
RS10503669	8p21.3	19847690	LPL	HDL-C[9], TG[9]	In LD with rs328			No
RS10761731	10q21.2	64697616	JMJD1C	TG[5]	Did not fit in i-plex			No
RS10889353	1p31.1	63118196	DOCK7	TC[4]		0.99	0.155	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS10903129	1p36.11	25768937	TMEM57	TC[4]		0.96	0.104	Yes
RS11065987	12q24.12	110556807	BRAP	TC[5]	Did not fit in i-plex			No
RS11136341	8q24.3	145115531	PLEC1	LDL-C[5]	Did not fit in i-plex			No
RS11206510	1p32.3	55496039	PCSK9	LDL-C[9, 10], MI (early onset)[3]		0.99	0.155	Yes
RS11220462	11q24.2	125749162	ST3GAL4	LDL-C[5]	Did not fit in i-plex			No
RS11869286	17q12	37813856	STARD3	HDL-C[5]	Did not fit in i-plex			No
RS12272004	11q23.3	116603724	APOA1, APOA4, APOA5, APOC3	LDL-C[4], TG[4]				Yes
RS12670798	7p15.3	21607352	DNAH11	TC[5], LDL-C[4]				Yes
RS12678919	8p21.3	19844222	LPL	TG[5, 10], HDL-C[10], LDL-C[10]				Yes
RS12740374	1p13.3	109817590	CELSR2, PSRC1, SORT1	LDL-C[10]				Yes
SNPs selected for their association with anthropomorphic related traits								
RS6265	11p14.1	27679916	BDNF	BMI[31]		0.94	0.888	Yes
RS29941 (rs29939)	19q13.11	39001372	KCTD15, CHST8	Weight[31]		0.96	0.873	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS925946	11p14.1	27667202	BDNF	BMI[31]		1.00	0.003	Yes
RS987237	6p12.3	50803050	TFAP2B	Adiposity[32]		1.00	1.000	Yes
RS1121980	16q12.2	52366748	FTO	BMI[33]	No PCR product	0.00		No
RS1558902	16q12.2	53803574	FTO	WC[34]	Out of HWE	1.00	<0.001	No
RS2568958	1p31.1	72765116	NEGR1	Weight[31]		0.97	0.962	Yes
RS2605100	1q41	219644224	LYPLAL1	Adiposity[32]		0.99	0.017	Yes
RS2844479	6p21.33	31572956	AIF1, NCR3	Weight[31]	No PCR product	0.00		No
RS3764261	16q13	56993324	CETP	WC[35]		0.99	1.000	Yes
RS6499640	16q12.2	53769677	FTO	BMI[31]		0.96	0.401	Yes
RS6548238	2p25.3	634905	TMEM18	BMI[36]		0.96	0.422	Yes
RS7498665	16p11.2	28883241	SH2B1	BMI[31, 36]		0.97	0.094	Yes
RS7561317	2p25.3	644953	TMEM18	BMI[31]		0.93	0.466	Yes
RS7647305	3q27.2	185834290	SFRS10, ETV5, DGKG	BMI[31]		0.95	0.421	Yes
RS7826222	8q23.1	110184852	TRHR	WC[32]	No PCR product	0.00		No
RS9930506	16q12.2	52387966	FTO	BMI[15, 31]	In LD with rs8050136			No
RS10146997	14q31.1	79945162	NRXN3	WC[34]		1.00	0.385	Yes
RS10838738	11p11.2	47663049	MTCH2	BMI[36]		0.95	0.313	Yes
RS10913469	1q25.2	177913519	SEC16B, RASAL2	Weight[31]		0.93	0.049	Yes

Selected SNP (proxy SNP)	Locus	BP position	Candidate genes	Trait	Reason not (successfully) genotyped	Succes rate	P-value HWE	Succesfully genotyped
RS10938397	4p12	45182527	GNPDA2	BMI[36]		0.95	0.005	Yes
RS11084753	19q13.11	34322137	KCTD15	BMI[36]		0.94	0.874	Yes
RS12970134	18q21.32	57884750	MC4R	WC[31]		0.96	0.952	Yes
RS17782313	18q21.32	56002077	MC4R	Obesity[33, 36]	In LD with rs17782313			No

Table s2. Weights for the SNPs selected for their association with CHD. The weights are based on the log odds ratio as published in Schunker et al. 2011 or Peden et al 2011.

		Subcohort		Cases				
SNP	**RA**	**N**	**RAF**	**N**	**RAF**	**OR**	**Weighting factor***	***Source***
RS10757278[1]	G	1832	0.49	623	0.55	1.29	1.00	Schunkert *et al.* (2011)
RS10953541	C	1902	0.76	641	0.75	1.08	0.30	Peden *et al.* (2011)
RS11206510	T	1881	0.81	632	0.81	1.08	0.30	Schunkert *et al.* (2011)
RS1122608	G	1887	0.76	640	0.76	1.14	0.51	Schunkert *et al.* (2011)
RS11556924	C	1902	0.60	641	0.64	1.09	0.34	Schunkert *et al.* (2011)
RS12190287	C	1901	0.64	641	0.65	1.08	0.30	Schunkert *et al.* (2011)
RS12413409	G	1901	0.92	641	0.90	1.12	0.45	Schunkert *et al.* (2011)
RS12526453	C	1865	0.66	616	0.65	1.10	0.37	Schunkert *et al.* (2011)
RS12936587	G	1900	0.56	641	0.58	1.07	0.27	Schunkert *et al.* (2011)
RS1412444	A	1900	0.33	639	0.35	1.09	0.34	Peden *et al.* (2011)
RS17114036	A	1890	0.91	635	0.92	1.17	0.62	Schunkert *et al.* (2011)
RS1746048	C	1838	0.86	622	0.87	1.09	0.34	Schunkert *et al.* (2011)
RS17465637	C	1880	0.73	636	0.75	1.14	0.51	Schunkert *et al.* (2011)
RS17609940	G	1897	0.81	637	0.81	1.07	0.27	Schunkert *et al.* (2011)
RS216172	G	1898	0.35	637	0.35	1.07	0.27	Schunkert *et al.* (2011)
RS2505083[2]	C	1897	0.41	637	0.43	1.05	0.19	Schunkert *et al.* (2011)
RS2895811	C	1901	0.41	641	0.41	1.07	0.27	Schunkert *et al.* (2011)

RS3825807	T	1901	0.58	641	0.57	1.08	0.30	Schunkert *et al.* (2011)
RS4380028	G	1902	0.61	641	0.6	1.07	0.27	Peden *et al.* (2011)
RS46522	T	1901	0.54	642	0.55	1.06	0.23	Schunkert *et al.* (2011)
RS4773144	G	1901	0.45	642	0.46	1.07	0.27	Schunkert *et al.* (2011)
RS579459	C	1890	0.23	633	0.23	1.10	0.37	Schunkert *et al.* (2011)
RS646776	A	1887	0.79	641	0.82	1.11	0.41	Schunkert *et al.* (2011)
RS6725887	G	1811	0.13	611	0.13	1.14	0.51	Schunkert *et al.* (2011)
RS6922269	A	1866	0.26	615	0.27	1.06	0.23	Schunkert *et al.* (2011)
RS964184	C	1880	0.14	639	0.14	1.13	0.48	Schunkert *et al.* (2011)
RS974819	A	1901	0.31	642	0.31	1.07	0.27	Peden *et al.* (2011)
RS9818870	T	1869	0.17	638	0.16	1.12	0.45	Schunkert *et al.* (2011)
RS9982601	T	1859	0.13	626	0.15	1.18	0.65	Schunkert *et al.* (2011)

[1] Instead of rs10757278, rs4977574 was selected for replication in Schunkert et al.

[2] Instead of rs2505083, rs7920682 was selected for replication in Schunkert et al.

[*] The weighting factor is determined by dividing the log odds ratio of a SNP by the log odds ratio of rs10757278 (the SNP with the highest log odds ratio).

Table s3. . The association between the 179 successfully genotyped SNPs and incident CHD in the CAREMA study.

		Subcohort		Cases		CHD		CHD adjusted*	
SNP	**RA**	**N**	**RAF**	**N**	**RAF**	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
SNPs selected for their association with coronary artery diseases									
RS10757278	G	1832	0.49	623	0.55	**1.34 (1.17-1.53)**	**<0.001**	**1.29 (1.05-1.59)**	**0.015**
RS10953541	C	1902	0.76	641	0.75	**0.98 (0.84-1.14)**	**0.779**	**0.87 (0.77-0.99)**	**0.028**
RS11206510	T	1881	0.81	632	0.81	0.99 (0.83-1.17)	0.861	1.02 (0.88-1.17)	0.840
RS1122608	G	1887	0.76	640	0.76	1.01 (0.87-1.18)	0.869	1.01 (0.89-1.15)	0.919
RS11556924	C	1902	0.60	641	0.64	1.16 (1.01-1.32)	0.033	1.23 (0.93-1.63)	0.155
RS12190287	C	1901	0.64	641	0.65	1.05 (0.92-1.21)	0.451	1.06 (0.94-1.20)	0.323
RS12413409	G	1901	0.92	641	0.90	0.88 (0.69-1.14)	0.334	0.86 (0.71-1.04)	0.124
RS12526453	C	1865	0.66	616	0.65	0.92 (0.80-1.06)	0.255	1.00 (0.89-1.13)	0.991
RS12936587	G	1900	0.56	641	0.58	1.08 (0.95-1.24)	0.247	1.04 (0.93-1.17)	0.505
RS1412444	A	1900	0.33	639	0.35	1.07 (0.93-1.24)	0.325	1.15 (0.77-1.73)	0.489
RS17114036	A	1890	0.91	635	0.92	1.06 (0.85-1.33)	0.598	0.99 (0.80-1.23)	0.924
RS1746048	C	1838	0.86	622	0.87	1.01 (0.82-1.23)	0.960	1.05 (0.88-1.25)	0.568
RS17465637	C	1880	0.73	636	0.75	1.14 (0.99-1.32)	0.064	1.04 (0.92-1.17)	0.535
RS17609940	G	1897	0.81	637	0.81	0.99 (0.83-1.17)	0.864	0.88 (0.63-1.24)	0.459
RS216172	G	1898	0.35	637	0.35	1.02 (0.88-1.17)	0.809	1.09 (0.93-1.28)	0.286
RS2505083	C	1897	0.41	637	0.43	1.07 (0.93-1.23)	0.322	1.00 (0.89-1.12)	0.942
RS2895811	C	1901	0.41	641	0.41	0.99 (0.87-1.14)	0.905	0.99 (0.89-1.11)	0.864
RS3825807	T	1901	0.58	641	0.57	1.00 (0.87-1.14)	0.956	1.04 (0.93-1.17)	0.447

RS4380028	G	1902	0.61	641	0.60	0.99 (0.87-1.13)	0.890	1.06 (0.95-1.18)	0.337
RS46522	T	1901	0.54	642	0.55	1.10 (0.96-1.26)	0.168	1.07 (0.94-1.21)	0.327
RS4773144	G	1901	0.45	642	0.46	0.99 (0.87-1.13)	0.846	0.99 (0.88-1.11)	0.818
RS579459	C	1890	0.23	633	0.23	0.97 (0.84-1.13)	0.727	0.99 (0.87-1.12)	0.821
RS646776	A	1887	0.79	641	0.82	**1.27 (1.09-1.47)**	**0.002**	**1.25 (1.07-1.45)**	**0.004**
RS6725887	G	1811	0.13	611	0.13	1.05 (0.85-1.30)	0.637	1.08 (0.90-1.29)	0.395
RS6922269	A	1866	0.26	615	0.27	1.06 (0.91-1.24)	0.431	1.07 (0.92-1.25)	0.386
RS964184	C	1880	0.14	639	0.14	0.96 (0.8-1.170)	0.704	0.80 (0.56-1.14)	0.215
RS974819	A	1901	0.31	642	0.31	1.02 (0.88-1.18)	0.812	0.96 (0.85-1.09)	0.552
RS9818870	T	1869	0.17	638	0.16	0.94 (0.79-1.11)	0.452	0.92 (0.79-1.07)	0.285
RS9982601	T	1859	0.13	626	0.15	1.19 (0.96-1.49)	0.114	1.16 (0.98-1.37)	0.092
SNPs selected for their association with blood pressure									
RS1004467	T	1894	0.91	639	0.90	0.89 (0.70-1.14)	0.347	**0.83 (0.69-0.99)**	**0.038**
RS11014166	A	1895	0.66	639	0.68	1.05 (0.91-1.20)	0.526	1.14 (1.00-1.30)	0.051
RS11191548	T	1899	0.92	641	0.91	0.92 (0.72-1.18)	0.523	0.90 (0.75-1.09)	0.282
RS12946454	T	1900	0.26	642	0.24	0.91 (0.79-1.06)	0.218	0.83 (0.68-1.03)	0.084
RS1378942	G	1899	0.32	642	0.34	1.11 (0.96-1.28)	0.163	1.00 (0.88-1.12)	0.937
RS1530440	C	1898	0.18	641	0.18	1.05 (0.88-1.25)	0.622	0.89 (0.77-1.03)	0.112
RS16948048	G	1900	0.38	642	0.39	1.06 (0.93-1.22)	0.390	1.09 (0.97-1.23)	0.132
RS16998073	T	1899	0.29	641	0.31	1.06 (0.91-1.23)	0.460	1.08 (0.95-1.22)	0.250
RS17367504	A	1897	0.15	642	0.15	1.03 (0.85-1.26)	0.735	1.15 (0.98-1.36)	0.082
RS2681472	T	1900	0.84	642	0.84	0.93 (0.77-1.13)	0.464	**0.85 (0.73-0.99)**	**0.038**
RS3184504	T	1900	0.51	642	0.50	0.98 (0.86-1.12)	0.731	0.92 (0.81-1.06)	0.241

RS381815	T	1894	0.26	639	0.28	1.16 (0.99-1.36)	0.075	**1.22 (1.08-1.38)**	**0.002**
RS9815354	A	1896	0.17	641	0.17	0.99 (0.83-1.17)	0.865	1.06 (0.92-1.22)	0.437
SNPs selected for their association with diabetes type II									
RS10811661	T	1874	0.81	622	0.80	0.94 (0.79-1.12)	0.491	0.91 (0.76-1.08)	0.284
RS10923931	A	1821	0.09	620	0.10	1.16 (0.90-1.50)	0.253	1.14 (0.81-1.59)	0.455
RS10946398	C	1861	0.31	624	0.31	1.01 (0.88-1.17)	0.867	1.06 (0.94-1.20)	0.316
RS1111875	C	1861	0.60	622	0.61	1.03 (0.89-1.18)	0.726	1.11 (0.81-1.53)	0.512
RS12779790	G	1818	0.18	613	0.19	1.04 (0.87-1.25)	0.655	1.08 (0.92-1.25)	0.350
RS13266634	C	1871	0.69	620	0.69	1.02 (0.88-1.18)	0.838	0.99 (0.87-1.12)	0.881
RS2237892	C	1831	0.94	619	0.94	0.94 (0.70-1.27)	0.703	0.80 (0.59-1.09)	0.153
RS2237897	C	1848	0.96	613	0.96	0.98 (0.69-1.39)	0.913	0.89 (0.54-1.45)	0.635
RS4402960	A	1880	0.32	638	0.34	1.12 (0.97-1.30)	0.124	1.11 (0.98-1.25)	0.109
RS4506565	T	1847	0.34	631	0.35	1.01 (0.88-1.17)	0.853	0.94 (0.72-1.22)	0.637
RS4607103	C	1870	0.75	618	0.77	1.12 (0.97-1.30)	0.126	1.09 (0.96-1.24)	0.181
RS4689388	A	1900	0.59	641	0.60	1.07 (0.94-1.23)	0.311	1.15 (0.89-1.49)	0.291
RS4712524	G	1838	0.31	617	0.32	1.04 (0.89-1.20)	0.643	1.09 (0.97-1.24)	0.152
RS5215	G	1856	0.37	627	0.36	0.93 (0.81-1.06)	0.270	1.03 (0.92-1.16)	0.639
RS5219	A	1789	0.36	608	0.36	0.94 (0.82-1.09)	0.417	1.06 (0.95-1.20)	0.305
RS6931514	G	1843	0.27	621	0.28	1.02 (0.88-1.19)	0.801	1.09 (0.96-1.23)	0.174
RS7578597	A	1869	0.89	632	0.89	1.03 (0.83-1.27)	0.806	1.05 (0.87-1.26)	0.608
RS7961581	C	1836	0.30	623	0.31	1.08 (0.93-1.26)	0.295	1.04 (0.91-1.19)	0.541
RS8050136	A	1815	0.36	622	0.40	**1.19 (1.04-1.38)**	**0.014**	**1.20 (1.07-1.34)**	**0.002**
RS864745	T	1857	0.52	631	0.50	0.91 (0.79-1.04)	0.159	0.90 (0.80-1.02)	0.090

SNPs selected for their association with blood lipid levels									
RS10401969	A	1888	0.93	641	0.92	0.85 (0.64-1.14)	0.279	0.76 (0.44-1.33)	0.338
RS1042034	A	1901	0.78	641	0.75	0.88 (0.74-1.04)	0.123	0.83 (0.73-0.94)	0.005
RS10468017	C	1863	0.72	625	0.70	0.95 (0.81-1.10)	0.460	1.00 (0.88-1.13)	0.965
RS1084651	T	1898	0.85	637	0.85	0.88 (0.72-1.08)	0.234	0.92 (0.77-1.08)	0.302
RS10889352	A	1882	0.66	637	0.68	1.13 (0.98-1.30)	0.082	1.07 (0.95-1.21)	0.249
RS10903129	G	1820	0.55	620	0.53	0.96 (0.84-1.10)	0.594	1.05 (0.93-1.17)	0.437
RS11206510	T	1881	0.81	632	0.81	0.99 (0.83-1.17)	0.861	1.02 (0.88-1.17)	0.840
RS1167998	T	1877	0.66	633	0.68	1.11 (0.97-1.28)	0.138	1.07 (0.95-1.21)	0.256
RS12272004	A	1855	0.07	606	0.06	0.85 (0.66-1.09)	0.193	0.77 (0.60-0.99)	0.043
RS1260326	T	1863	0.42	615	0.44	1.07 (0.93-1.23)	0.343	0.91 (0.72-1.15)	0.426
RS12670798	C	1888	0.23	640	0.24	1.01 (0.86-1.17)	0.949	0.91 (0.80-1.04)	0.173
RS12678919	A	1857	0.89	625	0.90	1.19 (0.96-1.46)	0.107	1.02 (0.84-1.24)	0.831
RS12740374	C	1854	0.79	626	0.82	1.26 (1.08-1.47)	0.003	1.24 (1.07-1.43)	0.004
RS12916	C	1896	0.43	632	0.41	0.96 (0.84-1.10)	0.562	0.96 (0.85-1.09)	0.528
RS1495741	G	1902	0.23	641	0.24	1.05 (0.89-1.24)	0.534	1.09 (0.95-1.24)	0.233
RS1501908	G	1845	0.63	634	0.66	1.10 (0.96-1.26)	0.194	1.11 (0.99-1.25)	0.085
RS1532085	G	1883	0.62	638	0.60	0.93 (0.81-1.06)	0.276	0.95 (0.84-1.06)	0.355
RS1532624	C	1860	0.57	620	0.57	1.00 (0.88-1.15)	0.973	0.94 (0.83-1.05)	0.270
RS1564348	T	1873	0.85	628	0.86	1.05 (0.88-1.25)	0.624	1.13 (0.97-1.32)	0.123
RS157580	A	1844	0.62	630	0.63	1.01 (0.88-1.17)	0.842	0.99 (0.88-1.12)	0.837
RS1689800	G	1893	0.37	637	0.37	1.00 (0.87-1.15)	0.998	0.93 (0.83-1.04)	0.225
RS17216525	C	1846	0.92	633	0.91	1.02 (0.80-1.29)	0.901	0.97 (0.79-1.18)	0.729

RS174547	G	1871	0.33	618	0.33	1.05 (0.90-1.21)	0.555	1.04 (0.62-1.72)	0.896
RS174570	C	1793	0.86	602	0.86	0.95 (0.78-1.16)	0.591	1.02 (0.86-1.20)	0.832
RS1800562	G	1901	0.94	641	0.94	1.03 (0.79-1.34)	0.822	0.77 (0.62-0.96)	0.020
RS1800588	T	1890	0.23	640	0.22	0.97 (0.82-1.13)	0.675	1.00 (0.87-1.16)	0.961
RS1800961	T	1857	0.03	621	0.03	1.15 (0.72-1.85)	0.558	0.80 (0.30-2.14)	0.649
RS181362	T	1890	0.21	634	0.24	1.17 (0.99-1.40)	0.071	1.24 (1.09-1.41)	0.001
RS1883025	A	1836	0.25	609	0.25	1.02 (0.88-1.20)	0.766	0.91 (0.79-1.03)	0.140
RS2000999	A	1901	0.20	642	0.24	1.32 (1.08-1.60)	0.006	1.21 (1.02-1.43)	0.027
RS2068888	G	1896	0.55	637	0.54	0.96 (0.84-1.10)	0.568	0.86 (0.75-0.99)	0.034
RS2072183	G	1890	0.23	635	0.22	1.01 (0.86-1.19)	0.900	0.96 (0.84-1.11)	0.582
RS2075650	G	1859	0.14	624	0.15	1.01 (0.83-1.24)	0.893	1.00 (0.85-1.18)	1.000
RS2081687	T	1901	0.35	641	0.37	1.10 (0.96-1.27)	0.176	1.13 (1.01-1.27)	0.036
RS2083637	A	1874	0.74	622	0.75	1.01 (0.86-1.18)	0.933	0.90 (0.68-1.18)	0.433
RS2228671	C	1883	0.88	640	0.89	1.22 (1.01-1.46)	0.037	1.16 (0.97-1.38)	0.101
RS2240466	G	1873	0.90	621	0.89	0.91 (0.72-1.15)	0.412	0.95 (0.73-1.22)	0.670
RS2255141	T	1902	0.30	641	0.32	1.06 (0.91-1.23)	0.450	1.12 (0.85-1.47)	0.425
RS2271293	A	1817	0.87	620	0.87	1.02 (0.84-1.24)	0.837	0.93 (0.75-1.15)	0.503
RS2277862	C	1901	0.85	642	0.83	0.90 (0.74-1.10)	0.317	0.97 (0.83-1.13)	0.697
RS2290159	G	1895	0.80	637	0.80	0.93 (0.78-1.11)	0.440	0.88 (0.76-1.02)	0.078
RS2293889	A	1890	0.41	635	0.41	1.01 (0.89-1.16)	0.853	1.00 (0.89-1.11)	0.958
RS2304130	A	1886	0.91	640	0.91	1.01 (0.80-1.27)	0.933	0.93 (0.76-1.12)	0.426
rs2241213	G	1834	0.47	623	0.45	0.95 (0.83-1.08)	0.444	0.99 (0.86-1.13)	0.860
RS2412710	A	1901	0.02	641	0.03	1.08 (0.68-1.70)	0.758	0.82 (0.58-1.16)	0.251

RS2479409	G	1895	0.36	635	0.36	0.99 (0.86-1.13)	0.865	0.98 (0.88-1.10)	0.728
RS2650000	A	1839	0.34	623	0.34	1.01 (0.88-1.17)	0.858	1.00 (0.89-1.13)	0.981
RS2652834	A	1886	0.20	633	0.19	0.98 (0.83-1.16)	0.776	0.93 (0.80-1.08)	0.346
RS2737229	T	1901	0.67	642	0.68	1.00 (0.87-1.15)	0.979	0.92 (0.79-1.09)	0.343
RS2814944	T	1890	0.15	635	0.13	0.91 (0.76-1.09)	0.307	0.93 (0.77-1.11)	0.400
RS2814982	G	1902	0.90	641	0.90	0.95 (0.75-1.19)	0.643	0.98 (0.81-1.19)	0.870
RS2902940	A	1890	0.70	635	0.71	1.04 (0.90-1.19)	0.607	1.06 (0.61-1.85)	0.828
RS2923084	G	1890	0.19	635	0.18	0.93 (0.79-1.09)	0.351	0.97 (0.84-1.12)	0.707
RS2925979	T	1890	0.30	635	0.33	1.23 (1.05-1.44)	0.012	1.13 (0.93-1.38)	0.217
RS2929282	A	1901	0.05	641	0.05	0.99 (0.74-1.33)	0.952	1.01 (0.79-1.30)	0.927
RS2954029	A	1870	0.52	619	0.54	1.07 (0.94-1.23)	0.310	1.05 (0.94-1.18)	0.388
RS2972146	T	1898	0.64	637	0.67	1.19 (1.05-1.36)	0.009	1.06 (0.94-1.19)	0.353
RS328	C	1835	0.89	620	0.90	1.17 (0.95-1.44)	0.146	1.03 (0.84-1.26)	0.770
RS3757354	C	1898	0.79	642	0.80	1.05 (0.89-1.23)	0.551	1.00 (0.86-1.17)	0.994
RS3764261	C	1889	0.69	634	0.70	1.09 (0.95-1.25)	0.215	1.06 (0.94-1.20)	0.327
RS3905000	T	1798	0.12	614	0.10	0.87 (0.72-1.06)	0.161	0.86 (0.54-1.39)	0.544
RS4129767	G	1898	0.51	637	0.52	1.01 (0.89-1.16)	0.840	0.97 (0.86-1.09)	0.575
RS4148008	C	1890	0.30	635	0.32	1.13 (0.97-1.31)	0.113	0.97 (0.85-1.10)	0.583
rs6544713	T	1832	0.30	630	0.30	1.06 (0.91-1.23)	0.475	0.96 (0.85-1.08)	0.491
RS439401	C	1838	0.66	602	0.65	0.99 (0.86-1.13)	0.827	0.98 (0.87-1.10)	0.678
RS442177	T	1888	0.59	635	0.59	1.03 (0.90-1.18)	0.622	1.11 (0.99-1.24)	0.086
RS4660293	C	1861	0.25	633	0.25	0.97 (0.83-1.13)	0.686	0.82 (0.72-0.94)	0.003
RS471364	G	1841	0.12	612	0.13	1.08 (0.87-1.34)	0.464	1.08 (0.91-1.28)	0.369

RS4731702	C	1863	0.52	631	0.52	1.01 (0.88-1.15)	0.944	1.01 (0.91-1.13)	0.802
RS4765127	G	1893	0.68	637	0.70	1.04 (0.89-1.20)	0.648	1.02 (0.88-1.17)	0.832
RS4846914	G	1825	0.41	602	0.43	1.08 (0.94-1.25)	0.256	0.98 (0.86-1.11)	0.755
RS492602	G	1899	0.44	640	0.45	1.05 (0.92-1.19)	0.517	1.04 (0.93-1.16)	0.483
RS4939883	C	1777	0.84	602	0.86	1.15 (0.97-1.36)	0.106	1.20 (1.04-1.40)	0.015
RS514230	T	1890	0.52	634	0.54	1.14 (1.00-1.31)	0.051	1.19 (0.94-1.51)	0.151
RS581080	C	1897	0.19	636	0.19	1.01 (0.85-1.19)	0.938	0.97 (0.84-1.12)	0.639
RS6029526	A	1888	0.47	632	0.46	0.97 (0.85-1.11)	0.690	1.01 (0.90-1.13)	0.925
RS629301	A	1897	0.79	639	0.82	1.25 (1.07-1.45)	0.004	1.23 (1.06-1.42)	0.005
RS6450176	T	1893	0.26	633	0.26	0.98 (0.85-1.14)	0.789	0.86 (0.76-0.98)	0.019
RS645040	T	1890	0.79	634	0.79	0.98 (0.83-1.16)	0.808	0.91 (0.80-1.05)	0.195
RS646776	A	1887	0.79	641	0.82	1.27 (1.09-1.47)	0.002	1.25 (1.07-1.45)	0.004
RS6511720	G	1867	0.12	615	0.11	0.80 (0.67-0.95)	0.013	0.85 (0.71-1.01)	0.062
RS6544713	T	1832	0.30	630	0.30	1.06 (0.91-1.23)	0.475	0.96 (0.85-1.08)	0.491
RS6754295	T	1839	0.76	632	0.74	0.92 (0.78-1.08)	0.294	0.85 (0.75-0.97)	0.016
RS6756629	C	1888	0.94	641	0.93	0.97 (0.74-1.27)	0.817	0.81 (0.65-1.01)	0.057
RS6882076	G	1900	0.63	641	0.66	1.11 (0.97-1.27)	0.138	1.13 (1.00-1.27)	0.044
RS693	T	1874	0.48	620	0.48	1.01 (0.89-1.16)	0.860	0.93 (0.83-1.04)	0.191
RS6987702	C	1890	0.72	641	0.75	1.16 (1.01-1.34)	0.038	1.19 (1.04-1.34)	0.008
RS7134375	C	1901	0.59	641	0.60	1.01 (0.88-1.15)	0.912	0.88 (0.79-0.98)	0.019
RS7255436	C	1897	0.49	640	0.51	1.05 (0.93-1.20)	0.439	1.10 (0.89-1.36)	0.375
RS737337	C	1826	0.08	616	0.10	1.12 (0.87-1.45)	0.393	1.11 (0.92-1.33)	0.275
RS7395662	G	1871	0.59	632	0.58	0.89 (0.77-1.02)	0.091	0.88 (0.56-1.38)	0.574

RS7515577	T	1901	0.77	642	0.78	1.05 (0.90-1.23)	0.569	1.03 (0.90-1.18)	0.691
RS7557067	A	1853	0.76	620	0.74	0.92 (0.79-1.08)	0.329	**0.87 (0.76-0.99)**	**0.029**
RS7570971	A	1901	0.41	641	0.44	1.09 (0.95-1.24)	0.217	1.14 (1.02-1.27)	0.019
RS7679	C	1807	0.20	622	0.19	1.02 (0.85-1.21)	0.869	1.04 (0.90-1.20)	0.595
RS780093	T	1860	0.41	628	0.44	1.11 (0.96-1.27)	1.530	0.95 (0.85-1.07)	0.426
RS7819412	C	1866	0.51	612	0.49	0.93 (0.81-1.06)	0.271	0.93 (0.82-1.04)	0.194
RS7941030	C	1901	0.38	642	0.39	1.03 (0.89-1.18)	0.725	0.96 (0.86-1.07)	0.463
RS8017377	A	1901	0.46	642	0.46	0.99 (0.86-1.13)	0.855	0.92 (0.82-1.03)	0.158
RS838880	A	1901	0.69	641	0.71	1.03 (0.90-1.19)	0.643	1.01 (0.89-1.14)	0.890
RS9488822	T	1899	0.67	640	0.67	1.00 (0.87-1.15)	0.990	1.09 (0.97-1.23)	0.149
RS964184	C	1880	0.14	639	0.14	0.96 (0.80-1.17)	0.704	0.80 (0.56-1.14)	0.215
RS9686661	T	1902	0.19	641	0.20	1.08 (0.91-1.29)	0.378	1.15 (1.00-1.32)	0.050
RS9987289	A	1896	0.07	639	0.07	0.88 (0.69-1.11)	0.274	0.68 (0.53-0.88)	0.003
SNPs selected for their association with anthropomorphic related traits									
RS10146997	G	1900	0.19	642	0.21	1.11 (0.93-1.32)	0.262	1.07 (0.93-1.23)	0.341
RS10838738	C	1813	0.32	612	0.32	1.05 (0.91-1.22)	0.506	1.02 (0.90-1.16)	0.720
RS10913469	G	1764	0.19	599	0.21	1.19 (0.98-1.44)	0.081	**1.29 (1.10-1.51)**	**0.001**
RS10938397	G	1810	0.43	615	0.42	0.94 (0.83-1.08)	0.371	0.96 (0.86-1.07)	0.478
RS11084753	G	1794	0.68	609	0.67	0.94 (0.81-1.09)	0.389	**0.81 (0.70-0.94)**	**0.004**
RS1260326	T	1863	0.42	615	0.44	1.07 (0.93-1.23)	0.343	0.91 (0.72-1.15)	0.426
RS12970134	A	1839	0.26	621	0.28	1.10 (0.94-1.29)	0.239	1.01 (0.89-1.15)	0.833
RS2568958	A	1854	0.61	622	0.61	0.96 (0.84-1.11)	0.572	0.95 (0.83-1.10)	0.495
RS2605100	G	1895	0.70	639	0.71	1.04 (0.90-1.19)	0.616	1.10 (0.98-1.24)	0.115

rs29939	G	1837	0.69	615	0.68	0.91 (0.78-1.06)	0.212	**0.79 (0.66-0.93)**	**0.006**
RS3764261	C	1889	0.69	634	0.70	1.09 (0.95-1.25)	0.215	1.06 (0.94-1.20)	0.327
RS6265	G	1792	0.79	612	0.78	0.91 (0.76-1.08)	0.267	0.99 (0.86-1.12)	0.825
RS6499640	G	1841	0.57	614	0.57	1.03 (0.89-1.18)	0.728	1.04 (0.93-1.16)	0.551
RS6548238	C	1831	0.83	618	0.82	0.96 (0.80-1.16)	0.694	0.91 (0.79-1.06)	0.225
RS7498665	G	1859	0.36	626	0.38	1.06 (0.92-1.22)	0.455	1.04 (0.90-1.21)	0.562
RS7561317	G	1767	0.83	596	0.82	0.98 (0.82-1.18)	0.858	0.93 (0.80-1.09)	0.368
RS7647305	G	1824	0.81	602	0.80	1.01 (0.85-1.20)	0.902	0.97 (0.84-1.12)	0.671
RS8050136	A	1815	0.36	622	0.40	**1.19 (1.04-1.38)**	**0.014**	**1.20 (1.07-1.34)**	**0.002**
RS925946	T	1896	0.28	641	0.30	1.14 (0.98-1.32)	0.091	**1.29 (1.10-1.51)**	**0.002**
RS987237	G	1899	0.19	641	0.18	0.95 (0.81-1.12)	0.561	0.96 (0.83-1.12)	0.613

SNP, single nucleotide polymorphism; BP, base pair; RA, risk allele; RAF, risk allele frequency; CHD, coronary heart disease; HR, hazard ratio; CI, confidence interval; P, P-value; MI, myocardial infarction; CAD, coronary artery disease

* Adjusted for sex, current smoking, systolic blood pressure, total cholesterol, HDL cholesterol, body mass index, diabetes and family history of MI, age was used as the time scale variable

Table s4. Information about the subset of SNPs, their coefficients and the appropriate weighting factors. This subset of SNPs was selected using penalized regression (lasso) method.

SNP	Locus	Nearby genes	Coefficient	weighting factor*	Previously associated with:
rs10757278	9p21.3	CDKN2A, CDKN2B	0.159	1.000	CHD
rs2925979	16q23.2	CMIP	0.066	0.414	HDL
rs6882076	5q33.3	TIMD4	0.006	0.039	TC
rs2954029	8q24.13	TRIB1	0.001	0.007	TG
rs6987702	8q24.13	TRIB1	0.011	0.067	TC
rs10889352	1p31.1	DOCK7	0.023	0.142	TC
rs2972146	2q36.3	IRS1	0.050	0.316	HDL-C
rs11556924	7q32.2	ZC3HC1	0.031	0.198	CAD
rs514230	1q42.3	IRF2BP2	0.021	0.135	TC
rs8050136	16q12.2	FTO	0.063	0.394	DM2, BMI
rs181362	22q11.21	UBE2L3	0.015	0.096	HDL
rs646776	1p13.3	CELSR2, PSRC1, SORT1	0.065	0.406	MI, LDL-C
rs925946	11p14.1	BDNF	0.022	0.136	BMI
rs2000999	16q22.2	HPR	0.128	0.804	TC

* The weighting factor is set to 1 for the SNP with the highest coefficient

Table s5a. Reclassification of subjects when a Genetic Risk Score (GRS) composed of all 179 SNPs was used in addition to traditional risk factors (age, sex, current smoking, total cholesterol, HDL cholesterol, BMI, parental history of MI and self reported diabetes).

Model without Overall GRS	**Model with Overall GRS**				
	<2%	**2%-<5%**	**5%-<10%**	**≥10%**	**Total**
Incident cases					
<2%	61	1	0	0	62
2%-<5%	1	128	4	0	133
5%-<10%	0	1	126	2	129
≥10%	0	0	4	106	110
Total	62	130	134	108	434
Non-cases					
<2%	1341	12	0	0	1353
2%-<5%	13	389	7	0	409
5%-<10%	0	11	157	4	172
≥10%	0	0	3	75	78
Total	1354	412	167	79	2012

Table s5b. Reclassification of subjects when a Genetic Risk Score (GRS) composed of the 153 CHD risk factor SNPs was used in addition to traditional risk factors (age, sex, current smoking, total cholesterol, HDL cholesterol, BMI, parental history of MI and self reported diabetes).

Model without Risk Factor GRS	**Model with Risk Factor GRS**				
	<2%	**2%-<5%**	**5%-<10%**	**≥10%**	**Total**
Incident cases					
<2%	61	1	0	0	62
2%-<5%	0	131	2	0	133
5%-<10%	0	1	126	2	129
≥10%	0	0	1	109	110
Total	61	133	129	111	434
Non-cases					
<2%	1347	6	0	0	1353
2%-<5%	8	398	3	0	409
5%-<10%	0	2	169	1	172
≥10%	0	0	1	77	78
Total	1355	406	173	78	2012

Table s5c. Reclassification of subjects when a Genetic Risk Score (GRS) composed of the 29 CHD risk SNPs was used in addition to traditional risk factors (age, sex, current smoking, total cholesterol, HDL cholesterol, BMI, parental history of MI and self reported diabetes).

Model without GRS	**Model with CHD GRS**				
	<2%	**2%-<5%**	**5%-<10%**	**≥10%**	**Total**
Incident cases					
<2%	58	4	0	0	62
2%-<5%	2	126	5	0	133
5%-<10%	0	4	116	9	129
≥10%	0	0	4	106	110
Total	60	134	125	115	434
Non-cases					
<2%	1330	23	0	0	1353
2%-<5%	24	374	11	0	409
5%-<10%	0	18	150	4	172
≥10%	0	0	5	73	78
Total	1354	415	166	77	2012

Table s5d. Reclassification of subjects when a Genetic Risk Score (GRS) composed of the 29 weighted CHD risk SNPs was used in addition to traditional risk factors (age, sex, current smoking, total cholesterol, HDL cholesterol, BMI, parental history of MI and self reported diabetes).

Model without Risk Factor GRS	Model with Risk Factor GRS				
	<2%	**2%-<5%**	**5%-<10%**	**≥10%**	**Total**
Incident cases					
<2%	58	4	0	0	62
2%-<5%	5	116	12	0	133
5%-<10%	0	3	114	12	129
≥10%	0	0	9	101	110
Total	63	123	135	113	434
Non-cases					
<2%	1326	27	0	0	1353
2%-<5%	34	361	14	0	409
5%-<10%	0	22	144	6	172
≥10%	0	0	8	70	78
Total	1360	410	166	76	2012

Table s5e. Reclassification of subjects when a weighted Genetic Risk Score (GRS) composed of SNPs selected using LASSO regression was used in addition to traditional risk factors (age, sex, current smoking, total cholesterol, HDL cholesterol, BMI, parental history of MI and self reported diabetes).

Model without LASSO GRS	**Model with LASSO GRS**				
	<2%	**2%-<5%**	**5%-<10%**	**≥10%**	**Total**
Incident cases					
<2%	47	15	0	0	62
2%-<5%	14	97	22	0	133
5%-<10%	0	15	83	31	129
≥10%	0	1	17	92	110
Total	61	128	122	123	434
Non-cases					
<2%	1311	42	0	0	1353
2%-<5%	88	289	32	0	409
5%-<10%	0	47	113	12	172
≥10%	0	0	17	61	78
Total	1399	378	162	73	2012

References to supplementary documents

1. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AF, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Carlquist JF, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Diemert P, Do R, Doering A, Eifert S, Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJ, Khaw KT, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Muhleisen TW, Muhlestein JB, Munzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nothen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schafer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoep JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WH, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann HE, Wild PS, Willenborg C, Witteman JC, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, Marz W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333-338

2. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443-453

3. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, 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, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Daly MJ, Nemesh J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burtt N, Gabriel SB, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall AS, Linsel-Nitschke P, Lieb W, Ziegler A, Konig I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein SE, Scheffold T, Berger K, Huge A, Martinelli N, Olivieri O, Corrocher R, McKeown P, Erdmann E, Konig IR, Holm H, Thorleifsson G, Thorsteinsdottir U, Stefansson K, Do R, Xie C, Siscovick D. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat.Genet.* 2009;41:334-341

4. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T, Martin NG, Pedersen NL, Kyvik KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllensten U, Campbell H, Rudan I, Johansson A, Marroni F, Hayward C,

Vitart V, Jonasson I, Pattaro C, Wright A, Hastie N, Pichler I, Hicks AA, Falchi M, Willemsen G, Hottenga JJ, de Geus EJ, Montgomery GW, Whitfield J, Magnusson P, Saharinen J, Perola M, Silander K, Isaacs A, Sijbrands EJ, Uitterlinden AG, Witteman JC, Oostra BA, Elliott P, Ruokonen A, Sabatti C, Gieger C, Meitinger T, Kronenberg F, Doring A, Wichmann HE, Smit JH, McCarthy MI, van Duijn CM, Peltonen L. Loci influencing lipid levels and coronary heart disease risk in 16 european population cohorts. *Nat Genet*. 2009;41:47-55

5. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, 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 JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, Konig IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Doring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA, Jr., Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713

6. A genome-wide association study in europeans and south asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43:339-344

7. Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet DA, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, El Mokhtari NE, Schafer A, Marz W, Renner W, Bugert P, Kluter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H. New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat.Genet.* 2009;41:280-282

8. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH,

Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491-1493

9. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*. 2008;40:161-169

10. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat.Genet.* 2009;41:56-65

11. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Kottgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FU, Smith AV, Taylor K, Scharpf RB, Hwang SJ, Sijbrands EJ, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JC, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, van Duijn CM. Genome-wide association study of blood pressure and hypertension. *Nat.Genet.* 2009

12. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, Papadakis K, Voight BF, Scott LJ, Zhang F, Farrall M, Tanaka T, Wallace C, Chambers JC, Khaw KT, Nilsson P, van der HP, Polidoro S, Grobbee DE, Onland-Moret NC, Bots ML, Wain LV, Elliott KS, Teumer A, Luan J, Lucas G, Kuusisto J, Burton PR, Hadley D, McArdle WL, Brown M, Dominiczak A, Newhouse SJ, Samani NJ, Webster J, Zeggini E, Beckmann JS, Bergmann S, Lim N, Song K, Vollenweider P, Waeber G, Waterworth DM, Yuan X, Groop L, Orho-Melander M, Allione A, Di GA, Guarrera S, Panico S, Ricceri F, Romanazzi V, Sacerdote C, Vineis P, Barroso I, Sandhu MS, Luben RN, Crawford GJ, Jousilahti P, Perola M, Boehnke M, Bonnycastle LL, Collins FS, Jackson AU, Mohlke KL, Stringham HM, Valle TT, Willer CJ, Bergman RN, Morken MA, Doring A, Gieger C, Illig T, Meitinger T, Org E, Pfeufer A, Wichmann HE, Kathiresan S, Marrugat J, O'Donnell CJ, Schwartz SM, Siscovick DS, Subirana I, Freimer NB, Hartikainen AL, McCarthy MI, O'Reilly PF, Peltonen L, Pouta A, de Jong PE, Snieder H, van Gilst WH, Clarke R, Goel A, Hamsten A, Peden JF, Seedorf U, Syvanen AC, Tognoni G, Lakatta EG, Sanna S, Scheet P, Schlessinger D, Scuteri A, Dorr M, Ernst F, Felix SB, Homuth G, Lorbeer R, Reffelmann T, Rettig R, Volker U, Galan P, Gut IG, Hercberg S, Lathrop GM, Zelenika D, Deloukas P, Soranzo N, Williams FM, Zhai G, Salomaa V, Laakso M, Elosua R, Forouhi NG, Volzke H, Uiterwaal CS, van der Schouw YT, Numans ME, Matullo G, Navis G, Berglund G, Bingham SA, Kooner JS, Connell JM, Bandinelli S, Ferrucci L, Watkins H, Spector TD, Tuomilehto J, Altshuler D, Strachan DP, Laan M, Meneton P, Wareham NJ, Uda M, Jarvelin MR, Mooser V, Melander O, Loos RJ, Elliott P, Abecasis GR, Caulfield M, Munroe PB. Genome-wide association study identifies eight loci associated with blood pressure. *Nat.Genet.* 2009

13. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS,

McCarthy MI, Hattersley AT. Replication of genome-wide association signals in uk samples reveals risk loci for type 2 diabetes. *Science*. 2007;316:1336-1341

14. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*. 2007;316:1341-1345

15. Timpson NJ, Lindgren CM, Weedon MN, Randall J, Ouwehand WH, Strachan DP, Rayner NW, Walker M, Hitman GA, Doney AS, Palmer CN, Morris AD, Hattersley AT, Zeggini E, Frayling TM, McCarthy MI. Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes*. 2009;58:505-510

16. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat.Genet.* 2008;40:638-645

17. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316:1331-1336

18. Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, Hirota Y, Mori H, Jonsson A, Sato Y, Yamagata K, Hinokio Y, Wang HY, Tanahashi T, Nakamura N, Oka Y, Iwasaki N, Iwamoto Y, Yamada Y, Seino Y, Maegawa H, Kashiwagi A, Takeda J, Maeda E, Shin HD, Cho YM, Park KS, Lee HK, Ng MC, Ma RC, So WY, Chan JC, Lyssenko V, Tuomi T, Nilsson P, Groop L, Kamatani N, Sekine A, Nakamura Y, Yamamoto K, Yoshida T, Tokunaga K, Itakura M, Makino H, Nanjo K, Kadowaki T, Kasuga M. Variants in kcnq1 are associated with susceptibility to type 2 diabetes mellitus. *Nat.Genet.* 2008;40:1092-1097

19. Unoki H, Takahashi A, Kawaguchi T, Hara K, Horikoshi M, Andersen G, Ng DP, Holmkvist J, Borch-Johnsen K, Jorgensen T, Sandbaek A, Lauritzen T, Hansen T, Nurbaya S, Tsunoda T, Kubo M, Babazono T, Hirose H, Hayashi M, Iwamoto Y, Kashiwagi A, Kaku K, Kawamori R, Tai ES, Pedersen O, Kamatani N, Kadowaki T, Kikkawa R, Nakamura Y, Maeda S. Snps in kcnq1 are associated with susceptibility to type 2 diabetes in east asian and european populations. *Nat.Genet.* 2008;40:1098-1102

20. Takeuchi F, Serizawa M, Yamamoto K, Fujisawa T, Nakashima E, Ohnaka K, Ikegami H, Sugiyama T, Katsuya T, Miyagishi M, Nakashima N, Nawata H, Nakamura J, Kono S, Takayanagi R, Kato N. Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the japanese population. *Diabetes*. 2009;58:1690-1699
21. Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, Bacot F, Balkau B, Belisle A, Borch-Johnsen K, Charpentier G, Dina C, Durand E, Elliott P, Hadjadj S, Jarvelin MR, Laitinen J, Lauritzen T, Marre M, Mazur A, Meyre D, Montpetit A, Pisinger C, Posner B, Poulsen P, Pouta A, Prentki M, Ribel-Madsen R, Ruokonen A, Sandbaek A, Serre D, Tichet J, Vaxillaire M, Wojtaszewski JF, Vaag A, Hansen T, Polychronakos C, Pedersen O, Froguel P, Sladek R. Genetic variant near irs1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat.Genet.* 2009;41:1110-1115
22. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-678
23. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. A variant in cdkal1 influences insulin response and risk of type 2 diabetes. *Nat Genet*. 2007;39:770-775
24. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445:881-885
25. Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B, Hypponen J, Korhonen VP, Asikainen J, Devine C, Tuomainen TP, Luedemann J, Nauck M, Kerner W, Stephens RH, New JP, Ollier WE, Gibson JM, Payton A, Horan MA, Pendleton N, Mahoney W, Meyre D, Delplanque J, Froguel P, Luzzatto O, Yakir B, Darvasi A. Type 2 diabetes whole-genome association study in four populations: The diagen consortium. *Am J Hum Genet.* 2007;81:338-345
26. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet*. 2008;40:189-197
27. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, Sovio U, Ruokonen A, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Gabriel S, Elliot P, McCarthy MI, Daly MJ, Jarvelin MR, Freimer NB, Peltonen L. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat.Genet.* 2009;41:35-46
28. Sandhu MS, Waterworth DM, Debenham SL, Wheeler E, Papadakis K, Zhao JH, Song K, Yuan X, Johnson T, Ashford S, Inouye M, Luben R, Sims M, Hadley D, McArdle W, Barter P, Kesaniemi YA, Mahley RW, McPherson R, Grundy SM, Bingham SA, Khaw KT, Loos RJ, Waeber G, Barroso I, Strachan DP, Deloukas P, Vollenweider P, Wareham NJ, Mooser V. Ldl-cholesterol concentrations: A genome-wide association study. *Lancet*. 2008;371:483-491
29. Ridker PM, Pare G, Parker AN, Zee RY, Miletich JP, Chasman DI. Polymorphism in the cetp gene region, hdl cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18 245 initially healthy women from the women's genome health study. *Circ Cardiovasc Genet*. 2009;2:26-33

30. Heid IM, Boes E, Muller M, Kollerits B, Lamina C, Coassin S, Gieger C, Doring A, Klopp N, Frikke-Schmidt R, Tybjaerg-Hansen A, Brandstatter A, Luchner A, Meitinger T, Wichmann HE, Kronenberg F. Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based kora study sheds new light on intergenic regions. *Circ Cardiovasc Genet*. 2008;1:10-20

31. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, Styrkarsdottir U, Gretarsdottir S, Thorlacius S, Jonsdottir I, Jonsdottir T, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Jonsson F, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Lauritzen T, Aben KK, Verbeek AL, Roeleveld N, Kampman E, Yanek LR, Becker LC, Tryggvadottir L, Rafnar T, Becker DM, Gulcher J, Kiemeney LA, Pedersen O, Kong A, Thorsteinsdottir U, Stefansson K. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat.Genet.* 2009;41:18-24

32. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, Speliotes EK, Thorleifsson G, Willer CJ, Herrera BM, Jackson AU, Lim N, Scheet P, Soranzo N, Amin N, Aulchenko YS, Chambers JC, Drong A, Luan J, Lyon HN, Rivadeneira F, Sanna S, Timpson NJ, Zillikens MC, Zhao JH, Almgren P, Bandinelli S, Bennett AJ, Bergman RN, Bonnycastle LL, Bumpstead SJ, Chanock SJ, Cherkas L, Chines P, Coin L, Cooper C, Crawford G, Doering A, Dominiczak A, Doney AS, Ebrahim S, Elliott P, Erdos MR, Estrada K, Ferrucci L, Fischer G, Forouhi NG, Gieger C, Grallert H, Groves CJ, Grundy S, Guiducci C, Hadley D, Hamsten A, Havulinna AS, Hofman A, Holle R, Holloway JW, Illig T, Isomaa B, Jacobs LC, Jameson K, Jousilahti P, Karpe F, Kuusisto J, Laitinen J, Lathrop GM, Lawlor DA, Mangino M, McArdle WL, Meitinger T, Morken MA, Morris AP, Munroe P, Narisu N, Nordstrom A, Nordstrom P, Oostra BA, Palmer CN, Payne F, Peden JF, Prokopenko I, Renstrom F, Ruokonen A, Salomaa V, Sandhu MS, Scott LJ, Scuteri A, Silander K, Song K, Yuan X, Stringham HM, Swift AJ, Tuomi T, Uda M, Vollenweider P, Waeber G, Wallace C, Walters GB, Weedon MN, Witteman JC, Zhang C, Zhang W, Caulfield MJ, Collins FS, Davey SG, Day IN, Franks PW, Hattersley AT, Hu FB, Jarvelin MR, Kong A, Kooner JS, Laakso M, Lakatta E, Mooser V, Morris AD, Peltonen L, Samani NJ, Spector TD, Strachan DP, Tanaka T, Tuomilehto J, Uitterlinden AG, van Duijn CM, Wareham NJ, Hugh W, Waterworth DM, Boehnke M, Deloukas P, Groop L, Hunter DJ, Thorsteinsdottir U, Schlessinger D, Wichmann HE, Frayling TM, Abecasis GR, Hirschhorn JN, Loos RJ, Stefansson K, Mohlke KL, Barroso I, McCarthy MI. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS.Genet.* 2009;5:e1000508

33. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy RM, Attwood AP, Beckmann JS, Berndt SI, Jacobs KB, Chanock SJ, Hayes RB, Bergmann S, Bennett AJ, Bingham SA, Bochud M, Brown M, Cauchi S, Connell JM, Cooper C, Smith GD, Day I, Dina C, De S, Dermitzakis ET, Doney AS, Elliott KS, Elliott P, Evans DM, Sadaf Farooqi I, Froguel P, Ghori J, Groves CJ, Gwilliam R, Hadley D, Hall AS, Hattersley AT, Hebebrand J, Heid IM, Lamina C, Gieger C, Illig T, Meitinger T, Wichmann HE, Herrera B, Hinney A, Hunt SE, Jarvelin MR, Johnson T, Jolley JD, Karpe F, Keniry A, Khaw KT, Luben RN, Mangino M, Marchini J, McArdle WL, McGinnis R, Meyre D, Munroe PB, Morris AD, Ness AR, Neville MJ, Nica AC, Ong KK, O'Rahilly S, Owen KR, Palmer CN, Papadakis K, Potter S, Pouta A, Qi L, Randall JC, Rayner NW, Ring SM, Sandhu MS, Scherag A, Sims MA, Song K, Soranzo N, Speliotes EK, Syddall HE, Teichmann SA, Timpson NJ, Tobias JH, Uda M, Vogel CI, Wallace C, Waterworth DM, Weedon MN, Willer CJ, Wraight, Yuan X, Zeggini E, Hirschhorn JN, Strachan DP, Ouwehand WH, Caulfield MJ, Samani NJ, Frayling TM, Vollenweider P, Waeber G, Mooser V, Deloukas P, McCarthy MI, Wareham NJ, Barroso I, Kraft P, Hankinson SE, Hunter DJ, Hu FB, Lyon HN, Voight BF, Ridderstrale M, Groop L, Scheet P, Sanna S, Abecasis GR, Albai G, Nagaraja R, Schlessinger D, Jackson AU, Tuomilehto J, Collins FS, Boehnke M, Mohlke KL. Common variants near mc4r are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008;40:768-775

34. Heard-Costa NL, Zillikens MC, Monda KL, Johansson A, Harris TB, Fu M, Haritunians T, Feitosa MF, Aspelund T, Eiriksdottir G, Garcia M, Launer LJ, Smith AV, Mitchell BD, McArdle PF, Shuldiner AR, Bielinski SJ, Boerwinkle E, Brancati F, Demerath EW, Pankow JS, Arnold AM, Chen YD, Glazer NL, McKnight B, Psaty BM, Rotter JI, Amin N, Campbell H, Gyllensten U, Pattaro C, Pramstaller PP, Rudan I, Struchalin M, Vitart V, Gao X, Kraja A, Province MA, Zhang Q, Atwood LD, Dupuis J, Hirschhorn JN, Jaquish CE, O'Donnell CJ, Vasan RS, White CC, Aulchenko YS, Estrada K, Hofman A, Rivadeneira F, Uitterlinden AG, Witteman JC, Oostra BA, Kaplan RC, Gudnason V, O'Connell JR, Borecki IB, van Duijn CM, Cupples LA, Fox CS, North KE. Nrxn3 is a novel locus for waist circumference: A genome-wide association study from the charge consortium. *PLoS.Genet.* 2009;5:e1000539

35. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, Balding D, Scott J, Kooner JS. Common genetic variation near mc4r is associated with waist circumference and insulin resistance. *Nat Genet*. 2008;40:716-718

36. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C, Lettre G, Lim N, Lyon HN, McCarroll SA, Papadakis K, Qi L, Randall JC, Roccasecca RM, Sanna S, Scheet P, Weedon MN, Wheeler E, Zhao JH, Jacobs LC, Prokopenko I, Soranzo N, Tanaka T, Timpson NJ, Almgren P, Bennett A, Bergman RN, Bingham SA, Bonnycastle LL, Brown M, Burtt NP, Chines P, Coin L, Collins FS, Connell JM, Cooper C, Smith GD, Dennison EM, Deodhar P, Elliott P, Erdos MR, Estrada K, Evans DM, Gianniny L, Gieger C, Gillson CJ, Guiducci C, Hackett R, Hadley D, Hall AS, Havulinna AS, Hebebrand J, Hofman A, Isomaa B, Jacobs KB, Johnson T, Jousilahti P, Jovanovic Z, Khaw KT, Kraft P, Kuokkanen M, Kuusisto J, Laitinen J, Lakatta EG, Luan J, Luben RN, Mangino M, McArdle WL, Meitinger T, Mulas A, Munroe PB, Narisu N, Ness AR, Northstone K, O'Rahilly S, Purmann C, Rees MG, Ridderstrale M, Ring SM, Rivadeneira F, Ruokonen A, Sandhu MS, Saramies J, Scott LJ, Scuteri A, Silander K, Sims MA, Song K, Stephens J, Stevens S, Stringham HM, Tung YC, Valle TT, van Duijn CM, Vimaleswaran KS, Vollenweider P, Waeber G, Wallace C, Watanabe RM, Waterworth DM, Watkins N, Witteman JC, Zeggini E, Zhai G, Zillikens MC, Altshuler D, Caulfield MJ, Chanock SJ, Farooqi IS, Ferrucci L, Guralnik JM, Hattersley AT, Hu FB, Jarvelin MR, Laakso M, Mooser V, Ong KK, Ouwehand WH, Salomaa V, Samani NJ, Spector TD, Tuomi T, Tuomilehto J, Uda M, Uitterlinden AG, Wareham NJ, Deloukas P, Frayling TM, Groop LC, Hayes RB, Hunter DJ, Mohlke KL, Peltonen L, Schlessinger D, Strachan DP, Wichmann HE, McCarthy MI, Boehnke M, Barroso I, Abecasis GR, Hirschhorn JN. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat.Genet.* 2009;41:25-34

Chapter 9

General Discussion: Summary and Synthesis and Future perspective

Summary and synthesis

With the success of recent genome-wide association (GWA) studies, many genuine common genetic variants underlying plasma lipid levels and coronary heart disease (CHD) risk have been identified in the past 5 years. These genetic loci, especially the novel ones, should provide the foundation to develop a broad biological understanding of lipid metabolism and CHD pathophysiology, and most importantly, to identify new therapeutic opportunities (drug targets) for the prevention and treatment of CHD. At the same time, these discoveries have been pursued to use genetic markers to usher in a new era of personalized (genetic) medicine by incorporating genetic information into formulas for risk prediction, including those to be used in the primary prevention of CHD. These new genetic markers, however, must be held to the same standards as used for other biomarkers before they could be brought to the general public and the patients. In order to explore these applications, this research was conducted focusing on lipid metabolism and genetic CHD risk prediction.

Common genetic variants contributing to inter-individual variations in plasma lipid levels

Chapter 2, 3, and 4 explored common genetic variants involved in plasma cholesterol metabolism using a pathway-driven approach. Common genetic variants in the genes along known cholesterol metabolic pathways, such as bile acid and bile metabolic pathways, the HDL cholesterol metabolic pathway, and the plasma total cholesterol metabolic pathway, are involved in determining plasma cholesterol levels. The modest effect associated with each individual variant, however, caused the amount of heritability explained by them in aggregate to be relatively small: 13 single nucleotide polymorphisms (SNPs) explained 4% of inter-individual variation in HDL cholesterol levels (Chapter 3), whereas 12 SNPs explained 6.9% of inter-individual variation in total cholesterol levels (Chapter 4). These observations are consistent with the results recently reported by the Global Lipids Genetics Consortium: 12.4% (total cholesterol), 12.2% (LDL cholesterol), 12.1% (HDL cholesterol), and 9.6% (triglycerides (TG)) of the total variance in each trait are explained by 121 loci in total. This corresponds to 25-30% of the genetic variance for each trait (1). One of the explanations for the limited heritability explained by identified common variants are the low-frequency variants (0.5% < minor allele frequency < 5%) and the rare variants (minor allele frequency < 0.5%) that are not well covered by current genotyping platforms (2). This has been demonstrated in a genetic study of plasma HDL levels long before the widespread use of GWA studies (3), and is further corroborated by later studies (4-6). At present, the emphasis has shifted towards the rare variant-common disease (trait) hypothesis. This is in accordance with the presence of a gradient of genetic effects that at one extreme, when the effects are the largest, result in single gene disorders with Mendelian pattern of inheritance, and at the other extreme, when the effects are modest or small, result in polygenic disorders with no simple pattern of inheritance (2, 7). The paradigm in the genetic studies of lipid traits is now shifting towards identification of low-frequency and rare variants with large effects (8). This shift has been partly accelerated by the availability of the next generation sequencing platforms, which enable identification of the low-frequency and rare variants through whole exome and whole genome sequencing.

New insight in plasma lipid metabolism and CHD pathophysiology

In chapter 5, we found that genetic variants in the FADS1 gene potentially interact with dietary polyunsaturated fatty acid (PUFA) intake to affect plasma cholesterol levels. A high intake of omega-3 PUFA was associated with increased plasma non-HDL cholesterol levels, consistent with increased plasma LDL cholesterol levels observed in fish oil intervention studies. Since the LDL receptor is the master regulator of plasma LDL cholesterol levels (9), increased LDL cholesterol levels could be due to hepatic downregulation of the LDL receptor gene (LDLR) in subjects with high omega-3 PUFA intakes. This is further confirmed by the findings described in Chapter 6 that the hepatic LDLR gene was significantly downregulated in fish oil treated mice. This study also confirmed PUFAs to be weak PPAR ligands. The increased plasma HDL cholesterol levels in the subjects with high PUFA intakes in Chapter 5 could be due to upregulated peroxisome proliferator-activated receptor (PPAR) mediated genes that are directly involved in HDL lipoprotein metabolism (10, 11). All the above may explain the changes in blood cholesterol levels upon PUFA intake observed in human studies.

In Chapter 6, with a comparative transcriptomic and metabolic analysis of fenofibrate and fish oil treatment, we found that not only downregulation in the hepatic lipogenic pathway but also upregulation in hepatic fatty acid oxidation pathways are involved in lowering plasma TG levels upon fish oil treatment. The striking parallel between fenofibrate and fish oil in hepatic downregulation of blood coagulation and fibrinolysis pathways suggest that hepatic activation of PPARα is potentially one of the mechanisms responsible for anticoagulation effects of fish oil treatment observed in humans. However, definitive proof of this concept should come from well-powered human intervention studies on hepatic gene expression patterns with fish oil treatment.

In Chapter 7, a protective effect of high δ-5 desaturase activity against CHD risk was observed, a finding which merits attention. Increased δ-5 desaturase activity contributes to the intracellular increase of EPA and especially arachidonic acid levels. Arachidonic acid and its derived eicosanoids have been regarded as negative factors in atherosclerosis, based on potential pro-inflammatory and pro-thrombotic activity. But arachidonic acid plays an important role in the normal growth and development of infants and protection of digestive tract epithelium (12). In non-fish eating populations, arachidonic acid is the predominant tissue long-chain PUFA, and could reach 80% of total PUFA (12, 13). Despite the potential pro-coagulant and pro-inflammatory effects of increased exposures of arachidonic acid and its derived eicosanoid metabolites (12, 14), there is no evidence of increased CHD risk with increased habitual arachidonic acid intake so far (15). Some of the lipoxygenation metabolites of arachidonic acid were even found to have anti-inflammatory and pro-resolving actions (16). High dietary n-6 PUFA intakes or high plasma n-6 PUFA levels are associated with increased blood HDL cholesterol levels and reduced blood TG levels (15, 17, 18). All these point to a potential cardiovascular protective effect of n-6 PUFAs. Thus, the long-term net effects of increased n-6 PUFA (including arachidonic acid) exposure in the pathogenesis of CHD remain to be destablished in future large-scale random controlled trials. The fact that increased DHA levels associated with increased δ-5 desaturase activity protected against CHD was consistent with the current established cardiovascular protective

effect of increased n-3 PUFA exposure, especially EPA and DHA. However, we could not rule out other unidentified pleiotropic cardiovascular protective effects of increased δ-5 desaturase activity, for example in immune cells that are important in atherosclerotic CHD progression.

Common genetic variants and CHD risk prediction

In Chapter 8, we constructed several gene risk scores (GRS) for CHD that consisted of SNPs directly associated with CHD or intermediate CHD risk factors in GWA studies, and tested their relationship to incident CHD and their potential to improve risk prediction. The weighted GRS based on 29 CHD SNPs predicted future CHD independently from established traditional risk factors. However, the GRS based on 153 SNPs associated with intermediate risk factors and the GRS based on the total 179 SNPs did not. None of them improved risk discrimination. Risk classification of CHD, measured by the net reclassification index, improved only when the GRS based on the 29 CHD SNPs was used. These results are generally consistent with the results from other recent studies that took a similar approach as ours (19, 20). Although the final conclusions on GRS application could not be drawn at this early stage, several important insights into the future of CHD GRSs can be obtained. First, the predictive utility of GRSs that explain only a small fraction of the heritability will likely be marginal. To significantly improve risk prediction, genetic risk assessment will have to be significantly refined. It is suggested that approximately 20% of the heritability needs to be explained to provide similar discrimination as obtained from standard risk prediction model (21, 22). Second, incremental improvement by addition of a genetic biomarker must be shown beyond well-validated risk scores by use of standard metrics to evaluate their clinical performance. Therefore, useful genetic markers for risk prediction will need to be sufficiently uncorrelated with known CHD risk factors to provide independent information on risk (22). Third, it could be argued that any genetic marker should also provide incremental risk information over and above a model that incorporates family history, given that such information is often readily available (22).

Future perspectives

Since current (genome-wide) association studies have detected the common SNPs that explained the highest fraction of the genetic variance, the remaining variants to be found will explain exponentially smaller proportions of the remaining genetic variance. In view of the precipitous decline in the cost of DNA sequencing (23), approaches based on whole exome and whole genome sequencing are expected to dominate genetic studies of plasma lipid traits and CHD in the coming years. These studies will elucidate whether low-frequency and rare variants account for a significant component of the "missing" heritability or alternative mechanism, such as epistasis, gene-environment interactions, and/or epigenetics might also play larger roles (2).

The initial results of the 1000 genome projects indicate that each genome has about 250 to 300 loss-of-function variants in the annotated genes and 50 to 100 variants previously implicated in inherited disorders (24). In addition, each genome has about 35 to 49 *de novo* germline base substitutions, a finding that indicates a germline mutation rate of 1×10^{-8} per base-pair per generation (24, 25). Furthermore, each genome has several large (>50 Kbp)

and about 100 heterozygous copy number variants covering about 3 Mbp (26-28). Collectively, the data indicate that humans differ in about 0.12% of their genomes, or about 4 million DNA sequence variants, comprising about 3.5 million SNP and several hundred thousand structural variations (28). Empirical evidence demonstrated that both common and rare variants participated in determining common diseases (traits) (29-31). Each of us inherited some variants that confer risk and some variants that provide protection, and they will therefore have an overall risk around the average. A small proportion of us, however, will have inherited mainly variants that confer risk of developing the disease (32). Each of us is perhaps, genetically susceptible to certain disease across the whole disease spectrum (2, 32); but, most of us do not have overt diseases, either because the genetic variants act in a recessive way, or for other reasons such as the presence of modifier genes (33, 34) or absence of environmental triggers.

Clinical significant improvements in predictive performance that are also cost-effective should represent the threshold for clinical utility of genetic risk prediction. Marginal improvements that meet an arbitrary threshold for statistical significance will not suffice for translation to clinical use. CHD may belong to one of the particular difficult phenotypes for genetic risk prediction. It may represent the culmination of multiple potential causal pathways (such as endothelial dysfunction, accelerated atherosclerosis, or thrombosis, etc), with each pathway having its own set of genetic associations. The allelic architecture of CHD in the general population is still not unresolved. If most of the genetic variation resides in rare variants not captured by HapMap SNPs, this variation would not be detected by GWA study. Current efforts to sequence the genome in large cohorts of individuals are under way (28), which will undoubtedly lead to uncovering many novel genetic variants that are biologically linked with disease. However, unless they identify rare variants with strong effects that also explain relatively large proportions of genetic variance, their value for risk prediction will likely be limited (2). So, at present, many questions remain about the feasibility of genetic risk prediction of CHD. Whether “we will get there” for genetic CHD prediction remains an open question (22). We hope, with a great understanding of the genetic architecture of CHD in the future, many challenges faced for genetic CHD risk prediction could be resolved. Now, clinicians should continue to inquire about family history for risk prediction, because this represents a simple, cheap, and useful risk factors for CHD that likely represents the net integrated effects from hundreds of genetic risk variants (that have yet to be discovered) with the accompanied “environmental” factors .

References:

1. 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., et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466:707-713.
2. 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., et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747-753.
3. Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305:869-872.

4. Romeo, S., Pennacchio, L.A., Fu, Y., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H.H., and Cohen, J.C. 2007. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 39:513-516.
5. Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., et al. 2010. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42:684-687.
6. Cohen, J.C., Pertsemlidis, A., Fahmi, S., Esmail, S., Vega, G.L., Grundy, S.M., and Hobbs, H.H. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Nat Acad Sci* 103:1810-1815.
7. Plomin, R., Haworth, C.M., and Davis, O.S. 2009. Common disorders are quantitative traits. *Nat Rev Genet* 10:872-878.
8. Musunuru, K., Pirruccello, J.P., Do, R., Peloso, G.M., Guiducci, C., Sougnez, C., Garimella, K.V., Fisher, S., Abreu, J., Barry, A.J., et al. 2010. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 363:2220-2227.
9. Horton, J.D., Shimano, H., Hamilton, R.L., Brown, M.S., and Goldstein, J.L. 1999. Disruption of LDL receptor gene in transgenic SREBP-1a mice unmasks hyperlipidemia resulting from production of lipid-rich VLDL. *J Clin Invest* 103:1067-1076.
10. Duval, C., Muller, M., and Kersten, S. 2007. PPARalpha and dyslipidemia. *Biochim Biophys Acta* 1771:961-971.
11. Akiyama, T.E., Sakai, S., Lambert, G., Nicol, C.J., Matsusue, K., Pimprale, S., Lee, Y.H., Ricote, M., Glass, C.K., Brewer, H.B., Jr., et al. 2002. Conditional disruption of the peroxisome proliferator-activated receptor gamma gene in mice results in lowered expression of ABCA1, ABCG1, and apoE in macrophages and reduced cholesterol efflux. *Mol Cell Biol* 22:2607-2619.
12. Calder, P.C. 2007. Dietary arachidonic acid: harmful, harmless or helpful? *Brit J Nutr* 98:451-453.
13. Harris, W.S., Poston, W.C., and Haddock, C.K. 2007. Tissue n-3 and n-6 fatty acids and risk for coronary heart disease events. *Atherosclerosis* 193:1-10.
14. Schmitz, G., and Ecker, J. 2008. The opposing effects of n-3 and n-6 fatty acids. *Prog Lipid Res* 47:147-155.
15. Harris, W.S., Mozaffarian, D., Rimm, E., Kris-Etherton, P., Rudel, L.L., Appel, L.J., Engler, M.M., Engler, M.B., and Sacks, F. 2009. Omega-6 fatty acids and risk for cardiovascular disease: a science advisory from the American Heart Association Nutrition Subcommittee of the Council on Nutrition, Physical Activity, and Metabolism; Council on Cardiovascular Nursing; and Council on Epidemiology and Prevention. *Circulation* 119:902-907.
16. Spite, M., and Serhan, C.N. 2010. Novel lipid mediators promote resolution of acute inflammation: impact of aspirin and statins. *Circ Res* 107:1170-1184.
17. Mensink, R.P., Zock, P.L., Kester, A.D., and Katan, M.B. 2003. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *The American journal of clinical nutrition* 77:1146-1155.
18. Motoyama, K.R., Curb, J.D., Kadowaki, T., El-Saed, A., Abbott, R.D., Okamura, T., Evans, R.W., Nakamura, Y., Sutton-Tyrrell, K., Rodriquez, B.L., et al. 2009. Association of serum n-6 and n-3 polyunsaturated fatty acids with lipids in 3 populations of middle-aged men. *Am J Clin Nutr* 90:49-55.
19. Ripatti, S., Tikkanen, E., Orho-Melander, M., Havulinna, A.S., Silander, K., Sharma, A., Guiducci, C., Perola, M., Jula, A., Sinisalo, J., et al. 2010. A multilocus genetic risk score

for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 376:1393-1400.

20. Paynter, N.P., Chasman, D.I., Pare, G., Buring, J.E., Cook, N.R., Miletich, J.P., and Ridker, P.M. 2010. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303:631-637.
21. Amin, N., van Duijn, C.M., and Janssens, A.C. 2009. Genetic scoring analysis: a way forward in genome wide association studies? *Eur J Epidemiol* 24:585-587.
22. Thanassoulis, G., and Vasan, R.S. 2010. Genetic cardiovascular risk prediction: will we get there? *Circulation* 122:2323-2334.
23. Mardis, E.R. 2011. A decade's perspective on DNA sequencing technology. *Nature* 470:198-203.
24. Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., Gibbs, R.A., Hurles, M.E., and McVean, G.A. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061-1073.
25. Conrad, D.F., Keebler, J.E., Depristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V., et al. 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43:712-714.
26. Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., Sampas, N., Bruhn, L., Shendure, J., and Eichler, E.E. 2010. Diversity of human copy number variation and multicopy genes. *Science* 330:641-646.
27. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* 470:59-65.
28. Marian, A.J., and Belmont, J. 2011. Strategic approaches to unraveling genetic causes of cardiovascular diseases. *Circ Res* 108:1252-1269.
29. Hegele, R.A. 2009. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet* 10:109-121.
30. Johansen, C.T., Wang, J., Lanktree, M.B., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., Schwartz, S.M., et al. 2011. An Increased Burden of Common and Rare Lipid-Associated Risk Alleles Contributes to the Phenotypic Spectrum of Hypertriglyceridemia. *Arterioscler Thromb Vasc Biol*.
31. Kotowski, I.K., Pertsemlidis, A., Luke, A., Cooper, R.S., Vega, G.L., Cohen, J.C., and Hobbs, H.H. 2006. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet* 78:410-422.
32. Donnelly, P. 2008. Progress and challenges in genome-wide association studies in humans. *Nature* 456:728-731.
33. Klassen, T., Davis, C., Goldman, A., Burgess, D., Chen, T., Wheeler, D., McPherson, J., Bourquin, T., Lewis, L., Villasana, D., et al. 2011. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. *Cell* 145:1036-1048.
34. Lehner, B. 2011. Molecular mechanisms of epistasis within and between genes. *Trends Genet* 27:323-331.

Summary

Coronary heart disease (CHD) continues to be a leading cause of morbidity and mortality among adults worldwide. Deregulated lipid metabolism (dyslipidemia) that manifests as hypercholesterolemia, hypertriglyceridemia, low high-density-lipoprotein (HDL) cholesterol levels or a combination of those, is an established risk factor for CHD among other established risk factors. Linoleic acid (LA, C18:2n-6) and alpha-linolenic acid (ALA, C18:3n-3) are polyunsaturated fatty acids (PUFAs) that cannot be synthesized *de novo* by human or animal cells, and therefore must be obtained from the diet. From these two PUFAs, two series of long-chain PUFAs are formed; the omega-6 series that are synthesized from LA, and the omega-3 series that are from ALA. Formation of these long-chain PUFAs involves a series of alternate desaturation and elongation processes. These PUFAs, especially, omega-3 PUFAs, have long been observed to reduce CHD risk. In contrast to the consistently observed cardiovascular protective effects of omega-3 PUFAs, accumulating evidence suggests a potential pro-atherogenic effect of omega-6 PUFAs, although this is still under debate.

It has been estimated that genetic factors account for 26%-69% of inter-individual variation in CHD risk. These genetic factors are thought to influence CHD risk both directly and through effects on known CHD risk factors such as plasma lipid levels. The heritability of plasma lipid levels (total cholesterol, LDL cholesterol, HDL cholesterol, and TG) is estimated to be about 50% (ranging from 28%-78%). The completion of the Human Genome Project and the International Haplotype Map Project has made it possible to perform genome-wide screens for common DNA sequence variants that are associated with phenotypes of interest, including CHD and its risk factors. This approach has substantially expanded our knowledge of the genetic basis of CHD and plasma lipid levels, with 24 and 95 unequivocal genetic loci recently identified to be associated with CHD and plasma lipid levels, respectively.

Chapter 2, 3, and 4 explored common genetic variants involved in plasma cholesterol metabolism using a pathway-driven approach. For this we used data from 3575 men and women from the Doetinchem cohort, examined thrice over 11 years. They were genotyped on 384 single nucleotide polymorphisms (SNPs) across 251 genes in regulatory pathways that control fatty acid, glucose, cholesterol and bile salt homeostasis. Common genetic variants in the genes along known cholesterol metabolic pathways, such as bile acid and bile metabolic pathways, cholesterol biosynthetic pathway, VLDL metabolic pathway, LDL metabolic pathway, and HDL metabolic pathway, are involved in determining plasma cholesterol levels. The modest effect associated with each individual variant, however, caused the amount of heritability explained by them, in aggregate, to be relatively small. One of the explanations for the limited heritability explained by identified common variants is the fact that the low-frequency variants and the rare variants are not well covered by current genotyping platforms. In view of the precipitous decline in the cost of DNA sequencing, approaches based on whole exome and whole genome sequencing are expected

to dominate genetic studies of plasma lipid traits and CHD in the coming years. These studies will elucidate whether low-frequency and rare variants account for a significant component of the "missing" heritability or alternative mechanism, such as epistasis, gene-environment interactions, and/or epigenetics might also play larger roles.

In Chapter 5, the potential mechanism underlying the influence of dietary PUFA intake on plasma cholesterol levels is explored using a population-based genetic approach using the same data as described earlier. We found that genetic variants in the FADS1 gene potentially interact with dietary PUFA intake to affect plasma cholesterol levels. A high intake of omega-3 PUFA was associated with increased plasma non-HDL cholesterol levels, consistent with increased plasma LDL cholesterol levels observed in fish oil intervention studies. Increased LDL cholesterol levels could be due to hepatic downregulation of the LDL receptor gene (LDLR) in subjects with high omega-3 PUFA intakes. This is further confirmed by the findings described in Chapter 6 that the hepatic LDLR gene was significantly downregulated in fish oil treated mice. This study also confirmed PUFAs to be weak PPAR ligands. The increased plasma HDL cholesterol levels in the subjects with high PUFA intakes in Chapter 5 could be due to PPARs-mediated genes that are directly involved in HDL lipoprotein metabolism. All these may explain the changes in blood cholesterol levels upon PUFA intake observed in human studies.

In Chapter 6, novel mechanistic insights on the TG-lowering and anti-thrombotic effects of fish oil (omega-3 PUFAs) treatment is explored in comparison with fenofibrate treatment in mice using a genomic approach. We found that not only downregulation in the hepatic lipogenic pathway but also upregulation in hepatic fatty acid oxidation pathways are involved in lowering plasma TG levels upon fish oil treatment. The striking parallel between fenofibrate and fish oil in hepatic downregulation of blood coagulation and fibrinolysis pathways suggest that hepatic activation of PPARα is potentially one of the mechanisms responsible for anticoagulation effects of fish oil treatment observed in humans.

In Chapter 7, the genetic determinants of plasma PUFA levels, and potential effects of omega-6 and omega-3 PUFAs and desaturase enzyme activities on CHD risk are explored using a population-based genetic approach. Data were used from the CAREMA cohort that involved 15,236 middle-aged subjects and was followed up for a median of 12.1 years. For the first time, a protective effect of high δ-5 desaturase activity against CHD risk was observed. Increased δ-5 desaturase activity could contribute to the intracellular increase of EPA and especially arachidonic acid levels. Despite the potential pro-coagulant and pro-inflammatory effects of increased exposures of arachidonic acid and its derived eicosanoid metabolites, there is no evidence of increased CHD risk with increased habitual arachidonic acid intake so far. Some of the oxygenated metabolites of arachidonic acid were found to have anti-inflammatory and pro-resolving actions. High dietary n-6 PUFA intakes or high plasma n-6 PUFA levels are associated with increased blood HDL cholesterol levels and reduced TG levels. All these point to a potential cardiovascular protective effect of n-6 PUFAs. Thus, the long-term net effects of increased n-6 PUFA (including arachidonic acid) exposure in the pathogenesis of CHD remains to be delineated in future large-scale

randomized controlled trials. The fact that increased EPA and/or DHA levels associated with increased δ-5 desaturase activity protect against CHD is consistent with the current established cardiovascular protective effect of increased n-3 PUFA exposure, especially EPA and DHA.

Individual genetic variants identified from the genome-wide association studies on the association with CHD and its risk factors, have been tested to improve CHD risk prediction. In Chapter 8, the current known common genetic variants associated with CHD risk factors (blood pressure, obesity, blood lipid levels, and type 2 diabetes) and CHD itself from published genome-wide association studies are examined to see whether they provide additional value in CHD risk prediction beyond established traditional CHD risk factors. We constructed several gene risk scores (GRS) for CHD that consisted of SNPs directly associated with CHD or intermediate CHD risk factors in GWA studies, and tested their relationship to incident CHD and their potential to improve risk prediction. The weighted GRS based on 29 CHD SNPs predicted future CHD independently from established traditional risk factors. However, the GRS based on 153 SNPs associated with intermediate risk factors and the GRS based on the total 179 SNPs did not. None of them improved risk discrimination. Risk classification of CHD, measured by the net reclassification index, improved only when the GRS based on the 29 CHD SNPs was used. These results are generally consistent with the results from other recent studies that took a similar approach as ours. However, the final conclusions on GRS application could not be drawn at this early stage. With a great understanding of the genetic architecture of CHD in the future, more research should be done on this topic.

In conclusion, our studies in this thesis demonstrated that common genetic variants along the known candidate cholesterol metabolic pathways are involved in determining the plasma cholesterol levels. PUFAs are not only weak PPARα ligands, but also inhibit SREBPs' activities. All these could explain part of the cardiovascular protective effects (increased HDL cholesterol levels and reduced TG levels) of PUFAs, increased LDL cholesterol levels upon fish oil treatment in humans, and potentially reduced CHD risk of high δ-5 desaturase activities. At present, many questions remain about the feasibility of genetic risk prediction of CHD. Clinicians should continue to inquire about family history of CHD for risk prediction, because this represents a simple, cheap, and useful risk factor for CHD that likely represents the net integrated effects from hundreds of genetic risk variants.

Samenvatting

Hart en vaatziekte (HVZ) is wereldwijd bij volwassenen één van de meest voorkomende ziektes en een leidende doodsoorzaak. Een alom bekende risicofactor voor HVZ is een slecht functionerend vetzuurmetabolisme. Een slecht functionerend vetzuurmetabolisme wordt gekenmerkt door een hoog totaal cholesterol, hoge triglyceride niveaus, een laag HDL-cholesterol of door een combinatie van deze factoren. Linolzuur (C18:2n-6) en alfa-linoleenzuur (C18:3n-3) zijn meervoudig onverzadigde vetzuren. Omdat deze meervoudig onverzadigde vetzuren niet door de menselijke of dierlijke cel zelf gesynthetiseerd kunnen worden, moeten deze vetzuren verkregen worden uit het dieet. Uit linolzuur en linoleenzuur worden, door een serie van afwisselende desaturatie en elongatie stappen, langketenige meervoudig onverzadigde vetzuren gevormd. Uit linolzuur worden omega-6 vetzuren gevormd en uit alfa-linoleenzuur worden omega-3 vetzuren gevormd. Van deze meervoudig onverzadigde vetzuren, in het bijzonder van de meervoudig onverzadigde omege-3 vetzuren, is het al langer bekend dat ze het risico op HVZ reduceren. De wetenschappelijke literatuur is eenduidig over het preventieve cardiovasculaire effect van omega-3 vetzuren. Voor omega-6 vetzuren daarentegen, alhoewel er over dit onderwerp nog steeds veel discussie is, is er opkomend bewijs voor een pro-atherogeen effect.

De door genetische factoren te verklaren inter-individuele variatie in HVZ risico wordt geschat op 26%-69%. Men denkt, dat deze genetische factoren HVZ zowel direct als indirect door een effect op HVZ risico factoren, zoals plasma lipiden niveaus, beïnvloeden. De erfelijkheid van plasma lipide niveaus (totaal cholesterol, LDL-cholesterol, HDL-cholesterol en triglyceriden) wordt geschat op zo'n 50% (variërend van 28%-78%). De voltooiing van het "Human Genome Project" en het "International Haplotype Map Project" hebben het mogelijk gemaakt om genoom-wijde associatie studies naar veel voorkomende DNA varianten, die geassocieerd zijn met een bepaald fenotype, bijv. HVZ, uit te voeren. Deze genoom wijde benadering heeft geleid tot een substantiële toename in kennis over de genetische basis van HVZ en lipide niveaus.

In hoofdstuk 2, 3 en 4 worden veelvoorkomende genetische varianten betrokken in het cholesterol metabolisme bestudeerd door middel van een mechanistische benadering. Voor deze studies maakten we gebruik van 3575 mannen en vrouwen uit het Doetinchem cohort. Deze mannen en vrouwen werden 3 keer in elf jaar onderzocht. In deze mannen en vrouwen werden 384 single nucleotide polymorphisms (SNPs) bepaald. Deze 384 SNPs waren gelokaliseerd in 251 genen, die vetzuur niveaus, glucose, cholesterol metabolisme en gal zout homeostasis reguleren. Veelvoorkomende genetische varianten in genen betrokken bij verschillende facetten van het cholesterol metabolisme, zoal het galzuur en gal metabolisme, de biosyntese van cholesterol, het VLDL metabolisme, het LDL metabolisme, het HDL metabolisme, zijn van invloed op cholesterol niveaus. Omdat het effect van alle individuele genetische varianten erg klein is verklaren alle bekende genetische varianten gezamenlijk maar een relatief klein deel van de erfelijkheid. Een van de mogelijke verklaringen voor de beperkte erfelijkheid van de reeds geïdentificeerde veel voorkomende genetische varianten, is dat de zeldzame genetische varianten niet goed gedekt zijn door de huidige genotyperings platforms. De drastische daling van de kosten

voor DNA sequencing in ogenschouw nemend wordt er verwacht dat sequencing gebaseerd op het gehele exoom of op het hete gehele genoom, de methode is die de genetische associatie studies naar lipiden niveaus en HVZ zal domineren in de komende jaren. Deze studies zullen ophelderen of lage frequente en zeldzame varianten verantwoordelijk zijn voor de significante component van “missende” erfelijkheid of dat een alternatief mechanisme, zoals epistasis, gen-omgevings interacties en/of epigenetica, misschien ook een rol speelt.

In hoofdstuk 5 wordt het potentiele mechanisme verantwoordelijk voor de invloed van uit het dieet afkomstige meervoudig onverzadigde vetzuren op het cholesterol metabolisme bestudeerd met behulp van een populatie representatieve genetische benadering, gebruik makend van eerder beschreven data. We vonden dat genetische varianten in het FADS1 gen en uit het dieet afkomstige meervoudige onverzadige vetzuren, mogelijk een interactie effect laten zien op plasma cholesterol niveaus. Een hoge inneming van omega-3 meervoudig onverzadige vetzuren was geassocieerd met een verhoogd plasma cholesterol niet zijnde HDL-cholesterol. Deze resultaten zijn consistent met de verhoogde plasma LDL cholesterol niveaus gevonden in visolie interventie studies. Verhoogde LDL cholesterol niveaus kunnen veroorzaakt worden door een neerwaartse regulatie in de lever van het LDL receptor gen (LDLR) in personen met hoge omega-3 meervoudig onverzadigde vetzuur inneming. Dit wordt verder bevestigd door de bevinding in hoofdstuk 6, dat het LDLR gen neerwaarts is gereguleerd inlevers van muizen die behandeld werden met visolie. Deze studie bevestigd ook dat meervoudig onverzadigde vetzuren zwakke liganden zijn voor PPAR. De verhoogde plasma HDL cholesterol niveaus in personen met een hoge meervoudig onverzadigde vetzuur inneming in hoofdstuk 5, kunnen veroorzaakt zijn door PPAR-gemedieerde genen, die betrokken zijn in het HDL lipoproteïne metabolisme. Dit alles zou de door meervoudig onverzadigde vetzuren geïnduceerde veranderingen in bloed cholesterol niveaus in mensen kunnen verklaren.

In hoofdstuk 6 worden, gebruik makende van een genomics benadering nieuwe mechanistische inzichten omtrent trilglyceriden verlagende and antitrombotische effecten van een visolie (omega-3 meervoudig onverzadigde vetzuren) behandeling in vergelijking met een fenofibraat behandeling in muizen besproken. We vonden dat niet alleen de neerwaartse regulatie van de hepatische productie van vetzuren, maar ook de opwaartse regulatie van het hepatisch vetzuur oxidatie mechanisme, betrokken is bij de verlaging van plasma triglyceride niveaus door een visolie behandeling. Er bestaat een opvallende parallel tussen fenofibraat en visolie. Beide induceren namelijk een neerwaartse regulatie van bloed stollings- en fibrinolytische mechanismen in de lever. Dit suggereert dat de hepatische activatie van PPARά een van de mogelijke mechanisme is, dat verantwoordelijk is voor het antistollings effect van visolie gevonden in humane studies.

In hoofstuk 7, worden de genetische determinanten van plasma meervoudig onverzadigde vetzuren en de potentiële effecten van omega-6 en omega-3 meervoudig onverzadigde vetzuren en desaturase enzym activiteiten op HVZ beschreven, gebruik makend van een populatie representatieve genetische benadering. De gebruikte data waren afkomstig van het CAREMA cohort, een studie met 15,236 mensen van middelbare leeftijd, die gevolgd

zijn gedurende een mediane tijd van 12.1 jaar. Voor het eerst, is er een beschermend effect van een hoge δ-5 desaturase activiteit op HVZ gevonden. Een verhoogde δ-5 desaturase activiteit kan bijdrage aan de intracellulaire verhoging van EPA en in het bijzonder, arachidonzuur niveaus. Ondanks de potentiele stollings en pro-inflammatoire effecten van een verhoogde blootstelling aan arachidonzuur en van arachidonzuur afgeleide metabolieten, is er tot nu toe geen bewijs, dat een verhoogde dagelijkse inname van arachidonzuur geassocieerd is met HVZ. Enkele van de van arachidonzuur afgeleide geoxideerde metabolieten hebben een anti-inflammatoir effect en een verhoogd oplossend vermogen. Een hoge inneming van uit de voeding afkomstige n-6 meervoudig onverzadigde vetzuren of van hoge plasma n-6 meervoudig onverzadigde vetzuur niveaus zijn geassocieerd met een verhoogd HDL-cholesterol en met gereduceerde triglyceride niveaus. Al het bewijs wijst in de richting van een potentieel cardiovasculair beschermend effect van n-6 meervoudig onverzadigde vetzuren. Daarom is het van belang het netto lange termijn effect van een verhoogde blootstelling aan n-6 meervoudig onverzadigde vetzuren (inclusief archidonzuur) op de pathogenesis van HVZ te onderzoeken in een toekomstig, grootschalig, gerandomiseerd en gecontroleerd onderzoek. Het feit dat verhoogde EPA en/of DHA niveaus geassocieerd zijn met verhoogde δ-5 desaturase activiteit en beschermend werken tegen HVZ, is consistent met het al reeds bewezen cardiovasculair beschermende effect van een verhoogde blootstelling aan n-3 meervoudig onverzadigde vetzuren, in het bijzonder aan EPA en DHA.

Er is getest of individuele genetische varianten, die in genoom-wijde associatie studies geassocieerd zijn met HVZ en met risicofactoren voor HVZ, de risicopredictie van HVZ verbeteren. In hoofdstuk 8, wordt voor genetische varianten, die in reeds gepubliceerde genoom-wijde associatie studies geassocieerd zijn met risicofactoren voor HVZ (bloeddruk, obesitas, bloed lipide niveaus en type 2 diabetes) of met HVZ zelf, getest of ze de risicopredictie gebaseerd op alleen traditionele risicofactoren voor HVZ kunnen verbeteren.

We hebben verschillende genetische risico scores voor HVZ risicopredictie, bestaande uit SNPs die direct geassocieerd waren met HVZ in GWAS of uit SNPs die geassocieerd waren met een intermediaire risicofactor voor HVZ in GWAS, geconstrueerd. Voor deze genetische risico scores hebben we onderzocht of ze geassocieerd waren met incidente HVZ en of ze de potentie hadden om de risicopredictie voor HVZ te verbeteren. De gewogen genetische risico score gebaseerd op 29 HVZ SNPs voorspelde toekomstige HVZ onafhankelijk van traditionele risicofactoren. Daarentegen, de genetische risico score gebaseerd op 153 SNPs geassocieerd met intermediaire risico factoren en de genetische risico scores gebaseerd op 179 SNPs deden dit niet. Geen van deze risicoscores verbeterden het risico onderscheidend vermogen voor HVZ. Risico classificatie voor HVZ, gemeten met de net reclassificatie index, verbeterde alleen bij de gen risico score gebaseerd op 29 HVZ SNPs. Deze resultaten komen over het algemeen overeen met gelijksoortige, recent verschenen studies. Desalniettemin kunnen we in dit vroege stadium nog geen definitieve conclusies trekken over de toepassing van gen risico scores. Om ons begrip van de genetische architectuur van HVZ in de toekomst te verbeteren, zouden we meer onderzoek naar dit onderwerp moeten doen.

Concluderend, de studies in dit proefschrift tonen aan dat veel voorkomende genetische varianten in genen in het cholesterol metabolisme, betrokken zijn bij de bepaling van cholesterol niveaus. Meervoudig onverzadigde vetzuren zijn niet alleen zwakke PPARά liganden, maar remmen ook SREBP activiteit. Dit alles verklaart gedeeltelijk het cardiovasculair beschermend effect (verhoogde HDL-cholesterol niveaus en gereduceerde triglyceride niveaus) van meervoudig onverzadigde vetzuren. Ook verklaard dit alles gedeeltelijk de verhoging van LDL cholesterol door een visolie behandeling in mensen en het potentieel verminderd HVZ risico van een hoge δ-5 desaturase activiteit. Momenteel zijn er nog veel vragen over de haalbaarheid van genetische risicovoorspelling voor HVZ. Clinici zouden door moeten gaan met navraag naar familiegeschiedenis van HVZ, omdat dit een simpele, goedkope en bruikbare risicofactor voor HVZ is, die waarschijnlijk het netto geïntegreerde effect van honderden genetische varianten vertegenwoordigd.

Acknowledgements

When looking back, I suddenly realized I have just gone through this long-dreamed PhD training process. Yes, it is hard but very rewarding. From being unrealistically optimistic at the beginning, slowly confidence-dipping day after day, till final confidence regaining with more publications, I have experienced all as a PhD student.

I should first thank **Professor Frans Kok** and **Dr. JK Zhao (Button)** who introduced Wageningen University to me in 2001 when I was preparing to pursue my further education. Yes, this country and Wageningen University gave me a chance to fulfill my biggest dream so far in my life. I should also thank **Professor Pieter van't Veer** and **Professor Ellen Kampman** who taught me the most essence of epidemiological research during my master training, which lets me know how to critically evaluate an epidemiological research. I should also thank **Ellen** who suggested me to pursue my 2nd master thesis in NMG lab at a critical moment, which seeded the chance of my PhD training here.

I should thank **Professor Edith JM Feskens** who gave me a chance to pursue what I want to do in my life. I have learned from her not only in academic skills in epidemiology and genetic epidemiology but also in how to deal with these unprepared situations when things get bad. Edith, thanks for your encouragement from time to time, and also tolerating my irrational complains sometimes.

I should thank **Professor Michael Müller** who gave me a chance to learn from him about molecular biology and nutrigenomics that really intrigued me the most when I first heard of Wageningen University. His ideas and knowledge have been inspiring me all the time and I will certainly benefit from them in my future academic career. I know it is not easy to combine metabolism with inflammation or immunology in genetics. I will try my best not to let you down.

I should thank **Dr. Jolanda MA Boer** for introducing RIVM and the people who work there to me, supervising my research in RIVM, teaching me how to prepare and write a good manuscript, and giving me suggestions and encouragement all those years. Thanks, Jolanda.

I should thank **Dr. Martijn ET Dollé** and **Sandra** for teaching me the bioinformatics and some skills in genetics and genotyping. **Sandra**, thanks for your hard work for APOE genotyping. Without your work, we could not get the manuscript of "genetics of total cholesterol levels" published. It also immediately made my record of rejected submissions under control ☺. I should also thank **Saskia** for introducing me the Doetinchem dataset that contributed three research articles to my PhD thesis, for teaching me how to use SAS matrix codes to do the multi-level analysis and sharing with me your methodology of your manuscript even before yours was published. I should also thank **Mariken** who helped me from time to time when I was working in RIVM and passed me some wonderful SAS loop codes and showed me that small magic SAS book. I should also thank **Henk** for giving me

suggestions when I encountered the computation problems in RIVM and for the nice talk and sincere encouragement during those NuGo weeks. Thanks, Henk.

I should thank **Dr. Mark Boekschoten** for supervising my PhD research in NMG and teaching me the bioinformatic and genomic skills, and how to use Ingenuity and other bioinformatic tools to analyse and visualize microarray data. Mark, I will definitely ask your help in the future on these skills. I feel I am still not good enough.

I should thank **Dr. Guido Hooiveld** who taught me molecular skills and how to think (as a molecule) in molecular science when you mentored my 2nd master thesis. You asked me the most important question in my academic life so far: "there are always difficulties in research, and the question is whether you like what you are doing or not". Thanks, Guido. I really asked myself several times. I am glad I struggled through these difficulties. I think I will definitely ask myself again and again when things do not go in my way in the future.

I should thank **Professor Sander Kersten** who taught me the most of the metabolism in his course of "advanced metabolic aspect of nutrition" that is one of the best courses I did here in Wageningen, together with "gene technology" and "nutritional genetics and genomics". I have really learnt a lot about genomics and lipid metabolism from you when you supervised my first genomic manuscript. Your expertise and experience in lipid metabolism and genomics make me immediately feel I was in a safe hand when working with you. Thanks, Sander.

I should thank **Dr. Lydia Afman** who asked me the most important questions when I was preparing the fat challenge research proposal although this project was cancelled at a late time. I think I will definitely ask myself these multiple "why" questions in the future when I am capable enough to win the research funding by myself.

I should thank **Cecile** for talking to me, informing me some news from time to time and helping me translate the Dutch letters, and also giving me suggestions when I was not very confident in myself when preparing those review papers. Thanks, Cecile, I think you will finish and defend your PhD research successfully and wonderfully next year. I should thank **Du** for agreeing to share an office with me and telling me some important things in PhD research. Yes, life is not easy for a foreign PhD student here alone. I should thank **Olga** for introducing me to these complicated haplotype-based association analysis and multiple testing concept and your wonderful PhD thesis. That really saved me a lot of time to crack out those basics in statistical genetics. I should also thank **Martinette** for teaching me some tricks in epidemiology, especially the last one (the choice of time unit in Cox regression model). You really saved my PhD thesis; otherwise, I could have more sleepless nights (worrying about the deadline), and also thank you for telling me many things about Wageningen and this country when we shored a corridor long time ago. I should also thank **Sandra** for sending me those forms for preparing my PhD defence, and good luck with your research and life in IARC. I also thank **Marielle and Ondine** for helping me out in Dutch language translation, manuscript submission, preparing for the travel to Tufts (not realized), and some casual talks from time to time.

I should thank **Geert** for teaching me randomForest and entropy-based interaction analysis. Your teaching saved me a lot of time in my first two years' PhD research. Unfortunately, I have not made that interaction manuscript published as I promised. I hope everything now is going on well with you in the States. I should thank **Ananda** for teaching me these wonderful Bayesian graphic networks. Sorry for my stubbornness. I still have not understood all the codes till now. I should thank **Ohid** for teaching me so many R skills and how to use PCA to do the multivariate analysis. I will definitely try this method in my future research. I also thank **Anastasia** and **Susanne** for telling me how to make those demanding figures in genomics, also **Anastasia** for sharing my complaining about the bad winter season.

I should also thank **Jan** who has been sharing an office with me in my last years of research here and for helping me out from time to time. You are so kind and considerate, tolerating my speechless working all the time. I should also thank **Lidwien** for informing me so many important things during my whole PhD research here. Thanks for your patience and time, and also, **Marie, Karen, Gabrielle, Cornelia, Gea,** and **Eric**. Thanks for the help you have given to me from time to time. I should also thank **Robert** and **Paul** who really pushed hard for these "precious" fatty acid data for my last manuscript. Thanks.

Finally, we would like to express my thanks to **Dr. Bastiaan T Heijmans** and **Professor Eline Slagboom** for giving me a chance to learn from them in genetics and **Anika** for working in LUMC together with me for this wonderful NHS project.

Last but not least, I should thank my beloved family members, especially my wife and my mum who support me and encourage me over these years, and my dear angel (my daughter), who has been the hope for me all the time.

Best!

Kevin

About the author

Yingchang (Kevin) Lu was born on October 4 1974 in Jianhu, P. R. China. After completing his senior high school education at Shanggang high School in Jianhu in 1993, he entered Shanghai Medical University (now affiliated with Fudan University) to receive his medical training. He graduated with an award of “one of the best graduates in Shanghai in 1998”. From 1998 to 2004, he worked in the Department of Chronic non-communicable disease of Jiangsu provincial Center for Disease control and Prevention, mainly involved in the population-based epidemiological researches. In 2004, he started his Msc training in the Division of Human Nutrition, Wageningen University. He was trained in both epidemiology and public health and nutritional physiology and nutritional genomics. His first master thesis was “Smoking and dietary factors in lung cancer risk in two counties in Jiangsu Province, China” supervised by Professor Pieter van’t Veer, and his second master thesis was “Research on iron bioavailability of foods using Caco-2 cell model” supervised by Dr. Guido Hooiveld. In 2007, he was appointed as a PhD fellow to the Division of Human Nutrition at the Wageningen University and Centre for Nutrition and Health at National Institute for Public Health and the Environment (RIVM) to perform research on “Lipid and fatty acid metabolism: integrated analysis of genetic variation, biomarkers and environmental factors to identify new classifiers of incident MI risk and to maximize biological understanding”. This research was supported by Netherland Heart Foundation and he was supervised by Professor Edith JM Feskens, Professor Michael Müller, and Dr. Jolanda MA Boer. After graduation in 2011, he started his post-doctoral training in the Department of Dermatology in UCSF, working on the genetics and genomics of psoriasis.

List of publications

Publication in English journals

Yingchang Lu, Mark V. Boekschoten, Suzan Wopereis, Michael Müller, Sander Kersten. Comparative Transcriptomic and Metabolomic Analysis of Fenofibrate and Fish Oil Treatments in Mice. Physiological Genomics (In press).

Lu Y, Feskens EJ, Boer JM, Imholz S, Verschuren WM, Wijmenga C, Vaarhorst A, Slagboom E, Müller M, Dollé ME. Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study. Atherosclerosis. 2010;213(1):200-5.

Lu Y, Feskens EJ, Dollé ME, Imholz S, Verschuren WM, Müller M, Boer JM. Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem Cohort Study. Am J Clin Nutr. 2010;92(1):258-65.

Lu Y, Feskens EJ, Boer JM, Müller M. The potential influence of genetic variants in genes along bile acid and bile metabolic pathway on blood cholesterol levels in the population. Atherosclerosis. 2010;210(1):14-27.

Lu Y, Dollé ME, Imholz S, van 't Slot R, Verschuren WM, Wijmenga C, Feskens EJ, Boer JM. Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations. J Lipid Res. 2008;49(12):2582-9.

Wu M, Zhao JK, Hu XS, Wang PH, Qin Y, **Lu YC** et al. Association of smoking, alcohol drinking and dietary factors with esophageal cancer in high-and low-risk areas of Jiangsu Province. World J Gastroenterol. 2006;12(11):1686-93.

Submitted:

Anika A.M. Vaarhorst, **Yingchang Lu**, Bastiaan T. Heijmans, Martijn E.T. Dollé, Stefan Böhringer, Hein Putter, Sandra Imholz, Audrey H.H. Merry, Marleen M. van Greevenbroek, J. Wouter Jukema, Anton P.M. Gorgels, Piet A. van den Brandt, Michael Müller, Leo J. Schouten, Edith J.M. Feskens, Jolanda M.A. Boer, P. Eline Slagboom. Literature-based genetic risk scores for coronary heart disease; the CAREMA prospective-cohort study. (Submitted).

Yingchang Lu, Jolanda MA Boer, Roza M Barsova, Olga Favorova, Michael Müller, Edith JM Feskens. TGFB1 genetic polymorphisms and coronary heart disease risk: a meta-analysis. (Submited).

Y Lu, A Vaarhorst, AHH Merry, MET Dolle, R Hovenier, S Imholz, LJ Schouten, BT Heijmans, M Muller, PE Slagboom, PA van den Brandt, AP Gorgels, JMA Boer, EJM Feskens. Endogenous desaturase activity and risk of coronary heart disease in the CAREMA cohort study. (Submitted).

Publications in Chinese Journals:

Lu YC, et al. Application of Geographic Information System in Spatial Distribution of Stroke in Jingsu Province. Chinese Journal of Chronic Disease Control and Prevention, 2005;13(3),103-7.

J Zhao, **Y Lu** et al. Establishment of geographic information system for distribution of water iodine content in Jiangsu. Chinese Journal of Endemiology. 2005;24(1),101-4.

J Zhao, M Wu, A Liu, L Shen, **Y Lu**, et al. Matched case-control study of esophageal cancer in a high risk county of Jiangsu Chinese. Journal of Chronic Disease Prevention and Control. 2005;13(1),17-19.
Lu YC, et al. Factors influencing Spatial and Geographic Distribution of Hypertension in Jiangsu Province. Journal of Chinese Epidemiology (Zhonghua Liu Xing Bing Xue Za Zhi).2004;25(7),673-9.
Lu YC, et al. Application of GIS Technology in Public Health. Journal of Foreign Medical Science, Section of Geographic Medicine. 2004;25(4),185-8.
Lu YC, et al. Advancement of Study on Tobacco Consumption. Journal of Foreign Medical Science, Section of Social Science. 2003; 20(1),29-32.
Lu YC, et al. Geographic Information System of Air Pollution and Mortality of Lung Cancer in Jiangsu Province. Bulletin of Chinese Cancer. 2003;12(7),374-7.
Wu M, Hu XS, Zhao JK, Wang PH, Qin Y, **Lu YC**. The mortality of common Malignancies in 1990s in Jiangsu Province. Bulletin of Chinese Cancer. 2003;12(7),378-80.
Qin Y, Hu XS, Zhao JK, **Lu YC**, et al. Regression analysis on liver cancer and geographic factors in Jiangsu Province. Bulletin of Chinese Cancer. 2003;12(11),630-2.
Lu YC, et al. Current Status and Distribution of Cancer Death in Northern Part of Jiangsu Province from 1996 to 1998. Bulletin of Chinese Cancer. 2001;10(9),512-4.
Wang XS, Sun RK, Li YC, Zhao JK, Hu J, **Lu YC**. Analysis of the mortality between 1996 and 1998 in Ganyu County. Journal of Chinese Public Health. 2000;17(1),69-70.

Overview of complete educational training

Discipline specific activities

Courses

Masterclass Nutrigenomics, Wageningen, The Netherlands, 2007
SNPs and Human Diseases, Rotterdam, The Netherlands, 2007
VLAG-NZO MasterClass "Diet and Cancer", Wageningen, The Netherlands, 2007
Advances in Population-based Studies of Complex Genetic Disorders, Rotterdam, The Netherlands, 2008
Systems Biology: "Statistical analysis of ~omics data", Wageningen, The Netherlands, 2008
Workshop: "Browsing Genes and Genomes with Ensemble IV", Rotterdam, The Netherlands, 2009
Translation of genetic variation to gene function in biology and epidemiology, Arlanda, Sweden, 2009
Bayesian Statistics, Wageningen, The Netherlands, 2009
Mendelian Randomization and Application of Bayesian Graphical Modelling in Genetic Epidemiology, Rotterdam, The Netherlands, 2010
Advanced Course: "Molecular Immunology", Rotterdam, The Netherlands, 2011
MasterClass "Multilevel Analysis", Wageningen, The Netherlands, 2011
Workshop "Browsing Gene and Genomes with UCSC", Rotterdam, The Netherlands, 2011

Meetings

Benelux Nuclear Receptor Meeting, Utrecht, The Netherlands, 2008
ICBL Conference on the Bioscience of Lipids, Maastricht, The Netherlands, 2008
NZO-WUR Symposium "state of the art in Human Nutrigenomics", Wageningen, The Netherlands, 2009
Symposium "Lipids and Insulin Resistance", Amsterdam, The Netherlands, 2009
Wageningen Nutritional Sciences Forum: Too much Too little, Arnhem, The Netherlands, 2009
Symposium "Macrophages, Lipid Handling, and Inflammation", Rotterdam, The Netherlands, 2009
7th Annual Center for Medical Systems Biology Symposium, Leiden, The Netherlands, 2010
The Generation R Symposium "Genetics in Child Cohort Studies", Rotterdam, The Netherlands, 2010
European Human Genetics Conference, Amsterdam, The Netherlands, 2011
NuGo Week, 2008, 2009 and 2011 (Potsdam, Germany; Montecatini, Italy; Wageningen, The Netherlands)

General Courses

VLAG PhD Week, Bilthoven, The Netherlands, 2008

Optional Activities

NMG Scientific Meetings (every week)