A GENETIC EPIDEMIOLOGIC STUDY OF LIPIDS AND DEPRESSIVE SYMPTOMS

AYŞE DEMİRKAN

Printing of this thesis was supported by CMSB and Erasmus University.

The work presented in this thesis was conducted iat the Genetic Epidemiology Unit, Departments of Epidemiology and Clinical Genetics, Erasmus University Medical Center, Rotterdam. The ERF study was supported by grants from the Netherlands Organisation for Scientific Research (Pionier, 047.016.009, 047.017.043), Erasmus MC and the Centre for Medical Systems Biology (CMSB; National Genomics Initiative).

ISBN 978-94-6169-323-5

A GENETIC EPIDEMIOLOGIC STUDY OF LIPIDS AND DEPRESSIVE SYMPTOMS

EEN GENETISCH EPIDEMIOLOGISCHE STUDIE VAN VETTEN EN DEPRESSIEVE SYMPTOMEN

Proefschrift

Ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.G.Schmidt
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 14 november 2012, 11:30 uur

door

Ayşe Demirkan
Geboren te Istanbul, Turkije

Promotoren

Prof.dr. Ir. C.M. van Duijn

Prof.dr. B.A. Oostra

Overige leden

Prof.dr. D.I. Boomsma

Prof.dr. T. Hankemeier

Prof.dr. H.W. Tiemeier

Co-promotor

Assoc. Prof.dr. Aaron Isaacs

"For small creatures such as we are, the vastness is bearable only through love"
Carl Sagan

To the beloved and far away

PUBLICATIONS AND MANUSCRIPTS BASED ON THE STUDIES DESCRIBED IN THIS THESIS

Chapter 2.1

Ayşe Demirkan, Brenda WJH Penninx, Karin Hek, Naomi R Wray, Najaf Amin, Yurii S Aulchenko, Richard van Dyck, Eco JC de Geus,Albert Hofman, André G Uitterlinden, Jouke-Jan Hottenga, Willem A Nolen, Ben A Oostra, Patrick F Sullivan, Gonneke Willemsen, Frans G Zitman, Henning Tiemeier, A Cecile JW Janssens, Dorret I Boomsma, Cornelia M van Duijn, Christel M Middeldorp. *Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry. Jul;16(7):773-783.*

Chapter 2.2

Ayşe Demirkan, Najaf Amin, Aaron Isaacs, Marjo-Riitta Jarvelin, John B Whitfield, Heinz-Erich Wichmann, Kirsten Ohm Kyvik, Igor Rudan ,Christian Gieger, Andrew A Hicks, Åsa Johansson, Jouke-Jan Hottenga, Johannes J Smith, Sarah H Wild, Nancy L Pedersen, Gonneke Willemsen, Massimo Mangino, Caroline Hayward, André G Uitterlinden, Albert Hofman, Jacqueline Witteman, Grant W Montgomery, Kirsi H Pietiläinen, Taina Rantanen, Jaakko Kaprio, Angela Döring, Peter P Pramstaller,Ulf Gyllensten, Eco JC de Geus, Brenda W Penninx, James F Wilson , Fernando Rivadeneria, Patrik KE Magnusson, Dorret I Boomsma, Tim Spector, Harry Campbell, Birgit Hoehne, Nicholas G Martin, Ben A Oostra, Mark McCarthy, Leena Peltonen-Palotie, Yurii Aulchenko, Peter M Visscher, Samuli Ripatti, A Cecile JW Janssens, Cornelia M van Duijn, for the ENGAGE CONSORTIUM. *Genetic architecture of circulating lipid levels. Eur J Hum Genet. Jul;19(7):813-819.*

Chapter 3.1

Ayşe Demirkan, Najaf Amin, Irina V Zorkoltseva, Aaron Isaacs, Antoine van der Sloot, Karin Hek, Lennart C Karssen, Peter Henneman, Michael A Nalls, Nese Direk, Henning Tiemeier, CHARGE Consortium Depression Work Group, UK Brain Expression Consortium, North American Brain Expression Consortium, Abdel Abdellaoui, Andre G Uitterlinden, Albert Hofman, Gert Jan B van Ommen, Mirjam C G M van den Hout, Gisela M Terwindt, Ko Willems van Dijk, Arn M J M van den Maagdenberg, Manfred Kayser, Dorret I Boomsma, Wilfred F J van IJcken, Yurii S Aulchenko, Tatiana I Axenovich, Ben A Oostra, Cornelia M van Duijn. *New candidates for depressive symptoms: Results from a quantitative linkage and exome sequencing in a Dutch centralized facility.(Submitted)*

Chapter 3.2

Karin Hek, **Ayşe Demirkan**, Jari Lahti, Toshiko Tanaka, Alexander Teumer, Marylin C Cornelis, Najaf Amin, Erin Bakshis, Jens Baumert, Jingzhong Ding, Yongmei Liu, Kristin Marciante, Osorio Meirelles, Michael A Nalls, Yan Sun, Nicole Vogelzangs, Lei Yu, Stefania Bandinelli, Emelia J

Benjamin, David A Bennett, Dorret Boomsma, Alessandra Cannas, Laura H Coker Eco de Geus, Philip L De Jager, Ana Diez-Roux MD[7], Shaun Purcell, Frank B Hu, Eric B Rimm, David J Hunter, Majken K Jensen, Gary Curhan, Kenneth Rice, Alan D Penman, Jerome I Rotter, Nona Sotoodehnia, Rebecca Emeny, Johan G Eriksson, Denis A Evans, Luigi Ferrucci, Myriam Fornage, Vilmundur Gudnason, Albert Hofman, Thomas Illig , Sharon Kardia, Margaret Kelly-Hayes, Karestan Koenen, Peter Kraft, Maris Kuningas, Joseph M Massaro, David Melzer, Antonella Mulas, Cornelis Mulder, Anna Murray, Ben A Oostra, Aarno Palotie, Brenda Penninx, Astrid Petersmann, Bruce Psaty, Rajesh Rawal, Eric M Reiman, Andrea Schulz, Joshua M Shulman, Andrew B Singleton, Albert Smith, Angelina R Sutin, André G Uitterlinden, Henry Völzke, Elisabeth Widen, Kristine Yaffe, Alan Zonderman, Francesco Cucca, Tamara Harris, Karl-Heinz Ladwig, David J Llewellyn, Katri Räikkönen, Antonio Terracciano, Cornelia M van Duijn, Hans J Grabe, Lenore J Launer, Kathryn L Lunetta, Thomas H Mosley, Anne B Newman, Henning Tiemeier, Joanne Murabito. *A Genome-Wide Association Study of Depressive Symptoms. (Accepted for publication in Biological Psychiatry)*

Chapter 3.3

Ayşe Demirkan, Jari Lahti, Nese Direk, Alexander Viktorin, Kathryn L Lunetta, Antonio Terracciano, Toshiko Tanaka, Michael A Nalls, Myriam Fornage, Jürgen Wellmann, Hanna M Ollila, Lei Yu, Aaron Isaacs, Aarno Palotie, Karin Hek, Wei V Zhuang, Alan Zonderman, Angelina Sutin, Osorio Meirelles, Antonella Mulas, Albert Hofman, Andre Uitterlinden, Fernando Rivadeneira , Markus Perola, Veikko Salomaa, Kristine Yaffe, Annemarie I Luik, Yongmei Liu, Jingzhong Ding, Paul Lichtenstein, Mikael Landén, Elisabeth Widen, Johan G Eriksson, Patrik KE Magnusson, Luigi Ferrucci, Thomas H Mosley, Francesco Cucca, Ben A Oostra, David A Bennett, Tiina Paunio, Klaus Berger, Tamara B Harris, Nancy L Pedersen, Joanne M Murabito, Henning Tiemeier, Cornelia M van Duijn, Katri Räikkönen. *Somatic complaints of depression and melatonin receptor MTNR1A gene. Results from the meta-GWAS of CHARGE Consortium. (Under preparation)*

Chapter 4.1

Ayşe Demirkan, Aaron Isaacs, Irina Zorkolsteva, Najaf Amin, Peter Ugocsai, Gerhard Liebisch, Rutger W W Brouwer, Wilfred F J van IJcken, Gerd Schmitz, Ben A Oostra, Tatiana Axenovich, Cornelia M van Duijn. *Genome wide linkage scan on plasma phospho-sphingolipids identifies multiple quantitative trait loci. (Under preparation)*

Chapter 4.2

Ayşe Demirkan, Cornelia M van Duijn, Peter Ugocsai, Aaron Isaacs, Peter P Pramstaller, Gerhard Liebisch, James F Wilson, Åsa Johansson, Igor Rudan, Yurii S Aulchenko, Anatoly V Kirichenko, A Cecile JW Janssens, Ritsert C Jansen, Carsten Gnewuch, Francisco S Domingues, Cristian Pattaro, Sarah H Wild, Inger Jonasson, Ozren Polasek, Irina V Zorkoltseva, Albert Hofman, Lennart C Karssen, Maksim Struchalin, James Floyd, Wilmar Igl, Zrinka Biloglav, Linda Broer, Arne Pfeufer,

Irene Pichler , Susan Campbell, Ghazal Zaboli, Ivana Kolcic, Fernando Rivadeneira, Jennifer Huffman, Nicholas D Hastie, Andre Uitterlinden, Lude Franke, Christopher S Franklin, Veronique Vitart, DIAGRAM Consortium, Christopher P Nelson, Michael Preuss, CARDIoGRAM Consortium, Joshua C Bis, Christopher J O'Donnell, Nora Franceschini, CHARGE Consortium, Jacqueline CM Witteman, Tatiana Axenovich, Ben A Oostra, Thomas Meitinger, Andrew A Hicks, Caroline Hayward, Alan F Wright, Ulf Gyllensten, Harry Campbell, Gerd Schmitz, on behalf of the EUROSPAN consortium. *Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. PLoS Genet. Feb;8(2):e1002490.*

Chapter 5

Ayşe Demirkan, Aaron Isaacs, Peter Ugocsai, Gerhard Liebisch, Maksim Struchalin, A Cecile JW Janssens, Gerd Schmitz, Ben A Oostra, Cornelia M van Duijn. *Plasma phosphatidylcholine and sphingomyelin concentrations are associated with depression and anxiety symptoms in a Dutch family-based lipidomics study.(Submitted)*

TABLE OF CONTENTS

1

General Introduction

INTRODUCTION

Discovery of lipids in human brain and plasma

In 1719, Hensing published the first monograph on the chemical composition of the brain, in which he isolated phosphorus from cerebral tissue. His work is considered a cornerstone of modern neurochemistry[1]. One century later, Vauquelin demonstrated that the phosphorus was actually bound to fat in the brain[2]. Meanwhile, cholesterol, a marker for all lipoproteins, was first discovered in bile and in gallstones by Poulletier de la Salle in 1769[3] and then rediscovered in 1815 by Chevreul, who named it "cholesterine"[4]. In 1823, Chevreul's work resulted in his masterpiece on lipids "*Recherches chimiques sur les corps gras d'origine animale*" where he described for the first time several fatty acids (margaric, oleic, stearic, butyric and caproic acids), including isovaleric acid (he named it "acide phocénique"), the first branched-chain fatty acid to be isolated from the head oil of dolphins and from porpoise oil[5]. Ten years later, in 1833, Boudet found cholesterol in blood[6]. Finally, in 1884, Johann Ludwig Wilhelm Thudichum published another fundamental work "*A treatise on the chemical constitution of the brain*", in which he described the presence of sphinosine in brain, but also the presence of a choline and sphingosine containing phospholipid without glycerol, which he named sphingomyelin. He additionally described the presence of cerebrosides and sulfatides in brain extracts and isolated cephalin (phosphatidylethanolamine) as a phospholipid distinct from lecithin (phosphatidylcholine).

The discovery of plasma lipoproteins, however, dates to one century later. In 1929, Michel Macheboeuf, working at the Pasteur Institute in Paris, reported the isolation of a stable, water-soluble lipoprotein from horse serum. This lipoprotein contained 59% protein and 41% lipids, which consisted of 18% cholesterol and 23% phospholipid[7] and it had the same composition as the α-lipoprotein that we now recognize as high-density lipoprotein (HDL). Today, we know that phospholipids constitute 7% of chylomicrons, 16% of very low density lipoproteins, 18% of lipoprotein-a and 22% of low density lipoprotein (LDL) particles, in addition to proteins, triglyceride and cholesterol. Meanwhile, the discovery of lipase enzymes, starting with Claude Bernard's description of pancreatic lipase in 1846, followed by the isolation of the first phospholipase by Bokay in 1877[8], helped to build the first hypothesis on fat transport throughout the body. Today, it is broadly accepted that HDL accomplishes reverse cholesterol transport from extrahepatic tissues to the liver, whereas LDL carries the majority of cholesterol through the bloodstream and delivers it to the cells of the body. The first association studies by Thannhauser and Magendantz on both atherosclerosis and xanthoma with high serum cholesterol levels followed in 1938[9].

After the discovery of docosahexaenoic acid (DHA) in fish oil in 1942, and other unsaturated fatty acids, a new numbering system; the "omega (ω) nomenclature" was proposed by Holman[10]. He also described a case of linolenic acid deficiency involving neuronal abnormalities; linolenic acid is recognized as essential for humans[11]. The beneficial effects of fish oils were

translated into the realm of cardiovascular health when experiments in dogs showed reduced thickening of veins[12].

Studies on the cell membrane, meanwhile, underlined the importance of lipid molecules in systems other than fat transport and energy accumulation. The discovery of lipid membrane asymmetry, and the final model of cell membrane architecture, known as the fluid mosaic model, was proposed by Singer[13]. This was followed by the discovery of ATP-dependent aminophospholipid-specific transporters (translocases), which transport phosphatidylserine and phosphatidylethanolamine from the outer to inner leaflet of plasma membranes[14]. Studies on cell membrane and receptor mediated pathways demonstrated the involvement of lipid mediators in signal transduction[15, 16], such as ceramide's role as a potent agent in apoptosis and arachidonoylethanolamide as a cannabinoid/anandamide receptor agonist[17, 18].

These findings were inevitably followed by association studies that demonstrated a tight connection between the intensity of blood ω-3 fatty acid depletion and the severity of depressive symptoms in patients[19, 20] and, finally, elevated sphingomyelinase activity in depressed patients[21]. Clinical trials showed that the ω -3 fatty acids from fish oil improved and stabilized mood in patients with bipolar disorder and that they improved cardiovascular health[22, 23]; this resulted in the current marketing and consumption of these species as daily supplements.

Depressive disorder

The first hypothesis that mood disorders would be due to imbalances in "body fluids" dates back to Hippocrates (460-377 BC), who favored a 'somatic' etiology of mental illness. According to him, melancholia (from the Ancient Greek μέλας (*melas*), "dark, black" + χολή (*kholé*), "bile") was a distinct disease with particular mental and physical manifestations. In his *Aphorisms*, he characterized all "fears and despondencies, if they last a long time" as being symptomatic of melancholia.

In modern medicine, this term mostly refers to Major Depressive Disorder (MDD), which will become the second leading source of disability across all ages by 2020, with a lifetime prevalence of 25%[24, 25]. The most widely used criteria for diagnosing depressive conditions are found in the American Psychiatric Association's revised fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR) and the World Health Organization's *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), which uses the name *recurrent depressive disorder*. Both the DSM-IV-TR and ICD-10 describe typical depressive symptoms. ICD-10 defines three (depressed mood, anhedonia, and reduced energy), two of which should be present to determine depressive disorder diagnosis. According to the DSM-IV-TR, there are two main symptoms: depressed mood and anhedonia. At least one of these must be present to diagnose a major depressive episode. A valid biomarker for depression still does not exist. This has hampered genetic research. Despite the standardized criteria for depression, it has been argued that there is not a cut–off of distinguishing those healthy from the diseased

ones. This has led to the proposal to study quantitative endophenotypes rather than discrete outcomes in psychiatric research. The work presented in this thesis focuses on symptoms of depression measured in a quantitative way, which enabled us to include individuals below the clinical threshold and made it possible to perform population wide studies[26].

The complex etiology of depression is not yet understood. The disease is more prevalent among women, with contributions from environmental factors such as stressful life events and educational level, and has also been shown to co-occur with anxiety disorders, dementia, cardiovascular problems and metabolic syndrome. Consistent evidence has shown the clustering of depression among families with moderate heritability; however, no major genes have been described thus far.

Aim of this thesis

The main aim of this thesis is to show possible causal pathways and candidate genes for depressive disorders while dissecting its association with circulating lipids as potential endophenotypes in addition to depressive symptoms.

Recent years have seen a great success in the field of complex trait genetics. This is mainly due to genome-wide association studies (GWAS), which provided important clues for understanding complex diseases. Common genetic variation, however, is not sufficient to explain the complete genetic variation of complex traits. Figure 1 shows three broad categories of risk alleles that are likely to be discoverable by genetic approaches in the near future: high-risk-rare

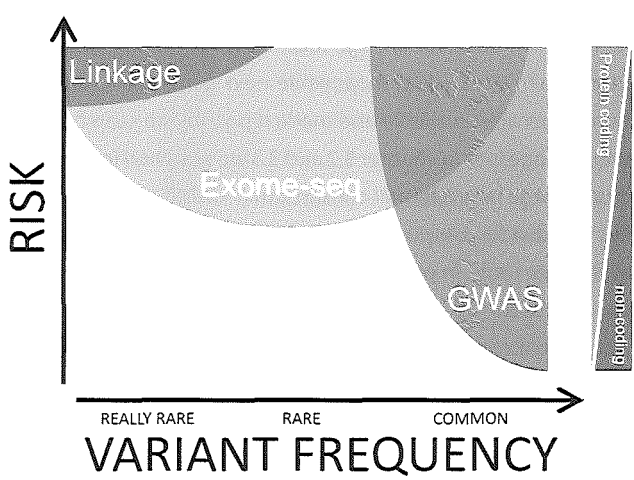


FIGURE 1. Genetic architecture.

Schematic landscape of the genetic architecture of disease and the applicability of current methods to finding risk and causative alleles. Exome-seq; whole-exome sequencing. Modified from the concept presented by Manolio and colleagues, adopted from Singleton, AB et al, 2010

alleles, which lead to Mendelian or near Mendelian disease, moderate-risk-low-frequency alleles, and low-risk-common alleles, which have modest effects on predisposition to disease[27, 28]. Understanding the genetic model behind a disease or a trait is of great importance for the design of gene discovery studies.

Chapters 2.1 and 2.2 focus on the genetic architecture of depressive symptoms and "classical" lipids carried by circulating lipoproteins (high density lipoprotein cholesterol, HDL-C; low density lipoprotein cholesterol, LDL-C; and triglycerides, TG). The genetic architecture of a complex trait is comprised of four factors: the number of risk alleles that contribute to the trait in the population, the frequency of those risk alleles, their effect sizes, and the way in which risk alleles act together, additively or with interaction[29]. In both studies, we tried to describe the traits with a score analysis based on the polygenic inheritance model. The standard polygenic model of biometrical genetics can be motivated by considering a quantitative trait determined by a large number of loci acting independently and additively[30].

Our studies summarized in chapters 3 and 4 follow similar designs for gene discovery and utilize a wide range of methods ranging from linkage, to association, to exome sequencing, and, finally, to meta-analysis in large consortia. Results from these studies were evaluated by gene-network and pathway enrichment analysis as well as gene expression databases from human and mouse. Chapter 3 mainly focuses on gene discovery for depressive symptoms and includes three studies in which different approaches for gene discovery were utilized. Linkage analyses of continuous measures of depressive symptoms resulted in wide peaks harboring several genes. For this reason, we combined linkage with association in order to fine-map these linkage peaks. This is followed by conditional linkage analysis which confirms if the associated variants are the reason for the linkage signal. The association panels that are used for GWAS tag only the common variation, however linkage method is for finding genes with large effects, which is more likely to be rare. Luckily, meanwhile next generation exome sequencing data for a part of our sample became available and made it possible to search also for rare missense mutations within our linkage loci. In chapter 3.2 and 3.3 we searched for depression genes using association data from a large international consortium (CHARGE). First, in chapter 3.2 we meta-analyzed 17 studies worldwide, comprising more than 50 000 participants in which depression was measured by different tools, in quantitative way. Finally, in chapter 3.3 we focused on the subscales of the widely used CES-D questionnaire, again within the same consortia. In this study we hypothesized that certain people would express their disease in different ways, leading to different symptom clusters such as somatic complaints of interpersonal problems; and that studying such clusters may be reliable phenotypes.

Gene discovery efforts for a wide range of phospho- and sphingolipid species are described in chapter 4. In total 148 lipid species were quantified by mass spectrometry methodology, including sphingomyelines, phosphatidylcholines, plasmalogens, phosphatidylethanolamines, lysophosphatidylcholines and ceramides. For the study in chapter 4.1 we performed quantitative linkage and estimated the heritability for each single metabolic trait measured. Chapter 4.2

summarizes the genome wide meta-analysis of EUROSPAN lipidomics data. This part of the thesis is unique considering the number of phenotypes studied (N = 356) and computation time spent for one single study. We have constructed separate kinship matrices for each of these traits and performed mixed model linear analysis on the 2.5 million HapMap imputed SNP data for five European isolates (ERF, SAMI, ORKNEY, VIS and TYROL) and finally meta-analyzed them all. This work could not be done without cloud computing facilities.

Chapter 5 includes one study in which we searched for phenotypic correlations between circulating phospho- and sphingolipids and depressive symptoms. There have been studies on the causal relationship of polyunsaturated fatty acids, especially ω-3 type. However, plain analyses of free fatty acids in plasma yields limited information. Using the mass spectrometry data of 148 lipid species we performed for the first time a detailed biochemical analysis of human circulating lipidome and its relation to depression and anxiety symptoms using correlation method. This study also includes a replication sample for the same lipids, measured on a different platform (BIOCRATES Life Sciences).

Finally, the results of the thesis are discussed in chapter 6. Chapter 7 includes a short summary of the thesis. The supplementary material generated during the thesis as well as extended methods that are too long to be considered for publication are collected in chapter 8.

REFERENCES

1. Hensing JT. Cerebri examen chemicum ex eodemque Phosphorus singularem omnia inflammabilia accendentem... Vulpius, Université de Giessen; 1719.
2. Vauquelin LN. Analyse de la matière cérébrale de l'homme et de quelques animaux. *Ann Mus Hist Nat* 1811;18:212.
3. Dam H. Historical introduction to cholesterol. In: Cook RP, ed. *Chemistry, Biochemistry and Pathology*. New York, NY. : Academic Press; 1958.
4. Chevreul ME. Recherches chimiques sur plusieurs corps gras, et particulièrement sur leurs combinations avec les alcalis. Cinquième mémoire. Des corps qu'on ont appelés adipocire, c'est-à-dire, de la substance cristallisée des calculs biliaires humains, du spermacéti et de la substance grasse des cadavres. *Ann. Chim.* 1815;95:5-50.
5. Carmichael EB. Michel Eugene Chevreul. Experimental chemist and physicist: lipids and dyes. *Ala J Med Sci.* Apr 1973;10(2):223-232.
6. Boudet MF. Nouvelle recherches sur la composition du serum du sang humain. *Ann. Chim. Phys.* 1833;52:337-348.
7. Macheboeuf M. *Bull Soc Chim Biol* 1929(11):485.
8. Bokay A. *Z physiol Chem* 1877;1:157.
9. Thannhauser SJ, Magendantz, H. *Ann. intern. Med.* 1938;11:1662.
10. Holman R. *Fed Proc* 1964;23:1062.
11. Holman RT, Johnson SB, Hatch TF. A case of human linolenic acid deficiency involving neurological abnormalities. *Am J Clin Nutr.* Mar 1982;35(3):617-623.
12. Landymore RW, Kinley CE, Cooper JH, MacAulay M, Sheridan B, Cameron C. Cod-liver oil in the prevention of intimal hyperplasia in autogenous vein grafts used for arterial bypass. *J Thorac Cardiovasc Surg.* Mar 1985;89(3):351-357.
13. Singer SJ, Nicolson GL. The fluid mosaic model of the structure of cell membranes. *Science.* Feb 18 1972;175(4023):720-731.
14. Seigneuret M, Devaux PF. ATP-dependent asymmetric distribution of spin-labeled phospholipids in the erythrocyte membrane: relation to shape changes. *Proc Natl Acad Sci U S A.* Jun 1984;81(12):3751-3755.
15. Hannun YA, Loomis CR, Merrill AH, Jr., Bell RM. Sphingosine inhibition of protein kinase C activity and of phorbol dibutyrate binding in vitro and in human platelets. *J Biol Chem.* Sep 25 1986;261(27):12604-12609.
16. Merrill AH, Jr., Stevens VL. Modulation of protein kinase C and diverse cell functions by sphingosine--a pharmacologically interesting compound linking sphingolipids and signal transduction. *Biochim Biophys Acta.* Feb 9 1989;1010(2):131-139.
17. Felder CC, Briley EM, Axelrod J, Simpson JT, Mackie K, Devane WA. Anandamide, an endogenous cannabimimetic eicosanoid, binds to the cloned human cannabinoid receptor and stimulates receptor-mediated signal transduction. *Proc Natl Acad Sci U S A.* Aug 15 1993;90(16):7656-7660.
18. Obeid LM, Linardic CM, Karolak LA, Hannun YA. Programmed cell death induced by ceramide. *Science.* Mar 19 1993;259(5102):1769-1771.
19. Adams PB, Lawson S, Sanigorski A, Sinclair AJ. Arachidonic acid to eicosapentaenoic acid ratio in blood correlates positively with clinical symptoms of depression. *Lipids.* Mar 1996;31 Suppl:S157-161.
20. Maes M, Smith R, Christophe A, Cosyns P, Desnyder R, Meltzer H. Fatty acid composition in major depression: decreased omega 3 fractions in cholesteryl esters and increased C20: 4 omega 6/C20:5 omega 3 ratio in cholesteryl esters and phospholipids. *J Affect Disord.* Apr 26 1996;38(1):35-46.

21. Kornhuber J, Medlin A, Bleich S, et al. High activity of acid sphingomyelinase in major depression. *J Neural Transm.* Nov 2005;112(11):1583-1590.
22. Stoll AL, Severus WE, Freeman MP, et al. Omega 3 fatty acids in bipolar disorder: a preliminary double-blind, placebo-controlled trial. *Arch Gen Psychiatry.* May 1999;56(5):407-412.
23. Ginty AT, Conklin SM. Preliminary evidence that acute long-chain omega-3 supplementation reduces cardiovascular reactivity to mental stress: a randomized and placebo controlled trial. *Biol Psychol.* Jan;89(1):269-272.
24. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Jama.* Jun 18 2003;289(23):3095-3105.
25. Murray CJ, Lopez AD. Evidence-based health policy--lessons from the Global Burden of Disease Study. *Science.* Nov 1 1996;274(5288):740-743.
26. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. *Mol Psychiatry.* May 22.
27. Singleton AB, Hardy J, Traynor BJ, Houlden H. Towards a complete resolution of the genetic architecture of disease. *Trends Genet.* Oct;26(10):438-442.
28. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* Oct 8 2009;461(7265):747-753.
29. Falconer DS. The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann Hum Genet.* Aug 1967;31(1):1-20.
30. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.* 1918;52:399-433.

2

Evaluating the genetic architecture of lipids and depression

Chapter 2.1

Genetic risk profiles for depression and anxiety in adult and elderly cohorts

Ayşe Demirkan, Brenda WJH Penninx, Karin Hek, Naomi R Wray, Najaf Amin, Yurii S Aulchenko, Richard van Dyck, Eco JC de Geus, Albert Hofman, André G Uitterlinden, Jouke-Jan Hottenga, Willem A Nolen, Ben A Oostra, Patrick F Sullivan, Gonneke Willemsen, Frans G Zitman, Henning Tiemeier, A Cecile JW Janssens, Dorret I Boomsma, Cornelia M van Duijn, Christel M Middeldorp.

Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry. Jul;16(7):773-783.

ABSTRACT

The first generation of genome-wide association studies (GWAS) for psychiatric disorders has led to new insights regarding the genetic architecture of these disorders. We now start to realize that a larger number of genes, each with a small contribution, are likely to explain the heritability of psychiatric diseases. The contribution of a large number of genes to complex traits can be investigated with genome-wide profiling. In a discovery sample a genetic risk profile for depression was defined based on a GWAS of 1738 adult cases and 1802 controls. The genetic risk scores were tested in two population based samples of elderly participants. The genetic risk profiles were evaluated for depression and anxiety in the Rotterdam Study cohort and the Erasmus Rucphen Family (ERF) study. The genetic risk scores were significantly associated to different measures of depression and explained up to ~0.7% of the variance in depression in Rotterdam Study and up to ~1 % in ERF study. The genetic score for depression was also significantly associated to anxiety explaining up to 2.1% in Rotterdam study. These findings suggest the presence of many genetic loci of small effect that influence both depression and anxiety. Remarkably, the predictive value of these profiles was as large in the sample of elderly participants as in the middle-aged samples.

INTRODUCTION

Genetic factors play an important role in the susceptibility to depression. A meta-analysis of twin studies on Major Depressive Disorder (MDD) estimated the heritability at 37%[1]. However, the success of studies aiming to find genes underlying the vulnerability for depression has been limited. An overview of promising results of linkage studies on MDD and neuroticism, a related personality trait, shows some overlap in regions of interest, but, so far, no single locus has been identified[2]. Candidate gene studies, mostly focusing on genes involved in neurotransmitter circuits or in reactions to stress, have also not been able to unambiguously identify a genetic variants explaining differences in the vulnerability for depression[3, 4].

An important issue in research on the etiological factors of MDD has been the frequent comorbidity with anxiety disorders. In the National Comorbidity Survey Replication, 59% of the subjects with a lifetime diagnosis of MDD also fulfilled the criteria for a lifetime anxiety disorder diagnosis.[5] A review of twin and family studies indicated that this comorbidity might be explained by shared, mostly genetic factors.[6] Still, an overview of promising results of linkage studies of anxiety only showed one overlapping region of interest with MDD and neuroticism[2].

The recent success of genome-wide association studies (GWAS) has fuelled expectations on finding genes for MDD. One of the first GWAS of depression showed evidence for the role of the presynaptic protein piccolo (*PCLO*) gene[7, 8].Recently, this result was replicated with a *P*-value of 2×10^{-9} in a meta-analysis of the results in three population based samples but not when five clinical samples were also included[7, 9]. However, the first GWAS of MDD as well as those of other psychiatric phenotypes have also demonstrated that genome-wide significant findings are rare and explain a small part of heritability[7, 10, 11]. This might be due to a lack of power. The Genetic Association Information Network (GAIN) MDD-GWAS, for example including ~1700 cases and 1800 controls, had 80% statistical power to detect relative risks of 1.59, 1.40 and 1.35 with a *P*-value of 1×10^{-7}, for minor allele frequencies of 0.10, 0.25 and 0.40. This is well comparable to other first generation GWAS. However, the first results of GWAS suggest that the strongest odds ratios may be < 1.2[12]. Another explanation for the scarce genome-wide significant findings could be that there is not a distinct number of genes for MDD with moderate to small risks but rather a large number of variations spread over the genome, each with small effects. Such a polygenic model predicts that the more markers are used, the better the disease is predicted and it implies that everybody carries risk variants but patients carry more than non-diseased people. We examined if a polygenetic component influences MDD implying that a large number of genetic variants are involved in explaining its heritability.

The evidence for a polygenic origin has recently been examined for schizophrenia and the hypothesis of a polygenic component was directly tested using GWAS data[13]. In this approach, the joint effect of multiple single nucleotides (SNPs) is tested rather than the effects of individual SNPs. These individual SNPs are not required to reach a genome-wide significance level by themselves. This approach aims to test whether the genetic disease liability reflects, at least

in part, the additive effect of a large number of variants spread across the genome whose joint action may be captured in a genome-wide genetic risk score. To obtain this risk score, a discovery set is used to select SNPs based on specific *P*-value thresholds (e.g. 0.0001, 0.001, 0.01 and so on) from a genome-wide scan for the disease of interest. In the target samples, genetic risk scores are calculated for each individual for each set of SNPs. The selected SNPs will contain false positives but if they are enriched with true associated variants with low effect size, then the genetic risk score might still be significantly associated to the disease in an independent sample. The problem is to distinguish truly associated SNPs from the false positives which occur massively around liberal *P*-value thresholds. In the schizophrenia study, the genetic risk scores based on the multiple SNPs in the discovery sample were associated to schizophrenia in three independent samples. The variance explained by the risk scores increased as more SNPs were included, i.e. risk scores based on SNPs that had a *P*-value below 0.5 in the discovery sample explained more variance in the replication sample than the risk scores based on SNPs selected at *P*-value below 0.1[13]. Moreover, the genetic risk scores for schizophrenia were also significantly associated with bipolar disorder assessed in two samples suggesting that the genes influencing schizophrenia and bipolar disorder partly overlap.

The study applied the risk score approach to investigate whether a polygenic component can be detected for depression and whether this polygenic component also influences anxiety. As the study samples differ in age, differences in the effect of the polygenic component may indicate that the genetic factors influencing depression and/or anxiety differ across the lifespan. Twin studies have already shown that the relative influence of genetic factors for depression decreases with age[14-16], but that the genes influencing depression remain the same across the years[17]. This can be investigated directly in this study.

The discovery set consisted of the GAIN-MDD sample, including 1738 cases and 1802 controls[7, 18] with over 400 000 SNPs genotyped. The target sets consisted of two independent Dutch samples. The first sample was based on the Rotterdam Study cohort and consisted of 178 patients with DSM-IV defined depressive disorder and 915 controls at low liability of depression as well as 222 cases for anxiety and 290 controls at low liability for anxiety. The second target sample was the Erasmus Rucphen Family (ERF) study in which symptoms of anxiety and depression were measured in 1886 participants. The subjects in the GAIN-MDD sample and the ERF sample were around 45 years of age. The Rotterdam sample was an elderly sample with a mean age of around 70. Height and intraocular pressure (IOP), phenotypes unrelated to psychiatric disorders, were additionally investigated to examine if the association with the genetic risk scores was specific to depression and anxiety.

MATERIAL AND METHODS

Discovery Sample

The discovery sample consisted of subjects from two large-scale longitudinal studies: the Netherlands Study of Depression and Anxiety (NESDA)[19] and the Netherlands Twin Registry (NTR)[20]. The chances of success of genetic risk score analyses depend primarily on the size of the discovery or training set. If the sample size is too small, the risk profiles will be based on random noise and are not expected to explain variance in the target set. To increase the chances of success, the power of the discovery set should therefore be maximized[21]. The size of the GAIN-MDD study made it more suitable to be used as the discovery set than the Rotterdam and ERF studies, which thus supplied the target samples. NESDA and NTR studies were approved by the Medical Ethics Committee of all participating institutes. Collection of the phenotype data and quality control of the genotype data as well as the statistical methods are described in detail elsewhere[7, 18]. In brief, inclusion criteria for MDD cases were a lifetime diagnosis of DSM-IV MDD[22] assessed with the Composite International Diagnostic Interview[23], age of 18-65 years and self-reported western European ancestry. Persons who were not fluent in Dutch and those with a primary diagnosis of schizophrenia or schizoaffective disorders, obsessive-compulsive disorder, bipolar disorder or severe substance use dependence were excluded. Inclusion criteria for control subjects were availability of biological samples and survey data with assessments of depression, anxiety and neuroticism, no report of MDD at any measurement occasion and low genetic liability for MDD based on the survey data. In addition, controls and their parents were required to have been born in the Netherlands or Western Europe. Only one control per family was selected. The cases and controls were carefully matched on age and sex.

The genotypic data used in the discovery sample were part of one of the six initial GAIN studies sponsored by the Foundation for the NIH[20]. Individual genotyping was conducted by Perlegen Sciences (Mountain View, CA, USA) using a set of four proprietary, high-density oligonucleotide arrays. The SNPs on these arrays were selected to tag common variation in the HapMap European and Asian panels. Of the 3820 samples sent to Perlegen, genotypes were delivered for 3761 samples (98.5%) of which 3540 subjects passed quality controls and were available in the final analysis dataset including 1738 MDD cases and 1802 controls. The SNP quality control process and the precautions against population stratification are detailed in Sullivan et al.[7]. A total of 427 037 SNPs on chromosome 1 to chromosome 22 met all selection criteria and were included in the final association analyses, which were performed in PLINK[24].

Target Samples

Rotterdam Study

The Rotterdam Study is a prospective cohort study in the district Ommoord of Rotterdam[25]. In 1990 all inhabitants aged 55 years and over were invited and 7983 persons agreed to participate. The Medical Ethics Committee of the Erasmus Medical Center, Rotterdam, approved the study. Written informed consent was obtained from all participants.

Ascertainment of depressive symptoms and incident depressive disorders was described previously[26]. Depression and anxiety symptoms were assessed with the Center for Epidemiological Studies Depression Scale (CES-D) and Hospital Anxiety and Depression Scale (HADS). The CES-D scale consists of 20 items with scores ranging from 0 to 60. A score of 16 or higher on the CES-D is considered indicative of a depressive disorder. The HADS-Depression (HADS-D) and HADS-Anxiety (HADS-A) scales each consist of 7 items with scores ranging from 0 to 21 with higher scores indicating more symptoms of depression. These questionnaires are valid and reliable self report measures of symptoms of depression[27, 28] .The HADS was assessed during the second visit in a randomly selected subgroup of individuals (N = 2231). Depression was measured with the CES-D 3 times during the follow-up.

Among 7983 subjects who agreed to participate, 5974 were successfully genotyped, 524 died before depression screening and 747 did not participate in depression screening. In the remaining sample, 587 persons scoring higher than 16 on the CES-D in the third or fourth visit were invited for a semi-structured interview with the Present State Examination[29] by a clinician. In addition, general practitioner records and specialist letters were surveyed actively for the occurrence of depression. Furthermore, physicians conducted repeated interviews to assess self-reported depression in the interval period. This effort identified 178 persons with current depressive disorder (145 MDD, 15 dysthymia and 18 depression-not otherwise specified cases) and eligible genotype data. The control group consisted of 915 persons, who scored in the lowest quartile (CES-D="0") on CES-D in the third visit (N = 3879) and who did not report any depressive symptoms during the follow up. Anxiety disorders were assessed during the fourth visit in the total sample by trained lay-interviewers who conducted a slightly adapted version of the Munich Composite International Diagnostic Interview (M-CIDI) to obtain DSM-IV diagnoses of generalized anxiety disorder (GAD), panic disorder, agoraphobia, social phobia and specific phobia. Obsessive compulsive disorder and post-traumatic stress disorder were not included. The current sample is selected from the 2779 persons who had valid M-CIDI assessment and genotype data. Out of 2779, 222 persons were anxiety disorder cases. The control group consists of 290 persons who did not have any anxiety disorder and scored in the lowest quartile (HADS-A = 0) on the HADS-A measured in 1322 persons of the interviewed and sample with eligible genotype data during the second visit.

Genome-wide SNP data were available from the Illumina HumanHap550K array for all cases and controls. Data were excluded based on call rate < 97.5%, sex mismatch, excess autosomal

heterozygosity, and outliers identified by the clustering analysis. MACH 1.0 software (v1.0.16)[30, 31] was used to impute to ~2.54 million SNPs based on the HapMap CEU phased haplotypes (release 22). SNPs included in imputation met thresholds of MAF ≥ 1%, HWE *P*-value ≥ 10^{-6} and SNP call rate ≥ 98.0%. GWA analysis of MDD was performed with Mach2Dat (logistic regression on allele dosage) using the GRIMP interface[30, 32]. Age and sex were included as covariates in the analysis.

ERF Study

The Erasmus Rucphen Family study is part of the Genetic Research in Isolated Population program. The study population essentially consists of one extended family of descendents from 20 related couples who lived in the isolate between 1850 and 1900 and had at least 6 children baptized in the community church. The detailed information about ERF isolate can be found elsewhere[33-35]. The Medical Ethical Committee of the Erasmus Medical Center, Rotterdam approved the study and informed consent was obtained from all participants.

Symptoms of depression and anxiety were assessed using the HADS and the CES-D[27, 28] in 2385 participants who also underwent an extensive medical examination. Among 2385 persons, high density genotype data were available for 1886 subjects. Data on height and IOP were collected during the medical examination. The height of participants was determined using a stadiometer and bilateral IOP measurements were performed using Goldmann applanation tonometry[36].

The genotype data were available for this population on 4 different genotyping platforms which were Illumina 6k, Illumina 318K, Illumina 370K and Affymetrix 250K (Affymetrix, Inc., Santa Clara, CA, USA). These were then merged and ~ 2.54×10^6 SNPs were imputed using MACH 1.0 software (v1.0.16)[30, 31], using build 36 HapMap (release 22) CEU population as reference. Within each genotyping batch, only SNPs showing a call rate >98%, MAF>1% and HWE *P*-value > 10^{-6} were used for imputations. As the ERF study included related individuals, GWA analyses were performed using a mixed model by the "mmscore" option in GenABEL software[37] which combines the FASTA method of Abecasis et al[38], and kinship matrix estimated from genotyped SNPs[39].

Risk Score Profiling

The score profiling method tests the association of a genetic score variable that reflects a combined effect of a number of selected SNPs with a trait. For a more detailed description of the method, we refer to Purcell et al[13]. In brief, SNPs were selected using the results from the GAIN-MDD GWAS[7] (the "discovery sample"). These sets of SNPs were used to calculate the genetic scores in the target samples. SNPs were selected on the basis of their nominal *P*-value ($P_{discovery}$) for association with MDD in the discovery sample. Genetic risk scores were calculated for $P_{discovery}$ thresholds ranging from 0.00001 to 1.0. Only those SNPs were included that were

directly genotyped in the discovery set (N = 427 049 SNPs). To avoid ambiguity A/T –G/C SNPs were excluded. Since an A to T or G to C change will result in the same nucleotides on the opposite strands, this change might be missed during the genotype analysis. SNPs for which the quality of imputation had an $R^2 < 0.95$ in target samples were also excluded. After all quality checks and exclusions, a total of 181 582 SNPs that were available in both ERF and Rotterdam study samples were selected for calculations of genetic risk scores. For each individual in the two independent target samples, the genetic score was calculated by multiplying the number of risk alleles per SNP (0, 1 or 2) with the log odds ratio, summed over all SNPs in the considered set of SNPs[40]. We calculated individual scores for each set of SNPs using the PLINK (v1.06) software[24].

Logistic regression models were used to test the association of the individual genetic risk scores for depressive and anxiety disorders. Linear regression models were used to test the association between genetic risk scores and the total CES-D, HADS-D and HADS-A scores as well as for height and IOP. Sex and age were used as covariates. As an alternative control for false positivity, 10% of the non-associated cluster of SNPs with $P_{discovery} > 0.9$ (N = 1569 SNPs) in the discovery set was selected and used for computing the risk profile in both target samples. For the Rotterdam Study regression analysis were performed in SPSS 15.0 for Windows (SPSS Inc., Chicago, IL, USA). As the ERF sample includes relatives, data are not independent, which can lead to biased estimates of standard errors and test statistics if this dependency between measures is not taken into account[41]. Association analysis of genetic risk score and the traits in ERF population were performed in SOLAR (Sequential Oligogenic Linkage Analysis Routines) 4.1.5 software package (Southwest Foundation for Biomedical Research, San Antonio, Texas, USA)[42] using the "polygenic" option to adjust for pedigree kinship. Among 1886 people both genotyped and phenotyped, 1697 were clustered into pedigrees (using the pedigree splitting algorithm PedCut[43]) and included in the family-based analysis. The remaining persons were not included in the analysis because they were also (distantly) related. The difference of the explained variance in the null and alternative model was considered as the variance explained by the genetic score. A genetic risk score with a P-value < 0.05 in the model was considered as significantly associated to the trait.

RESULTS

Table 1 shows the descriptive data for the case-control studies of GAIN-MDD and the Rotterdam Study, and Table 2 for the ERF study. Since in the Rotterdam study subjects were ascertained on the basis of age 55 or more, the mean age was 74 years. This was higher than in the GAIN-MDD and ERF study in which the mean ages were around 45 years. Level of education was higher in the GAIN-MDD sample than in the other two samples. In the target samples, subjects diagnosed with a depressive disorder or an anxiety disorder were more often women and were older. In the discovery sample cases and controls were matched based on age and sex.

TABLE 1. Descriptive data of case-control samples

	GAIN MDD		Rotterdam Study DD		Rotterdam Study AD	
	Cases (N = 1738)	Controls (N = 1802)	Cases (N = 178)	Controls (N = 915)	Cases (N = 222)	Controls (N = 290)
Age, mean (sd)	42.6 (12.6)	45.1 (14.1)	67.7 (6.8)*	64.8 (6.5)*	75.4 (5.82)	74.6 (5.32)
Women(%)	69.6	62	75.7	43.5	78	43
Education(%)						
Elementary	7.8	5.7	41.8	22.7	37.2	19.4
Intermediate	62	56.3	53.7	64.5	56.4	64.6
Higher	32.2	38.1	4.5	12.8	6.4	16.0
Antidepressant medication (%)	34.5	0	12.4*	0*	12.1	2
Co-morbid AD / MDD (%)	69.9	0	35.8**	3.6***	10.8	0.7

AD: Anxiety disorder, DD: Depressive Disorder, *Recorded at baseline, **Data available in 108 out of 178 cases, ***Data available in 701 of 915 controls, sd, standart deviation.

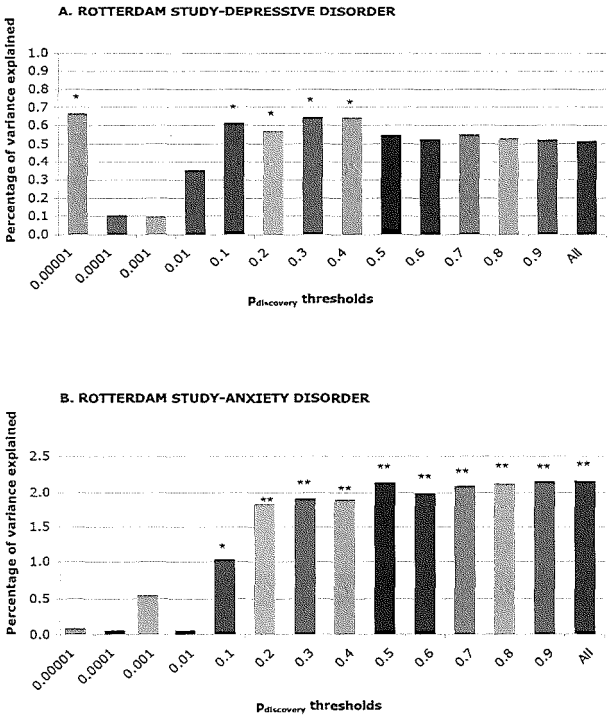
TABLE 2. Descriptive data of ERF study

	Mean (sd)
CES-D	10.6 (9.6)
HADS-D	5.9 (4.3)
HADS-A	6.7(4.5)
Age, mean (sd)	48.2 (14.7)
Women(%)	57.4
Education(%)	
Elementary	30.8
Intermediate	63.8
Higher	5.4
Antidepressant medication (%)	5.0

CES-D: Center for Epidemiological Studies Depression Scale. HADS-D: Hospital Anxiety and Depression Scale-Depression subscale. HADS-A: Hospital Anxiety and Depression Scale-Anxiety subscale. sd, standard deviation.

In the ERF study, CES-D, HADS-D and HADS-A scores were highly correlated ($r \sim 0.7$ pair wise for all three). Figure 1A shows the variance explained by the genetic risk scores in the logistic regression analyses performed in the Rotterdam Study using depressive disorder as dependent variable. The genetic score based on the first cluster of 6 SNPs ($P_{discovery} < 0.00001$) significantly explained 0.66% of the variance in depressive disorder in the Rotterdam Study (P-value = 0.03). This association is explained for a large part by a cluster of 3 SNPs (rs2715148, rs2522833, rs2522840) in the *PCLO* gene as after removing the *PCLO* SNPs in linkage disequilibrium (LD), the risk score was not significantly associated with depression in the target sample anymore. More importantly, the scores based on SNPs with $P_{discovery} < 0.1$ to $P_{discovery} < 0.4$ were associated

FIGURE 1. Percentage of variance explained by genetic risk scores in Rotterdam Study.



Percentage of variance represented as difference in Nagelkerke R^2 after adjustment for age and sex. A, Analyses based on comparison of MDD persons (N = 178) to persons scoring in the lowest quartile of CES-D (CES-D ="0") scale and who did not report any depressive complaints during the follow-up (N = 915) B, Analyses based on comparison of persons with anxiety disorder (N = 222) to persons scoring in the lowest quartile of HADS-A (HADS-A="0") scale and who did not report any depression or anxiety symptoms during the follow-up (N = 290).

* *P*-value < 0.05, ** *P*-value < 0.001

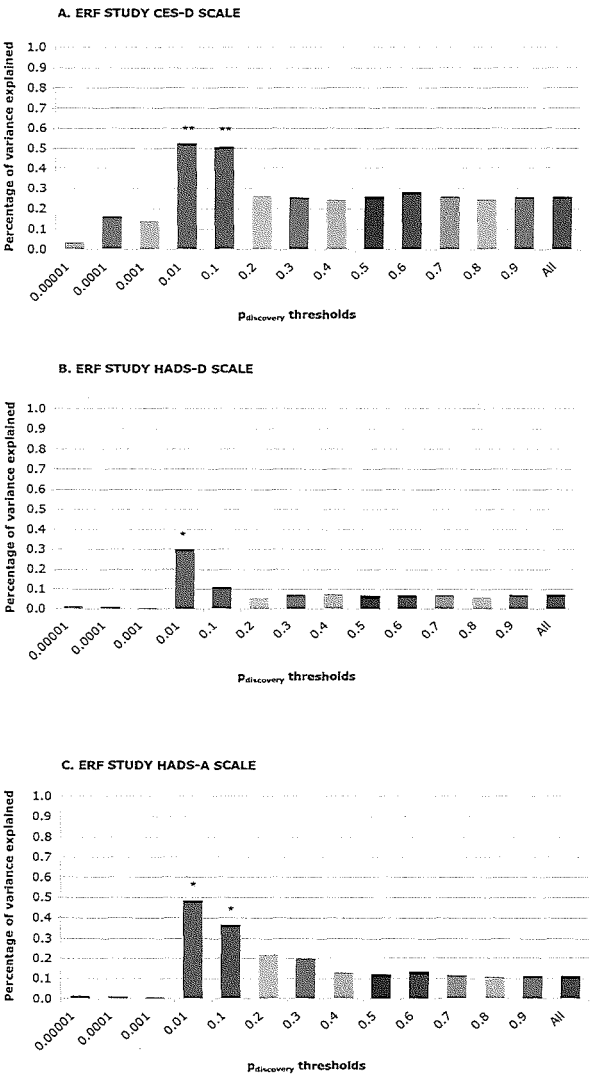
to depressive disorder in the Rotterdam Study explaining up to 0.65% of the variance, with a *P*-value < 0.05. As shown in Figure 1B, the Rotterdam Study anxiety disorder case-control sample analysis yielded the highest percentage of variance explained with the genetic risk scores from GAIN –MDD study. The risk scores based on SNPs with $P_{discovery} < 0.1$ to $P_{discovery} < 1.0$ significantly explained up to 2.1% of the variance (*P*-value = 0.0025). For $P_{discovery}$ values of 0.1, 0.2 and 0.3, the percentage of variance increased from 1% to 2% when a higher number of SNPs were included in risk scoring.

Figure 2A-C show the linear regression results for the analysis using the continuous scores on the CES-D, HADS-D and HADS-A in the ERF study. For CES-D, the scores based on SNPs with $P_{discovery}$ values < 0.01 and 0.1 explained ~0.5% of the variance (*P*-value = 0.007 and *P*-value =0.008). For the HADS-D, the score based on SNPs with $P_{discovery} < 0.01$ significantly explained

0.3% of the variance (*P*-value = 0.03). The MDD based genetic score was also significantly associated to anxiety measured with the HADS-A explaining up to 0.5% of the variance (*P*-value=0.01).

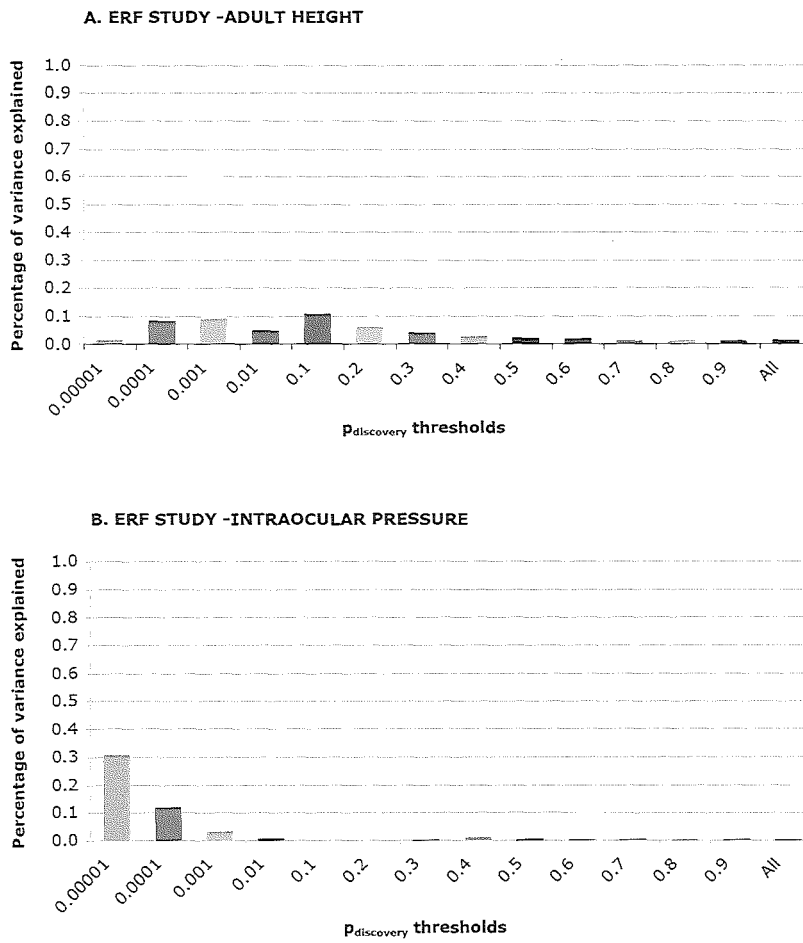
To examine whether these results were due to chance, we tested whether the MDD based genetic risk score also predicted variation in height and IOP measured in ERF. Heritability of IOP

FIGURE 2. Percentage of variance explained in depression and anxiety symptoms in ERF study by the genetic risk scores.



Percentage of variance represented as difference in R^2 after adjusting for age sex and family relations. A, B and C, Analyses of continuous scales. * *P*-value < 0.05
** *P*-value < 0.01

CHAPTER 2.1 Figure 3. Predicting height and Intraocular Pressure in ERF study.



Linear regression analysis of height and Intraocular Pressure. Percentage of variance represented as difference in R^2 after adjusting for age sex and family relations

was 35% and for height was 86%[36, 44] and none of the traits was correlated to depression or anxiety in the ERF study. The genetic risk score for MDD failed to predict IOP and height (Figure 3A and B) suggesting that this relation is specific to depression and anxiety. Moreover, a genetic score of SNPs with $P_{discovery} < 0.9$ in the discovery set did not show significant association with any of the phenotypes in the target samples (data not shown).

DISCUSSION

The aim of this study was to investigate the genetic architecture of depression and the potential overlap in genetic risk factors with anxiety. Owing to the availability of an elderly cohort, it was also possible to examine whether the genetic factors influencing anxiety and depression change across the life span. Using genetic risk scores derived from the association results of the GAIN-MDD study in two independent target samples we evaluated the evidence for a genome-wide signature for several measures of depression and anxiety used as outcome variables in the target samples. For depression, either diagnosed according to the DSM-IV or measured with the CES-D or HADS-D, we could explain up to ~1% of the variance with the genetic risk scoring approach. Moreover, the genetic risk scores for depression were also associated to anxiety explaining up to 2.1% of the variance when approximately half of the genome-wide SNP data were included in the score. The explained variance was highest in the elderly sample indicating that the genetic factors influencing anxiety and depression hardly change with age. No significant results were found for the control variables height and IOP, implying that the association of the genetic risk score with depression and anxiety does not reflect chance alone. Overall, these findings suggest the presence of many loci, each with a small effect influencing depression as well as anxiety.

We checked whether our results were only due to SNPs in high LD segregating together. We performed a strict LD pruning (200 SNPs sliding window with $r^2_{snp\text{-}snp}$ threshold of 0.25). Considering the CES-D scale in ERF study, percentage of variance explained by the risk scores based on SNPs with $P_{discovery} < 0.01$ dropped slightly after LD pruning from 0.52 to 0.49 but remained significant (P-value = 0.01) whereas a less strict pruning with an $r^2_{snp\text{-}snp}$ threshold of 0.50 improved the percentage of explained variance to 0.62 (P-value = 0.0003). Results with HADS-A and HADS-D scales were similar which shows that LD pruning itself does not add a major difference to the method (data not shown). Excluding the SNPs with minor allele frequency < 0.05 did not change the explained variance. This finding suggest the common disease common variant hypothesis is explaining MDD heritability, on the other hand, the power to detect the effect of rare variants in the discovery and target sets, was low and such rare variants may be detected by other approaches such as linkage or deep sequencing.

Our results are in agreement with the results from the International Schizophrenia Consortium[13] that pointed out a polygenic component influencing schizophrenia as well as bipolar disorder. There was a somewhat higher amount of explained variance for schizophrenia (3.2% compared to %1). This may be due to power issues such as differences in sample size (~3300 cases for ISC vs ~1800 cases for GAIN-MDD), MDD being a common disease with clear non-genetic influence because of life events, and lower heritability compared to schizophrenia (~40% vs ~ 80%).

The percentage of explained variance in anxiety (2.1%), supports the idea of shared genetic background between these disorders. This has already been suggested by twin studies[45] and is

confirmed by our results. The trend of increase in R^2 for anxiety with different $P_{discovery}$ thresholds is different from the trend that we observe in depression, pointing out that the effect sizes are different, but the direction of effect is the same. We would like to stress that difference in explained variance between the target samples can evenly well be explained by chance. It is important to note that 70% of the GAIN-MDD cases had a co-morbid lifetime anxiety diagnosis. This could explain the overlap with anxiety. However, this high co-morbidity is exactly what is expected if two disorders are influenced by similar genes and diagnoses are not mutually exclusive. Future research, preferably with a more balanced proportion of pure depressed and comorbid cases, can shed more light on the overlap in genetic factors influencing anxiety and depression.

A limitation of this study was that there were some differences between the discovery set and the target samples. Different instruments were used to measure depression and anxiety. In the GAIN–MDD study, the Composite International Diagnostic Interview was used to diagnose MDD and anxiety disorders, while in the Rotterdam study, the Present State Examination (PSE) was used. However, both instruments aim to make diagnoses according to the DSM-IV criteria and have adequate agreement for overall syndromes[46]. In the ERF population, symptoms of depression and anxiety were measured using the CES-D and HADS. Several validation studies on various types of patients using different diagnostic tools have shown that HADS performs well in assessing the symptom severity and case status of anxiety disorders and depression in both somatic, psychiatric and primary care patients and in the general population[47]. The HADS-D subscale has shown high sensitivity (~ 0.9)and specificity (~ 0.7) for MDD as diagnosed by DSM-IV in various studies[48]. The CES-D scale was found to be satisfactory in a semi-clinical sample of the elderly and in general population (sensitivity = 0.9 and specificity = 0.6) for life-time MDD and also performed excellent for 1 month of prevalence of MDD as diagnosed by DSM-IV (sensitivity = 1.0 and specificity= 0.9) among elderly Dutch[49-51]. Considering the HADS-A subscale, the sensitivity and specificity for DSM-IV generalized anxiety disorder was reported to be 0.9 and 0.8, respectively[52]. In addition, the discovery set in this study included lifetime MDD cases, whereas the Rotterdam study recorded depressive disorders during a 9-year follow-up rather than lifetime MDD. Similarly, CES-D and HADS measure depressive and anxious symptoms in the last week. This means that subjects in the control groups in the target samples may be non-current but life-time MDD or anxiety cases. To summarize, although the measurements of anxiety and depression used in the three study samples are definitely related to each other, the fact that they are not entirely similar implies some heterogeneity, biasing the results toward the null hypothesis. Another point involves the difference in gender ratios between discovery and target samples. In the discovery sample, the cases and controls were carefully matched on age and sex. Meta-analysis of twin studies suggests that genetic factors that influence depression are mostly shared between men and women[1, 53, 54]. Sex was also used as a covariate when predicting depression or anxiety in the target samples. Thus, it seems unlikely that the gender ratio may have a major effect in the replication of the findings. There was also heterogeneity in

education level as a measure of socio economical status. In spite of these differences, we still found a significant effect of the genetic risk score suggesting that the effects of the risk scores are actually even stronger. In both the International Schizophrenia Consortium study and the current studies, the low variance explained compared with the heritability of the disorders will also reflect that the analyses did not include the X chromosome, that gene–gene or gene–environment interactions are not considered and that the current generation of genotyping platforms do not fully tag genomic variance[55]. This study is the second study showing direct evidence for a polygenic component influencing the susceptibility for a psychiatric disorder as well as overlap in genetic risk factors with another psychiatric condition. In addition, this study suggests that the genetic factors influencing anxiety and depression hardly change with age. The results imply that causal SNPs or the SNPs in LD with such variants do exist, but have lower effect sizes than the first generation of GWA studies on psychiatric disorders was powered to detect. This provides optimism that variants associated at genome-wide levels of significance will be detectable as sample sizes increase in the next generation of GWA studies and their meta/mega analyses. Moreover, it confirms that genome-wide profiling is a useful approach to analyze the genetic architecture of disorders, that is, similarities and differences in genetic factors influencing several disorders or influencing the same disorder across the lifespan.

ACKNOWLEDGEMENTS

GAIN-MDD: We acknowledge support from NWO: genetic basis of anxiety and depression (904-61-090); resolving cause and effect in the association between exercise and well-being (904-61-193); twin-family database for behavior genomic studies (480-04-004); twin research focusing on behavior (400-05-717), Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192); Neuroscience Campus Amsterdam (NCA-VU); genome-wide analyses of European twin and population cohorts (EU/QLRT-2001-01254); genome scan for neuroticism (NIMH R01 MH059160); Geestkracht program of ZonMW (10-000-1002); matching funds from universities and mental health care institutes involved in NESDA. Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). CM Middeldorp is financially supported by NWO (VENI grant 916-76-125). Genotype data were obtained from dbGaP(http://www.ncbi.nlm.nih.gov/dbgap, accession number phs000020.v1.p1). Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is financially supported by the NWO (480-05-003). Dr Sullivan was also supported by R01s MH074027 and MH077139. Dr Wray is supported by the Australian National Health and Medical Research Council (grant 496688). Dr Middeldorp was supported by NWO (VENI grant 916-76-125). Rotterdam Study: Genome-wide genotyping of the Rotterdam Study was supported by NWO (175.010.2005.011). ERF study: The ERF study was supported by grants from The Netherlands Organisation for Scientific Research, Erasmus MC,

The Netherlands Brain Foundation (HsN) and the Centre for Medical Systems Biology (CMSB). The genotyping was founded by EUROSPAN (European Special Populations Network) Consotium. We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.

REFERENCES

1. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry.* Oct 2000;157(10):1552-1562.
2. Middeldorp C, Boomsma D. Genetics and Psychopathology. In: Berntson GG CJ, ed. *Handbook of Neuroscience for Behavioral Sciences.* New York: Wiley; 2009.
3. Levinson DF. The genetics of depression: a review. *Biol Psychiatry.* Jul 15 2006;60(2):84-92.
4. Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, et al. Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry.* Aug 2008;13(8):772-785.
5. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Jama.* Jun 18 2003;289(23):3095-3105.
6. Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol Med.* May 2005;35(5):611-624.
7. Sullivan PF, de Geus EJ, Willemsen G, et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry.* Apr 2009;14(4):359-375.
8. Bochdanovits Z, Verhage M, Smit AB, et al. Joint reanalysis of 29 correlated SNPs supports the role of PCLO/Piccolo as a causal risk factor for major depressive disorder. *Mol Psychiatry.* Jul 2009;14(7):650-652.
9. Hek K, Mulder CL, Luijendijk HJ, et al. The PCLO gene and depressive disorders: replication in a population-based study. *Hum Mol Genet.* Nov 26 2009.
10. Psychiatric GCCC, Cichon S, Craddock N, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry.* May 2009;166(5):540-556.
11. Muglia P, Tozzi F, Galwey NW, et al. Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry.* Dec 23 2008.
12. Craddock N, O'Donovan MC, Owen MJ. Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes. *Mol Psychiatry.* 2008;13(7):649-653.
13. International Schizophrenia C, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* Aug 6 2009;460(7256):748-752.
14. Baldwin RC, Tomenson B. Depression in later life. A comparison of symptoms and risk factors in early and late onset cases. *Br J Psychiatry.* Nov 1995;167(5):649-652.
15. Jansson M, Gatz M, Berg S, et al. Gender differences in heritability of depressive symptoms in the elderly. *Psychol Med.* Apr 2004;34(3):471-479.
16. Johnson W, McGue M, Gaist D, Vaupel JW, Christensen K. Frequency and heritability of depression symptomatology in the second half of life: evidence from Danish twins over 45. *Psychol Med.* Oct 2002;32(7):1175-1185.
17. Gillespie NA, Kirk KM, Evans DM, Heath AC, Hickie IB, Martin NG. Do the genetic or environmental determinants of anxiety and depression change with age? A longitudinal study of Australian twins. *Twin Res.* Feb 2004;7(1):39-53.
18. Boomsma DI, Willemsen G, Sullivan PF, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet.* Mar 2008;16(3):335-342.
19. Penninx BW, Beekman AT, Smit JH, et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res.* 2008;17(3):121-140.
20. Boomsma DI, de Geus EJ, Vink JM, et al. Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet.* Dec 2006;9(6):849-857.

21. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 2007;17(10):1520-1528.
22. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders 4th edition.* 4 ed: APA; 1994
23. Organization WH. *Composite International Diagnostic Interview (version 2.1)* Geneva: WHO; 1992.
24. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* Sep 2007;81(3):559-575.
25. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol.* Sep 2 2009.
26. Luijendijk HJ, van den Berg JF, Dekker MJ, et al. Incidence and recurrence of late-life depression. *Arch Gen Psychiatry.* Dec 2008;65(12):1394-1401.
27. Radloff. The CES-D scale: a self report depression scale for research in theb general population. *Appl Pshycol Measurement.* 1977;3:385-401.
28. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* Jun 1983;67(6):361-370.
29. Wing J, Cooper, JE and Sartorius, N. Measurement and Classification of Psychiatric Symptoms. An Instruction Manual for the PSE and CATEGO Programme. Cambridge: Cambridge University Press; 1974.
30. Li Y, Ding J, Abecasis G. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. Paper presented at: American Society of Human Genetics; 9–13 October 2006, 2006; New Orleans, Louisiana.
31. Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive evaluation of SNP genotype imputation. *Hum Genet.* Mar 2009;125(2):163-171.
32. Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. GRIMP: A web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics.* Aug 28 2009.
33. Lopez-Leon S, Chi Choy W, Aulchenko YS, et al. Genetic Factors Influence the Clustering of Depression among Individuals with Lower Socioeconomic Status. *PLoS ONE.* 2009;4(3):e5069.
34. Santos R, Zillikens M, Rivadeneira F, et al. Heritability of fasting glucose levels in a young genetically isolated population. *Diabetologia.* 2006;49(4):667-672.
35. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet.* 2004;12(7):527-534.
36. van Koolwijk LME, Despriet DDG, van Duijn CM, et al. Genetic Contributions to Glaucoma: Heritability of Intraocular Pressure, Retinal Nerve Fiber Layer Thickness, and Optic Disc Morphology. *Investigative Ophthalmology Visual Science.* 2007;48(8):3669-3676.
37. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* May 15 2007;23(10):1294-1296.
38. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* Jan 2002;30(1):97-101.
39. Amin N, van Duijn CM, Aulchenko YS. A genomic background based method for association analysis in related individuals. *PLoS One.* 2007;2(12):e1274.
40. Evans DM, Visscher PM, Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum Mol Genet.* Sep 15 2009;18(18):3525-3531.
41. Rebollo I, de Moor MH, Dolan CV, Boomsma DI. Phenotypic factor analysis of family data: correction of the bias due to dependency. *Twin Res Hum Genet.* Jun 2006;9(3):367-376.

42. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* May 1998;62(5):1198-1211.
43. Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet.* Jul 2008;16(7):854-860.
44. Axenovich TI, Zorkoltseva IV, Belonogova NM, et al. Linkage analysis of adult height in a large pedigree from a Dutch genetically isolated population. *Hum Genet.* May 24 2009.
45. Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry.* Sep 2003;60(9):929-937.
46. Andrews G, Peters L. The psychometric properties of the Composite International Diagnostic Interview. *Soc Psychiatry Psychiatr Epidemiol.* Feb 1998;33(2):80-88.
47. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res.* Feb 2002;52(2):69-77.
48. Silverstone PH. Concise assessment for depression (CAD): a brief screening approach to depression in the medically ill. *J Psychosom Res.* Aug 1996;41(2):161-170.
49. Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. *Psychol Med.* Jan 1997;27(1):231-235.
50. Donker T, van Straten A, Marks I, Cuijpers P. Brief self-rated screening for depression on the Internet. *J Affect Disord.* Aug 11 2009.
51. Haringsma R, Engels GI, Beekman AT, Spinhoven P. The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. *Int J Geriatr Psychiatry.* Jun 2004;19(6):558-563.
52. Olsson I, Mykletun A, Dahl AA. The Hospital Anxiety and Depression Rating Scale: a cross-sectional study of psychometrics and case finding abilities in general practice. *BMC Psychiatry.* 2005;5:46.
53. Kendler KS, Gatz M, Gardner CO, Pedersen NL. A Swedish national twin study of lifetime major depression. *Am J Psychiatry.* Jan 2006;163(1):109-114.
54. Middeldorp CM, Birley AJ, Cath DC, et al. Familial clustering of major depression and anxiety disorders in Australian and Dutch twins and siblings. *Twin Res Hum Genet.* Dec 2005;8(6):609-615.
55. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature.* Sep 10 2009;461(7261):272-276.

Chapter 2.2

Genetic architecture of circulating lipid levels

Ayse Demirkan, Najaf Amin, Aaron Isaacs, Marjo-Riitta Jarvelin, John B Whitfield, Heinz-Erich Wichmann, Kirsten Ohm Kyvik, Igor Rudan, Christian Gieger, Andrew A Hicks, Åsa Johansson, Jouke-Jan Hottenga, Johannes J Smith, Sarah H Wild, Nancy L Pedersen, Gonneke Willemsen, Massimo Mangino, Caroline Hayward, André G Uitterlinden, Albert Hofman, Jacqueline Witteman, Grant W Montgomery, Kirsi H Pietiläinen, Taina Rantanen, Jaakko Kaprio, Angela Döring, Peter P Pramstaller, Ulf Gyllensten, Eco JC de Geus, Brenda W Penninx, James F Wilson , Fernando Rivadeneria, Patrik KE Magnussön, Dorret I Boomsma, Tim Spector, Harry Campbell, Birgit Hoehne, Nicholas G Martin, Ben A Oostra, Mark McCarthy, Leena Peltonen-Palotie, Yurii Aulchenko, ter M Visscher, Samuli Ripatti, A Cecile JW Janssens, Cornelia M van Duijn, for the ENGAGE CONSORTIUM.

Genetic architecture of circulating lipid levels. Eur J Hum Genet. Jul;19(7):813-819.

ABSTRACT

Serum concentrations of low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG) and total cholesterol (TC) are important heritable risk factors for cardiovascular disease. Although genome-wide association studies (GWAS) of circulating lipid levels have identified numerous loci, a substantial portion of the heritability of these traits remains unexplained. Evidence of unexplained genetic variance can be detected by combining multiple independent markers into additive genetic risk scores. Such polygenic scores, constructed using results from the ENGAGE Consortium GWAS on serum lipids, were applied to predict lipid levels in an independent population-based study, the Rotterdam Study-II (RS-II). We additionally tested for evidence of a shared genetic basis for different lipid phenotypes. Finally, the polygenic score approach was used to identify an alternative genome-wide significance threshold prior to pathway analysis and those results were compared to those based on the classical genome–wide significance threshold. Our study provides evidence suggesting that many loci influencing circulating lipid levels remain undiscovered. Cross-prediction models suggested a small overlap between the polygenic backgrounds involved in determining LDL-C, HDL-C and TG levels. Pathway analysis utilizing the best polygenic score for TC uncovered extra information compared to using only genome-wide significant loci. These results suggest that the genetic architecture of circulating lipids involves a number of undiscovered variants with very small effects, and that increasing GWAS sample sizes will enable the identification of novel variants that regulate lipid levels.

INTRODUCTION

Serum concentrations of low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG) and total cholesterol (TC) are highly heritable phenotypes associated with the risk of cardiovascular morbidity and mortality[1-4]. A number of genome-wide association studies (GWAS) successfully identified multiple genes influencing circulating lipid levels[5-12]. There are currently over 100 established loci that include both common variants with relatively small effects as well as a considerable number of rare variants with large effects[13]. Despite these successes, a substantial proportion of the heritability of each trait remains unexplained, suggesting that many determinants have yet to be identified[14].

Several plausible explanations may underlie the unexplained heritability of lipid traits, including the presence of both unknown common variants with small effects and novel rare variants with larger effects. The ENGAGE GWAS[5] was one of the first large population based studies designed to find variants associated with circulating lipid levels. The study, based on 16 European cohorts including up to 22 562 individuals, identified 6 novel loci, in addition to replicating 16 previously known loci. However, as demonstrated by the recent GWAS from the Global Lipid Genetics Consortium (GLGC), numerous additional variants passed the genome-wide significance threshold as a result of increased sample size[15]. The GLGC GWAS, which included over 100 000 individuals of European ancestry, reported 95 loci, with 59 reaching genome-wide significance for the first time. These results raise an interesting question: if common variants remain to be discovered, how many should we expect? Are there still a limited number of loci or can we expect a polygenic mechanism that involves a very large number of variants with very small effects? In the latter case, these variants would contribute to a continuous spectrum of alleles spanning the genome and single genes involved in this complex polygenic model might not be detectable by GWAS, regardless of sample size[16]. Evidence for this type of genetic architecture can be shown using a genome-wide scoring approach, as was recently demonstrated for a number of psychiatric outcomes[17-19]. Additionally, these polygenic scores may provide extra information useful in determining *P*-value thresholds for pathway analysis.

The current study aimed to explore the extent to which common variation accounts for the unexplained heritability of circulating lipid levels using the genome-wide scoring method. We also evaluated the evidence for a common polygenic effect underlying different lipid traits, using the same risk scoring approach. Finally, we examined the utility of genome-wide polygenic scores for identifying pathways beyond those identified using a classical GWAS approach.

MATERIALS AND METHODS

The polygenic risk score approach involves using results from a discovery set to explore the genetic architecture of an independent target sample. Our discovery set consisted of the meta-analysis of 16 European populations from the ENGAGE Lipid Consortium (N = 17 798 – 22 562)

TABLE 1. Descriptive data of discovery and replication samples

	ENGAGE		RS-II	
	Men	**Women**	**Men**	**Women**
Number of subjects	8403	14159	1061	1253
HDL-C (mmol/L)	1.3 (0.3)	1.6 (0.4)	1.2 (0.3)	1.5 (0.4)
LDL-C (mmol/L)	3.4 (0.9)	2.3 (0.9)	3.6 (0.9)	1.5 (0.8)
TG (mmol/L)	1.6 (1.1)	1.1 (0.7)	1.6 (0.9)	1.5 (0.8)
TC (mmol/L)	5.6 (0.9)	5.9 (0.9)	5.6 (1.1)	5.7 (1.1)

Means and standard deviations (sd) are given.

(Table 1). A detailed description of this study, including populations, genotyping information and statistical analysis, was previously published[5].

The target sample consisted of RS-II, an extension of the Rotterdam Study, a prospective cohort study started in 1990 in the Ommoord district of the city of Rotterdam. RS-II, which was not a part of the ENGAGE discovery set, consists of 3011 participants (out of 4472 invitees) who were 55 years or older during the recruitment period (2000-2001)[20]. Of the 3011, 2540 persons were successfully genotyped with an Illumina 610K array. Fasting HDL-C, TG and TC were measured with enzymatic colorimetric tests on a Roche/Hitachi 911 analyzer (Roche Diagnostics, Meylan, France). LDL-C was estimated using the Friedewald formula[21].

SNPs included in the construction of the polygenic scores were based on the results from the ENGAGE study. We selected different clusters of SNPs for the calculation of the scores using several *P*-value thresholds ($P_{discovery}$) ranging from 5×10^{-8} to 0.5. We calculated genetic scores for those various clusters of SNPs in the target sample by multiplying the number of risk alleles for each SNP (0, 1 or 2) by the effect sizes from the discovery set, and summing them up across all the SNPs in that cluster. For this analysis we used the PLINK "profile scoring" option. SNPs that had a call rate < 90% or HWE *P*-value < 1×10^{-8} were excluded from these computations. A/T and G/C polymorphisms were also excluded to avoid potential strand inconsistencies. SNPs in linkage disequilibrium (LD) were pruned over 200 SNP sliding windows using a pair wise r^2 threshold of 0.25 in PLINK[22]. LD pruning was performed per SNP cluster. (See Supplementary Table 2 for the number of SNPs remaining in each cluster and used for analysis)

The associations between these scores and serum lipid levels were tested in SPSS using linear regression models with sex, age and age^2 as covariates (the same covariates as included in the discovery GWAS). The proportion of total variance explained by the genetic score, referred to here as the percentage of explained variance (PEV), was determined by comparing models with/without the risk score.

To evaluate whether the PEV results were driven by the GWAS hits, we also constructed a variable comprised of only the significant GWAS variants and included it as a covariate in our original models. When calculating the polygenic scores for these analyses, we also removed SNPs within 2 mega-base (Mb) windows surrounding the GWAS hits. We employed exactly the same pruning approach for this analysis.

TABLE 2 Correlation matrix of circulating lipids and genetic risk scores in RS-II.

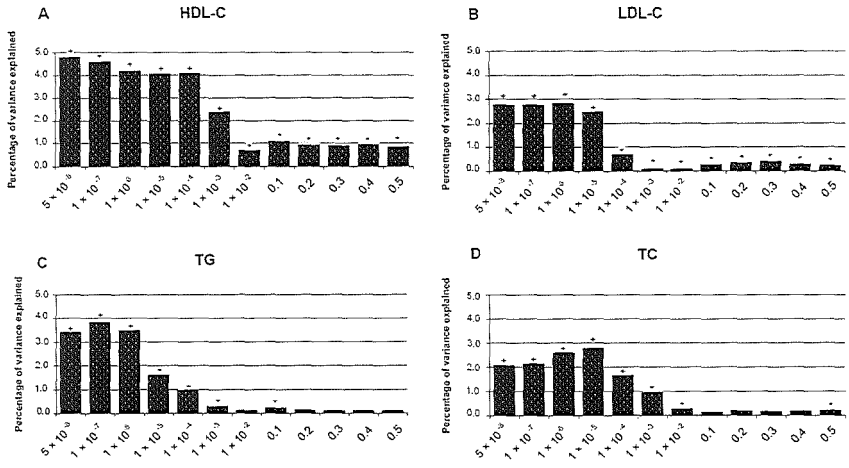
Correlation between the phenotypes		*Correlation between the genetic risk scores*			
		HDL-C	**LDL-C**	**TG**	**TC**
HDL-C	5×10^{-8}		0.01	-0.20 **	0.02
	1×10^{-7}		0.01	-0.17 **	0.03
	1×10^{-6}		-0.01	-0.09 **	0.07 **
	1×10^{-5}		0.02	-0.04 *	0.05 *
LDL-C	5×10^{-8}	-0.1 **		0.01	0.76 **
	1×10^{-7}			0.02	0.75 **
	1×10^{-6}			0.05 *	0.81 **
	1×10^{-5}			0.03	0.71 **
TG	5×10^{-8}	-0.5 **	0.1 **		0.13 **
	1×10^{-7}				0.13 **
	1×10^{-6}				0.12 **
	1×10^{-5}				0.08 **
TC	5×10^{-8}	0.1 **	0.9 **	0.3 **	
	1×10^{-7}				
	1×10^{-6}				
	1×10^{-5}				

Lower-left side of the matrix shows the phenotypic correlation between circulating lipid levels, adjusted by age, age^2 and sex. Upper-right side of the matrix shows the correlation between the genetic risk scores of four circulating lipids, for the first four risk scores with $p_{discovery} < 5 \times 10^{-8}$, $p_{discovery} < 1 \times 10^{-7}$, $p_{discovery} < 1 \times 10^{-6}$ and $p_{discovery} < 1 \times 10^{-5}$. * Correlation significant at *P*-value < 0.05. **Correlation significant at *P*-value < 0.001.

In order to search for evidence for a shared genetic background between various lipid traits, we tested additional models in which we used the polygenic score for a particular lipid and tried to predict the others, for instance, utilizing the HDL-C polygenic score to predict TG, and vice versa.

The score which yielded the highest PEV for a given lipid trait ostensibly includes the most valuable genetic information; therefore we selected these thresholds to utilize in pathway analysis (in contrast to using only genome-wide significant loci). For these analyses, we used the PANTHER tools (http://www.pantherdb.org)[23]. We first tested the genome-wide significant SNPs ($P_{discovery} < 5 \times 10^{-8}$) from the ENGAGE GWAS in the pathway analysis. These results were then compared with those obtained using alternative $P_{discovery}$ thresholds selected on the predictive ability of the polygenic scores. After SNP selection, SNPs within gene regions were converted to gene symbols using the "SCAN SNP and CNV annotation database" (http://www.scandb.org). Gene lists were tested for enrichment in three PANTHER categories: (1) pathways, (2) biological processes and (3) molecular functions. Testing for enrichment basically involves comparing one gene list to the reference list to statistically determine over- or under- representation of PANTHER classification categories. Based on the reference list, an expected value is computed (the number of genes one would expect in the list for a particular PANTHER category) and it is assumed that, under the null hypothesis, genes in the tested list are sampled

FIGURE 1. Percentage of explained variance of circulating lipids with risk scores by different $p_{discovery}$ thresholds.



Adjusted for age, sex and age^2
+ P-value $< 5 \times 10^{-8}$, * $5 \times 10^{-8} < P$-value < 0.05

from the same distribution as genes from the reference set. The *Homo sapiens* gene list from National Center for Biotechnology Information was used as the reference gene list. To avoid bias caused by multiple testing, PANTHER's Bonferroni correction option was implemented. (See Supplementary Figure 1 for the over all flowchart of the study)

RESULTS

Table 1 shows summary statistics for the discovery and target samples. The female/male ratio in the discovery set was significantly higher compared to the target set (1.6 vs 1.2, P-value < 0.001). Genome-wide significant SNPs from the ENGAGE GWAS were checked for their associations in the target sample using linear regression. Generally, evidence of association between those SNPs and lipid levels were marginally significant or non-significant (Supplementary Table 1). The GWAS of circulating lipids in RS-II did not show any genome-wide significant findings except the CETP gene region SNPs which were associated to HDL-C (rs7499892, P-value =3.4 $\times 10^{-13}$). Manhattan plots for the GWAS of the HLD-C, LDL-C, TG and TC can be found in the Supplementary Figure 2.

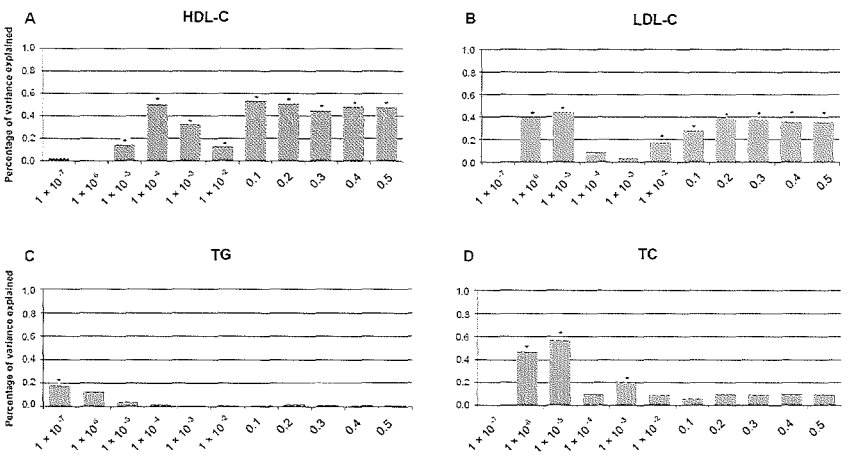
Prediction

Figure 1 shows the PEV obtained for each lipid trait using the polygenic scores generated for a number of P-value thresholds in the target sample (RS-II). For HDL-C, the polygenic score

computed using 19 genome-wide significant SNPs from 8 gene regions ($P_{discovery} < 5 \times 10^{-8}$) resulted in the maximum PEV compared to the null model (4.75 %, *P*-value = 3.6×10^{-30}; Figure 1 A). For LDL-C, (Figure 1 B), the maximum PEV was observed with the polygenic score that included 21 SNPs with a $P_{discovery} < 1 \times 10^{-6}$ (2.6 %, *P*-value = 5.1×10^{-16}). Figure 1 C shows PEVs for TG levels; the score that included 12 SNPs from 8 regions with $P_{discovery} < 1 \times 10^{-7}$ (3.8 %, *P*-value = 2.8×10^{-21}) was the best predictor. For these traits, the variance explained decreased with the inclusion of additional SNPs in the polygenic score selected using more liberal $P_{discovery}$ thresholds (Figure 1 A to C). Finally, for TC, the highest PEV was obtained using 46 SNPs from 24 regions with $P_{discovery} < 10^{-5}$ (2.7%, p = 1.4×10^{-16}). This was higher than the PEV obtained using only the genome-wide significant SNPs (PEV = 2.1%, *P*-value = 8.2×10^{-13}, N = 20 SNPs from 11 regions; Figure 1 D). As with HDL-C, LDL-C, and TG, the explained variance for TC dropped when more liberal $P_{discovery}$ thresholds were used to construct the polygenic score. For LDL-C, HDL-C and TC, all scores were significant (up to a threshold of $P_{discovery} < 0.5$). We observed similar patterns when we used unpruned data (Supplementary Figure 3).

Figure 2 shows the results from the second approach, in which models were adjusted for genome-wide significant variants. For HDL-C (Figure 2 A), the PEV increased as SNPs were added, up to 0.5 % with $P_{discovery} < 0.1$ (*P*-value = 1.0×10^{-4}) and remained significant until $P_{discovery} < 0.5$ (p=2.3×10^{-4}). A similar pattern was observed with LDL-C (Figure 2 B, explained variance was up to 0.4 % (*P*-value = 0.002) with $P_{discovery}$ threshold of 0.2. In contrast, the polygenic score for TG, when the effects of known variants were excluded, was not associated with TG levels in

FIGURE 2. Percentage of explained variance of circulating lipids when the top regions are excluded.



Adjusted for age, sex age^2 and risk score computed from genome wide significant findings. The lack of association in the first cluster of SNPs are due to the exclusion of SNPs within 2MB window region surrounding the top findings, since there were only a few SNPs to be included in the analysis after excluding the top regions.

* *P*-value < 0.05

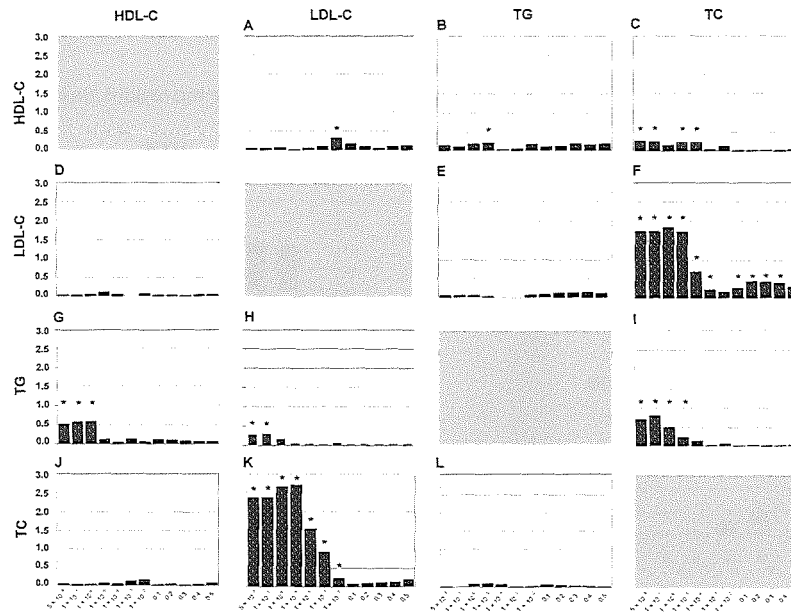
the target population (Figure 2 C). For TC (Figure 2 D), the maximum PEV was observed with $P_{discovery} < 1\times10^{-5}$, (0.6 % and *P*-value = 1.8×10^{-4}).

Cross prediction

Table 2 shows the phenotypic correlations for the four outcomes studied, and additionally shows the correlations between the polygenic scores for different $P_{discovery}$ thresholds. Correlations between the traits were modest, with the exceptions of TC and LDL-C, (r = 0.9) and TG and HDL-C (r = -0.5). The correlations between the polygenic scores were weaker than the phenotypic correlations (0.8 for TC/LDL-C and -0.2 for TG/HDL for $P_{discovery} < 5 \times 10^{-8}$)

To evaluate the evidence for common polygenetic effects underlying lipid levels, we performed cross-prediction analyses (Figure 3). The highest PEV was based on the TC score at $P_{discovery\ (TC)} < 1\times10^{-5}$, which explained up to 2.7% of the variance in circulating LDL-C (*P*-value = 2.0×10^{-5}; Figure 3 K). Similarly, LDL-C risk profiles explained up to 1.8 % of the variance in TC when we selected all SNPs with a $P_{discovery\ (LDL\text{-}C)} < 10^{-6}$ (*P*-value = 1.4×10^{-11}; Figure 3 F). These findings are inline with the high phenotypic correlations between those variables. Figures 3 G

FIGURE 3. Cross prediction across different lipids.



Evaluation of the evidence for a joint polygenic effect underlying various lipids: A to C: Prediction based on HDL risk scores imposed on LDL, triglycerides and total cholesterol. D to F; prediction based on LDL risk scores imposed on HDL, triglycerides and total cholesterol , G to I; prediction based on triglyceride risk scores, J to L, prediction based on total cholesterol risk scores.* *P*-value < 0.05

TABLE 3 Pathway analysis

		NCBI	Observed	Expected	Over/ Under	*P*-value	*P*-value *
PATHWAYS	***n.s***						
BIOLOGICAL PROCESS							
HDL-C	***Lipid, fatty acid and steroid metabolism***	770	5	0.42	+	4.05×10^{-5}	1.26×10^{-3}
	Lipid and fatty acid transport	131	3	0.07	+	4.77×10^{-5}	6.91×10^{-3}
LDL-C	***Lipid, fatty acid and steroid metabolism***	770	4	0.51	+	1.46×10^{-3}	4.52×10^{-2}
	Lipid and fatty acid transport	131	3	0.09	+	8.81×10^{-5}	1.28×10^{-2}
TG	***n.s***						
TC	***Lipid, fatty acid and steroid metabolism***	770	6	1.21	+	1.22×10^{-3}	3.78×10^{-2}
	Lipid and fatty acid transport	131	4	0.21	+	5.55×10^{-5}	8.05×10^{-3}
	Transport	1306	8	2.05	+	8.47×10^{-4}	2.63×10^{-2}
MOLECULAR FUNCTION							
HDL-C	***Lipase***	75	3	0.04	+	9.11×10^{-6}	1.47×10^{-3}
	Apolipoprotein	23	2	0.02	+	1.10×10^{-4}	1.77×10^{-2}
LDL-C	***Transfer/carrier protein***	327	3	0.22	+	1.26×10^{-3}	3.66×10^{-2}
TG	***n.s***						
TC	***n.s***						

Enrichment of a particular "pathway", "biological process" or "molecular function" PANTHER categories were tested by pathway analysis. SNPs that are included in the pathway analysis are selected based on their $p_{discovery}$ values which were 10^{-6} for LDL-C, 10^{-5} total cholesterol, 5×10^{-8} for HDL-C and 10^{-7} for Triglycerides. NCBI: Number of genes that belong to the particular category. Observed:Number of genes that belong to the given particular category among GWAS results. Expected:Expected value for number of genes that belong to the particular pathway among GWAS results. Over/Under: Stands for "overrepresented /under-represented".

* *P*-value corrected for multiple testing.

n.s.: no significant findings.

to I show the predictions based on a TG score which explained up to 0.8 % of the variance in other lipids. HDL-C scores explained up to 0.3 % of the variance in other lipids (Figure 3 A to C).

Pathway analysis

Pathways analyses using only genome-wide significant SNPs was compared to the analogous analyses using SNPs from the polygenic scores which yielded the highest PEV for each trait

(Figure 1). These scores used thresholds of *P*-value < 1×10^{-6} for LDL-C, *P*-value < 1×10^{-5} for TC, *P*-value < 5×10^{-8} for HDL-C and *P*-value < 1×10^{-7} for TG. Table 3 shows the findings from the pathway analysis, based on alternatives to a *P*-value threshold of 5.0×10^{-8}. None of the pathways among categories defined by the PANTHER tool were significant after strict adjustment for multiple testing (Bonferroni correction). With respect to biological processes the *lipid and fatty acid transport,* and *lipid, fatty acid and steroid metabolism* pathways were two biological processes enriched in the HDL-C and LDL-C GWAS findings. At the level of molecular function, genes with an *apolipoprotein* and *transfer/carrier* function were enriched in LDL-C, while genes with a *lipase* function were observed to be significantly enriched among the top GWAS results for HDL-C. For HDL-C and TG, we were not able to select alternative *P*-value thresholds since the highest PEVs were observed with *P*-value < 5×10^{-8}. With respect to LDL-C, the pathway analysis utilizing two different p-value thresholds (*P*-value < 1×10^{-6} and *P*-value < 5×10^{-8}) resulted in the same findings. No additional pathways were identified by using extra information from the risk profiles for LDL-C, TG and HDL-C. For TC, on the other hand, the *lipid, fatty acid and steroid metabolism, lipid and fatty acid transport* and *transport* terms additionally emerged among biological processes tested using the alternative threshold (Table 3).

DISCUSSION

Using prediction modelling, we could explain up to 4.8% of the variance in HDL-C, 2.6% in LDL-C, 3.8% in TG and 2.7% in TC. These PEVs are very similar to those from similar studies[5,9] and much higher than the single SNP analysis of genome-wide significant SNPs from the ENGAGE GWAS (Supplementary Table 1)

However, these proportions are much lower than those identified by GLGC, which were estimated to explain 12.4% (TC), 12.2% (LDL-C), 12.1% (HDL-C), and 9.6% (TG) of the variance in the Framingham Heart Study sample, as mentioned by Teslovich et al.[24]. This is expected since increases in sample size lead to better estimation of the effect sizes of the SNPs and GLGC had a sample size 5 times larger than the ENGAGE sample, which we used as a discovery set in our study.

For all of the traits, the PEV reached a maximum and then decreased with the use of more liberal $P_{discovery}$ thresholds to calculate the polygenic scores (Figure 1). This is most likely explained by the inclusion of more and more biologically non-relevant SNPs, so that the effects of true positive findings are diluted and this is reflected by the decreases in PEV. For all of the studied traits, we found the highest PEV when the polygenic score was based on SNPs with a low $P_{discovery}$ value (5×10^{-8} for HDL-C, 1×10^{-7} for TG, 1×10^{-6} for LDL-C and 1×10^{-5} for TC).

Including the top regions from the ENGAGE GWAS dataset as a separate predictor in the models (Figure 2) uncovered a residual polygenic component which does not explain more than 1% of HDL-C, LDL-C and TC levels. These findings suggest that there are unknown genes with much smaller effects involved in determining these outcomes. However, the PEVs for

these additional variants were small when compared with those for the top findings. For TG, on the contrary, excluding the top regions from the polygenic score resulted in non-significant findings. For TC, which is highly heterogeneous compared to the other traits, it seems that some variants remain to be discovered ($P_{discovery} < 1\times10^{-5}$).

It is of note that among newly discovered loci for HDL-C by GLGC, leading SNPs from 10 loci had *P*-value > 0.05 in the ENGAGE HDL-C analysis. Similar findings were observed for 10 loci for LDL-C, 3 loci for TG and for 9 loci in TC[24]. It is already known that monogenic disorders[25] and rare variants also account for variation in circulating lipid levels[26-32]. This may help to explain why the explained variance is small compared to the high heritability of the traits, especially since many rarer variants are population specific, and might not have been well-represented in our European dataset, or not well tagged by the common SNPs under study. For instance, *APOE* gene variations are tagged by the *CEACAM16-TOMM40* region among the ENGAGE GWAS top findings, and SNPs from this region were not associated to LDL-C levels in RS-II, however, *APOE* ε2 carrier status explains 2.6 % of the phenotypic variation in LDL-C levels in RS-II. Additionally, the gender ratio difference between the discovery and target samples may have been a limitation to the current study, since some loci show different effect sizes for males and females[5].

Our findings have implications for gene discovery and suggest that GWAS of much larger samples may be needed to discover additional variants with small effects for HDL-C and LDL-C. However, at the same time, this study suggests that many of the unknown SNPs have relatively large effects and that is confirmed by the GLGC data. Our findings suggest that GWAS on serum lipids in the future will still be successful as sample sizes increase[14]Our cross prediction results are interesting from a biological perspective. These findings showed very little overlap between the polygenic scores for different circulating lipids. A strong inverse relationship exists between low HDL-C and elevated plasma TG (r = -0.5 in RS-II). Low HDL-C levels are strongly associated with hypertriglyceridemia since high levels of plasma triglycerides drive an exchange reaction for HDL-C cholesteryl esters mediated by *CETP*[33]. In addition, the triglyceride and phospholipids in HDL-C are hydrolyzed by *LIPC*[13, 33]. However, using our genetic evaluation it was not possible to predict a large proportion of the variance in TG levels using HDL-C risk profiles despite the correlation between the two lipids. The polygenic score for TG was slightly better in predicting HDL-C than when we used the top SNPs, however, the PEV did not exceed 0.6 % and was lower than the variance explained by HDL-C SNPs and also lower than the variance explained in circulating TG by TG SNPs. Thus, our data implies that common genetic variants involved in determining both TG and HDL-C levels do not explain the phenotypic correlation between these traits, suggesting that the correlation may be influenced strongly by environmental factors, and/or restricted to a few genes. An alternative explanation may be that we tested the polygenic effects of common variants weighted by their effect size from the initial GWAS. When there are strong causal variants among the top hits that are specific to HDL-C but not to TG, this may dilute the effect of genes with small effect sizes on both outcomes. Also, the current

analyses do not account other forms of genetic variation, such as rare variants or copy number variations (CNVs).

As expected, we also found evidence for a number of genes that regulate both HDL-C and LDL-C (Figure 3A) and a similar overlap between TG and LDL-C (Figure 3 H). TC SNPs were able to explain up to 2.7% of the variation in LDL-C, suggesting that the genes determining LDL-C and TC are for a large part overlapping. This result is in-line with the high phenotypic correlation between the two measures. Genome-wide significant findings from the ENGAGE GWAS harboured two loci (*APOB* and *LPL*) influencing both HDL-C and TG, 2 loci influencing both TG and TC (*DOCK7* and *CEACAM16-TOMM40* regions) and 7 loci influencing both LDL-C and TC (*CELSR2, APOB, ABCG5, HMGCR, FADS2/3, LDLR* and *CEACAM16-TOMM40*). A limitation here is that LDL-C was not directly measured but calculated with the Friedewald formula in the RS-II sample and so, by definition, depends directly on TC, HDL-C and TG. This may cause a potential bias in findings for LDL-C and may inflate the association between lipids in cross prediction findings with this phenotype.

We investigated whether the polygenic score approach can be used as a tool for selecting SNPs of interest in order to further evaluate them in a pathway analysis. First, we evaluated the genome-wide significant SNPs from an existing GWAS and compared the results to those obtained using the SNPs from the polygenic model with the maximum PEV. Neither of the approaches yielded any novel pathways/biological processes (only those already known to be involved in lipid metabolism, such as *cholesterol biosynthesis; lipid and fatty acid transport; and lipid, fatty acid and steroid metabolism).* Also, we see that, although the use of the polygenic score approach did not provide extra information concerning LDL-C, HDL-C or TG, for TC, pathway analysis based on the best predicting polygenic score (with $P_{discovery} < 1 \times 10^{-5}$) was more informative than analysis based solely on the genome-wide significant findings. Including TC SNPs up to a more liberal threshold of 1×10^{-5} suggested three processes which are already biologically known but were not detectable with the 5×10^{-8} discovery threshold. This finding shows that for complex traits like TC, the risk scoring approach might be used to select the SNP cluster which harbours a large number of true positives that are not significant at the genome-wide level. Taken together with the polygenic component analysis results, it is likely that ENGAGE TC-GWAS results harbor undiscovered associated variants distributed between $1\times10^{-6} < P_{discovery} < 1\times10^{-5}$.

Using a gene scoring approach, we tested the evidence of a polygenic component for the heritable circulating lipids. We concluded that a polygenic form of inheritance exists for HDL-C, LDL-C, TG and TC. These findings may be useful for future gene discovery efforts for lipids. We also tested for possible genetic overlap between biologically related lipid traits and compared two different approaches for pathway analysis. This study gives an example of utilizing the risk scoring approach to search for the common genetic background of different quantitative traits, thus; it may also be an example for more sophisticated future studies.

ACKNOWLEDGEMENTS

ORCADES was supported by the Chief Scientist Office of the Scottish Government, the Royal Society and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney. For the MICROS study, we thank the primary care practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. In South Tyrol, the study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano and the South Tyrolean Sparkasse Foundation. Genome-wide genotyping of the Rotterdam Study was supported by NWO (175.010.2005.011). The ERF study was supported by grants from The Netherlands Organisation for Scientific Research, Erasmus MC and the Centre for Medical Systems Biology (CMSB). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.

REFERENCES

1. Isaacs A, Sayed-Tabatabaei FA, Aulchenko YS, et al. Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. *Eur J Epidemiol.* 2007;22(2):99-105.
2. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J, 3rd. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. *Ann Intern Med.* Jul 1961;55:33-50.
3. Kuulasmaa K, Tunstall-Pedoe H, Dobson A, et al. Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations. *Lancet.* Feb 26 2000;355(9205):675-687.
4. Namboodiri KK, Green PP, Kaplan EB, et al. The Collaborative Lipid Research Clinics Program Family Study. IV. Familial associations of plasma lipids and lipoproteins. *Am J Epidemiol.* Jun 1984;119(6):975-996.
5. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* Jan 2009;41(1):47-55.
6. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet.* Feb 2008;40(2):189-197.
7. 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.
8. Kooner JS, Chambers JC, Aguilar-Salinas CA, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet.* Feb 2008;40(2):149-151.
9. 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.* Jan 2009;41(1):35-46.
10. Sandhu MS, Waterworth DM, Debenham SL, et al. LDL-cholesterol concentrations: a genome-wide association study. *Lancet.* Feb 9 2008;371(9611):483-491.
11. Wallace C, Newhouse SJ, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet.* Jan 2008;82(1):139-149.
12. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet.* Feb 2008;40(2):161-169.
13. Hegele RA. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet.* Feb 2009;10(2):109-121.
14. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* Oct 8 2009;461(7265):747-753.
15. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* Aug 5;466(7307):707-713.
16. Visscher PM. Sizing up human height variation. *Nat Genet.* May 2008;40(5):489-490.
17. Demirkan A, Penninx BWJH, Hek K, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Molecular Psychiatry (Accepted for publication).* 2010.
18. International Schizophrenia C, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* Aug 6 2009;460(7256):748-752.
19. Amin N, van Duijn CM, Janssens AC. Genetic scoring analysis: a way forward in genome wide association studies? *Eur J Epidemiol.* 2009;24(10):585-587.
20. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol.* 2009;24(9):553-572.

21. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem.* Jun 1972;18(6):499-502.
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* Sep 2007;81(3):559-575.
23. Thomas PD, Kejariwal A, Campbell MJ, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* Jan 1 2003;31(1):334-341.
24. Teslovich TM, Musunuru K, Smith AV, et al. Meta-analysis of >100,000 individuals identifies 63 new loci associated with serum lipid concentrations. *The American Society of Human Genetics.* Hawaii; 2009.
25. Rader DJ, Cohen J, Hobbs HH. Monogenic hypercholesterolemia: new insights in pathogenesis and treatment. *J Clin Invest.* Jun 2003;111(12):1795-1803.
26. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science.* Aug 6 2004;305(5685):869-872.
27. Cohen JC, Pertsemlidis A, Fahmi S, et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A.* Feb 7 2006;103(6):1810-1815.
28. Gilbert B, Rouis M, Griglio S, de Lumley L, Laplaud P. Lipoprotein lipase (LPL) deficiency: a new patient homozygote for the preponderant mutation Gly188Glu in the human LPL gene and review of reported mutations: 75 % are clustered in exons 5 and 6. *Ann Genet.* Jan-Mar 2001;44(1):25-32.
29. Kotowski IK, Pertsemlidis A, Luke A, et al. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet.* Mar 2006;78(3):410-422.
30. Simha V, Garg A. Inherited lipodystrophies and hypertriglyceridemia. *Curr Opin Lipidol.* Aug 2009;20(4):300-308.
31. Slatter TL, Jones GT, Williams MJ, van Rij AM, McCormick SP. Novel rare mutations and promoter haplotypes in ABCA1 contribute to low-HDL-C levels. *Clin Genet.* Feb 2008;73(2):179-184.
32. Talmud PJ. Rare APOA5 mutations--clinical consequences, metabolic and functional effects: an ENID review. *Atherosclerosis.* Oct 2007;194(2):287-292.
33. Genest JJ, Jr., Martin-Munley SS, McNamara JR, et al. Familial lipoprotein disorders in patients with premature coronary artery disease. *Circulation.* Jun 1992;85(6):2025-2033.

3

Gene discovery studies on depressive symptoms

Chapter 3.1

New candidates for depressive symptoms: Results from a quantitative linkage and exome sequencing in a Dutch centralized facility

Ayse Demirkan, Najaf Amin, Irina V Zorkoltseva, Aaron Isaacs, Antoine van der Sloot, Karin Hek, Lennart C Karssen, Peter Henneman, Michael A Nalls, Nese Direk, Henning Tiemeier, CHARGE Consortium Depression Work Group, UK Brain Expression Consortium, North American Brain Expression Consortium, Abdel Abdellaoui, Andre G Uitterlinden, Albert Hofman, Gert Jan B van Ommen, Mirjam C G M van den Hout, Gisela M Terwindt, Ko Willems van Dijk, Arn M J M van den Maagdenberg, Manfred Kayser, Dorret I Boomsma, Wilfred F J van IJcken, Yurii S Aulchenko, Tatiana I Axenovich, Ben A Oostra, Cornelia M van Duijn.

(submitted)

ABSTRACT

Depressive symptoms are known to be heritable and may serve as phenotypes for gene discovery. We conducted a linkage study using Hospital Anxiety and Depression Scale (HADS-D) and Centre for Epidemiological Studies Depression Scale (CES-D), followed by exome sequencing. Linkage analysis were conducted among participants from the Erasmus Rucphen (ERF) family study, a Dutch isolate that includes over 23 000 people with available genealogy of 23 generations. 2203 living participants with eligible phenotype data were genotyped on the *6K Illumina Linkage IV Panels®*. Variance components linkage analyses were performed using Merlin software. We observed a total of three significant (LOD > 3.3) linkage regions on 9q21 (LOD = 3.68 for HADS-D), 13q33 (LOD = 3.8 for CES-D) and 16q21 (LOD = 3.36 for HADS-D). Also 5 loci 1p36, 3p14, 9q32, 10q22 and 22q11 for HADS-D and one locus 5q34 for CES-D showed suggestive evidence (LOD > 1.9) for linkage. Common SNPs and rare coding variations within the loci were further analysed for association to depressive symptoms. We suggest the involvement of common SNPs within 9q21 and 16q21 region in depression pathology, particularly related to genes *PRUNE2, C16ORF80* and possibly its downstream protein *NDRG4*, based on significant linkage and association, as well as evidence from brain eQTL data. We also show association with a rare missense variation resulting in Asn105Ser change in the *ATP10B* gene on 5q34, observed among 19 carriers from the ERF isolate. Our findings support the hypothesis of independent contribution of common variants with small effects and of family specific rare variants with larger effects in depressive symptoms.

INTRODUCTION

Major depressive disorder (MDD) is one of the most common and disabling diseases in Western societies[1]. Its lifetime risk has been estimated to range from 15 to 30%[2, 3]. MDD results from a complex interplay between several factors, such as stress and socio-economic status, and interaction of those with candidate gene products[4, 5]. The total contribution of genetic factors in the origin of disease, the heritability, is estimated to be 31-42% and the disease is shown to cluster in families[6]. It has been extremely difficult to unravel genes involved in MDD. Candidate gene studies, genome wide linkage and genome wide association studies (GWAS) have not been able to identify robust genetic variants for depression[7, 8]. One explanation for this is the polygenic architecture of depression phenotype which may involve many variants with small effects that are not detectable with the current association studies. A similar hypothesis has been put forward for anxiety, bipolar disorder and schizophrenia[9].

Another explanation may relate to difficulties in diagnosing patients. On one hand there may be (genetic) subgroups of patients within a differential diagnosis[10]. On the other hand there is growing evidence for a joint genetic origin of different psychiatric disorders with very different clinical phenotypes. Although clinically major (unipolar) depression is the relevant outcome,

studies including genome wide association studies (GWAS) have focused on outcomes based on depressive symptomatology and endophenotypes (See chapter 3.2). Depressive symptoms can be measured by questionnaires resulting in scales of depression, such as the Hospital Anxiety and Depression Scale-Depression subscale (HADS-D) and the Centre for Epidemiological Studies Depression Scale (CES-D). Both scales have shown to be heritable in the ERF population ($H^2 = 0.24$, P-value = 0.001 for HADS-D and $H^2 = 0.22$, P-value = 0.001 for CES-D)[11]. Also for these outcomes genetic architecture is complex and involves additive effects of multiple common variants[12]. There also may be major genes leading to depressive symptoms and related traits such as neuroticism, which do not necessarily meet the DSM criteria of MDD, for instance chromosomes 1, 4, 7, 11, 12, and 13 showed significant linkage signals (LOD > 3) in earlier studies indicating these loci may be surrounding genes with larger effect sizes[13]. The advantage of working with depressive symptoms instead of a differential diagnosis of major depression is that depressive symptoms can be used as a quantitative outcome. This has proven to be a very powerful approach in GWAS[14]. Rare variants that run across families also may be responsible for some portion of the heritability of quantitative outcomes and may add upon the already existing underlying genetic component since the common SNPs are insufficient to explain the variation[15].

We studied depressive symptoms of 2203 participants of the Erasmus Rucphen Family (ERF) study using HADS-D and CES-D. The series was earlier used in a large consortium for genome wide association studies of several phenotypes[14, 16]. However, the family based design also allows linkage analyses. The major problem interpreting the linkage findings is that arge regions are identified and genes explaining the linkage are not easily pinpointed. To uncover the genetic variants with relatively large effects, linked regions were analysed in detail using regional association analysis of common SNPs. We also searched for variants with large effects, integrating the exome sequence data for the candidate regions.

METHODS

Study population

The ERF study, which is a part of the Genetic Research in Isolated Populations (GRIP) Program, is a family-based study including over 3000 participants descending from 22 couples that lived in the Rucphen region in the southwest Netherlands in the 19th century. All descendants of those couples were invited to visit the clinical research centre in the region where they were examined in person[17]. The Medical Ethical Committee of the Erasmus Medical Center Rotterdam approved the study and informed consent was obtained from all participants[18]. Participants filled out questionnaires about depressive symptoms and underwent extensive medical examinations in which medication use was assessed[17, 19]. Symptoms of depression were assessed in 2385 participants using depression subscales, namely the HADS-D and the

CES-D[20, 21]. Both questionnaires are validated and are reliable self-report measures of symptoms of depression[22]. The HADS-D consists of 7 items with scores ranging from 0 to 21 and the CES-D of 20 items with total scores ranging from 0 to 60, with higher scores indicating more symptoms of depression[21, 23].

Genotyping and statistical analysis of the linkage study

Illumina's HumanHap6k Genotyping BeadChip (6K Illumina Linkage IV Panels®) was used for genotyping for the linkage analysis. All genotyping procedures were performed according to the manufacturer's protocols. Only markers with MAF > 0.05 were selected for further analysis. Genotyping errors leading to Mendelian inconsistencies were detected using PedCheck[24]. Unlikely double recombination events were detected using MERLIN[25]. All detected errors were eliminated from the data. A total of 5250 autosomal SNPs with a call rate greater than 95% were included in the linkage analyses. ERF is a single large pedigree with multiple loops including 23 612 people, spanning 23 generations. A total of 2203 persons with genotype and phenotype data were included in the initial analysis[25, 26]. Variance component multipoint linkage analysis was performed using --vc option in MERLIN v.1.0.1 software[25]. This program calculates exact IBD sharing probabilities using the Lander-Green algorithm, applying restriction on the pedigree size. Because of this, the pedigree was split in non-overlapping fragments of no more than 18 bits with the help of the PedSTR program[27]. All traits were adjusted for sex and age and rank transformed prior to linkage analysis.

Regions of interest with LOD > 1.9 were selected for further analysis. Borders of the linkage areas were defined as LOD score minus 1 support intervals (LOD-1 SI) around the linkage peaks which were extracted using the "qtl" package implemented in R[28]. Genes within the LOD-1 SI were annotated using SCAN (SNP and CNV Annotation Database), available at http://www.scandb.org/newinterface/index.html. Ingenuity Pathway Tools were used to evaluate the possible biological plausibility of the genes within our linkage regions (©2000-2012 Ingenuity Systems, Inc.)[29]. Gene expression from human brain tissue was extracted from the brain eQTL database of UK Brain Expression Consortium and North American Brain Expression Consortium[30, 31] (See supplementary Text for the methods). The gene expression data in mouse amygdala was extracted from the open source data from the "INIA Amygdala Cohort Affy MoGene 1.0 ST (Mar11) RMA" dataset, which is stored at http://www.genenetwork.org/.

Genotyping and statistical analysis of the association study

Among 2385 phenotyped persons, dense genotypes were available for 1886 subjects, typed on 3 different genotyping platforms (Illumina 318K, Illumina 370K and Affymetrix 250K), which were merged first and then ~ 2.54×10^6 SNPs were imputed using MACH 1.0 software (v1.0.16)[32, 33], using build 36 HapMap (release 22) CEU population as reference. Within each genotyping

batch, only SNPs showing a call rate > 98%, MAF > 1% and HWE P-value > 10^{-6} were used for imputations. Only SNPs that are directly typed at least on one of the platforms were included in the association analysis. Association results of SNPs within the linkage loci were corrected by FDR-based Q-value technic[34]. SNPs that significantly associate with depressive symptoms were checked for their cis effects on brain mRNA expression in frontal lobe and cerebellum[30, 31, 35, 36]. These SNPs were also looked up in data sets of the CHARGE consortium CES-D meta-GWAS which is the largest analysis on quantitative depression scales to date (unpublished) and long term follow up study of anxious depression and neuroticism in the Netherlands Twin Registry (NTR)[37]

Exome sequencing

Six hundred exomes from the ERF pedigree were sequenced "in-house" at the Center for Biomics of the Cell Biology department of the Erasmus MC, The Netherlands, using the Agilent version V4 capture kit on an Illumina Hiseq2000 sequencer using the TruSeq Version 3 protocol. The sequence reads were aligned to the human genome build 19 (hg19) using BWA and the NARWHAL pipeline[38, 39]. Subsequently, the aligned reads were processed further using the IndelRealigner, MarkDuplicates and TableRecalibration tools from the Genome Analysis Toolkit (GATK)[40] and Picard (http://picard.sourceforge.net) to remove systematic biases and to recalibrate the PHRED quality scores in the alignments. After processing, genetic variants were called using the Unified Genotyper tool from the GATK. The effects of the called variants on the respective protein sequences were determined with a custom variant annotation script. For each sample, at least 4 Gigabases of sequence was aligned to the genome. Further for comparison and prediction of the functionality of the variants annotations were also performed using the dbNSFP (database of human non-synonymous SNPs and their functional predictions) and Seattle databases available at: http://snp.gs.washington.edu/SeattleSeqAnnotation131/. These databases gave functional prediction results from four different programs including polyPhen2, SIFT, MutationTaster and LRT, apart from gene and variant annotations. In total about 1.2 million Single Nucleotide Variants (SNVs) were called. After removing the low quality variants (QUAL < 100) we retrieved about 700 000 high-quality SNVs were included in the analysis in this study. Of the 600 individuals with exome sequencing data, 540 had data on the depression scales. Residual depression scores after adjusting for age and sex effects were compared using T-test between mutation carriers and non-carriers. We employed a Bonferroni correction for the number of deleterious mutations selected ($P\text{-value}^{\text{Bonferroni}} = 0.05 / 48 = 0.001$). Gene wise set-based rare variant association analysis were performed using Sequence Kernel Association Test (SKAT)[41]. Current version of SKAT does not consider family relatedness. For this reason the traits were adjusted for relatedness using Grammar approach[42] as implemented in GenABEL[43] procedure "polygenic" prior to the SKAT analysis.

TABLE 1. Descriptives of the study populations

	Linkage study				Association Studies			
	ERF population (N = 2203)		ERF population (N =1889)		CHARGE Depressive Symptoms Meta-GWAS *		NTR (N = 5259)***	
	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range*	Mean (SD)
HADS-D	0-21	6.0 (4.3)	0-21	5.9 (4.3)	ND		0-32	6.95 (5.60)
CES-D	0-58	10.7 (9.6)	0-58	10.63 (9.7)	0-59	7.41 (7.8) **		
Age, years	18.1-90.9	53.1 (14.9)	18.1-89.7	48.19 (14.7)	50-92	66.5 (7.6)	18-63	na
Women (%)	1243 (55.9)		1082 (57.4)		21768 (63.6)		3321 (63.1)	
AD use (%)	156 (7)		130 (6.9)		2661 (7.7)		153 (2.9)	
Education (%)								
Elementary	622 (30.8)		519 (29.9)		5129 (14.8)		891 (16.9)	
Intermediate	1283 (63.5)		1118 (64.4)		14365 (41.6)		1619 (34.9)	
Higher	117 (5.8)		99 (5.7)		15022 (43.4)		2134 (40.6)	

Range, mean and standard deviation (SD) are given for the continuous measurements and percentage is given for the categorical variables. AD; anti-depressant medication. *Details of CHARGE Meta-GWAS can be found in the Supplementary Table 2. **Mean values computed for the CES-D 20 item cohorts. *Anxious depression scale.

RESULTS

The linkage and association data sets of the ERF sample overlapped to large extent (85%) and did not differ in terms of depression scores, female ratio, socio-economical status or depressive medication (Table 1). Figure 1 shows the chromosome plots to which depressive symptoms have been linked significantly (i.e. LOD > 3.3) or suggestively (i.e. 3.3 > LOD >1.9). Although significant peaks varied for HADS-D and CES-D scales, LOD scores over the genome were correlated ($r = 0.34$). Three regions showed significant LOD scores for either of the scales. The highest LOD score of 3.8 was observed for 13q33 to which CES-D was linked. For the HADS-D, a 4 MB region on 9q21 reached a LOD score of 3.68 and a ~5MB region on 16q21 reached a LOD score of 3.36 (details are given in the Supplementary Table 1). A total of 6 regions showed suggestive LOD scores for either of the scales. HADS-D was suggestively linked to 1p36 (LOD = 2.86), 3p14 (LOD = 2.4), 9q32 (LOD = 2.26), 10q22 (LOD = 3.18), and 22q11 (LOD = 2.05) and CES-D suggestively linked to 5q34 (LOD = 2.47).

FIGURE 1. Chromosomes with significant and suggestive linkage peaks.

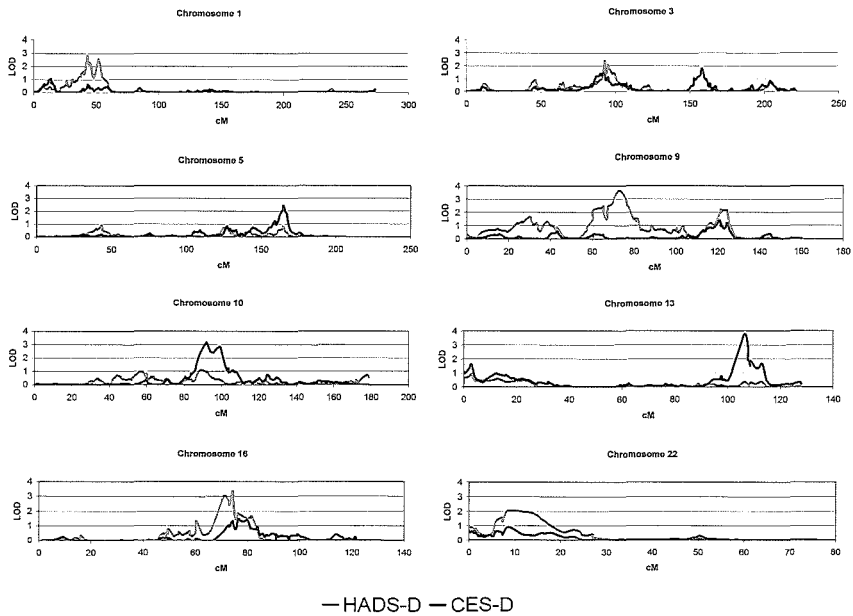


Figure 1 shows the chromosomes with significant and suggestive linkage peaks. Blue lines refer to the CES-D scale whereas pink lines refer to the HADS-D scale.

Common sequence variations within linkage intervals of interest

To narrow down the wide linkage peaks we performed association analysis within the linkage peaks using the imputed genetic SNP data from ERF. Table 2 shows the results from association analyses of SNPs within the linkage peaks. Regional association plots for the selected 9 regions are provided in Supplementary Figure 1. Five variants including rs311452 on 1p36 (*P*-value = 2.09 ´ 10^{-3}), rs2272600 on 5q34 (*P*-value = 0.02), rs7034735 on 9q21 (*P*-value = 7.76 ´ 10^{-3}), rs1989775 (*P*-value = 0.02) on 9q32 and rs9937047 on 16q21 (*P*-value = 5.93 ´ 10^{-4}) showed significant evidence for association with the depression scales in ERF, using FDR-based correction for multiple testing. For each of these five regions, we re-ran linkage analysis conditioning additionally on the associated variant to see if the associated variant explained the observed linkage signals (see Table 2). For HADS-D, for the 9q21 and 16q21 linkage regions the LOD scores dropped in the conditional analysis (to 2.6 and 1.6, respectively) indicating that the associated variants (rs7034735 and rs9937047, respectively) or neighbouring variants in linkage disequilibrium (LD) with them most likely explain the linkage signals. As these were common SNPs, we also analyzed their association to depressive symptoms in the CHARGE consortium and in anxious depression and neuroticism in an ongoing long term follow-up study in the NTR. rs9937047 was nominally significantly associated (*P*-value = 0.01) in the CHARGE CES-D data

TABLE 2. Association findings in the linkage regions in the discovery population (ERF) and the impact on brain eQTL

				ERF study			eQTL effect [30]		
Region	SNP	MAF	Gene	Distance	P-value	Q-value	P-value^{frontal}	P-value^{cerebellar}	Transcript
1p36	rs311452	0.25	RUNX3	95	1.64×10^{-6}	2.09×10^{-3}	0.08	0.86	TMEM50A
3p14	rs4428187	0.36	FRMD4B	Intronic	1.29×10^{-4}	0.10	-	-	-
5q34	rs2272600	0.14	GABRG2	Intronic	2.86×10^{-5}	0.02	0.07	0.71	GABRA1
9q21	rs7034735	0.13	PRUNE2	Intronic	1.42×10^{-5}	7.76×10^{-3}	1.00	0.20	VPS13A
9q32	rs1989775	0.11	PAPPA	262	3.20×10^{-5}	0.02	0.13	0.62	TRIM32
10q22	rs2812541	0.37	C10ORF35	68	2.20×10^{-4}	0.06	-	-	-
13q33	rs954580	0.35	DAOA	10	4.40×10^{-4}	0.09	-	-	-
16q21	rs9937047	0.31	C16ORF80	2	8.08×10^{-7}	5.93×10^{-4}	0.82	0.08	NDRG4
22q11	rs6005346	0.16	SLC7A4	10	1.30×10^{-4}	0.06	-	-	-

Table 2 shows the best association findings in the linkage regions. Gene; closest gene, distance; distance from the closest gene in kilobases. *P*-value; nominal *P*-value from the ERF association study; *Q*-value, false discovery rate for correction for multiple testing; *P*-value^{CHARGE}; *P*-value for the SNP from CHARGE-CES-D meta GWAS. *P*-value$^{frontal\ /\ cerebellar}$; *P*-value for the effect of genotype on gene expression in frontal/cerebellar brain, region wise corrected by the number of probes tested.

and rs1989775 in the NTR maximum anxious depression score (*P*-value = 0.03), with the same direction of effect. (Results shown in Supplementary Table 2). rs9937047 also correlated with *NDRG4* expression in cerebellum in the brain eQTL database. However, the *P*-value region-wide adjusted for multiple testing was only borderline significant ($0.05 < P\text{-value} < 0.1$).

Rare variants in the coding sequence

As common variations did not explain the linkage peaks 1p36, 5q34 and 9q32, we next explored the hypothesis whether the linkage is explained by rare exonic variants in these regions. The results of the search for coding rare variants for these three loci are summarized in Table 3. We focused on relatively rare (frequency < 1-5%) missense that were predicted to be deleterious by at least two of the prediction software used and non-sense variations that were. This selection yielded 48 variants in 38 genes in the linkage intervals of 1p36, 5q34, 9q32, which we further analysed in relation to depression scales. Supplementary Figure 2 shows the means and standard errors for the residual depression scores of mutation carriers and non-carriers after regressing out the effects of age and sex. This effort uncovered a T>C variation among 9 carriers (Figure 2) at position 160113242 in the *ATP10B* (on 5q34) gene that is significantly associated with higher depression scores when adjusted for multiple testing (*P*-value = 0.0001). For this mutation, we screened 600 more individuals in ERF and identified 10 additional carriers. All 19

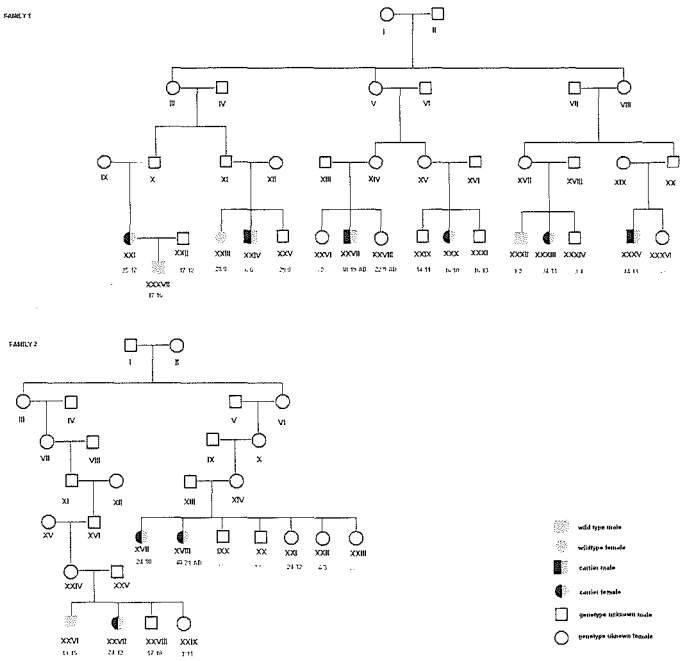
TABLE 3. Selection of the coding variations

Loci	Variants in the coding region				MAF 1%-5%		Predicted to be damaging		
	Synonymous	Missense	Non-sense	Total	Missense	Non-sense	Missense	Non-sense	Genes
1p36	380	504	4	888	115	2	29	2	*E2F2, ID3, MYOM3, RHD, MAN1C1, PAQR7, PAFAH2, TRIM63, UBXN11, AIM1L, DHDDS, GPATCH3, MAP3K6, PTAFR, PHACTR4, GMEB1, EPB41, NKAIN1, SERINC2, COL16A1, SPOCD1, CCDC28B, TMEM234, PHC2, CSMD2*
5q34	38	52	1	91	13	0	7	0	*TTC1, FABP6, C5orf54, SLU7, ATP10B, GABRA6, CCNG1, HMMR*
9q32	97	108	0	205	28	1	9	1	*AMBP, DFNB31, TNC, PAPPA, ASTN2*
Total	515	664	5	1184	156	3	45	3	

Table 3 shows the exonic variants within the LOD-1 support intervals of the linkage loci. Coding region variants were filtered based on their frequency and their deleterious potentials based on three prediction softwares; polyPhen2, Mutation Taster and LRT. Variations that are predicted to be damaging by at least 2 of the 3 prediction softwares were selected for further analysis. MAF; minor allele frequency

carriers could be linked to a single section of the large ERF pedigree (Supplementary Figure 3). The residual CES-D score among the 19 carriers in ERF was 8.23 and was significantly higher than that of non-carriers who had a residual mean score of 1.18 (*P*-value = 0.004). Also the rare variant burden analysis including the five coding variants in *ATP10B* yielded a significant gene-wise SKAT (Sequence Kernel Association test) *P*-value of 0.026 supporting the association of the rare variants with depressive symptoms. None of the other rare variants within the linkage intervals of 1p36, 5q34 and 9q32 reached the Bonferroni corrected *P*-value threshold of significance (*P*-value = 0.001) (Supplementary Table 3).

FIGURE 2. Familial segregation of the *ATP10B* 5:160113242T>C mutation across two pedigrees within ERF.



ATP10B mutation carriers are clustered in two families within the large ERF pedigree. Carriers are indicated in black. CES-D and HADS-D depression scores were given in red beneath the individuals (CES-D / HADS-D). AD; anti-depressive medication use.

DISCUSSION

We performed a quantitative linkage study on depression symptoms and identified 3 significant (9q21, 13q33, 16q21) and 6 suggestive regions (1p36, 3p14, 5q34, 9q32, 10q22, 22q11) either for CES-D or HADS-D scales. Five common variants in these regions were associated with depressive symptoms in ERF. Two of them resulted in a partial drop of the regional LOD scores (i.e. rs7034735 for 9q21 and rs9937047 for 16q21) in the conditional analysis. Of these two SNPs, rs9937047 showed suggestive effect on the expression of a neighbouring gene, *NDRG4* (16q21), in human cerebellum. Our rare variant analysis in the other three regions in which no association to a common variant was found, uncovered a missense coding variation in *ATP10B gene*. In the ERF population, carriers of this missense mutation clustered into a large subfamily in which the carriers had a significantly increased score for depressive symptoms.

Of note, for the most significant locus (13q33; LOD = 3.8 for CES-D) we did not find a possible causal variant. The LOD-1 SI for this region spans 1.2 MB and harbours only one gene; D-amino acid oxidase activator gene (*DAOA*), which is known to associate with schizophrenia and bipolar

disorder[41,42]. The best SNP (rs954580) for this region only showed suggestive (Q-value=0.09) association. This SNP is located in DAOA-anti sense RNA-1 (DAOA-AS1) which may be regulating the DAOA gene[44]. Including this SNP as a covariate in the conditional linkage analysis did not lower the linkage scores, implying that it is unlikely the variant that explains the linkage signals. There were no missense or non-sense variants present in *DAOA* in our exome data. Enlarging the 13q33 linkage interval, i.e. by investigating the region to LOD-2 SI, yielded gene *EFNB2* that had only one coding variant that did not associate with depression scores. As neither association nor exome sequencing elucidated no effect, deep sequencing the region is necessary. Our second best locus was on 9q21 (LOD = 3.62) for HADS-D. The best association SNP at this locus was rs7034535 and when its effect was regressed out, we observed only a relatively small drop in the LOD score for this linkage peak (conditional LOD = 2.61), suggesting that this variant (or perhaps another variant) is in incomplete LD with the causal variant partially explained the quantitative linkage of the trait at this chromosomal location. The SNP is located in *PRUNE2*, a brain expressed gene, which encodes a neuronal protein that has been associated with Alzheimer disease[45, 46] and again deep sequencing may be necessary to further elucidate the gene. HADS-D showed significant evidence for linkage to the 16q21 locus (LOD = 3.36). The best SNP from the ERF association analysis is rs9937047 and is located in the downstream region of the transcription factor *C16ORF80*. Including rs9937047 as a covariate in the linkage analysis dropped the LOD to 1.6, suggesting that rs9937047 or variants is in incomplete LD and can be responsible for the linkage signal. Rs9937047 also shows marginally evidence for association to CES-D in a meta-analysis of GWAS of CES-D,(*P*-value = 0.01) including 35 000 persons from 17 cohorts as part of the CHARGE consortium. This SNP shows suggestive effect on the expression level of *NDRG4* in cerebellum *(P*-value = 0.08). *NDRG4* encodes a cytoplasmic stress response protein which is highly expressed in human brain as well as in heart tissue. It is shown that *NDRG4* deficient mice show decreased levels of brain-derived neurotrophic factor (*BDNF*)[47]. *NDRG4* did not show any coding sequence mutation in the ERF exome data. However, a burden analyses including the intronic, 3'-UTR and 5'-UTR SNPs of *NDRG4* yielded a gene-wise *P*-value of 0.04, suggesting that within the regulating region of the gene there may be relevant variants for depression symptoms.

Among the suggestive linkage regions, the findings at 5q34 were of interest since the region surrounded a cluster of (gamma-amino butyric acid) GABA receptors that function as ligand-gated chloride channels that are dependent on binding of GABA, the major inhibitory neurotransmitter in the mammalian brain. GABA receptors have been implicated in several neurologic and psychiatric phenotypes[47]. Rs2272600 was the best SNP for association in ERF and is intronic to *GABRG2*. This SNP is marginally associated with *GABRA1* expression in frontal brain tissue ($0.05 < P\text{-value} < 0.1$). However, regressing out the effect of this SNP did not change the LOD score, indicating that the SNP does not co-segregate with the linkage variant, which does not exclude the relevance of the association but makes it unlikely that the rs2272600 explains the linkage result at 5q34.

Our analysis of rare coding variants spanning the three linkage regions on 5q34, 9q21 and 16q21 uncovered one mutation which was consistently predicted to be functionally damaging by the three prediction programs. The variant is located in *ATP10B*, which is also located on 5q34, near the GABA receptor cluster. The gene encodes a class V ATPase, with a phospholipid flippase domain. The mutation 5:160113242T>C (rs184217288 in 1000 Genomes) in *ATP10B* results in damaging Asn105Ser change in the transmembrane domain of the protein which takes part in maintaining the phospholipid asymmetry of the cell membrane. Knowledge on class V ATPase proteins is restricted to paralogs of *ATP10B*, *ATP10A* and *ATP10D*, which also cluster together with GABA receptor subunit genes on 15q12 and 4p12[48]. It is of interest that the class V ATPase coding genes and GABA receptor subunit coding genes seem duplicated together, suggesting a possible functional link as well (see Supplementary Figure 4). These three GABA receptor regions are associated with neurologic and psychiatric outcomes, including depressive symptoms and alcohol dependence, Angelman Syndrome and schizophrenia with evidence for involvement of genomic imprinting[49-52]. *GABRB3* on 15q12 is among the top findings of a the largest GWAS for major depressive disorder[53].

Another region of interest is the 1p36. The best associated SNP is located 95 Kb upstream of *RUNX3*. *RUNX3* is a leukemia-related transcription factor and the transcript related to rs311452 is *TMEM50A*, a transmembrane protein with unknown function. Even though the association between depression scores rs311452 had a nominal *P*-value of 1.64×10^{-6} in the discovery ERF study, the SNP did not explain the linkage peak in the conditional linkage analysis. The linkage signal apparently comes from another gene in the region. The LOD-1 SI of the region includes various biologically plausible genes. Gene *HTR1D*, which has been extensively studied for major depressive disorder, is also located in the 1p36 linkage region and maps into the gene network of "*Cell Death, Nervous System Development and Function, Cellular Compromise*" together with genes *BAI2* (brain-specific angiogenesis inhibitor 2), *CNR2* (cannabinoid receptor 2) and *OPRD1* (opioid receptor, delta 1), all located in the same region. However, a coding region analysis of these genes did not yield any SNV that may explain the linkage results asking for a more in-depth analysis.

One limitation of our study at present is that it was not possible to perform conditional linkage analysis for *ATP10B* mutation, since the total number of sequenced people for this mutation was less than the linkage panel (1200 versus 2203). Also, since we do not have complete sequence data on parents, we cannot analyse the parent-of-origin effects that are known to occur in Angelman syndrome. Another limitation is that, taking into account the tissue-specific involvement of depression, gene expression in human amygdala was not measured in our brain eQTL set. Therefore, we extended our search to mouse brain tissue[54]. *NDRG4* and *ATP10B* expression are correlated with each other ($r = 0.382$, *P*-value=0.005) and with *HTR1D* (5-hydroxytryptamine (serotonin) receptor 1D) in mouse amygdala tissue ($r = 0.599$, *P*-value= 2.0×10^{-7} for *NDRG4* and $r = 0.630$, *P*-value = 2.4×10^{-7} for *ATP10B*).

For the linkage locus at 13q33 the best SNP from the ERF association study failed to explain the linkage peaks in the conditional linkage analysis. Neither significant rare coding variants were found within the LOD -1 SI of these regions, except for 5q34. There may be several explanations. Linkage peaks are not precise in highlighting the location of the causal variant. At some loci, the region of interest cannot be easily pinpointed by association analyses. Other reasons may be that we did not take into account alternative mechanisms such as structural and copy number variations (CNVs) or repeats in the linkage regions. Causal rare variants may be located outside the coding sequence, which we did not include in our sequencing analyses.

Combining the resources of linkage and association analysis with those of eQTL data in mouse and human, we found strong evidence that *PRUNE2, C16ORF80, NDRG4 and ATP10B* determine the susceptibility to depressive symptoms. From a genetic perspective the Asn105Ser variant raises the question whether phospholipid asymmetry in the cell is relevant for the occurrence of depressive symptoms. Of these genes *ATP10B* and *NDRG4* can be linked to *HTR1D* in mice, suggesting a role in serotonin pathology.

ACKNOWLEDGEMENTS

ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 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). High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection. North American Brain Expression Consortium (NABEC) was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, part of the US Department of Health and Human Services; project number ZIA AG000932-04. In addition this work was supported by a Research Grant from the Department of Defense, W81XWH-09-2-0128. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov). NABEC scientists are Andrew B Singleton, Mark R Cookson, J. Raphael Gibbs, Dena G Hernandez, Alissa Dilman, Michael A Nalls, Alan B Zonderman, Sampath Arepalli, Luigi Ferrucci , Robert Johnson, Dan L Longo, Richard O'Brien, Bryan Traynor, Juan Troncoso and Marcel van der Brug.

UK Brain Expression Consortium (UKBEC) scientists are John Hardy, Michael E Weale, Mina Ryten, Adaikalavan Ramasamy, Daniah Trabzuni, Colin Smith, H. Ronald Zielke and Robert

Walker. UKBEC was supported by the MRC through the MRC Sudden Death Brain Bank (C.S.), by a Project Grant (G0901254 to J.H. and M.W.) and by a Fellowship award (G0802462 to M.R.). D.T. was supported by the King Faisal Specialist Hospital and Research Centre, Saudi Arabia. Computing facilities used at King's College London were supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. We would like to thank AROS Applied Biotechnology AS company laboratories and Affymetrix for their valuable input.

REFERENCES

1. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Global Burden of Disease and Risk factors; 2006.
2. Kruijshaar ME, Barendregt J, Vos T, de Graaf R, Spijker J, Andrews G. Lifetime prevalence estimates of major depression: an indirect estimation method and a quantification of recall bias. Eur J Epidemiol. 2005;20(1):103-111.
3. Paykel ES, Brugha T, Fryers T. Size and burden of depressive disorders in Europe. Eur Neuropsychopharmacol. Aug 2005;15(4):411-423.
4. Kendler KS, Karkowski LM, Prescott CA. Causal relationship between stressful life events and the onset of major depression. Am J Psychiatry. Jun 1999;156(6):837-841.
5. Lorant V, Deliege D, Eaton W, Robert A, Philippot P, Ansseau M. Socioeconomic inequalities in depression: a meta-analysis. Am J Epidemiol. Jan 15 2003;157(2):98-112.
6. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. Oct 2000;157(10):1552-1562.
7. Levinson DF. The genetics of depression: a review. Biol Psychiatry. Jul 15 2006;60(2):84-92.
8. Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, et al. Meta-analyses of genetic studies on major depressive disorder. Mol Psychiatry. Aug 2008;13(8):772-785.
9. International Schizophrenia C, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. Aug 6 2009;460(7256):748-752.
10. Lamers F, Rhebergen D, Merikangas KR, de Jonge P, Beekman AT, Penninx BW. Stability and transitions of depressive subtypes over a 2-year follow-up. Psychol Med. Feb 17:1-11.
11. Lopez-Leon S, Aulchenko YS, Tiemeier H, Oostra BA, van Duijn CM, Janssens AC. Shared genetic factors in the co-occurrence of symptoms of depression and cardiovascular risk factors. J Affect Disord. May;122(3):247-252.
12. Demirkan A, Penninx BW, Hek K, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry. Jul;16(7):773-783.
13. Boomsma DI, Willemsen G, Sullivan PF, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur J Hum Genet. Mar 2008;16(3):335-342.
14. 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.
15. Lubke GH, Hottenga JJ, Walters R, et al. Estimating the Genetic Variance of Major Depressive Disorder Due to All Single Nucleotide Polymorphisms. Biol Psychiatry. Apr 18.
16. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. Oct 6;478(7367):103-109.
17. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet. Jul 2004;12(7):527-534.
18. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. Ann Hum Genet. May 2005;69(Pt 3):288-295.
19. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22(11):819-829.
20. Radloff. The CES-D scale: a self report depression scale for research in the genreal population. Appl Psychol Measurement. 1977;3:385-401.
21. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. Jun 1983;67(6):361-370.

22. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. Am J Epidemiol. Sep 1977;106(3):203-214.
23. Radloff. The CES-D scale: a self report depression scale for research in theb general population. Appl Pshycol Measurement. 1977;3:385-401.
24. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet. Jul 1998;63(1):259-266.
25. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. Jan 2002;30(1):97-101.
26. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. Nat Genet. May 2000;25(1):12-13.
27. Kirichenko AV, Belonogova NM, Aulchenko YS, Axenovich TI. PedStr software for cutting large pedigrees for haplotyping, IBD computation and multipoint linkage analysis. Ann Hum Genet. Sep 2009;73(Pt 5):527-531.
28. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. May 1 2003;19(7):889-890.
29. Jimenez-Marin A, Collado-Romero M, Ramirez-Boo M, Arce C, Garrido JJ. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. BMC Proc. 2009;3 Suppl 4:S6.
30. Hernandez DG, Nalls MA, Moore M, et al. Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. Neurobiol Dis. Jul;47(1):20-28.
31. Gibbs JR, van der Brug MP, Hernandez DG, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. May;6(5):e1000952.
32. Li Y, Ding J, Abecasis G. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. Paper presented at: American Society of Human Genetics; 9–13 October 2006, 2006; New Orleans, Louisiana.
33. Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive evaluation of SNP genotype imputation. Hum Genet. Mar 2009;125(2):163-171.
34. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. Aug 5 2003;100(16):9440-9445.
35. Trabzuni D, Wray S, Vandrovcova J, et al. MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. Hum Mol Genet. Jul 18.
36. Trabzuni D, Ryten M, Walker R, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. J Neurochem. Oct;119(2):275-282.
37. Middeldorp CM, de Geus EJ, Willemsen G, Hottenga JJ, Slagboom PE, Boomsma DI. The serotonin transporter gene length polymorphism (5-HTTLPR) and life events: no evidence for an interaction effect on neuroticism and anxious depressive symptoms. Twin Res Hum Genet. Dec;13(6):544-549.
38. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. Jul 15 2009;25(14):1754-1760.
39. Brouwer RW, van den Hout MC, Grosveld FG, van Ijcken WF. NARWHAL, a primary analysis pipeline for NGS data. Bioinformatics. Jan 15;28(2):284-285.
40. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. Sep;20(9):1297-1303.
41. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. Jul 15;89(1):82-93.
42. Aulchenko YS, de Koning DJ, Haley C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics. Sep 2007;177(1):577-585.

43. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. May 15 2007;23(10):1294-1296.
44. Chumakov I, Blumenfeld M, Guerassimenko O, et al. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proc Natl Acad Sci U S A. Oct 15 2002;99(21):13675-13680.
45. Li S, Itoh M, Ohta K, et al. The expression and localization of Prune2 mRNA in the central nervous system. Neurosci Lett. Oct 10;503(3):208-214.
46. Potkin SG, Guffanti G, Lakatos A, et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. PLoS One. 2009;4(8):e6501.
47. Yamamoto H, Kokame K, Okuda T, Nakajo Y, Yanamoto H, Miyata T. NDRG4 protein-deficient mice exhibit spatial learning deficits and vulnerabilities to cerebral ischemia. J Biol Chem. Jul 22;286(29):26158-26165.
48. Paulusma CC, Oude Elferink RP. The type 4 subfamily of P-type ATPases, putative aminophospholipid translocases with a role in human disease. Biochim Biophys Acta. Jun 30 2005;1741(1-2):11-24.
49. Vollenweider I, Smith KS, Keist R, Rudolph U. Antidepressant-like properties of alpha2-containing GABA(A) receptors. Behav Brain Res. Feb 2;217(1):77-80.
50. Kertes DA, Kalsi G, Prescott CA, et al. Neurotransmitter and neuromodulator genes associated with a history of depressive symptoms in individuals with alcohol dependence. Alcohol Clin Exp Res. Mar;35(3):496-505.
51. Dan B, Boyd SG. Angelman syndrome reviewed from a neurophysiological perspective. The UBE3A-GABRB3 hypothesis. Neuropediatrics. Aug 2003;34(4):169-176.
52. Seal JL, Gornick MC, Gogtay N, et al. Segmental uniparental isodisomy on 5q32-qter in a patient with childhood-onset schizophrenia. J Med Genet. Nov 2006;43(11):887-892.
53. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry. Apr 3.
54. Chesler EJ, Wang J, Lu L, Qu Y, Manly KF, Williams RW. Genetic correlates of gene expression in recombinant inbred strains: a relational model system to explore neurobehavioral phenotypes. Neuroinformatics. 2003;1(4):343-357.

Chapter 3.2

Genome-Wide Association Study of Depressive Symptoms

Karin Hek, Ayse Demirkan, Jari Lahti, Toshiko Tanaka, Alexander Teumer, Marylin C Cornelis, Najaf Amin, Erin Bakshis, Jens Baumert, Jingzhong Ding, Yongmei Liu, Kristin Marciante, Osorio Meirelles, Michael A Nalls, Yan Sun, Nicole Vogelzangs, Lei Yu, Stefania Bandinelli, Emelia J Benjamin, David A Bennett, Dorret Boomsma, Alessandra Cannas, Laura H Coker Eco de Geus, Philip L De Jager, Ana Diez-Roux, Shaun Purcell, Frank B Hu, Eric B Rimm, David J Hunter, Majken K Jensen, Gary Curhan, Kenneth Rice, Alan D Penman, Jerome I Rotter, Nona Sotoodehnia, Rebecca Emeny, Johan G Eriksson, Denis A Evans, Luigi Ferrucci, Myriam Fornage, Vilmundur Gudnason, Albert Hofman, Thomas Illig, Sharon Kardia, Margaret Kelly-Hayes, Karestan Koenen, Peter Kraft, Maris Kuningas, Joseph M Massaro, David Melzer, Antonella Mulas, Cornelis Mulder, Anna Murray, Ben A Oostra, Aarno Palotie, Brenda Penninx, Astrid Petersmann, Bruce Psaty, Rajesh Rawal, Eric M Reiman, Andrea Schulz, Joshua M Shulman, Andrew B Singleton, Albert Smith, Angelina R Sutin, André G Uitterlinden, Henry Volzke, Elisabeth Widen, Kristine Yaffe, Alan Zonderman, Francesco Cucca, Tamara Harris, Karl-Heinz Ladwig, David J Llewellyn, Katri Räikkönen, Antonio Terracciano, Cornelia M van Duijn, Hans J Grabe, Lenore J Launer, Kathryn L Lunetta, Thomas H Mosley, Anne B Newman, Henning Tiemeier, Joanne Murabito.

A Genome-Wide Association Study of Depressive Symptoms. (Accepted for publication in Biological Psychiatry)

ABSTRACT

Depression is a heritable trait that exists on a continuum of varying severity and duration. Yet, the search for genetic variants associated with depression has had few successes. We exploit the entire continuum of depression to find common variants for depressive symptoms. In this genome-wide association study we combined the results of 17 population-based studies assessing depressive symptoms with the Center for Epidemiological Studies Depression scale (CES-D). Replication of the independent top hits (P-value $< 1 \times 10^{-5}$) was performed in five studies assessing depressive symptoms with other instruments. In addition, we performed a combined meta-analysis of all 22 discovery and replication studies. The discovery sample comprised 34 549 individuals (mean age of 66.5) and no loci reached genome-wide significance (lowest P-value = 1.05×10^{-7}). Seven independent SNPs were considered for replication. In the replication set (N = 16 709), we found suggestive association of one SNP with depressive symptoms (rs161645, 5q21, P-value = 9.19×10^{-3}). This 5q21 region reached genome-wide significance (P-value = 4.78×10^{-8}) in the overall meta-analysis combining discovery and replication studies (N = 51 258). The results suggest that only a large sample comprising more than 50 000 subjects may be sufficiently powered to detect genes for depressive symptoms.

INTRODUCTION

Major depressive disorder (MDD) is a complex disease with an underlying heritable component. Family and twin studies report a high familial tendency of the disorder and heritability estimates of 31-42%[1, 2]. However, the long search for genetic variants associated with depression has had few successes. Several linkage studies for major depressive disorder have been performed, and these identified only one relevant locus[3, 4]. In addition, hundreds of candidate genes have been investigated in association studies, but only six variants have been confirmed in meta-analyses[5, 6]. Recent efforts to find new candidate genes via genome-wide association studies (GWAS) have also been largely unsuccessful[7-15]. GWASs identified interesting regions, but associations with MDD reached standard levels of genome-wide significance at only one locus[15]. Furthermore, only few previously reported candidate genes were replicated in genome-wide association studies[7, 13, 16].

Depression exists on a continuum of varying severity and duration. Depressive symptoms (measured on a continuous scale) and MDD (measured on a dichotomous scale) are associated with similar patterns of risk factors suggesting shared etiology with varying severity[17]. The ability to detect genetic predictors might therefore be improved by analyzing depression quantitatively[18], defining MDD as a diagnostic entity applied to the extreme of the depression continuum[19]. Using the phenotypic variation within cases and controls by analyzing depression quantitatively has been shown to greatly increase the power to detect genetic variants[20]. In fact, a GWAS of the depression facet of personality (a continuous trait), identified several candidate genes. However, the sample size was small and findings remain to be confirmed[21].

In the current study, we exploit the entire continuum of depression defined as the number and severity of depressive symptoms a person experiences. We assessed depressive symptoms with one of the most widely used instruments in the general population, namely the Center for Epidemiological Studies Depression (CES-D) scale. This scale assesses the following major dimensions of depression: depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite and sleep disturbance. The CES-D detects cases of MDD with high sensitivity and specificity[22] and has proven to be relatively stable over time, 82% of older adults had stable CES-D scores over four measurement rounds in ten years[23, 24]. In addition, a high CES-D score, like a diagnosis of MDD, is associated with cardiovascular disease and mortality[25, 26]. Moreover, heritability estimates of depressive symptoms as measured with the CES-D range from 15 to 34%[27-29].

We present the results of a meta-analysis combining genome-wide association results of depressive symptoms from 17 population-based studies of European ancestry (N = 34 549). In addition, we sought to replicate our findings in five samples that used instruments other than the CES-D to quantify depressive symptoms (N = 16 709). Finally, we performed a combined meta-analysis of all discovery and replication studies that included 51 258 individuals.

METHODS

Discovery samples

This discovery set included results from 17 population-based studies comprising a total of 34 549 persons of European descent. The following studies collaborating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium[30] in the United States and Europe were included: the Atherosclerosis Risk In Communities study (ARIC1 and ARIC2[31]), the Cardiovascular Health Study (CHS[32]), the Framingham Heart Study (FHS[33, 34]), and the Rotterdam Study I, II and III (RS-I, RS-II and RS-III[35]). The following population-based studies joined the discovery analyses: the Baltimore Longitudinal Study of Aging (BLSA[36]), The Erasmus Rucphen Family (ERF[37]) study, the Health, Aging and Body Composition (Health ABC) study, the Invecchiare in Chianti (Aging in the Chianti area, InCHIANTI [38]) study, Helsinki Birth Cohort Study (HBCS[39]), Multi-Ethnic Study of Atherosclerosis (MESA[40]), Nurses' Health Study (NHS[41]), Rush Memory and Aging Project (MAP[42]), Religious Orders Study (ROS[43]), and SardiNIA study[44]. All studies were approved by their local institutional review boards and all participants provided written informed consent.

Phenotype definition

Depressive symptoms were measured with the CES-D scale (10 item version (CHS, NHS, RUSH MAP, RUSH ROS), 11 item version (ARIC1), or 20 item version (ARIC2, BLSA, ERF, FHS, HBCS, Health ABC, InCHIANTI, MESA, RS-I, RS-II, RS-III, SardiNIA)). The CES-D scale was designed for use in the general population. All three CES-D versions used here detect the same four latent factors[45]: depressed affect, somatic symptoms, positive affect and interpersonal problems. Each item is scored from 0 to 3 depending on the frequency of the symptoms during the past week. A higher score corresponds to more depressive symptoms. Scores from one examination round per study were used, but CES-D scores have been shown to be relatively stable over time[23, 24]. In studies with multiple CES-D assessments, the round with the largest number of participants (generally the first examination round) was chosen. Persons with schizophrenia or bipolar disorder were excluded based on records, interviews or medication use (these disorders probably have a distinct genetic component). In addition, persons with a Mini Mental State Examination (MMSE) score < 22, indicative of dementia, were excluded. We included persons with genotype data and depressive symptom score who were aged 40 years and older.

Adjustment for use of anti-depressants

In the search for common variants for depressive symptoms in a population-based sample, persons using anti-depressants, who most likely had had depression or depressive symptoms,

increase genetic information. We thus did not exclude these persons from the analysis, but we chose to adjust their total depressive symptoms score for medication use. However, response to anti-depressants is highly variable. In addition, information on compliance is often not available in population-based studies. We therefore used a non-parametric imputation algorithm to adjust the CES-D score for treatment effect. We made two assumptions: 1) the CES-D score of a person using anti-depressants is a right-censored value; the score is lower than the untreated value would be, and 2) persons with a high CES-D score, on average, responded less to their medication than persons with a lower CES-D score. We replaced the score of a person on anti-depressants with the mean depressive symptom score of all persons using anti-depressants that had the same or a higher depressive symptom score. This procedure was performed separately for men and women and was based on an algorithm used for adjustment of blood pressure for persons on antihypertensive drugs[46]. Anti-depressant medication was defined by each study separately to account for differences between countries.

Genotyping and imputation

Genome-wide genotyping was performed by the individual studies on Illumina or Affymetrix platforms. All studies imputed their genotype data to ~2.5 million SNPs to account for the different genotyping platforms. HapMap release 22 CEU build 36 was generally used as reference for imputation (two studies used build 35). Genotype and imputation quality control was performed in each study separately. Each studies genotype and quality control procedure can be found in Supplementary Table 1.

Data analysis

A linear regression was performed on total depressive symptom score, adjusted for age, and gender. The distribution of CES-D scores is skewed, but linear regression is fairly robust to non-normallity. CHS and ARIC additionally adjusted for field study site, NHS for disease status, SardiNIA for self-report versus tester-read and reported answers, FHS for cohort (Offspring, Generation 3). Furthermore, FHS used linear mixed effect models to account for familial correlations. In the ERF study, kinship matrix was used to correct for relatedness.

Meta-analysis

We performed a *P*-value based meta-analysis weighted by sample size. This is a valid approach to account for the different CES-D versions to measure depressive symptoms and for the different distributions of depressive symptoms. The meta-analysis test-statistic was computed using sample size weighted method as implemented in METAL (http://www.sph.umich.edu/csg/abecasis/metal/ [47]. The beta (β) of each individual study *i* was matched to a common coded

allele (the minor allele) for each SNP across all studies. SNPs with a minor allele frequency (MAF) less than 2.5% or an observed to expected variance ratio (imputation quality) less than 0.30 were excluded on a per-study basis. SNPs for which the total sample size was lower than 5000 were removed from the results. Genomic control correction was applied to each study's results.

Replication

Independent top SNPs with a *P*-value $< 1 \times 10^{-5}$ in the discovery meta-analysis were selected with the clumping function in PLINK[48] ($r^2 < 0.05$, 500 Kb) for replication in five studies that measured depressive symptoms with other instruments (total N = 16 709). Persons included in the replication studies were independent from those in the discovery studies. Although replication with other instruments than the CES-D might introduce some heterogeneity, all instruments measure depressive symptoms. Further, a positive replication would ensure that our top hits are not instrument-dependent.

Age, Gene, Environment Susceptibility – Reykjavik Study (AGES-RS[49]), the ARIC study (ARIC3[31]), Monitoring of Trends and Determinants of Cardiovascular Disease/Cooperative Health Research in the Region of Augsburg F3 and F4 (MONICA/KORA[50]), and the Study of Health in Pomerania (SHIP[51, 52]) measured depressive symptoms with Geriatric Depression Scale (GDS), Maastricht Questionnaire (MQ), Patient Health Questionnaire (PHQ-9) and the Beck Depression Inventory-II (BDI-II), respectively. The BDI-II, GDS and PHQ-9 aim to screen for depression and are highly correlated[53, 54]. The BDI-II is based on the DSM-IV criteria for MDD and comprises 21 items on a scale of 0 to 3 with higher scores indicating more severe depressive symptoms over the past two weeks. The PHQ-9 is like the BDI-II based on the DSM-IV criteria for MDD, but it consists of nine items on a scale of 0 to 3 to assess depressive symptoms over the past two weeks. The GDS was specifically designed to screen for depression in older adults and comprised 15 items answered with 'yes' or 'no'. The Maastricht Questionnaire (21 items), although designed to measure vital exhaustion, correlates with measures of depressive symptoms [55] and was previously used to assess depressive symptoms[56, 57].

Replication was considered significant if the Bonferroni corrected *P*-value for testing 7 SNPs was ≤ 0.050 (uncorrected *P*-value $\leq 7.1 \times 10^{-3}$).

Pathway analysis

Protein ANalysis THrough Evolutionary Relationships (PANTHER[58]) was used to identify and classify biological processes among the SNPs associated with *P*-value $<10^{-4}$ from the overall meta-analysis (N = 51 258). After SNP selection, SNP ids were annotated to genes and/or flanking genes with the SCAN SNP and CNV Annotation Database (http://www.scandb.org). PANTHER then compares this gene list to a reference list (Homo Sapiens gene list from the National Cen-

ter for Biotechnology Information) using the binomial test. Results were Bonferroni-corrected to account for multiple testing.

Candidate gene search

Altogether 17 SNPs variants previously reported to be associated to depression were selected: 1 SNP that has been found genome-wide significantly associated with depressive phenotypes after replication[7, 59], 4 top SNPs from the largest MDD meta-analysis so far[13], and 12 top SNPs from the only published GWAS that studied a depressive trait continuously[21]. SNPs were tested for association in the discovery meta-analysis (N = 34 549) and in the overall meta-analysis including all studies that measured depressive symptoms (N = 51 258).

RESULTS

Meta-analysis of depressive symptoms

Table 1 shows the characteristics of the study populations. Mean age in the discovery studies ranged between 55.9 and 80.8 years. The percentage of women varied between 44.6% and 100%. In line with the population-based design of the studies, median depressive symptoms scores ranged between two and ten for the CES-D 20 item version. This is well below the cut-off of 16 at which major depression cases in older adults can be identified with high specificity and sensitivity[22]. The percentage of persons scoring above this cut-off varied between 4.7% and 27.1%. Distributions of CES-D scores differed between studies and so a Z-score-based meta-analysis was used to combine the individual studies' results. Anti-depressant use ranged from 3.0% to 14.0%. On average, CES-D scores for persons on anti-depressants more than doubled after imputation.

The genomic control inflation factor lambda (λ_{gc}) for each study ranged between 0.997 and 1.024. A meta-analysis of 17 studies (N = 34 549) with depressive symptoms measured by CES-D was performed (Q-Q and Manhattan plots in Supplementary Figure 1). The total number of SNPs analyzed was 2 391 896. No association reached the pre-specified genome-wide significance level of 5×10^{-8} for the association with the depressive symptom score. However, we identified 117 SNPs with a *P*-value $< 1 \times 10^{-5}$, which included seven independent top SNPs ($r^2 <$ 0.05 in 500 Kb, Table 2). The SNP with the lowest *P*-value was rs8020095 (*P*-value $=1.05 \times 10^{-7}$) and maps to an intronic region of *GPHN* on chromosome 14. Of the seven top SNPs none had a heterogeneity p-value (tested by Cochran's Q) below 0.05 in the discovery meta-analysis.

We reran the analysis for the independent top SNPs excluding people on anti-depressants. *P*-value of the top SNPs shifted towards one (e.g. rs8020095 *P*-value 1.56×10^{-6}, rs161645 *P*-value 1.71×10^{-3}). Adding five points to the total score for people using anti-depressants in

TABLE 1: Study sample characteristics of discovery and replication samples

			Depressive symptom score										International Standard Classification of Education**				
Sample	Instrument	N	Mean	(Sd)	Median	(range)	≥ 16 %*	Anti-Depressant users %	Mean Age	(SD)	Female %	Current smokers %	Level 0/1 %	Level 2 %	Level 3 %	Level 4 %	Level 5/6 %
	Discovery Studies (n=34 549)																
ARIC1	CES-D 11	393	3.80	(3.57)	3	(0-18)	9.92	14.0	72.7	(5.46)	59.5	19.6	2.0	8.1	35.4	7.9	46.6
ARIC2	CES-D 20	614	8.52	(7.41)	6	(0-34)	16.1	11.1	71.0	(5.60)	49.7	19.7	3.1	8.3	34.7	11.7	42.2
BLSA	CES-D 20	764	6.90	(6.5)	5	(0-55)	8.51	NA	71.6	(13.8)	44.6	3.0	0.4	1.5	11.0	12.4	74.8
CHS	CES-D 10	3155	4.27	(4.29)	3	(0-26)	11.3	3.11	72.2	(5.29)	61.2	11.0	2.5	12.3	38.6	9.3	37.2
ERF	CES-D 20	1297	12.7	(10.9)	10	(0-59)	27.1	8.20	55.9	(10.1)	56.7	43.2	40.4	42.5	13.6	NA	3.5
FHS	CES-D 20	4956	7.25	(8.21)	4	(0-53)	10.3	10.4	56.1	(10.5)	53.3	14.7	0.5	3.1	32.2	24.9	39.2
HABC	CES-D 20	1654	4.93	(5.78)	3	(0-43)	4.70	3.60	73.8	(2.80)	47.1	6.4	11.9	NA	34.4	53.6	NA
InCHIANTI	CES-D 20	942	11.8	(8.24)	10	(0-46)	24.6	3.40	70.4	(9.85)	52.8	18.5	73.5	11.2	7.3	4.6	3.4
RSI	CES-D 20	3791	4.86	(7.35)	2	(0-52)	7.30	3.80	72.7	(7.21)	58.5	16.4	31.4	29.0	29.8	NA	9.8
RSII	CES-D 20	2093	5.81	(7.90)	3	(0-48)	9.70	5.00	64.8	(8.03)	54.5	19.6	21.6	35.6	27.1	NA	15.7
HBCS	CES-D 20	1386	9.58	(8.68)	7	(0-53)	19.4	4.70	63.4	(2.86)	59.7	23.0	33.0	18.4	26.0	NA	22.5
MESA	CES-D 20	2423	6.93	(6.87)	5	(0-50)	10.0	12.2	62.7	(10.2)	52.2	11.4	1.6	3.4	16.5	28.4	50.1
NHS	CES-D 10	5891	6.36	(4.50)	6	(0-26)	15.9	13.3	71.7	(6.70)	100	5.5	0	0	0	72.6	27.4
RSIII	CES-D 20	2041	6.32	(8.22)	3	(0-53)	9.90	6.90	56.0	(5.67)	56.1	22.4	9.8	35.0	28.4	NA	26.8
RUSH MAP	CES-D 10	825	1.38	(1.75)	1	(0-8)	20.1	13.6	80.8	(6.53)	73.0	2.4	1.7	27.4	19.9	42.8	8.2
RUSH ROS	CES-D 10	778	1.10	(1.51)	1	(0-8)	13.9	9.00	75.5	(7.24)	66.5	2.1	1.3	5.4	3.1	46.0	44.2
SardiNIA	CES-D 20	1438	11.9	(8.20)	10	(0-53)	25.2	3.00	58.0	(11.4)	59.5	NA	28.9	50.3	16.1	NA	4.8

TABLE 1: Study sample characteristics of discovery and replication samples (continued)

		Depressive symptom score										International Standard Classification of Education**				
Sample	Instrument	N	Mean	(Sd)	Median	(range)	≥ 16 %*	Anti-Depressant users %	Mean Age	(SD)	Female %	Current smokers %	Level 0/1 %	Level 2 %	Level 3 %	Level 4 %	Level 5/6 %
Replication studies (n=16 709)																	
AGES-RS	GDS	2855	2.58	(2.26)	2	(0-15)	9.92	13.8	76.4	(5.46)	58.0	12.7	22.1	16.8	NA	33.3	27.8
ARIC3	MQ	8918	10.2	(8.79)	8	(0-42)	9.39	4.04	57.2	(5.67)	52.7	23.8	4.8	10.2	36.4	9.2	39.4
MK F3	PHQ-9	1433	3.52	(3.54)	3	(0-26)	6.80	NA	60.5	(9.13)	51.3	14.3	12.1	56.4	17.6	0.8	13.1
MK F4	PHQ-9	1807	3.36	(3.3)	3	(0-27)	5.50	NA	60.9	(8.85)	51.5	14.6	10.0	52.4	22.6	1.1	14.0
SHIP	BDI-II	1696	6.44	(7.11)	4	(0-58)	8.90	NA	59.4	(11.6)	51.4	25.5	5.1	0.3	60.4	15.9	18.4

Abbreviations: ARIC, Atherosclerosis Risk in Communities study; BLSA, Baltimore Longitudinal Study of Aging; CHS, Cardiovascular Health Study; ERF, Erasmus Rucphen Family study; FHS, Framingham Heart Study; HABC, Health, Aging and Body Composition study; InCHIANTI, Invecchiare in Chianti; RS, Rotterdam Study; HBCS, Helsinki Birth Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses Health Study; RUSH MAP, RUSH Memory and Aging Project; RUSH ROS, RUSH Religious Orders Study; SardiNIA, SardiNIA study; AGES-RS, Age, Gene, Environment Susceptibility – Reykjavik Study; MK, Monitoring of trends and determinants of cardiovascular disease/cooperative health research in the region of Augsburg (MONICA/KORA); SHIP, Study of Health In Pomerania; CES-D, Center for Epidemiologic Studies Depression scale; GDS, Geriatric Depression Scale; MQ, Maastricht Questionnaire; PHQ-9, Patient Health Questionnaire-9 items; BDI-II, Beck Depression Inventory –II; CIDI, Composite International Diagnostic Interview; sd, standard deviation. ARIC1, ARIC2, ARIC3, RSI, RSII , RSIII and MKF3, MKF4 included unique individuals.

* Cut-off for screen positives was 9 for ARIC1, 8 for CHS, 9 for NHS, 3 for RUSH MAP and RUSH ROS, 6 for AGES-RS, 24 for ARIC3, and 17 for SHIP.

** Level 0: pre-primary education, level 1: primary education or first stage of basic education, level 2: lower secondary education or second stage of basic education, level 3: (upper) secondary education, level 4: post-secondary non-tertiary education, level 5: first stage of tertiary education, level 6: second stage of tertiary education.

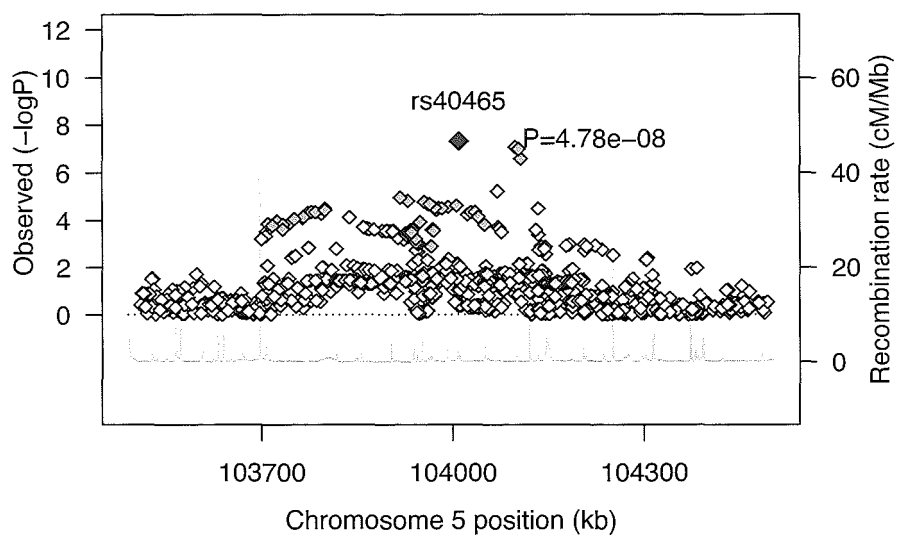
a subsample (RS-I, RS-II, RS-III, N = 7 925) resulted in the same top SNPs and similar *P*-value for the top SNPs tested here.

Replication

Table 2 presents the results of the replication analysis and the overall meta-analysis across discovery sample and replication sample. The mean observed to expected variance ratio for the seven top SNPs across all cohorts ranged between 0.91 and 0.98 (Supplementary Table 2). In the replication sample, a SNP on chromosome 5 showed an association with depressive symptoms (5q21, rs161645, *P*-value = 9.19×10^{-3}, Table 2), but this association did not reach the pre-defined threshold for multiple testing (corrected for multiple testing *P*-value = 0.064). This SNP resides in a gene desert, with the closest gene *NUDT12* more than 1000 Kb away.

In the overall meta-analysis including discovery and replication samples (N = 51 258), SNP rs40465 reached genome-wide significance (*P*-value = 4.78×10^{-8}). This SNP is in high LD with SNP rs161645 ($R^2 = 0.80$). Rs40465 had a *P*-value of 2.58×10^{-6} in the discovery meta-analysis and a *P*-value of 5.00×10^{-3} in the meta-analysis of replication studies. An association plot of the 5q21 region is presented in Figure 1.

FIGURE 1: Association results in the 5q21 region.



Summary of the association of SNPs on chromosome 5 (base 103 500 000 to 104 500 000) with depressive symptoms from the overall meta-analysis (N = 51 258). The SNP with the strongest association (rs40465) is highlighted in blue and its corresponding *P*-value is given. Other SNPs are coloured according to their degree of linkage disequilibrium with rs40465, ranging from high LD (orange, r^2 0.5-1.0) to low LD (white, $r^2 < 0.2$).

TABLE 2: Meta-analysis results of CES-D depressive symptom score in discovery studies, replication of results in studies that measured depressive symptoms with other instruments, and overall meta-analysis of all studies

								Discovery Meta-analysis CES-D N = 34 549		Replication Other instruments N = 16 709		Overall Meta-analysis N = 51 258	
SNP[#]	Chr	Position	SNPs (N)[##]	Closest Gene	Distance (bp)	Allele	MAF	Direction	*P*-value	Direction	*P*-value	Direction	*P*-value
rs8020095	14	66 523 611	2	GPHN	intron	A/G	0.17	+(+++-+++++++-+++-+?)	1.05×10^{-7}	-(-?--+)	0.79	+	3.04×10^{-6}
rs8038316	15	52 560 732	3	UNC13C	intron	A/G	0.05	-(-?--------+----+-)	1.24×10^{-6}	-(----+)	0.42	-	9.64×10^{-6}
rs161645	5	104 097 816	3	NUDT12	1 171 427	A/G	0.34	+(++++++++++++-+++-+?)	2.32×10^{-6}	+(+++-+)	9.19×10^{-3}	+	8.39×10^{-8}[###]
rs357282	5	38 904 792	0	OSMR	intron	T/G	0.13	+(+++++++++-++++++)	7.56×10^{-6}	+(-+--+)	0.87	+	1.60×10^{-4}
rs4653635	1	223 662 313	3	LBR	intron	A/G	0.16	-(---+----++-------)	8.14×10^{-6}	+(-+++-)	0.55*	-	8.89×10^{-4}
rs4594522	20	30 718 645	5	COMMD7	35 508	C/T	0.36	-(----+---+------+--)	9.29×10^{-6}	-(-+++--)	0.80	-	1.56×10^{-4}
rs13137117	4	94 673 387	9	GRID2	intron	T/A	0.25	+(+++-+++-+++++++++-)	9.77×10^{-6}	+(-+-+-)	0.97	+	2.63×10^{-4}

[#]Independent SNPs with a p-value<$1x10^{-5}$ in the discovery meta-analysis.

[##]Supporting SNPs: Number of SNPs in linkage disequilibrium with the topSNP (r^2>0.8), with a *P*-value < 10^{-4}.

[###]Lowest *P*-value of the overall meta-analysis *P*-value = 4.78×10^{-8} for SNP rs40465 (G/T) that is in LD (r^2=0.80) with rs161645, discovery *P*-value = 2.58×10^{-6} (++++++++++++++-+?), replication *P*-value = 5.00×10^{-3} (+++-+).

* heterogeneity *P*-value < 0.05

** The total N for SNP rs8020095 was 40 902, for rs8038316 48 103, for rs161645 49 820 and for the other SNPs 51 258. The mean observed versus expected variance ratio (measure of imputation quality) for imputed SNPs ranged between 0.91 and 0.99. Supplementary Table S2 includes this information detailed per SNP.

Chr: chromosome; Allele: minor/major on the + strand, the minor allele is the coded allele; MAF: minor allele frequency.

Direction; Overall direction and direction of effect for each study are shown. Direction of effect Discovery: FHS, CHS, RS-I/RS-II/ RS-III, ARIC1, ARIC2, ERF, InCHIANTI, Health ABC, BLSA, HBCS, MESA, NHS-Brc, NHS-CHD, NHS-KID, NHS-T2D, RUSH-MAP, RUSH-ROS, SardiNIA. ?=not tested. Direction of effect Replication: AGES-RS, ARIC3, MK F3, MK F4, SHIP. ?=not tested.

In contrast, the strength of the associations of the other top SNPs with depressive symptoms was attenuated as judged by the *P*-value. All SNPs with a *P*-value < 1×10^{-4} from the overall meta-analysis (N = 51 258) are presented in Supplementary Table 3.

Pathway analysis

One hundred and four functional genes of the 170 genes that were annotated were mapped to biological processes. Relevant processes that were overrepresented amongst top SNPs (*P*-value < 10^{-4}) of the overall meta-analysis were neurotransmitter secretion (Bonferroni-corrected *P*-value = 9.84×10^{-3}), vitamin transport (Bonferroni-corrected *P*-value = 0.014), and synaptic transmission (Bonferroni-corrected *P*-value = 0.037). A complete list of biological processes that were significantly overrepresented is presented in Table 3.

TABLE 3. Pathway analysis

Biological Process	NCBI	Observed	Expected	Over/under	Adjusted P-value*
neurotransmitter secretion	346	6	1.81	+	9.84×10^{-3}
vitamin transport	95	3	0.5	+	0.014
protein metabolic process	3240	26	16.92	+	0.015
synaptic transmission	594	7	3.10	+	0.037
transport	2857	22	14.92	+	0.038
vesicle-mediated transport	1160	11	6.06	+	0.040
cation transport	621	7	3.24	+	0.045
cell-cell signaling	1331	12	6.95	+	0.045
protein transport	1646	14	8.60	+	0.048
intracellular protein transport	1646	14	8.60	+	0.048

Enrichment of biological processes among our top results (overall meta-analysis *P*-value < 10^{-4}) was statistically tested with a binomial test.
NCBI: number of genes in a biological process (reference). Observed: number of genes that belong to a biological process amongst the GWAS results. Expected: expected number of genes that belong to a biological process in the GWAS results. Over/under: over-representation or under-representation of the genes in our results. *A Bonferroni-correction was applied to correct for multiple testing.

Candidate gene search

None of the 17 tested candidate genes were replicated in the current study (Supplementary Table 4). Nine out of seventeen associations had the same direction in our overall meta-analysis as in the published study, and none of the nine was significant (uncorrected for multiple testing).

DISCUSSION

In this GWAS of depressive symptoms we combined the results of 17 population-based studies with 34 549 individuals to find common variants for depressive symptoms. Including the five replication studies, this effort comprised data from 51 258 independent individuals. Of the seven SNPs we attempted to replicate, we found suggestive evidence for the observed association of one SNP in the 5q21 region with depressive symptoms. This region reached genome-wide significance when tested over all studies (N = 51 258).

Although evidence shows that depression can be well represented by a continuum of depressive symptoms we observed a genome-wide significant hit in this large GWAS only when pooling all studies with depressive symptoms. This difficulty of finding signals is in line with GWASs of major depression. Nine GWASs of depression, of which the largest comprised ~6000 MDD cases and ~7000 controls, yielded only one genome-wide significant finding[15].

The approach of studying depression on a continuum has the advantage that not only information on extremes is used, but that all available information is exploited. Van der Sluis et al.[20] showed that if the phenotypic variation among cases as well as the variation among controls is used this greatly increases the power to detect genetic variants. However, studying depression along a continuum in population-based studies implies that many individuals have a low depressive symptoms score and that few persons score high. Therefore, it remains to be validated whether the results presented here are generalizable to clinical depression cases. In addition, the CES-D measures current depressive symptoms, and not remitted depressive symptomatology. This introduced false-negatives, but in this population-based approach in which low depressive symptomatology is overrepresented the resulting bias would be conservative. Furthermore, the distribution of depressive symptoms differed between cohorts. We therefore performed a p-value based meta-analysis, which is a valid approach, but has the consequence that we cannot draw conclusions on effect sizes.

Differences in depressive symptoms distribution do not impact on the validity of the findings. People with high depressive symptoms are more likely to carry risk variants, but this should not depend on the number of people with a high score. Furthermore, the distribution of I^2, a measure of heterogeneity[60], of the results combining all samples did not differ from the distribution of I^2 of the results when meta-analyzing samples with low or high depression prevalence were analyzed separately. No excess heterogeneity was observed, which suggests that depressive symptoms can be analyzed linearly. However, some genetic main effects may be more detectable in very homogeneous populations. Observed differences in distributions of depressive symptoms may have resulted from environmental factors and if these in turn interact with specific genetic variants, only very homogeneous studies could also detect a genetic main effect.

Environmental factors, like education level, differed among cohorts. In observational research, one would have controlled for such possible confounders. In genetic studies

confounding by environmental factors is unlikely to occur, but controlling for environmental factors can also be done to increase precision, i.e. reduce the variance in depressive symptoms. However, environmental factors explain very little variance in depressive symptoms. Therefore, the benefit of performing additional controlled analyses will be negligible and offset by running several models with the risk of multiple testing.

In the current study, depressive symptoms scores for people using anti-depressants were imputed to take into account the high variability in response to anti-depressants. In an analysis of depressive symptoms, people on anti-depressants, who most likely had had depression or depressive symptoms, are particularly informative. Therefore excluding this group a priori may have changed the results. In a subsample, the imputation algorithm used in the current study yielded similar results as adding an arbitrary score of five points to the depressive symptoms scores of people using anti-depressants.

This study was performed in older adults. Cerebrovascular burden and cognitive impairment, which have a relatively high prevalence in old age, are known to be associated with depressive symptoms. In addition, while a high CES-D score indicates depressive symptoms it can also be suggestive of, for example, anxiety[61]. In other words, the level of depressive symptoms is a clinically heterogeneous phenotype. However, the genetic background of clinically heterogeneous phenotypes like anxiety and depression may be more uniform than the clinical presentation suggests[62]. In addition, while non-genetic determinants of depression may differ with age, genetic determinants were shown to be stable at different ages[63, 64]. Therefore, the results presented here are presumably generalizable to younger populations.

We combined results from studies that measured depressive symptoms with instruments other than the CES-D to replicate the association between depressive symptoms and seven independent top SNPs. In an overall meta-analysis we tested whether any variation introduced by different instruments was offset by the increased power. In the replication effort, one SNP (5q21 region) reached a *P*-value below 0.05, but did not pass this threshold when controlling for multiple testing. Another SNP in the 5q21 region, however, reached genome-wide significance when the association across discovery and replication studies was tested (N =51 258). The 5q21 region resides in a gene desert with the closest gene, *NUDT12*, lying more than 1000 Kb away, and which has not been previously implicated in psychiatric disorders.

Although we observed suggestive association of the 5q21 region with depressive symptoms, genome-wide significance was observed only after pooling the results of the discovery and replication studies. Also, we could not replicate associations with candidate genes that previously have been reported to be associated with depression. Several explanations are plausible.

A first explanation for these observations is that the top SNPs identified in this study are false-positive findings. However, the discovery set was large and although we did not find any genome-wide significant hits, true hits are expected to be found amongst the top findings. A pathway analysis on the results of the overall meta-analysis showed that biological processes that play a role in depression were overrepresented amongst our top hits.

Second, the replication sample was smaller than the discovery sample and may be underpowered to detect true effects with moderate effect sizes which might have been overestimated in the discovery analysis (winner's curse). Indeed, we found suggestive evidence of association for only one of seven SNPs, but the direction of association was compatible for five out of seven SNPs.

Third, lack of replication might be related to heterogeneity of the replication phenotype. In the replication approach, we combined the results of studies that measured depressive symptoms with different instruments. Instruments were also administered at different timepoints across studies. However, the instruments have been reported to be highly correlated (correlations between 0.77 and 0.86) and relatively stable genetic determinants over the life span were observed in an Australian Twin study[53, 54, 63, 65, 66].

Several other factors can hinder the search for common variants associated with depressive symptoms. Population stratification for example, can result in false-positive findings. To avoid population-stratification, only individuals from European descent were included. Including only individuals from European descent also minimized measurement error caused by cultural differences in responses to the CES-D[67]. Other possible explanations are the presence of genetic heterogeneity[68], gene-gene interactions[69], and gene-environment interactions. The interaction between candidate genes and life events has been repeatedly studied for depression[70]. However, to study this phenomenon in a genome-wide approach requires much larger data sets[13]. In addition, it is suggested that the gain of gene-environment interaction studies over studies of main effects for complex diseases like depression is minimal[71]. The study described here focused on common genetic variation, but rare variants or copy number variations (CNVs) not tagged by SNPs might play a role in depression[72, 73]. Using a larger reference panel, like the haplotypes generated by the 1000 genomes project, would have improved the yield of rare variants. Harmonizing imputation reference and imputation tools might have further increased the power of the study to detect associations. Also, not single SNPs, but many SNPs collectively, each with a very small effect, may affect the susceptibility for depressive symptoms[64].

In conclusion, the efforts of a large collaboration to identify common variants associated with depressive symptoms yielded no genome-wide significant hit in the discovery sample. In the replication approach we found suggestive evidence for a SNP in the 5q21 region. When analyzing the discovery and replication samples, one genome-wide significant hit in this region was observed. Further investigation of the 5q21 region is necessary to verify the association with depressive symptoms and to pinpoint the possible functional variant. Such a future study of depressive symptoms could analyze this phenotype stratified by gender and incorporate longitudinal information with repeated measures of depressive symptoms to provide more power to our search for potential candidate genes.

ACKNOWLEDGMENTS

CHARGE: The authors acknowledge the essential role of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium in development and support of this manuscript. CHARGE members include the Netherland's Rotterdam Study (RS), the NHLBI's Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), the NHLBI's Atherosclerosis Risk in Communities (ARIC) Study, and the NIA's Iceland Age, Gene/Environment Susceptibility (AGES) Study. AGES-RS: The Age, Gene/Environment Susceptibility Reykjavik Study has been funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). ARICThe Atherosclerosis Risk in Communities Study: The research is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, and grants R01-HL087641, R01-HL093029 and R01-HL70825; National Human Genome Research Institute contract U01-HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. BLSA: This research was supported entirely by the Intramural Research Program of the NIH, National Institute on Aging. CHS: This CHS research was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133 and NHLBI grants HL080295, HL075366, HL087652, HL105756 with additional contribution from NINDS. Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the NIA. See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by CTSI Grant UL1RR033176 to the Cedars-Sinai General Clinical Research Center Genotyping core, National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center, and the Governors' Chair in Medical Genetics (JIR). ERF: 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. The genotyping for the ERF study was supported by EUROSPAN (European Special Populations Research Network) and the European Commission FP6 STRP grant (018947; LSHG-CT-2006-01947). The ERF study was further supported by grants from the Netherlands Organisation for Scientific Research, Erasmus MC, the Centre for Medical Systems Biology (CMSB) and the Netherlands Brain Foundation (HersenStichting Nederland). We are grateful to all participating individuals and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection. FHS: The phenotype-genotype association analyses were supported by R01-AG29451. "This research was conducted in part using data and

resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center." HBCS: Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, European Science Foundation (EUROSTRESS), Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation, Juho Vainio Foundation, and Wellcome Trust (grant number WT089062). We thank all study participants as well as everybody involved in the Helsinki Birth Cohort Study. Health ABC: Research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. Dr. Yaffe is supported by NIH grant R01 MH086498. InCHIANTI:The Invechhiare in Chianti (InCHIANTI). The InCHIANTI)Study was supported as a "targeted project" (ICS 110.1RS97.71) by the Italian Ministry of Health, by the U.S. National Institute on Aging (Contracts N01]AG]916413, N01]AG] 821336, 263 MD 9164 13, and 263 MD 821336) and in part by the Intramural Research Program, National Institute on Aging, National Institutes of Health, USA. MESA: The MESA SHARe project is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159 through N01-HC-95169 and UL1-RR-024156. Funding for genotyping was provided by NHLBI Contract N02-HL-6-4278 and N01-HC-65226. Funding for this project was also provided by # R01 HL101161. The MONICA/KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, and supported by grants from the German Federal Ministry of Education and Research (BMBF). Furthermore, the research was supported within the Munich Center of Health Sciences (MC Health) as part of LMU. NHS: The Nurses' Health Studies are supported by NIH grants CA 65725, CA87969 (National Cancer Institute), CA49449, CA67262, CA50385 and 5UO1CA098233. RS: 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 MC and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. Henning Tiemeier was supported by the VIDI grant of ZonMw (2009-017.106.370). Karin Hek was supported by a grant from BavoEuropoort. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database, and Karol Estrada and Maksim V. Struchalin for their support in creation and analysis of imputed data. We would like to thank Karol Estrada, Dr. Fernando Rivadeneira, Dr. Tobias A. Knoch, Anis Abuseiris, Luc V. de Zeeuw, and Rob de Graaf (Erasmus MC Rotterdam, The Netherlands), for their help in creating GRIMP, and BigGRID, MediGRID, and Services@MediGRID/D-Grid, (funded by the German Bundesministerium für Forschung und Technology; grants 01 AK 803 A-H, 01 IG 07015 G) for access to their grid computing resources. RUSH:

The Rush Memory and Aging Project is supported by NIA grants R01AG15819, R01AG17917, and K08AG34290, and the Translational Genomics Research Institute.The Rush Religious Orders Study is supported by NIA grants P30AG10161, R01AG15819, R01AG30146, and K08AG34290 and the Translational Genomics Research Institute. We thank the study participants and the staff of the Rush Alzheimer's Disease Center and Rush Alzheimer's Disease Center. SardiNIA: This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. Funding was also provided through contract NO1-AG-1-2109 from the NIA-NIH. SHIP: SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG. This work was also funded by the German Research Foundation (DFG: GR 1912/5-1); Federal Ministry of Education and Research Germany; the Humboldt Foundation, and the German Research Foundation.

REFERENCES

1. Farmer A, Harris T, Redman K, Sadler S, Mahmood A, McGuffin P. Cardiff depression study. A sib-pair study of life events and familiality in major depression. *Br J Psychiatry.* Feb 2000;176:150-155.
2. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry.* Oct 2000;157(10):1552-1562.
3. Breen G, Webb BT, Butler AW, et al. A genome-wide significant linkage for severe depression on chromosome 3: the depression network study. *Am J Psychiatry.* Aug 2011;168(8):840-847.
4. Pergadia ML, Glowinski AL, Wray NR, et al. A 3p26-3p25 genetic linkage finding for DSM-IV major depression in heavy smoking families. *Am J Psychiatry.* Aug 2011;168(8):848-852.
5. Lopez Leon S, Croes EA, Sayed-Tabatabaei FA, Claes S, Van Broeckhoven C, van Duijn CM. The dopamine D4 receptor gene 48-base-pair-repeat polymorphism and mood disorders: a meta-analysis. *Biol Psychiatry.* May 1 2005;57(9):999-1003.
6. Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, et al. Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry.* Aug 2008;13(8):772-785.
7. Sullivan PF, de Geus EJ, Willemsen G, et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry.* Apr 2009;14(4):359-375.
8. Muglia P, Tozzi F, Galwey NW, et al. Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry.* Jun 2010;15(6):589-601.
9. Shyn SI, Shi J, Kraft JB, et al. Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry.* Feb 2011;16(2):202-215.
10. Shi J, Potash JB, Knowles JA, et al. Genome-wide association study of recurrent early-onset major depressive disorder. *Mol Psychiatry.* Feb 2 2011;16(2):193-201.
11. Lewis CM, Ng MY, Butler AW, et al. Genome-wide association study of major recurrent depression in the U.K. population. *Am J Psychiatry.* Aug 2010;167(8):949-957.
12. Rietschel M, Mattheisen M, Frank J, et al. Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biol Psychiatry.* Sep 15 2010;68(6):578-585.
13. Wray NR, Pergadia ML, Blackwood DH, et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry.* Nov 2 2010.
14. Huang J, Perlis RH, Lee PH, et al. Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am J Psychiatry.* Oct 2010;167(10):1254-1263.
15. Kohli MA, Lucae S, Saemann PG, et al. The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron.* Apr 28 2011;70(2):252-265.
16. Bosker FJ, Hartman CA, Nolte IM, et al. Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry.* May 2011;16(5):516-532.
17. Ayuso-Mateos JL, Nuevo R, Verdes E, Naidoo N, Chatterji S. From depressive symptoms to depressive disorders: the relevance of thresholds. *Br J Psychiatry.* May 2010;196(5):365-371.
18. Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS. A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry.* May 2006;163(5):857-864.
19. Kendler KS, Gardner CO, Jr. Boundaries of major depression: an evaluation of DSM-IV criteria. *Am J Psychiatry.* Feb 1998;155(2):172-177.
20. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. *Mol Psychiatry.* May 22 2012.

21. Terracciano A, Tanaka T, Sutin AR, et al. Genome-wide association scan of trait depression. *Biol Psychiatry.* Nov 1 2010;68(9):811-817.
22. Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. *Psychol Med.* Jan 1997;27(1):231-235.
23. Harlow SD, Goldberg EL, Comstock GW. A longitudinal study of risk factors for depressive symptomatology in elderly widowed and married women. *Am J Epidemiol.* Sep 1 1991;134(5):526-538.
24. Kuchibhatla MN, Fillenbaum GG, Hybels CF, Blazer DG. Trajectory classes of depressive symptoms in a community sample of older adults. *Acta Psychiatr Scand.* Nov 25 2011.
25. Gump BB, Matthews KA, Eberly LE, Chang YF. Depressive symptoms and mortality in men: results from the Multiple Risk Factor Intervention Trial. *Stroke.* Jan 2005;36(1):98-102.
26. Ariyo AA, Haan M, Tangen CM, et al. Depressive symptoms and risks of coronary heart disease and mortality in elderly Americans. Cardiovascular Health Study Collaborative Research Group. *Circulation.* Oct 10 2000;102(15):1773-1779.
27. Carmelli D, Swan GE, Kelly-Hayes M, Wolf PA, Reed T, Miller B. Longitudinal changes in the contribution of genetic and environmental influences to symptoms of depression in older male twins. *Psychol Aging.* Sep 2000;15(3):505-510.
28. Jansson M, Gatz M, Berg S, et al. Gender differences in heritability of depressive symptoms in the elderly. *Psychol Med.* Apr 2004;34(3):471-479.
29. Choy WC, Lopez-Leon S, Aulchenko YS, et al. Role of shared genetic and environmental factors in symptoms of depression and body composition. *Psychiatr Genet.* Feb 2009;19(1):32-38.
30. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet.* Feb 2009;2(1):73-80.
31. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol.* Apr 1989;129(4):687-702.
32. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* Feb 1991;1(3):263-276.
33. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* Dec 1975;4(4):518-525.
34. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health.* Mar 1951;41(3):279-281.
35. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. *Eur J Epidemiol.* Aug 2011;26(8):657-686.
36. Shock NW, Greulich, R.C., Costa, P.T., Andres, R., Lakatta, E.G., Arenberg, D., Tobin, J.D. *Normal Human Aging: The Baltimore Study of Aging.* Washington, DC: US Government Printing Office; 1984.
37. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet.* Jul 2004;12(7):527-534.
38. Ferrucci L, Bandinelli S, Benvenuti E, et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc.* Dec 2000;48(12):1618-1625.
39. Barker DJ, Osmond C, Forsen TJ, Kajantie E, Eriksson JG. Trajectories of growth among children who have coronary events as adults. *N Engl J Med.* Oct 27 2005;353(17):1802-1809.
40. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol.* Nov 1 2002;156(9):871-881.

41. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer.* May 2005;5(5):388-396.
42. Bennett DA, Schneider JA, Buchman AS, Mendes de Leon C, Bienias JL, Wilson RS. The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology.* 2005;25(4):163-175.
43. Bennett DA, Schneider JA, Arvanitakis Z, et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology.* Jun 27 2006;66(12):1837-1844.
44. Pilia G, Chen WM, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* Aug 25 2006;2(8):e132.
45. Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. *J Aging Health.* May 1993;5(2):179-193.
46. Levy D, DeStefano AL, Larson MG, et al. Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertension.* Oct 2000;36(4):477-483.
47. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* Sep 1 2010;26(17):2190-2191.
48. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* Sep 2007;81(3):559-575.
49. Harris TB, Launer LJ, Eiriksdottir G, et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol.* May 1 2007;165(9):1076-1087.
50. Holle R, Happich M, Lowel H, Wichmann HE, Group MKS. KORA--a research platform for population based health research. *Gesundheitswesen.* Aug 2005;67 Suppl 1:S19-25.
51. Volzke H, Alte D, Schmidt CO, et al. Cohort Profile: The Study of Health in Pomerania. *Int J Epidemiol.* Feb 18 2010.
52. John U, Greiner B, Hensel E, et al. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz Praventivmed.* 2001;46(3):186-194.
53. Milette K, Hudson M, Baron M, Thombs BD, Canadian Scleroderma Research G. Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: internal consistency reliability, convergent validity and clinical correlates. *Rheumatology (Oxford).* Apr 2010;49(4):789-796.
54. Shean G, Baldwin G. Sensitivity and specificity of depression questionnaires in a college-age sample. *J Genet Psychol.* Sep 2008;169(3):281-288.
55. Kopp MS, Falger PR, Appels A, Szedmak S. Depressive symptomatology and vital exhaustion are differentially related to behavioral risk factors for coronary artery disease. *Psychosom Med.* Nov-Dec 1998;60(6):752-758.
56. Wojciechowski FL, Strik JJ, Falger P, Lousberg R, Honig A. The relationship between depressive and vital exhaustion symptomatology post-myocardial infarction. *Acta Psychiatr Scand.* Nov 2000;102(5):359-365.
57. Wattanakit K, Williams JE, Schreiner PJ, Hirsch AT, Folsom AR. Association of anger proneness, depression and low social support with peripheral arterial disease: the Atherosclerosis Risk in Communities Study. *Vasc Med.* Aug 2005;10(3):199-206.
58. Thomas PD, Kejariwal A, Campbell MJ, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.* Jan 1 2003;31(1):334-341.59. Hek K, Mulder CL, Luijendijk HJ, et al. The PCLO gene and depressive disorders: replication in a population-based study. *Hum Mol Genet.* Feb 15 2010;19(4):731-734.
60. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* Sep 6 2003;327(7414):557-560.

61. Breslau N. Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. *Psychiatry Res.* Jul 1985;15(3):219-229.
62. Hettema JM. What is the genetic relationship between anxiety and depression? *Am J Med Genet C Semin Med Genet.* May 15 2008;148C(2):140-146.
63. Gillespie NA, Kirk KM, Evans DM, Heath AC, Hickie IB, Martin NG. Do the genetic or environmental determinants of anxiety and depression change with age? A longitudinal study of Australian twins. *Twin Res.* Feb 2004;7(1):39-53.
64. Demirkan A, Penninx BW, Hek K, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry.* Jul 2011;16(7):773-783.
65. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol.* Jan 2006;62(1):123-146.
66. Agrell B, Dehlin O. Comparison of six depression rating scales in geriatric stroke patients. *Stroke.* Sep 1989;20(9):1190-1194.
67. Bernert S, Matschinger H, Alonso J, et al. Is it always the same? Variability of depressive symptoms across six European countries. *Psychiatry Res.* Jul 30 2009;168(2):137-144.
68. McClellan J, King MC. Genomic analysis of mental illness: a changing landscape. *Jama.* Jun 23 2010;303(24):2523-2524.
69. Zhang J, Chen Y, Zhang K, et al. A cis-phase interaction study of genetic variants within the MAOA gene in major depressive disorder. *Biol Psychiatry.* Nov 1 2010;68(9):795-800.
70. Yang C, Xu Y, Sun N, et al. The combined effects of the BDNF and GSK3B genes modulate the relationship between negative life events and major depressive disorder. *Brain Res.* Oct 8 2010;1355:1-6.
71. Zammit S, Owen MJ, Lewis G. Misconceptions about gene-environment interactions in psychiatry. *Evid Based Ment Health.* Aug 2010;13(3):65-68.
72. Glessner JT, Wang K, Sleiman PM, et al. Duplication of the SLIT3 locus on 5q35.1 predisposes to major depressive disorder. *PLoS One.* 2010;5(12):e15463.
73. Rucker JJ, Breen G, Pinto D, et al. Genome-wide association analysis of copy number variation in recurrent depressive disorder. *Mol Psychiatry.* Nov 1 2011.

Chapter 3.3

Somatic complaints of depression and the melatonin receptor MTNR1A gene: results from the CHARGE Consortium.

Ayşe Demirkan, Jari Lahti, Nese Direk, Alexander Viktorin, Kathryn L Lunetta, Antonio Terracciano, Toshiko Tanaka, Michael A Nalls, Myriam Fornage, Jurgen Wellmann, Hanna M Ollila, Lei Yu, Aaron Isaacs, Aarno Palotie, Karin Hek, Wei V Zhuang, Alan Zonderman, Angelina Sutin, Osorio Meirelles, Antonella Mulas, Albert Hofman, Andre Uitterlinden, Fernando Rivadeneira, Markus Perola, Veikko Salomaa, Kristine Yaffe, Annemarie I Luik, Yongmei Liu, Jingzhong Ding, Paul Lichtenstein, Mikael Landén, Elisabeth Widen, Johan G Eriksson, Patrik KE Magnusson, Luigi Ferrucci, Thomas H Mosley, Francesco Cucca, Ben A Oostra, David A Bennett, Tiina Paunio, Klaus Berger, Tamara B Harris, Nancy L Pedersen, Joanne M Murabito, Henning Tiemeier, Cornelia M van Duijn, Katri Raikkonen.

(under preparation)

ABSTRACT

Diagnosis of major depressive disorder (MDD) is based on heterogeneous symptoms which can be classified into different domains such as somatic symptoms, negative affect, lack of positive affect, and interpersonal problems. Despite the moderate heritability of MDD, genome-wide association studies (GWAS) for MDD, as well as related continuous outcomes, conducted thus far have shown consistent results. Attempts to elucidate the genetic basis of MDD may be hindered by difficulties in diagnosis. The Centre for Epidemiological Studies Depression Scale (CES-D) provides a widely used tool for measuring depressive symptoms clustered in four different domains. We performed meta-analyses of GWAS of the CES-D symptom clusters. We recruited 10 cohorts with the 20 item CES-D for the discovery stage and 2 cohorts with the 10 item CES-D for the replication stage (22 200 and 9900 persons, respectively). One SNP, located in the brain-expressed melatonin receptor (*MTNR1A*) gene, was associated with the somatic domain of depression symptoms, with genome-wide significance ($P\text{-value}_{stage1\&2} = 4.56 \times 10^{-8}$). The SNP was analyzed in an additional 8 cohorts for the third stage analysis (11 480 persons). However, the association was not consistent among the third stage samples ($P\text{-value}_{stage1,2\&3} = 1.58 \times 10^{-5}$) with an evidence of heterogeneity ($P\text{-value}_{het} = 0.03$) Despite the effort to harmonize the phenotypes across cohorts and participants, our study is still underpowered to detect consistent association for depression, even through symptom classification.

INTRODUCTION

Genetic factors play an important role in the susceptibility to depression. A meta-analysis of twin studies on Major Depressive Disorder (MDD) estimated a heritability between 31 and 42%[1]. Despite this, the success of genome-wide association studies (GWAS) aiming to find genes underlying vulnerability for depression has been limited; the most promising findings to date are poorly replicated and explain only a small portion of this heritability[2-4]. This may be explained by the polygenic architecture of the trait[5] as well as difficulties in diagnosis. A validated biomarker for depression does not exist and the diagnosis is based solely on symptoms. These include psychological symptoms, such as depressed or sad moods, loss of interest in activities, feelings of worthlessness or inappropriate guilt, recurrent thoughts of death, and poor concentration, and somatic symptoms, such as changes in appetite, sleep patterns, fatigue or weight loss[6,7]. The diverse clusters of complaints, which result in variations in measurements between people and populations, lead to problems for gene discovery. A focus on outcomes based on depressive symptoms and endophenotypes has been shown to increase power in association studies[8,9]. Additionally, for these outcomes, the genetic architecture is complex and involves the additive effects of multiple common variants[5].

Depressive symptoms can be measured by questionnaires, such as the Centre for Epidemiological Studies Depression Scale (CES-D), which is known to be heritable ($H^2 = 0.22$, P-value = 0.001)[10]. The original version of the CES-D scale consists of 20 items with scores ranging from 0 to 60, and measures symptoms clustered in somatic, positive, negative and interpersonal domains, which is usually summed up to one single score[11]. The subscales of the CES-D can also be analyzed separately in order to focus on the specific symptom clusters. We conducted a meta-analysis of GWAS of specific symptom clusters measured by CES-D across the CHARGE consortium cohorts, including a total of ~32 100 persons. The discovery set consisted of 10 cohorts with the 20 item CES-D and the replication set consisted of 2 cohorts with the 10 item CES-D (22 200 and 9900 persons respectively)

MATERIALS AND METHODS

Table 1 summarizes the characteristics of the discovery and replication cohorts. The discovery sample consisted of the CHARGE cohorts with eligible 20 question CES-D information (CES-D 20). These cohorts were the Baltimore Longitudinal Study of Aging (BLSA)[12], the Dortmund Health Study (DHS)[13,14], the Erasmus Rucphen Family Study (ERF)[15,16], NHLBI's Framingham Heart Study (FHS)[17-19], the Helsinki Birth Cohort Study (HBCS)[20], European ancestry participants from the Health, Aging and Body Composition study (HEALTH ABC-Eur/Am)[21], the Rotterdam Study I-II-III (RS I-II-III)[22] and SardiNIA[23]. Two studies in which the symptoms of depression were measured with the short version of CES-D (CES-D 10), the Atherosclerosis Risk In Communities study (ARIC) and the Swedish Twin Registry (STR)[24], were utilized for replication. FINRISK[25], the

TABLE 1 Study sample characteristics of discovery and replication cohorts

	Somatic items	Negative items	Positive items	Interpersonal items	Age (years)	N	Women %
Discovery sample [20 item]	**Mean(sd)**	**Mean(sd)**	**Mean(sd)**	**Mean(sd)**	**Mean(sd)**		
BLSA	2.92 (2.81)	1.42(2.41)	10.31(2.45)	0.22(0.65)	71.6(13.8)	827	45.1
DHS	2.92(3.18)	1.54(2.91)	7.03(3.25)	0.22(0.72)	52.4(13.7)	991	52.6
ERF	3.78(3.76)	2.07(3.36)	8.43(3.40)	0.40(0.89)	55.0(10.1)	1107	55.2
FHS	1.05(0.751)	0.61(0.75)	0.596(0.739)	0.149(0.35)	56.1(10.5)	6636	51.8
HBCS	3.79 (3.31)	2.00 (3.05)	9.22 (2.40)	0.37(0.79)	63.4(2.9)	1360	59.4
HEALTH ABC (Eur)	1.68 (2.13)	0.93(1.85)	correct	0.13(0.49)	73.8(2.8)	1520	46.4
RS I	1.52 (2.61)	1.24(2.65)	10.36(2.59)	0.09(0.42)	72.7(7.2)	3709	58.1
RS II	1.98 (2.78)	1.34(2.66)	10.19(2.60)	0.15(0.53)	64.8(8.0)	1995	53.3
RS III	2.66 (3.32)	1.17(2.58)	10.37(2.41)	0.18(0.58)	56.0(5.7)	1917	55.1
SardiNIA	3.27(2.91)	2.51(3.08)	4.81(2.53)	0.48(0.85)	58.0(11.4)	2608	58.1
Replication sample [10 item]							
ARIC	2.31(2.15)	1.19(1.72)	5.62(0.95)	0.21(0.65)	72.7(5.5)	384	42.1
STR	1.10(1.60)	0.79(1.43)	1.22(1.27)	0.21(0.58)	57.7 (8.9)	9474	52.7
3rd stage sample							
FINRISK [10 item]	1.74(1.63)	-	-	-	53.1(13.4)	605	49.7
HRS [10 item]	1.38(1.50)	-	-	-	69.3(5.5)	3753	58.1
INCHIANTI [20 item]	3.23(3.18)	-	-	-	66.0(15.0)	1019	47.0
NHS (BC) [10 item]	1.91(1.56)	-	-	-	72.5 (6.4)	1537	100
NHS (CHD) [10 item]	1.97(1.58)	-	-	-	73.7 (6.4)	733	100
NHS (T2DM) [10 item]	1.94(1.59)	-	-	-	71.3 (6.7)	2397	100
RUSH MAP [10 item]	0.50(0.80)	-	-	-	80.8(6.5)	721	71.7
RUSH ROS [10 item]	0.57(0.82)	-	-	-	75.5(7.2)	715	65.9

Mean; mean value of each scale, sd, Standard deviation of the mean, N; number of subjects included. BLSA;Baltimore Longitudinal Study of Aging, DHS; Dortmunt Health Study, ERF; Erasmus Rucphen Family Study, FHS;Framingham Heart Study, HBCS; Helsinki Birth Cohort Study, HEATLH ABC(Eur); Health, Aging and Body Composition study (of European ancestors), RS I-II-III; Rotterdam study first, second and third waves, SardiNIA; SardiNIA study, ARIC; Atherosclerosis Risk in Communities study, STR; Sweedish Twin Registrym, RUSH MAP, RUSH Memory and Aging Project; RUSH ROS, RUSH Religious Orders Study, FINRISK; National FINRISK Study of Finland.

Health and Retirement Study (HRS), INCHIANTI, the Nurses' Health Study (NHS)[26] (breast cancer, type2 diabetes and cardiovascular disease sub-samples), and the Memory and Aging Project and Religious Order Study of Rush Alzheimer's Disease Centre (RUSH-ROS and RUSH-MAP)[27, 28] were used as 3rd stage replication samples (See Supplementary Text 1 for the study descriptions and Supplementary text 2 for the items of CES-D scale).

GWA analyses were performed individually by the study centres, according to the same analysis plan; each study excluded dementia cases (MMSE score < 22), and anti-depressive medication users, since the effect of anti-depressive medication on the scales was not consistent across the studies. There was no restriction on age. Each study centre computed the subscales

of the CES-D questionnaire that resulted in four separate scores for each individual, measuring different clusters of complaints. Each study implemented linear regression models, adjusted for age, age-square and sex, under the assumption of an additive genetic model, regressing each subscale on allele dosage and reported the summary statistics. The genotyping and imputation methods for each study are given in Supplementary Table 1. Additional study site specific adjustments included linear mixed effect models to account for familial correlations in FHS and ERF, adjustment for disease status in NHS, adjustment for the top 3 eigen vectors in RUSH-MAP, RUSH-ROS, STR, NHS (T2DM) and NHS (CHD) and adjustment for the top 4 eigen vectors in NHS (BC) studies.

Prior to meta-analysis, all SNP IDs were mapped to dbSNP Build 129. Possible measurement and scoring differences across different study centers were checked through extracting median standard error from the GWAS summary statistics of each study centre and plotting it towards the square-root of the sample size. Allele frequencies for all SNPs were compared to HapMap frequencies. Stratified Q-Q plots were generated for minor allele frequency and imputation quality strata to assess potential sources of inflation. Meta-analyses were performed using the fixed effects inverse variance method as implemented in the METAL software package[29]. The over-all meta analysis in which CES-D 20 item and CES-D 10 item cohorts were combined, were performed with sample size weighted method as implemented in METAL. We also performed sex stratified analyses. Meta-analysis of interpersonal domain, which consists of only two questions, was not performed due to differences in the median standard errors across the cohorts. SNPs with a MAF less than 2.5% or an observed to expected variance ratio (imputation quality) less than 0.30 were excluded on a per-study basis. SNPs for which the total sample size was lower than 5000 were removed from further analysis. Genomic control correction was applied to each study's results[30].

RESULTS

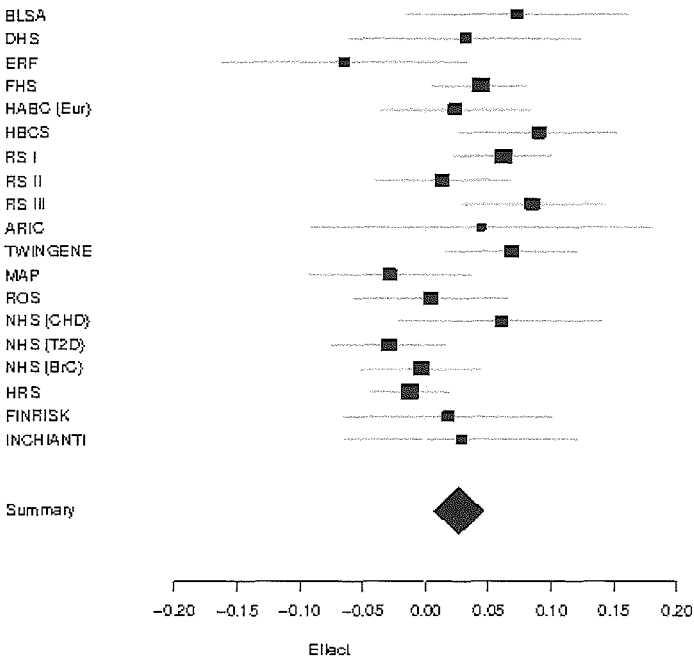
The inflation factors for the discovery GWAS of the three scales varied between 1.026 and 0.984. We did not observe any genome-wide significant findings for any of the scales in the discovery set. Table 2 shows the top findings (*P*-value < 1.00×10^{-6}) from each meta-analysis and their *P*-values in the replication set for the combined and sex-stratified analysis. Among the top findings, one SNP (rs713224) showed significant association in the same direction in the replication set and resulted in a genome-wide significant *P*-value (*P*-value $_{stage1\&2} = 4.56 \times 10^{-8}$) when the stage 1 and stage 2 samples were analysed together. The analysis of this particular SNP was further extended to a third stage, which included 11 480 persons from 8 study samples, as shown in Table 1. The analysis included 3 case-control studies of people with somatic diseases: cardiovascular, diabetes and cancer. Figure 1 shows the forest plots for rs713224 across all 19 studies. The overall analysis yielded a non-significant result (*P*-value $_{stage1,2\&3} = 1.58 \times 10^{-5}$). Testing for heterogeneity showed evidence for outliers (*P*-value $_{het} = 0.03$) in the combined

TABLE 2 Association of the top SNPs in the replication sample

	Discovery							Replication								
	Domain	A1	A2	Effect	SE	P-value	Direction	Effect	StdErr	P-value	Direction	CHR	Position	MAF	Gene region	Overall P-value
rs713224	S/C	c	g	-0.05	0.01	4.99×10^{-7}	--+----?--	-0.07	0.03	0.009	--	4	187412367	0.24	MTNRA1	4.60×10^{-8}
rs2375800	S/C	a	g	0.05	0.01	6.80×10^{-7}	++++++?++	0.07	0.03	0.009	++	4	187414185	0.24	MTNRA1	6.15×10^{-8}
rs16909586	S/C	a	c	-0.10	0.02	7.20×10^{-7}	---------	0.04	0.05	0.456	++	11	22536809	0.03	SLC17A6	
rs13149231	P/F	t	c	-0.05	0.01	1.21×10^{-7}	+--------	0.00	0.02	0.982	-+	4	38716483	0.19	KLF3	
rs6845639	P/F	a	g	0.05	0.01	1.82×10^{-7}	-+++++++++	0.00	0.02	0.967	+-	4	38716721	0.19	KLF3	
rs1000096	P/F	t	c	-0.05	0.01	9.39×10^{-7}	+--------	0.00	0.02	0.937	-+	4	38693335	0.32	KLF3	
rs3209570	P/F	a	g	-0.05	0.01	9.51×10^{-7}	+--------	0.00	0.02	0.896	-+	4	38699657	0.26	KLF3	
rs4730076	S/M	a	c	0.08	0.01	3.25×10^{-7}	+++++-++++	0.02	0.04	0.645	++	7	104951685	0.12	SRPK2	

S/C, somatic items GWAS in the combined sample; P/F, positive items GWAS in the women only sample; S/M somatic items GWAS in the men only sample. Effect, pooled effect estimate from the meta-analysis; SE, standart error of the pooled effect estimate; MAF, minor allele frequency of the SNP. Overall *P*-value, *P*-value from the stage 1 and 2 combined analysis

FIGURE 1. Forest plots of all study sample for SNP rs713224.



analysis, compared the discovery phase (P-value $_{het} = 0.17$). This was attributed to the cohorts that included case control samples (NHS) and RUSH.

Lastly, we meta-analysed the GWAS of the stage 1 and stage 2 samples together (N = 32100) to increase power. Q-Q plots and Manhattan plots of this analysis are presented in the Supplementary Figures 1 and 2. This analysis did not yield any significant results, with the exception of rs713224. (P-value $_{stage1\&2} = 4.56 \times 10^{-8}$)

To see if any of our top hits provided evidence for association to MDD disease status, we performed a look up using the open access data from the Psychiatric Genetics Consortium (PGC) MDD mega-GWAS[31]. The SNPs with P-value $< 1.00 \times 10^{-4}$ and their status in the PGC-MDD GWAS are shown in the Supplementary Table 2. We did not find any overlap between our findings on symptom dimensions and results from the MDD GWAS.

DISCUSSION

We conducted a GWAS on dimensions of depressive symptoms in which we combined the results of 12 population-based studies with more than 32 000 individuals to find common variants that increase the vulnerability to a particular symptom cluster (somatic symptoms, positive

effect and negative effect). Of the 8 SNPs (*P*-value < 1.00×10^{-6}) we attempted to replicate, we found evidence for 2 neighboring SNPs in the *MTNR1A* gene. One of the SNPs, rs713224, reached genome-wide significance in the meta-analysis of the combined stage 1 and stage 2 samples ($P\text{-value}_{stage1\&2} = 4.56 \times 10^{-8}$). Rs713224 was further analyzed in 8 separate samples, which resulted in an increase of the overall *P*-value for this SNP ($P\text{-value}_{stage1,2\&3} = 1.58 \times 10^{-5}$).

Our analysis yielded only one genome-wide significant SNP, for the combined analysis of the somatic complaints domain. Somatic symptoms are very common in depression and vital sense, fatigue, and changes in sleep and appetite are included as diagnostic criteria for MDD. These symptoms might be more easily measured than mood alterations because they are straightforward to recognize by patients[32]. This SNP was located in the *MTNR1A* gene, which encodes a melatonin receptor expressed in brain. The melatonin receptor pathway is known to be involved in depression[33-38] and its relationship with somatic complaints, and vitality in general, makes it a biologically plausible gene. However, lack of replication in the 3rd stage sample both in the separate study results and the whole meta-analysis may imply that our finding for this SNP is maybe a false positive. There were important differences between stage 1&2 and the stage 3 samples. Individuals from population based studies of European descent also minimized measurement error caused by cultural differences in response to the CES-D[39]. This was not true for the stage 3 sample. We observed increased heterogeneity when we included the 3rd stage samples in the analysis. Another problem with the stage 3 sample is that it includes patients of severe disease. For instance, the NHS study included only women and cardiovascular disease, breast cancer and type 2 diabetes cases, which may confound the association with somatic complaints even though the analysis were adjusted for case-control status. Also RUSH study reported very low mean values for the somatic items, included more women and older persons (Table 1). For these reasons, we additionally performed sensitivity analysis, excluding these studies separately and together. Excluding the case-control samples improved the *P*-value ($P\text{-value}_{stage1,2\&3} = 0.03$) and yielding a non-significant *P*-value for heterogeneity test ($P\text{-value}_{het} = 0.07$), excluding both studies yielded a $P\text{-value}_{stage1,2\&3} = 4.96 \times 10^{-7}$ and increased the *P*-value from the heterogeneity test ($P\text{-value}_{het} = 0.08$). rs713224 is a G to C transition which may cause strand errors for this SNP. It is of note that PGC-MDD GWAS also showed association with rs4478239, located within 800 KB of MTNR1A with recurrent depression (*P*-value $= 4.7 \times 10^{-7}$) in a study including 6743 cases and 9519 controls[31]. However, the proxy for our top SNP in that region (rs2375800) was not associated with MDD status in the PGC-MDD GWAS results (Supplementary Table 2).

The difficulty in finding GWAS signals for major depression has been a common experience both for depressive symptoms and MDD. A previous study of depressive symptoms of the CHARGE Consortium (*Hek, K* et al, unpublished) on a partially overlapping sample suggested a region on 5q21 in a combined analysis of more than 50 000 persons. A meta-analysis of eight GWAS of MDD status, of which the largest comprised ~6000 MDD cases and ~7000 controls,

yielded only one genome-wide significant finding[40], while the recent PGC mega-analysis pointed out one region on 3p21.1 that reached genome–wide significance.

To conclude, our efforts in a large collaboration utilizing phenotypes defined by symptom clustering yielded no genome-wide significant hit in the discovery sample. The best result, a SNP associated with somatic complaints, reached genome-wide significance in the combined sample and suggested the involvement of *MTNR1A* in the melatonin signaling pathway, but was not further replicated. Our results suggest that GWAS for depression in large population based samples remains underpowered due to phenotypic and genetic heterogeneity.

ACKNOWLEDGEMENTS

ARIC: The research is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, and grants R01-HL087641, R01-HL093029 and R01-HL70825; National Human Genome Research Institute contract U01-HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. BLSA: BLSA research was supported by the Intramural Research Program of the NIH, National Institute on Aging. Blood collection in the Dortmund Health Study was funded by the Institute of Epidemiology and Social Medicine University of Muenster and genotyping with the HumanOmni chip by the German Ministry of Research and Education (BMBF, 01ER0816).The collection of socio-demographic and clinical data in the Dortmund Health Study was supported by the German Migraine & Headache Society (DMKG) and by unrestricted grants of equal share from Almirall, Astra Zeneca, Berlin Chemie, Boehringer, Boots Health Care, Glaxo-Smith-Kline, Janssen Cilag, McNeil Pharma, MSD Sharp & Dohme and Pfizer to the University of Muenster. ERF: The genotyping for the ERF study was supported by EUROSPAN (European Special Populations Research Network) and the European Commission FP6 STRP grant (018947; LSHG-CT-2006-01947). The ERF study was further supported by grants from the Netherlands Organisation for Scientific Research, Erasmus MC, the Centre for Medical Systems Biology (CMSB) and the Netherlands Brain Foundation (HersenStichting Nederland). We are grateful to all participating individuals and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection. FHS: The phenotype-genotype association analyses IN FHS was supported by R01-AG29451. This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association

Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. FINRISK :The Finrisk Study has been funded by Sigrid Juselius foundation, Jalmari and Rauha Ahokas foundation and Biomedicum Helsinki foundation. HABC: This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. HBCS: We thank all study participants as well as everybody involved in the Helsinki Birth Cohort Study. Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, European Science Foundation (EUROSTRESS), Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation, Juho Vainio Foundation, and Wellcome Trust (grant number WT089062). HRS:HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded as a separate award from the National Institute on Aging (RC2 AG036495). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were perfomed by the Genetics Coordinating Center at the University of Washington. INCHIANTI: The Invechhiare in Chianti (InCHIANTI). The InCHIANTI)Study was supported as a "targeted project" (ICS 110.1RS97.71) by the Italian Ministry of Health, by the U.S. National Institute on Aging (Contracts N01]AG]916413, N01]AG] 821336, 263 MD 9164 13, and 263 MD 821336) and in part by the Intramural Research Program, National Institute on Aging, National Institutes of Health, USA. NHS: The Nurses' Health Studies are supported by NIH grants CA 65725, CA87969 (National Cancer Institute), CA49449, CA67262, CA50385 and 5UO1CA098233. RS: The generation and management of the genotype data for the RS are supported by the Netherlands Organization of Scientific Research Investments (#175.010.2005.011, 911-03-012). 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 and RIDE2); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); the Municipality of Rotterdam; and the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research project number 050-060-810. F.P. Velders is supported by the Sophia Foundation for Scientific Research (SKZ Foundation) (grant no. 491) and Netherlands Organization for the Health Research and Development (ZonMw) (grant no. 10.000.1003). RUSH-ROS: The ROS study was supported by

National Institute on Aging grants P30AG10161 and the Illinois Department of Public Health. RUSH-MAP: The MAP study was supported by National Institute on Aging grants R01AG17917, R01AG15819, and the Illinois Department of Public Health. SARDINIA: SARDINIA research was supported by the Intramural Research Program of the NIH, National Institute on Aging. Funding was also provided through contract NO1-AG-1-2109 from the NIA-NIH. TWINGENE: STR study was supported by The Ministry for Higher Education, the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, the Swedish Foundation for Strategic Research (SSF). Heart and Lung foundation no. 20070481.

REFERENCES

1. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry.* Oct 2000;157(10):1552-1562.
2. Psychiatric GCCC, Cichon S, Craddock N, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry.* May 2009;166(5):540-556.
3. Muglia P, Tozzi F, Galwey NW, et al. Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry.* Dec 23 2008.
4. Sullivan PF, de Geus EJ, Willemsen G, et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry.* Apr 2009;14(4):359-375.
5. Demirkan A, Penninx BW, Hek K, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry.* Jul;16(7):773-783.
6. Association AP. *Diagnostic and statistical manual of mental disorders 4th edition.* 4 ed: APA; 1994
7. NICE. *NICE Draft Guidelines. Depression: the Management of Depression in Primary and Secondary Care.* London: National Institute for Clinical Excellence; December 2003.
8. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. *Mol Psychiatry.* May 22.
9. 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.
10. Lopez-Leon S, Aulchenko YS, Tiemeier H, Oostra BA, van Duijn CM, Janssens AC. Shared genetic factors in the co-occurrence of symptoms of depression and cardiovascular risk factors. *J Affect Disord.* May;122(3):247-252.
11. Radloff. The CES-D scale: a self report depression scale for research in theb general population. *Appl Pshycol Measurement.* 1977;3:385-401.
12. Sutin AR, Zonderman AB. Depressive symptoms are associated with weight gain among women. *Psychol Med.* Apr 5:1-10.
13. Pfaffenrath V, Fendrich K, Vennemann M, et al. Regional variations in the prevalence of migraine and tension-type headache applying the new IHS criteria: the German DMKG Headache Study. *Cephalalgia.* Jan 2009;29(1):48-57.
14. Vennemann MM, Hummel T, Berger K. The association between smoking and smell and taste impairment in the general population. *J Neurol.* Aug 2008;255(8):1121-1126.
15. Lopez-Leon S, Chi Choy W, Aulchenko YS, et al. Genetic Factors Influence the Clustering of Depression among Individuals with Lower Socioeconomic Status. *PLoS ONE.* 2009;4(3):e5069.
16. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet.* 2004;12(7):527-534.
17. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* Dec 1975;4(4):518-525.
18. Splansky GL, Corey D, Yang Q, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.* Jun 1 2007;165(11):1328-1335.
19. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health.* Mar 1951;41(3):279-281.
20. Barker DJ, Osmond C, Forsen TJ, Kajantie E, Eriksson JG. Trajectories of growth among children who have coronary events as adults. *N Engl J Med.* Oct 27 2005;353(17):1802-1809.
21. Cesari M, Penninx BW, Newman AB, et al. Inflammatory markers and cardiovascular disease (The Health, Aging and Body Composition [Health ABC] Study). *Am J Cardiol.* Sep 1 2003;92(5):522-528.

22. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. *Eur J Epidemiol.* Aug;26(8):657-686.
23. Pilia G, Chen WM, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* Aug 25 2006;2(8):e132.
24. Rahman I, Bennet AM, Pedersen NL, de Faire U, Svensson P, Magnusson PK. Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet.* Jun 2009;12(3):286-294.
25. Vartiainen E, Laatikainen T, Peltonen M, et al. Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol.* Apr;39(2):504-518.
26. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer.* May 2005;5(5):388-396.
27. Bennett DA. Postmortem indices linking risk factors to cognition: results from the Religious Order Study and the Memory and Aging Project. *Alzheimer Dis Assoc Disord.* Jul-Sep 2006;20(3 Suppl 2):S63-68.
28. Bennett DA, Schneider JA, Buchman AS, Mendes de Leon C, Bienias JL, Wilson RS. The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology.* 2005;25(4):163-175.
29. Willer C, Li Y, Abecasis G. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* Sep 1 2010;26(17):2190-2191.
30. Zheng G, Freidlin B, Li Z, Gastwirth JL. Genomic control for association studies under various genetic models. *Biometrics.* Mar 2005;61(1):186-192.
31. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* Apr 3.
32. Kapfhammer HP. Somatic symptoms in depression. *Dialogues Clin Neurosci.* 2006;8(2):227-239.
33. Wetterberg L, Beck-Friis J, Kjellman BF, Ljunggren JG. Circadian rhythms in melatonin and cortisol secretion in depression. *Adv Biochem Psychopharmacol.* 1984;39:197-205.
34. Goldstein JA. Melatonin as depression marker. *Biol Psychiatry.* May 1985;20(5):585.
35. Galecka E, Szemraj J, Florkowski A, et al. Single nucleotide polymorphisms and mRNA expression for melatonin MT(2) receptor in depression. *Psychiatry Res.* Oct 30;189(3):472-474.
36. Carman JS, Post RM, Buswell R, Goodwin FK. Negative effects of melatonin on depression. *Am J Psychiatry.* Oct 1976;133(10):1181-1186.
37. Bourin M, Prica C. Melatonin receptor agonist agomelatine: a new drug for treating unipolar depression. *Curr Pharm Des.* 2009;15(14):1675-1682.
38. Anderson G. The role of melatonin in post-partum psychosis and depression associated with bipolar disorder. *J Perinat Med.* Nov;38(6):585-587.
39. Bernert S, Matschinger H, Alonso J, et al. Is it always the same? Variability of depressive symptoms across six European countries. *Psychiatry Res.* Jul 30 2009;168(2):137-144.
40. Kohli MA, Lucae S, Saemann PG, et al. The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron.* Apr 28 2011;70(2):252-265.

4

Gene discovery studies on circulating lipids

Chapter 4.1

Genome wide linkage scan on plasma phospho-sphingolipids identifies multiple quantitative trait loci.

Ayşe Demirkan, Aaron Isaacs, Irina Zorkoltseva, Najaf Amin, Peter Ugocsai, Gerhard Liebisch, Rutger W W Brouwer, Wilfred F J van IJcken, Gerd Schmitz, Ben A Oostra, Tatiana Axenovich, Cornelia M van Duijn.

(under preparation)

ABSTRACT

Phospho- and sphingolipids are crucial cellular and intracellular compounds. These lipids are required for active transport, enzymatic processes, membrane formation and cell signalling. Disruption of their metabolism leads to diverse neurological, psychiatric, and metabolic consequences, as well as several rare and common diseases. We conducted linkage analyses in the Erasmus Rucphen Family Study (ERF), using mass spectrometry lipid measurements from 820 individuals in 91 families. Twenty-four sphingomyelins (SPM), 9 ceramides (CER), 57 phosphatidylcholines (PC), 20 lysophosphatidylcholines (LPC), 27 phosphatidylethanolamines (PE) and 16 PE-based plasmalogens (PLPE), as well as their proportions in each major class, were analysed by variance component and two-point linkage methods. Heritability estimations were performed separately for each trait, as well as for their proportions among their own class. This effort uncovered 9 quantitative trait loci with LOD > 3.3: 1p22.2 (LOD = 4.22), 1p31.1 (LOD = 4.26), 1p36.32 (LOD = 3.7), 18p11.21 (LOD = 3.77), 17q31 (LOD = 3.67), 19p12 (LOD = 3.42), 6q16.3 (LOD = 3.35), 11q12.3 (LOD = 3.35), 11q21 (LOD = 3.35). Additionally, there were 330 suggestive linkage findings with 1.9 < LOD < 3.3. Heritability estimates ranged from 0% to 55% and were generally higher for within class proportions than for plasma levels. LOD support intervals for several loci included previously known genes, such as *FADS1-2-3* (% PC 36:4) and *LPAR2* (% CER 18:0), but also novel regions containing plausible candidate genes such as *PLCH2* (PE 38:5), *PRKACB* (PLPE 18:1 / 22:6), and *PKN2* (% PLPE 18:1 / 22:6)

INTRODUCTION

Phospho- and sphingolipids are key determinants of cell behaviour and function[1-3]. Phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), lysophosphatidylcholines (LPCs) and PE-based plasmalogens (PLPE) take part in cell survival and inflammation[4-6]. Sphingolipids are believed to play critical roles in cell surface protection, protein and lipid transport, and cellular signalling cascades[7]. In plasma PC, SPM and PE are carried within HDL-C and LDL-C, as well as in platelets[8-11]. In addition to their involvement in the pathology of common diseases such as type 2 diabetes and dementia[8-14], disruption of their metabolism also causes rare monogenic diseases such as Niemann–Pick and Farber[15].

Recent genome wide association studies (GWAS) on the human lipidome identified many loci with strong statistical significance and uncovered common SNPs that explain up to 23% of the variance of particular lipid traits[16-19]. For many circulating phospho- and sphingolipids, however, a substantial proportion of their heritability remains unexplained. Some of this may be due to rare variants that GWAS do not have power to identify. These rare variants can be detected by linkage methods in family studies. In an effort to discover loci that may include rare variants responsible for determining phospho- and sphingolipid levels, we conducted the first linkage study on 147 plasma phospho- and sphingolipid species using a pedigree containing 820 individuals.

MATERIALS AND METHODS

Study population

The ERF study, which is a part of the Genetic Research in Isolated Populations (GRIP) Program, is a family-based study including over 3000 participants descendant from 22 couples living in the Rucphen region in the southwest Netherlands in the 19th century. All descendants of those couples were invited to visit the clinical research centre in the region where they were examined in person and where fasting blood was drawn[20]. The Medical Ethical Committee of the Erasmus Medical Center Rotterdam approved the study and informed consent was obtained from all participants[21].

Lipidomics analysis

Lipid species were quantified by electrospray ionization tandem mass spectrometry (ESIMS/MS) using methods validated and described previously[22, 23]. In brief, samples were analyzed by direct flow injection using a precursor ion scan of m/z 184 specific for phosphocholine containing lipids including PC, SPM[23] and LPC[22]. A neutral loss scan of m/z 141 was used for PE [9] and PLPE and was analyzed according to the principles described by Zemski-Berry[24]. Fragment ions

of m/z 364, 380 and 382 were used for the PE p16:0, p18:1 and p18:0 species, respectively. Quantification was achieved by calibration lines generated by the addition of naturally occurring lipid species to plasma and internal standards belonging to the same lipid class (PC 14:0 / 14:0, PC 22:0 / 22:0, PE 14:0 / 14:0, PE 20:0 / 20:0, LPC 13:0, LPC 19:0). Calibration lines were generated for the following naturally occurring species: PC 34:1, 36:2, 38:4, 40:0 and PC O 16:0 / 20:4; LPC 16:0, 18:1, 18:0; PE 34:1, 36:2, 38:4, 40:6 and PLPE 16:0 / 20:4. Correction of isotopic overlap of lipid species, as well as data analysis, was performed by self programmed Excel macros for all lipid classes according to the principles described previously [23]. Nomenclature of sphingomyelin species is based on the assumption that d18:1 (dihydroxy 18:1 sphingosine) is the main base of plasma SPM species, where the first number refers to the number of carbon atoms in the chain and the second number to the number of double bonds in the chain. The performed analysis does not always allow an exact assignment. In this case, an "O" is added to the subspecies name, e.g., PC O 36:5 and PC O 32:1. This denotes that the two species are most likely to be assigned to PC species containing an ether bond (alkyl) and may constitute plasmalogens. However, we cannot exclude the possibility that PC O 36:5 may be assigned to PC 35:5, an unlikely odd carbon number species. Similarly, PC O 32:1 may be assigned to PC31:1.

Genotyping and linkage analysis

Illumina's HumanHap6k Genotyping BeadChip (6K Illumina Linkage IV Panels®) was used for genotyping. All genotyping procedures were performed according to the manufacturer's protocols. Only markers with MAF > 0.05 were selected for further analysisGenotyping errors leading to Mendelian inconsistencies were detected using PedCheck [25]. Unlikely double recombination events were detected using MERLIN [26]. All detected errors were eliminated from the data. A total of 5250 autosomal SNPs with call rate greater than 95% were included in the linkage analyses.

ERF is a large pedigree with multiple loops including 23 612 people, spanning 23 generations. Due to computational limitations, the complete pedigree was split into non-overlapping sub-pedigrees using the PedSTR program[27]. A total of 820 persons with genotype and phenotype data were included in the initial analyses[26, 28]. Variance component multipoint linkage analysis was performed using the "vc" option in the MERLIN v.1.0.1 software, which calculates exact IBD sharing probabilities using the Lander-Green algorithm[26].

All traits were adjusted for sex and age and rank transformed prior to linkage analysis. For the traits with point-mass distributions, only "mass" phenotypes were included in the variance component multipoint linkage analysis. Sixteen sphingolipids with point-mass distributions were analyzed using the GADS software for parametric linkage analysis[29].

Regions of interest with LOD > 3.3 were selected for further evaluation. LOD score minus 1 support intervals (LOD-1 SI) around the linkage peaks were extracted using the "qtl" package

implemented in R[30]. Genes encompassed by the LOD-1 SI were annotated using SCAN (SNP and CNV Annotation Database), available at http://www.scandb.org/newinterface/index.html

Exome sequencing analysis

Six hundred exomes from ERF pedigree are sequenced "in-house" (Center for Biomics of the Cell-biology department, Erasmus MC) using the Agilent version V4 capture kit on an Illumina Hiseq2000 sequencer using the TruSeq Version 3 protocol. The sequence reads were aligned to the human genome build 19 (hg19) using BWA and the NARWHAL pipeline[31, 32]. Subsequently, the aligned reads were processed further using the IndelRealigner, MarkDuplicates and TableRecalibration tools from the Genome Analysis Toolkit (GATK) and Picard (http://picard.sourceforge.net) to remove systematic biases and to recalibrate the PHRED quality scores in the alignments. After processing, genetic variants were called using the Unified Genotyper tool from the GATK[33]. The effects of the called variants on the protein sequences were determined with a custom variant annotation script. For each sample, at least 4 Gigabases of sequence was aligned to the genome. Further for comparison and prediction of the functionality of the variants annotations were also performed using the dbNSFP (database of human non-synonymous SNPs and their functional predictions) and Seattle databases available at: http://snp.gs.washington.edu/SeattleSeqAnnotation131/. These databases gave functional prediction results from four different programs including polyPhen2, SIFT, MutationTaster and LRT, apart from gene and variant annotations. About 1.2 million Single Nucleotide Variants (SNVs) were called. After removing the low quality variants (QUAL < 100) we retrieved about 700 000 high quality SNVs were included in the analysis in this study. Among the sequenced 600 individuals, 251 had lipidomics measurements. Residual lipid levels, after adjusting for age and sex, were compared using a T-test contrasting mutation carriers and non-carriers.

RESULTS

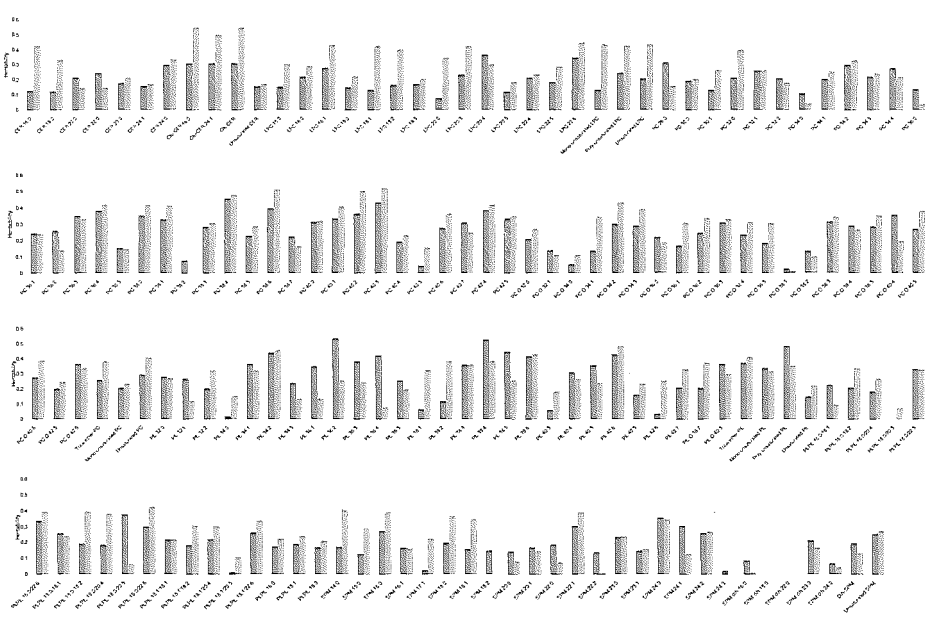
Descriptives of the ERF study population are provided in Table 1. Supplementary Table 1 shows the mean and standard deviations of the lipid species included in this study. The heritability estimates of the plasma phospho– and sphingolipids varied between 0 and 55% and are shown in Figure 1. The highest estimates in each class of lipids species were observed for Glu CER 16:0 (55% for the proportion and 31% for the absolute value), LPC 22:6 (46% for the proportion and 35% for the absolute value), PC 40:3 (52% for the proportion and 44% for the absolute value), PE 40:6 (49% for the proportion and 43% for the absolute value), PLPE 18:0 / 22:6 (43% for the proportion and 31% for the absolute value), SPM 24:0 (41% for the proportion and 17% for the absolute value). In general, within class proportions yielded higher heritability estimates compared to the absolute plasma levels of the same species (in 101 out of 147 measurements). For instance, the heritability of the proportions of CER 16:0, LPC 20:0, PC 40:5, PE 42:6, PLPE

TABLE 1. Descriptives of the ERF population (N= 800)

	Males	Females
Age, years	49.62 (15.66)	50.69 (16.44)
BMI, kg/m2	27.28 (4.13)	26.67 (5.12)
TC, mmol/l	5.48 (1.09)	5.62 (1.09)
HDL-C, mmol/l	1.12 (0.30)	1.39 (0.36)
LDL-C, mmol/l	3.73 (0.97)	3.71 (0.98)
TG, mmol/l	1.49 (0.91)	1.23 (0.64)
Total SPM, µM	498.2 (106.11)	554.41 (114.42)
Total PC, µM	2129.2 (518.42)	2242.84 (447.59)
Total PLPE, µM	53.2 (15.57)	52.50 (17.82)
Total PE, µM	34.67 (18.07)	39.50 (17.24)
Total LPC, µM	273.42 (67.18)	242.99 (64.67)
Total CER, µM	8.66 (2.30)	8.36 (2.01)

Means and standard deviations (sd) are given for the continuous measurements. BMI; body mass index, TC; total cholesterol, HDL-C, high density lipoprotein cholesterol, LDL-C; low density lipoprotein cholesterol, TG; triglycerides, SPM; sphingomyelin, PC; phosphatidylcholine, PLPE; phosphotidylethanolamine plasmologen, PE; phosphatidylethanolamine, LPC; lysophosphatidylcholine, CER; ceramide.

FIGURE 1. Heritability estimates of circulating phosho- sphingolipids in ERF population.



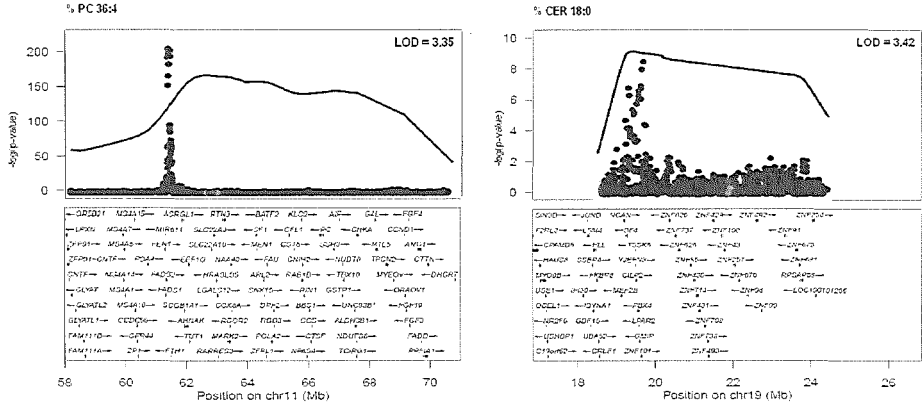
X axis show the species studied, whereas y axis shows the heritability estimates. Absolute values are presented in dark blue, and proportions are presented in turquoise.

18:0/20:4 and SPM dih 24:0 were 2.1 to 9 times higher than the heritability of the absolute values of the same lipids. For a limited number of species the estimates were less than 5%; these were: PLPE 16:0/20:5, SPM 24:3, SPM dih 18:0, SPM dih 22:0, SPM 22:2, SPM dih 16:0, PC O 38:1, PLPE 18:1/20:5, PE 34:0, SPM 24:3, SPM 17:0, PE 42:6, PC 36:0, PC 34:0 and PC 40:5.

Linkage analysis

Nine significant linkage signals (LOD > 3.3) were observed (see Supplementary Table 2 for the markers and genes annotated inside the LOD-1 SI). Only two of them were identified earlier by GWAS: 11q12 and 19p13 as shown in the regional association plots in Figure 2[34]. 11q12.3 has a LOD score of 3.35 for % PC 36:4 and harbors genome-wide significant SNPs (rs102275, *P-value* = 9.88×10^{-204}) for the same lipid. The 19p13 region (LOD = 3.42) was similarly detected by GWAS for % CER 18:0 (rs2304130, *P-value* = 5.85×10^{-9}). The influence of those two GWAS hits on the LOD scores were further assessed by including them as covariates in the linkage analysis. These conditional analyses yielded a non-significant LOD score of 0.03 for the marker rs1525064 on 11q12.3, and also decreased in the heritability estimate of % PC 36:4 (29%). The LOD score on 19p13 for % CER 18:0 did not change when the GWAS SNP was included. For this region, rare coding variants in the exome sequencing data were examined as possible sources of the observed linkage. Within the LOD -1 support interval of the linkage peak, a total of 412 damaging variations were observed (166 synonymous, 242 missense and 4 stop codons). We focused on the rare (1-5%) coding variants which were expected to be damaging by at least 2 of the 3 prediction softwares (Polyphen, LRT and Mutation taster). This selection yielded 1 stop and 7 missense mutations in the genes *ATP13A1, ZNF254, CILP2, CCNE1, SLC25A42, NR2C2AP*

FIGURE 2. Regional association plots of common variations inside 11q12 and 19p12 linkage loci.



P-values are extracted from EUROSPAN lipidomics GWAS. Linkage peak for %PC36:4 is explained by the FADS region SNP rs102275. However linkage peak for % CER 18:0 is not explained by rs23004130.

and *YJEFN3*. Different values for % CER 18:0 were not observed among the carriers of these mutations compared to the non-carriers. None of the other 8 loci (LOD -1 SI) overlapped with the findings from our previous report.

We next explored whether multiple lipids were linked to particular loci. A cluster of PE species provided significant evidence of linkage to 1p36.32, the most remarkable was had LOD = 3.7 for PE 38:5 and is among our nine significant regions. Analysis of PLPE 18:1/22:6 resulted in two different peaks, on 1p31.1 (LOD = 4.26, absolute level) and 1p22.2 (LOD = 4.22, proportion). Two ether-PCs, PC O 36:1 and PC O 34:3 were linked to 18p11.21 and 17q31 (LOD = 3.77 and 3.43 respectively). With respect to sphingolipids, two loci showed significant evidence of linkage: 6q16.3 (LOD = 3.35 for total poly-unsaturated SPMs) and 11q21 (LOD = 3.35 for % of saturated SPMs). We did not observe any rare damaging variants in these regions that were associated with lipid levels.

Figure 3 depicts the genome-wide linkage signals for the clusters of biologically related species linked to significant and suggestive loci for the chromosomes 1, 6, 11, 17, 18 and 19 (LOD > 1.9). Of those, the most obvious was the 1p36.32 locus that was linked to several PE species, but also to total PE levels and % PLPE 16:0 / 22:5. Exact LOD scores are given in Supplementary Table 3.

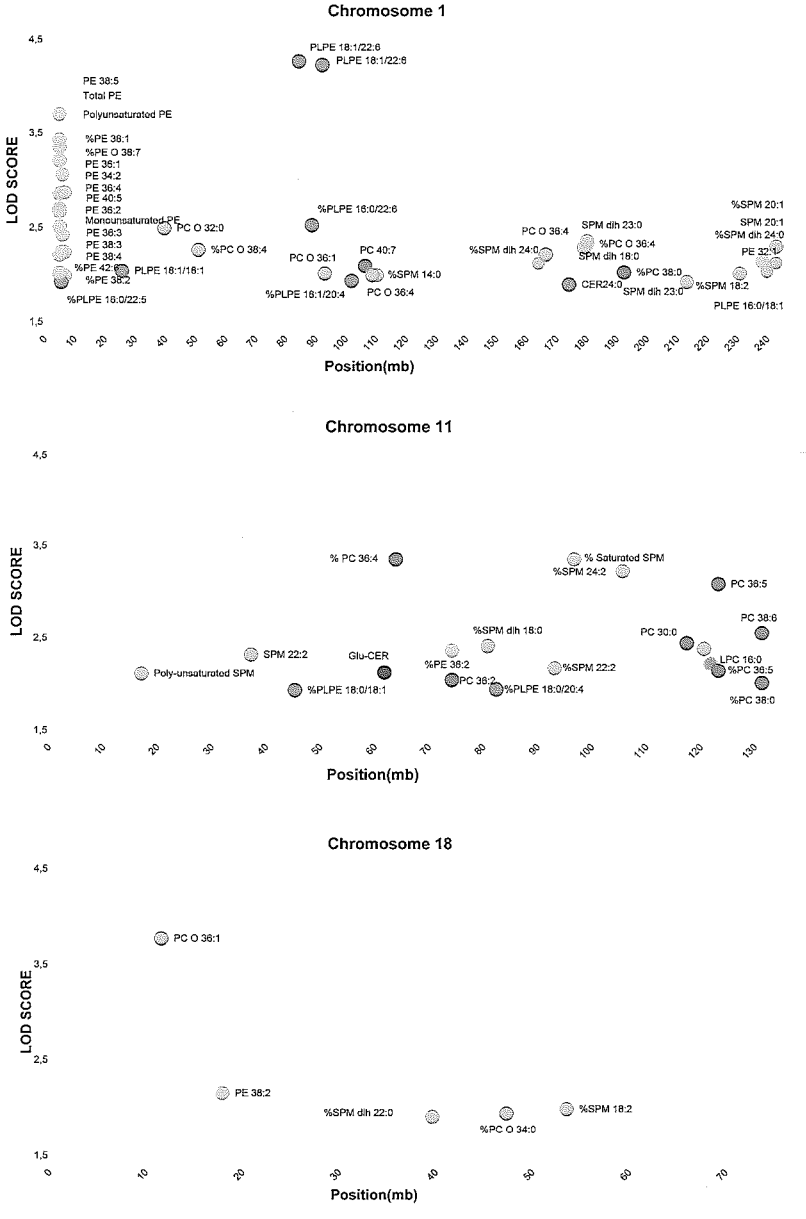
DISCUSSION

Using quantitative linkage methods on human circulating lipidome, heritabilities for 147 circulating species were estimated and 9 quantitative loci with LOD > 3.3 were identified. Two of these loci harboured previously known genes from association studies of phospho- and sphingolipids (*FADS 1-2-3* and *LPAR2*). In addition, 330 loci with suggestive LOD scores (1.9 < LOD < 3.3) were identified.

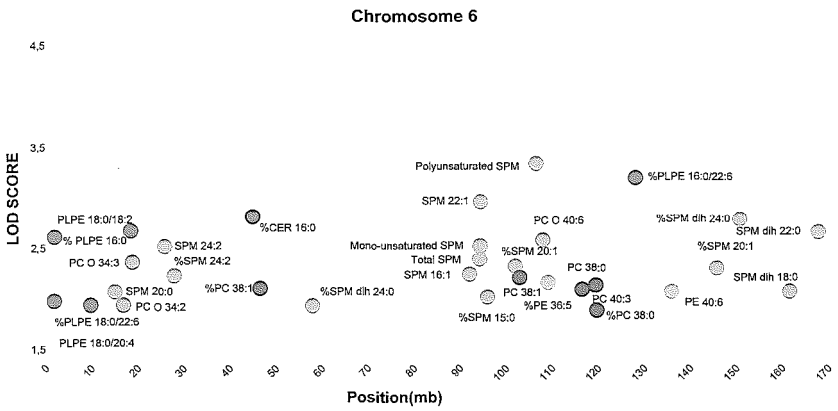
Using family-based methods, we estimated moderate to high heritability for most of the lipid species, although the heritability estimates were lower than 5% some of the traits, which makes them unsuitable for genetic association studies. Two of those lipids; PLPE 16:0 / 20:5 and PLPE 18:1 / 20:5 are of particular interest because of their ω-3 eicosapentaenoic acid (EPA) content. It is of note that EPA is an essential fatty acid which comes mainly from a sea-food diet, but can also be converted from α-linolenic acid (18:3, ω-3) with an efficiency of only a few percent.[35] On the other hand, for another EPA-PLPE, PLPE 18:0 / 20:5, the heritability was estimated as 38 % for the absolute plasma level and 7% for the proportion. This finding points out different genetic and environmental mechanisms for accumulation of EPA, such as the variation in 16:0 and 18:1 PLPEs might be due to non-genetic factors, mainly diet.

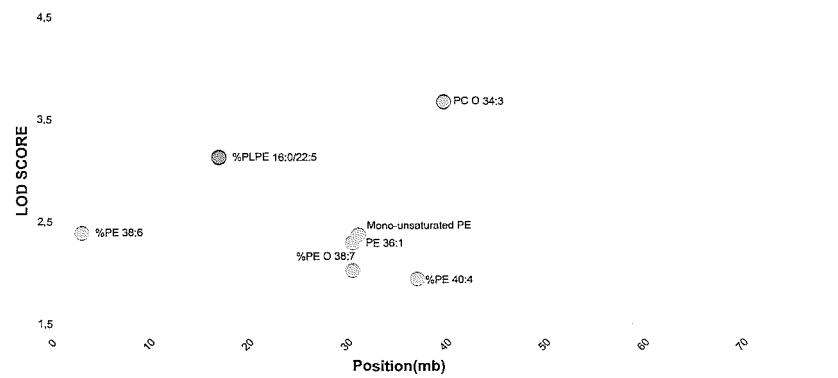
Linkage analysis yielded 9 significant loci on 1p22.2, 1p31.1, 1p36.32, 6q16.3, 11q12.3, 11q21, 17q31, 18p11.21, and 19p12. Of those, genes located on 11q12.3 (*FADS 1-2-3* region) and 19p12 (*LPAR2*) are already known to harbour common variation associated with the same lipids (PC 36:4 and CER 18:0, respectively) [16]. Our conditional analyses demonstrated that rs102275,

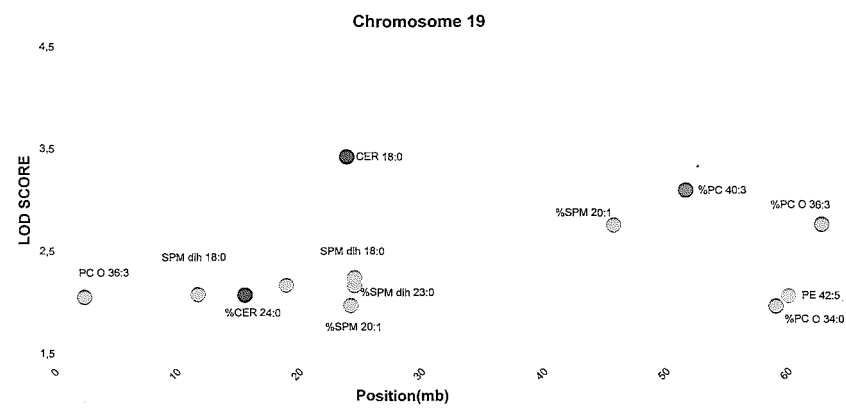
FIGURE 3. Chromosomal locations of the linkage hits for biologically related species.



In each graph X axis show the chromosomal location in base-pair, y axis show the LOD score. PE species are depicted in light blue dots, PLPEs in pink, SPMs in orange, ether-PCs in green, ceramides in dark blue, LPCs in blue and PCs are depicted in purple.







In each graph X axis show the chromosomal location in base-pair, y axis show the LOD score. PE species are depicted in light blue dots, PLPEs in pink, SPMs in orange, ether-PCs in green, ceramides in dark blue, LPCs in blue and PCs are depicted in purple.

or variants in strong LD with it, may be responsible for the linkage signal in this region. As for the 19p12 region, the common variant did not explain the linkage peak in conditional analysis. A search for rare coding variants within this region did not provide additional information. One limitation here is the small sample size in the current exome data set which includes only 251 phenotyped individuals.

Some biologically plausible genes are located inside the LOD-1 SI of the significant linkage peaks; for instance, *PKN2* (cardiolipin-activated protein kinase) is located under the 1p22.2 peak and is known to be activated by lipids, particularly cardiolipin, and to a lesser extent by other acidic phospholipids and unsaturated fatty acids. Our analysis linked % PLPE 18:1 / 22:6 to this locus. The linkage peak for this lipid partially overlaps with the peak on 1p31.1, where *PRKACB* (cAMP-dependent protein kinase catalytic subunit beta) is located. *PRKACB* is also known to be involved in the regulation of lipid and glucose metabolism and is a component of the signal transduction mechanism of certain G-protein couples receptors[36]. The 1p36.32 locus that is linked to a number of PE species harbors *PLCH2*, which is a member of the PLC-η family of the phosphoinositide-specific phospholipase C superfamily of enzymes that cleave phosphatidylinositol to generate secondary messengers inositol 1,4,5-trisphosphate and diacylglycerol[37]. We also observed three loci implicated in carcinogenesis: 6q16.3-21 (polyunsaturated SPMs), 11q21 (% Saturated SPMs) and 17q21.31 (PC O 34:3). Of these, the 6q16.3-21 locus includes tumor suppressors *HACE1* and *PRDM1*[38]. On 11q21, the linkage SNP rs483884 maps inside the *MAML2* gene which is involved in the carcinogenesis pathway through co-activation of *NOTCH* proteins [39]. Finally, the 17q21.31 locus includes *BRCA1*, which is a well known risk gene for early onset breast cancer but has also been shown to influence fatty acid biosynthesis and lipogenesis in normal cells[40, 41], This locus also includes the tau protein gene *MAPT* which incorporates a well characterized copy number variant[42].

We additionally observed 330 suggestively linked loci. Some of those are worth mentioning, since they were linked to multiple biologically related traits: 1q41-44, 2p25.3-p14, 3p14.2-q21.3, 4p15, 5q11.2-q14.3, 6p15-q16, 8q13.1-24.12, and 19p13 were linked to clusters of SPMs. These loci are novel, except for 19p13, which harbors *LASS4* that was previously uncovered by GWAS. 14q23.1- 14q24.2, linked to PC O 34:2, PC O 36:2, PC O 38:3 and PC O 36:3, includes *PLEKHH1*, which was shown to associate with other alkyl-acyl PCs, PC O 36:5 and PC O 32:1, in our recent GWAS[16].

Genetic studies of complex disease conducted to date discovered common variants typically responsible for small effects. These variants explain a small part of the heritability, and sometimes do not directly link to biologically plausible underlying pathways. Studying the circulating phospho- and sphingolipids as specific metabolic endophenotypes may provide better clues to the causal genes involved compared to studies on complex disease end points. However our exome analysis did not reveal any evidence for major mutations suggesting more complex mechanisms outside the coding regions. Taken together with increasing resources and interest

on rare human genetic variation research, family-based studies on these endophenotypes will serve as reliable starting points for the upcoming sequencing era and personalized genomics.

ACKNOWLEDGEMENTS

ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 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). High-throughput genome-wide association analysis of the data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). Lipidomic analysis was supported by the European Commission FP7 grant LipidomicNet (2007-202272). The ERF study was supported by grants from the Netherlands Organisation for Scientific Research (Pionier, 047.016.009, 047.017.043), Erasmus MC and the Centre for Medical Systems Biology (CMSB; National Genomics Initiative). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.

REFERENCES

1. Holthuis JC, Pomorski T, Raggers RJ, Sprong H, Van Meer G. The organizing potential of sphingolipids in intracellular membrane transport. *Physiol Rev.* Oct 2001;81(4):1689-1723.
2. Lajoie P, Goetz JG, Dennis JW, Nabi IR. Lattices, rafts, and scaffolds: domain regulation of receptor signaling at the plasma membrane. *J Cell Biol.* May 4 2009;185(3):381-385.
3. Merrill AS, K. Sphingolipids: metabolism and cell signaling. In: Vance DV, JE, ed. *Biochemistry of Lipids, Lipoproteins and Membranes*; 2002:373-407.
4. Bakovic M, Fullerton MD, Michel V. Metabolic and molecular aspects of ethanolamine phospholipid biosynthesis: the role of CTP:phosphoethanolamine cytidylyltransferase (Pcyt2). *Biochem Cell Biol.* Jun 2007;85(3):283-300.
5. Poli G, Leonarduzzi G, Biasi F, Chiarpotto E. Oxidative stress and cell signalling. *Curr Med Chem.* May 2004;11(9):1163-1182.
6. van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol.* Feb 2008;9(2):112-124.
7. Zheng W, Kollmeyer J, Symolon H, et al. Ceramides and other bioactive sphingolipid backbones in health and disease: lipidomic analysis, metabolism and roles in membrane structure, dynamics, signaling and autophagy. *Biochim Biophys Acta.* Dec 2006;1758(12):1864-1884.
8. Farooqui AA, Horrocks LA, Farooqui T. Glycerophospholipids in brain: their metabolism, incorporation into membranes, functions, and involvement in neurological disorders. *Chem Phys Lipids.* Jun 2000;106(1):1-29.
9. Brugger B, Erben G, Sandhoff R, Wieland FT, Lehmann WD. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc Natl Acad Sci U S A.* Mar 18 1997;94(6):2339-2344.
10. Hodge AM, English DR, O'Dea K, et al. Plasma phospholipid and dietary fatty acids as predictors of type 2 diabetes: interpreting the role of linoleic acid. *Am J Clin Nutr.* Jul 2007;86(1):189-197.
11. 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.* Apr 2008;43(4):289-299.
12. Mannheim D, Herrmann J, Versari D, et al. Enhanced expression of Lp-PLA2 and lysophosphatidylcholine in symptomatic carotid atherosclerotic plaques. *Stroke.* May 2008;39(5):1448-1455.
13. Matsumoto T, Kobayashi T, Kamata K. Role of lysophosphatidylcholine (LPC) in atherosclerosis. *Curr Med Chem.* 2007;14(30):3209-3220.
14. Wang L, Folsom AR, Zheng ZJ, Pankow JS, Eckfeldt JH. Plasma fatty acid composition and incidence of diabetes in middle-aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Clin Nutr.* Jul 2003;78(1):91-98.
15. Kolter T, Sandhoff K. Sphingolipid metabolism diseases. *Biochim Biophys Acta.* Dec 2006;1758(12):2057-2079.
16. Demirkan A, van Duijn CM, Ugocsai P, et al. Genome-Wide Association Study Identifies Novel Loci Associated with Circulating Phospho- and Sphingolipid Concentrations. *PLoS Genet.* Feb;8(2):e1002490.
17. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet.* Feb 2009;42(2):137-141.
18. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* Nov 2008;4(11):e1000282.
19. Hicks AA, Pramstaller PP, Johansson A, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* Oct 2009;5(10):e1000672.

20. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet.* Jul 2004;12(7):527-534.
21. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet.* May 2005;69(Pt 3):288-295.
22. Liebisch G, Drobnik W, Lieser B, Schmitz G. High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clin Chem.* Dec 2002;48(12):2217-2224.
23. Liebisch G, Lieser B, Rathenberg J, Drobnik W, Schmitz G. High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochim Biophys Acta.* Nov 8 2004;1686(1-2):108-117.
24. Zemski Berry KA, Murphy RC. Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. *J Am Soc Mass Spectrom.* Oct 2004;15(10):1499-1508.

25. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet.* Jul 1998;63(1):259-266.
26. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* Jan 2002;30(1):97-101.
27. Kirichenko AV, Belonogova NM, Aulchenko YS, Axenovich TI. PedStr software for cutting large pedigrees for haplotyping, IBD computation and multipoint linkage analysis. *Ann Hum Genet.* Sep 2009;73(Pt 5):527-531.
28. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet.* May 2000;25(1):12-13.
29. Axenovich TI, Zorkoltseva IV. GADS software for parametric linkage analysis of quantitative traits distributed as a point-mass mixture. *Comput Biol Chem.* Feb;36:13-14.
30. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* May 1 2003;19(7):889-890.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* Jul 15 2009;25(14):1754-1760.
32. Brouwer RW, van den Hout MC, Grosveld FG, van Ijcken WF. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics.* Jan 15;28(2):284-285.
33. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* Sep;20(9):1297-1303.
34. Vural B, Yakar F, Derin D, et al. Evaluation of Glutathione S-Transferase P1 Polymorphisms (Ile105Val and Ala114Val) in Patients with Small Cell Lung Cancer. *Genet Test Mol Biomarkers.* Feb 17.
35. Anderson BM, Ma DW. Are all n-3 polyunsaturated fatty acids created equal? *Lipids Health Dis.* 2009;8:33.
36. Higuchi H, Yamashita T, Yoshikawa H, Tohyama M. PKA phosphorylates the p75 receptor and regulates its localization to lipid rafts. *Embo J.* Apr 15 2003;22(8):1790-1800.
37. Zhou Y, Wing MR, Sondek J, Harden TK. Molecular cloning and characterization of PLC-eta2. *Biochem J.* Nov 1 2005;391(Pt 3):667-676.
38. Zhang L, Anglesio MS, O'Sullivan M, et al. The E3 ligase HACE1 is a critical chromosome 6q21 tumor suppressor involved in multiple cancers. *Nat Med.* Sep 2007;13(9):1060-1069.
39. Nemoto N, Suzukawa K, Shimizu S, et al. Identification of a novel fusion gene MLL-MAML2 in secondary acute myelogenous leukemia and myelodysplastic syndrome with inv(11)(q21q23). *Genes Chromosomes Cancer.* Sep 2007;46(9):813-819.
40. Ortega FJ, Moreno-Navarrete JM, Mayas D, et al. Breast Cancer 1 (BrCa1) May Be behind Decreased Lipogenesis in Adipose Tissue from Obese Subjects. *PLoS One.*7(5):e33233.
41. Ray H, Suau F, Vincent A, Dalla Venezia N. Cell cycle regulation of the BRCA1/acetyl-CoA-carboxylase complex. *Biochem Biophys Res Commun.* Jan 16 2009;378(3):615-619.

42. Rovelet-Lecrux A, Campion D. Copy number variations involving the microtubule-associated protein tau in human diseases. *Biochem Soc Trans.* Aug 1;40(4):672-676.

Chapter 4.2

Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations.

Ayşe Demirkan, Cornelia M van Duijn, Peter Ugocsai, Aaron Isaacs, Peter P Pramstaller, Gerhard Liebisch, James F Wilson, Åsa Johansson, Igor Rudan, Yurii S Aulchenko, Anatoly V Kirichenko, A Cecile JW Janssens, Ritsert C Jansen, Carsten Gnewuch, Francisco S Domingues, Cristian Pattaro, Sarah H Wild, Inger Jonasson, Ozren Polasek, Irina V Zorkoltseva, Albert Hofman, Lennart C Karssen, Maksim Struchalin, James Floyd, Wilmar Igl, Zrinka Biloglav, Linda Broer, Arne Pfeufer, Irene Pichler, Susan Campbell, Ghazal Zaboli, Ivana Kolcic, Fernando Rivadeneira, Jennifer Huffman, Nicholas D Hastie, Andre Uitterlinden, Lude Franke, Christopher S Franklin, Veronique Vitart, DIAGRAM Consortium, Christopher P Nelson, Michael Preuss, CARDIoGRAM Consortium, Joshua C Bis, Christopher J O'Donnell, Nora Franceschini, CHARGE Consortium, Jacqueline CM Witteman, Tatiana Axenovich, Ben A Oostra, Thomas Meitinger, Andrew A Hicks, Caroline Hayward, Alan F Wright, Ulf Gyllensten, Harry Campbell, Gerd Schmitz, on behalf of the EUROSPAN consortium.

PLoS Genet. Feb;8(2):e1002490.

ABSTRACT

Phospho- and sphingolipids are crucial cellular and intracellular compounds. These lipids are required for active transport, a number of enzymatic processes, membrane formation and cell signalling. Disruption of their metabolism leads to several diseases, with diverse neurological, psychiatric, and metabolic consequences. A large number of phospholipid and sphingolipid species can be detected and measured in human plasma. We conducted a meta-analysis of five European family-based genome-wide association studies (N = 4034) on plasma levels of 24 sphingomyelins (SPM), 9 ceramides (CER), 57 phosphatidylcholines (PC), 20 lysophosphatidylcholines (LPC), 27 phosphatidylethanolamines (PE) and 16 PE-based plasmalogens (PLPE), as well as their proportions in each major class. This effort yielded 25 genome-wide significant loci for phospholipids (smallest *P*-value = 9.88×10^{-204}) and 10 loci for sphingolipids (smallest *P*-value = 3.10×10^{-57}). After a correction for multiple comparisons (*P*-value < 2.2×10^{-9}), we observed four novel loci significantly associated with phospholipids (*PAQR9, AGPAT1, PKD2L1, PDXDC1*) and two with sphingolipids (*PLD2* and *APOE*) explaining up to 3.1% of the variance. Further analysis of the top findings with respect to within class molar proportions uncovered three additional loci for phospholipids (*PNLIPRP2, PCDH20* and *ABDH3*) suggesting their involvement in either fatty acid elongation/saturation processes or fatty acid specific turnover mechanisms. Among those, 14 loci (*KCNH7, AGPAT1, PNLIPRP2, SYT9, FADS1-2-3, DLG2, APOA1, ELOVL2, CDK17, LIPC, PDXDC1, PLD2, LASS4,* and *APOE*) mapped into the glycerophospholipid and 12 loci *(ILKAP, ITGA9, AGPAT1, FADS1-2-3, APOA1, PCDH20, LIPC, PDXDC1, SGPP1, APOE, LASS4, and PLD2)* to the sphingolipid pathways. In large meta-analyses, associations between *FADS1-2-3* and carotid intima media tickness, *AGPAT1* and type 2 diabetes and *APOA1* and coronary artery disease were observed. In conclusion, our study identified nine novel phospho- and sphingolipid loci, substantially increasing our knowledge of the genetic basis for these traits.

INTRODUCTION

Phospho- and sphingolipids are present in all eukaryotic cell membranes and contribute to organelle structure and signalling events that influence cell behaviour and function[55-57]. Phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), lysophosphatidylcholines (LPCs) and PE-based plasmalogens (PLPE) are major classes of phospholipids that play an important role in several key processes such as cell survival and inflammation[58-60]. Sphingolipids are also essential components of plasma membranes and endosomes and are believed to play critical roles in cell surface protection, protein and lipid transport and sorting, and cellular signalling cascades[61]. In plasma, PC, PE and SPM are included in the structure of lipoproteins; they constitute more than two-thirds of the total phospholipid content in HDL-C and LDL-C, as well as in platelets[62, 63]. Remarkable differences in plasma lipoprotein acceptor affinities for the phospholipids exist (LDL-C is the major acceptor for SPM, whereas HDL-C is the predominant acceptor for PC)[63]. Altered concentrations of circulating phospholipids have been implicated in the pathology of type 2 diabetes, dyslipidemia and cardiovascular disease[64-69], as well as a wide range of other common diseases including dementia and depression[70].

Identifying genetic variants that influence phospho- and sphingolipid concentrations will be an important step towards understanding pathways contributing to common human disease. Earlier studies of these metabolites identified a number of genetic loci associated with their levels in blood[71-73]. We conducted a meta-analysis of genome-wide association studies (GWAS) on plasma levels of 24 SPMs, 9 CERs, 57 PCs, 20 LPCs, 27 PEs and 16 PLPEs in five European populations: (1) the Erasmus Rucphen Family (ERF) study, conducted in the Netherlands, (2) the MICROS study from the Tyrol region in Italy, (3) the Northern Swedish Population Health Survey (NSPHS) in Norrbotten, Sweden, (4) the Orkney Complex Disease Study (ORCADES) in Scotland, and (5) the CROAS (CROATIA_Vis) study conducted on Vis Island, Croatia.

The top findings were further analysed by adjusting for plasma HDL-C, LDL-C, TG and TC levels. The influences of these top hits on within class lipid ratios were also assessed, to help elucidate potential mechanisms. Finally, the variants that were associated with plasma phospho- and sphingolipid levels were tested for association with carotid intima media thickness (IMT), type 2 diabetes (T2DM), and coronary-artery disease (CAD) using large consortia meta-analysis results.

MATERIALS AND METHODS

All studies were approved by the local ethical committees. Detailed descriptions of the study populations that contributed to the meta-analysis, as well as detailed information on ethical statements, genotyping, lipid measurements and pathway analysis, are presented in the Text S1. Briefly, lipid species were quantified by electrospray ionization tandem mass spectrometry (ESIMS/ MS) using methods validated and described previously [74, 75]. For each lipid molecule,

we adopted the naming system where lipid side chain composition is abbreviated as *Cx:y*, where *x* denotes the number of carbons in the side chain and y the number of double bonds. For example, PC 34:4 denotes an acyl-acyl phosphatidylcholine with 34 carbons in the two fatty acid side chains and 4 double bonds in one of them. Lipid traits were analysed individually as well as aggregated into groups of species with similar characteristics (e.g. unsaturated ceramides). These were then analyzed as both absolute concentrations (μM) and as molar percentages within lipid sub-classes (mol%) (calculated as the proportion of each lipid molecule among its own class (e.g. PC, PE, PE-pls, LPC)). The additive value of the analyses of molar proportions is that it may bring to light genes involved in the transition of one species to another, such as through fatty acid chain elongation or (de)saturation. We also performed single SNP association analyses for each novel locus and the ratio of the index lipid (for example, PC 34:1) to the other lipids in the same class (in the example, PC 34:1/PC 36:1, PC 34:1/PC 38:1) so that we could determine whether the SNP might be involved in elongation or (de)saturation.

DNA samples were genotyped according to the manufacturer's instructions on Illumina Infinium HumanHap300v2, HumanHap300v1 or HumanCNV370v1 SNP bead microarrays. Genotype data for these five populations were imputed using MACH 1.0 (v1.0.16)[33,76] using the HapMap CEU population (release 22, build 36).

As all of the studies included related individuals, testing for association between lipid and allele dosage were performed using a mixed model approach as implemented with the 'mmscore' option in the GenABEL software [43]. Results from the five populations were combined using inverse variance weighted fixed-effects model meta-analyses using the METAL software [77]. To correct for multiple testing, we adopted a Bonferroni correction for the number of phenotypes studied. Since most of the lipid values are correlated with each other, we used the number of principal components (n = 23) that accounted for 79% of the phenotypic variance for this correction and applied it to the classical genome-wide significance threshold (5×10^{-8}).

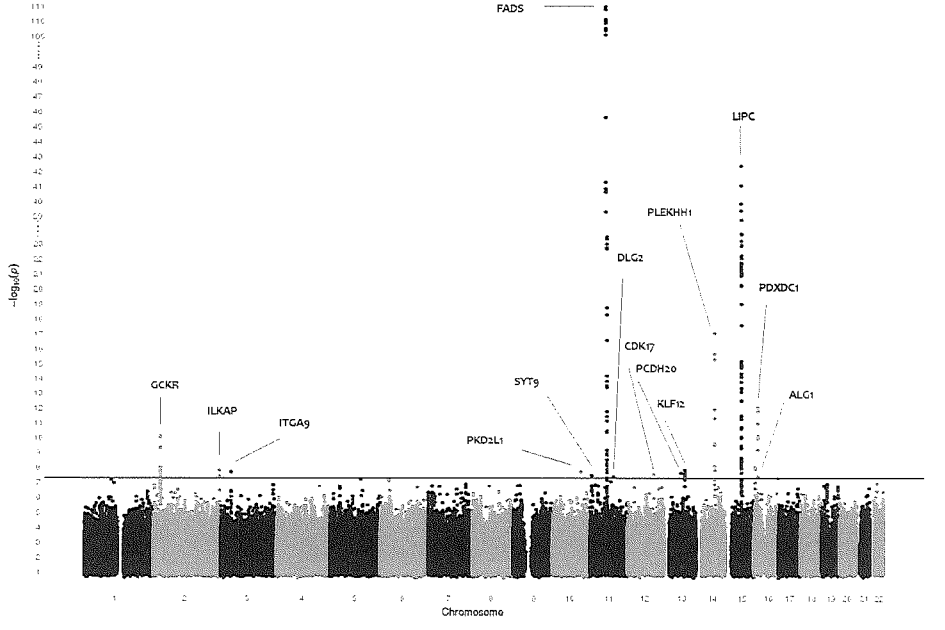
RESULTS

Table 1 provides an overview of the study populations. The mean age, gender ratio and mean values of major classes of phospho- and sphingolipids were comparable among the 5 populations. Means for the individual species are presented in Table S1. Figures 1A and 1B show the combined Manhattan plot for the meta-analyses of the absolute values and proportions of all phospholipid traits, respectively; Figures 2A and 2B provide the same for the sphingolipids. Out of 357 meta-analyses performed, 202 outcomes yielded genome-wide significant findings, most of which were located around two genes, *FADS* and *LIPC*, which were identified previously [71,73] as key lipid regulators and are associated with a large number of species (Tables 2 and 3). Q-Q plots for the lipid GWAS that yielded significant associations are provided in Supplementary Figure 1.

TABLE 1. Study population

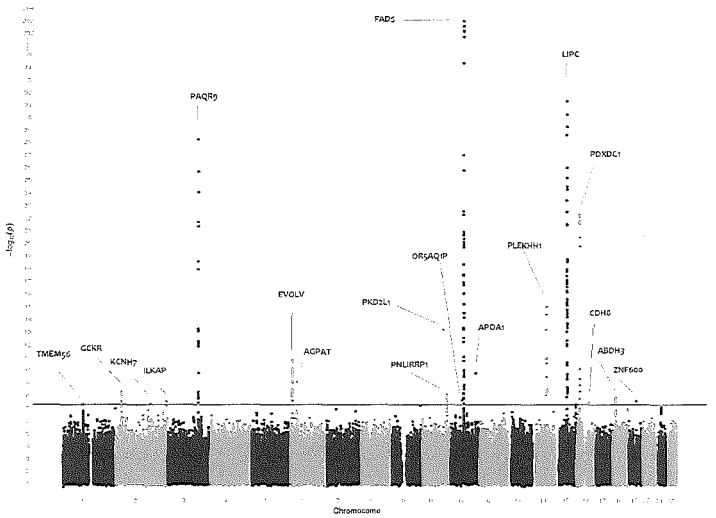
	ERF	MICROS	NSPHS	ORCADES	VIS
N = 4034	800	1086	654	714	780
%Female(56.4)	60.13	56.63	52.75	53.36	57.94
Mean(sd)					
Age	49.65(15.20)	45.26(16.08)	46.98(20.70)	53.59(15.71)	56.55(15.36)
Total SPM	532.21(109.39)	587.03(114.19)	516.98(121.98)	468.25(96.4)	499.01(105.26)
Total PC	2198.55(444.36)	2527.88(457.25)	2249.18(493.67)	1941.73(372.96)	2066.56(427.91)
Total PE-pls	52.68(14.91)	58.21(17.05)	62.28(24.28)	42.15(13.16)	55.60(15.39)
Total PE	37.22(16.69)	37.15(18.12)	20.58(7.25)	25.02(11.19)	34.34(14.55)
Total CER	8.45(1.95)	9.26(2.12)	9.30(2.52)	7.12(1.83)	9.05(2.23)

FIGURE 1A. Genome-wide association results for the plasma levels of 115 phospholipid species.



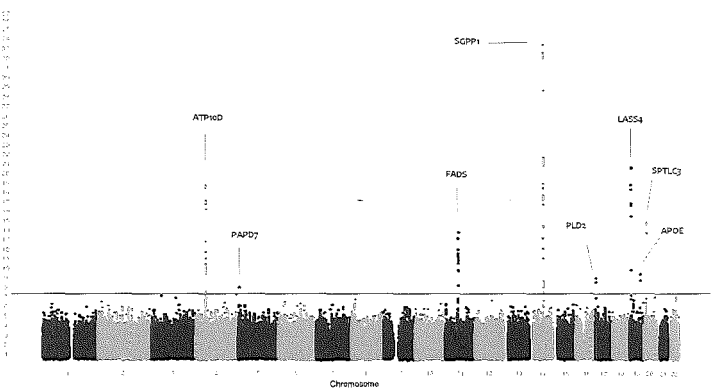
Manhattan plots show the combined association signals (–log 10 of *p*-value) on the y axis versus SNPs according to their position in the genome on the x axis (build 36). Novel genes are represented in red, while previously known loci are represented in black.

FIGURE 1B. Genome-wide association results for the within-class percentages of 115 plasma phospholipids.



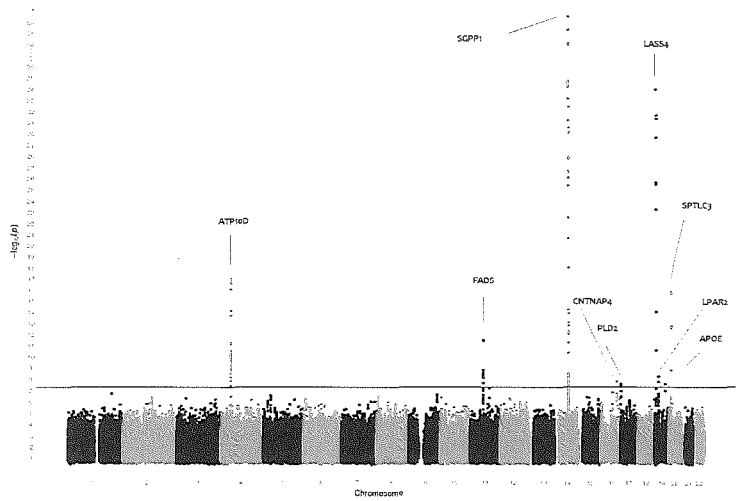
Manhattan plots show the combined association signals (–log 10 of *p*-value) on the y axis versus SNPs according to their position in the genome on the x axis (build 36). Novel genes are represented in red, while previously known loci are represented in black.

FIGURE 2A. Genome-wide association results for the plasma levels of 33 sphingolipid species.



Manhattan plots show the combined association signals (–log 10 of *p*-value) on the y axis versus SNPs according to their position in the genome on the x axis (build 36). Novel genes are represented in red, while previously known loci are represented in black.

FIGURE 2B. Genome-wide association results for the within-class percentages of 33 sphingolipid species.



MANHATTAN PLOTS show the combined association signals (–log 10 of *p*-value) on the y axis versus SNPs according to their position in the genome on the x axis (build 36). Novel genes are represented in red, while previously known loci are represented in black.

TABLE 2A. Variants significantly associated with circulating phospholipid levels

SNP	Chromosomal region	Position	Gene	Distance (kb)	#SNPs	$P_{Nominal}$	$P_{Corrected}$	% Explained variance	MAF	Species associated
rs4666002 *	2p23.3	27840640	GCKR	97	33	7.48×10^{-11}	1.72×10^{-09}	1.1-0.7	0.21	PC 34:4
rs12472274⋈	2q37.3	239095422	ILKAP	intronic	3	1.52×10^{-08}	3.49×10^{-07}	0.8	0.18	PC 40:3
rs197770⋈	3p22.2	37515827	ITGA9	intronic	1	1.95×10^{-08}	4.49×10^{-07}	0.8	0.14	PLPE 18:0/22:6
rs603424*⋈☢	10q24.31	102075479	PKD2L1	intronic	1	6.41×10^{-14}	1.47×10^{-12}	1.4	0.17	LPC 16:1
rs10769780⋈	11p15.4	7367095	SYT9	intronic	3	3.31×10^{-08}	7.61×10^{-07}	0.8	0.38	PC O 42:6
rs102275*	11q12.2	61557803	FADS-1-2-3	21	122	9.88×10^{-204}	2.27×10^{-202}	0.8-23.0	0.30	PE 32:1, PE 34:1, PE 34:2, PE 34:3. See supporting text file S1 for the full list.
rs17148090⋈	11q14.1	85017476	DLG2	intronic	1	4.14×10^{-08}	9.52×10^{-07}	0.7	0.05	PC O 42:5
rs12423247⋈	12q23.1	96858362	CDK17	22	1	3.09×10^{-08}	7.11×10^{-07}	0.8	0.04	PC O 42:6
rs7337573⋈☢	13q21.2	61513669	PCDH20	470	1	2.39×10^{-08}	5.49×10^{-07}	0.8	0.00	PC 32:1
rs17718828⋈	13q22.1	75128114	KLF12	422	5	1.49×10^{-08}	3.43×10^{-07}	0.8	0.05	PC O 42:5
rs1077989*	14q24.1	67975822	PLEKHH1	23	15	9.29×10^{-18}	2.14×10^{-16}	0.8-1.8	0.47	PC O 36:5, PC O 32:1
rs10468017*	15q21.3	58678512	LIPC	20	95	6.59×10^{-43}	1.52×10^{-41}	0.8-4.7	0.30	PE 36:4, PE 34:2, PE 38:5. See supporting text file S1 for the full list.
rs870288⋈	16p13.3	5585852	ALG1	410	2	9.57×10^{-09}	2.20×10^{-07}	0.8	0.24	PC30:1
rs4500751*⋈	16p13.11	15140211	PDXDC1	intronic	21	5.59×10^{-23}	1.29×10^{-21}	0.8-2.4	0.30	LPC 20:3, PC 38:3

TABLE 2B. Variants significantly associated with circulating phospholipid proportions

SNP	Chromosomal region	Position	Gene	Distance (kb)	#SNPs	$P_{Nominal}$	$P_{Corrected}$	% Explained variance	MAF	Species associated
rs9437689⋈	1p21.3	95549536	ALG14	11	1	4.92×10^{-08}	1.13×10^{-06}	0.7	0.46	%LPC 18:0
rs4666002 *	2p23.3	27840640	GCKR	97	33	7.48×10^{-11}	1.72×10^{-09}	1.1-0.7	0.21	%PC 34:4
rs1424760⋈	2q24.3	163781787	KCNH7	85	1	4.11×10^{-08}	9.46×10^{-07}	0.7	0.29	%PC O 36:5
rs12472274⋈	2q37.3	239095422	ILKAP	intronic	3	1.52×10^{-08}	3.49×10^{-07}	0.8	0.18	%PC 40:3
rs9832727*⋈	3q23	142649110	PAQR9	18	28	5.88×10^{-29}	1.35×10^{-27}	0.8-3.1	0.31	%PE34:1, %PE36:1, %Mono-unsaturated PEs
rs17606561*	6p24.2	10982359	ELOVL2	3'UTR	52	1.49×10^{-11}	3.44×10^{-10}	0.8-1.1	0.21	%PC 38:5, %PE 38:5, %LPC 22:5
rs1061808*⋈	6p21.32	32136547	AGPAT1	3'UTR	3	7.78×10^{-10}	1.79×10^{-08}	0.9	0.45	%PC 32:0
rs603424*⋈☢	10q24.31	102075479	PKD2L1	intronic	1	6.41×10^{-14}	1.47×10^{-12}	1.4	0.17	%LPC 16:1
rs10885997⋈☢	10q25.3	118397971	PNLIPRP2	synonymous	7	7.88×10^{-09}	1.81×10^{-07}	0.8	0.38	%PC 36:1
rs2945816⋈	11q12.1	55862091	OR8I2	0.3	1	2.15×10^{-08}	4.93×10^{-07}	0.8	0.01	%PC 26:0
rs102275*	11q12.2	61557803	FADS-1-2-3	21	122	9.88×10^{-204}	2.27×10^{-202}	0.8-23.0	0.30	%PC 36:4, %PC 38:4, %PE 38:4. See supporting text file S1 for the full list.
rs964184*	11q23.3	116648917	APOA5	11	1	1.71×10^{-10}	3.93×10^{-09}	1.0	0.13	PC/SPM, %PC O 34:1, %PC O 32:0, % polyunsaturated PCs
rs1077989*	14q24.1	67975822	PLEKHH1	23	15	9.29×10^{-18}	2.14×10^{-16}	0.8-1.8	0.47	%PC O 36:5
rs10468017*	15q21.3	58678512	LIPC	20	95	6.59×10^{-43}	1.52×10^{-41}	0.8-4.7	0.30	%PE32:0, %PE32:2, %PE34:0. See supporting text file S1 for the full list.
rs4500751*⋈	16p13.11	15140211	PDXDC1	intronic	21	5.59×10^{-23}	1.29×10^{-21}	0.8-2.4	0.30	%LPC 20:3, %PC 34:2, %PC 36:3, %PC 38:3, %PE 38:3,
rs9932186⋈	16q21	60314656	CDH8	1300	1	3.33×10^{-08}	7.67×10^{-07}	0.8	0.24	%Saturated LPCs
rs11662721⋈☢	18q11.2	19261413	ABHD3	intronic	25	1.36×10^{-08}	3.13×10^{-07}	0.7-0.8	0.26	%PC32:2
rs10404486⋈	19q13.41	53284135	ZNF600	intronic	1	8.53×10^{-09}	1.96×10^{-07}	0.8	0.16	PC/LPC

P-value $_{Bonferroni}$ Genome-wide association p-value after adjustment for number of independent vectors.

* Loci significantly associated to lipid levels

⋈ Loci associated to phospholipids for the first time

☢ Loci significantly (*P*-value<2.2×10^{-9}) associated to within class phospholipid ratios

TABLE 3A. Variants significantly associated with circulating sphingolipid concentrations

SNP	Chromosomal region	Position	Gene	Distance (kb)	#SNPs	$P_{Nominal}$	$P_{Corrected}$	% Explained variance	MAF	Species associated
rs13106975*	4p12	47551863	ATP10D	intronic	80	1.93×10^{-19}	4.45×10^{-18}	0.7-2.0	0.20	GluC16:0, GluC24:1, Glu-C
rs1566039ᴹ	5p15.31	6821914	PAPD7	62	1	1.09×10^{-08}	2.51×10^{-07}	0.8	0.40	SPM16:1-OH
rs174479*	11q12.2	61678754	FADS-1-2-3	17	32	1.99×10^{-14}	4.58×10^{-13}	0.8-1.5	0.49	SPM16:1, SPM18:1, SPM20:1, SPM22:1
rs17101394*	14q23.2	64232386	SGPP1	37	106	3.10×10^{-57}	7.13×10^{-56}	07-6.3	0.15	SPM14:0, SPM15:0, SPMdih16:0
rs12051548*ᴹ☢	17p13.2	4683035	PLD2	24	2	1.21×10^{-09}	2.78×10^{-08}	0.7-0.9	0.03	SPM23:0
rs7258249*	19p13.2	8271721	LASS4	2	9	1.09×10^{-34}	2.52×10^{-33}	0.8-3.7	0.48	C20:0, SPM20:1, SPM20:0, SPM18:1, SPM18:0
rs7259004*ᴹ	19q13.32	45432557	APOE-C1-C2-C4	9	2	4.89×10^{-10}	1.12×10^{-08}	0.8-1.0	0.11	SPM24:0, SPM22:0
rs680379*	20p12.1	12969400	SPTLC3	20	8	1.61×10^{-16}	3.70×10^{-15}	0.8-1.7	0.29	C22:0, C23:0, C24:1, C24:0, Sat-C, Tot-C, Unsa-C, Glu-C, SPM 16:1-OH

TABLE 3B. significantly associated with circulating sphingolipid concentrations

SNP	Chromosomal region	Position	Gene	Distance (kb)	#SNPs	$P_{Nominal}$	$P_{Corrected}$	% Explained variance	MAF	Species associated
rs13106975*	4p12	47551863	ATP10D	intronic	80	1.93×10^{-19}	4.45×10^{-18}	0.7-2.0	0.20	%GluC16:0, Glu-C, %Glu-C
rs174479*	11q12.2	61678754	FADS-1-2-3	17	32	1.99×10^{-14}	4.58×10^{-13}	0.8-1.5	0.49	% SPM16:1, spm/C
rs17101394*	14q23.2	64232386	SGPP1	37	106	3.10×10^{-57}	7.13×10^{-56}	07-6.3	0.15	% SPM 14:0, %SPM22:1, %SPM23:0,
rs4485401ᴹ	16q23.1	76566157	CNTNAP4	intronic	1	1.60×10^{-08}	3.68×10^{-07}	0.8	0.45	% Glu C24:1, %Glu-C
rs12051548*ᴹ☢	17p13.2	4683035	PLD2	24	2	1.21×10^{-09}	2.78×10^{-08}	0.7-0.9	0.03	% SPM23:0, SPM23:0
rs2304130ᴹ	19p13.11	19789528	LPAR2	49	2	5.85×10^{-09}	1.34×10^{-07}	0.8	0.06	%C18:0
rs7258249*	19p13.2	8271721	LASS4	2	9	1.09×10^{-34}	2.52×10^{-33}	0.8-3.7	0.48	%C20:0,%SPM20:1, %SPM20:0, %SPM24:1, %SPM18:1, %SPM18:0
rs7259004*ᴹ	19q13.32	45432557	APOE-C1-C2-C4	9	2	4.89×10^{-10}	1.12×10^{-08}	0.8-1.0	0.11	%SPM22:0
rs680379*	20p12.1	12969400	SPTLC3	20	8	1.61×10^{-16}	3.70×10^{-15}	0.8-1.7	0.29	%C16:0, %Glu C16:0, spm/C, % SPM20:0, %SPM16:1, %SPM16:1-OH

$P\text{-value}_{Bonferroni}$ Genome-wide association p-value after adjustment for number of independent vectors.

* Loci significantly associated to lipid levels after Bonferroni correction

ᴹ Loci associated to sphingolipids for the first time

Phospholipids

As shown in Table 2, 25 loci were nominally associated (*P*-value < 5×10^{-8}) with absolute plasma levels and/or proportions of the phospholipid species. Among those loci, previously reported relationships between the *FADS1, LIPC, PLEKHH1, GCKR, APOA1-5,* and *ELOVL2* loci and phospholipids were successfully replicated [71, 73]. Four novel genome-wide significant loci were also detected after a multiple testing correction to adjust for the approximate number of independent genotypes and phenotypes (n = 23) studied (*P*-value < 2.2×10^{-9}). These included *PAQR9* on 3q23 (associated with %PE 34:1 and %PE 36:1), *AGPAT1* on 6p21.32 (associated with PC 32:0), *PKD2L1* on 10q24.31 (LPC 16:1), and *PDXDC1* on 16p13.11 (LPC 20:3, PC 34:2, PC 36:3 and PC 38:3). Fifteen additional regions provided suggestive evidence of association (2.2×10^{-9} < *P*-value < 5×10^{-8}) with phospholipids including the *PNLIPRP2* locus, associated with %PC 36:1; *ZNF600* with PC/LPC ratio; *ALG1* with PC 30:1; *ABHD3* with %PC 32:2; *KLF12* and *DLG2*, both associated with PC O 42:5; *ILKAP* with PC 40:3 and %PC 40:3; *ITGA9* with PLPE 18:0/22:6; *OR8I2* with %PC 26:0; *PCDH20* with PC 32:1; *CDK17* and *SYT9*, both associated with PC O 42:6; *CDH8* with the proportion of saturated LPC; *KCNH7* with %PC O 36:5; and *ALG14* with %LPC 18:0. Regional association plots for all phospholipid loci are presented in Supplementary Figure 2.

Many of the genome-wide significant and suggestive loci in Table 2 were associated with the percentage of each lipid molecule within its own class (mol%) rather than to absolute values. Single SNP analysis of ratios showed that rs4500751 *(PDXDC1)* was strongly associated with PC 36:3/PC 34:2 (*P*-value = 4.37×10^{-25}) and LPC 20:3/LPC 16:1 (*P*-value = 6.84×10^{-23}) (Supplementary Table 2). Further, rs11662721 *(ABHD3)* was associated with the ratio of PC 32:2 to PC 36:2 (*P*-value = 9.35×10^{-10}), but also to PC 36:3 (*P*-value = 1.80×10^{-9}) and PC 38:3 (*P*-value = 6.71×10^{-9}). rs9437689 *(ALG14)* and rs603424 *(PKD2L1)* were associated with the ratios of LPC 16:0 to LPC 18:0 (*P*-value = 2.70×10^{-8}) and LPC 16:1(*P*-value = 2.25×10^{-15}), respectively. SNP rs10885997 *(PNLIPRP2)* was associated with PC 36:1/PC 34:1 (*P*-value = 3.28×10^{-10}) and PC 36:1/PC 34:3 (*P*-value = 1.15×10^{-9}). SNP rs7337585 *(PCDH20)* was associated with the ratio of PC 32:1 to several ether bound PC species (the strongest association was with PC 32:1/PC O 32:0; *P*-value = 1.82×10^{-18}) and, finally, rs2945816 *(OR8I2)* was associated with the ratio of PC26:0 to several long chain PCs (the strongest association was with PC 26:0/PC 36:1; *P*-value = 2.93×10^{-9}).

Sphingolipids

Table 3 shows the 10 loci that were associated with either absolute plasma levels (panel A) or percentages (panel B) of sphingomyelin species or ceramides. Among those loci, 5 *(ATP10D, FADS1-3, SGPP1, SPTLC3, LASS4)* were previously described in genome-wide analyses [72, 73]. These loci retained significance after adjustment for the number of genotypes and phenotypes tested.

In addition, five novel loci were identified at a nominal *P*-value of 5×10^{-8} (*PAPD7*, *CNTNAP4*, *PLD2*, *LPAR2*, and *APOE*). Two of these, *APOE* on 19q13.32 (associated with SPM 24:0 and SPM 22:0) and *PLD2* on 17p13.2 (associated with SPM 23:0), remained significant after correction for the number of phenotypes tested. The other three showed suggestive evidence of association ($2.2 \times 10^{-9} < P\text{-value} < 5 \times 10^{-8}$) to either sphingomyelins or ceramides: *PAPD7* on 5p15.31 (SPM 17:0), the *CNTNAP4* region on 16q23.1 (% Glu-CER 24:1, %Glu-CER) and *LPAR2* on 19p13.11 (% C 18:0). Regional association plots for the sphingolipid loci are presented in Figure S3.

When studying the ratios of the index lipid to the other lipids within the same class, the strongest association for rs12051548 *(PLD2)* was found with the SPM 23:0/SPM 16:1 ratio (*P*-value = 2.43×10^{-10}). SNP rs7259004 in the *APOE* locus was strongly associated with the ratio of SPM 24:0 to SPM 24:2 (*P*-value = 5.11×10^{-9}) and SPM 16:1 (*P*-value = 4.79×10^{-8}) but also with the ratio of SPM 22:0 to the same lipids (SPM 24:2: *P*-value = 2.91×10^{-8} and SPM 16:1: *P*-value = 1.98×10^{-8}).

HDL-C, LDL-C, TG and TC

As a point of reference, the genome-wide significant findings (*P*-value < 5×10^{-8}) from the GWAS of TC, LDL-C, HDL-C, and TG in these samples are provided in Supplementary Table 3. *CETP* was associated with HDL-C levels (*P*-value = 8.5×10^{-20}), *APOE* was associated with LDL-C (*P*-value = 9.2×10^{-26}) and TC levels (*P*-value = 4.6×10^{-11}). *APOA1-5* (*P*-value = 1.6×10^{-8}) and *PDCD11* (*P*-value = 2.7×10^{-10}) were associated with TG levels. Except for the *PDCD11* locus, these associations have all been previously reported [14].

To determine if the associations of the phospho- and sphingolipid loci were mediated by these major classes of plasma lipoproteins, conditional analyses were performed. Supplementary Table 4 shows the effect size, standard error, and *P*-values for the genome-wide significant loci when adjusted for HDL-C, LDL-C, TG and TC. Only the association of the *APOE* locus (rs7259004) with SPMs was greatly affected by the incorporation of LDL-C and TC. No other major differences were observed in effect size or *P*-value.

Pathway analyses

Finally, we investigated whether the genes from the GWAS fit into previously known sphingolipid and glycerophospholipid pathways, which are available among the canonical pathways from various data bases provided by ConsensusPathDB [78]. By testing for enrichment of known pathways, glycerolipid metabolism *(P*-value = 0.002; KEGG), chylomicron-mediated lipid transport (*P*-value = 0.003; Reactome), triglyceride biosynthesis (*P*-value = 0.006; Reactome), metabolism of lipids and lipoproteins (*P*-value = 0.002; Reactome) and biosynthesis of the N-glycan precursor (*P*-value = 0.005; Reactome) were found to be significantly enriched among the phospholipid related loci. Considering the sphingolipid associated loci, the same analysis

implicated the sphingolipid metabolism (*P*-value =1.0 × 10^{-5}; Reactome), metabolism of lipids and lipoproteins (*P*-value = 1.0 × 10^{-5}; Reactome), and LPA receptor mediated events (*P*-value = 0.002; PID) pathways. These analyses suggested that, among genes from the same locus, *SRD5A1* is a more likely candidate than *PAPD7* and *LPAR2* is a more likely candidate than neighbouring *ZNF101* and *ATP13A1* (Supplementary Tables 5 and 6).

Supplementary Figure 4 places all of the nearest, or most likely, genes from genome-wide significant and suggestive loci in the Ingenuity glycerophospholipid metabolism pathway [29]. Of the 25 loci associated with phospholipids at a nominal *P*-value < 5 × 10^{-8}, 13 genes *(KCNH7, AGPAT1, PNLIPRP2, SYT9, FADS2, DAGLA, DLG2, APOA1, APOC3, ELOVL2, CDK17, LIPC* and *PLA2G10)* from 11 loci can be mapped to the glycerophospholipid metabolism pathway; among the 10 loci associated with sphingomyelins or ceramides, 6 genes (*FADS2, DAGLA, PLD2, LASS4, APOE, APOC2*) from 4 loci can be mapped to the same pathway (Supplementary Figure 4). Supplementary Figure 5 maps the same genes onto the Ingenuity sphingolipid metabolism pathway. Of the 10 sphingomyelin or ceramide loci, 9 genes from 5 loci (*FADS1, FADS2, C11orf10, SGPP1, APOE, APOC1, APOC2, ,LASS4*, and *PLD2*) can be placed in this pathway, as was the case for 12 genes from 8 loci implicated in phospholipids (*ILKAP, ITGA9, AGPAT1, FADS1, FADS2, C11orf10, APOA1, APOA5, APOC3, PCDH20, LIPC*, and *PDXDC1*).

Association with IMT, T2DM, and CAD risk

The top 35 SNPs were assessed for association with IMT, T2DM, and CAD using the GWAS results from the CHARGE [79], DIAGRAM [80] and CARDIoGRAM [81] consortia, respectively. For IMT, we observed a significant association (*P*-value = 7 × 10^{-4}) with the *FADS1-2-3* locus SNP rs102275 (Supplementary Table 7). rs1061808, located in the *HLA* region on chromosome 6, and two SNPs from the *FADS1-2-3* region (rs174479 and rs102275) were associated with T2DM risk (nominal *P*-value < 0.05) (Supplementary Table 8). rs964184 from the *APOA1-5* region was previously reported to be associated with CAD risk (*P*-value = 8.02 × 10^{-10}) by the CARDIoGRAM meta-analysis study (Supplementary Table 9). For all three outcomes, the observed *P*-value distribution differed significantly from that expected under the null hypothesis (Kolmogorov Smirnov *P*-value ≤ 3.3 × 10^{-16}; Supplementary Figure 6).

DISCUSSION

This genome-wide association study of 148 phospho- and sphingolipid measurements in five European populations yielded 25 loci associated with phospholipids and 10 loci associated with sphingolipids using a nominal *P*-value of 5 ×10^{-8}. After correction for the number of independent phenotypes, the novel genome-wide significant loci included: *PAQR9, AGPAT1, PKD2L1, PDXDC1, APOE* and *PLD2*. In addition, further analysis of suggestive SNPs with lipid ratios showed significant association for an additional 3 loci *(ABDH3, PNLIPRP2,* and *PCDH20)*.

The strongest association in the *PAQR9* locus was observed between rs9832727 and the proportion of mono-unsaturated PEs, especially with the ratios PE 34:1/PE 34:2 and PE 36:1/PE 36:2. The protein coded by*PAQR9* is an integral membrane receptor and functions as receptor for the hormone adiponectin, suggesting a molecular link with obesity and T2DM [82]. However, we did not observe an association between T2DM risk and this variant.

In the *AGPAT1* locus, rs1061808 was associated with the proportion of PC 32:0, and, especially, with the ratio of PC 32:0/PC 34:1. *AGPAT1* is directly connected to phospholipid metabolism (Supplementary Figures 4 and 5), as the product of this gene converts lysophosphatidic acid (LPA) into phosphatidic acid (PA) [83]. The locus lies 400 kb distant from the *HLA-DRB1* gene which was previously associated with insulin secretion [84]. A suggestive association between rs1061808 and increased T2DM risk was observed in the DIAGRAM consortium meta-analysis results.

We found two loci that strongly influence plasma LPC levels: *PKD2L1* and *PDXDC1*. An intronic variant, rs603424 in the *PKD2L1* gene, was strongly associated with LPC 16:1. Pathway analyses suggest that another gene in the same region, *SCD* (*FADS-5*), 25 kb away, may be a better candidate since it encodes the stearoyl-CoA desaturase (delta-9-desaturase) enzyme which is involved in fatty acid desaturation. Other members of the *FADS* family are the strongest genetic regulators of phospholipid metabolism identified to date. In the *PDXDC1* locus, the strongest association was observed for intronic SNP rs4500751. This variant is 300 kb distant from *PLA2G10*, a gene that plays a major role in releasing arachidonic acid from cell membrane phospholipids [85] and the protein can be mapped to both the glycerophospholipid and the sphingolipid metabolism pathways by Ingenuity (Supplementary Figures 4 and 5). In our study, the variant was strongly associated with the ratios of 20:3 fatty acid carrying LPCs, as well as PEs, and PCs, but not with the others, suggesting a fatty-acid specific mechanism for this enzyme.

Another index SNP (rs7259004), associated with SPMs, maps to the well known *APOE* locus, which also includes three other lipid genes (*APOC1*, *APOC2* and *APOC4*). Results from the conditional analyses (Supplementary Table 4) suggest that the effect of this variant on SPM 22:0 levels is dependent on plasma LDL-C levels and that SPM 22:0 and SPM 24:0 are likely be abundant in LDL-C particles, which can also be inferred from their high phenotypic correlations with LDL-C ($r = 0.6$, *P*-value $= 2.8 \times 10^{-68}$ for SPM 22:0 and $r = 0.6$, *P*-value $= 2.8 \times 10^{-66}$.for SPM 24:0).

A second locus associated with the SPMs is *PLD2 (phospholipase D2). PLD2* catalyzes the hydrolysis of PC to produce phosphatidic acid and choline and the *PLD2* signalling pathway is involved in the destabilization of *ABCA1* and, therefore, plays role in generation of plasma HDL-C particles [86]. *PLD2*-related processes may be responsible, in part, for determining the SPM content of HDL-C. Unexpectedly, we did not observe an association between PC levels and the *PLD2* locus.

The analysis of the ratios of the phospholipids uncovered three additional associations significant at the adjusted genome-wide threshold (*P*-value $< 2.2 \times 10^{-9}$): *ABDH3, PNLIPRP2* and

PCDH20. The exact function of the *ABDH3* and *PCDH20* proteins, and how they relate to phospholipid metabolism, has not been determined. *PNLIPRP2* (pancreatic lipase-related protein 2) fulfils a key function in dietary fat absorption by hydrolyzing triglycerides into diglycerides and, subsequently, into monoglycerides and free fatty acids (Supplementary Figure 4) [87]. We found that a synonymous coding SNP (rs10885997) in *PNLIPRP2* was associated with the ratios PC 36:1/PC 34:1 and PC 36:1/PC 34:3, suggesting a fatty-acid specific turnover between these lipids.

A closer examination of the findings published by Illig et al., supports the association signals within 100 kb of loci *PDXDC1* (same SNP, *P*-value = 2.8×10^{-7}), *AGPAT1* (*P*-value = 4.9×10^{-7}), *PNLIPRP2* (*P*-value = 2.7×10^{-7}), *KLF12* (*P*-value = 5.9×10^{-7}), *ALG1* (*P*-value = 4.7×10^{-3}), *CDH8* (*P*-value = 7.6×10^{-7}), *PLD2* (*P*-value = 9.4×10^{-4}) and *ZNF600* (*P*-value = 3.3×10^{-7}) for various phospho- and sphingolipid outcomes. SNP rs603424 in *PKD2L1* was previously associated with acylcarnitine C 16:1, although this result was not replicated [73].

The significant hits from the current study were further studied for potential associations with IMT, T2DM, and CAD. For all three outcomes, the *P*-value distributions differed significantly from the expected null distribution even after exclusion of nominally significant SNPs, suggesting that some of these variants contribute to these outcomes even when they do not achieve statistical significance.

Among our top hits, rs102275 from the *FADS* cluster was associated with IMT in the CHARGE meta-analysis results [79]. This finding demonstrates the involvement of the *FADS* locus in the development of atherosclerosis.

In addition, the top SNP from the *APOA1-5* locus was implicated in CAD risk in the CARDIoGRAM study [81]. This locus, previously associated with TG levels [14], influenced two ether bound PCs and the PC/SPM ratio in our study. APOA1 and APOA2 are the predominant proteins in HDL-C particles, which also transport TG. The association between the phospholipids and rs964184 remained significant after adjustment for TG levels, suggesting that this signal is not due solely to TG mediated effects. *APOA1* is also a cofactor for lecithin cholesterol acyltransferase (LCAT) which converts cholesterol and PC to cholesteryl esters and LPC on the surface of HDL-C [88] and it is possible that the association we observe here is due to *LCAT* mediated phospholipid cleavage.

Mapping the findings into the glycerophospholipid and sphingolipid metabolism pathways uncovered several enzymes, kinases, peptidases and G-protein coupled receptors that may also be relevant for phospho- and sphingolipid metabolism. Among those involved in sphingolipid metabolism (Supplementary Figure 5), *HNF4A* (*hepatocyte nuclear factor-4*) appears to be a common interacting factor for several genes *(PCDH20, APOC1, AGPAT1, ITGA9, PLD2, C11ORF10, APOC2, GCKR, APOE, APOC3* and *LIPC)* from our GWAS. It is already known that the extinction of many hepatic functions and their expression are correlated with expression of *HNF4A* which is a candidate transcription factor for further research on lipidomics [89].

In conclusion, we identified 15 previously undescribed loci that were suggestively associated ($2.2 \times 10^{-9} < P\text{-value} < 5 \times 10^{-8}$) with phospho- and sphingolipid levels. These included interesting candidate genes such as *LPAR2*. These loci will require follow-up to definitively establish their relationship with these phenotypes. We also identified nine novel loci below the corrected genome-wide significance threshold ($P\text{-value} < 2.2 \times 10^{-9}$). These loci considerably expand our knowledge of genes/regions involved in the determination of phospho- and sphingolipid concentrations and provide interesting avenues for future research into this important topic.

ACKNOWLEDGEMENTS

For the ERF Study, we are grateful to all subjects and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection. For the MICROS, we thank the primary care practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. For the Northern Swedish Population Health Study (NSPHS) we are grateful to the district nurse Svea Hennix, Karesuando, for her dedication in performing the clinical

exams and sample collection. For the ORCADES study we would like to acknowledge the invaluable contributions of Lorraine Anderson and the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney. Genotyping of the CROATIA_Vis study in was performed by the Genetics Core of the Wellcome Trust Clinical Research Facility, WGH, Edinburgh. We thank: Professor Pavao Rudan and his staff; Professor Ariana Vorko-Jovic and her staff; Dr Branka Salzer; local general practitioners and nurses. All analyses were run on the Genetic Cluster Computer, Amsterdam (http://www.geneticcluster.org). The DIAGRAM and CARDIoGRAM Consortia members are listed in the Supplementary Methods.

REFERENCES

1. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. *Global Burden of Disease and Risk factors*; 2006.
2. Kruijshaar ME, Barendregt J, Vos T, de Graaf R, Spijker J, Andrews G. Lifetime prevalence estimates of major depression: an indirect estimation method and a quantification of recall bias. *Eur J Epidemiol.* 2005;20(1):103-111.
3. Paykel ES, Brugha T, Fryers T. Size and burden of depressive disorders in Europe. *Eur Neuropsychopharmacol.* Aug 2005;15(4):411-423.
4. Kendler KS, Karkowski LM, Prescott CA. Causal relationship between stressful life events and the onset of major depression. *Am J Psychiatry.* Jun 1999;156(6):837-841.
5. Lorant V, Deliege D, Eaton W, Robert A, Philippot P, Ansseau M. Socioeconomic inequalities in depression: a meta-analysis. *Am J Epidemiol.* Jan 15 2003;157(2):98-112.
6. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry.* Oct 2000;157(10):1552-1562.
7. Levinson DF. The genetics of depression: a review. *Biol Psychiatry.* Jul 15 2006;60(2):84-92.
8. Lopez-Leon S, Janssens AC, Gonzalez-Zuloeta Ladd AM, et al. Meta-analyses of genetic studies on major depressive disorder. *Mol Psychiatry.* Aug 2008;13(8):772-785.
9. International Schizophrenia C, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* Aug 6 2009;460(7256):748-752.
10. Lamers F, Rhebergen D, Merikangas KR, de Jonge P, Beekman AT, Penninx BW. Stability and transitions of depressive subtypes over a 2-year follow-up. *Psychol Med.* Feb 17:1-11.
11. Lopez-Leon S, Aulchenko YS, Tiemeier H, Oostra BA, van Duijn CM, Janssens AC. Shared genetic factors in the co-occurrence of symptoms of depression and cardiovascular risk factors. *J Affect Disord.* May;122(3):247-252.
12. Demirkan A, Penninx BW, Hek K, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry.* Jul;16(7):773-783.
13. Boomsma DI, Willemsen G, Sullivan PF, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet.* Mar 2008;16(3):335-342.
14. 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.
15. Lubke GH, Hottenga JJ, Walters R, et al. Estimating the Genetic Variance of Major Depressive Disorder Due to All Single Nucleotide Polymorphisms. *Biol Psychiatry.* Apr 18.
16. Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* Oct 6;478(7367):103-109.
17. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet.* Jul 2004;12(7):527-534.
18. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet.* May 2005;69(Pt 3):288-295.
19. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol.* 2007;22(11):819-829.
20. Radloff. The CES-D scale: a self report depression scale for research in the genreal population. *Appl Psychol Measurement.* 1977;3:385-401.
21. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* Jun 1983;67(6):361-370.

22. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. *Am J Epidemiol.* Sep 1977;106(3):203-214.
23. Radloff. The CES-D scale: a self report depression scale for research in theb general population. *Appl Pshycol Measurement.* 1977;3:385-401.
24. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet.* Jul 1998;63(1):259-266.
25. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* Jan 2002;30(1):97-101.
26. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet.* May 2000;25(1):12-13.
27. Kirichenko AV, Belonogova NM, Aulchenko YS, Axenovich TI. PedStr software for cutting large pedigrees for haplotyping, IBD computation and multipoint linkage analysis. *Ann Hum Genet.* Sep 2009;73(Pt 5):527-531.
28. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* May 1 2003;19(7):889-890.
29. Jimenez-Marin A, Collado-Romero M, Ramirez-Boo M, Arce C, Garrido JJ. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proc.* 2009;3 Suppl 4:S6.
30. Hernandez DG, Nalls MA, Moore M, et al. Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol Dis.* Jul;47(1):20-28.
31. Gibbs JR, van der Brug MP, Hernandez DG, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* May;6(5):e1000952.
32. Li Y, Ding J, Abecasis G. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. Paper presented at: American Society of Human Genetics; 9–13 October 2006, 2006; New Orleans, Louisiana.
33. Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive evaluation of SNP genotype imputation. *Hum Genet.* Mar 2009;125(2):163-171.
34. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* Aug 5 2003;100(16):9440-9445.
35. Trabzuni D, Wray S, Vandrovcova J, et al. MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies. *Hum Mol Genet.* Jul 18.
36. Trabzuni D, Ryten M, Walker R, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem.* Oct;119(2):275-282.
37. Middeldorp CM, de Geus EJ, Willemsen G, Hottenga JJ, Slagboom PE, Boomsma DI. The serotonin transporter gene length polymorphism (5-HTTLPR) and life events: no evidence for an interaction effect on neuroticism and anxious depressive symptoms. *Twin Res Hum Genet.* Dec;13(6):544-549.
38. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* Jul 15 2009;25(14):1754-1760.
39. Brouwer RW, van den Hout MC, Grosveld FG, van Ijcken WF. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics.* Jan 15;28(2):284-285.
40. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* Sep;20(9):1297-1303.
41. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* Jul 15;89(1):82-93.
42. Aulchenko YS, de Koning DJ, Haley C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics.* Sep 2007;177(1):577-585.

43. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* May 15 2007;23(10):1294-1296.
44. Chumakov I, Blumenfeld M, Guerassimenko O, et al. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl Acad Sci U S A.* Oct 15 2002;99(21):13675-13680.
45. Li S, Itoh M, Ohta K, et al. The expression and localization of Prune2 mRNA in the central nervous system. *Neurosci Lett.* Oct 10;503(3):208-214.
46. Potkin SG, Guffanti G, Lakatos A, et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS One.* 2009;4(8):e6501.
47. Yamamoto H, Kokame K, Okuda T, Nakajo Y, Yanamoto H, Miyata T. NDRG4 protein-deficient mice exhibit spatial learning deficits and vulnerabilities to cerebral ischemia. *J Biol Chem.* Jul 22;286(29):26158-26165.
48. Paulusma CC, Oude Elferink RP. The type 4 subfamily of P-type ATPases, putative aminophospholipid translocases with a role in human disease. *Biochim Biophys Acta.* Jun 30 2005;1741(1-2):11-24.
49. Vollenweider I, Smith KS, Keist R, Rudolph U. Antidepressant-like properties of alpha2-containing GABA(A) receptors. *Behav Brain Res.* Feb 2;217(1):77-80.
50. Kertes DA, Kalsi G, Prescott CA, et al. Neurotransmitter and neuromodulator genes associated with a history of depressive symptoms in individuals with alcohol dependence. *Alcohol Clin Exp Res.* Mar;35(3):496-505.
51. Dan B, Boyd SG. Angelman syndrome reviewed from a neurophysiological perspective. The UBE3A-GABRB3 hypothesis. *Neuropediatrics.* Aug 2003;34(4):169-176.
52. Seal JL, Gornick MC, Gogtay N, et al. Segmental uniparental isodisomy on 5q32-qter in a patient with childhood-onset schizophrenia. *J Med Genet.* Nov 2006;43(11):887-892.
53. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* Apr 3.
54. Chesler EJ, Wang J, Lu L, Qu Y, Manly KF, Williams RW. Genetic correlates of gene expression in recombinant inbred strains: a relational model system to explore neurobehavioral phenotypes. *Neuroinformatics.* 2003;1(4):343-357.
55. Holthuis JC, Pomorski T, Raggers RJ, Sprong H, Van Meer G. The organizing potential of sphingolipids in intracellular membrane transport. *Physiol Rev.* Oct 2001;81(4):1689-1723.
56. Lajoie P, Goetz JG, Dennis JW, Nabi IR. Lattices, rafts, and scaffolds: domain regulation of receptor signaling at the plasma membrane. *J Cell Biol.* May 4 2009;185(3):381-385.
57. Merrill AS, K. Sphingolipids: metabolism and cell signaling. In: Vance DV, JE, ed. *Biochemistry of Lipids, Lipoproteins and Membranes*; 2002:373-407.
58. Bakovic M, Fullerton MD, Michel V. Metabolic and molecular aspects of ethanolamine phospholipid biosynthesis: the role of CTP:phosphoethanolamine cytidylyltransferase (Pcyt2). *Biochem Cell Biol.* Jun 2007;85(3):283-300.
59. Poli G, Leonarduzzi G, Biasi F, Chiarpotto E. Oxidative stress and cell signalling. *Curr Med Chem.* May 2004;11(9):1163-1182.
60. van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol.* Feb 2008;9(2):112-124.
61. Zheng W, Kollmeyer J, Symolon H, et al. Ceramides and other bioactive sphingolipid backbones in health and disease: lipidomic analysis, metabolism and roles in membrane structure, dynamics, signaling and autophagy. *Biochim Biophys Acta.* Dec 2006;1758(12):1864-1884.

62. Broekman MJ, Handin RI, Derksen A, Cohen P. Distribution of phospholipids, fatty acids, and platelet factor 3 activity among subcellular fractions of human platelets. *Blood.* Jun 1976;47(6):963-971.
63. Engelmann B, Kogl C, Kulschar R, Schaipp B. Transfer of phosphatidylcholine, phosphatidylethanolamine and sphingomyelin from low- and high-density lipoprotein to human platelets. *Biochem J.* May 1 1996;315 (Pt 3):781-789.
64. Brugger B, Erben G, Sandhoff R, Wieland FT, Lehmann WD. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc Natl Acad Sci U S A.* Mar 18 1997;94(6):2339-2344.
65. Hodge AM, English DR, O'Dea K, et al. Plasma phospholipid and dietary fatty acids as predictors of type 2 diabetes: interpreting the role of linoleic acid. *Am J Clin Nutr.* Jul 2007;86(1):189-197.
66. 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.* Apr 2008;43(4):289-299.
67. Mannheim D, Herrmann J, Versari D, et al. Enhanced expression of Lp-PLA2 and lysophosphatidylcholine in symptomatic carotid atherosclerotic plaques. *Stroke.* May 2008;39(5):1448-1455.
68. Matsumoto T, Kobayashi T, Kamata K. Role of lysophosphatidylcholine (LPC) in atherosclerosis. *Curr Med Chem.* 2007;14(30):3209-3220.
69. Wang L, Folsom AR, Zheng ZJ, Pankow JS, Eckfeldt JH. Plasma fatty acid composition and incidence of diabetes in middle-aged adults: the Atherosclerosis Risk in Communities (ARIC) Study. *Am J Clin Nutr.* Jul 2003;78(1):91-98.
70. Farooqui AA, Horrocks LA, Farooqui T. Glycerophospholipids in brain: their metabolism, incorporation into membranes, functions, and involvement in neurological disorders. *Chem Phys Lipids.* Jun 2000;106(1):1-29.
71. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* Nov 2008;4(11):e1000282.
72. Hicks AA, Pramstaller PP, Johansson A, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. *PLoS Genet.* Oct 2009;5(10):e1000672.
73. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet.* Feb 2009;42(2):137-141.
74. Liebisch G, Drobnik W, Lieser B, Schmitz G. High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clin Chem.* Dec 2002;48(12):2217-2224.
75. Liebisch G, Lieser B, Rathenberg J, Drobnik W, Schmitz G. High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochim Biophys Acta.* Nov 8 2004;1686(1-2):108-117.
76. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol.* Dec 2010;34(8):816-834.
77. Willer C, Li Y, Abecasis G. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* Sep 1 2010;26(17):2190-2191.
78. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB--a database for integrating human functional interaction networks. *Nucleic Acids Res.* Jan 2009;37(Database issue):D623-628.
79. Bis JC, Kavousi M, Franceschini N, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet.* 2011;43(10):940-947.
80. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet.* Jul 2010;42(7):579-589.
81. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* Apr 2011;43(4):333-338.

82. Tang YT, Hu T, Arterburn M, et al. PAQR proteins: a novel membrane receptor family defined by an ancient 7-transmembrane pass motif. *J Mol Evol.* Sep 2005;61(3):372-380.
83. Aguado B, Campbell RD. Characterization of a human lysophosphatidic acid acyltransferase that is encoded by a gene located in the class III region of the human major histocompatibility complex. *J Biol Chem.* Feb 13 1998;273(7):4096-4105.
84. Howson JM, Rosinger S, Smyth DJ, Boehm BO, Todd JA. Genetic Analysis of Adult-Onset Autoimmune Diabetes. *Diabetes.* Oct 2011;60(10):2645-2653.
85. Singh DK, Subbaiah PV. Modulation of the activity and arachidonic acid selectivity of group X secretory phospholipase A2 by sphingolipids. *J Lipid Res.* Mar 2007;48(3):683-692.
86. Wang Y, Oram JF. Unsaturated fatty acids phosphorylate and destabilize ABCA1 through a phospholipase D2 pathway. *J Biol Chem.* Oct 28 2005;280(43):35896-35903.
87. Giller T, Buchwald P, Blum-Kaelin D, Hunziker W. Two novel human pancreatic lipase related proteins, hPLRP1 and hPLRP2. Differences in colipase dependence and in lipase activity. *J Biol Chem.* Aug 15 1992;267(23):16509-16516.
88. Breslow JL, Ross D, McPherson J, et al. Isolation and characterization of cDNA clones for human apolipoprotein A-I. *Proc Natl Acad Sci U S A.* Nov 1982;79(22):6861-6865.
89. Chartier FL, Bossu JP, Laudet V, Fruchart JC, Laine B. Cloning and sequencing of cDNAs encoding the human hepatocyte nuclear factor 4 indicate the presence of two isoforms in human liver. *Gene.* Sep 30 1994;147(2):269-272.

5

Depressive symptoms and circulating lipids

Chapter 5

Plasma phosphatidylcholine and sphingomyelin concentrations are associated with depression and anxiety symptoms in a Dutch family-based lipidomics study.

Ayşe Demirkan, Aaron Isaacs, Peter Ugocsai, Gerhard Liebisch, Maksim Struchalin, A Cecile JW Janssens, Gerd Schmitz, Ben A Oostra, Cornelia M van Duijn.

(submitted)

ABSTRACT

The central nervous system has the second highest concentration of lipids after adipose tissue. Alterations in neural membrane phospho- and sphingolipid composition can influence crucial intra- and intercellular signalling and alter the membrane's properties. Recently, the polyunsaturated fatty acids (PUFA) hypothesis for depression suggests that phospho- and sphingolipid metabolism includes potential pathways for the disease. In 742 people from a Dutch family-based study, we assessed the relationships between 148 different plasma phospho- and sphingolipid species and depression/anxiety symptoms as measured by the Hospital Anxiety and Depression Scales (HADS-A and HADS-D) and the Center for Epidemiological Studies Depression Scale (CES-D). We observed significant differences in plasma sphingomyelins (SPM), particularly the SPM 23:1 / SPM 16:0 ratio, which was inversely correlated with depressive symptom scores. We observed a similar trend for plasma phosphatidylcholines (PC), particularly the molar proportion of PC O 36:4 and its ratio to ceramide CER 20:0. Absolute levels of PC O 36:4 were also associated with depression symptoms in an independent replication. To our knowledge this is the first study on depressive symptoms that focuses on specific phospho- and sphingolipid molecules in plasma rather than total PUFA concentrations. The findings of this lipidomic study suggests that plasma sphingomyelins and ether phospholipids should be further studied for their potential as biomarkers and for a better understanding of the underlying mechanisms of this systemic disease.

INTRODUCTION

Lipids make up roughly 50% of the brain's dry weight, making it the organ with the second highest lipid content next to adipose tissue[1,2]. It is known that alterations in neural membrane lipid components can influence crucial intracellular and intercellular signalling and alter membrane physical properties such as fluidity, phase transition temperature, bilayer thickness, and lateral domains[3]. Several psychiatric disorders have been related to disturbed lipids metabolism in neuronal tissue, in particular fatty acid and phospho- and sphingolipid metabolism[4-9]. Polyunsaturated fatty acids (PUFAs) comprise 30% of brain fatty acid mass and determine the properties of the phospholipid bilayer. Their synthesis in brain from shorter chain precursors is limited and some speculate that plasma is the PUFA source for brain[10]. How PUFAs are transported to brain and pass the blood brain barrier is unclear. Suggested mechanisms are most likely PUFA specific (e.g. ω-3 or-ω-6) and include diffusion of free fatty acids, fatty acid transport proteins, and, finally, by lipoproteins[10, 11].

The nervous system of mammals is also rich in species with ceramide backbones, including sphingomyelin (SPM)[12] and glycosphingolipids[13]. Sphingolipids naturally segregate laterally within membranes and they form membrane microdomains (or lipid rafts) that are also enriched with cholesterol. The conversion of SPM to ceramide (CER) leads to the formation of

ceramide-rich lipid rafts, thereby altering the membrane dynamics[14-17]. Plasma concentrations of these phospho- and sphingolipids may directly or indirectly relate to concentrations or turnover in brain[18].

The evidence of association between PUFA (ω-3 and-ω-6) and depressive disorder in different sources (serum, plasma, serum phospholipids) is inconsistent. A recent meta-analysis demonstrated a significant decrease in total ω-3, but not ω-6, PUFAs among depressed patients[19]. However, various PUFAs play roles in different cascades when incorporated in phospho- and sphingolipids. Further, these lipids are metabolized by different fatty acid specific enzymes raising the question whether specific plasma phospho- and sphingolipid species are implicated in depression[20]. Using a lipidomics approach examining 148 different plasma phospho- and sphingolipid species, we searched for association between these lipids and symptoms of depression and anxiety in the Erasmus Rucphen Family (ERF) study.

MATERIAL & METHODS

Study population

ERF is a family-based study which includes over 3000 individuals descendant from 22 couples living in the Rucphen region in the southwest Netherlands in the 19th century[21]. Symptoms of depression and anxiety were assessed using the Hospital Anxiety and Depression Scale (HADS-D for depression and HADS-A for symptoms of anxiety) and the Center for Epidemiological Studies Depression Scale (CES-D) questionnaires[22, 23].

A broad range of phospho- and sphingolipid species (n = 148) were measured in 820 participants. Lipid species were quantified by electrospray ionization tandem mass spectrometry (ESIMS/ MS) using methods validated and described previously[24-27]. Correction of isotopic overlap of lipid species as well as data analysis was performed by self-programmed Excel macros for all lipid classes according to the principles described previously[26]. Further details can be found in the supplementary methods. Additionally, in an independent sample of 1000 ERF individuals, targeted metabolite profiling by electrospray ionization (ESI) tandem mass spectrometry (MS/MS) was performed using a quantitative metabolomics platform (*Biocrates Life Sciences AG*, Austria). 753 individuals with both Biocrates measurements and depression questionnaire data served as a partial replication dataset for species measured in both platforms.

In addition to the measured lipids, the proportions of each lipid in its own class, (for example, the proportion of SPM 23:1 among total SPM) were calculated. These proportions are valuable in assessing differences in concentrations that are related to within-class turnover. Partial correlation analysis, corrected for age and sex, were performed using R (http://www.r-project.org/). Significant associations between the HADS-D, HADS-A and CES-D scales and lipids were further assessed with the SOLAR (Sequential Oligogenic Linkage Analysis Routines) 4.1.5 software package (Southwest Foundation for Biomedical Research, San Antonio, Texas, USA)[28] using

the "polygenic" option to adjust for relatedness. Due to computational limitations, the whole pedigree was split into non-overlapping sub pedigrees using the PedCut program[29,30]. Many of the lipid levels measured in our study are highly correlated with each other. For this reason, we used a data reduction strategy to estimate the number of independent observations. In total, 23 principal components accounted for 79% of the phenotypic variance of all lipids. As we studied 3 outcomes for depression and anxiety, we defined the threshold for statistical significance as 0.05 / (23 × 3), i.e. 7.25×10^{-4} for the single species and as 0.05 / (((23 × 23) - 23) /2); i.e. 6.59×10^{-5} for the lipid ratio matrix. Final stage meta-analysis of the discovery and replication set were performed using the correlation coefficients (*r*) and sample size, through Fisher's r to Z transformation[31].

RESULTS

Table 1 shows general characteristics of the study populations. In general, individuals in the discovery set were older, more likely to be women, had higher mean levels of depression scales and had lower education levels, which we think is a measure of socio-economic status. Phospho- and sphingolipids measured in the ERF discovery sample, and also their proportions in their own class, were tested for association with depression and anxiety symptoms, as measured by HADS-D, HADS-A and CES-D. This effort yielded correlations ranging between -0.14 and 0.10 (see Supplementary Table 1). After correcting for the estimated number of independent associations tested, two lipids were significantly correlated (*P-value* < 7.25×10^{-4}), which are independent of each other when jointly entered into a regression model. The most significant correlation was observed with %SPM 23:1 (r = -0.14, *P-value* = 1.3×10^{-4}, for the HADS-D scale). The same lipid was also suggestively associated, in the same direction, with HADS-A (r = -0.12, *P-value* = 7.6×10^{-4}) and CES-D (r = -0.11, *P-value* = 2.2×10^{-3}). The other significant association was observed with both the absolute value and percentage of PC O 36:4. The percentage of this lipid in its class significantly correlates with HADS-A (r = -0.14, *P-value* = 1.7×10^{-4}), but also suggestively in the same direction to the depression scales (r = -0.11, *P-value* = 3.4×10^{-3} for HADS-D and r = -0.11, *P-value* = 2.7×10^{-3} for CES-D). Similar findings were also observed for absolute PC O 36:4 levels (r = -0.13, *P-value* = 3.4×10^{-4} for HADS-A; r = -0.12, *P-value* = 7.5×10^{-4} for HADS-D; and r = -0.9, *P-value* = 0.01 for CES-D). When adjusting for relatedness, all of these associations remained significant (Table 2). Results from all partial correlation tests are presented in Supplementary Table 1.

We additionally analyzed the lipid / lipid ratios, which may be informative for the molecular turnover between species. Figure 1 shows the lipid / lipid ratios as a function of their correlation with the HADS-D scale. Supplementary Table 2 shows the findings from lipid / lipid ratio correlation analyses with *P-value* < 0.001. We observed that an increase in the SPM 23:1 / SPM 16:0 ratio was associated with decreased depression symptoms, measured by HADS-D (r = -0.16, *P-value* = 1.04×10^{-5},Supplementary Figure 1A), and also suggestively with the other scales (r =

TABLE 1 Descriptives of the study population.

	Discovery	Replication	
	Mean (SD)	Mean (SD)	*P*-value*
HADS-A	7.31 (4.66)	6.39 (4.21)	*< 0.001*
HADS-D	6.68 (4.55)	5.69 (4.05)	*< 0.001*
CES-D	11.65 (9.72)	9.96 (9.18)	*< 0.001*
Age, years	53.09 (14.97)	48.13 (14.00)	*< 0.001*
LDL-C, mmol/L	3.78 (0.98)	3.70 (1.02)	0.09
HDL-C, mmol/L	1.28 (0.35)	1.26 (0.37)	0.27
TG, mmol/L	1.42 (0.86)	1.35 (0.81)	0.11
TC, mmol/L	5.66 (1.08)	5.55 (1.12)	*0.03*
Women (%)	447 (60)	392 (52)	*0.002*
Anti-depressant use (%)	57 (7.6)	49 (6.5)	0.43
Education (%)			
Elementary	259 (36.7)	207 (28.6)	*< 0.001*
Intermediate	422 (59.9)	468 (64.7)	
Higher	23 (3.2)	48 (6.6)	

Means and standard deviations (SD) are given for the continuous measurements and percentage is given for gender. TC; total cholesterol, HDL-C, high density lipoprotein cholesterol, LDL-C; low density lipoprotein cholesterol, TG; triglycerides. *Independent samples t-test was used to test for the difference of means for continuous variables, Pearson's chi^2 test was used for the categorical variables.

TABLE 2. Lipids correlated with depression and anxiety symptoms

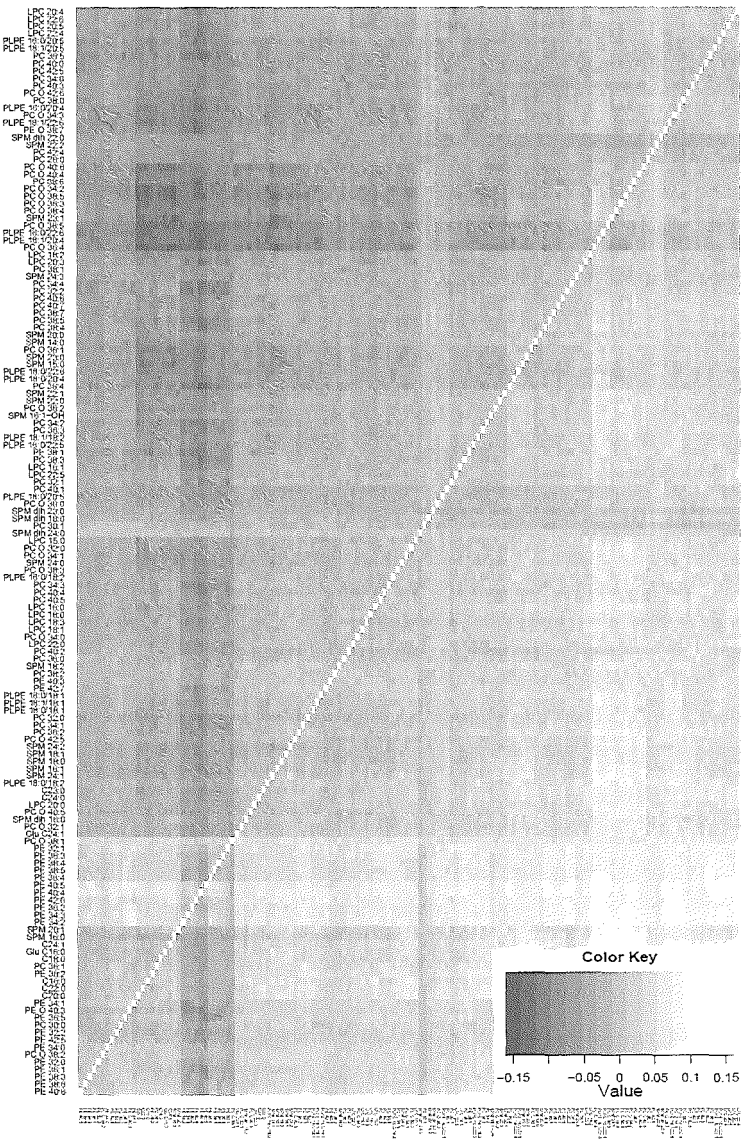
	HADS-D			CES-D			HADS-A		
	r	P-value	P-value*	r	P-value	P-value*	r	P-value	P-value*
%SPM 23:1	**-0.14**	**1.32×10^{-4}**	8.00×10^{-4}	-0.11	2.20×10^{-3}	1.40×10^{-3}	-0.12	7.55×10^{-4}	8.00×10^{-4}
PC O 36:4	-0.12	7.51×10^{-4}	3.27×10^{-2}	-0.09	1.13×10^{-2}	9.48×10^{-3}	**-0.13**	**3.37×10^{-4}**	1.12×10^{-3}
%PC O 36:4	-0.11	3.39×10^{-3}	2.04×10^{-2}	-0.11	2.65×10^{-3}	1.27×10^{-2}	**-0.14**	**1.69×10^{-4}**	5.39×10^{-3}

Significant findings after correction for multiple testing are shown in bold. *r*, correlation coefficient, *P-value*, nominal p-value, * *P-value* when adjusted for relatedness.

-0.13, *P-value* = 3.0×10^{-4} for HADS-A, and *r* = -0.12, *P-value* = 1.2×10^{-3} for CES-D). This finding is in line with the association with the %SPM 23:1 seen in Table 2. Similarly, the ratio of PC O 36:4 to CER 20:0 (*r* = -0.16, *P-value* = 1.08×10^{-5}, Supplementary Figure 1B) inversely correlates with depression as measured by HADS-D and was also associated in the same direction with the HADS-A and CES-D scales (*P-value* < 0.01).

In general, the ratios of some ether phospholipid species, especially those likely to carry arachidonic acid (such as PC O 36:4, PC O 36:5, PC O 38:4, PC O 38:5) and its precursor linoleic acid (PC O 34:2, PC O 36:3, PLPE 18:1/18:2) to some CERs, and acyl-acyl PUFA phospholipids (such as PE 40:4, PE 40:5, PE 40:6, PE 34, PE 36:2, PC 36:2), showed a trend of inverse correlation with depression (Figure 1). Among those ratios, SPM 23:1 / SPM 16:0, PC O 36:4 / CER 20:0, PC O 36:4 / SPM 16:0, PC O 36:4 / PE 40:4, PC O 36:4 / CER 22:0, PC O 38:5 / SPM 16:0 and PC O 38:5

FIGURE 1. Correlation between lipid/lipid ratios and HADS-D.



In the figure, the names on the y axis show the lipid in the numerator and the ones on the x axis show the denominator of each lipid fraction analyzed and plotted as a function of the correlation coefficient.

/ CER 20:0 for HADS-D and PC O 36:4/ PC 36:2 for HADS-A remained significant after correction for multiple testing (*P-value* < 6.59×10^{-5}), as shown in Supplementary Table 2.

In a second, independent dataset drawn from the ERF population, plasma levels of PC O 36:4 also showed suggestive evidence of association with depression and anxiety complaints with

the same direction of effect (r = -0.07, *P-value* = 0.040, for CES-D, r = -0.07, *P-value* = 0.052 for HADS-A), however, the significance with HADS-D did not reach the nominal cut-off of 0.05 (r = -0.03, *P-value* = 0.38). Meta analysis of the two samples (n = 1495) suggested association with all three scales (*P-value* = 0.001 for CES-D, *P-value* = 9.8×10^{-5} for HADS-A and *P-value* = 0.002 for HADS-D). Among the available lipid ratios in the replication sample (PC O 36:4 / SPM 16:0, PC O 38:5/ SPM 16:0 and PC O 36:4 / PC 36:2), PC O 36:4 / SPM 16:0 was also associated with CES-D, but not with the other scales (r =-0.07 *P-value* = 0.048 for CES-D, r =-0.05 *P-value* = 0.19 for HADS-D and r =-0.06 *P-value* = 0.09 for HADS-A).

DISCUSSION

These analyses showed that there is an inverse correlation between depression and anxiety symptoms and both PC O 36:4 and SPM 23:1. In an independent sample, the absolute concentrations of PC O 36:4, and its ratio to CER 20:0, were also nominally associated with CES-D. Combining the two samples yielded a highly significant and consistent correlation with the depression and anxiety scale with the *P-value* for the HADS-A scale improving (meta-analysis *P-value* = 9.8×10^{-5}).

PC O 36:4 is an alkyl-acyl PC (ether phospholipid) and generally includes an ether bound 16:0 fatty acid in the *sn-1* position and an ester bound arachidonic acid (20:4) in the *sn-2* position instead of ester bound fatty acids in both the *sn-1* and *sn-2* positions. Ether lipids constitute a very small fraction of the plasma phospholipids (~2% of total PC, predominantly PC O 36:4), but they have higher proportions among the brain phospholipids, especially in the form of PLPEs[6, 32]. PC O 36:4 is a potential target for phospholipases A2 (PLA2, EC 3.1.1.4), a superfamily of enzymes that specifically hydrolyze acyl (ester) bonds in the *sn-2* position, producing free fatty acids and LPCs. Some PLA2 isoforms were also found to exhibit a selectivity for ether lipids in purified protein extracts[33]. In the case of PC O 36:4, this fatty acid is arachidonic acid, which is then rapidly converted into inflammatory mediators, prostaglandins and leukotriens, but also directly modulates neural cell function by various mechanisms, such as altering the fluidity and polarization state of membranes, activating protein kinase C, stimulating calcium release, activating several other enzymes and regulating gene transcription[34]. Lithium, which is mostly used for bipolar depressive disorder (BPD) but is also used for treatment of unipolar depression[35] is thought to inhibit the over-activity of arachidonic acid-specific PLA2 in the brain in vivo[36, 37]. Altered expression of PLA2 was already shown in the cortex of post-mortem BPD patients[38].

In addition to PC O 36:4, we observed that the PC O 36:4 / CER 20:0 ratio inversely correlates with depression scores. It has been shown that ceramide increases sPLA2 (secretory PLA2) activity and alters the fatty acid specificity towards the arachidonyl at the *sn-2* position[39]. The PLA2 family includes vinyl ether phospholipid specific enzymes, specialized in arachidonic acid release from plasmalogens. However, the one responsible for the cleavage of alkyl-acyl

PC remains unknown. PLA2, it has been suggested, works in concert with another enzyme, CoA-independent transacylase, resulting in platelet activating factor (PAF) [40]. PAF is another tightly controlled potent inflammatory mediator and involved in neural cell migration, gene expression, calcium mobilization, and long-term potentiation in brain[41, 42].

We also observed a negative correlation between SPM 23:1 and depression scale scores, and, particularly, the ratio of SPM 23:1 to SPM 16:0. Previously, increased activity of the lipid metabolizing enzyme acid sphingomyelinase, which degragades SPM to phosphocholine and ceramide, was found among depressed patients, compared to healthy controls [4]. It was also suggested that antidepressant drugs induce a dose dependent decrease in acid spingomyelinase activity[5, 43]. Our findings are independent of the use of anti-depressive medication. However, the number of persons taking medication is relatively small (7.6 %). Our analysis suggests a mechanism involving sphingomyelin degradation in a fatty acid specific way. Of note, we also observed that HDL-C is inversely correlated with the depression scales (r = -0.06, *P-value* = 0.003, for HADS-D, r = -0.04, *P-value* = 0.042 for CES-D), but not with the anxiety scale. PC O 36:4 / CER 20:0 is strongly correlated with HDL-C levels (r = 0.30, *P-value* = 7.45^{-17}), but SPM 23:1 / SPM 16:0 is not.

Human studies to date have noted the importance of essential ω-3 fatty acids (20:5 ω-3 and 22:6 ω-3) for several neuropsychiatric traits, including depression, and focused on the ω-3/ ω-6 ratio in plasma, serum or brain tissue to understand their relationship with depressive pathology[19, 44]. To our knowledge, there is no study with which to directly compare our lipidomics findings that examined a wide range of phospholipids in a hypothesis free approach. Our results show a disruption in the distribution of PUFAs among phospho- and sphingolipids, apparently affecting alkyl-acyl PCs and SPMs; SPM 23:1, PC O 36:4, CER 20:0, SPM 16:0, PE 40:4, CER 22:0, PC O 38:5 and CER 20:0 in relation to depression pathogenesis. Of these phospholipids, the genetic origin of ether PUFA PCs is partly known[45] (e.g. PLEKHH1, a protein with unknown function but highly expressed in the brain and cortex) opening avenues for animal models and intervention studies.

One limitation of the study is that the measurement of the lipid species can not give exact information on the fatty acid chains involved and, for PC species, can not exclude the possibility that the annotation is not precise; in such case the mass of PC O 36:4 also fits to PC 35:4 (see Materials and Methods). Another general issue is the small difference in the baseline characteristics of the two study samples, such as age, education levels and depression scores, which may be due to random fluctuation but may have affected the power of our study. We did not observe a difference in the mean levels of PC O 36:4; however that does not exclude the fact that lipidomics measurement platforms between the discovery and replication sets were different.

The small changes we observe in human plasma might be a reflection of overactive ether lipid cleavage/turnover in brain and is likely to relate to the inflammatory path of depression pathogenesis, and may also provide a clue regarding the comorbidity of depression and

cardiovascular disease. Our study is unique in terms of the specific phospho- and sphingolipid molecules we studied with respect to depression and our findings suggest that focusing on fatty acid turnover mechanisms in phospho- and sphingolipid metabolism may lead us to a better understanding of the aetiology of depressive disease.

ACKNOWLEDGEMENTS

ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 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). High-throughput genome-wide association analysis of the data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). Lipidomic analysis was supported by the European Commission FP7 grant LipidomicNet (2007-202272) and by the BMBF network project "Systems Biology Consortium on Metabotypes (SysMBo) The ERF study was supported by grants from the Netherlands Organisation for Scientific Research (Pionier, 047.016.009, 047.017.043), Erasmus MC and the Centre for Medical Systems Biology (CMSB; National Genomics Initiative). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.

REFERENCES

1. Morrel P, Toews, A.D. Biochemistry of lipids. In: H.W. M, ed. *Handbook of Clinical Neurology.* Vol 22. Amsterdam: Elsevier; 1993:33-49
2. Watkins PA, Hamilton JA, Leaf A, et al. Brain uptake and utilization of fatty acids: applications to peroxisomal biogenesis diseases. *J Mol Neurosci.* Apr-Jun 2001;16(2-3):87-92; discussion 151-157.
3. Horrocks LA, Farooqui AA. Docosahexaenoic acid in the diet: its importance in maintenance and restoration of neural membrane function. *Prostaglandins Leukot Essent Fatty Acids.* Apr 2004;70(4):361-372.
4. Kornhuber J, Medlin A, Bleich S, et al. High activity of acid sphingomyelinase in major depression. *J Neural Transm.* Nov 2005;112(11):1583-1590.
5. Kornhuber J, Reichel M, Tripal P, et al. The role of ceramide in major depressive disorder. *Eur Arch Psychiatry Clin Neurosci.* Nov 2009;259 Suppl 2:S199-204.
6. Farooqui AA, Horrocks LA, Farooqui T. Glycerophospholipids in brain: their metabolism, incorporation into membranes, functions, and involvement in neurological disorders. *Chem Phys Lipids.* Jun 2000;106(1):1-29.
7. Horrobin DF. Phospholipid metabolism and depression: the possible roles of phospholipase A2 and coenzyme A-independent transacylase. *Hum Psychopharmacol.* Jan 2001;16(1):45-52.
8. Horrobin DF, Bennett CN. Depression and bipolar disorder: relationships to impaired fatty acid and phospholipid metabolism and to diabetes, cardiovascular disease, immunological abnormalities, cancer, ageing and osteoporosis. Possible candidate genes. *Prostaglandins Leukot Essent Fatty Acids.* Apr 1999;60(4):217-234.
9. Su KP, Shen WW, Huang SY. Effects of polyunsaturated fatty acids on psychiatric disorders. *Am J Clin Nutr.* Nov 2000;72(5):1241.
10. Spector AA. Plasma free fatty acid and lipoproteins as sources of polyunsaturated fatty acid for the brain. *J Mol Neurosci.* Apr-Jun 2001;16(2-3):159-165; discussion 215-121.
11. Mitchell RW, Hatch GM. Fatty acid transport into the brain: of fatty acid fables and lipid tails. *Prostaglandins Leukot Essent Fatty Acids.* Nov;85(5):293-302.
12. Cutler RG, Mattson MP. Sphingomyelin and ceramide as regulators of development and lifespan. *Mech Ageing Dev.* Jul 15 2001;122(9):895-908.
13. Tettamanti G. Ganglioside/glycosphingolipid turnover: new concepts. *Glycoconj J.* 2004;20(5):301-317.
14. Goni FM, Alonso A. Effects of ceramide and other simple sphingolipids on membrane lateral structure. *Biochim Biophys Acta.* Jan 2009;1788(1):169-177.
15. van Blitterswijk WJ, van der Luit AH, Veldman RJ, Verheij M, Borst J. Ceramide: second messenger or modulator of membrane structure and dynamics? *Biochem J.* Jan 15 2003;369(Pt 2):199-211.
16. Zhang Y, Li X, Becker KA, Gulbins E. Ceramide-enriched membrane domains--structure and function. *Biochim Biophys Acta.* Jan 2009;1788(1):178-183.
17. Jenkins RW, Canals D, Hannun YA. Roles and regulation of secretory and lysosomal acid sphingomyelinase. *Cell Signal.* Jun 2009;21(6):836-846.
18. Igarashi M, Kim HW, Chang L, Ma K, Rapoport SI. Dietary n-6 polyunsaturated fatty acid deprivation increases docosahexaenoic acid metabolism in rat brain. *J Neurochem.* Mar;120(6):985-997.
19. Lin PY, Huang SY, Su KP. A meta-analytic review of polyunsaturated fatty acid compositions in patients with depression. *Biol Psychiatry.* Jul 15;68(2):140-147.
20. Engelmann B, Wiedmann MK. Cellular phospholipid uptake: flexible paths to coregulate the functions of intracellular lipids. *Biochim Biophys Acta.* Jun;1801(6):609-616.
21. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet.* Jul 2004;12(7):527-534.

22. Radloff. The CES-D scale: a self report depression scale for research in theb general population. *Appl Pshycol Measurement.* 1977;3:385-401.
23. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* Jun 1983;67(6):361-370.
24. Brugger B, Erben G, Sandhoff R, Wieland FT, Lehmann WD. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc Natl Acad Sci U S A.* Mar 18 1997;94(6):2339-2344.
25. Liebisch G, Drobnik W, Lieser B, Schmitz G. High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clin Chem.* Dec 2002;48(12):2217-2224.
26. Liebisch G, Lieser B, Rathenberg J, Drobnik W, Schmitz G. High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochim Biophys Acta.* Nov 8 2004;1686(1-2):108-117.
27. Zemski Berry KA, Murphy RC. Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. *J Am Soc Mass Spectrom.* Oct 2004;15(10):1499-1508.
28. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* May 1998;62(5):1198-1211.
29. Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet.* Jul 2008;16(7):854-860.
30. Demirkan A, Penninx BW, Hek K, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry.* Jul;16(7):773-783.
31. Fisher RA. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika (Biometrika Trust).* 1915;10(4):507–521.
32. Brites P, Waterham HR, Wanders RJ. Functions and biosynthesis of plasmalogens in health and disease. *Biochim Biophys Acta.* Mar 22 2004;1636(2-3):219-231.
33. Farooqui AA. Studies on plasmalogen-selective phospholipase A2 in brain. *Mol Neurobiol.* Jun;41(2-3):267-273.
34. Farooqui AA, Horrocks LA. Brain phospholipases A2: a perspective on the history. *Prostaglandins Leukot Essent Fatty Acids.* Sep 2004;71(3):161-169.
35. Guzzetta F, Tondo L, Centorrino F, Baldessarini RJ. Lithium treatment reduces suicide risk in recurrent major depressive disorder. *J Clin Psychiatry.* Mar 2007;68(3):380-383.
36. Chang MC, Grange E, Rabin O, Bell JM, Allen DD, Rapoport SI. Lithium decreases turnover of arachidonate in several brain phospholipids. *Neurosci Lett.* Dec 20 1996;220(3):171-174.
37. Chang MC, Jones CR. Chronic lithium treatment decreases brain phospholipase A2 activity. *Neurochem Res.* Jun 1998;23(6):887-892.
38. Kim HW, Rapoport SI, Rao JS. Altered arachidonic acid cascade enzymes in postmortem brain from bipolar disorder patients. *Mol Psychiatry.* Apr;16(4):419-428.
39. Koumanov KS, Momchilova AB, Quinn PJ, Wolf C. Ceramides increase the activity of the secretory phospholipase A2 and alter its fatty acid specificity. *Biochem J.* Apr 1 2002;363(Pt 1):45-51.
40. Chilton FH, Fonteh AN, Sung CM, et al. Inhibitors of CoA-independent transacylase block the movement of arachidonate into 1-ether-linked phospholipids of human neutrophils. *Biochemistry.* Apr 25 1995;34(16):5403-5410.
41. Farooqui AA. Lipid mediators in the neural cell nucleus: their metabolism, signaling, and association with neurological disorders. *Neuroscientist.* Aug 2009;15(4):392-407.
42. Winkler JD, Bolognese BJ, Roshak AK, Sung CM, Marshall LA. Evidence that 85 kDa phospholipase A2 is not linked to CoA-independent transacylase-mediated production of platelet-activating factor in human monocytes. *Biochim Biophys Acta.* Jun 2 1997;1346(2):173-184.

43. Kolzer M, Werth N, Sandhoff K. Interactions of acid sphingomyelinase and lipid bilayers in the presence of the tricyclic antidepressant desipramine. *FEBS Lett.* Feb 13 2004;559(1-3):96-98.
44. Tiemeier H, van Tuijl HR, Hofman A, Kiliaan AJ, Breteler MM. Plasma fatty acid composition and depression are associated in the elderly: the Rotterdam Study. *Am J Clin Nutr.* Jul 2003;78(1):40-46.
45. Demirkan A, van Duijn CM, Ugocsai P, et al. Genome-Wide Association Study Identifies Novel Loci Associated with Circulating Phospho- and Sphingolipid Concentrations. *PLoS Genet.* Feb 2012;8(2):e1002490.

6

General discussion

This thesis applies a variety of methods for understanding the genetics of two complex phenotypes: lipids and depressive symptoms. The methods utilized include polygenic score, linkage and family- based association analysis, as well as next generation sequencing and meta-analysis of genome wide association studies (GWAS), depending on the study design. The main objective of this thesis was to discover genetic variants involved in depression and circulating lipids, and to explore if it is possible to discover new endophenotypes for depression using metabolomics data. In this chapter, the main findings are summarized and methodological issues are addressed.

MEASUREMENT OF DEPRESSION BY SELF QUESTIONNAIRES

Major depressive disorder (MDD) is an outcome of clinical and social interest, with a substantial burden of pathology to society in terms of medication and therapy costs and a diminished work force. In psychiatric research, there is increasing interest in quantitative phenotypes, as the dichotomization of affected and normal is artificial. Throughout this thesis, we used quantitative phenotypes, such as self reported questionnaires, rather than depression diagnoses. These outcomes are available in various epidemiological studies. Studying depression at the population level offers the opportunity to explore large datasets, and, additionally, the possibility to include subclinical persons with a liability to depression.

Using self-questionnaires also poses disadvantages compared to the MDD diagnosis. For instance, in terms of specificity, scoring high on these questionnaires could, by definition, be an indicator of various psychiatric conditions. The questionnaires only take into account a limited period in time (approximately two weeks), which means that they only measure current depressive symptoms and do not give any information about the age of onset or recurrence. A further limitation is that persons with bipolar depression can also score high for depressive symptoms depending on their episode, which could be either depressive or manic. Another problem is that individuals on anti-depressive medication because of their MDD diagnosis may score lower than expected and may, therefore, be classified as having less symptoms for depression or as not-depressed. This implies that we have to adjust, either by imputation or by exclusion of patients. The latter will result in a reduction of power as these are the most informative subjects. However, imputation is trivial and until now it has not been used widely.

In summary, using endophenotypes for gene discovery to increase power can be a useful approach and may help us discover new candidate genetic variations associated with depressive symptoms. However, the endophenotypes also come with a price and have limitations. Ultimately, new candidates can be tested in case-control studies of MDD to study their impact on the clinical phenotype.

LIPIDOMICS MEASUREMENTS

The measurement issues in lipidomics research are different. The lipidomics traits studied in chapters 4.1, 4.2 and 5 were measured by electrospray ionization tandem mass spectrometry (ESIMS/ MS) using validated previous methods[1-4]. Although more exact than the measurement of depressive symptoms, there are two limitations of these measurements. First, for species containing double fatty acid chains, it is not possible to determine exactly the fatty acids involved. For instance, PC 36:4 can be assigned to PC 16:0/20:4, but also to PC 18:0/18:4 or PC 18:1/18:3. We know that the dominant fatty acid in plasma in the sn-1 position of PCs is 16:0, followed by 18:0 and 20:4 in the sn-2 position but we have to accept that the traits measured with this method are fractions rather than single molecules. Another imitation of the measurement of lipid traits is the possible ambiguity of the ether bond containing species. The performed analysis does not always allow an exact assignment. In our study, an "O" is added to the subspecies name, e.g. PC O 36:5 and PC O 32:1. This denotes that the two species are most likely to be assigned to PC species containing an ether bond (alkyl) and may constitute plasmalogens. However, we cannot exclude the possibility that PC O 36:5 may be assigned to PC 35:5, an unlikely odd carbon number species. Similarly, PC O 32:1 may be assigned to PC 31:1.

In chapter 5, for replicating our findings with depressive symptoms, we used an independent sample from the same population, measured with a different quantitative metabolomics platform (*Biocrates Life Sciences AG*, Austria). Sixty-two lipid traits overlapped between our initial lipidomics dataset and the Biocrates platform, clustered in PCs, LPCs and SPMs. In order to check the consistency between the two measurements, 30 people were measured by both methods from the same blood samples. We observed high correlations for 34 species ($r > 0.7$) and unexpectedly low correlations for 6 species ($r < 0.1$) species. Our lipids of interest, PC O 36:4 and PC 36:4, are dominant species in their classes and they showed high correlations ($r = 0.845$ and $r = 0.847$, respectively) between platforms. For PC O 36:4, we think the ambiguity is limited and the measurements are consistent.

HUMAN GENETIC VARIATION

We have used 3 approaches to characterize the genome: (1) Illumina linkage panel, (2) Illumina 318K, Illumina 370K and Affymetrix 250K SNP panels, and (3) exome sequencing. The linkage panel we used for the studies in this thesis is based on 6000 SNPs. This coverage is sufficient for linkage in families as relatives share large segments of DNA. On the other hand, the GWAS and meta-analysis presented here are based on imputed genotypes, according to the HapMapII CEU reference panel of 60 individuals. Imputation with this panel yields 2.5 million SNPs, which is much denser compared to our genome-wide linkage panel. Such density is needed for genome-wide studies in populations.

GWAS are designed to assess the common SNPs in the human genome. Thus, studies with these genotypes naturally uncover associations with common SNPs (frequencies larger than 1%). The estimated effect sizes of these common SNPs have been sometimes disappointing. However, according to Fisher this is fully expected[5]. His multifactorial model states that: (1) several, but not an unlimited number, of loci are involved in the expression of the trait, (2) there is no dominance or recessivity at each of these loci, but rather a co-dominant effect, (3) the loci act in concert in an additive fashion, each adding or detracting a small amount from the phenotype, and (4) the environment interacts with the genotype to produce the final phenotype. Genome wide association is not facing its limits. There are still many more common variants that have not been discovered by GWAS, because their effect sizes are too small and need larger samples to be identified. A question is whether enlarging sample size will uncover genes for depression. Remarkably, the genomic architecture of the human lipidome shows loci with larger effects. The re-discovery of the *FADS 1-2-3* locus for polyunsaturated fatty acid metabolism shows a single common SNP that explains up to 23% of the variation (chapter 4.2). This corresponds to 54% of the narrow sense heritability. This finding highlights the fact that the lipidome is under strong genetic control. Studies of these endophenotypes may uncover genes relevant for clinical phenotypes such as cardiovascular, endocrinal (diabetes) and neurological (Alzheimer and MDD) diseases.

GWAS provide preliminary genetic information that is available for additional analysis. The variants identified by GWAS are often located in intergenic regions, since only 1% of the genome codes for proteins. They may not be the causal variant and the association may be due to another variant which is in incomplete LD with the GWAS hit. Alternatively, these SNPs may be involved in regulating genes either in close proximity (*cis*) or more distantly (*trans*). Dense genotyping of a candidate locus can be useful to identify the real variant. Due to the stringent *P*-values for association (*P*-value $< 10^{-7}$ or *P*-value $< 5 \times 10^{-8}$) many causal variants do not meet this criteria. Pathway enrichment analysis, for particular pathways can be used to evaluate this evidence of non-significant genes, as was mentioned in chapter 3.2.

A major problem encountered in linkage is the large regions identified, making it difficult to pinpoint the causal gene. We overcame this problem by combining linkage and association analysis in the ERF study (chapters 31. and 4.1). Dense numbers of SNPs in GWAS panels can be used to fine-map the linkage regions, and may help to find better proxies for the real causal variation.

Today, large scale sequencing is becoming more feasible and it is expected to be a common option in the future. Rare coding variants, which will be seen in only a small fraction of the affected individuals or in individuals with high values of the quantitative phenotype, will provide important information for the genetic etiology. On one hand, sequencing can be used to find novel loci. On the other hand, this technique opens opportunities to investigate many regions of interest identified by GWAS [6]. However, if we want to follow a hypothesis free approach, as has been done in GWAS, one of the most important limitations of rare variant analysis will

be the insufficient power to detect rare variations. Collapsing methods that combine multiple rare variants in a particular gene or genomic location are being developed for this reason. Discovery of rare variants with these methods also requires severe multiple testing corrections, for instance a genome-wide *P*-value threshold of 2.4×10^{-6} when taking into account 20 712 genes tested together. Until now, the exome data we have derived has been useful in families with major diseases. Exomes are much more difficult to interpret in the general population. The first sequence data from humans showed that not all rare variants are damaging and even stop codons can occur in healthy individuals.

MAIN FINDINGS

Genetic architecture

In Chapter 2.1, we sought to investigate the genetic architecture of depression and the potential overlap in genetic risk factors with anxiety. First, we showed the evidence for a genome-wide signature using genetic risk scores for depression and anxiety. This genome-wide signature was constructed of many common variants with small effects. One interesting finding in this study was that up to 2.1% of the anxiety phenotype was explained by the additive genetic score derived from the GAIN-MDD GWAS[7], suggesting a shared background between these two disorders.

In the study described in chapter 2.2, we employed the same methodology for the serum lipoproteins. This time, we used the ENGAGE lipid GWAS[8] as the discovery set. We could explain up to 4.8% of the variance in HDL-C, 2.6% in LDL-C, 3.8% in TG and 2.7% in TC, comparable to previous studies[9, 10], however lower than the 12.4% (TC), 12.2% (LDL-C), 12.1% (HDL-C), and 9.6% (TG) of the variance estimated in the Framingham Heart Study sample[11] using data from the GLGC GWAS[12]. We think this is due to the fact that increases in sample size lead to better estimation of the effect sizes of the SNPs and the GLGC had a sample size 5 times larger than the ENGAGE sample. For all of the studied traits, we found the highest explained variance when the polygenic score was based on SNPs with low *P*-values (5×10^{-8} for HDL-C, 1×10^{-7} for TG, 1×10^{-6} for LDL-C and 1×10^{-5} for TC) which led us to conclude that, for circulating lipids, GWAS has succeed the discovery of common variants and new studies with increased sample sizes are promising to dissect the true positive signals from common variants above the threshold. On the other hand, apart from those variants with considerable effect sizes, we also showed a residual polygenic component which was visible after adjusting for the top significant effects. We conclude that for classical lipids, there are also variants with very small effects, which we may be underpowered to discover.

Gene discovery studies for depressive symptoms

Chapter 3 mainly focuses on the genetics of depression as measured by depressive symptoms. In Chapter 3.1, we employed a linkage approach for gene discovery for depressive complaints, as measured by the CES-D and HADS-D scales. We identified 3 significant (9q21, 13q33, 16q21) and 6 suggestive regions (1p36, 3p14, 5q34, 9q32, 10q22, 22q11) either for CES-D or HADS-D scales. Both CES-D and HADS-D are used for scanning in the general population, but the latter does not contain questions on somatic complaints, which means that it also useful for hospital screenings. This difference is also reflected in our results; even though the LOD scores are correlated, the significance does not emerge in the same locus. Our search for common variation inside the linkage regions combined with brain eQTL data suggests that *PRUNE2*, *C16ORF80* and *NDRG4* may be involved in the pathogenesis of depression. Using exome sequence data, we also identified a damaging mutation inside the coding sequence of the *ATP10B* (5q34) gene, which relates to higher depression complaints among the carriers in a single large pedigree. *ATP10B* is brain expressed and encodes a phospholipid flippase protein and may be involved in neuronal signalling, but it was not our candidate gene inside the 5q34 region which also included gamma-amino butyric acid (GABA) receptors. GABA is the major inhibitory neurotransmitter in the mammalian brain and GABA receptors have been implicated in several neurologic and psychiatric outcomes[13-18]. Parologs of *ATP10B, ATP10A* and *ATP10D,* also cluster together with GABA receptor subunit genes on 15q12 and 4p12 which relate to Angelman Syndrome, recurrent MDD and alcohol dependence[14, 19-21]. A study of childhood onset schizophrenia also reported uniparental isodysomie of the 5q32-qter which also includes our candidate region of interest[22]. All taken together, 3 regions on human genome (4p12, 5q34 and 15q12) show sequence and functional homology, as well as association with psychiatric phenotypes, and two of them (15q12 and 5q34) show evidence for non-Mendelian inheritance (genomic imprinting). We think these regions are interesting candidate regions for psychiatric and/or neurologic research, but also suggest a functional link between phospholipid membrane dynamics and ion channels.

Chapter 3.2 describes the largest meta-analysis of depressive symptoms to date, which we performed within the CHARGE Consortium. In this GWAS of depressive symptoms, we combined the results of 17 population-based studies with 34 549 individuals to find common variants for depressive symptoms. Including the 5 replication cohorts, this effort comprised data from 51 258 independent individuals. One SNP in the 5q21 region showed suggestive evidence for association with depressive symptoms in the replication set and reached genome wide significance when tested over all cohorts. We could not replicate associations with candidate genes that were previously reported to be associated with depression. However, our pathway analysis on the results of the overall meta-analysis showed that biological processes that play a role in depression, such as neurotransmitter secretion, vitamin transport and synaptic transmis-

sion, were overrepresented amongst our top hits, but variants in these genes did not reach the genome-wide significance threshold.

The recent study presented in chapter 3.3 is conducted within a partially overlapping sample with the study in chapter 3.2, from the CHARGE Consortium. For this study, we performed GWAS on dimensions of depressive symptoms to find common variants that increase the vulnerability to a particular symptom cluster: somatic symptoms, positive effect and negative effect. rs713224, in the *MTNR1A* gene, showed genome-wide significant association in the meta-analysis of the combined sample. However, the association with rs713224 was not replicated in the additional 8 separate samples, with evidence for heterogeneity.

Gene discovery studies for the circulating human lipidome

Chapter 4 shows two different studies of gene discovery for circulating lipids based on the mass spectrometry data of 148 species of phospholipids and sphingolipids. In our study described in chapter 4.1, we performed quantitative linkage methods on the ERF pedigree. We also, for the first time, estimated the heritability of these traits. We determined that for PLPE 16:0/20:5, SPM 24:3, SPM dih 18:0, SPM dih 22:0, SPM 22:2, SPM dih 16:0, PC O 38:1, PLPE 18:1/20:5, PE 34:0, SPM, 24:3, SPM 17:0, PE 42:6, PC 36:0, PC 34:0 and PC 40:5, the heritability was estimated less than 5%, which makes them less suitable for genetic research. Two out of our top 9 linkage loci, were previously uncovered through association by both us and others (*FADS 1-2-3* on 11q23 and *LPAR2* on 19p12). Among the suggestive linkage loci, 14q23.1- 14q24.2 was linked to PC O 34:2, PC O 36:2, PC O 38:3 and PC O 36:3. This locus harbors *PLEKHH1*, which we previously showed was associated with other alkyl-acyl PCs, PC O 36:5 and PC O 32:1, in our recent GWAS[23], as described in chapter 4.2. Three linkage loci are interesting since literature search revealed evidence for their involvement in carcinogenesis. The 17q21.31 locus includes *BRCA1*, which is a well known risk gene for early onset breast cancer, but has also been shown to influence fatty acid biosynthesis and lipogenesis in normal cells[24, 25]. The 6q16.3-21 locus includes tumor suppressors *HACE1* and *PRDM1*. Finally on 11q21, the linkage SNP rs483884 maps to the *MAML2* gene, which is involved in the carcinogenesis pathway through co-activation of *NOTCH* proteins[26]. The phenotypic relationship between the lipid traits we studied and cancer is unknown, but knowing the involvement of the phospho- and sphingolipids in cell signalling, a causal relationship is not unlikely.

Chapter 4.2 summarizes our genome wide meta-analysis within the EUROSPAN Consortium. GWAS of targeted metabolomics data of related traits has been performed before, however with smaller sample sizes[27, 28]. Considering the overlapping metabolic species in our study and previous studies, we were directly able to observe the increased in power in our results, possibly due to the increased sample size. We replicated the previously known loci for particular lipid traits, such as *FADS1, LIPC, PLEKHH1, GCKR, APOA1-5,* and *ELOVL2* with smaller *P*-values. We also showed association for the first time with *PAQR9* on 3q23 (associated with %PE 34:1 and %PE

36:1), *AGPAT1* on 6p21.32 (associated with PC 32:0), *PKD2L1* on 10q24.31 (LPC 16:1), *PDXDC1* on 16p13.11 (LPC 20:3, PC 34:2, PC 36:3 and PC 38:3), *APOE* on 19q13.32 (associated with SPM 24:0 and SPM 22:0), *PLD2* on 17p13.2 (associated with SPM 23:0). Eighteen additional regions provided suggestive evidence of association ($2.2 \times 10^{-9} < P\text{-value} < 5 \times 10^{-8}$) including the *PNLIPRP2, ZNF600, ALG1, ABHD3, KLF12, DLG2, ILKAP, ITGA9, PLPE, OR8I2, PCDH20, CDK17, SYT9, CDH8, KCNH7, ALG14, PAPD7, CNTNAP4* and *LPAR2* regions. With a closer look of the findings published by Illig et al., we can see the association signals within 100 kb of the loci containing *PDXDC1* (same SNP, *P*-value = 2.8×10^{-7}), *AGPAT1* (*P*-value = 4.9×10^{-7}), *PNLIPRP2* (*P*-value = 2.7×10^{-7}), *KLF12* (*P*-value = 5.9×10^{-7}), *ALG1* (*P*-value = 4.7×10^{-3}), *CDH8* (*P*-value = 7.6×10^{-7}), *PLD2* (*P*-value = 9.4×10^{-4}) and *ZNF600* (*P*-value = 3.3×10^{-7}) for various phospho- and sphingolipid outcomes. Some of the phospho – and sphingolipid loci were also associated with disease end points in other studies. For instance, our top SNP in the *FADS* cluster (rs102275) was associated with IMT in the CHARGE meta-analysis results[29]. This SNP is considered to be the major controller of fatty acid desaturation and the path to the disease is supposed to be through an excess of saturated circulating lipids, as compared to the unsaturated ones. For some of the loci, it is difficult to determine the causal gene, for instance one SNP, rs1061808 in the *AGPAT1* locus on *6p21,* also lies 400 kb distant from the *HLA-DRB1* gene, which was previously associated with insulin secretion[30]. We observed a suggestive association between rs1061808 and increased T2DM risk in the DIAGRAM consortium meta-analysis[31]. However, the causal paths of these associations have not been proven thus far and can be proven by Mendelian randomisation. We also observed an association with the well-known *APOE* locus, which also relates to LDL-C and HDL-C. For this locus, we showed that the effect of the SNP on a particular sphingomyelin level is dependent on its effect on LDL-C. In addition, another SNP from the *APOA1-5* locus was implicated in CAD risk in the CARDIoGRAM study[32]. *APOA1* and *APOA2* are the predominant proteins in HDL-C particles, which also transport TG. For this locus, we observed that the association was not explained by TG levels, and we hypothesized that might be due to lecithin cholesterol acyltransferase (LCAT), which converts cholesterol and PC to cholesteryl esters and LPC on the surface of HDL-C[33].

The circulating lipidome: new biomarkers for depression?

In Chapter 5, we moved our search one step further and studied the phenotypic relationship between our two topics of interest, depression and anxiety symptoms and circulating lipids. The benefits of unsaturated fatty acids, especially fish oil, for the brain and circulation system are well-accepted knowledge in the 21st century. However, the causal relationship between types of PUFA (ω-3 and-ω-6) and depressive disorder in different sources (serum, plasma, serum phospholipids) has been difficult to understand, mainly due to the fact that both are essential for a healthy functioning brain. Our analyses showed that there is an inverse correlation between depression and anxiety symptoms and PC O 36:4 and SPM 23:1. PC O 36:4 was also

measured in an independent sample. Combining the two samples yielded a highly significant and consistent correlation with the depression and anxiety scale with the *P-value* for the HADS-A scale improving (meta-analysis *P-value* = 9.8×10^{-5}). However, even though the discovery and replication samples were randomly selected, we observed differences between them. The discovery sample had lower education, was older, and reported more complaints of depression. Another difference was the measurement platforms: in-house versus custom array. Ether lipids are known to be predominant in brain [34,35]. We suggest that an enzymatic hyperactivity towards hydrolysis of PC O 36:4 may lead to increased release of arachidonic acid, which is then rapidly converted into inflammatory mediators[36]. The enzyme involved in this path remains unknown but some phospholipases are known to relate to bipolar depression [37-40]. CoA-independent transacylase is also one of the candidates, since the product of its reaction on ether lipids result is platelet activating factor (PAF)[41]. PAF is another tightly controlled potent inflammatory mediator in brain and other tissues[42,43]. We also observed a negative correlation between SPM 23:1 and depression scores. This may be due to acid sphingomyelinase activity, which was previously found to be increased in depressed patients, compared to healthy controls, and also shown to be blocked by antidepressant drugs[44,45,46].

Future research

This thesis shows that lipids and depression are associated, but are likely to have different genetic architecture. For both lipids and depression, huge efforts have been performed by large international consortia to uncover the genes involved in these phenotypes. Following the first meta-analysis of lipids by the ENGAGE consortium (N = 23 000), which uncovered a total of 22 loci, the GLGC (N = 100 000) uncovered 95 loci below the genome-wide significance threshold. Now, this number is being further increased by the latest study from the same consortium, recruiting more individuals genotyped in depth with the "metabo-array" and reporting more novel loci[12]. There is no evidence that association using arrays has reached its limits for lipids and, therefore, enlarging datasets will bring to light new loci. On the other hand, the GWAS on depression continued to report lack of genome-wide significance, and lack of replication, despite the large consortia effort for both diagnosis and symptom measurement tools (*Hek, et al*, unpublished)[14,47]. Apparently, when searching for genes for depression through GWAS, we are still underpowered with 50 000 subjects to detect an association: the effect sizes for the common variants are too small. In line with our findings, the difficulty of finding and replicating signals in GWAS of depression has been a common experience both for depressive symptoms and MDD.

Enlarging the sample size is an avenue that may, in the end, bring success. Alternatively, our work suggests that we may have to live with the fact that many loci with (too) small effects are too difficult to identify. In that case, the full genome may be better at predicting MDD, and gene discovery efforts may be better focused on searching for loci with large effects in families with

multiple affected individuals. We have shown that lipidomics studies are more powerful to find new loci and this is a new approach.

Since the first GWAS of MDD was published in 2009, we still have not discovered any major genetic variation that would enlighten our knowledge on its etiology or epidemiology. Next generation sequencing in the general population may be an alternative, but also require enormous number of subjects to study in order to reach adequate power for gene discovery. The situation we experienced so far shows that we are still quite underpowered for hypothesis free association testing for depressive symptoms. The same experience has also been encountered by the Psychiatric Genetics Consortium[14]. One solution for that can be focusing on markers of causal pathways such as circulating lipids or brain imaging. Our first studies suggest that we may be able to use specific circulating lipids as endophenotypes. Our finding that PC O 36:4 levels in blood were associated with depressive symptoms suggests that genes determining these lipids may also be involved. Our findings on the *ATP10B* gene suggest that the phospholipid distribution in the cell membrane may be important . Further studies on the metabolom of patients may uncover further links and possible biomarkers. The work in this thesis was a first step towards linking lipidomics and depressive symptoms. Further studies involving the lipidome and metabolome may be helpful in unraveling the genetics of depression.

REFERENCES

1. Liebisch G, Drobnik W, Lieser B, Schmitz G. High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clin Chem.* Dec 2002;48(12):2217-2224.
2. Liebisch G, Lieser B, Rathenberg J, Drobnik W, Schmitz G. High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochim Biophys Acta.* Nov 8 2004;1686(1-2):108-117.
3. Zemski Berry KA, Murphy RC. Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. *J Am Soc Mass Spectrom.* Oct 2004;15(10):1499-1508.
4. Brugger B, Erben G, Sandhoff R, Wieland FT, Lehmann WD. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc Natl Acad Sci U S A.* Mar 18 1997;94(6):2339-2344.
5. Fisher RA. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Philosophical Transactions of the Royal Society of Edinburgh.* 1918;52:399-433.
6. Sanna S, Li B, Mulas A, et al. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet.* Jul;7(7):e1002198.
7. Boomsma DI, Willemsen G, Sullivan PF, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet.* Mar 2008;16(3):335-342.
8. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* Jan 2009;41(1):47-55.
9. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* Jan 2009;41(1):47-55.
10. 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.* Jan 2009;41(1):35-46.
11. Teslovich TM, Musunuru K, Smith AV, et al. Meta-analysis of >100,000 individuals identifies 63 new loci associated with serum lipid concentrations. *The American Society of Human Genetics.* Hawaii; 2009.
12. 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.
13. Roberts E, Sherman MA. GABA--the quintessential neurotransmitter: electroneutrality, fidelity, specificity, and a model for the ligand binding site of GABAA receptors. *Neurochem Res.* Apr 1993;18(4):365-376.
14. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry.* Apr 3.
15. Hashimoto T, Matsubara T, Lewis DA. [Schizophrenia and cortical GABA neurotransmission]. *Seishin Shinkeigaku Zasshi.*112(5):439-452.
16. Richter MA, Zai G, McBride JC, Mundo E, Swinson RP, Kennedy JL. The GABA(A)-receptor gamma2 (GABRG2) gene in obsessive-compulsive disorder. *Rev Bras Psiquiatr.* Dec 2009;31(4):328-331.
17. Pham X, Sun C, Chen X, et al. Association study between GABA receptor genes and anxiety spectrum disorders. *Depress Anxiety.* 2009;26(11):998-1003.
18. Henkel V, Baghai TC, Eser D, et al. The gamma amino butyric acid (GABA) receptor alpha-3 subunit gene polymorphism in unipolar depressive disorder: a genetic association study. *Am J Med Genet B Neuropsychiatr Genet.* Apr 1 2004;126B(1):82-87.
19. Vollenweider I, Smith KS, Keist R, Rudolph U. Antidepressant-like properties of alpha2-containing GABA(A) receptors. *Behav Brain Res.* Feb 2;217(1):77-80.

20. Kertes DA, Kalsi G, Prescott CA, et al. Neurotransmitter and neuromodulator genes associated with a history of depressive symptoms in individuals with alcohol dependence. *Alcohol Clin Exp Res.* Mar;35(3):496-505.
21. Dan B, Boyd SG. Angelman syndrome reviewed from a neurophysiological perspective. The UBE3A-GABRB3 hypothesis. *Neuropediatrics.* Aug 2003;34(4):169-176.
22. Seal JL, Gornick MC, Gogtay N, et al. Segmental uniparental isodisomy on 5q32-qter in a patient with childhood-onset schizophrenia. *J Med Genet.* Nov 2006;43(11):887-892.
23. Demirkan A, van Duijn CM, Ugocsai P, et al. Genome-Wide Association Study Identifies Novel Loci Associated with Circulating Phospho- and Sphingolipid Concentrations. *PLoS Genet.* Feb;8(2):e1002490.
24. Ortega FJ, Moreno-Navarrete JM, Mayas D, et al. Breast Cancer 1 (BrCa1) May Be behind Decreased Lipogenesis in Adipose Tissue from Obese Subjects. *PLoS One.*7(5):e33233.
25. Ray H, Suau F, Vincent A, Dalla Venezia N. Cell cycle regulation of the BRCA1/acetyl-CoA-carboxylase complex. *Biochem Biophys Res Commun.* Jan 16 2009;378(3):615-619.
26. Nemoto N, Suzukawa K, Shimizu S, et al. Identification of a novel fusion gene MLL-MAML2 in secondary acute myelogenous leukemia and myelodysplastic syndrome with inv(11)(q21q23). *Genes Chromosomes Cancer.* Sep 2007;46(9):813-819.
27. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* Nov 2008;4(11):e1000282.
28. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet.* Feb 2009;42(2):137-141.
29. Bis JC, Kavousi M, Franceschini N, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet.* 2011;43(10):940-947.
30. Howson JM, Rosinger S, Smyth DJ, Boehm BO, Todd JA. Genetic Analysis of Adult-Onset Autoimmune Diabetes. *Diabetes.* Oct 2011;60(10):2645-2653.
31. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet.* Jul;42(7):579-589.
32. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* Apr 2011;43(4):333-338.
33. Breslow JL, Ross D, McPherson J, et al. Isolation and characterization of cDNA clones for human apolipoprotein A-I. *Proc Natl Acad Sci U S A.* Nov 1982;79(22):6861-6865.
34. Farooqui AA, Horrocks LA, Farooqui T. Glycerophospholipids in brain: their metabolism, incorporation into membranes, functions, and involvement in neurological disorders. *Chem Phys Lipids.* Jun 2000;106(1):1-29.
35. Brites P, Waterham HR, Wanders RJ. Functions and biosynthesis of plasmalogens in health and disease. *Biochim Biophys Acta.* Mar 22 2004;1636(2-3):219-231.
36. Farooqui AA, Horrocks LA. Brain phospholipases A2: a perspective on the history. *Prostaglandins Leukot Essent Fatty Acids.* Sep 2004;71(3):161-169.
37. Chang MC, Grange E, Rabin O, Bell JM, Allen DD, Rapoport SI. Lithium decreases turnover of arachidonate in several brain phospholipids. *Neurosci Lett.* Dec 20 1996;220(3):171-174.
38. Chang MC, Jones CR. Chronic lithium treatment decreases brain phospholipase A2 activity. *Neurochem Res.* Jun 1998;23(6):887-892.
39. Kim HW, Rapoport SI, Rao JS. Altered arachidonic acid cascade enzymes in postmortem brain from bipolar disorder patients. *Mol Psychiatry.* Apr;16(4):419-428.
40. Guzzetta F, Tondo L, Centorrino F, Baldessarini RJ. Lithium treatment reduces suicide risk in recurrent major depressive disorder. *J Clin Psychiatry.* Mar 2007;68(3):380-383.

41. Chilton FH, Fonteh AN, Sung CM, et al. Inhibitors of CoA-independent transacylase block the movement of arachidonate into 1-ether-linked phospholipids of human neutrophils. *Biochemistry.* Apr 25 1995;34(16):5403-5410.
42. Farooqui AA. Lipid mediators in the neural cell nucleus: their metabolism, signaling, and association with neurological disorders. *Neuroscientist.* Aug 2009;15(4):392-407.
43. Winkler JD, Bolognese BJ, Roshak AK, Sung CM, Marshall LA. Evidence that 85 kDa phospholipase A2 is not linked to CoA-independent transacylase-mediated production of platelet-activating factor in human monocytes. *Biochim Biophys Acta.* Jun 2 1997;1346(2):173-184.
44. Kornhuber J, Medlin A, Bleich S, et al. High activity of acid sphingomyelinase in major depression. *J Neural Transm.* Nov 2005;112(11):1583-1590.
45. Kolzer M, Werth N, Sandhoff K. Interactions of acid sphingomyelinase and lipid bilayers in the presence of the tricyclic antidepressant desipramine. *FEBS Lett.* Feb 13 2004;559(1-3):96-98.
46. Kornhuber J, Reichel M, Tripal P, et al. The role of ceramide in major depressive disorder. *Eur Arch Psychiatry Clin Neurosci.* Nov 2009;259 Suppl 2:S199-204.
47. I. SS, Shi J, Kraft JB, et al. Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry.*

7

Summary

SUMMARY

The main aim of this thesis was to find candidate genes and pathways for depression while investigating its association with circulating lipids as potential endophenotypes (in addition to depressive symptoms). Chapter 1 includes a general introduction. In Chapter 2.1 we investigated the genetic architecture of depression and the potential overlap in genetic risk factors with anxiety. In two independent target samples, either diagnosed according to the DSM-IV or measured with the CES-D or HADS-D, we could explain up to ~1% of the variance for depression and 2.1% for anxiety. With this study, we directly showed a polygenic component influencing the susceptibility for a psychiatric disorder, as well as an overlap in genetic risk factors for another psychiatric condition (anxiety).

In the study described in chapter 2.2, we utilized the same methodology as in chapter 2.1 to study the serum lipoproteins. Using the ENGAGE lipid GWAS as the discovery set, we could explain up to 4.8% of the variance in HDL-C, 2.6% in LDL-C, 3.8% in TG and 2.7% in TC. For all of the studied traits, we found the highest explained variance when the polygenic score was based on SNPs with low *P*-values (5×10^{-8} for HDL-C, 1×10^{-7} for TG, 1×10^{-6} for LDL-C and 1×10^{-5} for TC), which led us to conclude that, for circulating lipids, GWAS has succeeded in the discovery of common variants.

Chapter 3.1 shows the results of an integrated approach in which we combined linkage, association and exome sequencing. For this study, we performed linkage analysis in the ERF population for depressive complaints, as measured by the CES-D and HADS-D scales. We identified three significant (9q21, 13q33, 16q21) and 6 suggestive (1p36, 3p14, 5q34, 9q32, 10q22, 22q11) loci for either the CES-D or HADS-D scales. To have a closer look at the genes in these loci, we narrowed down the regions using association methods with common variants. We identified 2 common SNPs related to the genes *PRUNE2*, *C16ORF80*, which are partially responsible for the linkage signals. Additionally, using exome sequence data, we identified a missense mutation inside the coding sequence of the *ATP10B* (5q34) gene, which is related to higher depression complaints among the carriers from a single large pedigree.

Chapter 3.2 describes the largest meta-analysis of depressive symptoms to date, which we performed within the CHARGE Consortium. In this study, we combined the GWAS results from 17 population-based studies, totaling 34 549 individuals, to find common variants associated with CES-D scale. Our top hits were further investigated in 5 replication cohorts in which symptoms of depression were measured by different instruments, yielding a combined dataset of 51 258 independent individuals. Only one SNP in the 5q21 region reached genome wide significance with depressive symptoms when tested over all cohorts, but we also showed that our top results are enriched by neurotransmission related genes.

The study described in chapter 3.3 was also conducted within the CHARGE Consortium. For this study, we performed a GWAS on dimensions of depressive symptoms to find common variants that increase the vulnerability to a particular symptom cluster: somatic symptoms,

positive effect and negative effect. rs713224, in the *MTNR1A* gene, showed genome-wide significant association in the meta-analysis of the combined sample. However, the association with rs713224 was not replicated in an additional 8 separate samples, with evidence for heterogeneity.

In the study described in chapter 4.1, we applied quantitative linkage methods to the lipidomics data in the ERF pedigree. We showed significant linkage to nine loci. Two of these nine had been previously detected through association by us and others and (*FADS 1-2-3* on 11q23 and *LPAR2* on 19p12). For *FADS*, we showed that the same common variation (rs102275), or variants in strong LD with it, was responsible for the linkage peak. We also, for the first time, estimated the heritability of these traits. We saw that, for PLPE 16:0/20:5, SPM 24:3, SPM dih 18:0, SPM dih 22:0, SPM 22:2, SPM dih 16:0, PC O 38:1, PLPE 18:1/20:5, PE 34:0, SPM, 24:3, SPM 17:0, PE 42:6, PC 36:0, PC 34:0 and PC 40:5, the heritability was estimated to be less than 5%, which makes them less suitable for genetic research.

Chapter 4.2 summarizes our genome wide a meta-analysis within the EUROSPAN Consortium. We replicated the previously known loci for these lipid traits, such as *FADS1, LIPC, PLEKHH1, GCKR, APOA1-5,* and *ELOVL2,* with smaller *P*-values. We also found novel associations with *PAQR9* on 3q23 (associated with %PE 34:1 and %PE 36:1), *AGPAT1* on 6p21.32 (associated with PC 32:0), *PKD2L1* on 10q24.31 (LPC 16:1), *PDXDC1* on 16p13.11 (LPC 20:3, PC 34:2, PC 36:3 and PC 38:3), *APOE* on 19q13.32 (associated with SPM 24:0 and SPM 22:0), and *PLD2* on 17p13.2 (associated with SPM 23:0). We also found eighteen additional regions suggestively associated ($2.2 \times 10^{-9} < P\text{-value} < 5 \times 10^{-8}$) including *PNLIPRP2, ZNF600, ALG1, ABHD3, KLF12, DLG2, ILKAP, ITGA9, PLPE, OR8I2, PCDH20, CDK17, SYT9, CDH8, KCNH7, ALG14, PAPD7, CNTNAP4* and *LPAR2*. Our top SNP in the *FADS* cluster (rs102275) was associated with IMT in the CHARGE meta-analysis results. This SNP is thought to be the major controller of fatty acid desaturation and the path to the atherosclerotic disease is supposed to be through an excess of saturated circulating lipids, as compared to unsaturated ones.

In Chapter 5.1, we moved our search one step further and studied the phenotypic relationship between our two topics of interest: depression and anxiety symptoms and circulating lipids. In our study, we analysed 148 phospho- and sphingolipid species, rather than the total fatty acid concentration or serum lipoproteins. Our analyses showed that there is an inverse correlation between depression and anxiety symptoms and PC O 36:4 and SPM 23:1. PC O 36:4 was also measured in an independent sample. Combining the two samples yielded a highly significant and consistent correlation with the depression and anxiety symptoms. The association may be due to increased *PLA2* or sphingomyelinase activity and genes coding for these enzymes are potential candidates for depressive symptoms research.

This thesis shows that the genetic architecture of two related outcomes (lipids and depression) is different. While the genetics of lipids involved genes with large effects, for depression we failed to find any. In chapter 6 (general discussion), we discuss the steps to further uncover the genetic aetiology of these traits and their relationship to each other.

SAMENVATTING

Het belangrijkste doel van dit proefschrift was het achterhalen van kandidaatgenen en biologische paden met betrekking tot depressie door de associatie met circulerende lipide als mogelijke endophenotypes (naast depressie symptomen) te onderzoeken. Hoofdstuk 1 bevat de algemene inleiding. In hoofdstuk 2.1 onderzoeken we de genetische architectuur van depressie als ook de potentiële overlap qua genetische risicofactoren met angstigheid. In twee onafhankelijke studies, òf gediagnosticeerd met de DSM-IV criteria òf gemeten met de CES-D of HADS-D vragenlijst, konden we maximaal ongeveer 1% van de variabiliteit van depressie en 2.1% van angstigheid verklaren. Met deze studie hebben we een poly-gen component gevonden die de vatbaarheid voor psychiatrische ziektes beïnvloedt, als ook een overlap in genetische risico-factoren voor andere een psychriatische eigenschap (angstigheid).

In de studie beschreven in hoofdstuk 2.2 gebruikten we dezelfde methodologie als in hoofdstuk 2.1 om de serum lipoproteïnen te onderzoeken. Door de ENGAGE vet GWAS als de ontdekkingsset te gebruiken konden we tot 4.8% van de variabiliteit in HDL-cholesterol, 2.6% in LDL-cholesterol, 3.8% in triglyceriden and 2.7% in totaal cholesterol verklaren. Voor al de onderzochten eigenschappen vonden we de grootste verklaarbare variabiliteit als de poly-gen score gebaseerd was op SNPs met lagen *P*-waardes (5×10^{-8} voor HDL-cholesterol, 1×10^{-7} voor triglyceriden, 1×10^{-6} voor LDL-cholesterol and 1×10^{-5} voor totaal cholesterol), wat ons doet concluderen dat, voor circulerende vetten, GWAS succesvol is geweest voor de ontdekking van veelvoorkomende varianten.

Hoofdstuk 3.1 toont de resultaten voor een geïntegreerde benadering waarin we linkage, associatie en exome sequencing combineren. Voor deze studie gebruikten we linkage analyse in de ERF populatie voor depressieve symptomen, zoals gemeten bij de CES-D en HADS-Dschalen. We hebben drie significante (9q21, 13q33, 16q21) en zes suggestieve regio's (1p36, 3p14, 5q34, 9q32, 10q22, 22q11) gevonden voor CES-D of HADS-D. Om een beter inzicht te krijgen in deze gebieden hebben we de gebieden verkleind door gebruik te maken van associatie methodes met veelvoorkomende varianten. We hebben twee veelvoorkomende SNPs gerelateerd aan de genen *PRUNE2 en C16ORF80* gevonden, welke gedeeltelijk verantwoordelijk zijn voor de linkage signalen. Boevendien hebben we door middel van exome sequentie data, een mis-sense mutatie gevonden in de coderende sequentie van het *ATP10B* (5q34) gen, welke gerelateerd is aan hogere depressie klachten bij de dragers in een grote stamboom.

Hoofdstuk 3.2 beschrijft de grootste meta-analyse van depressieve symptomen tot nu toe die we hebben uitgevoerd in het CHARGE Consortium. In deze studie hebben we de GWAS resultaten van 17 populatie gebaseerde studies gecombineerd, met een totaal van 34.549 deelnemers, om veelvoorkomende varianten te associëren met de CES-D schaal. Onze top resultaten werden verder onderzocht in vijf replicatiecohorten waarin de symptomen van depressie met andere instrumenten gemeten waren, hetgeen een gecombineerde dataset van 51.258 onafhankelijke deelnemers opleverde. Slechts één SNP in de 5q21 regio haalde

genoomwijde significantie met depressieve symptomen wanneer we alle cohorten bekeken, maar we toonden ook dat onze top resultaten verrijkt zijn met genen die betrokken zijn bij neurotransmissie.

De studie beschreven in hoofdstuk 3.3 is ook uitgevoerd in het CHARGE consortium. Voor deze studie voerden we een GWAS uit op aspecten van depressieve symptomen om veelvoorkomende varianten te identificeren die de ontvankelijkheid voor een specifieke symptomen cluster verhogen: somatische symptomen, positief en negatief effect. De SNP rs713224, in het *MTNR1A* gen, toonde een genoomwijde significante associatie in de meta-analyse in de gecombineerde dataset. Echter, de associatie van rs713224 werd niet gerepliceerd in een extra acht onafhankelijke studies met bewijs voor heterogeniteit.

In de studie beschreven in hoofdstuk 4.1 hebben we kwantitatieve linkage methoden gebruikt om de lipidomics data in de ERF stamboom te analyseren. We vonden significante linkage voor negen regio's. Twee van deze negen regio's werden eerder gevonden door middel van associatie door ons en anderen (*FADS 1-2-3* op 11q23 en *LPAR2* op 19p12). Voor FADS vonden we dat dezelfde veelvoorkomende variant (rs102275), of varianten in sterke LD met deze SNP, verantwoordelijk was voor de linkage piek. We hebben ook, voor de eerste keer, de erfelijkheid van deze eigenschappen bepaalt. We zagen dat voor PLPE 16:0/20:5, SPM 24:3, SPM dih 18:0, SPM dih 22:0, SPM 22:2, SPM dih 16:0, PC O 38:1, PLPE 18:1/20:5, PE 34:0, SPM, 24:3, SPM 17:0, PE 42:6, PC 36:0, PC 34:0 en PC 40:5, de erfelijkheid als minder dan 5% werd geschat, hetgeen ze minder geschikt maakt voor genetisch onderzoek.

Hoofdstuk 4.2 resumeert onze GWAS binnen het EUROSPAN consortium. We repliceerde de eerder beschreven regio's voor deze vet eigenschappen zoals *FADS1, LIPC, PLEKHH1, GCKR, APOA1-5,* en *ELOVL2,* met kleinere *P*-waardes. We vonden ook een nieuwe associatie met *PAQR9* op 3q23 (geassocieerd met %PE 34:1 en %PE 36:1), *AGPAT1* op 6p21.32 (geassocieerd met PC 32:0), *PKD2L1* op 10q24.31 (LPC 16:1), *PDXDC1* op 16p13.11 (LPC 20:3, PC 34:2, PC 36:3 en PC 38:3), *APOE* op 19q13.32 (SPM 24:0 en SPM 22:0), en *PLD2* op 17p13.2 (SPM 23:0). We hebben nog eens achttien suggestieve regio's geïdentificeerd ($2.2 \times 10^{-9} < P$-waarde $< 5 \times 10^{-8}$) inclusief *PNLIPRP2, ZNF600, ALG1, ABHD3, KLF12, DLG2, ILKAP, ITGA9, PLPE, OR8I2, PCDH20, CDK17, SYT9, CDH8, KCNH7, ALG14, PAPD7, CNTNAP4* en *LPAR2*. Onze top SNP in het FADS cluster (rs102275) werd geassocieerd met IMT in de CHARGE meta-analyse resultaten. Deze SNP is waarschijnlijk een belangrijke regelaar van vetzuur desaturatie en verantwoordelijk voor een overschot van verzadigde circulerende vetten, in plaats van een gezonde overschot aan onverzadigde vetten en daardoor waarschijnlijk een belangrijke stap naar atherosclerose plaques.

In hoofdstuk 5.1 gingen we een stap verder en onderzochten we de phenotypische relatie tussen onze twee onderwerpen: depressieve en angstigheids symptomen enerzijds en circulerende vetten anderzijds. In onze studie analyseerde we 148 phospho- en sphingovetsoorten, in plaats van de totale vetzuur concentraties in serum veteiwitten. Onze analyses toonden aan dat er een inverse correlatie bestaat tussen depressieve en angstigheids symptomen en PC O 36:4 en SPM 23:1 .In een onafhankelijke studie is PC O 36:4 ook gemeten. Wanneer we de twee

studies combineerden kregen we sterke significantie en een consistente correlatie met depressieve en angstigheids symptomen. De associatie kan misschien verklaard worden door meer *PLA2* of sphingomyelinaseactiviteit en genen die coderen voor deze enzymen zijn potentiële kandidaten voor onderzoek naar depressieve symptomen .

Dit Proefschrift toont aan dat de genetische architectuur van de twee gerelateerde uitkomsten (vetten en depressie) verschillend is. Terwijl de genetica van vetten door genen met grote effecten wordt gekarakteriseerd, konden we voor depressie geen enkel gen vinden. In hoofdstuk 6 (algemene discussie) bespreken we de stappen die genomen moeten worden om de genetische oorzaak van deze eigenschappen en hun onderlinge relatie verder te ontrafelen.

ÖZET

Bu tezin amacı depresyon için aday genler ve yolaklar bulmak ve depresif semptomlara ek potansiyel endofenotipler olan lipitler ve depresyon ilişkisini araştırmaktır. Bölüm 1 genel bir tanıtım içerir. Bölüm 2.1'de depresyonun genetik mimarisini ve anksiyete ile potansiyel olarak çakışan genetik risk faktörlerini araştırıldı. Bağımsız iki hedef örnekte ya DSM-IV 'e göre tanısı koyulan ya da CES-D veya HADS-D ile ölçülen depresyon için varyansın ~%1 kadarını, anksiyete için ise % 2. 1'ini açıklanabildi. Bu çalışma ile psikiyatrik bir hastalık için yatkınlığı değiştiren poligenik bir bileşen doğrudan gösterildi, ek olarak bir başka psikiyatrik durum (anksiyete) için genetik risk faktörlerindeki çakışma gösterilmis oldu.

Bölüm 2.2'de serum lipoproteinlerine odaklanan calısmada bölüm 2.1 ile aynı metodolojiden faydalanıldı. Bulgu seti olarak ENGAGE lipid GWAS kullanarak HDL-C'de %4.8, LDL-C'de %2.6, TG'de %2.7 maksimum varyans açıklandı. Calışılan tüm fenotipler için açıklanan en yüksek varyans poligenik skorlar düşük *P*-değeri olan SNP'lerden hesaplandığında (HDL-C için 5×10^{-8} ,TG için 1×10^{-7}, LDL-C için 1×10^{-6} ve TC için 1×10^{-6}) gösterilebildi. Bu sonuclar dolaşımda yer alan lipitler için GWAS'ın allel frekansı yüksek varyantları bulmakta başarılı olduğunu düsündürmektedir.

Bölüm 3.1, bağlantı, ilişkilendirme ve ekzom dizilemeyi birleştirdiğimiz entegre bir yaklaşımın sonuçlarını göstermektedir. Bu çalışma icin ERF popülasyonunda CES-D veya HADS-D ölçekleriyle ölçülmüs olan depresif şikayetler bağlantı analizi ile arastirildi. Bağlantı analizleri sonucunda CES-D veya HADS-D ölçekleri için toplamda üç tane anlamlı (9q21, 13q33, 16q21) ve 6 tane önemli (1p36, 3p14, 5q34, 9q32, 10q22, 22q11) lokus tanımlandı. Bu lokuslardaki genlere daha yakından bakmak üzere ortak allel frekansı yüksek SNPler için ilişkilendirme metotları kullanılarak detaylı haritalama denendi. Bu analizler sonucunda *PRUNE2*, *C16ORF80* genleriyle ilişkili olan ve bağlantı sinyallerinden kısmen sorumlu olan 2 yüksek frekanslı SNP tanımlandı. Ek olarak ekzom dizileme datasını kullanarak *ATP10B* (5q34) geninin kodlayan dizisinin içinde bir yanlış anlam mutasyonu tanımlandı. Bu mutasyon tek bir büyük soyağacından gelen taşıyıcılar arasında yüksek depresyon şikayetleri ile ilişkilidir.

Bölüm 3.2 depresif semptomlar üzerine bugüne kadar yapılmıs en geniş meta-analizi göstermektedir. Bu projeyi CHARGE konsorsiyumu içinde gerçekleştirdik. Bu çalışmada toplamda 34 549 bireyden oluşan 17 GWAS sonucunu birleştirerek CES-D ölçeği ile iliskili varyantlar bulmak hedeflendi. En kuvvetli aday SNPler 5 replikasyon kohortu içinde ileri düzeyde analiz edildi. Depresyon semptomlarının değişik cihazlarla ölçüldügu 5 calismanın daha ilavesiyle 51 258 bağımsız bireye ait birleştirilmiş veri seti elde edildi. Tüm kohortlar test edildiğinde yalnızca 5q21 bölgesindeki bir SNP genom çapında anlamlılığa ulaştı ancak, *P*-degeri en anlamlı olan SNPlerin yolak analizleri nörotransmisyon ile ilişkili genlerin listenin üst sıralarında zengince temsil edildiğini de gösterilmis oldu.

Bölüm 3.3'de açıklanan çalışma da CHARGE Konsorsiyumu içinde yürütüldü. Bu çalışmada depresif semptomların boyutlarına GWAS uygulayıp belirli bir semptom kümesinde hassasiyeti artıran ortak varyantları bulmayı amaçladık. Bu semptom kümesi somatik semptomlar, pozitif etki ve negatif etkiden oluşmaktadır. Birleştirilen örneğin meta-analizinde *MTNR1A* geni içinde rs713224 genom çapında anlamlı ilişki gösterdi. Oysa rs713224 ile ilişki ilave 8 ayrı örnekte tekrarlanmadı, ancak anlamlı derecede heterojenite de gözlendi.

Bölüm 4.1'de açıklanan çalışma da, ERF soyağacına ait lipidomik ölcümlere kantitatif bağlantı analizi metodları uygulandı. 9 lokustaki anlamılı bağlantı gösterildi. Bunlardan ikisinin (*FADS 1-2-3* on 11q23 and *LPAR2* on 19p12) ilişkisi baska arastırıcılar ve tarafımızdan daha önce gösterilmisti. Bu calısmamızda buna ek olarak *FADS* için aynı varyasyonun ya da onunla sıkı bir BD (bağlantı dengesizliği) içindeki başka bir varyasyonun bu pikten sorumlu olduğu gösterildi. Ayrıca ilk olarak bu özelliklerin kalıtımsal geçişleri de tahmin edilmiştir. PLPE 16:0/20:5, SPM 24:3, SPM dih 18:0, SPM dih 22:0, SPM 22:2, SPM dih 16:0, PC O 38:1, PLPE 18:1/20:5, PE 34:0, SPM, 24:3, SPM 17:0, PE 42:6, PC 36:0, PC 34:0 ve PC 40:5 için kalıtımsal geçişler %5'den daha az olduğu için genetik araştırmalar için elverisli olmadığı görüldü.

Bölüm 4.2'de EUROSPAN Konsorsiyumundaki genom boyu meta analizimiz özetlenmektedir. Bu lipid özellikleri için önceden belirlenmiş *FADS1, LIPC, PLEKHH1, GCKR, APOA1-5,* ve *ELOVL2* gibi lokusların daha küçük *P*-değerleri ile replikasyonu yapıldı. Ayrıca 3q23'deki *PAQR9* (%PE 34:1 ve %PE 36:1 ile ilişkili), 6p21.32'deki *AGPAT1* (PC 32:0 ile ilşikili), 10q24.31 'deki *PKD2L1* (LPC 16:1), 16p13.11'deki *PDXDC1* (LPC 20:3, PC 34:2, PC 36:3 ve PC 38:3), 19q13.32'deki *APOE* (SPM 24:0 ve SPM 22:0 ile ilişkili), 17p13.2'deki *PLD2*'de (SPM 23:0 ile ilişkili) yeni ilişkiler bulundu. Ayrıca *PNLIPRP2, ZNF600, ALG1, ABHD3, KLF12, DLG2, ILKAP, ITGA9, PLPE, OR8I2, PCDH20, CDK17, SYT9, CDH8, KCNH7, ALG14, PAPD7, CNTNAP4* ve *LPAR2* olmak üzere 18 ek bölge önemli olarak ilişkili bulundu. FADS kümesindeki en yüksek derecedeki SNP (rs102275) CHARGE meta analiz sonuçlarında IMT ile ilişkili bulunmuştur. Aterosklerotik hastalıkların doymamış lipidlerin aksine doymuş lipidlerin sirkülasyonundaki artış ile olduğu bilinmekte, bu SNP'in de yağ asidi desatürasyonunun temel düzenleyicisi olduğu düşünülmektedir

Bölüm 5.1'de araştırmamızı bir adım daha ileri götürüp ilgilendiğimiz iki konu arasındaki fenotipik ilişki incelenmiştir: depresyon ve anksiyete belirtileri ile dolaşımda bulunan lipitler. Çalışmamızda toplam yağ asidi konsantrasyonu veya serum lipoproteinlerinden ziyade 148

adet fosfolipit ve sfingolipit türü analiz edildi. Yaptığımız analizler depresyon ve anksiyete belirtileri ve PC O 36:4 ve SPM 23:1 arasında ters korelasyonun olduğunu göstermiştir. PC O 36:4 ayrıca bağımsız örneklerde de ölçülmüştür. İki örnek birleştirilerek depresyon ve anksiyete belirtileri ile son derece anlamlı ve tutarlı bir ilişki bulunmuştur. Bu ilşiki artmış *PLA2* ya da sfingomiyelinaz aktivitesine bağlı olabileceği gibi bu enzimleri kodlayan genler de depresif semptom araştırmaları için potansiyel adaylardır.

Bu tez, bulunan iki ilşkili sonucun (lipidler ve depresyon) genetik mimarisinin farklı olduğunu göstermektedir. Lipidlerin genetiğinde büyük etkileri olan genler bulunurken depresyon için bunlar bulunamamıştır. 6. Bölümde (genel tartışma) bu fenotiplerin genetik etiyolojisini ve birbiriyle olan ilişkilerini ortaya çıkarmak için gerekli olan adımlar tartışılmıştır.

8

Appendices

Detailed supplementary information regarding the published articles of this thesis can be found in the jounal pages:

Chapter 2.2: European Journal of Human Genetics (2011) 19, 813–819;
http://www.nature.com/ejhg/journal/v19/n7/suppinfo/ejhg201121s1.html?url=/ejhg/journal/v19/n7/full/ejhg201121a.html

Chapter 4.2: February 2012 Issue of PLOS Genetics:
http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002490#s5

CHAPTER 3.1

TABLE S1. Linkage regions with significant and suggestive LOD scores

Region	SNP	cM	Start	End	CES-D	HADS-D
1p36	rs1555024	42.87	rs6426747	rs536766	0.59	2.86
3p14	rs1490265	92.83	rs783512	rs1388276	1.3	2.4
5q34	rs923467	165	rs949602	rs878953	2.47	0.82
9q21	rs927632	72.97	rs2031197	rs722642	0.05	**3.68**
9q32	rs726657	121.5	rs14419	rs230146	1.03	2.26
10q22	rs877783	92.03	rs736594	rs1892498	3.18	0.95
13q33	rs981900	106.5	rs4476030	rs2039120	**3.8**	0.39
16q21	rs1982395	74.24	rs1946155	rs37358	1.35	**3.36**
22q11	rs374225	9.33	rs387399	rs6003823	0.5	2.05

TABLE S2. Association to common SNPs within the linkage intervals

					ERF study		CHARGE	NTR
Region	SNP	MAF	Gene	Distance	P-value	Q-value	P-value^{CHARGE}	P-value^{NTR}
1p36	rs311452	0.25	RUNX3	95	1.64×10^{-6}	2.09×10^{-3}	0.82	0.92
3p14	rs4428187	0.36	FRMD4B	Intronic	1.29×10^{-4}	0.10	0.78	0.92
5q34	rs2272600	0.14	GABRG2	Intronic	2.86×10^{-5}	0.02	0.11	0.29
9q21	rs7034735	0.13	PRUNE2	Intronic	1.42×10^{-5}	7.76×10^{-3}	0.09	0.61
9q32	rs1989775	0.11	PAPPA	262	3.20×10^{-5}	0.02	0.81	0.03
10q22	rs2812541	0.37	C10ORF35	68	2.20×10^{-4}	0.06	0.70	0.14
13q33	rs954580	0.35	DAOA	10	4.40×10^{-4}	0.09	0.90	0.66
16q21	rs9937047	0.31	C16ORF80	2	8.08×10^{-7}	5.93×10^{-4}	0.01	0.89
22q11	rs6005346	0.16	SLC7A4	10	1.30×10^{-4}	0.06	0.18	na

TABLE S3. CHARGE CES-D meta-GWAS Cohorts

		Depressive symptom score											International Standard Classification of Education**				
Sample	Instrument	N	Mean	(SD)	Median	(range)	≥ 16 %*	AD users %	Mean Age	(SD)	Female %	Current smokers %	Level 0/1 %	Level 2 %	Level 3 %	Level 4 %	Level 5/6 %
ARIC1	CES-D 11	393	3.80	(3.57)	3	(0-18)	9.92	14.0	72.7	(5.46)	59.5	19.6	2.0	8.1	35.4	7.9	46.6
ARIC2	CES-D 20	614	8.52	(7.41)	6	(0-34)	16.1	11.1	71.0	(5.60)	49.7	19.7	3.1	8.3	34.7	11.7	42.2
BLSA	CES-D 20	764	6.90	(6.5)	5	(0-55)	8.51	NA	71.6	(13.8)	44.6	3.0	0.4	1.5	11.0	12.4	74.8
CHS	CES-D 10	3155	4.27	(4.29)	3	(0-26)	11.3	3.11	72.2	(5.29)	61.2	11.0	2.5	12.3	38.6	9.3	37.2
ERF	CES-D 20	1297	12.7	(10.9)	10	(0-59)	27.1	8.20	55.9	(10.1)	56.7	43.2	40.4	42.5	13.6	NA	3.5
FHS	CES-D 20	4956	7.25	(8.21)	4	(0-53)	10.3	10.4	56.1	(10.5)	53.3	14.7	0.5	3.1	32.2	24.9	39.2
HABC	CES-D 20	1654	4.93	(5.78)	3	(0-43)	4.70	3.60	73.8	(2.80)	47.1	6.4	11.9	NA	34.4	53.6	NA
InCHIANTI	CES-D 20	942	11.8	(8.24)	10	(0-46)	24.6	3.40	70.4	(9.85)	52.8	18.5	73.5	11.2	7.3	4.6	3.4
RSI	CES-D 20	3791	4.86	(7.35)	2	(0-52)	7.30	3.80	72.7	(7.21)	58.5	16.4	31.4	29.0	29.8	NA	9.8
RSII	CES-D 20	2093	5.81	(7.90)	3	(0-48)	9.70	5.00	64.8	(8.03)	54.5	19.6	21.6	35.6	27.1	NA	15.7
HBCS	CES-D 20	1386	9.58	(8.68)	7	(0-53)	19.4	4.70	63.4	(2.86)	59.7	23.0	33.0	18.4	26.0	NA	22.5
MESA	CES-D 20	2423	6.93	(6.87)	5	(0-50)	10.0	12.2	62.7	(10.2)	52.2	11.4	1.6	3.4	16.5	28.4	50.1
NHS	CES-D 10	5891	6.36	(4.50)	6	(0-26)	15.9	13.3	71.7	(6.70)	100	5.5	0	0	0	72.6	27.4
RSIII	CES-D 20	2041	6.32	(8.22)	3	(0-53)	9.90	6.90	56.0	(5.67)	56.1	22.4	9.8	35.0	28.4	NA	26.8
RUSH MAP	CES-D 10	825	1.38	(1.75)	1	(0-8)	20.1	13.6	80.8	(6.53)	73.0	2.4	1.7	27.4	19.9	42.8	8.2
RUSH ROS	CES-D 10	778	1.10	(1.51)	1	(0-8)	13.9	9.00	75.5	(7.24)	66.5	2.1	1.3	5.4	3.1	46.0	44.2
SardiNIA	CES-D 20	1438	11.9	(8.20)	10	(0-53)	25.2	3.00	58.0	(11.4)	59.5	NA	28.9	50.3	16.1	NA	4.8

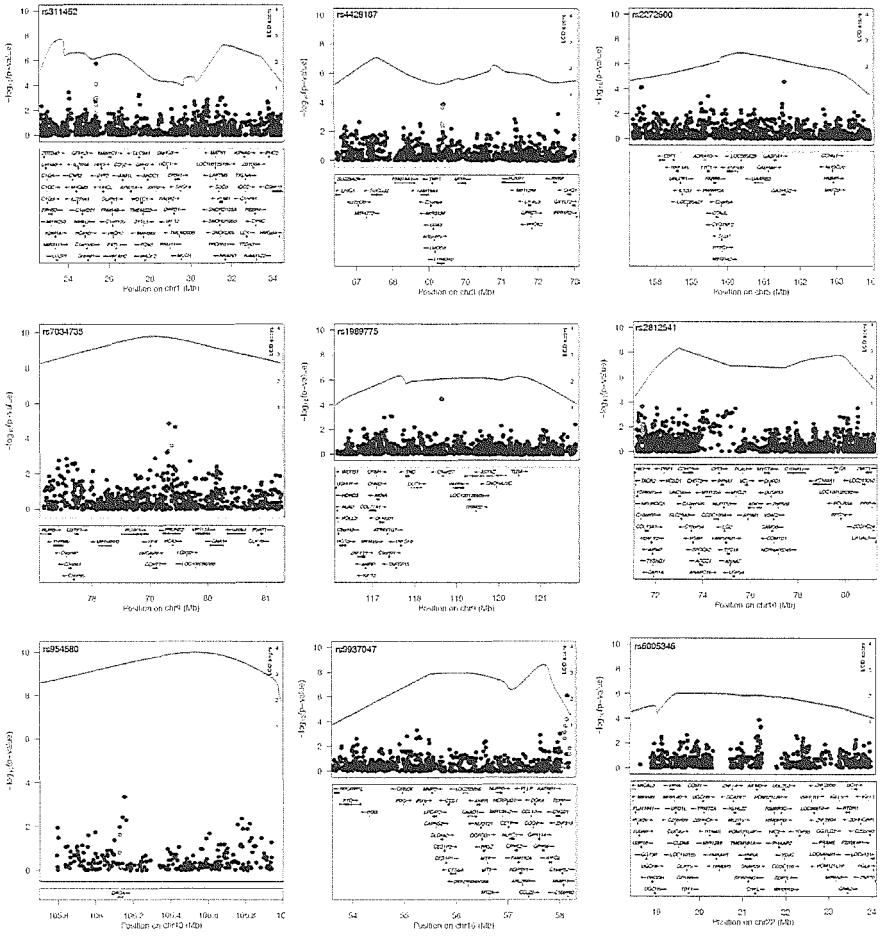
TABLE S4. Selected damaging variants in the coding regions

Chromosome	Position	$Mean_{wt}$	SE_{wt}	$Mean_{mut}$	SE_{mut}	*P*value	Number of observations	Novelty	Gene
1	23845566	0.08	0.21	-4.22	0.67	5.43E-03	10	.	E2F2
1	23885662	-0.01	0.21	0.70	2.34	7.46E-01	5	rs61749352	ID3
1	24192072	-0.05	0.21	4.12	2.24	3.65E-02	6	.	FUCA1
1	24406673	-0.01	0.21	0.61	0.88	6.61E-01	12	.	MYOM3
1	25628113	-0.04	0.21	4.44	2.09	3.97E-02	5	.	RHD
1	26110236	-0.05	0.21	2.81	2.18	6.48E-02	10	rs79507311	MAN1C1
1	26189517	-0.01	0.21	0.60	2.01	7.43E-01	7	rs6689014	PAQR7
1	26288532	0.00	0.21	-0.08	1.77	9.61E-01	9	.	PAFAH2
1	26303228	-0.04	0.21	3.85	1.08	7.50E-02	5	.	PAFAH2
1	26384907	0.05	0.21	-2.79	1.21	8.23E-02	9	rs61749355	TRIM63
1	26620806	0.03	0.21	-0.76	1.18	4.96E-01	18	rs56039743	UBXN11
1	26784304	0.02	0.21	-1.16	1.56	4.97E-01	8	.	DHDDS
1	27223967	0.05	0.21	-3.10	1.34	6.93E-02	8	rs35243557	GPATCH3
1	27690792	-0.09	0.21	3.59	1.33	6.90E-03	13	rs41291098	MAP3K6
1	28477192	-0.03	0.21	2.97	1.11	1.70E-01	5	.	PTAFR
1	28806965	0.04	0.21	-1.05	0.90	3.12E-01	21	rs72661785	PHACTR4
1	29010147	-0.03	0.21	2.33	1.80	2.01E-01	7	.	GMEB1
1	29342245	0.01	0.21	-0.56	1.60	6.63E-01	14	.	EPB41
1	31658150	0.02	0.21	-1.83	2.03	3.15E-01	7	.	NKAIN1
1	31898185	0.02	0.21	-1.03	1.26	5.46E-01	8	.	SERINC2
1	32145693	0.03	0.21	-2.39	1.67	1.90E-01	7	rs41263969	COL16A1
1	32257833	0.01	0.21	-0.84	2.19	7.00E-01	5	.	SPOCD1
1	32667609	0.03	0.21	-1.22	1.37	3.79E-01	12	rs1407134	CCDC28B
1	32682947	-0.01	0.21	0.52	1.65	6.85E-01	14	rs79700000	TMEM234
1	33832933	-0.05	0.21	0.95	1.11	2.88E-01	28	rs10914692	PHC2
1	33836164	-0.04	0.22	0.70	0.81	4.45E-01	27	rs41265897	PHC2
1	34066567	0.02	0.21	-1.72	1.63	3.45E-01	7	rs114879806	CSMD2
1	34164425	-0.03	0.21	2.52	3.00	2.01E-01	6	.	CSMD2
1	34166210	-0.02	0.21	1.05	1.32	5.38E-01	8	.	CSMD2
5	160016684	0.03	0.49	-1.03	3.60	7.29E-01	14	.	ATP10B
5	160042903	0.00	0.49	0.01	4.35	9.97E-01	9	rs61734665	ATP10B
5	160059137	0.00	0.49	-0.13	3.80	9.66E-01	13	rs56340994	ATP10B
5	160061565	0.15	0.49	-4.95	2.47	7.31E-02	16	.	ATP10B
5	160113242	-0.24	0.48	14.27	4.43	1.10E-04	9	.	ATP10B
5	161116672	0.02	0.49	-0.75	2.15	7.78E-01	17	rs3811993	GABRA6
5	162868905	-0.01	0.50	0.19	2.06	9.25E-01	29	rs2069352	CCNG1
9	116811629	-0.05	0.21	2.75	1.56	8.67E-02	9	.	ZNF618

TABLE S4. Selected damaging variants in the coding regions (continued)

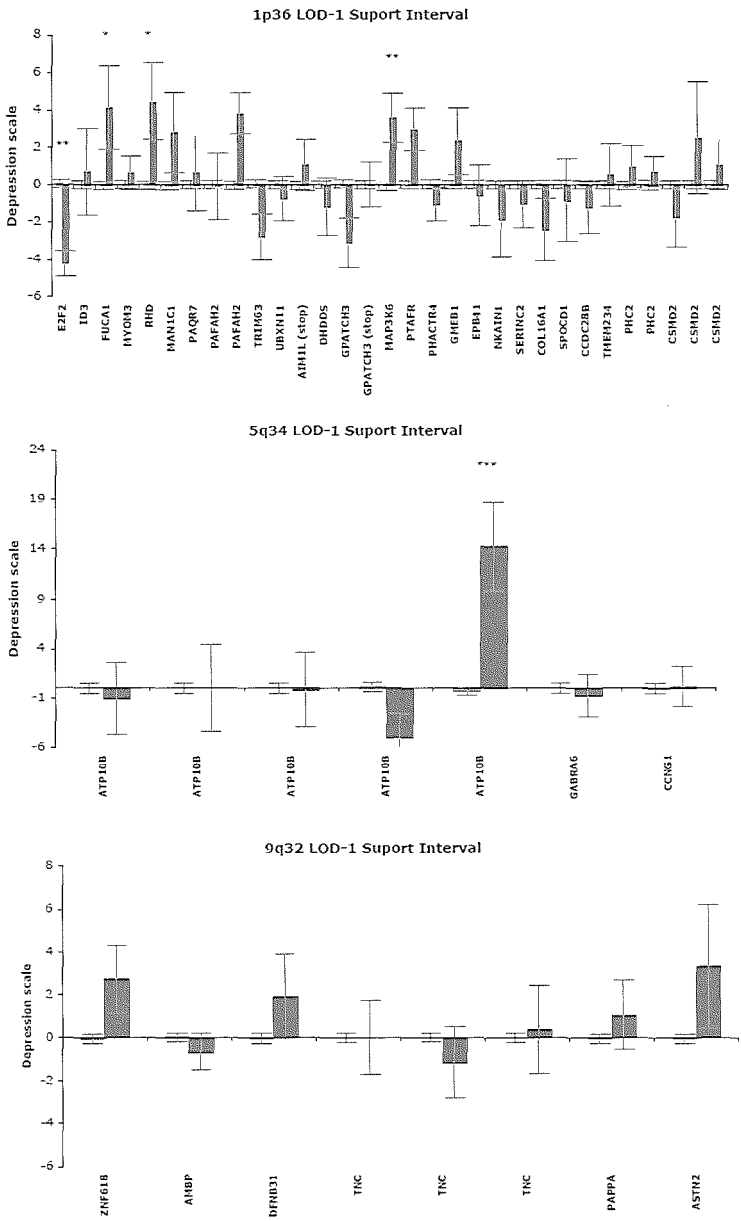
Chromosome	Position	$Mean_{wt}$	SE_{wt}	$Mean_{mut}$	SE_{mut}	*P*-value	Number of observations	Novelty	Gene
9	116832007	0.03	0.22	-0.65	0.84	5.01E-01	24	.	AMBP
9	117165140	-0.03	0.21	1.91	1.99	2.96E-01	7	.	DFNB31
9	117803271	0.00	0.21	0.00	1.72	9.99E-01	14	rs2274750	TNC
9	117808721	0.02	0.21	-1.13	1.66	4.80E-01	9	rs61734387	TNC
9	117848668	-0.01	0.21	0.39	2.05	8.31E-01	7	.	TNC
9	119028233	-0.03	0.21	1.07	1.61	4.03E-01	14	rs117124330	PAPPA
9	119976883	-0.03	0.21	3.33	2.87	1.24E-01	5	.	ASTN2

FIGURE S1. Association with common variants under the linkage regions.



Green line indicates the linkage peak. Dark blue dots indicate the nominal –log (*P*-values) for association of SNPs with depressive symptoms in ERF.

FIGURE S2. Depression scores among carriers of damaging deleterious coding variants under the linkage regions.



Bars indicate residual means and standard errors. * *P*-value < 0.05, ** *P*-value < 0.01, *** *P*-value < 0.001

FIGURE S3. Clustering of *ATP10B* 5:160113242T>C carriers in the single pedigree

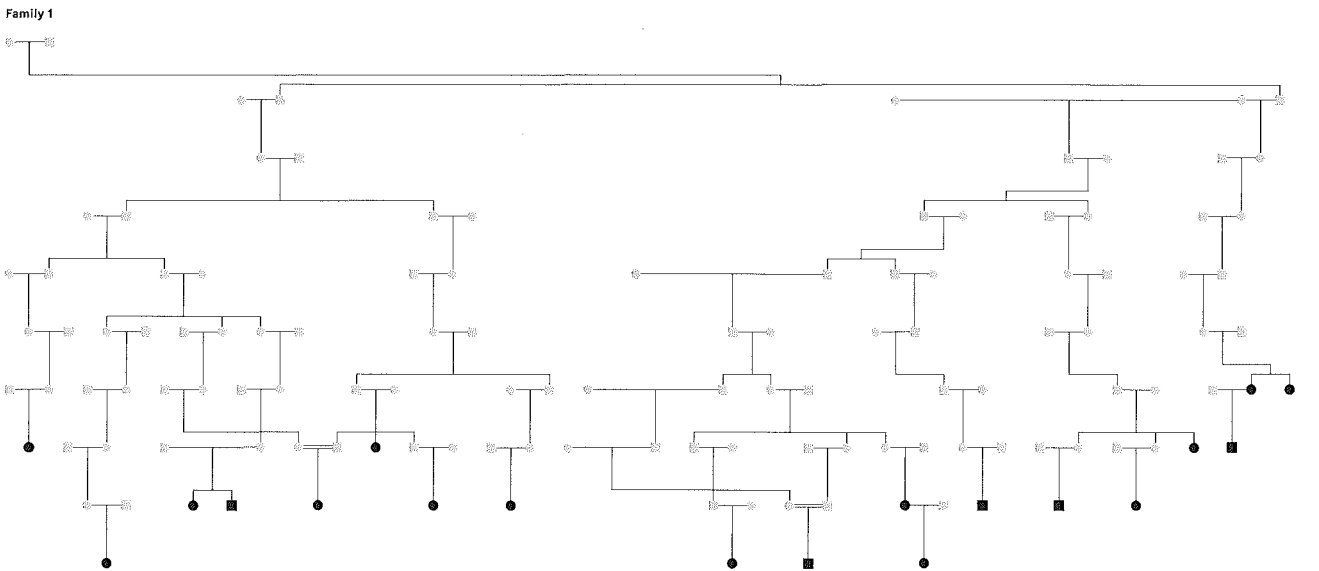
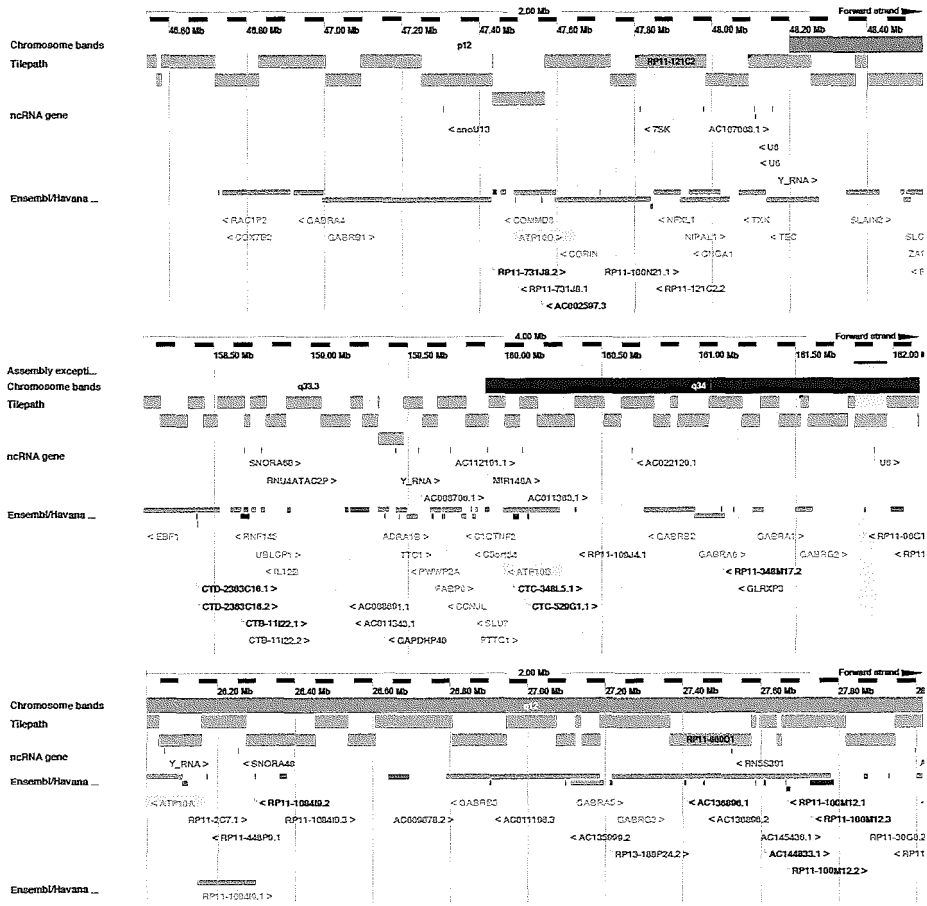


FIGURE S4. Clustering of ATP class V and GABA receptor subunit coding proteins together on chromosomes 4, 5 and 15.



Source: ensembl.org, Ensembl release 68 - July 2012

TABLE S1 Genotyping and Quality Control information

			Sample QC		Imputation QC				Imputation		Analysis	
Cohort	**Platform**	**Calling algorithm**	**Call rate**	**Other exclusions**	**MAF**	**HWE P-value**	**Call rate**	**Other**	**Software**	**Build**	**Software**	**Adjustment***
AGES-RS	Illumina 370 K	BeadStudio	<95%	Sex mismatch, sample failure, genotype mismatch with reference panel	>1%	>10-6	>97%		MACH	HapMap release 22 CEU (build 36)	ProbABEL	
ARIC1	Affymetrix 6.0	Birdseed	<95%	sex mismatch, 1st degree relatives, cryptic relatedness, genotype discordance, outliers in PCA	>1%	>10-6	>95%	-	MACH	HapMap release 22 CEU (build 36)	ProbABEL	study site
ARIC2	Affymetrix 6.0	Birdseed	<95%	sex mismatch, 1st degree relatives, cryptic relatedness, genotype discordance, outliers in PCA	>1%	>10-6	>95%	-	MACH	HapMap release 22 CEU (build 36)	ProbABEL	study site
ARIC3	Affymetrix 6.0	Birdseed	<95%	sex mismatch, 1st degree relatives, cryptic relatedness, genotype discordance, outliers in PCA	>1%	>10-6	>95%	-	MACH	HapMap release 22 CEU (build 36)	ProbABEL	study site
BLSA	Illumina HumanHap 550K	Beadstudio	<98.5 %	Non-European, sex mismatch	≥1%	≥10-4	>99%	-	MACH	HapMap release 22 CEU (build 36)	Merlin-offline, mach2dat	PCA

TABLE S1 Genotyping and Quality Control information (continued)

			Sample QC			Imputation QC			Imputation		Analysis	
Cohort	**Platform**	**Calling algorithm**	**Call rate**	**Other exclusions**	**MAF**	**HWE P-value**	**Call rate**	**Other**	**Software**	**Build**	**Software**	**Adjustment***
CHS	Illumina 370 CNV	BeadStudio	<95%	sex mismatch, sample failure	≥1%	≥10-5	>97%	>2 replicate errors or Mendelian inconsistencies (for reference CEPH trios), heterozygote frequency=0, not in HapMap	BimBam10	HapMap release 21A CEU (build36)	R	site
ERF	Illumina 6k, Illumina 318K, Illumina 370K and Affymetrix 250K	Beadstudio, BRLMM	-	-	>1%	>10-6	>98%	-	MACH	HapMap release 22 CEU (build 36)	Probable, R	Relatedness
FHS	Affymetrix 500K and MIPS 50K combined	BRLMM	<97%	subject heterozygosity >5 SD away from the mean, large Mendelian error rate	>1%	>10-6	>97%	Subject heterozygosity >5 SD away from the mean, large Mendelian error rate	MACH	HapMap release 22 CEU (build 36)	R packages kinship, GEE	Relatedness
HABC	Illumina Infinium Human1M-Duo BeadChip	BeadStudio	-	-	≥ 1%	≥10-6	≥97%	-	MACH	HapMap release 22 CEU (build 36)	R	none
HBCS	Illumina Infinium 610K Quad (Custom modified)	Beadstudio	-	HWE > 10-5	>1%	>10-6	>95%	-	MACH	HapMap release 22 CEU (build 36)	Plink & ProbABEL	none

TABLE S1 Genotyping and Quality Control information (continued)

			Sample QC		Imputation QC				Imputation		Analysis	
Cohort	**Platform**	**Calling algorithm**	**Call rate**	**Other exclusions**	**MAF**	**HWE P-value**	**Call rate**	**Other**	**Software**	**Build**	**Software**	**Adjustment***
InCHIANTI	Illumina HumanHap 550K	Beadstudio	-	gender mismatch, IBD analysis to exclude related individuals	>1%	>10-4	>98%	-	MACH	HapMap release 22 CEU (build 36)	Mach2qtl & Mach2dat	none
MESA	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed	<95%	Sample Level: gender mismatches, cryptic duplicates, and SNP level: monomorphic SNPs, SNPs with observed heterozygosity > 53%, and SNPs with missing rate > 5%	>1%	>10-4	>95%		IMPUTE	NCBI Build 36	SNPTest	first 10 PCs
MKF3	Affymetrix 500K	BRLMM	<93%	sex mismatch	-	-	-	-	MACH	HapMap release 21 CEU (build 35)	ProbAbel	none
MKF4	Affymetrix 6.0 (1000K)	Birdseed2	<93%	sex mismatch	-	-	>93%	-	MACH	HapMap release 22 CEU (build 36)	ProbAbel	none
NHSBrC	Illumina 550K	Beadstudio	<90%	SNP QC: MAF <0.01; Sample QC: call rate <90%, -duplicates and first/second degree relatives -ancestry outliers	-	-	-	-	MACH	HapMap release 22 CEU (build 36)	ProbABEL	disease status, top 4 EVs

TABLE S1 Genotyping and Quality Control information (continued)

			Sample QC		Imputation QC				Imputation		Analysis	
Cohort	**Platform**	**Calling algorithm**	**Call rate**	**Other exclusions**	**MAF**	**HWE P-value**	**Call rate**	**Other**	**Software**	**Build**	**Software**	**Adjustment***
NHSchd	Affymetrix 6.0	Birdseed	<98%	SNP QC: pHWE<10E-4, MAF <0.02 ; Sample QC: call rate <98%, --sex discrepancy with genetic data from X-linked markers -duplicates and first/second degree relatives -ancestry outliers -heterozygosity -missing phenotype information	-	-	-	-	MACH	HapMap release 22 CEU (build 36)	ProbABEL	disease status, top 3 EVs
NHSkid	Illumina 610Q	Beadstudio	<95%	SNP QC: MAF <0.01, pHWE<10E-5; Sample QC: call rate <95%, -duplicates and first/second degree relatives -ancestry outliers	-	-	-	-	MACH	HapMap release 22 CEU (build 36)	ProbABEL	disease status, top 4 EVs

TABLE S1 Genotyping and Quality Control information (continued)

			Sample QC		Imputation QC				Imputation		Analysis	
Cohort	**Platform**	**Calling algorithm**	**Call rate**	**Other exclusions**	**MAF**	**HWE P-value**	**Call rate**	**Other**	**Software**	**Build**	**Software**	**Adjustment***
NHSt2d	Affymetrix 6.0	Birdseed	<98%	SNP QC, pHWE<10E-4, MAF <0.02, >1 discordance/12 replicates, significant plate associations; Sample QC: call rate <98%, -sex discrepancy with genetic data from X-linked markers -duplicates and first/second degree relatives -ancestry outliers -heterozygosity -autosomal chromosome abberations	-	-	-	-	MACH	HapMap release 22 CEU (build 36)	ProbABEL	disease status, top 3 EVs
RS1	Version 3 Illumina Infinium II HumanHap 550 SNP chip array	Beadstudio	<98%	sex mismatch, excess autosomal heterozygosity >0,336, outliers identified by IBS clustering analysis	≥1%	>10-6	≥98%	-	MACH	HapMap release 22 CEU (build 36)	SPSS, Probabel, GRIMP, R	none

TABLE S1 Genotyping and Quality Control information (continued)

			Sample QC		Imputation QC				Imputation		Analysis	
Cohort	**Platform**	**Calling algorithm**	**Call rate**	**Other exclusions**	**MAF**	**HWE P-value**	**Call rate**	**Other**	**Software**	**Build**	**Software**	**Adjustment***
RS2	Version 3 Illumina Infinium II HumanHap 550 SNP chip array	Beadstudio	<98%	sex mismatch, excess autosomal heterozygosity >0,336, outliers identified by IBS clustering analysis	≥1%	>10-6	≥98%	-	MACH	HapMap release 22 CEU (build 36)	SPSS, Probabel, GRIMP, R	none
RS3	Version 3 Illumina Infinium II HumanHap 550 SNP chip array	Beadstudio	<98%	sex mismatch, excess autosomal heterozygosity >0,336, outliers identified by IBS clustering analysis	≥1%	>10-6	≥98%	-	MACH	HapMap release 22 CEU (build 36)	SPSS, Probabel, GRIMP, R	none
RUSH (MAP)	Affymetrix 6.0	Birdsuite	<95%	genotype-derived gender discordant with reported gender, inbreeding coefficient F>0.04	≥1%	>10-6	>95%	Exclusion: misshap test <10-9	MACH	HapMap release 22 CEU (build 36)	SAS, ProbABEL, R	First 3 principal components of eigenstrat
RUSH (ROS)	Affymetrix 6.0	Birdsuite	<95%	genotype-derived gender discordant with reported gender, inbreeding coefficient F>0.04	≥1%	>10-6	>95%	Exclusion: misshap test <10-9	MACH	HapMap release 22 CEU (build 36)	SAS, ProbABEL, R	First 3 principal components of eigenstrat
SardiNIA	Affymetrix 10K, 500K, 1000K	BRLMM	<95%	sex mismatch, Mendelian error	>1%	>10-6	>90%	-	MACH	HapMap CEU (build 36)	MERLIN	dummy variable: self-report vs. tester read and reported answers

TABLE S1 Genotyping and Quality Control information (continued)

			Sample QC		Imputation QC				Imputation		Analysis	
Cohort	**Platform**	**Calling algorithm**	**Call rate**	**Other exclusions**	**MAF**	**HWE P-value**	**Call rate**	**Other**	**Software**	**Build**	**Software**	**Adjustment***
SHIP	Affymetrix Human SNP Array 6.0	Birdseed V2	< 92%	duplicate samples (by IBS), reported/ genotyped gender mismatch	-	-	-	-	IMPUTE	HapMap release 22 CEU (Build 36)	SNPTEST, QUICKTES, R, PLINK	none

TABLE S2: Additional top SNP information

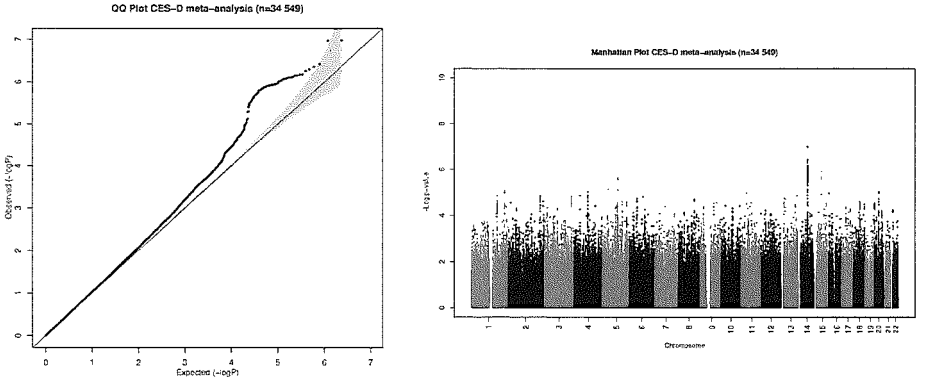
SNP	# genotyped*	# imputed**	r^{2***}
rs13137117	1	24	0.91
rs161645	0	24	0.97
rs357282	12	13	0.93
rs40465	1	23	0.93
rs4594522	9	16	0.97
rs4653635	1	24	0.99
rs8020095	9	14	0.91
rs8038316	0	24	0.98

* Number of cohorts that genotyped this SNP

** Number of cohorts that did not genotype this SNP

*** Observed versus expected variance ratio (measure of imputation quality) r^2 is based on SNPs that had not been genotyped.

FIGURE S1: Genome-wide association study results for depressive symptoms in the discovery sample.



Quantile-Quantile (A) and Manhattan (B) plot of total depressive symptoms score meta-analysis of discovery samples (N = 34 549)

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258).

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs40465	t	g	0.652	49820	-5.459	4.78E-08	----------+?---+-	0	0.601
rs161645	a	g	0.337	49820	5.359	8.39E-08	+++++++++++++++++?++++	0	0.752
rs6421926	t	c	0.338	49820	5.327	9.96E-08	+++++++++++++++?++++	0	0.739
rs60271	a	c	0.339	51258	5.155	2.54E-07	++++++++++++++++++++	0	0.652
rs1383605	a	t	0.206	51258	4.766	1.88E-06	+++++++++++++++++++	0	0.723
rs10279132	t	g	0.715	51256	4.735	2.19E-06	++++++++++++- ++++++++++	0	0.514
rs2242277	t	c	0.793	51258	-4.727	2.28E-06	-+------+-----++------	0	0.732
rs12679544	t	c	0.793	51258	-4.685	2.80E-06	-+------+-----++------	0	0.688
rs8020095	a	g	0.161	40902	4.668	3.04E-06	+++++++++++++++?-?--+	33.5	0.069
rs7152001	c	g	0.839	40902	-4.665	3.08E-06	---+------+---+?+?++-	33.1	0.072
rs11914750	t	c	0.671	51258	-4.542	5.58E-06	---++-+----------+--	0	0.527
rs1008813	a	g	0.519	51255	4.535	5.77E-06	+++++++++++++++- +++++++	15.1	0.256
rs1008812	a	g	0.482	51257	-4.534	5.80E-06	----+---+----+------	15.5	0.250
rs1976423	a	c	0.502	47397	-4.533	5.82E-06	-----+--+?-+--+?---+-	10.6	0.321
rs17026230	a	g	0.330	51258	4.531	5.86E-06	+++-+++++++++++++++	0	0.495
rs1873213	t	g	0.969	45537	-4.525	6.05E-06	+--??---?--?+?-+-----+?	25.1	0.165
rs8072065	a	g	0.829	51257	-4.523	6.10E-06	-------+++----------	0	0.679
rs1008814	a	t	0.519	51257	4.521	6.16E-06	++++-+++-++++++- +++++++	15.3	0.253
rs8000066	t	c	0.519	51257	4.506	6.60E-06	++++-+++-++++++- +++++++	14.5	0.263
rs7587554	t	c	0.474	51258	4.498	6.85E-06	+++---+++++++++++++++	0.9	0.447
rs10958604	t	g	0.786	51258	-4.497	6.88E-06	-+------+----++------	0	0.736
rs7339176	a	g	0.518	51257	4.493	7.01E-06	++++-+++-++++++- +++++++	15	0.257
rs12452091	c	g	0.830	51257	-4.492	7.07E-06	-------+++----------	0	0.677
rs12451111	t	c	0.169	51257	4.473	7.72E-06	++++++++-+-+++++++++- ++	0	0.750
rs12793618	a	g	0.043	41709	-4.469	7.86E-06	??---+-+------?--+-+	0	0.883
rs9900677	a	g	0.830	51257	-4.467	7.94E-06	-------++----------	0	0.694
rs17488749	a	g	0.145	49818	-4.461	8.16E-06	-----+-----++---?-----	39	0.033
rs17488784	a	t	0.145	49818	-4.458	8.28E-06	-----+-----++---?-----	38.9	0.033
rs9468252	a	g	0.969	45537	-4.454	8.42E-06	+--??---?--?+?-+-----+?	18.7	0.235
rs11784532	t	c	0.785	51258	-4.45	8.58E-06	-+------+----++------	0	0.656
rs12451588	c	g	0.832	51257	-4.44	8.98E-06	-------+-+---------+--	0	0.753
rs8038316	a	g	0.050	48103	-4.425	9.64E-06	-?-------+----+-----+	19.4	0.204
rs1592757	c	g	0.364	51258	4.406	1.05E-05	++++++-++--+++-++++--	28.7	0.099
rs1503389	t	c	0.931	48099	4.405	1.06E-05	+?+++++++++++++++++++	0	0.631
rs12452510	a	t	0.166	51257	4.402	1.07E-05	+++++++++-+- +++++++++++	0	0.813
rs10091355	t	c	0.846	51258	-4.386	1.16E-05	-+------+----++------	0	0.679
rs6900413	a	g	0.031	47317	4.382	1.18E-05	-++??+++?++--?+-++++++?	23	0.182

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258). (continued)

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs12449501	a	g	0.167	51253	4.371	1.24E-05	+++++++++- ++++++++++++	0	0.815
rs2409064	a	g	0.932	48103	4.36	1.30E-05	+?+++-++-++-+++-++++++	0	0.640
rs16966168	a	g	0.150	50250	4.359	1.31E-05	+++??++-+++++-++++++	0	0.551
rs7485858	t	c	0.369	50251	-4.349	1.37E-05	+--??----------+	0	0.728
rs1421669	c	g	0.628	51258	-4.334	1.47E-05	----+-+++-+-+--+	28.8	0.098
rs2168312	a	g	0.261	51258	4.333	1.47E-05	++++-++-++++++-++++	11.2	0.309
rs7978337	a	t	0.369	50250	-4.329	1.50E-05	+--??----------+	0	0.730
rs2139680	a	t	0.369	50251	-4.316	1.59E-05	+--??----------+	0	0.722
rs2447838	t	c	0.436	51258	4.311	1.63E-05	+++++++-++++++++-+++-	29.3	0.094
rs9535050	a	g	0.478	51257	4.299	1.72E-05	+++-+-++++++-+++++++-	0	0.827
rs2312971	c	g	0.516	51257	-4.293	1.76E-05	---++-+----++-+----+	0	0.683
rs9959343	t	c	0.261	51258	4.293	1.77E-05	+-++-++-++-++++--+++-+	13.4	0.278
rs10101533	a	g	0.151	51258	4.292	1.77E-05	+-++++++++++++--++++++	0	0.691
rs4942783	c	g	0.522	51257	-4.282	1.85E-05	---++-+----++-+----+	0	0.836
rs4754128	a	g	0.932	48102	4.282	1.86E-05	+?+++-++-++++-++++++	0	0.617
rs7004479	t	c	0.154	51258	4.28	1.87E-05	+-++++++-++++--++++++	0	0.670
rs6493686	c	g	0.950	48103	4.278	1.89E-05	+?++++++++++-+-+--+++-	22.6	0.167
rs8033074	a	c	0.950	48103	4.276	1.90E-05	+?++++++++-++-+--+++-	22.4	0.169
rs7107383	a	t	0.072	48103	-4.276	1.91E-05	-?---+-+-+---+----	0	0.730
rs937055	t	c	0.919	51258	-4.267	1.99E-05	-----+-++-+---+-++	11.9	0.298
rs254035	a	t	0.437	51258	4.264	2.01E-05	+++++++-++++++-+-+++-	31.1	0.079
rs12453488	a	g	0.173	51254	4.262	2.02E-05	+++++++++-+++++++++- ++	0	0.670
rs2447832	t	c	0.437	51258	4.259	2.05E-05	+++++++++++++++++-+++-	30.9	0.080
rs323105	c	g	0.965	43679	-4.247	2.17E-05	-?-----?-+++--???+-++	27.4	0.136
rs2312972	t	c	0.521	51257	-4.245	2.18E-05	---++-+----++-+----+	0	0.845
rs1520550	a	g	0.631	51258	4.24	2.23E-05	-++-+++++++++++++++-	0	0.762
rs7833452	a	g	0.847	51258	-4.236	2.27E-05	-+-----+----++----	0	0.702
rs10785027	a	t	0.369	51258	-4.234	2.30E-05	+--++----------+	0	0.757
rs2077781	t	g	0.261	51258	4.228	2.36E-05	+-++++-++++++-++++	11.9	0.298
rs33817	a	g	0.437	51258	4.226	2.38E-05	+++++++-++++++-++++-	27.1	0.115
rs8030855	c	g	0.056	51258	4.221	2.43E-05	+++++-++++++-+-++++++	0	0.599
rs16870152	t	c	0.843	51256	4.219	2.46E-05	++++++-++++- +++++++++	31.7	0.074
rs2276203	a	g	0.261	51258	4.214	2.51E-05	+-++-++-++++++-++++	12.7	0.288
rs7182991	t	g	0.055	51258	4.208	2.58E-05	+++++-++++++-+-++++++	0	0.684
rs4636213	a	g	0.200	51258	4.207	2.59E-05	+++++-+-++++++-+++++++++	0	0.966
rs7182611	c	g	0.055	51258	4.2	2.67E-05	+++++-++++++++-+++++++	0	0.679
rs185260	a	c	0.563	51258	-4.197	2.71E-05	-----+-----+-+--++	26.7	0.118
rs4489949	a	g	0.944	51258	-4.194	2.74E-05	----+-----+-+---+-	0	0.605
rs2414196	c	g	0.738	51258	4.191	2.78E-05	+-++++-++++++++++++-	35.7	0.047
rs1356893	t	c	0.605	51258	4.185	2.85E-05	-++-++-++++++++++++-	0	0.603

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258). (continued)

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs254025	c	g	0.563	51258	-4.184	2.87E-05	----+-----+---++	27.8	0.107
rs7177816	a	g	0.054	51258	4.183	2.88E-05	+++++++++++++++++++	0	0.679
rs11664693	c	g	0.262	51258	4.183	2.88E-05	+-++-++-++-++++-++++	6.7	0.369
rs12667152	t	g	0.156	49818	-4.182	2.88E-05	----+---++--?----	37.3	0.041
rs4941210	t	c	0.262	51258	4.179	2.92E-05	+++++++++++++-++++	8.5	0.346
rs254020	t	c	0.438	51258	4.178	2.94E-05	+++++++-++++++++-+++-	27.6	0.110
rs1073839	t	g	0.850	51257	-4.176	2.96E-05	---+-+---+-+-----	0	0.613
rs1353416	a	g	0.355	51258	4.176	2.97E-05	+++-+++++++++++++-++	0	0.480
rs2414195	a	g	0.738	51258	4.175	2.98E-05	++++++-++++++++++++-	36.9	0.040
rs11927424	a	g	0.644	51258	-4.172	3.02E-05	---++-+--------+-	0	0.522
rs9596054	a	g	0.478	51257	4.17	3.04E-05	+++-+-++++-+++++++-	0	0.768
rs3111816	a	g	0.352	51239	-4.17	3.04E-05	----+--------+	0	0.964
rs4073665	a	g	0.873	51258	-4.169	3.05E-05	----+++-----+---	26.1	0.124
rs1397164	a	c	0.637	49820	4.168	3.08E-05	-++++--++++++++?++++	22.3	0.170
rs7170422	t	c	0.945	51258	-4.167	3.08E-05	---+-----+-+---+-	0	0.664
rs10851526	t	c	0.261	51258	-4.167	3.09E-05	-+---+-----+---++	35.1	0.050
rs11977246	t	c	0.844	49818	4.166	3.10E-05	++++++-++++-+++?+++++	37.4	0.041
rs254023	t	c	0.563	51258	-4.166	3.11E-05	----+-----+-+--++	27.6	0.109
rs2919955	a	g	0.352	51257	4.165	3.12E-05	++++++-++-++++++-++++	0	0.660
rs7641985	a	c	0.644	51258	-4.163	3.14E-05	---++-+--------+-	0	0.476
rs13155692	t	c	0.748	49820	-4.161	3.18E-05	----------+++-?---++	17.1	0.233
rs9554349	t	c	0.033	38528	4.16	3.18E-05	+?+++-++++++++++??+??+	0	0.517
rs11147450	t	c	0.521	51257	-4.16	3.18E-05	---++-+----++-+-----+	0	0.795
rs2125659	t	c	0.946	51258	-4.158	3.21E-05	----+-----+-+---+-	0	0.698
rs10505424	a	c	0.199	51258	4.156	3.24E-05	+++++-+-++++-+++++++++	0	0.977
rs17750582	a	c	0.055	51258	4.155	3.25E-05	+++++-++++++++-+++++++	0	0.659
rs17553281	a	t	0.822	51257	-4.155	3.25E-05	---++-+----+----+---+	23.5	0.152
rs1442111	t	g	0.352	51258	4.155	3.26E-05	+++++++++-++++++++++	0	0.721
rs10879604	a	t	0.374	50251	-4.155	3.26E-05	+--??-+---------+	0	0.670
rs1587150	a	t	0.355	51258	4.154	3.27E-05	+++-+-++++++++++++-++	0	0.481
rs12955929	t	g	0.958	46834	-4.153	3.28E-05	-------?-++--???+-++-	16	0.258
rs8026763	t	c	0.262	51239	-4.152	3.30E-05	-+---+-----+---++	35.5	0.048
rs7308693	t	c	0.631	51258	4.149	3.33E-05	-++-++++++++++++-++++-	0	0.724
rs1011947	a	g	0.262	51258	4.148	3.36E-05	+++++++++++++++++++	5	0.393
rs10785028	t	g	0.631	51258	4.147	3.38E-05	-++-++++++++++++-+++-	0	0.723
rs7137885	t	c	0.368	51258	-4.144	3.42E-05	+--++--------+---+	0	0.735
rs2161097	t	c	0.438	51258	4.143	3.43E-05	+++++++++++++++++++-	29.4	0.093
rs7555997	a	c	0.101	50232	-4.141	3.46E-05	---??-+-+-+-------	0	0.808
rs2447828	t	c	0.438	51258	4.138	3.51E-05	+++++++-++++++-+++--	29.9	0.088
rs11927001	a	g	0.869	51258	4.131	3.62E-05	-++-++++-++++++++++	27.2	0.113
rs2125657	t	c	0.054	51258	4.131	3.62E-05	+++++++++++++-++++++	0	0.695
rs2112163	t	c	0.438	51258	4.13	3.62E-05	+++++++-++++++++-+++-	30.2	0.086
rs6551366	c	g	0.355	51258	4.128	3.66E-05	+++-+-++++++++++++++	0	0.487

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258). (continued)

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs1589595	t	c	0.738	51258	-4.128	3.66E-05	-+-+-+-+---+--+-	5.8	0.382
rs1522116	t	c	0.441	51257	4.126	3.69E-05	+++++-++++++-+++-++++	0	0.791
rs920623	a	g	0.352	51258	4.125	3.71E-05	+++++++-++-++++++++++	0	0.649
rs7205464	t	g	0.097	51256	4.122	3.75E-05	+++++++++++++++++--++++	0	0.945
rs7295470	t	c	0.372	51258	-4.122	3.75E-05	+--++----------+	0	0.723
rs1426134	c	g	0.353	51258	-4.12	3.78E-05	---------+---------+	0	0.915
rs16893023	a	c	0.194	51258	4.119	3.80E-05	++++-+-+++++-+++++++++	0	0.952
rs12050204	t	g	0.080	51252	-4.119	3.81E-05	----+--+-++-------+-	0	0.759
rs9903859	a	g	0.160	51257	4.118	3.82E-05	++++++++-+-+++++++++ ++	0	0.495
rs1045301	t	g	0.082	51258	-4.111	3.94E-05	----+--+-++-------+-	0	0.810
rs2337127	a	c	0.946	51258	-4.11	3.95E-05	-----+------+-+---+-	0	0.698
rs1687128	t	g	0.081	51258	4.105	4.05E-05	+++++++-+-++++++++++-	9	0.339
rs9303295	a	g	0.850	51257	-4.103	4.09E-05	---++--+----+-+------	0	0.654
rs4776080	t	c	0.269	51258	-4.101	4.12E-05	-+---++----------++	34.6	0.053
rs10879605	t	c	0.374	51258	-4.1	4.13E-05	+--++--+----------+	0	0.744
rs2414218	t	c	0.276	51257	-4.1	4.13E-05	-+----+----------++	31.8	0.073
rs325501	c	g	0.587	51258	-4.097	4.19E-05	-----+------+---++	17.3	0.227
rs2414217	t	c	0.276	51258	-4.093	4.26E-05	-+----+----------++	30.6	0.083
rs325481	a	g	0.586	51258	-4.092	4.27E-05	-----+------+---++	19.4	0.200
rs1583953	t	c	0.353	51258	4.087	4.37E-05	++++++++-+-++++++++-+	0	0.858
rs988542	a	g	0.628	51258	4.086	4.39E-05	-++--++-+++++++++++++-	0	0.759
rs1106420	a	t	0.318	51258	-4.086	4.39E-05	++--+-++----+-+----	1.4	0.442
rs8079016	t	c	0.161	51257	4.086	4.40E-05	++++++++-+-+++++++++- ++	0	0.494
rs7953276	a	c	0.377	51258	-4.084	4.43E-05	---++--+----------+	0	0.868
rs12955292	a	g	0.036	46834	4.083	4.44E-05	+++++++++?+--++???-+--+	22.3	0.184
rs768792	a	g	0.647	51258	-4.083	4.45E-05	-----+-++------+---+-	0	0.868
rs1687119	a	g	0.919	51258	-4.08	4.51E-05	-----+-++--+---+-+-+	8.9	0.339
rs9535127	t	c	0.481	51257	4.076	4.59E-05	+++-++-++++-+-++++++-	0	0.864
rs10748226	t	g	0.628	51258	4.075	4.60E-05	-++-+++++++++++++++++-	0	0.778
rs4760780	t	c	0.629	51258	4.075	4.60E-05	-++-++++++++++++++++-	0	0.740
rs6445194	t	g	0.489	51258	-4.075	4.61E-05	-+---+----+-+-+---	5.9	0.381
rs12441046	t	g	0.724	51258	4.072	4.65E-05	+++++++++++++++++++-	30.9	0.080
rs13177473	a	g	0.647	51258	-4.068	4.75E-05	-----+-++------+---+-	0	0.852
rs1542727	a	g	0.174	51258	4.065	4.81E-05	+++++-+++-++-+++++++++	0	0.536
rs2203976	t	c	0.628	51258	4.062	4.87E-05	-++-++++++++++++++++-	0	0.738
rs6582151	t	c	0.371	51258	-4.06	4.91E-05	+--++-+-------+---+	0	0.728
rs2139675	t	g	0.628	51258	4.059	4.93E-05	-++-++++++++++++++++-	0	0.737
rs13250310	a	t	0.732	48103	-4.056	4.99E-05	+?--+++------+----	8.7	0.344
rs1394309	a	g	0.932	48437	-4.053	5.05E-05	---+---?+++--++?--++-	38.6	0.038
rs139265	a	g	0.833	51256	-4.049	5.15E-05	+--+------+-+-+---+	0.4	0.454
rs16955611	a	g	0.949	51258	-4.044	5.25E-05	-----+------+-+----+-	0	0.773

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258). (continued)

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs7899547	t	g	0.360	51258	4.043	5.27E-05	+-++++++++-++-+++++++++	0	0.918
rs2028526	t	c	0.648	51251	-4.042	5.31E-05	---+--++----+--+-	0	0.748
rs9645898	a	c	0.706	50250	-4.041	5.33E-05	---??----+-+-------	0	0.784
rs3922857	t	c	0.097	51256	4.039	5.36E-05	+++++++++++++++---++++	0	0.921
rs2363065	t	c	0.628	51258	4.039	5.37E-05	-++--++-+++++++++++++++-	0	0.744
rs7976937	t	c	0.361	51258	-4.039	5.38E-05	+---+--+---------+	0	0.862
rs325485	a	g	0.388	51257	4.037	5.42E-05	+++++++-++-+-++++-+++--	12.8	0.287
rs6964185	a	g	0.126	51256	-4.032	5.53E-05	-+---+-+--+-++++----	0	0.675
rs2836021	t	c	0.836	51256	4.03	5.59E-05	+--+-+++++++++-++-+++-++	0	0.788
rs11179680	a	g	0.362	51258	-4.027	5.64E-05	+---+--+----------+	0	0.864
rs10984257	t	c	0.869	51258	-4.024	5.72E-05	----+---++----+-------	37.6	0.037
rs11683777	t	c	0.151	44606	4.022	5.77E-05	?+++++++++-+++++++++-?	0	0.823
rs12824659	a	g	0.333	51258	-4.021	5.79E-05	------+----------+	0	0.954
rs4889796	t	c	0.781	49819	4.02	5.83E-05	-+++++++++++++-?+++++	26.8	0.121
rs11611073	a	g	0.362	51258	-4.018	5.86E-05	+---+--+----------+	0	0.865
rs508760	t	g	0.070	48103	-4.018	5.87E-05	-?--+------+++-+---	0	0.631
rs7975033	a	t	0.372	51258	-4.015	5.94E-05	+--++--+-------+---+	0	0.704
rs6875442	a	c	0.625	51258	4.012	6.01E-05	+++++++-++++-+++-++++	0	0.958
rs9365900	a	g	0.071	40276	4.011	6.05E-05	??+-+-++++++++-+++?-+?++	23.5	0.171
rs2211846	a	g	0.165	50250	-4.01	6.08E-05	-+-??------+-----+-	0	0.769
rs10742719	a	c	0.525	50249	-4.008	6.13E-05	--+-??--+-------+-+---+-	0	0.590
rs2043475	t	c	0.168	51257	4.007	6.14E-05	++++++++++-++++++-++-++	0	0.604
rs2995807	t	c	0.140	51256	-4.007	6.15E-05	+----------+--+---	0	0.948
rs7964705	t	c	0.629	51258	4.006	6.17E-05	-++-++++++++++++++++-	0	0.697
rs7199995	t	c	0.140	51256	4.005	6.20E-05	++++++++-++++++-+-++++	0	0.964
rs2836012	a	t	0.165	51255	-4.005	6.21E-05	-+-++------+-----+--	0	0.833
rs2296561	a	t	0.879	49820	-4.005	6.22E-05	---------+---+-?+++-+	32.9	0.069
rs2084919	t	c	0.515	51258	4.004	6.23E-05	+-+++-+-+-++++++++-+	1.6	0.439
rs1915293	t	c	0.629	51258	4.004	6.24E-05	-++-++-+++++++++-++++-	0	0.699
rs12367407	a	t	0.637	51258	4.002	6.27E-05	-+++-+-+++++++++++++-	0	0.858
rs10942087	t	c	0.637	51258	4	6.33E-05	+++++-+-++++++-+++-++++	0	0.832
rs323097	a	c	0.959	46834	-4	6.33E-05	-------?-++--???+-++-	13.7	0.286
rs2047268	a	c	0.363	51258	-3.999	6.36E-05	----+-+----+--+----	0	0.826
rs13157155	c	g	0.626	51258	3.998	6.38E-05	+++++++-++++-+++-++++	0	0.958
rs2148473	a	g	0.869	51258	-3.998	6.39E-05	---+---++-+--+------	39.2	0.029
rs11917572	t	c	0.959	47020	3.997	6.42E-05	+++++-+-+?+-+--+++?+-+-	0	0.605
rs4627955	a	g	0.625	51258	3.996	6.43E-05	+++++++-+++-+-+++-+++++	0	0.955
rs12654558	t	c	0.636	51258	3.996	6.45E-05	+++++-+-++++-+++-++++	0	0.830
rs2836014	a	c	0.165	51257	-3.995	6.48E-05	-+-++------+-+---+-	0	0.806
rs1472763	c	g	0.318	50246	3.994	6.49E-05	+++??++++-++-++++--++++	0	0.684
rs12657561	t	c	0.364	51258	-3.992	6.55E-05	----+-+----+---+----	0	0.831
rs10984272	a	t	0.870	51258	-3.991	6.57E-05	----+--+--+-+-----	40.5	0.024

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258). (continued)

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs10984285	t	c	0.124	51258	3.99	6.61E-05	++++-+++-+-++-+++++++++	39.9	0.026
rs9516233	t	c	0.648	51257	3.987	6.69E-05	+++++-+++-+---++++++++	31.1	0.078
rs1443737	a	t	0.363	51258	-3.987	6.70E-05	+---+-+----------+	0	0.860
rs13013073	c	g	0.388	43950	-3.986	6.73E-05	-?---?-+-+-+---?-++	0	0.552
rs7825010	a	g	0.842	51258	-3.985	6.74E-05	-+------++-----++------	0	0.689
rs11179697	a	g	0.363	51258	-3.985	6.74E-05	+---+-+----------+	0	0.851
rs13163964	a	g	0.625	51258	3.983	6.80E-05	+++++++-+++-+-++-+++++	0	0.953
rs325506	c	g	0.433	49820	3.983	6.81E-05	+++++++-+-++++++?++++-	10.5	0.320
rs6551361	t	c	0.645	51258	-3.981	6.87E-05	---++-+------+-----+--	0	0.541
rs11038193	a	t	0.529	50250	-3.979	6.91E-05	-+-??-+---+--+-+--+-	0.6	0.450
rs9524069	a	c	0.648	51257	3.978	6.94E-05	+++++-++++----++++++	31.5	0.075
rs13339086	a	g	0.097	51256	3.977	6.99E-05	++++++++-++++++++----++++	0	0.930
rs995431	a	t	0.637	51258	3.976	7.00E-05	+++++-+++++-+++-++++++	0	0.831
rs9949310	t	c	0.959	46834	-3.976	7.02E-05	-------?-++--???+-++-	13.4	0.291
rs4072224	a	g	0.903	51256	-3.975	7.05E-05	-------+------+++----	0	0.860
rs139263	a	g	0.166	51256	3.972	7.13E-05	-++++++++++-+-++++++-	0	0.465
rs10413178	t	c	0.094	51258	-3.972	7.14E-05	---+--+---+----------	0	0.921
rs11179681	a	g	0.638	51258	3.972	7.14E-05	-+++-++-+++++++++++++++-	0	0.867
rs13181679	a	g	0.352	51258	3.971	7.15E-05	+++++++++-++++++-++++-	0	0.896
rs1443738	t	g	0.363	51258	-3.97	7.18E-05	+---+-+----------+	0	0.863
rs7467375	t	g	0.125	51258	3.97	7.20E-05	++++-+++--+++++++++++	39.7	0.027
rs7177989	t	c	0.726	51258	3.968	7.26E-05	+-+++++++++++++++++++-	36.2	0.044
rs973303	t	g	0.353	51257	-3.966	7.31E-05	----+---+-++++--+----	29.5	0.092
rs6881764	a	g	0.482	51258	3.961	7.48E-05	+++++++-+--+-++++++--	22.7	0.161
rs8008773	a	t	0.873	51258	-3.96	7.48E-05	---------+---+-++++++-	46.6	0.008
rs12541821	a	g	0.139	51258	3.96	7.51E-05	+-+++++++-++-++++-++++++	0	0.573
rs12134580	t	c	0.971	35673	-3.959	7.52E-05	----??+?-?????-???---+-	50.7	0.022
rs12368237	a	g	0.375	51258	-3.956	7.61E-05	+---+-+----------+	0	0.856
rs6421241	t	c	0.609	51258	3.955	7.65E-05	-++-++++++++++++++++-	0	0.740
rs1443742	t	c	0.362	51258	-3.955	7.66E-05	+---+-+----------+	0	0.863
rs4297682	a	g	0.904	51256	-3.955	7.67E-05	-------+------+++----	0	0.867
rs2049103	a	g	0.958	47020	3.954	7.68E-05	++++++-+-+?+-+++++?+-+-	0	0.687
rs11923274	a	g	0.958	47020	3.953	7.73E-05	++++++-+-+?+-+--+++?+-+-	0	0.656
rs10180695	t	c	0.424	51258	3.951	7.78E-05	++++++-+++++-+++-++++-++	0	0.940
rs4971723	t	c	0.435	51258	3.951	7.79E-05	++++++-+++++-+++-++++-++	0	0.904
rs11783005	t	c	0.284	48103	3.951	7.79E-05	+?++-+-++-++++++--++++	10.3	0.322
rs2414188	t	c	0.267	51258	-3.951	7.80E-05	-+---++----------++	33.7	0.059
rs11618590	t	c	0.837	51257	-3.95	7.81E-05	------+++-+-------+--	0	0.613
rs1545292	a	g	0.625	51258	3.95	7.83E-05	-++++++++++++++++++++-	0	0.855
rs4412846	a	t	0.353	51257	-3.949	7.84E-05	----+---+-++++--+----	30	0.088
rs7974278	c	g	0.624	51258	3.948	7.87E-05	-+++-++-++++++++++++++-	0	0.855
rs12521551	a	g	0.364	51258	-3.947	7.90E-05	----+-+----+--+-----	0	0.831
rs7651475	t	c	0.042	47020	-3.947	7.93E-05	----+-+-?-++++-?-+-+	0	0.664

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258). (continued)

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs2881577	a	g	0.958	47020	3.946	7.94E-05	+++++++-+?+-+---+++?+-+-	0	0.664
rs4738700	t	c	0.662	51258	-3.945	7.98E-05	----+-+------+-++	16.2	0.241
rs1373834	a	g	0.648	51257	3.944	8.01E-05	+++++++++-+----+++++++	33.1	0.064
rs13162928	t	c	0.648	51258	-3.943	8.04E-05	----+--++-----+---+-	0	0.894
rs1363179	t	g	0.583	51258	-3.943	8.05E-05	---+-+---+-+-+---++	30.4	0.085
rs6582152	c	g	0.368	51258	-3.941	8.11E-05	+--++-+-------+---+	0	0.667
rs8091788	a	g	0.829	51258	-3.94	8.14E-05	-+--+-+-----+-----	0	0.585
rs7628116	c	g	0.042	47020	-3.938	8.23E-05	----+-+-?-+-+---?-+-+	0	0.699
rs1421908	t	c	0.417	51258	3.937	8.25E-05	+++++-+-+++-+-+++++-+	30.8	0.081
rs7907283	a	g	0.087	51258	3.936	8.30E-05	+++++++----+-+++++++++	19.3	0.202
rs3787851	t	c	0.165	51257	-3.935	8.32E-05	-+-++------+--+---+-	0	0.791
rs12452350	a	g	0.833	51257	-3.932	8.41E-05	-------+-----+--+-	0	0.601
rs10769092	t	g	0.497	51258	-3.932	8.42E-05	-+--+---+---+-+---+-	0	0.555
rs6895949	a	g	0.279	51258	3.93	8.50E-05	+++-++++++++++++++-+-	27.2	0.114
rs10101647	a	g	0.160	51258	3.93	8.51E-05	+++++++-++++++-++++++	0	0.662
rs731428	t	c	0.936	49819	-3.929	8.53E-05	------+-+------?---+	0	0.560
rs4077278	c	g	0.475	51257	-3.928	8.56E-05	-------+---+---+-	0	0.939
rs7317531	t	g	0.457	51256	-3.927	8.59E-05	-+-------+-----+---	0	0.804
rs12519063	t	c	0.637	51258	3.926	8.64E-05	+++++-+-++++-+++-++++-	0	0.836
rs2111380	a	t	0.435	51258	3.926	8.65E-05	+++++++++++-+++-++++-	0	0.908
rs7822661	t	c	0.140	51258	3.925	8.66E-05	+-++++++-++-++++++++-	0	0.595
rs12209628	t	c	0.800	51258	-3.925	8.67E-05	---+--+++--++-+-----	0	0.970
rs1530303	t	c	0.350	51258	3.925	8.68E-05	+++++++-+-+-++++++-+++-	0	0.924
rs7620638	a	g	0.958	47020	3.924	8.70E-05	+++++++-+?+-+-+++?+-+-	0	0.661
rs12205387	t	c	0.200	51258	3.924	8.72E-05	++++-++---++--+-++++++	0	0.971
rs1501192	t	c	0.701	51258	-3.924	8.72E-05	-------+----+++---+	0	0.478
rs9898999	t	c	0.167	51257	3.922	8.80E-05	++++++++-+-++++++-++++	0	0.598
rs13248919	a	g	0.264	48103	3.918	8.92E-05	-?++--+-++-+++++--+++++	6.7	0.371
rs10742718	c	g	0.525	51258	-3.916	9.00E-05	-+--+--+------+-+---+-	0.3	0.456
rs10742725	t	c	0.475	50243	3.914	9.07E-05	+-+??++-++++++++-++++-	0	0.574
rs926300	a	t	0.828	51258	3.913	9.12E-05	++++++++++++++++++- +++-	0	0.703
rs6545190	t	g	0.564	51258	-3.91	9.25E-05	----+----+---+---+--	0	0.915
rs7631883	a	g	0.042	47020	-3.909	9.27E-05	----+-+-?-+-+----?-+-+	0	0.657
rs2352545	t	c	0.565	51258	-3.908	9.33E-05	-----+----+---+---+--	0	0.912
rs7930681	t	c	0.498	51258	3.904	9.45E-05	+-++-++-+++-+-+-++++-	9.6	0.330
rs2139686	a	c	0.957	47020	3.902	9.53E-05	+++++-+-+?+-+--+++?+-+-	0	0.673
rs2163946	a	g	0.529	48818	-3.901	9.57E-05	--+-??--+---+---+-+--?+-	0	0.464
rs1738819	t	c	0.133	51258	3.901	9.57E-05	-+++++-+-+++-++++-++++-	0	0.738
rs4901754	a	g	0.654	51257	-3.9	9.62E-05	++-----+---+-----++-	0	0.581
rs1395268	a	c	0.173	51258	3.899	9.66E-05	+-++-++-++++-++++-++++-	0	0.588
rs2836007	t	c	0.165	51257	-3.899	9.68E-05	-+-++-------+-----+--	0	0.834
rs7713437	a	g	0.720	51258	-3.898	9.68E-05	---++----+---+----+-+	25	0.136
rs2762089	a	t	0.355	51257	-3.898	9.69E-05	----+---+-++++-+----	31.2	0.078

TABLE S3: SNPs with a P-value <10-4 from the overall meta-analysis (discovery + replication. N=51 258). (continued)

SNP	Allele1	Allele2	Freq1	Weight	Zscore	P-value	Direction	HetISq	HetPVal
rs11179688	t	g	0.377	51258	-3.898	9.70E-05	?-?-?---------?	0	0.846
rs11746102	a	g	0.363	51258	-3.897	9.74E-05	---?-?---?-?---	0	0.837
rs11179690	a	g	0.623	51258	3.897	9.75E-05	?+?+?+?+?+?+?+?+?+?+?+?+?	0	0.843
rs7316126	a	t	0.624	51258	3.893	9.88E-05	?+?+?+?+?+?+?+?+?+?+?+?+?	0	0.834
rs4736893	a	g	0.231	51258	-3.893	9.91E-05	----?-----?-?----?	0	0.942
rs7728789	t	c	0.637	51258	3.891	9.97E-05	+++++?+?+?+?+?+?+?+?	0	0.835

TABLE S4. Replication of top SNPs (P-value <10-5) from previous GWASs.

				Original study		Current study - Discovery set (N = 34 549)		Current study – overall meta-analysis (N = 51 258)	
Study	SNP	Chr	Effective allele	Direction of effect	P-value	Direction of effect	P-value	Direction of effect	P-value
Sullivan 2009 (1738 ca, 1802 co)	rs2522833*	7	C	+	1.2e-06	+	0.26	+	0.83
Wray 2010 (5763 ca,6901co)**	rs11579964	1	T	-	4.4e-06	+	0.19	+	0.19
	rs7647854	3	G	+	4.6e-06	+	0.61	+	0.57
	rs12446956	16	C	+	1.1e-06	+	0.94	-	0.77
	rs12457996	18	C	-	5.7e-06	-	0.79	+	0.68
Terracciano 2010 (n=4811)***	rs12912233	15	T	+	6.3e-07	-	0.85	-	0.37
	rs8070473	17	T	-	1.5e-06	+	0.94	+	0.26
	rs349475	5	T	+	2.4e-06	+	0.72	+	0.28
	rs12420464	11	T	-	3.3e-06	+	0.17	+	0.51
	rs1927745	13	A	-	4.7e-06	-	0.69	+	0.78
	rs10514585	16	A	+	4.9e-06	-	0.053	-	0.011
	rs11009175	10	A	+	5.4e-06	-	0.067	-	0.17
	rs17864092	7	T	-	5.5e-06	-	0.60	-	0.33
	rs1449984	2	A	-	6.6e-06	+	0.90	-	0.76
	rs1924397	13	A	+	7.6e-06	+	0.66	+	0.21
	rs10744304	12	A	-	8.7e-06	+	0.31	-	0.80
	rs2017305	10	A	-	9.0e-06	-	0.99	-	0.63

*This SNP was tested for association in the current study as it was replicated previously

** Largest meta-analysis of MDD

*** Meta-analysis of trait depression, only independent top SNPs were tested here

CHAPTER 3.3

Text S1. Study descriptions

BLSA (Baltimore Longitudinal Study of Aging):The Baltimore Longitudinal Study of Aging is an ongoing multidisciplinary study of community-dwelling volunteers[1].

DHS (Dortmund Health Study): The DHS is a population based cohort, randomly recruited in 2003/4 in the city of Dortmund, Germany. 1312 participated in person and completed a physical examination, blood collection and medical interview, providing several phenotypes. Additionally 979 individuals participated through a mailed questionnaire. Genetic analyses are restricted to 1050 of those with a blood sample. The study was approved by the Ethic Board of the University of Muenster and all participants provided written informed consent[2, 3].
ERF (Erasmus Rucphen Family Study): The Erasmus Rucphen Family study is part of the Genetic Research in Isolated Population program. The study population essentially consists of one extended family of descendents from 20 related couples who lived in the isolate between 1850 and 1900 and had at least 6 children baptized in the community church. The detailed information about ERF isolate can be found elsewhere [4, 5]. The Medical Ethical Committee of the Erasmus Medical Center, Rotterdam approved the study and informed consent was obtained from all participants. The ERF study was approved by the Erasmus institutional medical-ethics committee in Rotterdam, The Netherlands.

FHS (NHLBI's Framingham Heart Study): For nearly 60 years, the Framingham Heart Study has examined the natural history, risk factors, and prognosis of cardiovascular, lung, and other diseases. Recruitment of the Original Cohort began in 1948. Twenty-three years later, 3548 children of the Original Cohort, along with 1576 of their spouses, enrolled in the Offspring Cohort. Beginning in 2002, 4095 adults having at least one parent in the Offspring Cohort enrolled in the Third Generation Cohort, along with 103 parents of Third Generation Cohort participants who were not previously enrolled in the Offspring Cohort. The objective of new recruitment was to complement phenotypic and genotypic information obtained from prior generations, with priority assigned to larger families. From a pool of 6553 eligible individuals, 1912 men and 2183 women consented and attended the first examination (mean age: 40 (standard deviation: 9) years; range: 19-72 years). The examination included clinical and laboratory assessments of vascular risk factors and imaging for subclinical atherosclerosis, as well as assessment of cardiac structure and function. The comparison of Third Generation Cohort data with measures previously collected from the first two generations will facilitate investigations of genetic and environmental risk factors for subclinical and overt diseases, with a focus on cardiovascular and lung disorders[6-8].

FINRISK:Finrisk07 belongs to the National FINRISK Study. The study has been conducted once in 5 years since 1972. Its focus is on studying chronic disease risk factors but includes also measures of mental health. The individuals were 25-74 years of age and came from the Eastern Finland, Turku, and Loimaa regions in Southwestern Finland, the metropolitan areas of the cities of Helsinki and Vantaa, and the province of Oulu in Northwestern Finland. The study was approved by the internal review board and received permission from the Uusimaa health region ethical committee. The participants provided informed written consent[9].

HEALTH ABC (Health, Aging and Body Composition study): The Health ABC study is a prospective cohort study investigating the associations between body composition, weight-related health conditions, and incident functional limitation in older adults. Health ABC enrolled well-functioning, community-dwelling African-American (n=1281) and white (n=1794) men and women aged 70-79 years between April 1997 and June 1998. Participants were recruited from a random sample of all Medicare eligible residents in the Pittsburgh, PA, and Memphis, TN, metropolitan areas. Eligibility requirements included no difficulty with activities of daily living, walking a quarter of a mile, or climbing 10 steps without resting. Participants have undergone annual exams and semi-annual phone interviews. CES-D questionairres and DNA extraction were carried out at baseline[10].

HBCS (The Helsinki Birth Cohort Study): HBCS is composed of 8760 individuals born between the years 1934-44 in one of the two main maternity hospitals in Helsinki, Finland. Between 2001 and 2003, a randomly selected sample of 928 males and 1075 females participated in a clinical follow-up study with a focus on cardiovascular, metabolic and reproductive health, cognitive function and depressive symptoms. In 2004, various psychological phenotypes were assessed, including the CESD. After exclusions, there were 1360 subjects with both valid phenotype and genotype data (59.1% women). The mean age of the subjects was 63.4 (SD=2.86). The mean age of the men was 63.3 (SD=2.68) and of the women was 63.6 (SD=3.00). Research plan of the HBCS was approved by the Institutional Review Board of the National Public Health Institute and all participants have signed an informed consent[11].

HRS (Health and Retirement Study): The Health and Retirement Study (HRS) is a longitudinal survey of a representative sample of Americans over the age of 50. The current sample is over 26 000 persons in 17 000 households. The study interviews respondents every two years about income and wealth, health and use of health services, work and retirement, and family connections. DNA was extracted from saliva collected during a face-to-face interview in the respondents' homes. These data represent respondents who provided DNA samples and signed consent forms in 2006 and 2008.

RUSH-ROS (The Religious Order Study of Rush Alzheimer's Disease Center): The Religious Orders Study (ROS), started in 1994, enrolled older Catholic priests, nuns and brothers from about 40

groups in 12 states of the United States. Since January 1994, more than 1,100 participants completed their baseline evaluation. The follow-up rate of survivors exceeds 90%. All participants were free of known dementia at enrollment and agreed to annual clinical evaluations including a medical history, cognitive function testing, neuropsychological examination, blood specimen collection, and brain donation at time-of-death. The study was approved by the Institutional Review Board of Rush University Medical Center. Informed consent and a signed anatomical gift act were obtained from each participant following a detailed presentation of the risks and benefits associated with participation[12].

RUSH-MAP (The Memory and Aging Project of Rush Alzheimer's Disease Center): The Rush Memory and Aging Project (MAP), started in 1997, enrolled older men and women from assisted living facilities in the Chicagoland area with no evidence on dementia at baseline. Since October 1997, more than 1,400 participants completed their baseline evaluation and the follow-up rate of survivors exceeds 90%. All participants agreed to annual clinical evaluations including a medical history, cognitive function testing, neuropsychological examination, blood specimen collection, and brain donation at time-of-death. The study was approved by the Institutional Review Board of Rush University Medical Center. Informed consent and a signed anatomical gift act were obtained from each participant following a detailed presentation of the risks and benefits associated with participation[13].

RS (Rotterdam Study I-II-III): The RS is an ongoing population-based cohort on risk factors for chronic diseases in the elderly which includes approximately 15,000 participants who live in Rotterdam. Detailed information on design, objectives and methods has been presented elsewhere. The Medical Ethics Committee of the Erasmus Medical Center approved the Rotterdam Study and written informed consent was obtained from all participants.[14]

SARDINIA: The SardiNIA is a family-based study includes over 6000 related individuals from four towns in the Ogliastra province of Sardinia, Italy[15].

STR (Sweedish Twin Registry): The TwinGene project of the STR, conducted between 2004 and 2008, is a population-based Swedish study of twins born between 1911 and 1958. The study participants have previously participated in a telephone interview called Screening Across the Lifespan of Twins cohort. To be included in TwinGene, both twins within a pair had to be alive. The zygosity of the twins was based on self-reported childhood resemblance, or by using DNA markers (for 18% of the total sample). A total of 9617 individuals who donated blood and answered questionnaires about life style and health, also passed the genotypic QC. The study was approved by the local ethics committee at Karolinska Institutet and all participants gave informed consent[16].

REFERENCES

1. Sutin AR, Zonderman AB. Depressive symptoms are associated with weight gain among women. *Psychol Med.* Apr 5:1-10.
2. Pfaffenrath V, Fendrich K, Vennemann M, et al. Regional variations in the prevalence of migraine and tension-type headache applying the new IHS criteria: the German DMKG Headache Study. *Cephalalgia.* Jan 2009;29(1):48-57.
3. Vennemann MM, Hummel T, Berger K. The association between smoking and smell and taste impairment in the general population. *J Neurol.* Aug 2008;255(8):1121-1126.
4. Lopez-Leon S, Chi Choy W, Aulchenko YS, et al. Genetic Factors Influence the Clustering of Depression among Individuals with Lower Socioeconomic Status. *PLoS ONE.* 2009;4(3):e5069.
5. Aulchenko YS, Heutink P, Mackay I, et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet.* 2004;12(7):527-534.
6. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* Dec 1975;4(4):518-525.
7. Splansky GL, Corey D, Yang Q, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.* Jun 1 2007;165(11):1328-1335.
8. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health.* Mar 1951;41(3):279-281.
9. Vartiainen E, Laatikainen T, Peltonen M, et al. Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol.* Apr;39(2):504-518.
10. Cesari M, Penninx BW, Newman AB, et al. Inflammatory markers and cardiovascular disease (The Health, Aging and Body Composition [Health ABC] Study). *Am J Cardiol.* Sep 1 2003;92(5):522-528.
11. Barker DJ, Osmond C, Forsen TJ, Kajantie E, Eriksson JG. Trajectories of growth among children who have coronary events as adults. *N Engl J Med.* Oct 27 2005;353(17):1802-1809.
12. Bennett DA. Postmortem indices linking risk factors to cognition: results from the Religious Order Study and the Memory and Aging Project. *Alzheimer Dis Assoc Disord.* Jul-Sep 2006;20(3 Suppl 2):S63-68.
13. Bennett DA, Schneider JA, Buchman AS, Mendes de Leon C, Bienias JL, Wilson RS. The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology.* 2005;25(4):163-175.
14. Hofman A, van Duijn CM, Franco OH, et al. The Rotterdam Study: 2012 objectives and design update. *Eur J Epidemiol.* Aug;26(8):657-686.
15. Pilia G, Chen WM, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* Aug 25 2006;2(8):e132.
16. Rahman I, Bennet AM, Pedersen NL, de Faire U, Svensson P, Magnusson PK. Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet.* Jun 2009;12(3):286-294.

Text S2. Items of CES-D subscales. * Indicates the item is also available in CES-D 10.

Positive items of the CES-D-20 questionnaire

4. I felt I was just as good as others
8. I felt hopeful about the future
12. I was happy*
16. I enjoyed life*

Negative items of the CES-D-20 questionnaire

3. I felt that I could not shake off the blues, even with help from my family and friends.
6. I felt depressed.*
9. I thought my life was a failure.
10. I felt fearful.
14. I felt lonely*
17. I had crying spells.
18. I felt sad. *

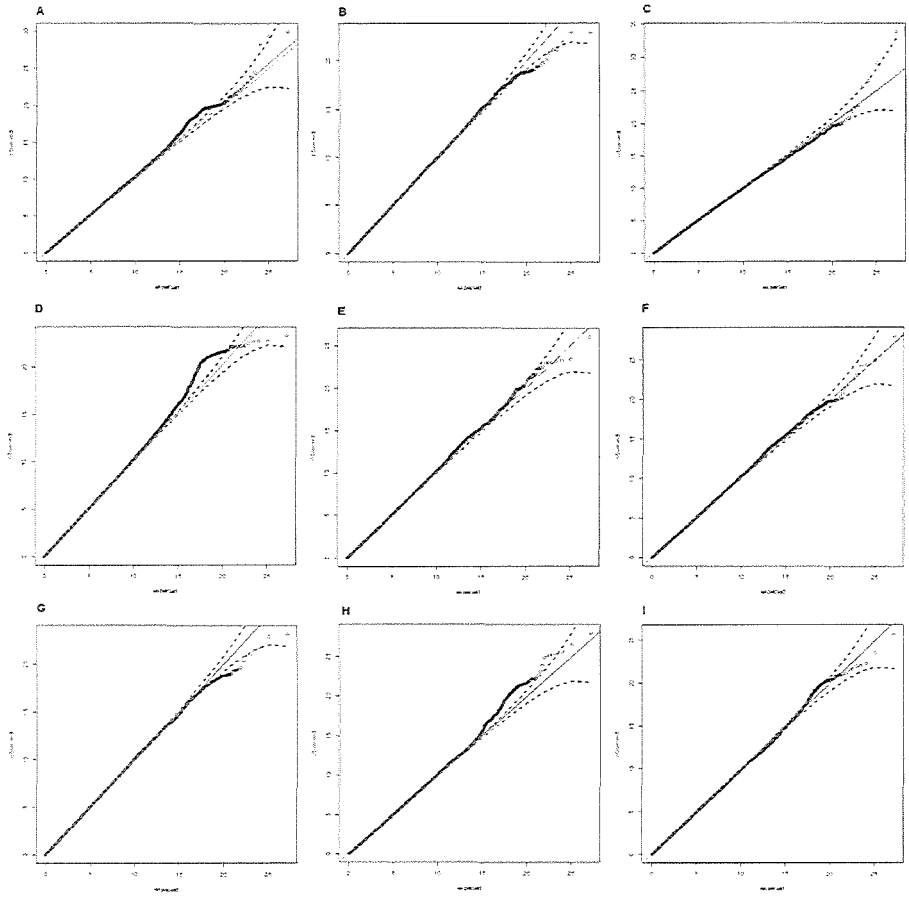
•*Somatic complaints in the CES-D-20 questionnaire*

1. I was bothered by things that usually do not bother me.
2. I did not feel like eating: my appetite was poor.
5. I had trouble keeping my mind on what I was doing.
7. I felt that everything I did was an effort.*
11. My sleep was restless.*
13. I talked less than usual.
20. I could not get going.*

- *Interpersonal problems*

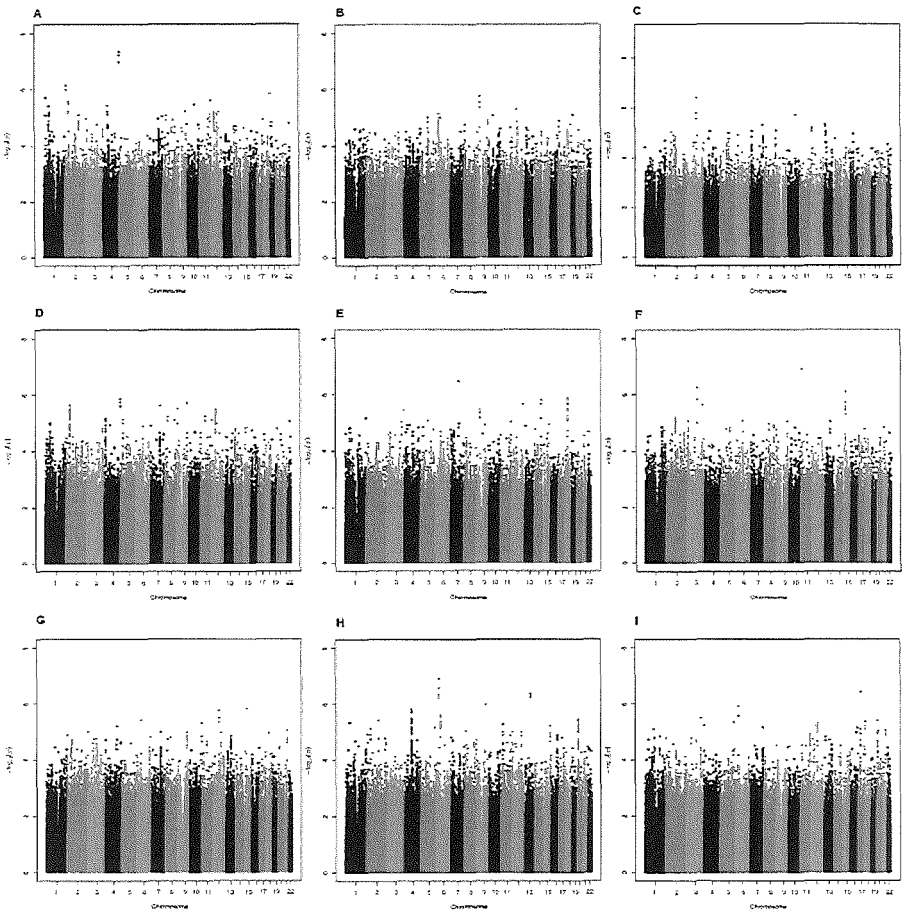
15. People were unfriendly.*
19. I felt that people disliked me.*

FIGURE S1. Q-Q plots of the over all meta-GWAS



A. Somatic items scale, combined analysis; B. Positive items scale, combined analysis; C. Negative items scale, D. Somatic items women-only analysis; E. Positive items women-only analysis; F. Negative items women-only analysis; G. Somatic items men-only analysis; H. Positive items men-only analysis; I. Negative items men-only analysis.

FIGURE S2. Manhattan plots of the over all meta-GWAS



A. Somatic items scale, combined analysis; B. Positive items scale, combined analysis; C. Negative items scale, D. Somatic items women-only analysis; E. Positive items women-only analysis; F. Negative items women-only analysis; G. Somatic items men-only analysis; H. Positive items men-only analysis; I. Negative items men-only analysis

TABLE S1. Genotyping and imputation information of the study samples

Study	Platform	Calling algorithm	QC filters for imputation				Data management and analysis			
			Call rate	MAF	HWE	Other	Software	Build	Software	Adjustment
ARIC	Affymetrix 6.0	Birdseed	<95%	>1%	>10-6	None	MACH	HapMap release 22 CEU (Build 36)	ProbABEL	Relatedness, study specific PCs, study site
BLSA	Illumina 550K	Beadstudio	>99%	>1%	<10E-4	None	MACH	HapMap release 22 CEU (Build 36)	merlin	none
DHS	Illumina HumanOmni2.5	TBFI	>97%	>1%	<10E-7	none	MACH	HapMap release 21 CEU II	SAS, PLINK with dosage module	none
ERF	Illumina 318K, Illumina 370K and Affymetrix 250K combined	Illumina Beadstudio, BRLMM	>98%	>1%	<10E-6	none	MACH	HapMap release 22 CEU (Build 36)	ProbABEL, R	Relatedness
FHS	Affymetrix 500K and MIPS 50K combined	BRLMM	>97%	>1%	<10E-6	Subject heterozygosity >5 SD away from the mean, large Mendelian error rate	MACH	HapMap release 22 CEU (Build 36)	R packages kinship, GEE	Relatedness
FINRISK	Illumina 6k	Illumina Beadstudio	>95%	>1%	tbfi	none	MACH	HapMap release 22 CEU (Build 36)	ProbABEL, R	none
HBCS	Illumina 610k, modified	Beadstudio	>95%,	>1%	>10E-5	none	MACH	HapMap release 22 CEU (Build 36)	Plink, ProbABEL	none
Health ABC-Eur	Illumina 1M and 1M duo	Illumina GenomeStudio	>98%	>1%	>10E-6		MACH	HapMap release 22 CEU (Build 36)	MACH2QTL, PLINK, MACH1	none
HRS	Illumina Omni2.5	Beadstudio	>98%	>1%	<10E-4		MACH	HapMap2 CEU (Build 36)	R, Plink	
InCHIANTI	Illumina HumanHap 550K	Beadstudio	-				MACH	HapMap release 22 CEU (build 36)	Mach2qtl & Mach2dat	none

TABLE S1. Genotyping and imputation information of the study samples (continued)

Study	Platform	Calling algorithm	QC filters for imputation		Data management and analysis			
NHSBrC	Illumina 550K	Beadstudio	<90%	SNP QC: MAF <0.01; Sample QC: call rate <90%, -duplicates and first/second degree relatives -ancestry outliers	MACH	HapMap release 22 CEU (build 36)	ProbABEL	disease status, top 4 EVs
NHSchd	Affymetrix 6.0	Birdseed	<98%	SNP QC: pHWE<10E-4, MAF <0.02 ; Sample QC: call rate <98%, --sex discrepancy with genetic data from X-linked markers -duplicates and first/ second degree relatives -ancestry outliers -heterozygosity -missing phenotype information	MACH	HapMap release 22 CEU (build 36)	ProbABEL	disease status, top 3 EVs

TABLE S1. Genotyping and imputation information of the study samples (continued)

Study	Platform	Calling algorithm	QC filters for imputation				Data management and analysis			
NHSt2d	Affymetrix 6.0	Birdseed	<98%			SNP QC, pHWE<10E-4, MAF <0.02, >1 discordance/12 replicates, significant plate associations; Sample QC: call rate <98%, -sex discrepancy with genetic data from X-linked markers -duplicates and first/ second degree relatives -ancestry outliers -heterozygosity -autosomal chromosome abberations	MACH	HapMap release 22 CEU (build 36)	ProbABEL	disease status, top 3 EVs
RS I	Illumina 550	Illumina Beadstudio, BRLMM	>98%	>1%	>10E-6	none	MACH	HapMap release 22 CEU (Build 36)	ProbABEL	none
RS II	Illumina 550 duo	Illumina Beadstudio, BRLMM	>98%	>1%	>10E-6	none	MACH	HapMap release 22 CEU (Build 36)	ProbABEL	none
RS III	Illumina 610 quad BeadChip version 3	Illumina Beadstudio, BRLMM	>98%	>1%	>10E-6	none	MACH	HapMap release 22 CEU (Build 36)	ProbABEL	none
RUSH-MAP	Affymetrix 6.0	Birdsuite	> 95%	>1%	>10E-6	misshap test < 10E-09	MACH	HapMap release 22 CEU (Build 36)	Plink, R	3 PCs
RUSH-ROS	Affymetrix 6.0	Birdsuite	> 95%	>1%	>10E-6	misshap test < 10E-09	MACH	HapMap release 22 CEU (Build 36)	Plink, R	3 PCs
SardiNIA	Affymetrix arrays	BRLMM	>90%	>1%	>10E-6	none	MACH	HapMap release 22 CEU (Build 36)	Merlin	none
STR TwinGene	Illumina OmniExpress	Illumina Beadstudio	>97%	>1%	>10E-7	none	IMPUTE2	HapMap release 22 CEU (Build 36)	SAS, PLINK	Relatedness, study specific PCs

Table S2. Top Meta-GWAS SNPs and their *P*-values in the PGC-MDD GWAS

A. SOMATIC ITEMS							
Discovery	**CES-D SUBSCALES GWAS**			**PGC MDD GWAS**			
SNP ID	Coded allele	Z-score	*P*-value	Coded allele	OR	SE	*P*-value
rs2375800	A	5.42	6.15E-08	A	0.96	0.03	9.87E-02
rs2375799	T	-5.31	1.09E-07	T	1.05	0.03	9.32E-02
rs453359	A	-4.95	7.44E-07	A	1.00	0.03	9.54E-01
rs1564630	T	4.88	1.05E-06	T	1.00	0.03	9.93E-01
rs12141569	T	4.76	1.97E-06	T	1.00	0.02	8.87E-01
rs12224700	T	4.71	2.48E-06	T	0.98	0.02	3.30E-01
rs10889125	A	4.62	3.88E-06	A	0.99	0.03	8.16E-01
rs2764928	T	-4.61	4.07E-06	T	1.01	0.03	7.48E-01
rs7563359	A	-4.54	5.72E-06	A	0.98	0.05	6.35E-01
rs1849785	T	4.52	6.15E-06	T	1.00	0.02	9.71E-01
rs2213180	T	4.52	6.20E-06	T	1.04	0.03	2.45E-01
rs4293597	T	4.50	6.90E-06	T	1.03	0.05	6.13E-01
rs1167078	T	-4.48	7.43E-06	T	1.00	0.02	8.84E-01
rs7127737	T	4.47	7.87E-06	T	1.00	0.03	8.76E-01
rs12022410	A	-4.47	7.94E-06	A	1.01	0.02	8.00E-01
rs7586242	T	-4.46	8.07E-06	T	1.08	0.05	1.47E-01
rs885813	T	-4.46	8.38E-06	T	1.00	0.02	8.36E-01
rs1656533	A	-4.45	8.45E-06	A	1.00	0.02	8.97E-01
rs7844381	A	4.45	8.53E-06	A	1.03	0.03	2.32E-01
rs1167096	T	4.44	8.88E-06	T	1.00	0.02	9.45E-01
rs1167080	A	4.44	8.88E-06	A	1.00	0.02	9.14E-01
rs1795939	T	4.44	8.92E-06	T	1.00	0.02	8.82E-01
rs2716122	A	4.44	9.02E-06	A	0.99	0.03	7.80E-01
rs1344828	A	-4.43	9.25E-06	A	1.00	0.02	8.97E-01
rs1167079	A	-4.43	9.52E-06	A	1.00	0.02	9.01E-01
rs1167103	A	-4.42	9.68E-06	A	1.00	0.02	9.29E-01
rs1167130	A	-4.42	9.79E-06	A	1.00	0.02	8.96E-01
rs2706326	A	-4.42	9.89E-06	A	1.00	0.02	9.01E-01
rs1167094	A	-4.42	9.92E-06	A	1.00	0.02	9.12E-01
rs10790965	A	-4.42	1.01E-05	A	1.01	0.03	8.66E-01
rs1167132	T	4.40	1.06E-05	T	1.00	0.02	8.99E-01
rs1167065	A	4.40	1.07E-05	A	1.00	0.02	9.97E-01
rs1185221	T	-4.40	1.10E-05	T	1.00	0.02	9.60E-01
rs10748349	A	4.39	1.14E-05	A	1.00	0.02	8.82E-01
rs877474	A	4.39	1.16E-05	A	0.96	0.09	6.37E-01
rs1204056	A	-4.38	1.17E-05	A	1.00	0.02	9.46E-01
rs4278783	A	-4.37	1.22E-05	A	1.10	0.06	1.02E-01
rs13278740	A	-4.37	1.26E-05	A	1.04	0.09	6.69E-01
rs1629029	T	-4.36	1.31E-05	T	1.00	0.02	9.06E-01
rs6699576	T	-4.35	1.37E-05	T	1.00	0.03	8.85E-01

rs1167152	T	4.35	1.40E-05	T	1.00	0.02	9.34E-01
rs3754971	A	-4.34	1.41E-05	A	1.07	0.06	2.50E-01
rs1184844	A	-4.33	1.50E-05	A	1.00	0.02	9.61E-01
rs10879636	A	-4.31	1.62E-05	A	0.99	0.02	6.32E-01
rs133592	T	4.31	1.65E-05	T	0.98	0.02	3.02E-01
rs1435707	A	4.30	1.70E-05	A	1.00	0.02	8.83E-01
rs3769975	T	4.30	1.74E-05	T	0.93	0.06	1.82E-01
rs7850959	A	4.29	1.77E-05	A	0.98	0.03	5.63E-01
rs12465423	T	-4.29	1.79E-05	T	1.07	0.06	2.13E-01
rs6439360	A	-4.29	1.81E-05	A	0.99	0.02	6.39E-01
rs3820998	A	-4.29	1.81E-05	A	1.08	0.06	1.71E-01
rs13116055	A	-4.28	1.84E-05	A	0.97	0.02	1.93E-01
rs889914	T	4.28	1.88E-05	T	0.93	0.06	1.94E-01
rs16845681	A	4.28	1.89E-05	A	0.94	0.06	2.49E-01
rs7543260	T	4.28	1.89E-05	T	0.99	0.03	8.08E-01
rs3769977	A	4.27	1.93E-05	A	0.93	0.06	1.88E-01
rs1953725	T	4.26	2.03E-05	T	0.97	0.02	2.10E-01
rs4370216	T	4.26	2.08E-05	T	1.03	0.02	2.43E-01
rs4510528	T	-4.25	2.10E-05	T	0.97	0.02	1.98E-01
rs3754974	A	4.25	2.12E-05	A	0.93	0.06	1.85E-01
rs11608269	A	-4.24	2.24E-05	A	0.96	0.03	1.58E-01
rs7600342	A	4.23	2.31E-05	A	0.92	0.06	1.59E-01
rs2358017	T	4.23	2.31E-05	T	0.93	0.06	1.95E-01
rs2631702	A	-4.22	2.43E-05	A	1.00	0.02	9.30E-01
rs13395640	T	4.22	2.44E-05	T	0.93	0.06	1.79E-01
rs2333727	A	4.22	2.44E-05	A	1.03	0.02	1.93E-01
rs13397244	A	4.22	2.46E-05	A	0.93	0.06	1.81E-01
rs12379445	T	-4.22	2.46E-05	T	0.99	0.02	5.35E-01
rs9299929	T	4.21	2.53E-05	T	0.97	0.02	2.62E-01
rs12819043	T	4.21	2.55E-05	T	0.99	0.04	7.45E-01
rs3769980	A	4.21	2.62E-05	A	0.94	0.06	2.39E-01
rs3769981	T	4.20	2.62E-05	T	0.93	0.06	1.93E-01
rs3769972	A	4.20	2.66E-05	A	0.93	0.06	2.29E-01
rs3769973	A	4.20	2.67E-05	A	0.93	0.06	2.22E-01
rs10241415	A	4.20	2.68E-05	A	1.00	0.02	8.87E-01
rs4082469	A	-4.19	2.74E-05	A	1.02	0.03	4.10E-01
rs1191543	A	4.19	2.77E-05	A	1.02	0.02	4.30E-01
rs8021587	T	4.18	2.93E-05	T	0.94	0.03	2.82E-02
rs3856792	A	-4.18	2.98E-05	A	1.02	0.03	4.17E-01
rs3773992	A	-4.17	3.07E-05	A	1.03	0.05	5.15E-01
rs2706328	T	-4.17	3.07E-05	T	1.00	0.02	9.47E-01
rs2363698	A	4.17	3.09E-05	A	1.04	0.03	1.09E-01
rs769211	A	4.16	3.17E-05	A	1.07	0.02	7.42E-03
rs987317	T	4.16	3.18E-05	T	1.00	0.02	9.68E-01

rs2146184	T	-4.16	3.25E-05	T	0.99	0.02	5.21E-01
rs12680146	T	4.15	3.28E-05	T	0.98	0.03	3.82E-01
rs6466057	T	4.15	3.29E-05	T	1.00	0.02	8.92E-01
rs1373221	A	4.15	3.37E-05	A	1.04	0.03	1.11E-01
rs11859036	A	-4.15	3.39E-05	A	1.01	0.02	7.82E-01
rs1953726	T	-4.14	3.49E-05	T	1.04	0.02	8.81E-02
rs13395957	T	4.12	3.87E-05	T	1.01	0.02	5.77E-01
rs9820308	T	-4.11	3.91E-05	T	0.96	0.03	2.60E-01
rs1204058	T	4.11	3.92E-05	T	1.00	0.02	9.43E-01
rs41562	T	-4.11	3.99E-05	T	1.00	0.02	9.28E-01
rs2040914	T	4.11	4.03E-05	T	1.00	0.02	9.19E-01
rs10277120	A	4.10	4.18E-05	A	1.00	0.02	9.59E-01
rs6740625	T	4.09	4.38E-05	T	1.00	0.02	9.94E-01
rs12378096	T	4.08	4.46E-05	T	0.99	0.03	6.26E-01
rs7038589	A	4.08	4.48E-05	A	0.99	0.03	6.26E-01
rs41563	A	4.08	4.51E-05	A	1.00	0.02	9.55E-01
rs3779210	T	4.08	4.57E-05	T	1.00	0.02	9.45E-01
rs1204077	T	4.07	4.67E-05	T	1.00	0.02	9.32E-01
rs3801999	A	-4.07	4.68E-05	A	1.00	0.02	8.40E-01
rs4687374	T	4.07	4.69E-05	T	1.00	0.02	9.85E-01
rs7517073	A	-4.07	4.75E-05	A	0.97	0.02	2.09E-01
rs4332348	A	4.07	4.77E-05	A	0.98	0.02	4.21E-01
rs2030776	C	-4.07	4.78E-05	C	1.00	0.02	8.83E-01
rs6504034	A	-4.06	4.92E-05	A	0.98	0.03	4.31E-01
rs12334245	T	-4.06	4.97E-05	T	1.00	0.02	9.47E-01
rs6466064	T	-4.06	4.97E-05	T	1.00	0.02	8.42E-01
rs9876410	T	-4.05	5.05E-05	T	0.96	0.03	2.08E-01
rs4804735	A	4.05	5.07E-05	A	1.06	0.07	3.78E-01
rs9566810	T	4.05	5.10E-05	T	0.97	0.02	1.80E-01
rs7305029	A	-4.05	5.16E-05	A	1.05	0.06	3.99E-01
rs7817195	A	-4.05	5.20E-05	A	1.06	0.05	2.27E-01
rs7732527	A	4.05	5.23E-05	A	1.01	0.02	6.01E-01
rs6752879	A	-4.04	5.35E-05	A	1.00	0.02	8.98E-01
rs2156697	T	-4.04	5.35E-05	T	1.00	0.02	9.23E-01
rs17187986	T	4.03	5.53E-05	T	1.01	0.05	9.10E-01
rs7582919	A	4.03	5.53E-05	A	1.00	0.02	9.51E-01
rs11662156	T	-4.03	5.57E-05	T	0.98	0.03	3.94E-01
rs1256341	T	4.03	5.59E-05	T	0.96	0.02	1.14E-01
rs17105259	T	4.03	5.61E-05	T	0.98	0.07	7.81E-01
rs7079860	C	4.01	6.05E-05	C	1.20	0.09	4.38E-02
rs2172180	A	4.01	6.10E-05	A	1.00	0.02	8.94E-01
rs3856791	A	4.01	6.17E-05	A	0.97	0.03	2.19E-01
rs9315851	T	4.01	6.17E-05	T	0.97	0.02	1.82E-01
rs10266871	A	4.01	6.18E-05	A	1.00	0.02	9.90E-01

rs1879040	A	-4.01	6.18E-05	A	1.00	0.03	8.79E-01
rs11686063	T	4.01	6.19E-05	T	1.00	0.02	9.84E-01
rs2078385	A	4.00	6.22E-05	A	1.03	0.05	5.60E-01
rs10451107	T	4.00	6.25E-05	T	0.95	0.05	3.07E-01
rs3827715	T	-4.00	6.26E-05	T	0.94	0.02	7.69E-03
rs3801282	A	4.00	6.28E-05	A	1.00	0.02	9.55E-01
rs11854143	A	-4.00	6.36E-05	A	1.10	0.08	2.53E-01
rs10817476	T	-4.00	6.36E-05	T	0.99	0.03	8.23E-01
rs717099	A	4.00	6.40E-05	A	1.03	0.02	2.50E-01
rs1898663	A	-3.99	6.51E-05	A	0.99	0.02	7.83E-01
rs2421599	A	3.99	6.58E-05	A	1.00	0.02	8.94E-01
rs1998056	C	3.99	6.58E-05	C	1.02	0.02	3.13E-01
rs17034004	T	3.99	6.59E-05	T	0.97	0.03	3.52E-01
rs2299319	T	3.99	6.65E-05	T	1.00	0.02	9.42E-01
rs7929966	A	3.99	6.65E-05	A	1.01	0.02	8.38E-01
rs13121332	A	-3.99	6.66E-05	A	0.97	0.02	1.11E-01
rs6466056	T	3.98	6.85E-05	T	1.00	0.02	9.65E-01
rs16835317	T	3.98	6.87E-05	T	0.83	0.16	2.55E-01
rs4461986	A	3.98	7.02E-05	A	1.00	0.02	8.55E-01
rs10783088	T	-3.97	7.06E-05	T	0.99	0.02	7.67E-01
rs904274	A	-3.97	7.19E-05	A	1.07	0.05	1.95E-01
rs17130163	T	3.97	7.19E-05	T	1.01	0.02	6.73E-01
rs1899684	A	3.97	7.22E-05	A	0.99	0.02	6.82E-01
rs7776707	A	3.97	7.29E-05	A	1.01	0.02	6.67E-01
rs7147892	T	-3.96	7.44E-05	T	0.98	0.02	3.88E-01
rs10238507	A	-3.96	7.46E-05	A	1.00	0.02	9.46E-01
rs16909586	A	-3.96	7.58E-05	A	0.90	0.10	3.22E-01
rs16946278	T	3.95	7.68E-05	T	1.00	0.04	9.52E-01
rs9325187	T	3.95	7.72E-05	T	1.02	0.06	7.88E-01
rs17108545	T	3.95	7.83E-05	T	0.86	0.08	8.16E-02
rs917114	T	-3.95	7.83E-05	T	1.00	0.02	9.79E-01
rs10487150	A	-3.94	8.02E-05	A	1.00	0.02	9.24E-01
rs2057884	T	3.94	8.09E-05	T	1.00	0.02	9.58E-01
rs2430514	A	-3.94	8.24E-05	A	0.98	0.02	4.10E-01
rs3801278	T	-3.94	8.29E-05	T	1.00	0.02	9.52E-01
rs10281856	A	-3.94	8.30E-05	A	0.98	0.02	3.46E-01
rs7143249	A	3.93	8.37E-05	A	1.02	0.02	4.38E-01
rs747145	T	-3.93	8.45E-05	T	1.04	0.06	5.16E-01
rs1974590	A	3.93	8.67E-05	A	1.02	0.03	4.74E-01
rs4738610	A	3.93	8.69E-05	A	1.03	0.02	1.37E-01
rs7196147	T	3.92	8.78E-05	T	1.00	0.02	8.53E-01
rs2536505	T	3.92	8.81E-05	T	1.03	0.03	3.57E-01
rs2601744	T	3.92	8.83E-05	T	1.03	0.02	1.45E-01
rs12080311	A	-3.92	9.01E-05	A	1.00	0.03	9.96E-01

rs4738613	A	-3.92	9.02E-05	A	0.97	0.02	1.18E-01
rs7508195	A	-3.92	9.05E-05	A	1.01	0.08	9.43E-01
rs13255034	T	3.91	9.26E-05	T	1.03	0.02	1.14E-01
rs11223955	T	-3.91	9.36E-05	T	1.02	0.03	3.97E-01
rs1034911	A	3.91	9.41E-05	A	1.02	0.03	4.48E-01
rs16878481	T	3.90	9.52E-05	T	1.04	0.04	3.61E-01
rs6777846	T	-3.90	9.81E-05	T	1.00	0.02	9.29E-01
rs887990	A	-3.90	9.82E-05	A	1.03	0.03	2.33E-01
rs7148151	C	-3.90	9.83E-05	C	0.98	0.02	4.14E-01
rs7547128	A	3.90	9.83E-05	A	0.99	0.02	6.46E-01
rs10281994	T	3.89	9.84E-05	T	1.00	0.03	8.98E-01
rs17506599	T	3.89	9.90E-05	T	1.02	0.03	4.45E-01

B. POSITIVE ITEMS

Discovery	**CES-D SUBSCALES GWAS**			**PGC MDD GWAS**			
SNP ID	Coded allele	Z-score	*P*-value	Coded allele	OR	SE	*P*-value
rs2810747	A	4.78	1.74E-06	A	0.98	0.02	4.29E-01
rs2585667	A	4.68	2.87E-06	A	0.98	0.02	4.46E-01
rs2810745	A	-4.60	4.27E-06	A	1.02	0.02	3.66E-01
rs2585668	T	-4.60	4.33E-06	T	1.02	0.02	3.82E-01
rs238516	T	-4.56	5.15E-06	T	0.98	0.02	2.60E-01
rs454214	T	4.42	9.86E-06	T	0.97	0.02	1.56E-01
rs10947531	T	-4.36	1.29E-05	T	0.97	0.02	2.14E-01
rs9296129	C	4.34	1.40E-05	C	1.03	0.02	2.07E-01
rs34316	A	-4.34	1.44E-05	A	1.03	0.02	1.70E-01
rs7739771	A	-4.34	1.45E-05	A	0.97	0.02	1.98E-01
rs7955793	A	-4.33	1.49E-05	A	1.00	0.03	9.12E-01
rs9469886	T	-4.31	1.67E-05	T	0.97	0.02	1.86E-01
rs9469914	A	4.30	1.69E-05	A	1.03	0.02	2.01E-01
rs7774697	A	4.30	1.73E-05	A	1.03	0.02	1.78E-01
rs1998702	A	4.30	1.73E-05	A	1.03	0.02	1.85E-01
rs3734264	A	-4.28	1.86E-05	A	0.97	0.02	2.59E-01
rs11875511	A	4.28	1.86E-05	A	0.98	0.02	2.61E-01
rs9462015	T	4.27	1.92E-05	T	1.03	0.02	1.82E-01
rs6457792	A	4.27	1.96E-05	A	1.02	0.02	3.57E-01
rs9462027	A	-4.26	2.09E-05	A	0.98	0.02	3.65E-01
rs11759151	T	-4.26	2.09E-05	T	0.97	0.02	2.01E-01
rs10118655	A	-4.23	2.37E-05	A	1.02	0.02	3.44E-01
rs11147763	T	4.22	2.41E-05	T	0.99	0.02	7.14E-01
rs9469917	A	-4.22	2.48E-05	A	0.97	0.02	2.24E-01
rs2269720	T	4.22	2.50E-05	T	1.03	0.02	2.36E-01
rs17163886	A	-4.21	2.52E-05	A	0.97	0.03	3.49E-01
rs921453	T	-4.21	2.58E-05	T	0.98	0.03	5.24E-01
rs10829613	T	-4.21	2.61E-05	T	0.99	0.02	7.14E-01

rs6457796	T	4.20	2.66E-05	T	1.02	0.02	3.94E-01
rs11661149	A	4.20	2.73E-05	A	0.94	0.08	4.49E-01
rs2814971	A	-4.19	2.80E-05	A	0.97	0.02	2.03E-01
rs905699	T	-4.18	2.93E-05	T	1.00	0.03	9.57E-01
rs6558668	A	4.17	3.08E-05	A	0.99	0.03	7.87E-01
rs9469857	T	-4.16	3.15E-05	T	0.97	0.02	2.08E-01
rs9469859	A	4.16	3.16E-05	A	1.03	0.02	2.00E-01
rs410671	A	-4.16	3.18E-05	A	1.03	0.02	1.98E-01
rs916725	T	-4.16	3.20E-05	T	1.02	0.02	4.36E-01
rs1390428	T	4.16	3.23E-05	T	1.01	0.03	6.79E-01
rs732292	A	4.16	3.24E-05	A	1.05	0.02	3.57E-02
rs6777846	T	-4.15	3.31E-05	T	1.00	0.02	9.29E-01
rs2764208	A	4.15	3.37E-05	A	1.03	0.02	2.08E-01
rs9906503	A	-4.14	3.44E-05	A	1.02	0.02	3.79E-01
rs4740783	A	4.14	3.48E-05	A	1.01	0.02	5.93E-01
rs4144603	T	-4.13	3.60E-05	T	1.05	0.02	4.86E-02
rs2764211	T	4.13	3.64E-05	T	1.03	0.02	1.93E-01
rs6831601	T	4.13	3.70E-05	T	1.02	0.03	5.14E-01
rs954510	A	4.12	3.74E-05	A	0.98	0.03	3.25E-01
rs2814959	T	-4.12	3.75E-05	T	0.97	0.02	1.94E-01
rs9383330	T	4.12	3.78E-05	T	0.98	0.02	3.80E-01
rs2477508	T	-4.12	3.79E-05	T	0.97	0.02	1.75E-01
rs7756405	A	-4.11	3.89E-05	A	0.97	0.02	1.92E-01
rs7438080	A	-4.11	3.97E-05	A	1.02	0.03	4.98E-01
rs6717562	A	4.11	4.00E-05	A	0.99	0.25	9.75E-01
rs1076993	T	4.11	4.04E-05	T	1.02	0.03	5.97E-01
rs7228898	T	-4.11	4.05E-05	T	0.98	0.03	6.10E-01
rs7076400	T	4.10	4.19E-05	T	1.03	0.02	1.21E-01
rs4578584	A	-4.09	4.37E-05	A	1.05	0.02	4.27E-02
rs9315851	T	4.08	4.55E-05	T	0.97	0.02	1.82E-01
rs9566810	T	4.08	4.57E-05	T	0.97	0.02	1.80E-01
rs11624336	A	4.07	4.70E-05	A	0.96	0.02	7.08E-02
rs11617401	T	-4.07	4.79E-05	T	1.06	0.03	1.14E-01
rs2744961	T	-4.06	4.81E-05	T	0.97	0.02	1.99E-01
rs1887136	A	4.06	4.90E-05	A	1.03	0.02	1.46E-01
rs11872359	C	4.06	4.96E-05	C	1.01	0.03	8.03E-01
rs2196985	T	4.05	5.04E-05	T	0.95	0.02	2.13E-02
rs9380455	A	4.05	5.05E-05	A	1.03	0.02	1.95E-01
rs1001122	T	4.05	5.07E-05	T	1.00	0.02	9.84E-01
rs4330979	T	4.05	5.22E-05	T	1.03	0.05	5.99E-01
rs13124137	A	4.04	5.42E-05	A	0.98	0.02	3.37E-01
rs1550721	T	-4.03	5.48E-05	T	0.95	0.06	4.58E-01
rs6507179	A	4.03	5.65E-05	A	1.01	0.03	6.81E-01
rs42850	A	4.03	5.66E-05	A	0.97	0.02	1.67E-01

rs6776500	T	-4.03	5.68E-05	T	1.00	0.02	9.84E-01
rs10804612	A	-4.02	5.76E-05	A	1.00	0.02	9.62E-01
rs1407947	C	4.02	5.88E-05	C	0.96	0.02	6.96E-02
rs34320	T	-4.02	5.91E-05	T	1.03	0.02	1.71E-01
rs6737675	A	-4.01	6.01E-05	A	0.95	0.02	1.01E-02
rs13101939	A	4.01	6.02E-05	A	1.03	0.02	2.23E-01
rs2814945	A	-4.01	6.08E-05	A	0.97	0.02	2.26E-01
rs9644396	A	-4.01	6.12E-05	A	0.97	0.05	5.99E-01
rs11205596	T	4.00	6.36E-05	T	1.02	0.08	8.04E-01
rs1378807	C	-4.00	6.44E-05	C	1.00	0.02	9.26E-01
rs8084815	C	3.99	6.53E-05	C	1.02	0.03	6.16E-01
rs579270	A	3.99	6.54E-05	A	0.99	0.02	5.65E-01
rs34318	A	3.99	6.58E-05	A	0.97	0.02	1.72E-01
rs809538	A	-3.99	6.60E-05	A	0.98	0.04	6.70E-01
rs1926262	T	-3.98	6.91E-05	T	0.99	0.05	8.90E-01
rs2825324	T	-3.98	6.94E-05	T	1.02	0.02	4.57E-01
rs503764	T	-3.98	6.95E-05	T	0.99	0.02	6.53E-01
rs7722746	T	-3.98	7.03E-05	T	1.00	0.02	8.55E-01
rs105154	A	-3.97	7.14E-05	A	0.84	0.09	5.69E-02
rs7708977	T	-3.97	7.21E-05	T	0.96	0.04	3.19E-01
rs2810729	T	-3.97	7.24E-05	T	1.00	0.02	9.94E-01
rs2814981	A	-3.96	7.43E-05	A	0.97	0.02	1.89E-01
rs2814992	A	3.96	7.53E-05	A	1.03	0.02	2.23E-01
rs216250	A	-3.96	7.62E-05	A	1.00	0.02	8.31E-01
rs17085784	A	3.95	7.68E-05	A	0.95	0.02	2.29E-02
rs16968087	A	-3.95	7.69E-05	A	0.97	0.03	3.94E-01
rs8086052	A	-3.95	7.77E-05	A	0.97	0.03	3.92E-01
rs941095	A	3.95	7.79E-05	A	1.01	0.03	8.10E-01
rs2512574	C	3.95	7.83E-05	C	0.99	0.02	6.10E-01
rs2644253	A	3.95	7.83E-05	A	1.01	0.03	6.41E-01
rs10885329	A	3.95	7.85E-05	A	1.03	0.02	1.24E-01
rs205484	T	-3.95	7.86E-05	T	1.00	0.05	9.99E-01
rs216244	T	3.95	7.88E-05	T	1.00	0.02	9.72E-01
rs11660710	A	-3.95	7.96E-05	A	0.98	0.03	5.94E-01
rs10872500	T	-3.94	8.16E-05	T	1.00	0.02	9.92E-01
rs1926562	A	3.94	8.17E-05	A	1.03	0.02	1.31E-01
rs2048472	T	-3.94	8.22E-05	T	1.03	0.03	2.98E-01
rs990462	A	3.93	8.35E-05	A	1.02	0.03	6.36E-01
rs7645254	A	3.93	8.37E-05	A	0.85	0.15	2.74E-01
rs755448	A	-3.93	8.47E-05	A	1.01	0.03	6.77E-01
rs11853808	A	-3.93	8.49E-05	A	0.97	0.07	6.07E-01
rs11164935	A	-3.93	8.50E-05	A	1.04	0.05	3.61E-01
rs2744949	T	-3.93	8.56E-05	T	0.97	0.02	2.08E-01
rs949681	A	-3.93	8.58E-05	A	1.04	0.09	6.35E-01

rs2235569	T	-3.93	8.58E-05	T	0.97	0.02	1.93E-01
rs7841009	T	3.93	8.59E-05	T	0.98	0.03	5.09E-01
rs2814953	A	3.93	8.60E-05	A	1.03	0.02	1.55E-01
rs1565976	A	-3.92	8.70E-05	A	0.99	0.03	6.38E-01
rs963911	T	3.92	8.70E-05	T	0.96	0.04	2.83E-01
rs1955513	T	-3.92	8.79E-05	T	1.06	0.04	1.98E-01
rs12636390	T	-3.92	8.85E-05	T	1.00	0.02	9.90E-01
rs761339	T	-3.92	8.92E-05	T	1.03	0.02	1.22E-01
rs325105	T	-3.92	8.98E-05	T	1.02	0.02	3.78E-01
rs4245331	A	3.92	8.99E-05	A	1.00	0.04	9.77E-01
rs12369132	A	3.92	9.00E-05	A	1.01	0.03	8.45E-01
rs12436991	T	3.92	9.00E-05	T	0.96	0.03	6.69E-02
rs1915886	A	3.92	9.00E-05	A	1.02	0.02	4.76E-01
rs616391	T	3.92	9.02E-05	T	0.97	0.02	1.63E-01
rs625970	T	-3.91	9.18E-05	T	1.03	0.02	1.40E-01
rs6502214	A	3.91	9.35E-05	A	1.00	0.02	9.02E-01
rs12049572	T	3.90	9.47E-05	T	1.02	0.03	4.20E-01
rs7221396	A	-3.90	9.48E-05	A	1.00	0.02	9.37E-01
rs2764928	T	-3.90	9.61E-05	T	1.01	0.03	7.48E-01
rs4791519	T	3.90	9.77E-05	T	1.00	0.02	9.64E-01
rs2659524	A	-3.90	9.80E-05	A	1.05	0.03	5.08E-02
rs6560804	A	3.90	9.80E-05	A	1.02	0.03	4.36E-01
rs11103693	T	-3.90	9.83E-05	T	1.01	0.03	8.40E-01
rs12236525	T	-3.89	9.92E-05	T	1.00	0.02	8.30E-01
rs2814984	T	3.89	9.95E-05	T	1.03	0.02	1.89E-01
rs11662615	T	3.89	9.99E-05	T	1.01	0.03	6.81E-01

C. NEGATIVE ITEMS

Discovery	**CES-D SUBSCALES GWAS**			**PGC MDD GWAS**			
SNP ID	Coded allele	Z-score	*P*-value	Coded allele	OR	SE	*P*-value
rs16859850	T	-4.79	1.71E-06	T	1.34	0.18	1.02E-01
rs9512880	A	4.57	4.79E-06	A	1.23	0.15	1.69E-01
rs2312971	C	-4.53	5.81E-06	C	1.00	0.02	8.36E-01
rs4937697	A	-4.46	8.11E-06	A	1.04	0.04	3.62E-01
rs2568654	T	-4.45	8.80E-06	T	0.99	0.02	7.33E-01
rs1320279	T	-4.44	9.02E-06	T	1.02	0.03	5.78E-01
rs6841066	A	-4.44	9.10E-06	A	1.01	0.02	6.94E-01
rs6545832	T	-4.43	9.27E-06	T	0.99	0.02	5.71E-01
rs9633969	A	4.43	9.54E-06	A	0.96	0.04	3.89E-01
rs2312972	T	-4.41	1.05E-05	T	1.00	0.02	8.19E-01
rs2328765	C	-4.37	1.25E-05	C	1.02	0.02	3.92E-01
rs12497924	T	-4.36	1.31E-05	T	0.95	0.03	9.89E-02
rs1404732	T	-4.31	1.64E-05	T	0.94	0.05	1.83E-01
rs11778919	T	4.30	1.69E-05	T	0.96	0.04	3.63E-01

rs15795	T	-4.30	1.73E-05	T	1.00	0.03	8.92E-01
rs17027255	T	-4.30	1.73E-05	T	0.98	0.02	4.94E-01
rs7785374	C	-4.30	1.75E-05	C	0.94	0.05	1.56E-01
rs261230	A	-4.29	1.75E-05	A	1.01	0.02	7.08E-01
rs261229	T	-4.29	1.81E-05	T	1.01	0.02	7.79E-01
rs7785839	T	4.27	1.94E-05	T	1.06	0.05	1.80E-01
rs10924341	T	4.26	2.00E-05	T	0.99	0.02	7.38E-01
rs13223313	T	-4.26	2.03E-05	T	0.94	0.05	1.80E-01
rs4851011	T	-4.26	2.03E-05	T	0.98	0.02	4.84E-01
rs9596054	A	4.23	2.36E-05	A	1.00	0.02	8.81E-01
rs10203477	A	-4.23	2.38E-05	A	1.02	0.02	2.72E-01
rs174415	A	-4.23	2.39E-05	A	1.00	0.02	9.05E-01
rs1557672	T	4.22	2.45E-05	T	0.99	0.03	7.62E-01
rs13225365	A	4.21	2.56E-05	A	1.07	0.05	1.71E-01
rs9832931	A	-4.21	2.61E-05	A	1.02	0.02	4.38E-01
rs10490204	A	4.16	3.12E-05	A	1.01	0.02	5.74E-01
rs2848852	A	4.16	3.21E-05	A	1.01	0.02	7.30E-01
rs17026974	A	-4.16	3.22E-05	A	0.99	0.02	7.30E-01
rs4677299	T	4.15	3.26E-05	T	1.01	0.02	6.08E-01
rs410739	T	4.15	3.28E-05	T	1.00	0.02	8.43E-01
rs41267	T	4.14	3.41E-05	T	1.00	0.03	9.90E-01
rs11781773	A	-4.13	3.58E-05	A	1.01	0.04	8.49E-01
rs261227	A	-4.13	3.71E-05	A	0.99	0.02	5.90E-01
rs3771150	A	-4.12	3.85E-05	A	0.99	0.02	5.61E-01
rs911271	A	-4.11	4.03E-05	A	1.01	0.02	7.51E-01
rs6788539	T	-4.10	4.15E-05	T	1.02	0.09	8.21E-01
rs2287034	A	-4.10	4.22E-05	A	0.99	0.02	7.73E-01
rs2377256	T	-4.09	4.24E-05	T	1.02	0.09	7.99E-01
rs2277669	A	-4.09	4.25E-05	A	1.09	0.02	1.47E-04
rs12720356	A	-4.09	4.31E-05	A	1.07	0.04	1.32E-01
rs17027166	A	-4.08	4.49E-05	A	0.99	0.02	5.64E-01
rs133592	T	4.08	4.60E-05	T	0.98	0.02	3.02E-01
rs6706689	A	-4.07	4.66E-05	A	1.02	0.02	2.80E-01
rs261226	T	-4.07	4.75E-05	T	1.00	0.02	8.71E-01
rs434342	A	-4.07	4.81E-05	A	1.05	0.03	1.26E-01
rs353951	A	4.05	5.06E-05	A	1.00	0.02	8.46E-01
rs759125	A	-4.05	5.14E-05	A	1.03	0.05	6.33E-01
rs11255224	T	-4.05	5.21E-05	T	1.09	0.06	1.76E-01
rs6135875	A	4.04	5.25E-05	A	1.00	0.03	9.18E-01
rs41493446	C	-4.04	5.35E-05	C	1.02	0.03	4.99E-01
rs2584980	T	4.03	5.65E-05	T	0.96	0.02	3.49E-02
rs3817092	A	4.02	5.95E-05	A	0.98	0.03	4.12E-01
rs16969899	T	4.01	5.98E-05	T	0.98	0.03	4.11E-01
rs405509	T	4.01	6.04E-05	T	1.00	0.02	9.34E-01

rs353952	C	4.01	6.15E-05	C	1.00	0.02	8.41E-01
rs6135876	T	-4.01	6.22E-05	T	1.00	0.03	9.77E-01
rs1053862	A	-4.00	6.24E-05	A	1.02	0.02	2.72E-01
rs2292116	T	4.00	6.36E-05	T	0.98	0.03	4.27E-01
rs6692618	T	4.00	6.39E-05	T	1.00	0.03	9.71E-01
rs13333300	A	4.00	6.42E-05	A	0.94	0.05	1.73E-01
rs185793	A	4.00	6.44E-05	A	1.00	0.02	9.69E-01
rs12905	A	-3.99	6.59E-05	A	0.99	0.02	7.32E-01
rs3743079	T	-3.99	6.62E-05	T	1.02	0.03	4.38E-01
rs2390879	T	3.99	6.68E-05	T	0.98	0.02	4.17E-01
rs4787645	T	3.98	6.80E-05	T	0.97	0.03	3.38E-01
rs6093035	A	-3.98	6.81E-05	A	0.99	0.02	7.17E-01
rs3761294	A	-3.98	6.87E-05	A	1.00	0.03	9.72E-01
rs1004592	A	-3.97	7.16E-05	A	1.00	0.02	9.60E-01
rs353940	A	-3.97	7.17E-05	A	0.99	0.02	8.05E-01
rs2269073	T	-3.97	7.19E-05	T	0.99	0.02	7.16E-01
rs12828607	A	-3.97	7.22E-05	A	1.10	0.04	1.53E-02
rs261218	T	-3.97	7.30E-05	T	1.01	0.02	6.69E-01
rs17773012	T	-3.96	7.58E-05	T	1.03	0.07	6.41E-01
rs1477223	T	3.96	7.65E-05	T	1.00	0.02	9.29E-01
rs6045664	T	-3.95	7.71E-05	T	0.95	0.05	3.12E-01
rs17027087	T	-3.95	7.72E-05	T	0.99	0.02	7.56E-01
rs281313	A	3.95	7.80E-05	A	1.00	0.02	8.58E-01
rs11975237	C	3.95	7.96E-05	C	1.00	0.02	9.58E-01
rs261999	T	3.94	8.12E-05	T	1.00	0.02	8.46E-01
rs17554795	A	3.94	8.14E-05	A	0.91	0.04	2.28E-02
rs7816485	T	-3.93	8.39E-05	T	1.06	0.03	3.35E-02
rs3732123	C	3.93	8.40E-05	C	1.01	0.02	6.96E-01
rs16981638	A	3.93	8.57E-05	A	1.06	0.06	2.85E-01
rs7414883	T	-3.92	8.73E-05	T	1.03	0.04	4.67E-01
rs6751658	A	-3.92	8.90E-05	A	1.00	0.02	9.82E-01
rs353958	A	-3.92	8.92E-05	A	0.98	0.02	2.66E-01
rs17027006	C	-3.92	8.99E-05	C	0.99	0.02	6.92E-01
rs403871	A	-3.92	9.06E-05	A	1.00	0.02	9.15E-01
rs1135354	T	3.91	9.07E-05	T	1.01	0.02	7.74E-01
rs6577902	A	-3.91	9.26E-05	A	1.06	0.03	3.43E-02
rs2015774	A	-3.91	9.37E-05	A	1.00	0.05	9.38E-01
rs12540014	A	3.91	9.42E-05	A	1.05	0.06	3.94E-01
rs1035130	T	-3.90	9.64E-05	T	1.00	0.02	8.49E-01
rs305578	A	-3.90	9.67E-05	A	1.02	0.04	6.52E-01
rs6944918	A	3.90	9.71E-05	A	1.00	0.02	9.47E-01
rs3091796	T	-3.89	9.95E-05	T	0.84	0.06	2.52E-03

CHAPTER 4.1

Table S1. Mean and standard deviations (SD) for the lipid species included in this study.

	Absolute values (μM)		Proportions	
	Mean	SD	Mean	SD
CER 16:0	0.978	0.252	0.118	0.025
CER 18:0	0.236	0.061	0.028	0.006
CER 20:0	0.207	0.049	0.025	0.005
CER 22:0	1.213	0.330	0.143	0.013
CER 23:0	0.959	0.258	0.114	0.014
CER 24:1	1.441	0.365	0.172	0.027
CER 24:0	3.403	0.931	0.401	0.039
Glu-CER 16:0	0.452	0.118	0.055	0.013
Glu-CER 24:1	0.575	0.168	0.070	0.022
LPC 15:0	1.165	0.352	0.005	0.001
LPC 16:0	133.052	34.628	0.523	0.038
LPC 16:1	2.476	0.866	0.010	0.002
LPC 18:0	46.685	14.720	0.182	0.025
LPC 18:1	21.200	6.969	0.083	0.012
LPC 18:2	32.088	11.630	0.125	0.028
LPC 18:3	0.637	0.263	0.002	0.001
LPC 20:0	0.375	0.111	0.002	0.000
LPC 20:3	2.545	0.990	0.010	0.003
LPC 20:4	9.564	4.825	0.038	0.015
LPC 20:5	0.890	0.499	0.003	0.002
LPC 22:4	0.593	0.304	0.002	0.001
LPC 22:5	0.852	0.481	0.003	0.001
LPC 22:6	2.018	0.977	0.008	0.003
PC 26:0	0.691	0.467	0.000	0.000
PC 30:0	4.297	2.020	0.002	0.001
PC 30:1	0.746	1.193	0.000	0.001
PC 32:0	16.295	3.777	0.007	0.001
PC 32:1	18.658	9.447	0.008	0.003
PC 32:2	4.935	2.110	0.002	0.001
PC 34:0	3.595	1.839	0.002	0.001
PC 34:1	246.240	68.305	0.112	0.018
PC 34:2	562.224	131.337	0.256	0.032
PC 34:3	18.313	6.432	0.008	0.002
PC 34:4	1.906	0.848	0.001	0.000
PC 36:0	2.964	1.372	0.001	0.001
PC 36:1	42.455	13.723	0.019	0.004
PC 36:2	292.093	72.008	0.133	0.017

TABLE S1. Mean and standard deviations (SD) for the lipid species included in this study. (continued)

	Absolute values (μM)		Proportions	
	Mean	**SD**	**Mean**	**SD**
PC 36:3	154.868	39.970	0.070	0.009
PC 36:4	217.312	57.647	0.099	0.016
PC 36:5	24.007	10.912	0.011	0.004
PC 38:0	6.960	2.709	0.003	0.001
PC 38:1	9.890	5.235	0.004	0.002
PC 38:2	5.815	2.872	0.003	0.001
PC 38:3	46.673	15.146	0.021	0.005
PC 38:4	132.768	38.692	0.060	0.013
PC 38:5	64.080	17.734	0.029	0.005
PC 38:6	90.872	28.265	0.041	0.011
PC 38:7	2.484	1.010	0.001	0.000
PC 40:0	1.400	0.704	0.001	0.000
PC 40:1	5.199	2.879	0.002	0.001
PC 40:2	2.838	1.545	0.001	0.001
PC 40:3	3.814	2.153	0.002	0.001
PC 40:4	4.446	1.555	0.002	0.001
PC 40:5	11.864	4.014	0.005	0.001
PC 40:6	31.109	11.038	0.014	0.004
PC 40:7	7.233	2.253	0.003	0.001
PC 42:4	0.981	0.648	0.000	0.000
PC 42:5	1.493	0.851	0.001	0.000
PC O 32:0	3.675	1.038	0.002	0.000
PC O 32:1	2.721	0.830	0.001	0.000
PC O 34:0	1.466	0.576	0.001	0.000
PC O 34:1	9.756	2.373	0.004	0.001
PC O 34:2	12.199	3.377	0.006	0.001
PC O 34:3	7.585	2.512	0.003	0.001
PC O 36:0	0.986	0.511	0.000	0.000
PC O 36:1	8.465	2.096	0.004	0.001
PC O 36:2	13.766	3.482	0.006	0.001
PC O 36:3	8.104	2.192	0.004	0.001
PC O 36:4	17.940	4.597	0.008	0.002
PC O 36:5	11.076	3.169	0.005	0.001
PC O 38:1	1.041	0.854	0.000	0.000
PC O 38:2	2.024	0.971	0.001	0.000
PC O 38:3	4.265	1.223	0.002	0.000
PC O 38:4	12.489	2.913	0.006	0.001
PC O 38:5	17.637	4.173	0.008	0.002
PC O 40:4	2.381	0.727	0.001	0.000
PC O 40:5	3.477	0.972	0.002	0.000
PC O 40:6	4.240	1.265	0.002	0.000

TABLE S1. Mean and standard deviations (SD) for the lipid species included in this study. (continued)

	Absolute values (µM)		Proportions	
	Mean	**SD**	**Mean**	**SD**
PC O 42:5	1.548	0.529	0.001	0.000
PC O 42:6	1.190	0.462	0.001	0.000
PE 32:0	0.130	0.056	0.004	0.002
PE 32:1	0.328	0.194	0.009	0.003
PE 32:2	0.130	0.052	0.004	0.002
PE 34:0	0.125	0.079	0.004	0.002
PE 34:1	2.835	1.591	0.074	0.015
PE 34:2	5.634	2.977	0.149	0.025
PE 34:3	0.411	0.206	0.011	0.003
PE 36:1	1.233	0.609	0.033	0.008
PE 36:2	6.715	3.197	0.181	0.028
PE 36:3	2.002	1.042	0.053	0.013
PE 36:4	2.987	1.430	0.080	0.011
PE 36:5	0.299	0.129	0.008	0.003
PE 38:1	0.193	0.060	0.006	0.002
PE 38:2	0.154	0.049	0.005	0.002
PE 38:3	0.561	0.289	0.015	0.003
PE 38:4	4.926	2.222	0.134	0.023
PE 38:5	1.577	0.729	0.043	0.008
PE 38:6	2.967	1.615	0.079	0.022
PE 40:3	0.047	0.025	0.001	0.001
PE 40:4	0.175	0.093	0.005	0.001
PE 40:5	0.511	0.276	0.014	0.003
PE 40:6	2.047	1.112	0.055	0.017
PE 42:5	0.034	0.019	0.001	0.001
PE 42:6	0.038	0.019	0.001	0.001
PE 42:7	0.057	0.026	0.002	0.001
PE O 38:7	0.181	0.067	0.006	0.003
PE O 40:3	0.739	0.267	0.021	0.005
PLPE 16:0/18:1	0.846	0.251	0.016	0.004
PLPE 16:0/18:2	2.381	0.835	0.045	0.010
PLPE 16:0/20:4	5.847	2.071	0.111	0.019
PLPE 16:0/20:5	0.692	0.365	0.013	0.005
PLPE 16:0/22:5	1.650	0.517	0.032	0.007
PLPE 16:0/22:6	3.637	1.120	0.070	0.016
PLPE 18:0/18:1	1.053	0.353	0.020	0.004
PLPE 18:0/18:2	3.945	1.394	0.075	0.017
PLPE 18:0/20:4	9.597	3.444	0.180	0.027
PLPE 18:0/20:5	1.265	0.856	0.023	0.011
PLPE 18:0/22:6	3.386	1.198	0.064	0.013
PLPE 18:1/18:1	0.790	0.238	0.015	0.003

TABLE S1. Mean and standard deviations (SD) for the lipid species included in this study. (continued)

	Absolute values (μM)		Proportions	
	Mean	**SD**	**Mean**	**SD**
PLPE 18:1/18:2	1.939	0.678	0.037	0.009
PLPE 18:1/20:4	5.256	1.913	0.099	0.017
PLPE 18:1/20:5	0.662	0.358	0.012	0.005
PLPE 18:1/22:6	2.048	0.683	0.039	0.009
SPM 14:0	12.878	3.595	0.024	0.005
SPM 15:0	6.728	1.933	0.013	0.003
SPM 16:0	146.601	30.203	0.044	0.005
SPM 16:1	23.545	5.355	0.002	0.002
SPM 17:0	4.186	1.278	0.011	0.003
SPM 18:0	28.980	7.351	0.030	0.005
SPM 18:1	16.037	4.546	0.003	0.001
SPM 18:2	1.384	0.601	0.008	0.002
SPM 20:0	18.428	7.019	0.034	0.011
SPM 20:1	8.583	4.413	0.016	0.007
SPM 22:0	39.580	10.853	0.064	0.008
SPM 22:1	34.008	8.766	0.005	0.005
SPM 22:2	2.790	2.654	0.003	0.004
SPM 23:0	17.029	4.615	0.028	0.004
SPM 23:1	14.771	4.000	0.003	0.003
SPM 24:0	27.731	7.733	0.152	0.015
SPM 24:1	80.996	18.934	0.069	0.010
SPM 24:2	36.661	9.486	0.008	0.004
SPM 24:3	4.029	2.177	0.002	0.001
SPM dih 16:0	5.762	1.823	0.276	0.021
SPM dih 18:0	1.466	1.089	0.003	0.002
SPM dih 22:0	1.598	1.488	0.074	0.012
SPM dih 23:0	0.878	0.670	0.032	0.005
SPM dih 24:0	0.853	0.802	0.052	0.009

TABLE S2. Significant Linkage findings (LOD> 3.3)

Lipid	Marker	Locus	Position	LOD	Nearest gene	Distance (kb)	Genes within the LOD-1 SI
PLPE 18:1/22:6	rs785605	1p31.1	82435415	4.26	LPHN2	263	CLCA1 COX6A1P CTBS GBP1 GBP2 GBP3 GNG5 GTF2B CYR61 PRKACB PKN2 PTGFR LMO4 BCL10 SEP15 CLCA3 CLCA2 HS2ST1 IFI44 IFI44L CLCA4 LPHN2 LPAR3 DDAH1 SH3GLB1 ZNHIT6 MCOLN3 CCBL2 ODF2L DNASE2B ELTD1 SPATA1 TTLL7 BXDC5 SYDE2 GBP4 GBP5 SSX2IP WDR63 SAMD13 C1orf52 GBP6 MCOLN2 COL24A1 ADHSP2 GBP7 UOX C1orf180
%PLPE 18:1/22:6	rs954145	1p22.2	90210561	4.22	ZNF326	3	GBP1 GBP2 GBP3 GTF2B PKN2 CDC7 LRRC8B LRRC8D CCBL2 ZNF644 LRRC8C GBP4 GBP5 GBP6 HFM1 ZNF326 BARHL2 HSP90B3P GBP7 FLJ27354 WDR82P2 PHKA1P1
PC O 36:1	rs906283	18p11.21	10918707	3.77	PIEZO2	0 (intronic)	C18orf1 CIDEA GATA6 GNAL IMPA2 LAMA3 MC2R MC5R NDUFV2 NPC1 PTPN2 RBBP8 ROCK1 SNRPD1 RNMT NAPG RIOK3 VAPA PPP4R1 SLMO1 RALBP1 AFG3L2 RAB31 ANKRD12 KIAA0802 CABYR C18orf8 CEP192 IMPACT SPIRE1 PSMG2 TWSG1 CHMP1B MIB1 HRH4 FAM38B CTAGE1 MPPE1 OR4K7P CEP76 KIAA1772 FLJ11996 SEH1L TXNDC2 TUBB6 C18orf45 CABLES1 ESCO1 OSBPL1A C18orf19 KRT18P8 TTC39C C18orf15 ANKRD29 APCDD1 ZNF519 ABHD3 RAB12 C18orf58 ANKRD30B FGF7P1 A26B2 OR4K8P CXADRP3 MGC26718 AMAC1L1
PE 38:5*	rs2606418	1p36,32	2894007	3.7	ACTRT2	66	PEX10 TNFRSF14 PLCH2 RER1 ARHGEF16 PANK4 PRDM16 MMEL1 MORN1 C1orf93 ACTRT2 HES5 TTC34

TABLE S2. Significant Linkage findings (LOD> 3.3) (continued)

Lipid	Marker	Locus	Position	LOD	Nearest gene	Distance (kb)	Genes within the LOD-1 SI
PC O 34:3	rs13851	17q31	38727370	3.67	BRCA1	50	ACLY AOC2 ARL4D ATP5G1 ATP6V0A1 BRCA1 FMNL1 CACNB1 CDC6 CDC27 CHAD CCR7 CNP COL1A1 COX11 CRHR1 CSF3 DHX8 DLX3 DLX4 DUSP3 ERBB2 ETV4 EZH1 GAST FZD2 G6PC KAT2A GFAP GIP GNGT2 CCR10 GRB7 GRN HCRT HLF HOXB1 HOXB2 HOXB3 HOXB4 HOXB5 HOXB6 HOXB7 HOXB8 HOXB9 HSD17B1 IFI35 IGFBP4 ITGA2B ITGA3 ITGB3 JUP KPNB1 KRT9 KRT10 KRT12 KRT13 KRT14 KRT15 KRT16 KRT17 KRT19 KRT31 KRT32 KRT33A KRT33B KRT34 KRT35 LASP1 NBR1 MAPT ADAM11 MEOX1 MLLT6 MPP2 MPP3 MYL4 NAGLU NEUROD2 NFE2L1 NGFR NME1 NME2 NMT1 NSF PDX2 PHB PNMT MED1 PPY PSMB3 PYY PSMD3 PTP4AP1 RAB5C RARA RNY4P2 RPL19 RPL27 SGCA SHCL1 SLC4A1 SMARCE1 SP2 STAT3 STAT5A STAT5B HNF1B MLX THRA TOP2A DNAJC7 TUBG1 UBTF WNT3 WNT9B PCGF2 RND2 PIP4K2B SPOP CNTNAP1 TCAP SKAP1 AOC3 BECN1 KRT41P KRT38 KRT37 KRT36 ABCC3 CACNA1G HAP1 MAP3K14 SPAG9 EFTUD2 RPL23 NPEPPS GOSR2 NR1D1 TBKBP1 PLEKHM1 MED24 LRRC37A HDAC5 TOM1L1 GJC1 TOB1 PSME3 EIF1 NBR2 SLC35B1 CALCOCO2 RAMP2 HOXB13 VAT1 SC6S HEXIM1 IGF2BP1 C1QL1 RUNDC3A STARD3 CBX1 MYST2 NXPH3 SNF8 CASC3 IKZF3 ZNF652 GPATCH8 KCNH4 MMD C17orf88 KRT23 RNU2P2 TUBG2 NKIRAS2 CCDC56 PSMC3IP SNX11 TBX21 SOCS7 SOST UTP18 RAPGEFL1 ABI3 COPZ2 MRPL27 ARL17P1 SLC25A39 CROP HIGD1B CRKRS KRT20 MBTD1 CCDC49 C17orf73 EPN3 LRRC37A4 PNPO KLHL11 TMEM100 RSAD1 LRRC59 GSDMB CA10 ATXN7L3 PLXDC1 ARHGAP23 FKBP10 XYLT2 SPATA20 UBE2Z WNK4 C17orf53 TMUB2 DHX58 ATAD4 ACBD4 DCAKD PLEKHH3 DBF4B ACSF2 CDK5RAP3 SP6 COASY SNIP AARSD1 FAM117A KRTAP1-3 KRTAP1-1 KRTAP9-9 KRTAP4-6 KRTAP2-1 TTC25 KRTAP4-12 KRTAP1-5 KRTAP3-1 KRTAP3-2 KRTAP9-2 KRTAP9-3 KRTAP9-8 KRTAP17-1 MYCBPAP PPP1R1B TBC1D3 C17orf37 MRPL45 VPS25 TMEM101 C17orf92 GHDC KRTAP4-4 KIF2B PPP1R9B TNS4 FBXL20 KRTAP9-4 KRTAP4-1 KRTAP4-5 KRTAP4-3 KRTAP4-2 KRTAP3-3 KRTAP2-4 KRTAP3P1 KRTAP9P1 KRTAP4P1 LRRC46 SCRN2 LOC90586 ANKRD40 G6PC3 ASB16 PERLD1 HSPB9 ORMDL3 PLCD3 TMEM106A OSBPL7 NT5C3L ZPBP2 C17orf46 HEXIM2 LSM12 CCDC43 CNTD1 WFIKKN2 FLJ40194 B4GALNT2 C17orf57 MRPL10 GJD3 KRT222P KRT40 RPL21P4 CD300LG KIF18B RUNDC1 EME1 C17orf69 WIPF2 KRT25 TMEM99 NAGS FAM134C TMEM92 PHOSPHO1 IMP5 KRT28 KRT24 ARHGAP27 ZNF385C SAMD14 STH STXBP4 TAC4 KIAA1267 FAM171A2 TTLL6 CISD3 GSDMA KRT42P PTRF KLHL10 C17orf65 MSL-1 KRT27 FLJ43826 STAC2 KRT26 MRPS21P9 HILS1 C17orf98 CCDC103 RPRML ARL5C KRT39 BRCA1P1 MGC57346 GPR179 LRRC37A2 FBXO47 SNORA21 ARL17 KRT223P HSD17BP1 NME1-NME2 TBC1D3H KRTAP1-4 TBC1D3F TBC1D3E SNORD124

TABLE S2. Significant Linkage findings (LOD> 3.3) (continued)

Lipid	Marker	Locus	Position	LOD	Nearest gene	Distance (kb)	Genes within the LOD-1 SI
%CER 18:0	rs295357	19p12	23229879	3.42	ZNF724P	8	CCNE1 COMP NCAN GDF1 MEF2B UPF1 UQCRFS1 ZNF14 ZNF708 ZNF43 ZNF56 ZNF66 ZNF85 ZNF90 ZNF91 ZNF99 ZNF208 RFXANK LPAR2 HOMER3 ZNF254 SFRS14 LASS1 POP4 COPE CRTC1 KIAA0892 NDUFA13 GMIP TM6SF2 DDX49 GATAD2A TMEM161A KLHL26 ZNF253 ATP13A1 ZNF492 SF4 PLEKHF1 ZNF430 PBX4 ZNF93 C19orf12 TSSK6 ZNF486 ZNF682 ARMC6 ZNF101 ZNF257 NR2C2AP CILP2 ZNF98 ZNF738 ZNF714 ZNF681 ZNF676 ZNF100 ZNF431 ZNF675 ZNF626 PCGF7P ZNF725 SLC25A42 MGC39821 ZNF493 VSTM2B ZNF429 YJEFN3 LASS1-GDF1 ZNF728 KRT18P40 HAPLN4 ZNF506ZNF724PPHF5CP ZNF826 ZNF726 ZNF730 ZNF737 ZNF849P
Poly-unsaturated SPM	rs201204	6q16.3	104842863	3.35	HACE1	440	AIM1 PRDM1 CCNC GRIK2 PREP SIM1 ATG5 ASCC3 BVES QRSL1 HACE1 PRDM13 POPDC3 MCHR2 RTN4IP1 USP45 LIN28B
% PC 36:4	rs1525064	11q12.3	62723953	3.35	SLC22A25	0 (intronic)	Explained by FADS1-2-3
% Saturated SPM	rs483884	11q21	95551040	3.35	MAML2	0 (intronic)	BIRC2 BIRC3 CASP1 CASP4 CASP5 FUT4 GRIA4 MMP1 MMP3 MMP7 MMP8 MMP10 MMP12 MMP13 MRE11A MTNR1B PGR TRPC6 JRKL MTMR2 MMP20 MED17 CEP57 YAP1 GPR83 SFRS2B ENDOD1 PANX1 C11orf54 CWC15 CNTN5 ANKRD49 JMJD2D C11orf75 KIAA1377 CARD18 MMP27 TAF1D OR2AL1P DYNC2H1 CCDC82 PDGFD TMEM133 DCUN1D5 MAML2 C11orf70 CARD16 TMEM123 SLC36A4 FAT3 CASP12 FAM76B SESN3 PIWIL4 FLJ32810 AMOTL1 CCDC67 ANGPTL5 HEPHL1 C11orf90 NDUFB11P LOC401705 DDI1 LOC440063 INCA SCARNA9 SNORA8 SNORA1 SNORA18 SNORA40 SNORA25 SNORA32 SNORD5 SNORD6 HPRTP4 UNQ6228

Positions are based on H. sapiens variation (dbSNP 134, ENSEMBL). Gene refers to the nearest gene to the SNP, even though it may not be candidate gene in the region. Distance refers to the distance in kb between the end or start of the gene and the marker.

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3)

Lipid	Marker	Chromosome	Position	LOD	
%PE 38:2	rs2477703	1	2458154	2.02	
PE 38:4	rs2606418	1	2894007	2.21	
PE 38:3	rs2606418	1	2894007	2.02	
PE 34:2	rs2606418	1	2894007	2.7	
Total PE	rs2606418	1	2894007	3.21	
PE 40:5	rs2027262	1	2949536	2.51	
PE 36:4	rs2027262	1	2949536	2.67	
%PE O 38:7	rs2027262	1	2949536	2.86	
%PLPE 16:0/22:5	rs2821039	1	3430164	1.92	
Monounsaturated PE	rs878063	1	3496085	2.24	
PE 36:2	rs4654522	1	3910010	2.42	
%PE 38:1	rs4654522	1	3910010	3.06	
PE 36:3	rs2035453	1	4615328	2.24	
PE 36:1	rs2035453	1	4615328	2.87	
%PE 42:6	rs2035453	1	4615328	1.99	
PLPE 18:1/18:1	rs1555024	1	23531509	2.04	
PC O 32:0	rs6619	1	37699789	2.49	
%PC O 38:4	rs1934405	1	48962059	2.26	
%PLPE 16:0/22:6	rs2075630	1	86664622	2.52	
PC O 36:1	rs698274	1	90956198	2.01	
%PLPE 18:1/20:4	rs834984	1	99835359	1.93	
PC 40:7	rs2065188	1	104310593	2.09	
PC O 36:4	rs2031504	1	106770393	1.99	
%SPM 14:0	rs1020812	1	108196674	1.99	
%SPM dih 24:0	rs968853	1	161878230	2.116	*
PC O 36:4	rs761076	1	164351640	2.21	
CER 24:0	rs2040427	1	172052589	1.89	
SPM dih 18:0	rs716760	1	176981817	2.285	*
%PC O 36:4	rs1020782	1	178010159	2.32	
SPM dih 23:0	rs1020782	1	178010159	2.353	*
%PC 38:0	rs2039759	1	190335126	2.02	
SPM dih 23:0	rs417774	1	211221820	1.918	*
%SPM 18:2	rs653891	1	228798772	2.007	*
PE 32:1	rs717227	1	236208640	2.14	
%SPM dih 24:0	rs1341446	1	237867929	2.027	*
PLPE 16:0/18:1	rs1025599	1	238623917	1.9	
SPM 20:1	rs3127479	1	240827510	2.121	*
%SPM 20:1	rs3102460	1	240902758	2.292	*
SPM dih 18:0	rs6548222	2	286812	1.909	*
SPM 23:0	rs1370548	2	11932342	2.35	
SPM 20:1	rs956596	2	12978258	2.127	*
%SPM 18:2	rs952275	2	21133051	2.165	*

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3) (continued)

Lipid	Marker	Chromosome	Position	LOD	
PC 40:3	rs2048983	2	35827779	2.2	
PC 38:0	rs1504	2	36977669	2.48	
PLPE 18:1/22:6	rs1014454	2	45530624	2.62	
Monounsaturated SPM	rs992214	2	63398080	2.01	
Total SPM	rs992214	2	63398080	1.91	
Poly-unsaturated SPM	rs1050676	2	65222482	1.95	
%PC O 32:0	rs1015117	2	86652521	2.22	
%PE 40:4	rs880721	2	128869534	2.45	
SPM 20:1	rs113906	2	137492332	2.32	
%SPM dih 24:0	rs892874	2	137734986	2.003	*
SPM dih 18:0	rs1424937	2	169391691	1.936	*
PC 26:0	rs711814	2	176794848	1.93	
PC 30:1	rs1882395	2	191510509	2.06	
PC 38:5	rs1840947	2	193766027	1.98	
% Polyunsaturated PE	rs796283	2	208836633	2.44	
%PC 38:4	rs291335	2	239884101	2.01	
%PC 36:2	rs1483844	3	25401283	2	
PC 36:1	rs1348979	3	27002501	2.35	
%PC 34:1	rs1406568	3	34414318	2.19	
%PC 42:4	rs536036	3	57301923	2.42	
SPM dih 23:0	rs1472653	3	59560632	2.434	*
%SPM dih 23:0	rs1472653	3	59560632	2.105	*
%SPM dih 22:0	rs1562499	3	76685084	2.847	*
%SPM 20:1	rs1470569	3	77614616	2.279	*
%SPM dih 22:0	rs1913081	3	78524925	2.36	*
%SPM dih 22:0	rs1352463	3	95702121	2.165	*
%SPM dih 22:0	rs1519159	3	96650302	2.055	*
%SPM dih 22:0	rs1567058	3	98969929	2.421	*
%SPM dih 22:0	rs591728	3	101916282	2.743	*
SPM dih 22:0	rs1147696	3	121602169	2.084	*
SPM dih 22:0	rs2053820	3	127978004	2.506	*
SPM dih 22:0	rs730257	3	128119054	2.617	*
%PC 38:1	rs1398775	3	147030054	2.31	
PE 32:2	rs482314	3	162947581	2.37	
%PC 38:0	rs1996489	3	167561367	2.22	
%PE 38:2	rs1996489	3	167561367	1.93	
%PC 32:1	rs953834	3	169305526	1.99	
SPM 20:0	rs1468924	3	180465671	2.27	
PC / LPC	rs823515	3	196746841	2.24	
SPM dih 22:0	rs2686085	3	198711008	2.855	*
PE 40:6	rs1981635	4	10279286	2.07	
%PC 34:2	rs1325107	4	17722768	2.31	

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3) (continued)

Lipid	Marker	Chromosome	Position	LOD	
SPM 20:1	rs216113	4	23196799	2.328	*
%SPM 18:2	rs216113	4	23196799	1.9	*
PC 32:2	rs902659	4	32011211	1.92	
PC 38:1	rs725292	4	36965713	2.22	
SPM 20:1	rs992832	4	85013466	2.021	*
PC 38:1	rs732299	4	99352567	2.19	
%PE 38:3	rs713455	4	102504823	2.45	
%PE 38:3	rs1548484	4	106107661	2.68	
%PC 38:1	rs721412	4	111305082	2.43	
%PE 38:3	rs814397	4	117357176	2.19	
%LPC 22:5	rs1157308	4	128040359	2.88	
%SPM 24:2	rs1425566	4	134640602	2.44	
%Glu-CER 16:0	rs6840033	4	141586466	2.43	
%SPM 18:2	rs1014381	4	176923931	1.977	*
CER 23:0	rs12649669	4	178681057	2.03	
Total CER	rs1510629	4	179686414	2.14	
%SPM 18:2	rs2044868	4	180296648	2.043	*
%SPM 17:0	rs158958	5	57992021	2.73	
%SPM dih 24:0	rs264739	5	64401809	2.615	*
SPM dih 18:0	rs1047530	5	75947155	2.194	*
SPM dih 18:0	rs173686	5	82847256	2.147	*
SPM 22:2	rs877826	5	138646696	1.918	*
PC 40:6	rs270664	5	158489316	2.03	
Mono-unsaturated SPM	rs1299048	5	162160289	2.26	
PC O 40:6	rs1030154	5	164983908	2.19	
PLPE 16:0/22:6	rs1030154	5	164983908	1.99	
Glu-CER	rs1445716	5	165352084	1.98	
%PC 38:3	rs1445716	5	165352084	2.16	
SPM-dih	rs1422528	5	166738569	1.93	
% Polyunsaturated PC	rs1054998	5	169548076	1.91	
PLPE 18:0/18:2	rs719065	6	110632	2.62	
PLPE 18:0/20:4	rs719065	6	110632	1.99	
%PLPE 18:0/22:6	rs2815155	6	8010229	1.95	
PC/SPM	rs1150613	6	13288350	2.1	
SPM 20:0	rs1150613	6	13288350	2.08	
PC O 34:2	rs1259078	6	15123286	1.95	
% PLPE 16:0	rs6900454	6	16653636	2.68	
PC O 34:3	rs6901750	6	17127405	2.37	
SPM 24:2	rs1355460	6	23999853	2.53	
% SPM 24:2	rs13161	6	26222681	2.24	
% CER 16:0	rs1537638	6	43204177	2.82	
% PC 38:1	rs1293467	6	44877182	2.11	

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3) (continued)

Lipid	Marker	Chromosome	Position	LOD	
% SPM dih 24:0	rs1925156	6	56215224	1.942	*
SPM 16:1	rs1179900	6	90377879	2.25	
Mono-unsaturated SPM	rs1979797	6	92549591	2.53	
Total SPM	rs1979797	6	92549591	2.4	
SPM 22:1	rs1498252	6	92706956	2.97	
% SPM 15:0	rs491112	6	94226368	2.03	
% SPM 20:1	rs2894891	6	100316628	2.332	*
PC 38:1	rs239189	6	101235508	2.22	
PC O 40:6	rs2040431	6	106177944	2.59	
% PE 36:5	rs1665914	6	107436099	2.17	
PC 40:3	rs1415428	6	114798204	2.1	
PC 38:0	rs1321807	6	117791685	2.14	
% PC 38:0	rs210617	6	117966751	1.9	
% PLPE 16:0/22:6	rs1569741	6	126425169	3.21	
PE 40:6	rs969282	6	134246785	2.08	
% SPM 20:1	rs736784	6	144113703	2.314	*
%SPM dih 24:0	rs4131485	6	149041429	2.799	*
SPM dih 18:0	rs409359	6	159915407	2.08	*
SPM dih 22:0	rs875591	6	166039736	2.671	*
% PC 32:0	rs1362136	7	2278808	2.78	
% PLPE 18:1/18:1	rs6463843	7	8611957	2.13	
% PE 42:7	rs2024046	7	20510506	2.44	
% PLPE 18:1/22:6	rs678798	7	70633763	1.91	
PC 42:4	rs3135677	7	73099782	1.97	
% PC 36:4	rs740158	7	76700487	2.68	
PC O 36:4	rs1799003	7	78210331	1.96	
% PC 32:1	rs9008	7	91147058	1.92	
SPM dih 23:0	rs12217	7	121308604	1.909	*
% PC 40:1	rs322812	7	127338738	2.32	
% Poly-unsaturated SPM	rs691183	7	128698220	2.3	
% PC O 36:1	rs1862083	7	130745104	2	
% Ether-PC	rs4728251	7	131285591	2.2	
% Mono-unsaturated PE	rs1468974	7	132036949	2.08	
% Poly-unsaturated PE	rs889826	7	133053126	2.3	
PC / SPM	rs757723	7	148884802	1.97	
PC 38:4	rs3112	8	2941755	2.41	
PC O 36:5	rs922798	8	3469026	2.22	
PC O 38:4	rs922798	8	3469026	2.21	
PC 40:4	rs922798	8	3469026	2.46	
Ether-PC	rs922798	8	3469026	2.54	
SPM 20:1	rs2442475	8	6344345	2.049	*
Ether-PC	rs310319	8	23746576	2.18	

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3) (continued)

Lipid	Marker	Chromosome	Position	LOD	
SPM 24:2	rs879958	8	25026922	2.88	
% SPM 18:0	rs1000236	8	38593641	2.23	
% SPM 24:3	rs1108	8	67543417	2.053	*
% SPM 14:0	rs906998	8	78693270	2.42	
% PC 36:4	rs20571	8	86539249	2.49	
% SPM dih 22:0	rs1902866	8	87922048	2.306	*
SPM 22:2	rs2034843	8	121412955	1.946	*
% SPM 22:2	rs2034843	8	121412955	1.969	*
SPM 22:2	rs2034844	8	121413208	1.969	*
% SPM 22:2	rs2034844	8	121413208	1.946	*
% PLPE 18:1/20:5	rs7386971	8	144387314	2.76	
% PC O 42:6	rs1475656	9	30901514	1.95	
%SPM dih 23:0	rs735741	9	37934489	1.929	*
%SPM dih 23:0	rs1138374	9	37964743	1.907	*
pep_poly	rs927632	9	76389547	2.35	
%PLPE 18:0/18:1	rs2378592	9	81724781	2.47	
SPM 24:1	rs1048510	9	94400682	2.76	
PC O 36:5	rs1338121	9	100094573	2.26	
%PLPE 16:0/20:4	rs1329088	9	100960039	2.28	
%PLPE 18:0/18:2	rs363717	9	104624255	2.88	
% CER 18:0	rs6477450	9	105518992	2.52	
CER 24:0	rs1405	9	116034178	2.12	
LPC 18:3	rs913932	9	122029298	2.05	
SPM 23:0	rs1054879	9	130075322	2.16	
SPM 18:0	rs705670	9	135694529	2.14	
%LPC 20:0	rs3814595	10	3191679	2.41	
PE 38:5	rs1993181	10	4881168	2.42	
%PC O 38:4	rs1041226	10	6683857	2.59	
PE 40:5	rs942434	10	7277013	2.29	
%PC 38:4	rs1623807	10	9116302	2.45	
%PC 36:4	rs913375	10	10780754	2.33	
%PE 40:4	rs1001201	10	14083458	1.94	
%PC O 32:0	rs927099	10	33565842	3.11	
%SPM 20:1	rs1345561	10	35810628	1.921	*
%SPM 16:0	rs1208789	10	37762478	1.97	
spmp_pol	rs722317	11	15880138	2.11	
SPM 22:2	rs1499511	11	36044429	2.315	*
%PLPE 18:0/18:1	rs4755782	11	44124572	1.93	
Glu-CER	rs612415	11	60616462	2.12	
PC 36:2	rs591804	11	73040858	2.04	
%PE 36:2	rs591804	11	73040858	2.36	
%SPM dih 18:0	rs1247726	11	79644830	2.407	*

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3) (continued)

Lipid	Marker	Chromosome	Position	LOD	
%PLPE 18:0/20:4	rs1459937	11	81203361	1.94	
%SPM 22:2	rs586699	11	91929382	2.168	*
%SPM 24:2	rs1785872	11	104474890	3.22	
PC 30:0	rs609177	11	116275727	2.44	
pcBYlpc	rs1073636	11	119477658	2.38	
LPC 16:0	rs575030	11	120597608	2.22	
PC 36:5	rs1894078	11	122104161	3.08	
%PC 36:5	rs1894078	11	122104161	2.14	
PC 38:6	rs2001625	11	130103423	2.55	
%PC 38:0	rs2001625	11	130103423	2.01	
%PLPE 18:0/18:2	rs1005394	12	1311398	2.37	
%PLPE 16:0/18:2	rs1420725	12	2621844	2.12	
%PC 40:1	rs2365567	12	6714108	1.97	
%PC 40:6	rs1560011	12	9714219	2.91	
PC 40:6	rs306657	12	27585118	2.02	
%PC 40:6	rs306657	12	27585118	2.31	
%PE O 40:3	rs1705772	12	34066023	2.32	
%E 40:6	rs965125	12	44008489	2.82	
%PLPE 18:0/22:6	rs1542707	12	46921444	2.03	
PLPE 18:0/22:6	rs1471998	12	46998269	1.92	
%PC 38:1	rs1867299	12	52616242	2.61	
%PE 38:2	rs1240267	12	68570395	1.97	
PC 30:1	rs1433251	12	71362298	2.18	
PC 38:7	rs2037581	12	75302109	2.31	
%PC 38:5	rs2061589	12	86796951	1.99	
%LPC 16:0	rs1470106	12	95309585	2.62	
lpcp_sat	rs1470106	12	95309585	2.51	
%18_0	rs1544921	12	100613641	1.91	
%LPC 18:2	rs1863527	12	106711018	1.98	
PE 36:5	rs1642031	12	108024955	1.91	
SPM 16:0	rs7295288	12	117318533	2.44	
spm_sat	rs7295288	12	117318533	2.27	
%18_1	rs977655	13	25202569	2.22	
PC O 42:6	rs1544295	13	42396689	1.91	
%PC O 42:6	rs299344	13	43161589	2.07	
SPM dih 24:0	rs1382461	13	54063651	2.48	*
PC O 34:2	rs1545963	13	63751543	1.91	
PC O 40:6	rs1372177	13	78469870	2.99	
%PC O 40:6	rs2042766	13	87635854	2.27	
SPM dih 23:0	rs1925121	13	95736781	2.279	*
%PC 38:3	rs536863	13	101627600	2.02	
LPC 22:4	rs1959344	14	19764442	1.91	

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3) (continued)

Lipid	Marker	Chromosome	Position	LOD	
%PC 36:2	rs2378944	14	31272229	1.96	
PC O 34:2	rs311848	14	58270833	2.33	
PC O 36:2	rs311848	14	58270833	2.25	
PC O 38:3	rs229670	14	64353254	1.92	
%PC O 36:3	rs1015023	14	67431984	2.95	
%PC O 38:3	rs6574106	14	72455419	2.16	
SPM dih 23:0	rs1125221	14	73982583	2.018	*
% CER 23:0	rs178384	14	79252594	2.13	
SPM dih 23:0	rs1846362	15	22119999	1.966	*
PC 40:7	rs969860	15	23262304	2.07	
SPM dih 18:0	rs1055356	15	32935394	1.904	*
%SPM 24:1	rs1989223	15	33914398	2.44	
SPM 24:1	rs1075456	15	61421458	2.55	
%PE 42:6	rs2001597	15	68208915	2.74	
LPC 16:1	rs1553650	15	78899504	2.1	
%PC 38:5	rs1007328	15	94504377	2.4	
%SPM 22:2	rs7168512	15	99004156	2.186	*
PC O 32:1	rs1203974	16	217459	1.91	
%PE 36:5	rs3178656	16	1653514	3	
PE 36:5	rs879416	16	2856655	2.02	
%PE 40:5	rs11228	16	3212392	2.79	
SPM dih 22:0	rs1012259	16	5892910	2.051	*
%PC O 36:3	rs1035564	16	6050548	2.65	
pep_alky	rs1035564	16	6050548	1.96	
%SPM dih 23:0	rs933478	16	6224150	2.18	*
%PC 32:1	rs1507038	16	7975776	1.95	
SPM dih 22:0	rs1646129	16	29048189	2.139	*
SPM 20:1	rs837529	16	54043795	2.301	*
PC 38:1	rs41383	16	55668503	1.96	
%PC 34:1	rs247041	16	56435141	2.38	
pcp_mono	rs247041	16	56435141	2.21	
lpc_sat	rs153672	16	61559144	2.03	
%PC 34:1	rs153672	16	61559144	2.39	
%LPC 22:4	rs328384	16	72397297	2.42	
%PE 34:3	rs877375	16	73292368	2.03	
SPM dih 23:0	rs1424234	16	78240204	2.509	*
PE 40:6	rs2059280	16	79891932	2.04	
%LPC 18:2	rs725514	16	81412395	2.21	
%PLPE 18:0/22:6	rs922450	16	82990535	2.14	
%PLPE 16:0/22:5	rs454087	16	83089803	3.09	
%PC O 40:6	rs2076962	16	85717982	2.5	
PE 36:3	rs3794668	16	86301992	2	

TABLE S3 Suggestive linkage findings (1.9>LOD>3.3) (continued)

Lipid	Marker	Chromosome	Position	LOD	
%PE 38:6	rs1984749	17	2196654	2.39	
%PLPE 16:0/22:5	rs178815	17	16033654	3.13	
PE 36:1	rs2017167	17	29554828	2.29	
%PE O 38:7	rs2017167	17	29554828	2.02	
pe_mono	rs758299	17	30162722	2.37	
%PE 40:4	rs1526601	17	36088662	1.94	
PE 38:2	rs1019977	18	17257904	2.15	
PE 38:2	rs291781	18	17258111	2.15	
%SPM dih 22:0	rs1878677	18	38984443	1.905	*
%PC O 34:0	rs1822459	18	46628125	1.94	
%SPM 18:2	rs732982	18	52878457	1.986	*
PC O 36:3	rs757528	19	1733798	2.05	
SPM dih 18:0	rs1529729	19	11024562	2.078	*
% CER 24:0	rs1015902	19	14883931	2.07	
SPM dih 18:0	rs1075403	19	18245950	2.169	*
%SPM 20:1	rs1973371	19	23515487	1.967	*
%SPM dih 23:0	rs450542	19	23834737	2.159	*
%SPM dih 23:0	rs251746	19	23841888	2.144	*
%SPM dih 23:0	rs251743	19	23848208	2.243	*
%SPM 20:1	rs718066	19	45013680	2.755	*
%PC 40:3	rs725660	19	50954126	3.09	
%PC O 34:0	rs1045764	19	58303285	1.96	
PE 42:5	rs36633	19	59338102	2.06	
%PC O 36:3	rs382592	19	62051612	2.76	
PC O 38:3	rs1469781	20	118642	3.24	
%PE 42:5	rs4816023	20	780763	1.97	
LPC 22:5	rs1010310	20	44268451	2.04	
SPM 20:1	rs511145	20	56337304	1.949	*
%PE 34:0	rs12034	21	17864094	2.02	
SPM 22:2	rs2830252	21	26776572	1.924	*
%PLPE 18:0/20:4	rs2831032	21	27917072	1.94	
PE 38:3	rs1044195	21	33743677	2.21	
%PE 40:5	rs2839576	21	43013026	1.93	
%PE 36:2	rs2838325	21	43838712	2.31	
PC O 40:6	rs2839377	21	46902240	2.04	
LPC 20:4	rs761793	22	47288589	2.59	

* Indicates that the results obtained by 2-point linkage analysis of traits with point-mass distribution.

Text S1. Supplementary methods

Study population

ERF is a family-based study which includes over 3000 individuals descendant from 22 couples living in the Rucphen region in the southwest Netherlands in the 19th century. All participants were invited to visit the clinical research centre in the region where they completed extensive physical examinations, questionnaires, and interviews and where fasting blood samples were drawn [1]. Symptoms of depression and anxiety were assessed using the Hospital Anxiety and Depression Scale (HADS-D for depression and HADS-A for symptoms of anxiety) and the Center for Epidemiological Studies Depression Scale (CES-D) questionnaires [2, 3]. The ERF study was approved by the Medical Ethics Review Committee of the Erasmus University Medical Centre in Rotterdam, the Netherlands and all participants provided written informed consent.

Lipidomics

A broad range of phospho- and sphingolipid species (n = 148) were measured in 820 participants. Lipid species were quantified by electrospray ionization tandem mass spectrometry (ESIMS/ MS) using methods validated and described previously [4, 5]. In brief, samples were analyzed by direct flow injection using a precursor ion scan of *m/z* 184 specific for phosphocholine containing lipids including phosphatidylcholine (PC), SPM [5] and lysophosphatidylcholine (LPC) [4]. A neutral loss scan of *m/z* 141 was used for phosphatidylethanolamine (PE) [6] and PE-based plasmalogens (PLPE) and was analyzed according to the principles described by Zemski-Berry [7]. In brief, fragment ions of *m/z* 364, 380 and 382 were used for PE p16:0, p18:1 and p18:0 species, respectively. Quantification was achieved by calibration lines generated by the addition of naturally occurring lipid species to plasma and internal standards belonging to the same lipid class (PC 14:0 / 14:0, PC 22:0 / 22:0, PE 14:0 / 14:0, PE 20:0 / 20:0, LPC 13:0, LPC 19:0). Calibration lines were generated for the following naturally occurring species: PC 34:1, 36:2, 38:4, 40:0 and PC O 16:0 / 20:4; LPC 16:0, 18:1, 18:0; PE 34:1, 36:2, 38:4, 40:6 and PLPE 16:0/20:4. Correction of isotopic overlap of lipid species as well as data analysis was performed by self-programmed Excel macros for all lipid classes according to the principles described previously [5]. Nomenclature of SPM species is based on the assumption that d18:1 (dihydroxy 18:1 sphingosine) is the main base of plasma SPM, where the first number refers to the number of carbon atoms in the chain and the second number to the number of double bonds in the chain. Further details can be found in the supplementary table. The performed analysis does not always allow an exact assignment. In this case, an "O" is added to the subspecies name, e.g., PC O 36:5 and PC O 32:1. This denotes that the two species are most likely be assigned to PC species containing an ether bond (alkyl) and may constitute plasmalogens. However, we cannot exclude the possibility that

PC O 36:5 might be assigned to PC 35:5, an unlikely odd carbon number species. Similarly, PC O 32:1 could be assigned to PC 31:1.

Additionally, in an independent sample of 1000 ERF individuals, targeted metabolite profiling by electrospray ionization (ESI) tandem mass spectrometry (MS/MS) was performed using a quantitative metabolomics platform (*Biocrates Life Sciences AG*, Austria). 753 individuals with both Biocrates measurements and depression questionnaire data served as a partial replication dataset for species measured in both platforms.

Statistical analysis

In addition to the measured lipids, the proportions of each lipid in its own class, (for example, the proportion of SPM 23:1 among total SPM) were calculated. These proportions are valuable in assessing differences in concentrations that are related to within-class turnover. Partial correlation analysis, corrected for age and sex, were performed using R (http://www.r-project.org/). Significant associations between the HADS-D, HADS-A and CES-D scales and lipids were further assessed with the SOLAR (Sequential Oligogenic Linkage Analysis Routines) 4.1.5 software package (Southwest Foundation for Biomedical Research, San Antonio, Texas, USA) [8] using the "polygenic" option to adjust for relatedness. Due to computational limitations whole pedigree was split in non-overlapping sub pedigrees using PedCut program [9] [10]. Many of the lipid levels measured in our study are highly correlated to each other. For this reason, we used a data reduction strategy to estimate the number of independent observations. In total 23 principal components accounted for 79% of the phenotypic variance of all lipids. As we he have studied 3 outcomes for depression and anxiety, we defined the threshold for statistical significance as $0.05/(23 \times 3)$, i.e. 7.25×10^{-4} for the single species. For the lipid ratio matrix, it was estimated as $0.05/(((23 \times 23) - 23)/2)$; 6.59×10^{-5}. Final stage meta-analysis of the discovery and replication set were performed using the correlation coefficients (*r*) and sample size, through Fisher's r to Z transformation [11].

REFERENCES

1 Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* 2004; **12**(7): 527-534.

2 Radloff. The CES-D scale: a self report depression scale for research in theb general population. *Appl Pshycol Measurement* 1977; **3:** 385-401.

3 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**(6): 361-370.

4 Liebisch G, Drobnik W, Lieser B, Schmitz G. High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clinical chemistry* 2002; **48**(12): 2217-2224.

5 Liebisch G, Lieser B, Rathenberg J, Drobnik W, Schmitz G. High-throughput quantification of phosphatidylcholine and sphingomyelin by electrospray ionization tandem mass spectrometry coupled with isotope correction algorithm. *Biochimica et biophysica acta* 2004; **1686**(1-2): 108-117.

6 Brugger B, Erben G, Sandhoff R, Wieland FT, Lehmann WD. Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 1997; **94**(6): 2339-2344.

7 Zemski Berry KA, Murphy RC. Electrospray ionization tandem mass spectrometry of glycerophosphoethanolamine plasmalogen phospholipids. *Journal of the American Society for Mass Spectrometry* 2004; **15**(10): 1499-1508.

8 Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American journal of human genetics* 1998; **62**(5): 1198-1211.

9 Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet* 2008; **16**(7): 854-860.

10 Demirkan A, Penninx BW, Hek K, Wray NR, Amin N, Aulchenko YS *et al.* Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Molecular psychiatry*; **16**(7): 773-783.

11 Fisher RA. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika (Biometrika Trust)* 1915; **10**(4): 507–521.

TABLE S1 Correlation between plasma phospho-sphingolipids and depression and anxiety symptoms.

A. Sphingomyelins

	ABSOLUTE VALUES						PROPORTIONS					
	HADS-D		CES-D		HADS-A		HADS-D		CES-D		HADS-A	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
SPM 14:0	-0.04	0.24	-0.04	0.27	-0.05	0.22	-0.04	0.31	-0.02	0.55	-0.03	0.48
SPM 15:0	-0.06	0.08	-0.06	0.12	-0.07	0.07	-0.09	**0.02**	-0.06	0.09	-0.07	0.07
SPM 16:1	0.01	0.78	0.01	0.69	-0.03	0.40	0.03	0.41	0.05	0.18	-0.02	0.56
SPM 16:0	0.04	0.31	-0.01	0.85	0.01	0.84	0.10	**0.01**	0.06	0.12	0.07	0.06
SPM dih 16:0	0.02	0.64	-0.01	0.85	0.00	0.92	0.02	0.55	0.00	0.99	0.00	0.96
SPM 16:1-OH	-0.03	0.47	-0.03	0.36	-0.08	**0.04**	-0.02	0.50	-0.02	0.51	-0.08	**0.02**
SPM 18:2	0.01	0.76	0.00	0.93	-0.05	0.21	0.03	0.46	0.01	0.71	-0.04	0.28
SPM 18:1	0.00	0.90	0.02	0.51	-0.04	0.33	0.01	0.83	0.04	0.27	-0.05	0.17
SPM 18:0	0.01	0.82	0.00	0.92	-0.03	0.37	0.02	0.64	0.03	0.47	-0.04	0.26
SPM dih 18:0	-0.01	0.69	0.00	0.95	-0.04	0.25	-0.01	0.85	0.01	0.87	-0.04	0.31
SPM 20:1	0.04	0.29	0.02	0.62	0.02	0.52	0.06	0.08	0.06	0.12	0.05	0.17
SPM 20:0	-0.04	0.29	-0.03	0.49	0.01	0.84	-0.04	0.26	-0.01	0.86	0.03	0.40
SPM 22:2	-0.05	0.20	-0.04	0.27	-0.08	**0.04**	-0.05	0.18	-0.05	0.21	-0.08	**0.03**
SPM 22:1	-0.03	0.43	-0.04	0.34	-0.04	0.30	-0.05	0.20	-0.05	0.17	-0.05	0.14
SPM 22:0	-0.03	0.36	-0.03	0.36	-0.01	0.78	-0.04	0.33	-0.01	0.78	0.02	0.58
SPM dih 22:0	-0.05	0.18	0.01	0.85	0.04	0.26	-0.05	0.15	0.01	0.79	0.05	0.18
SPM 23:1	-0.08	**0.03**	-0.07	0.05	-0.08	**0.03**	-0.14	**0.00**	-0.11	**0.00**	-0.12	**0.00**
SPM 23:0	-0.06	0.13	-0.06	0.08	-0.05	0.21	-0.08	**0.02**	-0.08	**0.03**	-0.05	0.16
SPM dih 23:0	-0.02	0.65	-0.03	0.40	-0.02	0.51	-0.02	0.67	-0.03	0.35	-0.03	0.45
SPM 24:3	-0.04	0.31	-0.04	0.23	-0.05	0.20	-0.04	0.31	-0.05	0.22	-0.04	0.28
SPM 24:2	-0.01	0.83	-0.01	0.84	0.00	0.91	-0.01	0.75	0.00	0.92	0.00	0.98
SPM 24:1	0.02	0.63	0.00	0.90	0.01	0.76	0.04	0.30	0.02	0.68	0.04	0.27
SPM 24:0	-0.02	0.67	-0.02	0.65	0.00	0.92	-0.03	0.42	-0.01	0.69	0.02	0.59
SPM dih 24:0	0.00	0.99	-0.02	0.53	-0.02	0.52	0.01	0.76	-0.01	0.72	-0.01	0.69
Saturated SPM	0.00	0.97	-0.02	0.53	-0.01	0.88	0.01	0.79	0.00	0.96	0.04	0.28
Monounsaturated SPM	0.00	0.99	-0.01	0.80	-0.02	0.65	0.00	0.96	0.00	0.97	-0.03	0.45
Poly-unsaturated SPM	-0.02	0.64	-0.02	0.64	-0.03	0.36	-0.05	0.22	-0.03	0.40	-0.06	0.10
Unsaturated SPM	-0.01	0.88	-0.01	0.78	-0.02	0.57	-0.01	0.79	0.00	0.96	-0.04	0.28
Dihydroxylated SPM	-0.03	0.45	-0.01	0.73	0.02	0.56	-0.02	0.57	0.00	0.90	0.03	0.35
TOTAL SPM	-0.01	0.89	-0.02	0.60	-0.01	0.69						

B. Phosphotidylcholines

	ABSOLUTE VALUES						PROPORTIONS					
	HADS-D		CES-D		HADS-A		HADS-D		CES-D		HADS-A	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
PC 26:0	-0.04	0.24	-0.04	0.22	-0.07	**0.04**	-0.03	0.37	-0.05	0.18	-0.06	0.08
PC 30:1	-0.01	0.86	0.04	0.26	-0.02	0.66	-0.01	0.85	0.04	0.25	-0.02	0.67
PC 30:0	0.03	0.36	0.00	0.93	0.00	0.91	0.05	0.16	0.00	0.97	0.00	0.94
PC O 32:1	0.01	0.75	-0.01	0.88	0.00	0.91	0.04	0.30	0.00	1.00	0.00	0.97
PC O 32:0	0.00	0.97	-0.03	0.37	-0.03	0.35	0.02	0.68	-0.04	0.28	-0.04	0.30
PC 32:2	-0.02	0.59	-0.01	0.82	-0.02	0.67	-0.01	0.76	-0.01	0.70	-0.02	0.64
PC 32:1	0.00	0.94	0.02	0.54	0.02	0.68	0.01	0.74	0.03	0.49	0.01	0.71
PC 32:0	0.01	0.87	0.01	0.71	-0.01	0.79	0.04	0.34	0.00	0.96	-0.01	0.69
PC O 34:3	-0.07	0.06	-0.06	0.09	-0.07	**0.04**	-0.06	0.10	-0.07	0.06	-0.07	0.05
PC O 34:2	-0.11	**0.003**	-0.10	**0.01**	-0.09	**0.02**	-0.07	0.08	-0.09	**0.02**	-0.07	0.08
PC O 34:1	-0.01	0.82	-0.01	0.86	-0.04	0.24	0.03	0.44	-0.01	0.80	-0.05	0.16
PC O 34:0	0.00	1.00	0.01	0.83	0.04	0.32	0.02	0.62	0.00	0.90	0.04	0.28
PC 34:4	-0.03	0.48	-0.02	0.55	-0.05	0.20	-0.02	0.60	-0.04	0.25	-0.07	0.07
PC 34:3	0.01	0.87	0.03	0.46	0.01	0.69	0.04	0.34	0.02	0.56	0.01	0.84
PC 34:2	-0.01	0.71	0.02	0.57	0.03	0.45	0.02	0.51	0.03	0.41	0.04	0.22
PC 34:1	0.01	0.79	0.02	0.54	0.02	0.66	0.05	0.16	0.03	0.39	0.03	0.37
PC 34:0	-0.04	0.25	-0.01	0.69	-0.02	0.53	-0.05	0.20	-0.02	0.51	-0.04	0.29
PC O 36:5	-0.09	**0.01**	-0.08	**0.03**	-0.08	**0.03**	-0.08	**0.03**	-0.09	**0.01**	-0.09	**0.01**
PC O 36:4	-0.12	**0.0007**	-0.09	**0.01**	-0.13	**0.0003**	-0.11	**0.003**	-0.11	**0.002**	-0.14	**0.0001**
PC O 36:3	-0.09	**0.01**	-0.06	0.12	-0.04	0.33	-0.07	0.06	-0.06	0.12	-0.03	0.43
PC O 36:2	-0.03	0.35	-0.04	0.29	-0.02	0.59	0.01	0.78	-0.03	0.46	-0.01	0.72
PC O 36:1	-0.06	0.13	-0.03	0.35	-0.05	0.17	-0.03	0.40	-0.03	0.39	-0.07	0.05
PC O 36:0	-0.03	0.37	-0.01	0.89	-0.03	0.48	-0.01	0.81	0.01	0.87	-0.02	0.58
PC 36:5	-0.04	0.23	-0.04	0.31	-0.03	0.49	-0.04	0.24	-0.06	0.08	-0.03	0.36
PC 36:4	-0.06	0.13	-0.02	0.56	-0.04	0.33	-0.06	0.09	-0.06	0.12	-0.07	0.07
PC 36:3	-0.01	0.80	0.03	0.41	0.02	0.55	0.00	0.99	0.01	0.75	0.03	0.47
PC 36:2	0.02	0.54	0.04	0.30	0.06	0.13	0.08	**0.03**	0.05	0.19	0.09	**0.02**
PC 36:1	0.05	0.15	0.04	0.27	0.04	0.27	0.10	**0.01**	0.05	0.20	0.05	0.14
PC 36:0	0.00	0.99	-0.02	0.59	-0.03	0.43	0.02	0.59	-0.02	0.57	-0.03	0.48
PC O 38:5	-0.10	**0.01**	-0.09	**0.01**	-0.11	**0.009**	-0.08	**0.02**	-0.11	**0.002**	-0.12	**0.001**
PC O 38:4	-0.08	**0.03**	-0.07	0.06	-0.10	**0.01**	-0.06	0.08	-0.08	**0.02**	-0.09	**0.01**
PC O 38:3	0.00	0.92	0.01	0.87	-0.01	0.69	0.03	0.39	0.00	0.95	-0.02	0.51
PC O 38:2	0.05	0.21	0.03	0.43	0.05	0.18	0.07	0.08	0.03	0.49	0.05	0.17

B. Phosphotidylcholines (continues)

	ABSOLUTE VALUES						PROPORTIONS					
	HADS-D		CES-D		HADS-A		HADS-D		CES-D		HADS-A	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
PC O 38:1	0.06	0.11	0.07	**0.05**	0.04	0.23	0.06	0.09	0.07	0.05	0.04	0.27
PC 38:7	-0.04	0.33	-0.01	0.69	-0.04	0.34	-0.03	0.40	-0.02	0.53	-0.04	0.28
PC 38:6	-0.07	**0.04**	-0.03	0.49	-0.05	0.19	-0.07	**0.04**	-0.05	0.19	-0.06	0.11
PC 38:5	-0.04	0.29	0.00	0.96	-0.02	0.50	-0.03	0.45	-0.02	0.54	-0.04	0.25
PC 38:4	-0.04	0.29	-0.01	0.77	-0.03	0.35	-0.01	0.68	-0.01	0.69	-0.05	0.22
PC 38:3	-0.01	0.84	0.02	0.67	-0.01	0.73	0.00	0.99	0.00	0.90	-0.02	0.64
PC 38:2	0.01	0.87	0.03	0.40	0.02	0.59	0.02	0.51	0.04	0.30	0.02	0.56
PC 38:1	-0.04	0.32	0.03	0.41	0.02	0.52	-0.03	0.46	0.02	0.50	0.03	0.48
PC 38:0	-0.06	0.11	0.00	0.92	-0.04	0.32	-0.04	0.23	0.00	0.99	-0.04	0.23
PC O 40:6	-0.09	**0.02**	-0.04	0.31	-0.07	0.05	-0.05	0.16	-0.03	0.38	-0.07	0.05
PC O 40:5	0.01	0.76	0.02	0.53	0.00	0.97	0.04	0.30	0.02	0.67	0.00	0.97
PC O 40:4	-0.08	**0.03**	-0.06	0.11	-0.05	0.14	-0.07	0.07	-0.07	0.06	-0.06	0.09
PC 40:7	-0.04	0.27	0.01	0.82	-0.02	0.58	-0.02	0.63	0.01	0.87	-0.02	0.61
PC 40:6	-0.04	0.27	0.00	0.98	-0.02	0.50	-0.03	0.45	0.00	0.97	-0.02	0.62
PC 40:5	0.00	0.97	0.02	0.59	0.01	0.75	0.02	0.53	0.03	0.35	0.02	0.65
PC 40:4	-0.01	0.68	0.00	0.89	0.00	0.92	0.01	0.74	0.00	0.96	0.01	0.82
PC 40:3	-0.06	0.11	0.00	0.94	-0.04	0.32	-0.05	0.14	-0.01	0.87	-0.04	0.25
PC 40:2	0.00	0.99	0.05	0.19	0.01	0.80	0.02	0.67	0.05	0.18	0.01	0.82
PC 40:1	-0.03	0.46	0.04	0.27	0.01	0.81	-0.01	0.73	0.04	0.32	0.01	0.75
PC 40:0	-0.04	0.31	-0.02	0.63	-0.02	0.55	-0.03	0.36	-0.03	0.42	-0.02	0.62
PC O 42:6	-0.05	0.15	-0.02	0.51	-0.03	0.37	-0.04	0.28	-0.02	0.59	-0.03	0.43
PC O 42:5	0.01	0.78	-0.02	0.50	0.00	1.00	0.02	0.61	-0.04	0.31	0.00	0.93
PC 42:5	-0.04	0.31	0.02	0.64	0.00	0.96	-0.04	0.25	0.01	0.80	0.00	1.00
PC 42:4	-0.04	0.26	-0.02	0.54	-0.04	0.24	-0.04	0.34	-0.04	0.32	-0.05	0.16
Saturated PC	-0.02	0.54	-0.01	0.85	-0.03	0.45	-0.02	0.63	-0.04	0.32	-0.06	0.11
Mono-unsaturated PC	0.01	0.74	0.03	0.44	0.02	0.51	0.06	0.11	0.04	0.28	0.04	0.28
Poly-unsaturated PC	-0.02	0.54	0.02	0.65	0.01	0.76	-0.02	0.68	0.01	0.76	0.00	0.92
Unsaturated PC	-0.01	0.71	0.02	0.50	0.02	0.60	0.06	0.09	0.09	**0.02**	0.09	**0.01**
Ether PC*	-0.09	**0.01**	-0.08	**0.04**	-0.08	**0.02**	-0.07	0.07	-0.09	**0.01**	-0.09	**0.02**
Total PC	-0.02	0.57	0.02	0.63	0.01	0.75						

C. Phosphotidylethanolamines

	ABSOLUTE VALUES						PROPORTIONS					
	HADS-D		CES-D		HADS-A		HADS-D		CES-D		HADS-A	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
PE 32:2	0.03	0.40	0.00	0.98	-0.04	0.33	-0.04	0.30	-0.08	0.02	-0.08	0.03
PE 32:1	0.06	0.08	0.07	0.05	0.03	0.40	0.00	0.98	0.00	0.96	-0.01	0.71
PE 32:0	0.05	0.16	0.02	0.58	-0.01	0.82	-0.01	0.70	-0.06	0.11	-0.05	0.20
PE 34:3	0.07	0.05	0.06	0.11	0.02	0.51	0.04	0.24	-0.01	0.88	-0.01	0.77
PE 34:2	0.07	**0.04**	0.07	**0.05**	0.05	0.15	0.06	0.09	0.05	0.16	0.05	0.19
PE 34:1	0.03	0.38	0.04	0.24	0.01	0.79	-0.05	0.22	-0.03	0.44	-0.06	0.11
PE 34:0	0.04	0.26	0.00	0.94	-0.03	0.47	-0.01	0.76	-0.05	0.15	-0.05	0.14
PE 36:5	0.03	0.39	0.04	0.31	0.01	0.88	-0.06	0.09	-0.07	**0.05**	-0.05	0.15
PE 36:4	0.05	0.19	0.05	0.16	0.02	0.60	-0.03	0.49	-0.02	0.66	-0.05	0.16
PE 36:3	0.06	0.10	0.06	0.09	0.05	0.21	0.01	0.80	0.01	0.72	0.02	0.53
PE 36:2	0.07	**0.04**	0.08	**0.02**	0.06	0.09	0.04	0.27	0.04	0.24	0.05	0.17
PE 36:1	0.04	0.30	0.06	0.11	0.02	0.59	-0.03	0.36	-0.03	0.49	-0.05	0.16
PE O 38:7	-0.05	0.15	-0.04	0.22	-0.05	0.17	-0.10	**0.01**	-0.10	**0.01**	-0.06	0.09
PE 38:6	0.05	0.21	0.07	0.07	0.04	0.29	-0.02	0.65	0.00	0.94	0.00	0.92
PE 38:5	0.05	0.17	0.06	0.11	0.04	0.23	-0.01	0.87	0.00	0.90	0.02	0.66
PE 38:4	0.05	0.17	0.06	0.09	0.04	0.31	0.01	0.73	0.00	0.90	0.01	0.89
PE 38:3	0.05	0.16	0.05	0.15	0.04	0.33	-0.02	0.50	-0.04	0.25	-0.03	0.39
PE 38:2	0.07	0.07	0.04	0.23	0.03	0.40	-0.03	0.48	-0.04	0.27	-0.02	0.65
PE 38:1	-0.03	0.41	-0.05	0.16	-0.01	0.71	-0.09	**0.01**	-0.12	**0.001**	-0.06	0.09
PE O 40:3	0.03	0.40	0.04	0.23	0.03	0.49	-0.08	**0.02**	-0.09	**0.01**	-0.07	0.05
PE 40:6	0.05	0.20	0.08	**0.02**	0.04	0.30	-0.01	0.77	0.03	0.37	0.01	0.79
PE 40:5	0.08	**0.03**	0.07	**0.04**	0.03	0.34	0.06	0.09	0.04	0.25	0.01	0.71
PE 40:4	0.09	**0.02**	0.07	0.06	0.02	0.50	0.05	0.14	0.02	0.59	-0.03	0.41
PE 40:3	0.02	0.56	-0.02	0.68	-0.04	0.22	-0.04	0.30	-0.08	**0.03**	-0.07	0.07
PE 42:7	0.01	0.70	0.01	0.79	-0.02	0.64	-0.05	0.16	-0.08	**0.04**	-0.07	**0.04**
PE 42:6	0.09	**0.01**	0.04	0.28	0.02	0.66	0.01	0.70	-0.05	0.19	-0.03	0.45
PE 42:5	0.03	0.39	0.02	0.65	-0.01	0.77	-0.03	0.37	-0.06	0.09	-0.06	0.10
Saturated PE	0.05	0.19	0.01	0.88	-0.03	0.38	-0.02	0.58	-0.07	0.07	-0.06	0.08
Mono-unsaturated PE	0.03	0.38	0.05	0.21	0.01	0.73	-0.07	0.06	-0.05	0.16	-0.07	**0.04**
Poly-unsaturated PE	0.06	0.08	0.07	0.05	0.05	0.21	0.10	**0.01**	0.10	**0.01**	0.10	**0.004**
Unsaturated PE	0.06	0.10	0.07	0.06	0.04	0.24	0.08	**0.03**	0.10	**0.01**	0.08	**0.02**
Ether PE	0.01	0.73	0.03	0.41	0.01	0.78	-0.10	**0.01**	-0.10	**0.004**	-0.07	**0.04**
Total PE	0.06	0.10	0.07	0.07	0.04	0.26						

D. Phosphotidylethanolamine-plasmologens

	ABSOLUTE VALUES						PROPORTIONS					
	HADS-D		CES-D		HADS-A		HADS-D		CES-D		HADS-A	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
PLPE 16:0/18:2	0.00	1.00	0.00	0.89	-0.01	0.80	0.06	0.10	0.05	0.20	0.06	0.13
PLPE 18:0/20:5	0.02	0.67	0.00	0.94	0.00	1.00	0.06	0.10	0.04	0.27	0.06	0.12
PLPE 16:0/20:4	-0.05	0.18	-0.03	0.35	-0.04	0.29	-0.04	0.25	-0.03	0.43	-0.01	0.74
PLPE 16:0/22:6	-0.06	0.09	-0.05	0.16	-0.09	**0.01**	-0.06	0.11	-0.05	0.21	-0.08	**0.03**
PLPE 18:1/18:1	-0.07	**0.05**	-0.05	0.15	-0.09	**0.01**	-0.07	**0.04**	-0.04	0.24	-0.08	**0.03**
PLPE 18:1/22:6	-0.01	0.70	-0.02	0.64	-0.06	0.09	0.03	0.45	0.01	0.87	-0.04	0.34
PLPE 16:0/20:5	-0.02	0.53	-0.01	0.83	0.02	0.60	0.03	0.43	0.03	0.49	0.11	**0.003**
PLPE 18:0/22:6	0.01	0.83	0.01	0.77	0.00	1.00	0.06	0.13	0.05	0.19	0.07	0.06
PLPE 18:0/18:1	-0.05	0.22	-0.05	0.14	-0.01	0.86	-0.04	0.30	-0.07	0.07	0.02	0.57
PLPE 18:1/18:2	-0.09	**0.01**	-0.07	0.06	-0.08	**0.03**	-0.10	**0.01**	-0.07	0.05	-0.05	0.21
PLPE 18:1/20:4	-0.06	0.11	-0.02	0.52	-0.04	0.32	-0.05	0.17	-0.01	0.84	-0.01	0.84
PLPE 16:0/18:1	0.02	0.53	0.02	0.59	0.02	0.64	0.10	**0.004**	0.08	**0.02**	0.10	**0.004**
PLPE 18:0/20:4	0.01	0.86	0.00	0.94	0.00	0.93	0.04	0.23	0.03	0.47	0.06	0.08
PLPE 18:0/18:2	-0.02	0.54	-0.04	0.31	-0.02	0.51	-0.02	0.68	-0.05	0.19	0.00	0.90
PLPE 16:0/22:5	-0.05	0.19	-0.04	0.25	-0.07	0.05	-0.02	0.59	0.00	0.98	-0.02	0.63
PLPE 18:1/20:5	-0.05	0.16	-0.02	0.56	-0.04	0.25	-0.03	0.40	0.01	0.72	0.00	0.96
Total 16:0 PLPE	-0.04	0.23	-0.04	0.26	-0.08	**0.03**	-0.02	0.65	-0.02	0.52	-0.06	0.10
Total 18:1 PLPE	-0.06	0.10	-0.05	0.19	-0.05	0.15	-0.07	0.08	-0.05	0.22	0.01	0.89
Total 18:0 PLPE	-0.03	0.45	-0.03	0.43	-0.05	0.19	0.08	**0.03**	0.06	0.08	0.08	**0.04**
Total PLPE	-0.04	0.24	-0.04	0.25	-0.06	0.08						

E. Lysophosphotidylcholines

	ABSOLUTE VALUES						PROPORTIONS					
	HADS-D		CES-D		HADS-A		HADS-D		CES-D		HADS-A	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
LPC 15:0	-0.03	0.36	-0.05	0.20	-0.04	0.29	0.00	0.96	-0.04	0.32	-0.08	0.04
LPC 16:1	-0.01	0.74	0.02	0.67	0.01	0.82	-0.02	0.67	0.04	0.33	0.00	0.92
LPC 16:0	0.00	0.94	0.00	0.98	-0.01	0.89	0.04	0.27	0.06	0.13	-0.01	0.83
LPC 18:3	0.00	1.00	-0.01	0.78	0.01	0.89	0.00	0.92	-0.01	0.86	-0.02	0.61
LPC 18:2	-0.05	0.15	-0.06	0.12	0.00	0.92	-0.06	0.12	-0.06	0.12	-0.02	0.50
LPC 18:1	-0.01	0.86	-0.01	0.87	0.02	0.60	0.01	0.75	0.00	0.93	0.04	0.34
LPC 18:0	0.00	0.99	0.00	0.92	0.01	0.86	0.05	0.19	0.02	0.57	0.02	0.50
LPC 20:5	-0.05	0.17	-0.08	**0.04**	-0.03	0.48	-0.06	0.10	-0.08	**0.02**	-0.03	0.46
LPC 20:4	-0.05	0.14	-0.07	0.07	-0.04	0.29	-0.06	0.10	-0.06	0.08	-0.05	0.17
LPC 20:3	-0.04	0.26	-0.04	0.34	-0.02	0.65	-0.06	0.13	-0.04	0.26	-0.04	0.28
LPC 20:0	0.02	0.61	0.01	0.88	0.04	0.27	0.04	0.27	0.03	0.48	0.04	0.31
LPC 22:6	-0.05	0.18	-0.05	0.19	-0.03	0.48	-0.05	0.17	-0.04	0.29	-0.04	0.31
LPC 22:5	0.00	0.95	-0.02	0.60	0.01	0.77	-0.01	0.86	-0.01	0.70	0.00	0.98
LPC 22:4	-0.04	0.24	-0.05	0.19	0.00	0.98	-0.03	0.44	-0.03	0.49	0.00	0.95
LPC 22:0	0.01	0.73	0.00	0.93	0.02	0.61	0.01	0.82	-0.01	0.85	0.00	0.96
Saturated LPC	0.00	0.95	0.00	0.98	0.00	0.97	0.07	0.06	0.07	0.06	0.02	0.50
Mono-unsaturated LPC	-0.01	0.86	0.00	0.96	0.02	0.56	0.01	0.76	0.01	0.81	0.03	0.37
Poly-unsaturated LPC	-0.06	0.09	-0.07	0.07	-0.01	0.81	-0.08	**0.03**	-0.08	**0.03**	-0.04	0.31
Unsaturated LPC	-0.05	0.20	-0.05	0.18	0.00	0.98	-0.07	0.06	-0.07	0.06	-0.02	0.50
Total LPC	-0.01	0.71	-0.01	0.71	0.00	0.93						

F. Ceramides

	ABSOLUTE VALUES				PROPORTIONS							
	HADS-D		CES-D		HADS-A		HADS-D		CES-D		HADS-A	
	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value	r	P-value
CER 16:0	0.06	0.10	0.03	0.38	-0.01	0.81	0.03	0.46	0.01	0.83	0.01	0.89
CER 18:0*	0.06	0.09	0.08	**0.04**	-0.02	0.68	0.01	0.75	0.05	0.15	0.00	0.99
CER 20:0*	0.09	**0.02**	0.05	0.14	0.00	0.94	0.02	0.51	0.00	0.90	-0.02	0.68
CER 22:0	0.06	0.09	0.05	0.21	0.00	0.93	0.05	0.20	0.05	0.18	0.03	0.43
CER 23:0	0.01	0.77	0.00	0.99	-0.04	0.27	-0.05	0.15	-0.05	0.14	-0.07	0.05
CER 24:1	0.05	0.16	0.03	0.48	-0.03	0.49	0.02	0.67	0.00	0.98	-0.02	0.51
CER 24:0	0.02	0.57	0.01	0.72	0.00	0.99	-0.04	0.24	-0.03	0.40	0.02	0.59
Glu CER 16:0	0.04	0.23	0.04	0.34	0.00	0.97	0.01	0.86	0.01	0.84	0.00	0.95
Glu CER 24:1	0.02	0.62	0.02	0.64	-0.02	0.67	-0.01	0.70	0.00	0.98	0.00	0.95
Saturated CER	0.04	0.27	0.02	0.50	-0.01	0.86	-0.02	0.67	0.00	0.98	0.02	0.51
Unsaturated CER	0.05	0.16	0.03	0.48	-0.03	0.49	0.02	0.67	0.00	0.98	-0.02	0.51
Glu-CER	0.03	0.43	0.02	0.55	-0.01	0.76	-0.01	0.87	0.00	0.90	0.00	0.89
Total CER	0.05	0.21	0.03	0.44	-0.01	0.81						

Supplementary Table 1 shows the age and sex adjusted correlation between the plasma phosphor-sphingolipids and complaints of depression and anxiety. Correlation results with *P-values* less than 0.05 was shown in bold.

TABLE S2. Correlation between lipid-lipid ratios and depression/ anxiety scales

			HADS-D		CES-D		HADS-A	
Lipid 1	/	Lipid 2	r	P-value	r	P-value	r	P-value
SPM 23:1	/	SPM 16:0	**-0.16**	**1.04E-05**	-0.12	1.21E-03	-0.13	2.68E-04
PC O 36:4	/	CER 20:0	**-0.16**	**1.08E-05**	-0.12	1.60E-03	-0.10	8.84E-03
PC O 36:4	/	SPM 16:0	**-0.16**	**1.12E-05**	-0.09	9.68E-03	-0.13	2.38E-04
PC O 36:4	/	PE 40:4	**-0.15**	**3.60E-05**	-0.12	1.03E-03	-0.09	1.76E-02
PC O 36:4	/	CER 22:0	**-0.15**	**3.83E-05**	-0.11	2.01E-03	-0.11	3.21E-03
PC O 38:5	/	SPM 16:0	**-0.15**	**5.36E-05**	-0.11	3.60E-03	-0.14	2.04E-04
PC O 38:5	/	CER 20:0	**-0.15**	**6.06E-05**	-0.12	1.32E-03	-0.09	1.48E-02
PLPE 18:1/20:4	/	PE 40:4	-0.14	7.34E-05	-0.11	2.20E-03	-0.07	6.81E-02
PC O 36:4	/	CER 16:0	-0.14	8.00E-05	-0.10	5.77E-03	-0.10	8.39E-03
PC O 36:4	/	SPM 16:1	-0.14	8.71E-05	-0.12	7.14E-04	-0.10	6.76E-03
PC O 34:2	/	CER 16:0	-0.14	9.85E-05	-0.11	3.51E-03	-0.07	6.94E-02
PC 36:4	/	PE 40:4	-0.14	1.02E-04	-0.12	1.64E-03	-0.06	8.63E-02
PC O 34:2	/	CER 20:0	-0.14	1.07E-04	-0.10	4.88E-03	-0.06	1.05E-01
PC O 34:2	/	SPM 16:0	-0.14	1.20E-04	-0.10	5.82E-03	-0.09	1.18E-02
PC O 36:4	/	CER 24:1	-0.14	1.35E-04	-0.10	4.22E-03	-0.10	7.38E-03
PC O 36:4	/	PE 38:2	-0.14	1.41E-04	-0.11	2.21E-03	-0.13	5.81E-04
PC O 38:5	/	SPM 24:1	-0.14	1.44E-04	-0.12	6.78E-04	-0.13	2.80E-04
PC O 36:4	/	CER 18:0	-0.14	1.44E-04	-0.13	2.92E-04	-0.09	1.93E-02
PC O 36:4	/	SPM 24:1	-0.14	1.73E-04	-0.10	8.47E-03	-0.12	1.28E-03
PC O 36:4	/	PE 40:5	-0.14	1.75E-04	-0.12	1.08E-03	-0.10	7.95E-03
PC O 36:5	/	SPM 16:0	-0.14	1.84E-04	-0.09	1.16E-02	-0.11	3.98E-03
PC O 36:4	/	Glu CER 16:0	-0.13	2.18E-04	-0.10	4.44E-03	-0.11	2.03E-03
PC O 36:5	/	PE 40:4	-0.13	2.25E-04	-0.11	1.90E-03	-0.07	6.35E-02
PC 34:4	/	PE 40:4	-0.13	2.26E-04	-0.11	1.99E-03	-0.09	1.88E-02
PLPE 16:0/20:4	/	PE 40:4	-0.13	2.34E-04	-0.11	2.21E-03	-0.08	2.91E-02
PLPE 18:1/20:4	/	PE 42:6	-0.13	2.41E-04	-0.07	6.12E-02	-0.06	7.89E-02
PC O 40:6	/	CER 20:0	-0.13	2.42E-04	-0.08	2.84E-02	-0.06	9.68E-02
PLPE 16:0/22:6	/	PE 38:2	-0.13	2.56E-04	-0.10	8.41E-03	-0.13	3.35E-04
PC O 36:4	/	PE 42:6	-0.13	2.69E-04	-0.07	4.91E-02	-0.08	2.65E-02
PLPE 16:0/22:6	/	PE 40:4	-0.13	2.88E-04	-0.10	5.96E-03	-0.07	4.47E-02
PC O 38:5	/	CER 22:0	-0.13	2.97E-04	-0.11	1.96E-03	-0.10	6.43E-03
PC 36:4	/	PE 40:5	-0.13	3.02E-04	-0.11	1.74E-03	-0.08	2.95E-02
SPM 23:1	/	CER 20:0	-0.13	3.18E-04	-0.10	6.85E-03	-0.07	6.28E-02
PC O 34:2	/	CER 22:0	-0.13	3.38E-04	-0.11	4.10E-03	-0.07	5.62E-02
PC O 38:4	/	C20:0	-0.13	3.68E-04	-0.11	4.04E-03	-0.08	3.51E-02
PC O 36:4	/	PC 36:1	-0.13	3.76E-04	-0.10	7.22E-03	-0.13	3.13E-04
PC O 36:5	/	PE 40:5	-0.13	3.73E-04	-0.12	1.61E-03	-0.08	2.55E-02
PC O 36:5	/	SPM 24:1	-0.13	3.77E-04	-0.10	9.10E-03	-0.10	8.48E-03
PC 36:5	/	PE 40:4	-0.13	3.78E-04	-0.11	1.76E-03	-0.05	1.88E-01
PC O 36:3	/	CER 20:0	-0.13	3.83E-04	-0.08	3.08E-02	-0.02	5.65E-01
PLPE 18:1/22:6	/	PE 42:6	-0.13	3.93E-04	-0.05	1.39E-01	-0.05	1.50E-01

TABLE S2. Correlation between lipid-lipid ratios and depression/ anxiety scales (continued)

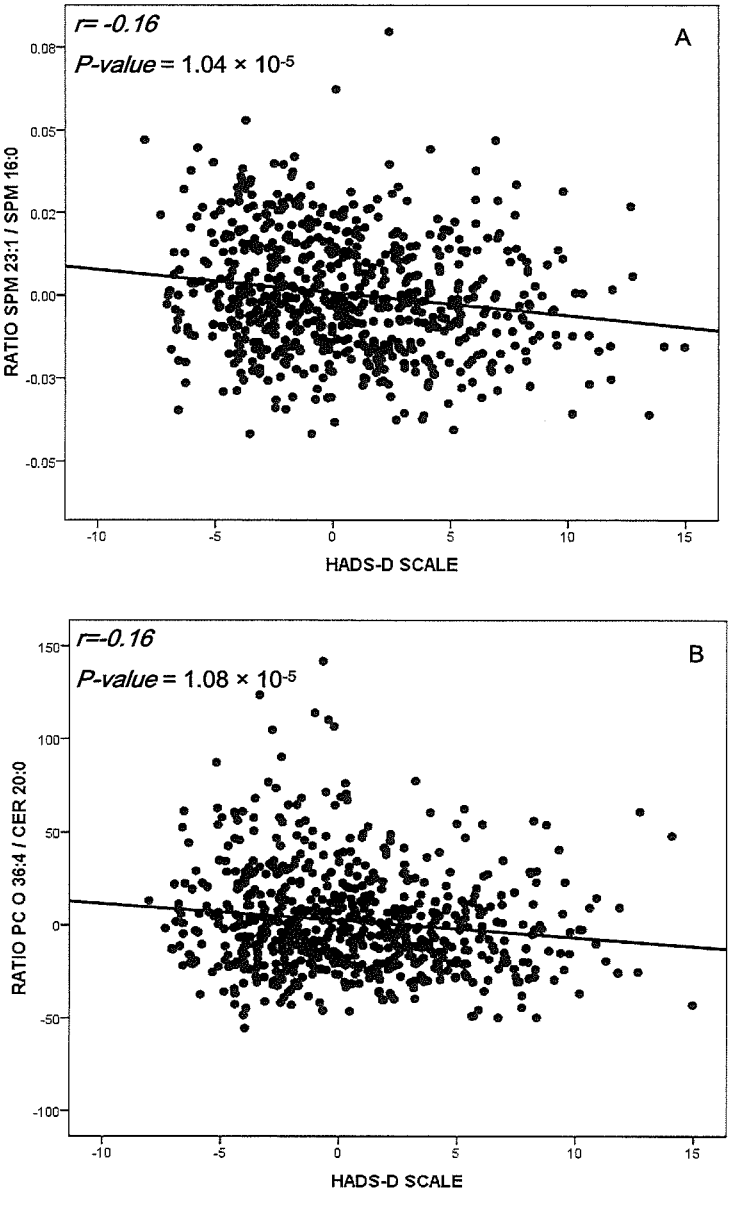
			HADS-D		CES-D		HADS-A	
SPM 23:1	/	SPM 24:1	-0.13	4.01E-04	-0.10	8.84E-03	-0.12	1.28E-03
PLPE 16:0/22:6	/	PE 42:6	-0.13	4.25E-04	-0.06	9.27E-02	-0.08	2.89E-02
PC O 36:5	/	CER 22:0	-0.13	4.27E-04	-0.11	3.65E-03	-0.08	2.92E-02
PC O 34:2	/	CER 24:1	-0.13	4.32E-04	-0.09	1.35E-02	-0.06	1.15E-01
PE O 38:7	/	PE 42:6	-0.13	4.68E-04	-0.07	7.10E-02	-0.06	1.33E-01
PC O 38:5	/	CER 16:0	-0.13	4.74E-04	-0.10	7.11E-03	-0.09	1.42E-02
PC 38:4	/	PE 40:4	-0.13	4.95E-04	-0.10	6.93E-03	-0.06	1.00E-01
PC O 36:5	/	CER 16:0	-0.13	5.14E-04	-0.10	9.35E-03	-0.07	4.28E-02
PLPE 18:1/20:4	/	PE 40:5	-0.13	5.21E-04	-0.10	4.23E-03	-0.08	3.33E-02
PC O 34:2	/	PE 42:6	-0.13	5.22E-04	-0.08	3.58E-02	-0.06	1.22E-01
PC O 36:4	/	PC 36:2	-0.13	5.33E-04	-0.11	3.35E-03	**-0.15**	**2.14E-05**
PC O 38:5	/	PE 40:4	-0.13	5.63E-04	-0.11	2.89E-03	-0.07	6.07E-02
PC O 36:3	/	SPM 16:0	-0.13	5.64E-04	-0.06	1.09E-01	-0.04	2.24E-01
PLPE 16:0/22:6	/	CER 20:0	-0.13	5.66E-04	-0.08	2.67E-02	-0.09	1.78E-02
PLPE 16:0/20:5	/	PE 42:6	-0.13	5.67E-04	-0.06	1.02E-01	-0.05	1.46E-01
PC O 38:5	/	CER 24:1	-0.13	5.86E-04	-0.11	2.90E-03	-0.09	1.32E-02
PC O 40:6	/	SPM 24:1	-0.13	5.91E-04	-0.06	1.29E-01	-0.10	5.10E-03
PC O 36:5	/	PE 38:2	-0.13	5.91E-04	-0.11	3.84E-03	-0.10	5.13E-03
PC O 36:3	/	CER 16:0	-0.13	5.91E-04	-0.08	4.04E-02	-0.03	3.73E-01
PC O 38:5	/	PE 38:2	-0.13	6.10E-04	-0.11	1.88E-03	-0.12	1.46E-03
PC O 36:5	/	CER 20:0	-0.12	6.23E-04	-0.10	6.92E-03	-0.07	6.05E-02
PC 38:5	/	PE 40:4	-0.12	6.39E-04	-0.09	1.22E-02	-0.05	1.70E-01
SPM 23:1	/	CER 22:0	-0.12	6.37E-04	-0.11	3.93E-03	-0.07	4.76E-02
SPM 15:0	/	SPM 16:0	-0.12	6.59E-04	-0.08	2.84E-02	-0.09	1.06E-02
PLPE 18:1/20:4	/	PE 38:2	-0.12	6.98E-04	-0.10	8.58E-03	-0.09	1.00E-02
PC 34:4	/	PE 40:5	-0.12	7.18E-04	-0.11	2.72E-03	-0.10	5.69E-03
SPM 23:1	/	SPM 16:1	-0.12	7.33E-04	-0.11	2.18E-03	-0.09	1.67E-02
PC O 38:4	/	PE 40:4	-0.12	7.32E-04	-0.11	2.94E-03	-0.07	6.79E-02
PC O 36:5	/	PE 42:6	-0.12	7.34E-04	-0.07	6.67E-02	-0.07	5.32E-02
PC O 38:5	/	Glu CER 16:0	-0.12	7.37E-04	-0.11	2.38E-03	-0.11	3.36E-03
PC 38:6	/	PE 40:4	-0.12	7.44E-04	-0.09	2.04E-02	-0.05	1.76E-01
PC O 40:6	/	CER 22:0	-0.12	7.38E-04	-0.07	4.21E-02	-0.07	6.42E-02
PC O 36:4	/	PE 34:2	-0.12	7.43E-04	-0.11	1.79E-03	-0.11	3.65E-03
PC O 36:5	/	SPM 16:1	-0.12	7.50E-04	-0.12	1.64E-03	-0.07	5.83E-02
PLPE 18:1/22:6	/	PE 38:2	-0.12	7.66E-04	-0.08	4.03E-02	-0.08	2.28E-02
PC O 38:5	/	CER 18:0	-0.12	7.80E-04	-0.14	1.55E-04	-0.08	2.60E-02
PC O 36:5	/	CER 24:1	-0.12	8.00E-04	-0.10	5.52E-03	-0.07	4.92E-02
PLPE 16:0/22:6	/	PE 40:5	-0.12	8.32E-04	-0.10	5.11E-03	-0.09	1.45E-02
PC O 34:2	/	PE 40:4	-0.12	8.53E-04	-0.10	4.21E-03	-0.05	1.83E-01
PC O 40:6	/	CER 18:0	-0.12	8.59E-04	-0.10	5.52E-03	-0.06	9.04E-02
PC O 34:2	/	Glu CER 16:0	-0.12	8.69E-04	-0.11	4.01E-03	-0.07	4.36E-02
PC O 40:6	/	CER 16:0	-0.12	8.72E-04	-0.06	1.01E-01	-0.05	1.37E-01

TABLE S2. Correlation between lipid-lipid ratios and depression/ anxiety scales (continued)

			HADS-D		CES-D		HADS-A	
PC O 36:4	/	PE 36:2	-0.12	8.84E-04	-0.12	1.11E-03	-0.11	1.83E-03
PC O 36:4	/	PE 38:4	-0.12	8.86E-04	-0.12	1.06E-03	-0.11	3.67E-03
PC O 34:2	/	PE 38:2	-0.12	8.99E-04	-0.11	2.03E-03	-0.08	2.46E-02
LPC 20:3	/	PE 40:4	-0.12	9.32E-04	-0.10	5.33E-03	-0.04	2.47E-01
PLPE 16:0/22:6	/	PE 36:2	-0.12	9.39E-04	-0.12	1.27E-03	-0.11	2.60E-03
PLPE 18:1/20:4	/	CER 20:0	-0.12	9.43E-04	-0.08	3.21E-02	-0.06	1.29E-01
SPM 15:0	/	CER 22:0	-0.12	9.56E-04	-0.11	2.06E-03	-0.08	3.10E-02
PC 36:5	/	PE 40:5	-0.12	9.57E-04	-0.11	3.53E-03	-0.06	1.20E-01
PC 32:2	/	PE 40:4	-0.12	9.66E-04	-0.09	1.87E-02	-0.05	2.00E-01
SPM 23:1	/	Glu CER 16:0	-0.12	9.60E-04	-0.12	1.04E-03	-0.09	1.42E-02
PC O 38:5	/	PE 42:6	-0.12	9.61E-04	-0.08	3.63E-02	-0.08	2.90E-02
PC O 36:3	/	CER 22:0	-0.12	9.68E-04	-0.08	2.67E-02	-0.03	3.83E-01
PC O 36:4	/	SPM 18:0	-0.12	9.79E-04	-0.09	1.16E-02	-0.07	5.52E-02
PLPE 18:1/20:4	/	PLPE 18:1/18:1	-0.12	9.72E-04	-0.08	2.08E-02	-0.09	1.07E-02
PC O 40:6	/	SPM 16:0	-0.12	9.82E-04	-0.05	2.06E-01	-0.09	1.34E-02

Age and sex adjusted correlation analysis between depression scales and lipid-lipid ratios. Results shown only for the ratios that have *P-values* less than 0.001 for the HADS-D scale. Lipid ratios with significant *P-values* are shown in bold. Only results with *P-value* <0.001 are shown.

FIGURE S1 .Correlation plots of lipid / lipid ratios and HADS-D.



Residual x-y plots are drawn after regression out the age and sex affects for lipid / lipid ratio and depression scale.

ACKNOWLEDGEMENTS

Finally I have the chance to thank to the special people who entered my life during these years. To the people who showed me the opportunities, shared ideas, thought and worked together with me and made this thesis possible, also to the large family of scientists, international collaborations that contributed their valuable data or effort and to the colleagues I worked together with. Eventually I sincerely want to thank to my friends and family for making life fun and the difficult days bearable.

On the very top of my list there is one amazing person in whom intelligence, experience, good will and unique personality combine. Dear Cornelia van Duijn, apart from being my smart supervisor, nice boss and tolerant teacher during the last five years, for me the most valuable was perhaps the never-ending positive energy you delivered. Your curiosity and your unlimited enthusiasm for the work we have been doing provided me encourage and confidence which I needed at the times of frustrations. You were sitting together with me for unlimited hours for dissecting tables, gene networks, thinking and writing and you rendered this thesis possible. These things I can never forget, hartelijk bedankt ! On the other side there was Ben Oostra, always knowing the right way to do things guiding me not only in the scientific approach but also through bureaucracy and rules, even after he retired. Dear Ben, I feel myself lucky that I had the chance to work with such an experienced scientist like you, thanks for sharing your valuable knowledge. And the third person, this perfectionist American: Aaron Isaacs. Dear Aaron: Thank you so much for reading my papers again and again, patiently, pointing out the little details carefully. Not only I have learned a lot of theory and practical tips from you on data analysis, but also you were a good example for planning and work attitude. My thesis committee: Dear Dorret Boomsma, I feel lucky that I had chance to work with you, I appreciated your constructive comments on our manuscripts and thanks for being in my thesis committee. Dear Henning Tiemeier, thank you for critically reading the several papers through the eye of a psychiatrist and emphasizing the clinical value of our research, also for accepting to be in my thesis reading committee. Dear Ko Willems van Dijk, thanks for the interest in our papers and your comments and accepting to be in the committee. I hope we will work together in exciting projects. Dear Thomas Hankemeier and Eric Sijbrands: thanks for being in my thesis committee.

Dear Cecile Janssens, we worked together on several papers, thank you for teaching me a lot on statistics and critically reading the manuscripts, you made the difficult questions easy and you cheered up the long meetings with your talent of humour. I also would like to thank the head of our successful department of epidemiology, to dear Albert Hofman, for the opportunities he provided.

A very large part this thesis, are based on analysis conducted within the Erasmus Rucphen Family data. Thanks to the general practitioners, geneologists who worked for the Erasmus Rucphen Family Study and the most importantly we would like to thank to the living descendants

of this large family who openly accepted to share their personal information for this unique study.

I also would like to thank to the network of scientists behind the large scale projects presented in this thesis. From the lipidomics studies in EUROSPAN: dear Andrew Hicks, Jim Wilson, Caroline Hayward, Ulf Gyllensten, Thomas Meitinger, Igor Rudan, Gerhard Liebisch, Gerd Schmitz, Harry Campbell and Asa Johansson; thanks for the data sharing and critics on the manuscripts. From DIAGRAM and CARDIOGRAM: Thanks to Mark McCarthy and Nilesh Samani for their interest in our lipidomics -GWAS and the contribution from their consortia. Samuli Ripatti from ENGAGE, thanks for the collaboration on the lipid paper. From QIMR, the experts on quantitative genetics: Peter Visscher and Naomi Wray, thanks for the advise and comments on several papers.

From Vrije Universiteit, Amsterdam: Dear Christel Middeldorp, we wrote a paper under your supervision, thanks for the advice and nice ideas. Thanks to Brenda Penninx for the help on NESDA data and Eco de Geus and Jouke-Jan Hottenga for the comments on manuscripts and finally Abdel Abdellaoui for the very quick look up he performed.

Our collaborators from Institute of Cytology and Genetics, Novosibirsk: Dear Tatiana Axenovich and Irina Zorkoltseva, thank you for the big support in computation and data analysis. Dear Yurii Aulchenko, I think you are a great teacher and expert and I learned a lot from you, thanks for all the tips and creative ideas, also for the nice dinners with Marianna Bevova.

Thanks to all reserchers from the CHARGE depressive symptoms working group: Thanks to, Hans Grabe, Joanne Murabito, Kathryn Lunetta and many others that I had chance to come across in the meetings and TCs. Thanks to Mike Nalls, for the extra analysis he performed and the help on brain expression data. Dear Karin (Hek), you have been the nicest fellow traveller on this way with your quick and clever responses to hundreds of not-so-depressing e-mails about depression we have been sending to each other, thank you so much and congratulations for the first fruit of the huge effort !. Dear Jari (Lahti), I enjoyed working with you for "happiness" and favoured your optimisim and factuality, but also as a friend, in Athens and Rotterdam, speak to you in the coming TCs! Dear Neşe (Direk), you have also been a loyal company of chat and analysis but you are also great psychiatrist, you should not drift apart from psychosis, patients miss you already!

From Max Planck institute for Psychiatry: Dear Bertram Müller-Myhsok thank you for the collaboration and nice lectures. Thanks to dear Martin Kohli and Suzanne Lucae for the (solute carrier) collaboration and thanks to Nazenin Karbalai for being the emergency replacement lecturer in GE03.From University Medical Center Groningen: Lude Frank and Harm-Jan Westra thanks for the advices on gene networks. Thanks to Tuomas Kilpeläinen, Ruth Loos and Adrienne Cupples, from GIANT consortium, Anna Kottgen from GUGC, thanks for the collaboration. From Leiden University Medical Center: Thanks to Gert-Jan van Ommen for the nice meetings he organized. Thanks to Harish Dharuri for his collaboration.

From Erasmus Medical Center: Vincent Jaddoe and Rob Taal from Generation R thanks for the collaboration on blood pressure. From the Biomics Department, Wilfred van IJcken, Rutger Brouwer and Mirjam van den Hout, thanks for the collaboration on the exome sequence data. From Epidemiology, Internal Medicine and Forensic Departments: Dear Karol Estrada, Abbas Deghan, Andre Uitterlinden, Fernando Rivedeneria, Germaine Verwoert, Frank van Rooij, Wishal Ramdas, Fan Liu and Manfred Kayser thank you for your collaboration.

Also special thanks to people who simplified a bunch of paperwork which was quite difficult for me: Hetty Gerritse, Yolanda Bekker, Solange Degenaar, Erica Kroos, Jeannette Lokkers and Astrid Vrakkings. Thank you Nano Suwarno, for writing your password on the screen every now and then, you promised me a new computer after I graduate!

Dear colleagues from genetic epidemiology, I am happy that we will be working together for some more time: dear Najaf, I can not forget your help on statistics in difficult times, I know that I can always ask for your advice, you are trustful friend and I assure that your cat can stay in my hostel whenever he wants. Dear Peter (Hennemann), was fun to share the same room with you and the data as well, good luck with your new job and I wish you all the best. My dear paranimphen: Linda and Maarten, thank you for translating the summary of this thesis into Dutch. Special thanks to Maarten who recovered back some hundred gigabytes of chapter 4.2 which I just trashed during the heat of making some space on the server. Dear Linda, we have been room buddies for a long time already, you have been the open-hearted listener of my very random complaints and improvised chat, was funny driving in your car to the zoo, to the meetings here and there and listening the discussions in tele-conferences as long as the telomeres themselves. I think we will do some more nice jobs together. Dear Lennart, you have been a great help and advisor on computing but more important you were understanding and relaxed, but also fun. Dear golden girls; Sara, Carla and Elisa, and the mermaids from overseas; Adriana and Claudia, dear Natalia of Rome and Revanius of Cameroon, also the genetics lab of epidemiology, dear Jeanette, Benardette, Andy, Andrea: I enjoyed working with you and our little chats. Dear Elza, much more than a secretary, you became my private doctor and a good friend, thanks for all the warm interest. Dear Jeanette and Elza, you solved the occupational back injury problem by phoning around and finding a proper furniture for me, so that I could keep on sitting behind computer and working. And finally my dear old friend Maksim: I am so happy that I know you and I deeply thank for everything, I wish the best for you and hope to see you soon.

I think it is also the right moment to mention some important people that contributed to my education before the Rotterdam. Dear Uğur Özbek, head of Institute for Experimental and Medical Research of Istanbul University: Dear Uğur Bey, thanks for all the opportunities you provided in several projects and especially for encouraging me to come to Rotterdam, to have classes in genetic epidemiology. Staying in Rotterdam was not the initial plan, but within those circumstances, I am still happy with the decision I made. Here I would also like to name some inspirational people for me who especially encouraged me in the way that I lead: Beyhan Tüysüz

from Cerrahpasa Medical Center, Child Genetics Department and dear Sevim Büyükdevrim from Istanbul Medical Faculty and his wife Işım Hanım, Osman Ünsal and Nebiha Kutaygil from Biologic Oceaonography and Hydrobiology Departments of Istanbul University.

Nihayet, DETAM genetik yıllari boyunca beraber calıstıgım arkadaslar: Elif, Özden, Müge, Mehves, Fatma ve digerleri; sizinle calısmak guzeldi. Gögüs Hastaliklarindan doktor Fatih (Yakar) ve Onkolojiden Pınar Saip Hoca ve Zeki Ustüner Hoca, biyokimyadan Pervin Vural Hanım, tesekkürler. Metabolik Sendrom projesinden sevgili Selcuk Görmez, tesekkurler. SEREX projelerinden sayin Ali Güre ve Ahmet Gül: ortak calismalarimiz ve sağladiginiz imkanlar icin tesekkurler. Sevgili Burcak, benim cok sevgili ve özel arkadasım, aynı zamanda hocam, umarım hersey gonlunce olur, yıllar boyu bitmeyen sıcak ilgin icin tesekkurler. Sevgili Naci, senden cok sey öğrendim, hersey gönlünce olsun. Sevgili Suzin, herzaman yanımda oldun, hersey icin cok tesekkürler, ayni zamanda sevgili Aris, özet kısmının Türkceye cevirisi ama en cok da leziz free-shop alisverislerin icin: tesekkürler. En kısa zamanda görüsmek uzere.

As the list never ends, mijn lieve huisgenoten from Poortgebouw: Robin, Anna, Yaman, Tom, Linda, Teun, Mik, Luuk and many many others: My computer, my manuscript and me have been the furniture of your living space for a while, you had to be quiet sometimes and finish your party early because I needed sleep to go back to my studies, sometimes I was too busy to cook and you cooked for me. All are much appreciated! I enjoyed the festivals we have been to, the music, movies, singing and cooking together. Yaman: Thanks for introducing long boarding to my life. Jan: Thanks for scanning the half of this thesis for typos. Tom: thanks for the updates on neuroscience articles, for sharing unique stuff and dinners as well as for the cover-art.

Last but not the least, my family: without questioning they always supported what I wanted to do and how I wanted to do and let me be of my own. And they never complained when I boiled already dead frogs (to clean the bones) in the kitchen pans, or when I kindly hosted scorpions, snakes and snails at home long before my biology training, during the early ages of curiosity. They supported me during the long years of education. Annecim, babacim: iyi ki varsiniz, sizleri cok seviyorum.

PHD PORTFOLIO SUMMARY

Name: Ayşe Demirkan
Erasmus MC Department: Epidemiology
Research School: NIHES

PhD period: 2008-2012
Promotor(s): Prof. C.M. van Duijn & Prof. B.A. Oostra. Co-promoter: Assoc. Prof. Aaron Isaacs

1. PHD TRAINING

	Year	Workload (hours/ECTS)
Courses		
Erasmus Summer Programme	2007	4.3
Erasmus Winter Programme	2008	4.6
CCO1 Study Design	2008	4.3
CCO2 Biostatistical Methods 1	2008	5.7
EP03 Biostatistical Methods 2	2008	4.3
GE02 Genetic Epidemiologic Research Methods	2008	5.1
GE08 SNPs and Human disease	2008	1.4
GE03 Advances in population based Studies of Complex Genetic disorders	2008	1.4
GE05 Genetic linkage analysis	2008	1.4
SC09 Linux for scientists	2011	1.4
R-course	2010	1
Presentations		
Genetic Epidemiology Unit of Department of Epidemiology: Circulating lipids and depressive symptoms	2008	1
Genetic Epidemiology Unit of Department of Epidemiology: Polygenic inheritance in depression	2009	1
6th CMSB symposium: Common genetic variants in depression and anxiety	2009	1
Genetic Epidemiology Unit of Department of Epidemiology: Next generation sequencing	2010	1
Genetic Epidemiology Unit of Department of Epidemiology: GWAS of the EUROSPAN Consortium on circulating phospholipids and sphingolipids	2010	1
Genetic Epidemiology Unit of Department of Epidemiology: Linkage analysis of depressive symptoms.	2011	1

Genetic Epidemiology Unit of Department of Epidemiology: An update of lipidomics projects : Linkage analysis of circulating phospholipids and sphingolipids	2012	1

Conferences and meetings

6th annual CMSB meeting, 2 October, Rotterdam	2009	1
17th MGC PhD student workshop, Brugge	2009	1
CHARGE Consortium meeting, Rotterdam	2009	1
ESF Meeting on Next Generation Sequencing August 29-September 1	2010	1
7th annual CMSB meeting, 15 October, Leiden	2010	1
World congress on Psychiatric Genetics, Athens	2010	1
MGC course "Technology Facilities", 9-11 February	2011	1
European Society of Human Genetics Meeting, 28-31 May, Amsterdam	2011	1
8th annual CMSB meeting, 4 October, Amsterdam	2011	1
PCDI PostDoc Retreat, 28 - 30 March, Kapellerput	2012	1
European Society of Human Genetics Meeting, 23-26 June, Nuremberg	2012	1
CHARGE Consortium meeting, Rejkyavik	2012	1
Spring NMC Metabolomics Meeting, May 7	2012	1

Seminars and workshops

Weekly scientific seminars of Genetic Epidemiology Unit of the Department of Epidemiology	2007-2012	1
Molecular Medicine Postgraduate School Workshop on "Browsing genes and genomes with UCSC Browser", 30 October	2007	1
Molecular Medicine Postgraduate School Workshop on "Browsing genes and genomes with Ensembl", 17-18 September	2008	1
Molecular Medicine Postgraduate School Workshop on "Bioinformatic Analysis, Tools and Services (BATS)", 28 February	2008	1

2. TEACHING ACTIVITIES

Teaching assistant for the GE02 course	2009-2011	1
Coordination of the GE03 course	2010-2011	1

3. OTHER

Reviewer	PLOS ONE	1
Reviewer	Journal of Psychosomatic Medicine	1
Reviewer	Journal of Psychopharmacology	1

PUBLICATIONS AND CO-AUTHORSHIPS

Anna Köttgen, Eva Albrecht, Alexander Teumer, Veronique Vitart, Jan Krumsiek, Claudia Hundertmark, Giorgio Pistis, Daniela Ruggiero, Conall M O'Seaghdha, Toomas Haller, Qiong Yang, Toshiko Tanaka, Andrew D Johnson, Zoltán Kutalik, Albert V Smith, Julia Shi, Maksim Struchalin, Rita PS Middelberg, Morris J Brown, Angelo L Gaffo, Nicola Pirastu, Guo Li, Caroline Hayward, Tatijana Zemunik, Jennifer Huffman, Loic Yengo, Jing Hua Zhao, **Ayse Demirkan**, Mary F Feitosa, Xuan Liu, Giovanni Malerba, Lorna M Lopez, Pim van der Harst, Xinzhong Li, Marcus E Kleber, Andrew A Hicks, Ilja M Nolte, Asa Johansson, Federico Murgia, Sarah H Wild, Stephan JL Bakker, John F Peden, Abbas Dehghan, Maristella Steri, Albert Tenesa, Vasiliki Lagou, Perttu Salo, Massimo Mangino, Lynda M Rose, Terho Lehtimäki, Owen M Woodward, Yukinori Okada, Adrienne Tin, Christian Muller, Christopher Oldmeadow, Margus Putku, Darina Czamara, Peter Kraft, Laura Frogheri, Gian Andri Thun, Anne Grotevendt, Gauti Kjartan Gislason, Tamara B Harris, Lenore J Launer, Patrick McArdle, Alan R Shuldiner, Eric Boerwinkle, Josef Coresh, Helena Schmidt, Michael Schallert, Nicholas G Martin, Grant W Montgomery, Michiaki Kubo, Yusuke Nakamura, Toshihiro Tanaka, Patricia B Munroe, Nilesh J Samani, David R Jacobs Jr, Kiang Liu, Pio D'Adamo, Sheila Ulivi, Jerome I Rotter, Bruce M Psaty, Peter Vollenweider, Gerard Waeber, Susan Campbell, Olivier Devuyst, Pau Navaro, Ivana Kolcic, Nicholas Hastie, Beverley Balkau, Philippe Froguel, Tõnu Esko, Andres Salumets, Kay Tee Khaw, Claudia Langenberg, Nicholas J Wareham, Aaron Isaacs, Aldi Kraja, Qunyuan Zhang, Philipp S Wild, Rodney J Scott, Elizabeth G Holliday, Elin Org, Margus Viigimaa, Stefania Bandinelli, Jeffrey E Metter, Antonio Lupo, Elisabetta Trabetti, Rossella Sorice, Angela Döring, Eva Lattka, Konstantin Strauch, Fabian Theis, Melanie Waldenberger, H-Erich Wichmann, Gail Davies, Alan J Gow, Marcel Bruinenberg, LifeLines Cohort Study, Ronald P Stolk, Jaspal S Kooner, Weihua Zhang, Bernhard R Winkelmann, Bernhard O Boehm, Susanne Lucae, BrendaPenninx, Johannes H Smit, Gary Curhan, Poorva Mudgal, Robert M Plenge, Laura Portas, Ivana Persico, Mirna Kirin, James F Wilson, Irene Mateo Leach, Wiek H van Gilst, Anuj Goel, Halit Ongen, Albert Hofman, Fernando Rivadeneira, Andre G Uitterlinden, Medea Imboden, Arnold von Eckardstein, Francesco Cucca, Ramaiah Nagaraja, Maria Grazia Piras, Matthias Nauck, Claudia Schurmann, Kathrin Budde, Florian Ernst, Susan M Farrington, Evropi Theodoratou, Inga Prokopenko, Michael Stumvoll, Antti Jula, Markus Perola, Veikko Salomaa, So-Youn Shin, Tim D Spector, Cinzia Sala, Paul M Ridker, Mika Kähönen, Jorma Viikari, Christian Hengstenberg, Christopher P Nelson, the CARDIoGRAM Consortium, the DIAGRAM Consortium, the ICBP Consortium, the MAGIC Consortium, James F Meschia, Michael A Nalls, Pankaj Sharma, Andrew B Singleton, Naoyuki Kamatani, Tanja Zeller, Michel Burnier, John Attia, Maris Laan, Norman Klopp, Hans L Hillege, Stefan Kloiber, Hyon Choi, Mario Pirastu, Silvia Tore, Nicole M Probst-Hensch, Henry Völzke, Vilmundur Gudnason, Afshin Parsa, Reinhold Schmidt, John B Whitfield, Myriam Fornage, Paolo Gasparini, David S Siscovick, Ozren Polasek, Harry Campbell, Igor Rudan, Nabila Bouatia-Naji, Andres Metspalu, Ruth JF Loos, Cornelia M van Duijn, Ingrid B Borecki, Luigi Ferrucci, Giovanni Gambaro, Ian J Deary, Bruce HR Wolffenbuttel,

John C Chambers, Winfried März, Peter P Pramstaller, Harold Snieder, Ulf Gyllensten, Alan F Wright, Gerjan Navis, Hugh Watkins on behalf of PROCARDIS, Jacqueline CM Witteman, Serena Sanna, Sabine Schipf, Malcolm G Dunlop, Anke Tönjes, Samuli Ripatti, Nicole Soranzo, Daniela, Toniolo, Daniel I Chasman, Olli Raitakari, WH Linda Kao, Marina Ciullo, Caroline S Fox, Mark Caulfield, Murielle Bochud, Christian Gieger. *Multiple novel loci highlighting metabolic control of urate production and Excretion are associated with Gout. (Accepted by Nature Genetics)* Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Mägi R, Waite L, Smith AV, Yerges-Armstrong LM, Monda KL, Hadley D, Mahajan A, Li G, Kapur K, Vitart V, Huffman JE, Wang SE, Palmer C, Esko T, Fischer K, Zhao JH, **Demirkan A**, Isaacs A, Feitosa MF, Luan J, Heard-Costa NL, White C, Jackson AU, Preuss M, Ziegler A, Eriksson J, Kutalik Z, Frau F, Nolte IM, Van Vliet-Ostaptchouk JV, Hottenga JJ, Jacobs KB, Verweij N, Goel A, Medina-Gomez A, Estrada K, Bragg-Gresham JL, Sanna S, Sidore C, Tyrer J, Teumer A, Prokopenko I, Mangino M, Lindgren CM, Assimes TL, Shuldiner AR, Hui J, Beilby JP, McArdle WL, Hall P, Haritunians T, Zgaga L, Kolcic I, Polasek O, Zemunik T, Oostra BA, Junttila MJ, Grönberg H, Schreiber S, Peters A, Hicks AA, Stephens J, Foad NS, Laitinen J, Pouta A, Kaakinen M, Willemsen G, Vink JM, Wild SH, Navis G, Asselbergs FW, Homuth G, John U, Iribarren C, Harris T, Launer L, Gudnason V, O'Connell JR, Boerwinkle E, Cadby G, Palmer LJ, James AL, Musk AW, Ingelsson E, Psaty BM, Beckmann JS, Waeber G, Vollenweider P, Hayward C, Wright AF, Rudan I, Groop LC, Metspalu A, Khaw KT, van Duijn CM, Borecki IB, Province MA, Wareham NJ, Tardif JC, Huikuri HV, Cupples AL, Atwood LD, Fox CS, Boehnke M, Collins FS, Mohlke KL, Erdmann J, Schunkert H, Hengstenberg C, Stark K, Lorentzon M, Ohlsson C, Cusi D, Staessen JA, Van der Klauw MM, Pramstaller PP, Kathiresan S, Jolley JD, Ripatti S, Jarvelin MR, de Geus EJC, Boomsma DI, Penninx B, Wilson JF, Campbell H, Chanock SJ, van der Harst P, Hamsten A, Watkins H, Hofman A, Witteman JC, Zillikens MC, Uitterlinden AG, Rivadeneira F, Kiemeney LA, Vermeulen SA, Abecasis GR, Schlessinger D, Schipf S, Stumvoll M, Tönjes A, Spector TD, North KE, Lettre G, McCarthy MI, Berndt SI, Heath AC, Madden PAF, Nyholt DR, Montgomery GW, Martin NG, McKnight B, Strachan DP, Hill WG, Snieder H, Ridker PM, Thorsteinsdottir U, Stefansson K, Frayling TM, Hirschhorn JN, Goddard ME & Visscher PM. *FTO genotype is associated with phenotypic variability of body mass index. Nature 16 September 2012,DOI: 10.1038/nature11401. [Epub ahead of print]*

Chasman DI, Fuchsberger C, Pattaro C, Teumer A, Böger CA, Endlich K, Olden M, Chen MH, Tin A, Taliun D, Li M, Gao X, Gorski M, Yang Q, Hundertmark C, Foster MC, O'Seaghdha CM, Glazer N, Isaacs A, Liu CT, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Johnson AD, Gierman HJ, Feitosa MF, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson A, Tönjes A, Dehghan A, Lambert JC, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Coassin S, Pistis G; CARDIoGRAM Consortium; ICBP Consortium; the CARe Consortium; WTCCC2, Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Cavalieri M, Rao M, Hu F, **Demirkan A**, Oostra BA, de Andrade M, Turner ST, Ding J, Andrews JS, Freedman BI, Giulianini F, Koenig W, Illig T, Meisinger

C, Gieger C, Zgaga L, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Stengel B, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Ketkar S, Colhoun H, Doney A, Robino A, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, Thun GA, Paulweber B, Haun M, Sala C, Mitchell P, Ciullo M, Kim SK, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Siscovick DS, van Duijn CM, Borecki IB, Kardia SL, Liu Y, Curhan GC, Rudan I, Gyllensten U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman J, Hayward C, Ridker PM, Parsa A, Bochud M, Heid IM, Kao WL, Fox CS, Köttgen A.

Integration of Genome-Wide Association Studies with Biological Knowledge Identifies Six Novel Genes Related to Kidney Function. Hum Mol Genet. 2012 Sep 8. [Epub ahead of print]

Pattaro C, Köttgen A, Teumer A, Garnaas M, Böger CA, Fuchsberger C, Olden M, Chen MH, Tin A, Taliun D, Li M, Gao X, Gorski M, Yang Q, Hundertmark C, Foster MC, O'Seaghdha CM, Glazer N, Isaacs A, Liu CT, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Johnson AD, Gierman HJ, Feitosa M, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson Å, Tönjes A, Dehghan A, Chouraki V, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Kollerits B, Pistis G; CARDIoGRAM Consortium; ICBP Consortium; CARe Consortium; Wellcome Trust Case Control Consortium 2 (WTCCC2), Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Cavalieri M, Rao M, Hu FB, **Demirkan A**, Oostra BA, de Andrade M, Turner ST, Ding J, Andrews JS, Freedman BI, Koenig W, Illig T, Döring A, Wichmann HE, Kolcic I, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Endlich K, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Ketkar S, Colhoun H, Doney A, Robino A, Giulianini F, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, Thun GA, Paulweber B, Haun M, Sala C, Metzger M, Mitchell P, Ciullo M, Kim SK, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Siscovick DS, van Duijn CM, Borecki I, Kardia SL, Liu Y, Curhan GC, Rudan I, Gyllensten U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman JC, Hayward C, Ridker P, Parsa A, Bochud M, Heid IM, Goessling W, Chasman DI, Kao WH, Fox CS. *Genome-wide association and functional follow-up reveals new loci for kidney function. PLoS Genet. 2012;8(3):e1002584. Epub 2012 Mar 29.*

Demirkan A, van Duijn CM, Ugocsai P, Isaacs A, Pramstaller PP, Liebisch G, Wilson JF, Johansson Å, Rudan I, Aulchenko YS, Kirichenko AV, Janssens AC, Jansen RC, Gnewuch C, Domingues FS,

Pattaro C, Wild SH, Jonasson I, Polasek O, Zorkoltseva IV, Hofman A, Karssen LC, Struchalin M, Floyd J, Igl W, Biloglav Z, Broer L, Pfeufer A, Pichler I, Campbell S, Zaboli G, Kolcic I, Rivadeneira F, Huffman J, Hastie ND, Uitterlinden A, Franke L, Franklin CS, Vitart V; DIAGRAM Consortium, Nelson CP, Preuss M; CARDIoGRAM Consortium, Bis JC, O'Donnell CJ, Franceschini N; CHARGE Consortium, Witteman JC, Axenovich T, Oostra BA, Meitinger T, Hicks AA, Hayward C, Wright AF, Gyllensten U, Campbell H, Schmitz G; EUROSPAN consortium.*Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. PLoS Genet. 2012 Feb;8(2):e1002490. Epub 2012 Feb 16.*

Vural B, Yakar F, Derin D, Saip P, Yakar A, **Demirkan A**, Karabulut A, Ugurel E, Cine N, Kilicaslan Z, Tüzün E, Ozbek U. *Evaluation of Glutathione S-Transferase P1 Polymorphisms (Ile105Val and Ala114Val) in Patients with Small Cell Lung Cancer. Genet Test Mol Biomarkers. 2012 Jul;16(7):701-6. Epub 2012 Feb 17.*

Taal HR, Verwoert GC, **Demirkan A**, Janssens AC, Rice K, Ehret G, Smith AV, Verhaaren BF, Witteman JC, Hofman A, Vernooij MW, Uitterlinden AG, Rivadeneira F, Ikram MA, Levy D, van der Heijden AJ; Cohort for Heart and Aging Research in Genome Epidemiology and Early Genetics and Lifecourse Epidemiology consortia, Jaddoe VW, van Duijn CM. *Genome-wide profiling of blood pressure in adults and children. Hypertension. 2012 Feb;59(2):241-7. Epub 2011 Dec 27.*

International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, Pihur V, Vollenweider P, O'Reilly PF, Amin N, Bragg-Gresham JL, Teumer A, Glazer NL, Launer L, Zhao JH, Aulchenko Y, Heath S, Sõber S, Parsa A, Luan J, Arora P, Dehghan A, Zhang F, Lucas G, Hicks AA, Jackson AU, Peden JF, Tanaka T, Wild SH, Rudan I, Igl W, Milaneschi Y, Parker AN, Fava C, Chambers JC, Fox ER, Kumari M, Go MJ, van der Harst P, Kao WH, Sjögren M, Vinay DG, Alexander M, Tabara Y, Shaw-Hawkins S, Whincup PH, Liu Y, Shi G, Kuusisto J, Tayo B, Seielstad M, Sim X, Nguyen KD, Lehtimäki T, Matullo G, Wu Y, Gaunt TR, Onland-Moret NC, Cooper MN, Platou CG, Org E, Hardy R, Dahgam S, Palmen J, Vitart V, Braund PS, Kuznetsova T, Uiterwaal CS, Adeyemo A, Palmas W, Campbell H, Ludwig B, Tomaszewski M, Tzoulaki I, Palmer ND; CARDIoGRAM consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen consortium; CHARGE-HF consortium, Aspelund T, Garcia M, Chang YP, O'Connell JR, Steinle NI, Grobbee DE, Arking DE, Kardia SL, Morrison AC, Hernandez D, Najjar S, McArdle WL, Hadley D, Brown MJ, Connell JM, Hingorani AD, Day IN, Lawlor DA, Beilby JP, Lawrence RW, Clarke R, Hopewell JC, Ongen H, Dreisbach AW, Li Y, Young JH, Bis JC, Kähönen M, Viikari J, Adair LS, Lee NR, Chen MH, Olden M, Pattaro C, Bolton JA, Köttgen A, Bergmann S, Mooser V, Chaturvedi N, Frayling TM, Islam M, Jafar TH, Erdmann J, Kulkarni SR, Bornstein SR, Grässler J, Groop L, Voight BF, Kettunen J, Howard P, Taylor A, Guarrera S, Ricceri F, Emilsson V, Plump A, Barroso I, Khaw KT, Weder AB, Hunt SC, Sun YV, Bergman RN, Collins FS, Bonnycastle LL, Scott LJ, Stringham HM, Peltonen L, Perola M, Vartiainen E, Brand SM,

Staessen JA, Wang TJ, Burton PR, Soler Artigas M, Dong Y, Snieder H, Wang X, Zhu H, Lohman KK, Rudock ME, Heckbert SR, Smith NL, Wiggins KL, Doumatey A, Shriner D, Veldre G, Viigimaa M, Kinra S, Prabhakaran D, Tripathy V, Langefeld CD, Rosengren A, Thelle DS, Corsi AM, Singleton A, Forrester T, Hilton G, McKenzie CA, Salako T, Iwai N, Kita Y, Ogihara T, Ohkubo T, Okamura T, Ueshima H, Umemura S, Eyheramendy S, Meitinger T, Wichmann HE, Cho YS, Kim HL, Lee JY, Scott J, Sehmi JS, Zhang W, Hedblad B, Nilsson P, Smith GD, Wong A, Narisu N, Stančáková A, Raffel LJ, Yao J, Kathiresan S, O'Donnell CJ, Schwartz SM, Ikram MA, Longstreth WT Jr, Mosley TH, Seshadri S, Shrine NR, Wain LV, Morken MA, Swift AJ, Laitinen J, Prokopenko I, Zitting P, Cooper JA, Humphries SE, Danesh J, Rasheed A, Goel A, Hamsten A, Watkins H, Bakker SJ, van Gilst WH, Janipalli CS, Mani KR, Yajnik CS, Hofman A, Mattace-Raso FU, Oostra BA, **Demirkan A**, Isaacs A, Rivadeneira F, Lakatta EG, Orru M, Scuteri A, Ala-Korpela M, Kangas AJ, Lyytikäinen LP, Soininen P, Tukiainen T, Würtz P, Ong RT, Dörr M, Kroemer HK, Völker U, Völzke H, Galan P, Hercberg S, Lathrop M, Zelenika D, Deloukas P, Mangino M, Spector TD, Zhai G, Meschia JF, Nalls MA, Sharma P, Terzic J, Kumar MV, Denniff M, Zukowska-Szczechowska E, Wagenknecht LE, Fowkes FG, Charchar FJ, Schwarz PE, Hayward C, Guo X, Rotimi C, Bots ML, Brand E, Samani NJ, Polasek O, Talmud PJ, Nyberg F, Kuh D, Laan M, Hveem K, Palmer LJ, van der Schouw YT, Casas JP, Mohlke KL, Vineis P, Raitakari O, Ganesh SK, Wong TY, Tai ES, Cooper RS, Laakso M, Rao DC, Harris TB, Morris RW, Dominiczak AF, Kivimaki M, Marmot MG, Miki T, Saleheen D, Chandak GR, Coresh J, Navis G, Salomaa V, Han BG, Zhu X, Kooner JS, Melander O, Ridker PM, Bandinelli S, Gyllensten UB, Wright AF, Wilson JF, Ferrucci L, Farrall M, Tuomilehto J, Pramstaller PP, Elosua R, Soranzo N, Sijbrands EJ, Altshuler D, Loos RJ, Shuldiner AR, Gieger C, Meneton P, Uitterlinden AG, Wareham NJ, Gudnason V, Rotter JI, Rettig R, Uda M, Strachan DP, Witteman JC, Hartikainen AL, Beckmann JS, Boerwinkle E, Vasan RS, Boehnke M, Larson MG, Järvelin MR, Psaty BM, Abecasis GR, Chakravarti A, Elliott P, van Duijn CM, Newton-Cheh C, Levy D, Caulfield MJ, Johnson T. *Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011 Sep 11;478(7367):103-9.*

Saip P, Sen F, Vural B, Ugurel E, **Demirkan A**, Derin D, Eralp Y, Camlica H, Ustuner Z, Ozbek U. *Glutathione S-transferase P1 polymorphisms are associated with time to tumor progression in small cell lung cancer patients. J BUON. 2011 Apr-Jun;16(2):241-6.*

Kohli MA, Lucae S, Saemann PG, Schmidt MV, **Demirkan A**, Hek K, Czamara D, Alexander M, Salyakina D, Ripke S, Hoehn D, Specht M, Menke A, Hennings J, Heck A, Wolf C, Ising M, Schreiber S, Czisch M, Müller MB, Uhr M, Bettecken T, Becker A, Schramm J, Rietschel M, Maier W, Bradley B, Ressler KJ, Nöthen MM, Cichon S, Craig IW, Breen G, Lewis CM, Hofman A, Tiemeier H, van Duijn CM, Holsboer F, Müller-Myhsok B, Binder EB. *The neuronal transporter gene SLC6A15 confers risk to major depression. Neuron. 2011 Apr 28;70(2):252-65.*

Gormez S, **Demirkan A**, Atalar F, Caynak B, Erdim R, Sozer V, Gunay D, Akpinar B, Ozbek U, Buyukdevrim AS. *Adipose tissue gene expression of adiponectin, tumor necrosis factor-α and leptin in metabolic syndrome patients with coronary artery disease. Intern Med. 2011;50(8):805-10.*

Demirkan A, Amin N, Isaacs A, Jarvelin MR, Whitfield JB, Wichmann HE, Kyvik KO, Rudan I, Gieger C, Hicks AA, Johansson Å, Hottenga JJ, Smith JJ, Wild SH, Pedersen NL, Willemsen G, Mangino M, Hayward C, Uitterlinden AG, Hofman A, Witteman J, Montgomery GW, Pietiläinen KH, Rantanen T, Kaprio J, Döring A, Pramstaller PP, Gyllensten U, de Geus EJ, Penninx BW, Wilson JF, Rivadeneria F, Magnusson PK, Boomsma DI, Spector T, Campbell H, Hoehne B, Martin NG, Oostra BA, McCarthy M, Peltonen-Palotie L, Aulchenko Y, Visscher PM, Ripatti S, Janssens AC, van Duijn CM; ENGAGE CONSORTIUM. *Genetic architecture of circulating lipid levels. Eur J Hum Genet. 2011 Jul;19(7):813-9.*

Fox ER, Young JH, Li Y, Dreisbach AW, Keating BJ, Musani SK, Liu K, Morrison AC, Ganesh S, Kutlar A, Ramachandran VS, Polak JF, Fabsitz RR, Dries DL, Farlow DN, Redline S, Adeyemo A, Hirschorn JN, Sun YV, Wyatt SB, Penman AD, Palmas W, Rotter JI, Townsend RR, Doumatey AP, Tayo BO, Mosley TH Jr, Lyon HN, Kang SJ, Rotimi CN, Cooper RS, Franceschini N, Curb JD, Martin LW, Eaton CB, Kardia SL, Taylor HA, Caulfield MJ, Ehret GB, Johnson T; International Consortium for Blood Pressure Genome-wide Association Studies (ICBP-GWAS),Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, Pihur V, Vollenweider P, O'Reilly PF, Amin N, Bragg-Gresham JL, Teumer A, Glazer NL, Launer L, Zhao JH, Aulchenko Y, Heath S, Sõber S, Parsa A, Luan J, Arora P, Dehghan A, Zhang F, Lucas G, Hicks AA, Jackson AU, Peden JF, Tanaka T, Wild SH, Rudan I, Igl W, Milaneschi Y, Parker AN, Fava C, Chambers JC, Kumari M, Go MJ, van der Harst P, Kao WH, Sjögren M, Vinay DG, Alexander M, Tabara Y, Shaw-Hawkins S, Whincup PH, Liu Y, Shi G, Kuusisto J, Seielstad M, Sim X, Nguyen KD, Lehtimäki T, Matullo G, Wu Y, Gaunt T, Onland-Moret NC, Cooper MN, Platou CG, Org E, Hardy R, Dahgam S, Palmen J, Vitart V, Braund P, Kuznetsova T, Uiterwaal CS, Campbell H, Ludwig B, Tomaszewski M, Tzoulaki I, Palmer ND, CARDIoGRAM consortium, CKDGen consortium, KidneyGen consortium, EchoGen consortium, CHARGE-HF consortium, Aspelund T, Garcia M, Chang YP, O'Connell JR, Steinle NI, Grobbee DE, Arking DE, Hernandez D, Najjar S, McArdle WL, Hadley D, Brown MJ, Connell JM, Hingorani AD, Day IN, Lawlor DA, Beilby JP, Lawrence RW, Clarke R, Collins R, Hopewell JC, Ongen H, Bis JC, Kähönen M, Viikari J, Adair LS, Lee NR, Chen MH, Olden M, Pattaro C, Hoffman Bolton JA, Köttgen A, Bergmann S, Mooser V, Chaturvedi N, Frayling TM, Islam M, Jafar TH, Erdmann J, Kulkarni SR, Bornstein SR, Grässler J, Groop L, Voight BF, Kettunen J, Howard P, Taylor A, Guarrera S, Ricceri F, Emilsson V, Plump A, Barroso I, Khaw KT, Weder AB, Hunt SC, Bergman RN, Collins FS, Bonnycastle LL, Scott LJ, Stringham HM, Peltonen L, Perola M, Vartiainen E, Brand SM, Staessen JA, Wang TJ, Burton PR, Soler Artigas M, Dong Y, Snieder H, Wang X, Zhu H, Lohman KK, Rudock ME, Heckbert SR, Smith NL, Wiggins KL, Shriner D, Veldre G, Viigimaa M, Kinra S, Prabhakaran D, Tripathy V, Langefeld CD, Rosengren A, Thelle DS, Corsi AM, Singleton A, Forrester T, Hilton G,

McKenzie CA, Salako T, Iwai N, Kita Y, Ogihara T, Ohkubo T, Okamura T, Ueshima H, Umemura S, Eyheramendy S, Meitinger T, Wichmann HE, Cho YS, Kim HL, Lee JY, Scott J, Sehmi J, Zhang W, Hedblad B, Nilsson P, Smith GD, Wong A, Narisu N, Stancáková A, Raffel LJ, Yao J, Kathiresan S, O'Donnell C, Schwartz SM, Ikram MA, Longstreth WT Jr, Seshadri S, Shrine NR, Wain LV, Morken MA, Swift AJ, Laitinen J, Prokopenko I, Zitting P, Cooper JA, Humphries S, Danesh J, Rasheed A, oel A, Hamsten A, Watkins H, Bakker SJ, van Gilst WH, Janipalli CS, Mani KR, Yajnik CS, Hofman A, Mattace-Raso FU, Oostra BA, **Demirkan A**, Isaacs A, Rivadeneira F, Lakatta EG, Orru M, Scuteri A, Ala-Korpela M, Kangas AJ, Lyytikäinen LP, Soininen P, Tukiainen T, Würz P, Ong RT, Dörr M, Kroemer HK, Völker U, Völzke H, Galan P, Hercberg S, Lathrop M, Zelenika D, Deloukas P, Mangino M, Spector TD, Zhai G, Meschia JF, Nalls MA, Sharma P, Terzic J, Kumar MJ, Denniff M, Zukowska-Szczechowska E, Wagenknecht LE, Fowkes FG, Charchar FJ, Schwarz PE, Hayward C, Guo X, Bots ML, Brand E, Samani NJ, Polasek O, Talmud PJ, Nyberg F, Kuh D, Laan M, Hveem K, Palmer LJ, van der Schouw YT, Casas JP, Mohlke KL, Vineis P, Raitakari O, Wong TY, Tai ES, Laakso M, Rao DC, Harris TB, Morris RW, Dominiczak AF, Kivimaki M, Marmot MG, Miki T, Saleheen D, Chandak GR, Coresh J, Navis G, Salomaa V, Han BG, Kooner JS, Melander O, Ridker PM, Bandinelli S, Gyllensten UB, Wright AF, Wilson JF, Ferrucci L, Farrall M, Tuomilehto J, Pramstaller PP, Elosua R, Soranzo N, Sijbrands EJ, Altshuler D, Loos RJ, Shuldiner AR, Gieger C, Meneton P, Uitterlinden AG, Wareham NJ, Gudnason V, Rettig R, Uda M, Strachan DP, Witteman JC, Hartikainen AL, Beckmann JS, Boerwinkle E, Boehnke M, Larson MG, Järvelin MR, Psaty BM, Abecasis GR, Elliott P, van Duijn C, Newton-Cheh C. Chakravarti A, Zhu X, Levy D. *Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. Hum Mol Genet. 2011 Jun 1;20(11):2273-84.*

Ramdas WD, Amin N, van Koolwijk LM, Janssens AC, Demirkan A, de Jong PT, Aulchenko YS, Wolfs RC, Hofman A, Rivadeneira F, Uitterlinden AG, Oostra BA, Lemij HG, Klaver CC, Vingerling JR, Jansonius NM, van Duijn CM. *Genetic architecture of open angle glaucoma and related determinants.*
J Med Genet. 2011 Mar;48(3):190-6.

Demirkan A, Penninx BWJH, Hek K, Wray NR, Amin N, Aulchenko YS, van Dyck R, de Geus EJ, Hofman A, Uitterlinden AG, Hottenga JJ, Nolen WA, Oostra BA, Sullivan PF, Willemsen G, Zitman FG, Tiemeier H, Janssens AC, Boomsma DI, van Duijn CM, Middeldorp CM. *Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry. 2011 Jul;16(7):773-83.*
Vural P, Degirmencioglu S, Saral NY, **Demirkan A**, Akgul C, Yildirim G, Issever H, Eroglu H. *Tumor necrosis factor alpha, interleukin-6 and interleukin-10 polymorphisms in preeclampsia. J Obstet Gynaecol Res.* 2010 Feb;36(1):64-71.

Hicks AA, Pramstaller PP, Johansson A, Vitart V, Rudan I, Ugocsai P, Aulchenko Y, Franklin CS, Liebisch G, Erdmann J, Jonasson I, Zorkoltseva IV, Pattaro C, Hayward C, Isaacs A, Hengstenberg

C, Campbell S, Gnewuch C, Janssens AC, Kirichenko AV, König IR, Marroni F, Polasek O, **Demirkan A**, Kolcic I, Schwienbacher C, Igl W, Biloglav Z, Witteman JC, Pichler I, Zaboli G, Axenovich TI, Peters A, Schreiber S, Wichmann HE, Schunkert H, Hastie N, Oostra BA, Wild SH, Meitinger T, Gyllensten U, van Duijn CM, Wilson JF, Wright A, Schmitz G, Campbell H. *Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet. 2009 Oct;5(10):e1000672.*

Vural B, **Demirkan A**, Ugurel E, Kalaylioglu-Wheeler Z, Esen BA, Gure AO, Gül A, Ozbek U. *Seroreactivity against PTEN-induced putative kinase 1 (PINK1) in Turkish patients with Behçet's disease.* Clin Exp Rheumatol. 2009 Mar-Apr;27(2 Suppl 53):S67-72.

Vural B, Atalar F, Ciftci C, **Demirkan A**, Susleyici-Duman B, Gunay D, Akpinar B, Sagbas E, Ozbek U, Buyukdevrim AS. *Presence of Fatty Acid Binding Protein 4 Expression in Human Epicardial Adipose Tissue in Metabolic Syndrome.* Journal Of Cardiovascular Pathology, Cardiovasc Pathol. 2008 Nov-Dec;17(6):392-8.

ABOUT THE AUTHOR

Ayşe Demirkan was born in Istanbul, Turkey on 3 March, 1980. She finished Burak Bora Anatolian High School (1997) and completed her bachelor education in Istanbul University, Department of Biology with a specification in biological oceanography (2002). Meanwhile she worked as an intern in the Genetic Unit of Institute for Experimental and Medical Research (DETAE) between 1999 and 2002 with Prof. Dr. Ugur Ozbek and Prof. Dr. Nihan Unaltuna. In 2002 she worked as an intern in St. Jude Childrens' Research Hospital, Memphis, USA, under supervision of Prof. Dr. Sosa-Pineda. In 2004, she completed her master of science thesis in Istanbul University of Health Sciences on genetics, with her thesis titled "Promotor methylation of RAR-β gene in small cell lung cancer", supported by Istanbul University Research Fund. During her master period she also performed minimal residual disease follow up for leukemic translocations and served for the rutine molecular genetic diagnosis. She spent short term internships in Clinical Genetics Department of Istanbul Medical Center (Prof. Dr. Sukru Palanduz) and Child Genetics Department of Cerrahpasa Medical Center (Prof. Dr. Beyhan Tuysuz) where she worked on clinical diagnosis of rare genetic diseases. She received a "Young Researcher Award" in the Turkish National Hematology Congress in 2004 and the "Cancer Research First Prize" fromTurkish Foundation Against Cancer in 2005. During her first PhD period which she did not complete, she was granted by The Scientific and Technological Research Council of Turkey (TUBITAK) for the project "Identification of specific immunogenic antigens in Behçet's Disease" (2006) and by European Union 6th Frame Project EPICURE on epilepsia genetics (2007). She also received an Erasmus Exchange Mobility Scholarship from European Union (2007) and two times Querido Chair Exchange Scholarship from Erasmus Medical Center (2008). She completed her second master of science programme in Netherlands Instutite of Health Sciences, with specification in Genetic Epidemiology. She currently works as a research fellow in the Genetic Epidemiology Unite of Erasmus Medical Center, focusing on the genetic analysis of complex traits, under supervision of Prof. Dr. Cornelia van Duijn. She is planning to continue her genetic research on metabolomic enophenotypes and their relation to health and disease for better understanding of complex human genetic traits.

Since 2009 she is member of the collective living group and the cultural breeding place Poortgebouw in Binnenhaven, Rotterdam and enjoys her three adopted cats, Sushi, Mao and Kim Jong-il.