OPEN ACCESS Freely available online PLoS GENETICS

A Genome-Wide Screen for Interactions Reveals a New Locus on 4p15 Modifying the Effect of Waist-to-Hip Ratio on Total Cholesterol

Ida Surakka[1,2☯], Aaron Isaacs[3,4☯], Lennart C. Karssen[3], Pirkka-Pekka P. Laurila[1,2,5], Rita P. S. Middelberg[6,7], Emmi Tikkanen[1,2], Janina S. Ried[8], Claudia Lamina[9], Massimo Mangino[10], Wilmar Igl[11], Jouke-Jan Hottenga[12], Vasiliki Lagou[13,14], Pim van der Harst[15], Irene Mateo Leach[15], Tõnu Esko[16,17], Zoltán Kutalik[18,19], Nicholas W. Wainwright[20,21], Maksim V. Struchalin[3], Antti-Pekka Sarin[1,2], Antti J. Kangas[22], Jorma S. Viikari[23], Markus Perola[1,2,16], Taina Rantanen[24], Ann-Kristin Petersen[8], Pasi Soininen[25], Åsa Johansson[11], Nicole Soranzo[10,20], Andrew C. Heath[26], Theodore Papamarkou[20,21], Inga Prokopenko[13,14], Anke Tönjes[27,28], Florian Kronenberg[9], Angela Döring[29,30], Fernando Rivadeneira[31,32,33], Grant W. Montgomery[6], John B. Whitfield[6], Mika Kähönen[34,35], Terho Lehtimäki[35,36], Nelson B. Freimer[37,38], Gonneke Willemsen[12], Eco J. C. de Geus[12], Aarno Palotie[1,5,20,39], Manj S. Sandhu[20,21], Dawn M. Waterworth[40], Andres Metspalu[16,17,41], Michael Stumvoll[42], André G. Uitterlinden[31,32,33], Antti Jula[43], Gerjan Navis[44], Cisca Wijmenga[45], Bruce H. R. Wolffenbuttel[46], Marja-Riitta Taskinen[47], Mika Ala-Korpela[22,25,48], Jaakko Kaprio[1,49,50], Kirsten O. Kyvik[51,52], Dorret I. Boomsma[12], Nancy L. Pedersen[53], Ulf Gyllensten[11], James F. Wilson[54], Igor Rudan[54,55], Harry Campbell[54], Peter P. Pramstaller[56,57,58,59], Tim D. Spector[10], Jacqueline C. M. Witteman[31,32,33], Johan G. Eriksson[60,61,62,63,64], Veikko Salomaa[65], Ben A. Oostra[66], Olli T. Raitakari[67,68], H.-Erich Wichmann[29,69,70], Christian Gieger[8], Marjo-Riitta Järvelin[71], Nicholas G. Martin[6], Albert Hofman[31,32,33], Mark I. McCarthy[13,14,72], Leena Peltonen[1,2,5,18,39†], Cornelia M. van Duijn[3,4,33], Yurii S. Aulchenko[3¶], Samuli Ripatti[1,2,20¶*], for the ENGAGE Consortium

1 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, **2** Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland, **3** Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands, **4** Centre for Medical Systems Biology, Netherlands Genomics Initiative, Leiden, The Netherlands, **5** Department of Medical Genetics, Haartman Institute, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland, **6** Queensland Institute of Medical Research, Brisbane, Australia, **7** Department of Medicine, Prince Charles Hospital, Chermside, Australia, **8** Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, **9** Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria, **10** Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, **11** Department of Immunology, Genetics, and Pathology, University of Uppsala, Uppsala, Sweden, **12** Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands, **13** Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, Oxford, United Kingdom, **14** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **15** Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, **16** The Estonian Genome Center and the Center of Translational Genomics of the University of Tartu, Tartu, Estonia, **17** The Institute of Molecular and Cellular Biology of the University of Tartu, Tartu, Estonia, **18** Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland, **19** Swiss Institute of Bioinformatics, Lausanne, Switzerland, **20** Genetic Epidemiology Group, Wellcome Trust Sanger Institute, Hinxton, United Kingdom, **21** Non-Communicable Disease Research Group, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom, **22** Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland, **23** Department of Medicine, University of Turku and Turku University Hospital, Turku, Finland, **24** Department of Health Sciences, Gerontology Research Centre, University of Jyväskylä, Jyväskylä, Finland, **25** NMR Metabonomics Laboratory, Department of Biosciences, University of Eastern Finland, Kuopio, Finland, **26** Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, United States of America, **27** Medical Department, University of Leipzig, Leipzig, Germany, **28** IFB AdiposityDiseases, University of Leipzig, Leipzig, Germany, **29** Institute of Epidemiology I, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, **30** Institute of Epidemiology II, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany, **31** Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands, **32** Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands, **33** Netherlands Genomics Initiative (NGI)–sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden, The Netherlands, **34** Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland, **35** Medical School, University of Tampere, Tampere, Finland, **36** Department of Clinical Chemistry, Tampere University Hospital, Tampere, Finland, **37** Department of Psychiatry, University of California Los Angeles, Los Angeles, United States of America, **38** Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, United States of America, **39** The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, United States of America, **40** Genetics, Medicines Discovery, and Development, GlaxoSmithKline, Philadelphia, Pennsylvania, United States of America, **41** The Estonian Biocentre, Tartu, Estonia, **42** Department of Medicine, University of Leipzig, Leipzig, Germany, **43** Department of Chronic Disease Prevention, National Institute for Health and Welfare, Turku, Finland, **44** Division of Nephrology, Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, **45** Department of Genetics, University Medical Centre Groningen and University of Groningen, Groningen, The Netherlands, **46** Department of Endocrinology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, **47** Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland, **48** Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, University of Oulu, Oulu, Finland, **49** Department of Public Health, University of Helsinki, Helsinki, Finland, **50** Unit for Child and Adolescent Mental Health, National Institute for Health and Welfare, Helsinki, Finland, **51** Institute of Regional Health Services Research, University of Southern Denmark, Odense, Denmark, **52** Odense Patient data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark, **53** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **54** Centre for Population Health Sciences, University of Edinburgh Medical School, Edinburgh, United Kingdom, **55** Croatian Centre for Global Health, University of Split Medical School, Split, Croatia, **56** Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC),

Bolzano, Italy, **57** Affiliated Institute of the University of Lübeck, Lübeck, Germany, **58** Department of Neurology, University of Lübeck, Lübeck, Germany, **59** Department of Neurology, Central Hospital of Bolzano, Bolzano, Italy, **60** Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland, **61** Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland, **62** Folkhälsan Research Centre, Helsinki, Finland, **63** Vasa Central Hospital, Vasa, Finland, **64** Department of Health Promotion and Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, **65** Unit of Chronic Disease Epidemiology and Prevention, National Institute for Health and Welfare, Helsinki, Finland, **66** Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands, **67** Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, **68** Clinical Physiology, University of Turku and Turku University Hospital, Turku, Finland, **69** Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, **70** Klinikum Grosshadern, Munich, Germany, **71** Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom, **72** Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom

Abstract

Recent genome-wide association (GWA) studies described 95 loci controlling serum lipid levels. These common variants explain ~25% of the heritability of the phenotypes. To date, no unbiased screen for gene–environment interactions for circulating lipids has been reported. We screened for variants that modify the relationship between known epidemiological risk factors and circulating lipid levels in a meta-analysis of genome-wide association (GWA) data from 18 population-based cohorts with European ancestry (maximum $N = 32,225$). We collected 8 further cohorts ($N = 17,102$) for replication, and *rs6448771* on 4p15 demonstrated genome-wide significant interaction with waist-to-hip-ratio (WHR) on total cholesterol (TC) with a combined *P*-value of 4.79×10^{-9}. There were two potential candidate genes in the region, *PCDH7* and *CCKAR*, with differential expression levels for *rs6448771* genotypes in adipose tissue. The effect of WHR on TC was strongest for individuals carrying two copies of G allele, for whom a one standard deviation (sd) difference in WHR corresponds to 0.19 sd difference in TC concentration, while for A allele homozygous the difference was 0.12 sd. Our findings may open up possibilities for targeted intervention strategies for people characterized by specific genomic profiles. However, more refined measures of both body-fat distribution and metabolic measures are needed to understand how their joint dynamics are modified by the newly found locus.

Citation: Surakka I, Isaacs A, Karssen LC, Laurila P-PP, Middelberg RPS, et al. (2011) A Genome-Wide Screen for Interactions Reveals a New Locus on 4p15 Modifying the Effect of Waist-to-Hip Ratio on Total Cholesterol. PLoS Genet 7(10): e1002333. doi:10.1371/journal.pgen.1002333

Editor: Greg Gibson, Georgia Institute of Technology, United States of America

Received June 23, 2011; **Accepted** August 23, 2011; **Published** October 20, 2011

Funding: This research was supported through funds from The European Community's Seventh Framework Programme (FP7/2007–2013), ENGAGE Consortium, grant agreement HEALTH-F4-2007-201413. SR was supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (213506 and 129680), Academy of Finland (251217), the Finnish foundation for Cardiovascular Research, and the Sigrid Juselius Foundation. The ATFS cohort was funded by the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498), the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, DP0343921), the EU 5th Framework Programme GenomEUtwin Project (QLG2-CT-2002-01254), and the U.S. National Institutes of Health (AA07535, AA10248, AA11998, AA13320, AA13321, AA13326, AA14041, AA17688, DA12854, MH66206). RPSM and GWM are supported by National Health and Medical Research Council (NHMRC) Fellowship Schemes. Special Population Research Network (EUROSPAN) was supported b' European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947). In South Tyrol, the MICROS Study was supported by the Ministry of Health of the Autonomous Province of Bolzano and the South Tyrolean Sparkasse Foundation. The Vis Study in the Croatian island of Vis was supported through the grants from the Medical research Council UK to HC and IR and from the Ministry of Science, Education, and Sport of the Republic of Croatia to IR (number 108-1080315-0302). Erasmus Ruchpen Family (ERF) was supported by grants from The Netherlands Organization for Scientific Research (NOW; Pionier Grant), Erasmus MC, and the Netherlands Genomics Initiative (NGI)–sponsored Center for Medical Systems Biology (CMSB). The Northern Swedish Population Health Study (NSPHS) was funded by the Swedish Medical Research Council (Project Number K2007-66X-20270-01-3) and the Foundation for Strategic Research (SSF). The KORA research platform was initiated and financed by the Helmholtz Center Munich, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Part of this work was financed by the German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823) and by the ''Genomics of Lipid-associated Disorders - GOLD'' of the ''Austrian Genome Research Programme GEN-AU.'' The KORA research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. The Northern Finland Birth Cohort 1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269 (SALVE), 114194, and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland, NHLBI grant 5R01HL087679 through the STAMPEED program (1RL1MH083268-01), ENGAGE project and grant agreement HEALTH-F4-2007-201413, the Medical Research Council (grant G0500539, centre grant G0600705, PrevMetSyn), and the Wellcome Trust (project grant GR069224), UK. The genotyping of NFBC1966 was funded by NHLBI grant 5R01HL087679, the Academy of Finland, and Biocentrum Helsinki. Helsinki Birth Cohort Study has been supported by grants from Academy of Finland (project numbers 114382, 126775, 127437, 129255, 129306, 130326, 209072, 210595, 213225, 216374), Finnish Diabetes Research Society, Samfundet Folkhälsan, Juho Vainio Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Päivikki and Sakari Sohlberg Foundation, Signe and Ane Gyllenberg Foundation, and Yrjö Jahnsson Foundation. The Young Finns Study has been financially supported by the Academy of Finland (grants 126925, 121584, and 124282), Finnish Cultural Foundation, Emil Aaltonen Foundation, the Social Institution of Finland, Kuopio, Tampere (grants for TL and MK) and Turku University Hospital Medical Funds, Juho Vainio Foundation, Paavo Nurmi Foundation, Finnish Foundation of Cardiovascular Research (TL and OTR). The GenomEUtwin project is supported by the European Commission under the programme 'Quality of Life and Management of the Living Resources' of 5th Framework Programme (no. QLG2-CT-2002-01254). JK has been supported by the Academy of Finland Centre of Excellence in Complex Disease Genetics. The Swedish Twin Cohort has been financially supported by the Swedish Research Council and Swedish Foundation for Strategic Research. The Danish Twin Registry has been supported by the Danish Medical Research Council, the Danish Diabetes Foundation, the Danish Heart Association, and the Novo Nordic Foundation. The TWINSUK study was funded by the Wellcome Trust (Grant ref. 079771); European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-ENGAGE and Framework 6 Project EUroClot. The study also receives support from the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. NS acknowledges financial support from the Wellcome Trust (Grant 091746/Z/10/Z). NTR, NTR2, and NLDTWIN funding was obtained from the Netherlands Organization for Scientific Research (NWO: MagW/ZonMW): Genetic basis of anxiety and depression (904-61-090); Genetics of individual differences in smoking initiation and persistence (NWO 985-10-002); Resolving cause and effect in the association between exercise and well-being (904-61-193); Twin family database for behavior genomics studies (480-04-004); Twin research focusing on behavior (400-05-717); Genetic determinants of risk behavior in relation to alcohol use and alcohol use disorder (Addiction-31160008); Genotype/phenotype database for behavior genetic and genetic epidemiological studies (911-09-032); Spinozapremie (SPI 56-464-14192); CMSB: Center for Medical Systems Biology (NWO Genomics); NBIC/BioAssist/RK/2008.024); BBMRI –NL: Biobanking and Biomolecular Resources Research Infrastructure; the VU University: Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA); the European Science Foundation (ESF): Genomewide analyses of European twin and population cohorts (EU/QLRT-2001-01254); European Community's Seventh Framework Program (FP7/2007-2013): ENGAGE (HEALTH-F4-2007-201413); the European Science Council

(ERC) Genetics of Mental Illness (230374); Rutgers University Cell and DNA Repository cooperative agreement (NIMH U24 MH068457-06); Collaborative study of the genetics of DZ twinning (NIH R01D0042157-01A); the Genetic Association Information Network, a public–private partnership between the NIH and Pfizer, Affymetrix, and Abbott Laboratories. The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. PREVEND genetics is supported by the Dutch Kidney Foundation (Grant E033), the EU project grant GENECURE (FP-6 LSHM CT 2006 037697), the National Institutes of Health (grant LM010098), The Netherlands organisation for health research and development (NWO VENI grant 916.761.70), and the Dutch Inter UniversityCardiology Institute Netherlands (ICIN). EGCUT was supported by the Estonian Ministry of E&R (SF0180142s08), EU FP7 OPENGENE (#245536) and ENGAGE (201413), by EurRDF grant to the Centre of Excellence in Genomics, Estonian Biocentre and University of Tartu and by Estoinian Research Infrastructure's Roadmap. Genmets was supported through funds from The European Community's Seventh Framework Programme (FP7/2007–2013), BioSHaRE Consortium, grant agreement 261433. VS was supported by the Sigrid Juselius Foundation, Finnish Foundation for Cardiovascular research, and the Finnish Academy (grant number 129494). The CoLaus study received financial contributions from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (33CSCO-122661). EPIC–Norfolk is supported by programme grants from the Medical Research Council UK (G9502233, G0300128) and Cancer Research UK (C865/A2883). Adipose Tissue eQTL dataset has been funded by Finnish Foundation for Cardiovascular Research, Helsinki University, Sigrid Juselius Foundation Central Hospital Research Foundation. Metabonomic datasets are supported by the Academy of Finland (grant number 137870 to PS) and the Responding to Public Health Challenges Research Programme of the Academy of Finland (grant number 129429 to MA-K), the Finnish Cardiovascular Research Foundation (MA-K), and the Jenny and Antti Wihuri Foundation (AJK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: One of the replication cohorts used in the study is partially funded by GlaxoSmithKline.

* E-mail: samuli.ripatti@fimm.fi

◔ These authors contributed equally to this work.

¶ These authors also contributed equally to this work.

† Deceased.

Introduction

Serum lipids are important determinants of cardiovascular disease and related morbidity [1]. The heritability of circulating lipid levels is estimated to be 40%–60% and recent genome-wide association (GWA) studies implicated a total of 95 loci associated with serum high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC), and triglyceride (TG) levels [2]. Currently identified common variants explain 10%–12% of the total variation in lipid levels, corresponding to ~25% of the trait heritability [2].

Epidemiological risk factors, such as alcohol consumption, smoking, physical activity, diet and body composition are known to affect lipid levels [3–5]. These risk factors also show moderate to high heritabilities, and over 120 loci with genome-wide significant association have been identified (http://www.genome.gov/26525384). To better understand the biological processes modifying lipid levels, several twin studies [6–8] and candidate gene studies [9–14] have tested for interactions between genes and epidemiological risk factors.

Interactions between genes and modifiable risk factors might help us develop new lifestyle interventions targeted to susceptible individuals based on their genetic information. The effects of genetic loci and risk factors have been studied widely separately, but to date no GWA studies for interactions on lipids have been reported.

Results

We conducted a genome-wide screen for interactions between 2.5 million genetic markers and sex, lifestyle factors (smoking and alcohol consumption), and body composition (BMI and WHR) in association to serum lipid levels (TC, TG, HDL-C, and LDL-C) in 18 population-based cohorts (max $N = 32{,}225$; Table S1A, Text S1). We defined interaction as a departure from a linear statistical model allowing for the additive main effects of both the SNP and the epidemiological risk factor.

18 SNPs with suggestive interactions for at least one of the trait – epidemiological factor combinations (P-value for the interaction $<10^{-6}$) in stage 1 analyses were taken forward to stage 2 analysis in eight additional cohorts (max $N = 14{,}889$; Table S1B, Text S1). In inverse variance meta-analyses combining the results from stage 1 and stage 2 (Table S2), the interaction between *rs6448771* in chromosome 4p15 and WHR on TC (Figure 1) was statistically genome-wide significant (stage 1 and 2 combined $P = 9.08 \times 10^{-9}$). This interaction was tested in stage 3 in two further cohorts ($N = 7{,}813$; Table S1C, Text S1), which showed an effect to the same direction. After combining results from all three stages (total $N = 43{,}903$), the P-value for interaction was 4.79×10^{-9}. The association between WHR and TC was strongest for individuals carrying two G alleles of *rs6448771*, for whom a one standard deviation (sd) difference in WHR corresponds to 0.19 sd difference (confidence interval 0.13–0.25) in TC concentration, while for individuals homozygous for the A allele the difference was 0.12 sd (confidence interval 0.09–0.16) (Table S3A, Figure S1). The effect corresponds to 0.5% and 0.2% of the total variance explained in a cohort of young individuals (YFS, mean age = 37.6) and an old cohort (HBCS, mean age = 61.49), respectively. Additionally, when looking at the effect of the SNP on TC in WHR tertiles, the estimates differed in a way that the estimated SNP effect is higher for the individuals with higher WHR (Table S3B). The SNP did not have a direct effect on either TC or WHR ($P = 0.46$ and $P = 0.51$, respectively, Figure 1). The SNP *rs6448771* is located 249 kb downstream of the protocadherin 7 (*PCDH7*) gene.

Since the polymorphisms associated with complex phenotypes often influence gene expression, we examined whether individuals carrying different genotypes of *rs6448771* have variation in their transcript profiles. As WHR reflects adipose tissue function, we selected 54 individuals from Finnish dyslipidemic families with available fat biopsies and GWA data. We used linear regression to find genes that were differentially expressed in adipose tissue depending on the *rs6448771* genotype. We found two potential candidate genes with nominally significant cis-eQTL effects, *PCDH7* ($P = 0.027$, distance from the *rs6448771* 250 kb) and *CCKAR* ($P = 0.017$, distance from the SNP 4.9 Mb). The region with *CCKAR* has previously been linked with obesity [15].

Author Summary

Circulating serum lipids contribute greatly to the global health by affecting the risk for cardiovascular diseases. Serum lipid levels are partly inherited, and already 95 loci affecting high- and low-density lipoprotein cholesterol, total cholesterol, and triglycerides have been found. Serum lipids are also known to be affected by multiple epidemiological risk factors like body composition, life-style, and sex. It has been hypothesized that there are loci modifying the effects between risk factors and serum lipids, but to date only candidate gene studies for interactions have been reported. We conducted a genome-wide screen with meta-analysis approach to identify loci having interactions with epidemiological risk factors on serum lipids with over 30,000 population-based samples. When combining results from our initial datasets and 8 additional replication cohorts (maximum $N = 17{,}102$), we found a genome-wide significant locus in chromosome 4p15 with a joint P-value of 4.79×10^{-9} modifying the effect of waist-to-hip ratio on total cholesterol. In the area surrounding this genetic variant, there were two genes having association between the genotypes and the gene expression in adipose tissue, and we also found enrichment of association in genes belonging to lipid metabolism related functions.

Additionally, using Ingenuity software (IPA), we conducted a pathway analysis for genes with eQTL P-value<0.01 (both trans- and cis-eQTLs). Among other diverse IPA-defined biological functions, there was an eQTL association enrichment among genes belonging to the 'degradation of phosphatidylcholine' (3 genes out of 6, $P = 6.64\times10^{-5}$, Benjamini-Hochberg corrected $P = 0.0138$) and 'degradation of phosphatidic acid' (4 genes out of 8, $P = 4.71\times10^{-4}$, B-H corrected $P = 0.0349$) functions, which are members of broader defined IPA categories "Lipid Metabolism" and "Carbohydrate Metabolism". These pathways were up-regulated in individuals carrying the G allele of *rs6448771*, possibly indicating a role for *rs6448771* in lipid and carbohydrate metabolism.

The associated SNP also shows evidence for interactions with WHR on LDL-C (effect estimate for the interaction = 0.03, $P = 0.0016$) and HDL-C (effect estimate = 0.02, $P = 0.029$) in our stage 1 meta-analysis and after adjusting for TC no residual interaction effect on LDL-C and a little on HDL-C remains ($P = 0.834$ and $P = 0.131$ respectively) when testing in data subset. Therefore we tested the SNP – WHR interaction also on a range of lipoprotein subclasses measured using NMR metabonomics platform [16] available in two cohorts (NFBC1966, $N = 4624$ mean age = 31.0; YFS, $N = 1889$, mean age = 37.6). The results show that the SNP has a positive interaction effect on large HDL particle concentration (combined effect for the interaction = 0.538, $P = 0.0186$) and a negative effect on large very-low-density lipoprotein (VLDL) particles (combined effect = −0.466, $P = 0.0291$) and total triglycerides (combined effect = −0.454, $P = 0.0343$) (Figure 2).

Discussion

Our genome-wide scan for interactions between SNP markers and traditional epidemiological risk factors in population-based random samples found a genome-wide significant locus, *rs6448771*, modifying the relationship between WHR and TC. The effect of WHR is estimated to be 64% stronger for individuals carrying two copies of the G allele than for individuals carrying two A alleles. The interaction explains around half a percent of the TC variance that is in par with the main effects of the strongest previously identified TC SNPs individually. This SNP also shows similar interaction effects on a cascade of more detailed lipid fractions suggesting broad involvement in lipid metabolism, which was also suggested by our eQTL association enrichment analysis with adipose tissue expression data.

The eQTL analysis pointed towards two potential candidate genes in the region. The first one of these was protocadherin 7 (*PCDH7*) gene, which produces a protein that is thought to function in cell-cell recognition and adhesion. The other candidate gene, cholecystokinin A receptor (*CCKAR*) regulates satiety and release of beta-endorphin and dopamine in the central and peripheral nervous system. It has been previously shown that rats with no expressed *CCKAR*s developed obesity, hyperglycemia and type 2 diabetes [17]. To test whether our eQTL finding was adipose tissue specific, we ran the eQTL analysis for *PCDH7* and *CCKAR* in another dataset with genome wide expression data from blood leukocytes ($N = 518$) available. *CCKAR* could not be tested due to its negligible expression in blood leukocytes, and no association was found for the *PCDH7* (P-value = 0.284) gene most likely indicating an adipose tissue specific eQTL for *PCDH7* as a function of *rs6448771*.

One interesting aspect of this study, given our large sample size, is that only one signal achieved genome-wide significance, where previously published lipid GWA studies have found close to a hundred. Although power to detect interaction is typically lower than for main effects, especially for rare exposures and SNPs, several of the exposures considered here (such as WHR, BMI, and gender) were common and available for a large proportion of the study sample. This suggests that the contribution of two-way G×E interactions to lipid levels, at least for the risk factors we examined, is rather small, or that our current measures of risk factors may not be robust enough for identifying interactions. More specific measures of both phenotypes and interacting risk factors would give better statistical power in future screens of G×E interactions.

Our findings allow us to draw several conclusions. First, to our knowledge, this is the first time an interaction between a genetic loci and a risk factor has been identified in a genome-wide scan using a stringent statistical threshold for genome-wide significance. Second, in our samples, *rs6448771* modified the relationship between WHR and TC, but was not associated with either WHR or TC alone. This observation suggests that genome-wide screens for interactions may be complementary to the current large-scale GWAS efforts for finding main effects. Third, in addition to careful harmonization of both risk factor data and phenotypes, large sample sizes are needed to identify interactions. In our study, 43,903 samples were combined to robustly identify the interaction. Our data, however, suggest that the contribution of G×E interaction using current phenotypes appears limited. Finally, from clinical point of view, the interaction may open up possibilities for targeted intervention strategies for people characterized by specific genomic profiles but more refined measures of both body-fat distribution and metabolic measures are needed to understand how their joint dynamics are modified by the newly found locus.

Materials and Methods

Participating studies

18 studies, with a combined sample size of over 30,000 individuals, participated in the discovery phase of this analysis; 8 studies were available for replication with over 14,000 individuals. In the discovery stage, only population-based cohorts not

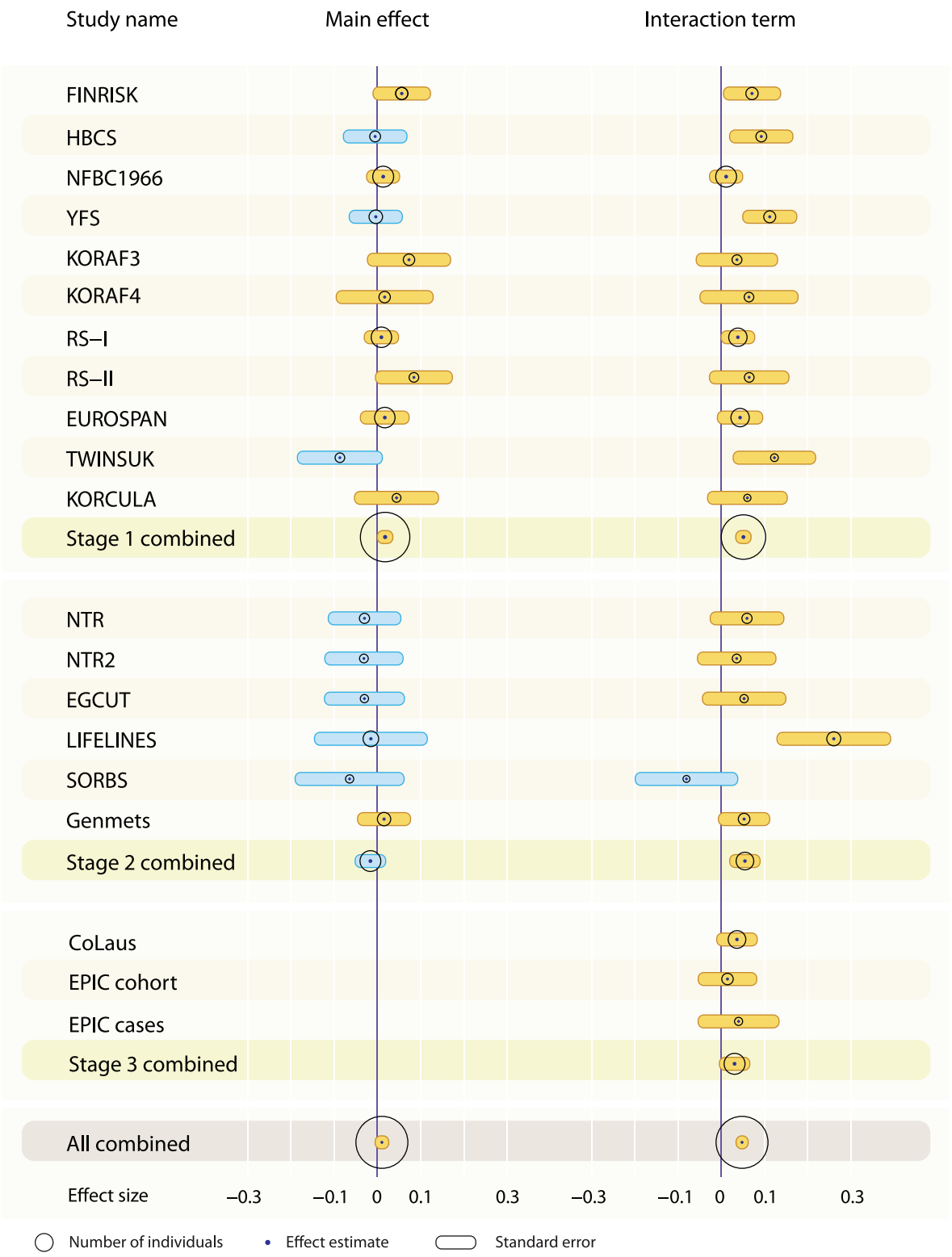

Figure 1. Forest plot of main and WHR interaction effect sizes of *rs6448771* on TC across the study cohorts. The circles in the plot are positioned at the effect estimates, betas, and the size corresponds to the number of individuals. The whiskers correspond to the standard errors of betas.
doi:10.1371/journal.pgen.1002333.g001

ascertained on the basis of phenotype, with a wide variety of well-defined epidemiological measures available, were included. In the replication datasets, the NTR cohort was selected on the basis of low risk for depression and the Genmets samples were selected for metabolic syndrome. In further replication of rs6448771, the EPIC cases were ascertained by BMI. Descriptive statistics for

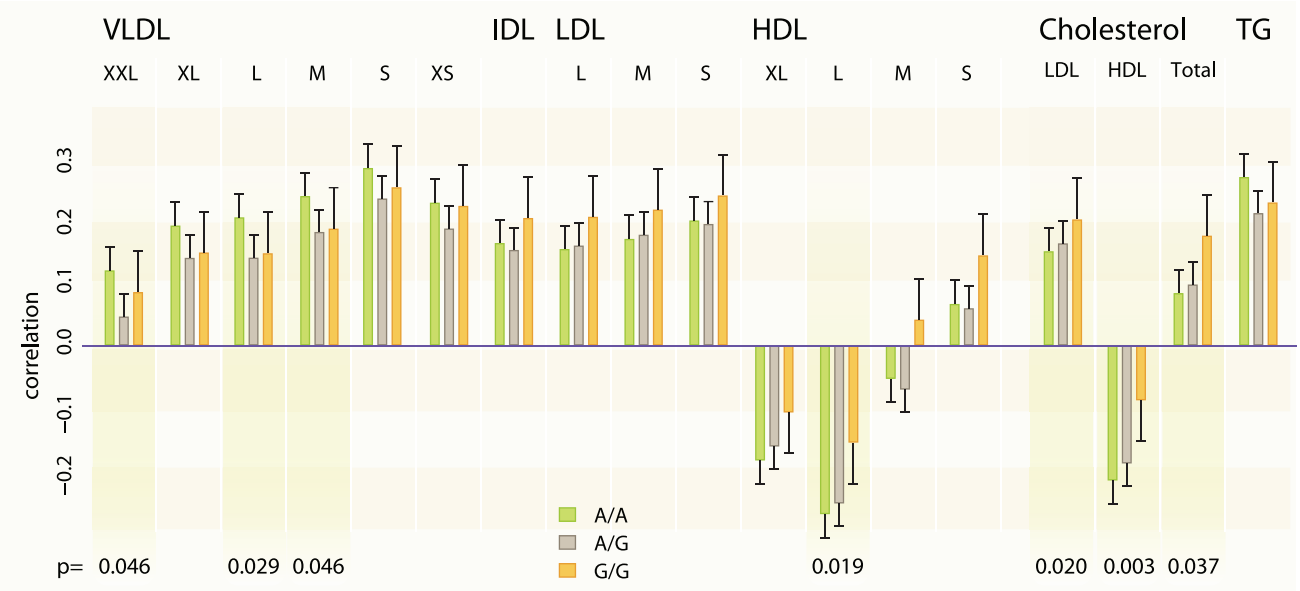

Figure 2. Lipoprotein subclass particle and key serum lipid concentration correlations with WHR for different genotypes of rs6448771. The height of the bar is the meta-correlation between the lipoprotein particle concentration and waist-to-hip ratio, and the whiskers correspond to standard error of the meta-correlation. The *P*-values have been taken from the interaction meta-analysis and only *P*-values<0.01 are shown in the figure. The two cohorts in which the lipid particle concentrations were measured with NMR metabonomics platform were YFS and NFBC1966 with combined number of samples of 6,500. XXL_VLDL: Chylomicrons and extremely large very low-density lipoprotein particles; XL: Very large, L: large, M: Medium, S: Small, XS: Very small; VLDL: very low-density lipoprotein; IDL: intermediate-density lipoprotein; LDL: low-density lipoprotein; HDL: High-density lipoprotein; TG: Triglycerides; TC: Total cholesterol.
doi:10.1371/journal.pgen.1002333.g002

these populations are detailed in Table S1A (discovery), S1B (replication) and S1C (further replication). Brief descriptions of the cohorts are provided in the Text S1 section "Short descriptions of the cohorts".

Phenotype determination

Individuals were excluded from analysis if they were not of European descent or were receiving lipid-lowering medication at the time of sampling. TC, HDL-C, and TG concentrations were measured from serum or plasma extracted from whole blood, typically using standard enzymatic methods. LDL-C was either directly measured or estimated using the Friedewald Equation (LDL-C = TC – HDL-C – 0.45×TG for individuals with TG≤4.52 mmol/l, samples with TG level higher than 4.52 were discarded in the calculation of LDL-C) [18].

Covariates and epidemiological risk factors were ascertained at the same time that blood was drawn for lipid measurements. BMI was defined as weight in kilograms divided by the square of height in meters. Waist circumference was measured at the mid-point between the lower border of the ribs and the iliac crest; hip circumference was measured at the widest point over the buttocks. Waist-to-hip ratio was defined as the ratio of waist and hip circumferences. Alcohol consumption and smoking habits were determined via interviews and/or questionnaires. Both behaviors were coded as dichotomous (abbreviations: ALC for drinker/abstainer and SMO for current smoker/current non-smoker) and semi-quantitative traits. Semi-quantitative alcohol usage (ALCq) was based on daily consumption in grams (0: 0 g/day; 1: >0 and ≤10 g/day; 2: >10 and ≤20 g/day; 3: >20 and ≤40 g/day; 4: >40 g/day). Semi-quantitative smoking (SMOq) was assessed based on the number of cigarettes per day (0: 0 cigarettes/day; 1: >0 and ≤10 cigarettes/day; 2: >10 and ≤20 cigarettes/day; 3: >20 and ≤30 cigarettes/day; 4: >30 cigarettes/day).

Genotyping and imputations

Affymetrix, Illumina or Perlegen arrays were used for genotyping in the discovery cohorts. Each study filtered both individuals and SNPs to ensure robustness for genetic analysis. After quality control, these data were used to impute genotypes for approximately 2.5 million autosomal SNPs based on the LD patterns observed in the HapMap 2 CEU samples. Imputed genotypes were coded as dosages, fractional values between 0 and 2 reflecting the estimated number of copies of a given allele for a given SNP for each individual. Cohort specific details concerning quality control filters, imputation reference sets and imputation software are described in Table S4.

In silico replication

Replication cohorts utilized genome-wide imputed data, as described above, where available. Details on the genotyping methods implemented in the replication samples are described in Table S4.

Serum NMR metabonomics, lipoprotein subclasses

Proton NMR spectroscopy was used to measure lipid, lipoprotein subclass and particle concentrations in native serum samples. NMR methods have been previously described in detail [16,19]. Serum concentrations of total triglycerides (TG), total cholesterol (TC) together with LDL-C and HDL-C were determined. In addition, total lipid and particle concentrations in 14 lipoprotein subclasses were measured. The measurements of these subclasses have been validated against high-performance liquid chromatography [20]. The subclasses were as follows: chylomicrons and largest VLDL particles (particle diameters from approx 75 nm upwards), five different VLDL subclasses: very large VLDL (average particle diameter 64.0 nm), large VLDL (53.6 nm), medium-size VLDL (44.5 nm), small VLDL

(36.8 nm), and very small VLDL (31.3 nm); intermediate-density lipoprotein (IDL) (28.6 nm); three LDL subclasses: large LDL (25.5 nm), medium-size LDL (23.0 nm), and small LDL (18.7 nm); and four HDL subclasses: very large HDL (14.3 nm), large HDL (12.1 nm), medium size HDL (10.9 nm), and small HDL (8.7 nm).

Statistical methods

Triglyceride concentrations were natural log transformed prior to analysis. BMI and WHR were transformed to normality using inverse-normal transformation of ranks. For analyses where sex was the epidemiological variable of interest, the phenotypes were defined as the rank-inverse normal transformed residuals resulting from the regression of the lipid measurement on age and age^2. For the other analyses, the phenotypes were defined as the inverse normal transformed residuals resulting from the regression of the lipid measurement on age, age^2, and sex.

Associations between the transformed residuals and epidemiological risk factors/SNPs were tested using linear regression models under the assumption of an additive (allelic trend) model of genotypic effect. The models regressed phenotypes on epidemiological factor, SNP, and epidemiological factor×SNP terms

$$\text{Transform(residuals)} \sim E + \text{SNP} + E \times \text{SNP}$$

and tested if the effect for E×SNP was 0 using 1 df Wald tests. In family-based cohorts, linear mixed modeling was implemented to control for relatedness among samples [21]. Analysis software used by the individual cohorts is described in Table S1A and S1B.

The interaction terms from the regression analyses were meta-analyzed using inverse variance weighted fixed-effects models [22]. Prior to meta-analysis, genomic control correction factors (λ_{GC}) [23], calculated from all imputed SNPs, were applied on a per-study basis to correct for residual bias possibly caused by population sub-structure. Meta-analyses were performed by two independent analysts using METAL (http://www.sph.umich.edu/csg/abecasis/Metal/index.html) and the R [24] package MetABEL (part of the GenABEL suite, http://www.genabel.org/). All results were concordant, reflecting a robust analysis. Results were selected for *in silico* replication if the meta-analysis *P*-value was less than 10^{-6}. Results passing the threshold of suggestive genome-wide association (*P*-value $\leq 5\times10^{-7}$) were selected for further replication by direct genotyping.

The commonly accepted genome wide level of significance (5×10^{-8}) reflects the estimated testing burden of one million independent SNPs in samples of European ancestry [25]. To address the multiple testing arising from testing interactions with multiple risk factors, we set the genome wide significance threshold to $5\times10^{-8}/3 = 1.67\times10^{-8}$ corresponding to three principal components explaining 97.8% of the total variation of the risk factors (Table S5).

Pathway analysis. The functional analyses were generated through the use of Ingenuity Pathways Analysis (Ingenuity Systems, www.ingenuity.com)." The Functional Analysis identified the biological functions and/or diseases that were most significant to the data set. Molecules which met the *P*-value cutoff of 0.01 for the rs6448771 – expression association in dataset of 54 Finnish individuals with both genotype and adipose tissue expression data, and were associated with biological functions and/or diseases in Ingenuity's Knowledge Base were considered for the analysis. Right-tailed Fisher's exact test was used to calculate a *P*-value determining the probability that each biological function and/or disease assigned to that data set is due to chance alone and Benjamini-Hochberg multiple test correction [26] was applied.

Supporting Information

Figure S1 Effect of waist-to-hip ratio on total cholesterol as a function of *rs6448771* genotypes. The bars in the plot are the effect estimates from three meta-analyzed linear models where total cholesterol (TC) has been explained using waist-to-hip ratio (WHR). The analyses were ran in three strata based on the *rs6448771* genotypes. The whiskers in the plot correspond to the confidence intervals of the effect estimates.
(DOC)

Table S1 Cohort characteristics. The number of study subjects with available phenotype and genotype (lower line) and summary statistics (upper line) for every cohort and trait. For continuous traits mean (standard deviation) is presented. For dichotomous traits number of individuals with phenotype present (%) is presented. TC: total cholesterol (mmol/l); HDL-C: high-density lipoprotein cholesterol (mmol/l); LDL-C: low-density lipoprotein cholesterol (mmol/l); TG: triglycerides (mmol/l); BMI: body-mass index; WHR: waist-to-hip ratio; NA: not available.
(DOC)

Table S2 Loci having *P*-value$<1\times10^{-6}$ in Stage 1 analyses and replication of the SNPs. Best SNP per locus having *P*-value$<1\times10^{-6}$ in the Stage 1 analysis combining 19 cohorts. The bolded number is the genome-wide significant *P*-value. *N*: number of individuals; *SE*: standard error of the effect estimate, Beta; LDL-C: low-density lipoprotein cholesterol; TC: total cholesterol; TG: triglycerides; HDL-C: high-density lipoprotein cholesterol; ALC: alcohol usage (drinker/abstainer); WHR: waist-to-hip ratio; BMI: body mass index; SMO: smoking (current/not);; SMOq: semi-quantitative smoking (0: 0 cigarettes/day; 1: >0 and ≤10 cigarettes/day; 2: >10 and ≤20 cigarettes/day; 3: >20 and ≤30 cigarettes/day; 4: >30 cigarettes/day); ALCq: semi-quantitative alcohol (0: 0 g/day; 1: >0 and ≤10 g/day; 2: >10 and ≤20 g/day; 3: >20 and ≤40 g/day; 4: >40 g/day).
(DOC)

Table S3 Effect of *rs6448771* on total cholesterol (TC) by waist-to-hip ratio (WHR) tertiles and effect of WHR on TC by SNP genotype classes. Section A shows the combined effect of waist-to-hip ratio (WHR) on total cholesterol (TC) stratified by the *rs6448771* genotype class from five Finnish cohorts (FINRISK, NFBC1966, YFS, Genmets and HBCS, combined number of individuals is 12,782) and section B shows the combined effect of the SNP on TC stratified by WHR tertiles from the same cohorts. The limit values for the waist-to-hip ratio (WHR) tertiles have been calculated using WHR values from all five datasets. Both analyses were ran using untransformed and standardized scales and were adjusted with age, age^2 and sex. Beta: effect estimate; *CI*: confidence interval.
(DOC)

Table S4 Details of GWA data in discovery and replication cohorts. QC: quality control; MAF: minor allele frequency; HWE: Hardy-Weinberg equilibrium.
(DOC)

Table S5 Proportions of variance explained by principal components. Principal components analysis (PCA) was run for the seven risk factors used in the screening. PC: Principal Component.
(DOC)

Text S1 Short descriptions of the cohorts and a full list of acknowledgements.
(DOC)

Acknowledgments

This manuscript is dedicated in memory of Prof. Leena Peltonen whose firm support and guidance had inspired this project immensely. The data annotation, exchange and deposition in public archives have been facilitated by the SIMBioMS platform (Krestyaninova et al, 2009)." A full list of acknowledgments is provided in the Text S1.

Author Contributions

Conceived and designed the experiments: LP CMD YSA SR. Performed the experiments: IS AI LCK. Analyzed the data: IS AI LCK PPL RPSM ET JSR CL MM WI JJH VL PH IML TE ZK NWW MVS APS AJK. Contributed reagents/materials/analysis tools: JSV MP TR AKP PS ÅJ NS ACH TP IP AT FK AD FR GWM JBW MK TL NBF GW EJCG AP MSS DW AM MS AGU AJ GN CW BHRW MRT MAK JK KOK DIB NLP UG JFW IR HC PPP TDS JCMW JGE VS BAO OTR HEW CG MRJ NGM AH. Wrote the paper: IS AI MIM CMD YSA SR.

References

1. Cooney M, Cooney H, Dudina A, Graham I (2010) Assesment of cardiovascular risk. Curr Hypertens Rep 12: 384–393.
2. Teslovich T, Musunuru K, Smith A, Edmondson A, Stylianou I, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466: 707–713.
3. Gaziano J, Manson J (1996) Diet and heart disease. The role of fat, alcohol, and antioxidants. Cardiol Clin 14: 69–83.
4. Bullen C (2008) Impact of tobacco smoking and smoking cessation on cardiovascular risk and disease. Expert Rev Cardiovasc Ther 6: 883–895.
5. Kraus W, Slentz C (2009) Exercise training, lipid regulation, and insulin action: a tangled web of cause and effect. Obesity (Silver Spring) 17 Suppl 3: S21–26.
6. Czerwinski S, Mahaney M, Rainwater D, Vandeberg J, MacCluer J, et al. (2004) Gene by smoking interaction: evidence for effects on low-density lipoprotein size and plasma levels of triglyceride and high-density lipoprotein cholesterol. Hum Biol 76: 863–876.
7. Greenfield J, Samaras K, Jenkins A, Kelly P, Spector T, et al. (2004) Do gene-environment interactions influence fasting plasma lipids? A study of twins. Eur J Clin Invest 34: 590–598.
8. Wang X, Ding X, Su S, Spector T, Mangino M, et al. (2009) Heritability of insulin sensitivity and lipid profile depend on BMI: evidence for gene-obesity interaction. Diabetologia 52: 2578–2584.
9. Sentí M, Aubo C, Bosch M (1998) The relationship between smoking and triglyceride-rich lipoproteins is modulated by genetic variation in the glycoprotein IIIa gene. Metabolism 47: 1040–1041.
10. Sentí M, Elosua R, Tomás M, Sala J, Masiá R, et al. (2001) Physical activity modulates the combined effect of a common variant of the lipoprotein lipase gene and smoking on serum triglyceride levels and high-density lipoprotein cholesterol in men. Hum Genet 109: 385–392.
11. Junyent M, Tucker K, Smith C, Garcia-Rios A, Mattei J, et al. (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.
12. Corbex M, Poirier O, Fumeron F, Betoulle D, Evans A, et al. (2000) Extensive association analysis between the CETP gene and coronary heart disease phenotypes reveals several putative functional polymorphisms and gene-environment interaction. Genet Epidemiol 19: 64–80.
13. Brand-Herrmann S, Kuznetsova T, Wiechert A, Stolarz K, Tikhonoff V, et al. (2005) Alcohol intake modulates the genetic association between HDL cholesterol and the PPARgamma2 Pro12Ala polymorphism. J Lipid Res 46: 913–919.
14. Marques-Vidal P, Bongard V, Ruidavets J, Fauvel J, Hanaire-Broutin H, et al. (2003) Obesity and alcohol modulate the effect of apolipoprotein E polymorphism on lipids and insulin. Obes Res 11: 1200–1206.
15. Arya R, Duggirala R, Jenkinson C, Almasy L, Blangero J, et al. (2004) Evidence of a novel quantitative-trait locus for obesity on chromosome 4p in Mexican Americans. Am J Hum Genet 74: 272–282.
16. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, et al. (2010) Metabonomic, transcriptomic, and genomic variation of a population cohort. Mol Syst Biol Dec 21: 441.
17. Moran T, Katz L, Plata-Salaman C, Schwartz G (1998) Disordered food intake and obesity in rats lacking cholecystokinin A receptors. Am J Physiol 273: R618–R625.
18. Friedewald W, Levy R, Fredrickson D (1972) Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 18: 499–502.
19. Soininen P, Kangas A, Würtz P, Tukiainen T, Tynkkynen T, et al. (2009) High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst 134: 1781–1785.
20. Okazaki M, Usui S, Ishigami M, Sakai N, Nakamura T, et al. (2005) Identification of unique lipoprotein subclasses for visceral obesity by component analysis of cholesterol profile in high-performance liquid chromatography. Thromb Vasc Biol 25: 578–584.
21. Aulchenko Y, Struchalin M, van Duijn C (2010) ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 16: 134.
22. de Bakker P, Ferreira M, Jia X, Neale B, Raychaudhuri S, et al. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 17: R122–128.
23. Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 600: 155–166.
24. R Development Core Team: R: A language and environment for statistical computing, Access date: 2010 Dec 13, http://R-project.org.
25. Pe'er I, Yelensky R, Altshuler D, Daly M (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32: 381–385.
26. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Static Soc Ser B 57: 289–300.