Genetic Determinants of Dyslipidemia

V Varga, Tibor

2016

Document Version:
Publisher's PDF, also known as Version of record

Genetic Determinants of Dyslipidemia

TIBOR V. VARGA
DEPARTMENT OF CLINICAL SCIENCES MALMÖ | LUND UNIVERSITY

Genetic Determinants of Dyslipidemia

Tibor V. Varga

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden will be publicly defended at Aulan, CRC, Jan Waldenströms gata 35, Skåne University Hospital, SE-205 02 Malmö, Sweden on Friday, 22nd April, 2016 at 9:00

Faculty opponent
Fredrik Karpe, MD, PhD
Professor of Metabolic Medicine
University of Oxford, UK

Organization	Document name
LUND UNIVERSITY	DOCTORAL DISSERTATION
	Date of issue: April 22nd, 2016
Author(s)	Sponsoring organization
Tibor V. Varga	

Title and subtitle: Genetic Determinants of Dyslipidemia

Abstract

Dyslipidemia is a chronic deviation from normal blood lipid levels that can lead to atherosclerosis and other cardiovascular diseases; dyslipidemia and its sequelae are caused by the complex interplay of genetic and environmental factors. Although circulating concentrations of total cholesterol (TC), triglycerides (TG), low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (LDL-C) have a strong genetic underpinning, not much is known about the genetic factors that affect long-term deteriorations in lipid concentrations. Through the work described in this thesis I sought to identify novel genetic loci associated with long-term lipid level changes and identify gene × environment interactions influencing blood lipid and lipoprotein concentrations.

In Papers I and II, large European prospective cohort studies with long-term follow-up were analyzed. The Gene–Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk (GLACIER) Study (N=3,495) was analyzed in the discovery phase of these studies. The MDC, PIVUS, ULSAM and MRC Ely studies (N_{max}=8,263) were utilized as replication cohorts. In Paper III, Scandinavian adults from the GLACIER, MDC, Inter99 and Health 2006 Studies were meta-analyzed (N_{max}=18,190). In Paper IV, analyses were conducted in the Diabetes Prevention Program (DPP) (N=2,993) multi-ethnic randomized clinical trial. Participants from the GLACIER Study and DPP, the two discovery studies intensively used in this thesis, were genotyped with the Illumina CardioMetaboChip array.

In Paper I, TC- and TG-specific genetic risk scores (GRSs) were robustly associated with TC- and TG level changes, respectively. Three genomic loci, *APOE*, *TRIB1* and *APOA1* were associated with either TC- or TG changes and were replicated in subsequent analyses. In Paper II, in addition to the findings of Paper I, seven further loci were associated with TC- or TG changes. Of these, variants at *CAPN3*, *HPR* and *SIX5* showed suggestive evidence for association with coronary artery disease. In Paper III, a robust sex-heterogeneous interaction between the TG-related GRS and body mass index was observed for circulating blood TG levels. In Paper IV, an interaction between the large HDL particle-associated GRS and the lifestyle intervention for large HDL particle concentrations was observed.

In conclusion, this thesis work shows genetic associations for long-term lipid changes and demonstrates examples of gene × environment interactions that influence blood lipid concentrations.

Key words: Coronary artery disease; Gene × environment interaction; Genetic associations studies; Genetic risk score; Genetics; HDL-C; Lipids; LDL-C; Prediction; Prospective studies; Randomized clinical trial; Single nucleotide polymorphism; Total cholesterol; Trait level change; Triglycerides.

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language: English
ISSN and key title: 1652-8220 Lund University, Faculty of Medicine Doctoral Dissertation Series 2016:32	ISBN: 978-91-7619-258-0
Recipient's notes	Number of pages / Price
	Security classification

Genetic Determinants of Dyslipidemia

Tibor V. Varga

LUND
UNIVERSITY

Genetic and Molecular Epidemiology (GAME) Unit,
Department of Clinical Sciences Malmö,
Faculty of Medicine, Lund University, Sweden

Cover picture by Péter Sántha and Tibor V. Varga

Lund University, Faculty of Medicine Doctoral Dissertation Series 2016:32
ISBN 978-91-7619-258-0
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2016

“The best thing for being sad," replied Merlin, beginning to puff and blow, "is to learn something. That's the only thing that never fails. You may grow old and trembling in your anatomies, you may lie awake at night listening to the disorder of your veins, you may miss your only love, you may see the world about you devastated by evil lunatics, or know your honour trampled in the sewers of baser minds. There is only one thing for it then — to learn. Learn why the world wags and what wags it. That is the only thing which the mind can never exhaust, never alienate, never be tortured by, never fear or distrust, and never dream of regretting. Learning is the only thing for you. Look what a lot of things there are to learn.”

— T.H. White, The Once and Future King

Contents

List of publications

Varga TV, Sonestedt E, Shungin D, Koivula RW, Hallmans G, Escher SA, Barroso I, Nilsson P, Melander O, Orho-Melander M, Renström F, Franks PW. *Genetic determinants of long-term changes in blood lipid concentrations: 10-year follow-up of the GLACIER Study*. PLoS Genetics. 2014 Jun 10(6):e1004388.

Varga TV, Kurbasic A, Aine M, Eriksson P, Ali A, Hindy G, Gustafsson S, Luan J, Shungin D, Chen Y, Schulz CA, Nilsson PM, Hallmans G, Barroso I, Deloukas P, Langenberg C, Scott RA, Wareham NJ, Lind L, Ingelsson E, Melander O, Orho-Melander M, Renström F, Franks PW. *Novel genetic loci associated with long-term deteriorations in blood lipid concentrations and coronary artery disease in European adults*. (manuscript)

Ali A*, **Varga TV***, Stojkovic I, Schulz CA, Hallmans G, Barroso I, Poveda A, Renström F, Orho-Melander M, Franks PW. *Do genetic factors modify the relationship between obesity and hypertriglyceridemia? Findings from the GLACIER and Malmö Diet and Cancer studies*. Circulation Cardiovascular Genetics, 2016 Feb 10. pii: CIRCGENETICS.115.001218.
*These authors contributed equally to this work

Varga TV, Winters AH, Jablonski KA, Horton ES, Khare-Ranade P, Knowler WC, Marcovina SM, Renström F, Watson KE, Goldberg R, Florez JC, Pollin TI, Franks PW. *Comprehensive analysis of established dyslipidemia-associated loci in the Diabetes Prevention Program*. (manuscript)

Publications not included in this thesis

1. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Magi R, Strawbridge RJ, Rehnberg E, Gustafsson S, Kanoni S, Rasmussen-Torvik LJ, Yengo L, Lecoeur C, Shungin D, Sanna S, Sidore C, Johnson PC, Jukema JW, Johnson T, Mahajan A, Verweij N, Thorleifsson G, Hottenga JJ, Shah S, Smith AV, Sennblad B, Gieger C, Salo P, Perola M, Timpson NJ, Evans DM, Pourcain BS, Wu Y, Andrews JS, Hui J, Bielak LF, Zhao W, Horikoshi M, Navarro P, Isaacs A, O'Connell JR, Stirrups K, Vitart V, Hayward C, Esko T, Mihailov E, Fraser RM, Fall T, Voight BF, Raychaudhuri S, Chen H, Lindgren CM, Morris AP, Rayner NW, Robertson N, Rybin D, Liu CT, Beckmann JS, Willems SM, Chines PS, Jackson AU, Kang HM, Stringham HM, Song K, Tanaka T, Peden JF, Goel A, Hicks AA, An P, Muller-Nurasyid M, Franco-Cereceda A, Folkersen L, Marullo L, Jansen H, Oldehinkel AJ, Bruinenberg M, Pankow JS, North KE, Forouhi NG, Loos RJ, Edkins S, **Varga TV**, Hallmans G, Oksa H, Antonella M, Nagaraja R, Trompet S, Ford I, Bakker SJ, Kong A, Kumari M, Gigante B, Herder C, Munroe PB, Caulfield M, Antti J, Mangino M, Small K, Miljkovic I, Liu Y, Atalay M, Kiess W, James AL, Rivadeneira F, Uitterlinden AG, Palmer CN, Doney AS, Willemsen G, Smit JH, Campbell S, Polasek O, Bonnycastle LL, Hercberg S, Dimitriou M, Bolton JL, Fowkes GR, Kovacs P, Lindstrom J, Zemunik T, Bandinelli S, Wild SH, Basart HV, Rathmann W, Grallert H, Replication DIG, Meta-analysis C, Maerz W, Kleber ME, Boehm BO, Peters A, Pramstaller PP, Province MA, Borecki IB, Hastie ND, Rudan I, Campbell H, Watkins H, Farrall M, Stumvoll M, Ferrucci L, Waterworth DM, Bergman RN, Collins FS, Tuomilehto J, Watanabe RM, de Geus EJ, Penninx BW, Hofman A, Oostra BA, Psaty BM, Vollenweider P, Wilson JF, Wright AF, Hovingh GK, Metspalu A, Uusitupa M, Magnusson PK, Kyvik KO, Kaprio J, Price JF, Dedoussis GV, Deloukas P, Meneton P, Lind L, Boehnke M, Shuldiner AR, van Duijn CM, Morris AD, Toenjes A, Peyser PA, Beilby JP, Korner A, Kuusisto J, Laakso M, Bornstein SR, Schwarz PE, Lakka TA, Rauramaa R, Adair LS, Smith GD, Spector TD, Illig T, de Faire U, Hamsten A, Gudnason V, Kivimaki M, Hingorani A, Keinanen-Kiukaanniemi SM, Saaristo TE, Boomsma DI, Stefansson K, van der Harst P, Dupuis J, Pedersen NL, Sattar N, Harris TB, Cucca F, Ripatti S, Salomaa V, Mohlke KL, Balkau B, Froguel P, Pouta A, Jarvelin MR, Wareham NJ, Bouatia-Naji N, McCarthy

MI, Franks PW, Meigs JB, Teslovich TM, Florez JC, Langenberg C, Ingelsson E, Prokopenko I and Barroso I. *Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways*. Nature Genetics. 2012;44:991-1005.

2. Albrechtsen A, Grarup N, Li Y, Sparso T, Tian G, Cao H, Jiang T, Kim SY, Korneliussen T, Li Q, Nie C, Wu R, Skotte L, Morris AP, Ladenvall C, Cauchi S, Stancakova A, Andersen G, Astrup A, Banasik K, Bennett AJ, Bolund L, Charpentier G, Chen Y, Dekker JM, Doney AS, Dorkhan M, Forsen T, Frayling TM, Groves CJ, Gui Y, Hallmans G, Hattersley AT, He K, Hitman GA, Holmkvist J, Huang S, Jiang H, Jin X, Justesen JM, Kristiansen K, Kuusisto J, Lajer M, Lantieri O, Li W, Liang H, Liao Q, Liu X, Ma T, Ma X, Manijak MP, Marre M, Mokrosinski J, Morris AD, Mu B, Nielsen AA, Nijpels G, Nilsson P, Palmer CN, Rayner NW, Renstrom F, Ribel-Madsen R, Robertson N, Rolandsson O, Rossing P, Schwartz TW, Group DESIRS, Slagboom PE, Sterner M, Consortium D, Tang M, Tarnow L, Tuomi T, van't Riet E, van Leeuwen N, **Varga TV**, Vestmar MA, Walker M, Wang B, Wang Y, Wu H, Xi F, Yengo L, Yu C, Zhang X, Zhang J, Zhang Q, Zhang W, Zheng H, Zhou Y, Altshuler D, t Hart LM, Franks PW, Balkau B, Froguel P, McCarthy MI, Laakso M, Groop L, Christensen C, Brandslund I, Lauritzen T, Witte DR, Linneberg A, Jorgensen T, Hansen T, Wang J, Nielsen R and Pedersen O. *Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes*. Diabetologia. 2013;56:298-310.

3. **Varga TV**, Hallmans G, Hu FB, Renstrom F and Franks PW. *Smoking status, snus use, and variation at the CHRNA5-CHRNA3-CHRNB4 locus in relation to obesity: the GLACIER study*. American Journal of Epidemiology. 2013;178:31-7.

4. Ahmad S, Rukh G, **Varga TV**, Ali A, Kurbasic A, Shungin D, Ericson U, Koivula RW, Chu AY, Rose LM, Ganna A, Qi Q, Stancakova A, Sandholt CH, Elks CE, Curhan G, Jensen MK, Tamimi RM, Allin KH, Jorgensen T, Brage S, Langenberg C, Aadahl M, Grarup N, Linneberg A, Pare G, InterAct C, Consortium D, Magnusson PK, Pedersen NL, Boehnke M, Hamsten A, Mohlke KL, Pasquale LT, Pedersen O, Scott RA, Ridker PM, Ingelsson E, Laakso M, Hansen T, Qi L, Wareham NJ, Chasman DI, Hallmans G, Hu FB, Renstrom F, Orho-Melander M and Franks PW. *Gene x physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry*. PLoS Genetics. 2013;9:e1003607.

5. Ahmad S, **Varga TV** and Franks PW. *Gene x environment interactions in obesity: the state of the evidence*. Human Heredity. 2013;75:106-15.

6. Stijnen P, Tuand K, **Varga TV**, Franks PW, Aertgeerts B and Creemers JW. *The association of common variants in PCSK1 with obesity: a HuGE review and meta-analysis*. American Journal of Epidemiology. 2014;180:1051-65.

7. Stijnen P, Tuand K, **Varga TV**, Franks PW, Aertgeerts B and Creemers JW. *The authors reply*. American Journal of Epidemiology. 2015;181:733-4.

8. Nead KT, Li A, Wehner MR, Neupane B, Gustafsson S, Butterworth A, Engert JC, Davis AD, Hegele RA, Miller R, den Hoed M, Khaw KT, Kilpelainen TO, Wareham N, Edwards TL, Hallmans G, **Varga TV**, Kardia SL, Smith JA, Zhao W, Faul JD, Weir D, Mi J, Xi B, Quinteros SC, Cooper C, Sayer AA, Jameson K, Grontved A, Fornage M, Sidney S, Hanis CL, Highland HM, Haring HU, Heni M, Lasky-Su J, Weiss ST, Gerhard GS, Still C, Melka MM, Pausova Z, Paus T, Grant SF, Hakonarson H, Price RA, Wang K, Scherag A, Hebebrand J, Hinney A, BioBank Japan A-BMIGC, Franks PW, Frayling TM, McCarthy MI, Hirschhorn JN, Loos RJ, Ingelsson E, Gerstein HC, Yusuf S, Beyene J, Anand SS and Meyre D. *Contribution of common non-synonymous variants in PCSK1 to body mass index variation and risk of obesity: a systematic review and meta-analysis with evidence from up to 331 175 individuals*. Human Molecular Genetics. 2015;24:3582-94.

9. Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA, Dauriz M, Hivert MF, Raghavan S, Lipovich L, Hidalgo B, Fox K, Huffman JE, An P, Lu Y, Rasmussen-Torvik LJ, Grarup N, Ehm MG, Li L, Baldridge AS, Stancakova A, Abrol R, Besse C, Boland A, Bork-Jensen J, Fornage M, Freitag DF, Garcia ME, Guo X, Hara K, Isaacs A, Jakobsdottir J, Lange LA, Layton JC, Li M, Hua Zhao J, Meidtner K, Morrison AC, Nalls MA, Peters MJ, Sabater-Lleal M, Schurmann C, Silveira A, Smith AV, Southam L, Stoiber MH, Strawbridge RJ, Taylor KD, **Varga TV**, Allin KH, Amin N, Aponte JL, Aung T, Barbieri C, Bihlmeyer NA, Boehnke M, Bombieri C, Bowden DW, Burns SM, Chen Y, Chen YD, Cheng CY, Correa A, Czajkowski J, Dehghan A, Ehret GB, Eiriksdottir G, Escher SA, Farmaki AE, Franberg M, Gambaro G, Giulianini F, Goddard WA, 3rd, Goel A, Gottesman O, Grove ML, Gustafsson S, Hai Y, Hallmans G, Heo J, Hoffmann P, Ikram MK, Jensen RA, Jorgensen ME, Jorgensen T, Karaleftheri M, Khor CC, Kirkpatrick A, Kraja AT, Kuusisto J, Lange EM, Lee IT, Lee WJ, Leong A, Liao J, Liu C, Liu Y, Lindgren CM, Linneberg A, Malerba G, Mamakou V, Marouli E, Maruthur NM, Matchan A, McKean-Cowdin R, McLeod O, Metcalf GA, Mohlke KL, Muzny DM,

Ntalla I, Palmer ND, Pasko D, Peter A, Rayner NW, Renstrom F, Rice K, Sala CF, Sennblad B, Serafetinidis I, Smith JA, Soranzo N, Speliotes EK, Stahl EA, Stirrups K, Tentolouris N, Thanopoulou A, Torres M, Traglia M, Tsafantakis E, Javad S, Yanek LR, Zengini E, Becker DM, Bis JC, Brown JB, Cupples LA, Hansen T, Ingelsson E, Karter AJ, Lorenzo C, Mathias RA, Norris JM, Peloso GM, Sheu WH, Toniolo D, Vaidya D, Varma R, Wagenknecht LE, Boeing H, Bottinger EP, Dedoussis G, Deloukas P, Ferrannini E, Franco OH, Franks PW, Gibbs RA, Gudnason V, Hamsten A, Harris TB, Hattersley AT, Hayward C, Hofman A, Jansson JH, Langenberg C, Launer LJ, Levy D, Oostra BA, O'Donnell CJ, O'Rahilly S, Padmanabhan S, Pankow JS, Polasek O, Province MA, Rich SS, Ridker PM, Rudan I, Schulze MB, Smith BH, Uitterlinden AG, Walker M, Watkins H, Wong TY, Zeggini E, Consortium EP-I, Laakso M, Borecki IB, Chasman DI, Pedersen O, Psaty BM, Tai ES, van Duijn CM, Wareham NJ, Waterworth DM, Boerwinkle E, Kao WH, Florez JC, Loos RJ, Wilson JG, Frayling TM, Siscovick DS, Dupuis J, Rotter JI, Meigs JB, Scott RA and Goodarzi MO. *Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility*. Nature Communications. 2015;6:5897.

10. Fretts AM, Follis JL, Nettleton JA, Lemaitre RN, Ngwa JS, Wojczynski MK, Kalafati IP, **Varga TV**, Frazier-Wood AC, Houston DK, Lahti J, Ericson U, van den Hooven EH, Mikkila V, Kiefte-de Jong JC, Mozaffarian D, Rice K, Renstrom F, North KE, McKeown NM, Feitosa MF, Kanoni S, Smith CE, Garcia ME, Tiainen AM, Sonestedt E, Manichaikul A, van Rooij FJ, Dimitriou M, Raitakari O, Pankow JS, Djousse L, Province MA, Hu FB, Lai CQ, Keller MF, Perala MM, Rotter JI, Hofman A, Graff M, Kahonen M, Mukamal K, Johansson I, Ordovas JM, Liu Y, Mannisto S, Uitterlinden AG, Deloukas P, Seppala I, Psaty BM, Cupples LA, Borecki IB, Franks PW, Arnett DK, Nalls MA, Eriksson JG, Orho-Melander M, Franco OH, Lehtimaki T, Dedoussis GV, Meigs JB and Siscovick DS. *Consumption of meat is associated with higher fasting glucose and insulin concentrations regardless of glucose and insulin genetic risk scores: a meta-analysis of 50,345 Caucasians*. The American Journal of Clinical Nutrition. 2015;102:1266-78.

11. Renstrom F, Koivula RW, **Varga TV**, Hallmans G, Mulder H, Florez JC, Hu FB and Franks PW. *Season-dependent associations of circadian rhythm-regulating loci (CRY1, CRY2 and MTNR1B) and glucose homeostasis: the GLACIER Study*. Diabetologia. 2015;58:997-1005.

12. Stitziel NO, Stirrups KE, Masca NGD, Erdmann J, Ferrario PG, König IR, Weeke PE, Webb TR, Auer PL, Schick UM, Lu Y, Zhang H, Dube M, Goel A, Farrall M, Peloso GM, Won H, Do R, van Iperen E, Kanoni S, Kruppa J, Mahajan A, Scott RA, Willenborg C, Braund PS, van Capelleveen JC, Doney ASF, Donnelly LA, Asselta R, Merlini P, Duga S, Marziliano N, Denny JC, Shaffer C, El-Mokhtari NE, Franke A, Gottesman O, Heilmann S, Hengstenberg C, Hoffmann P, Holmen OL, Hveem K, Jansson JH, Jöckel KH, Kessler T, Kriebel J, Laugwitz KL, Marouli E, Martinelli N, McCarthy MI, van Zuydam NR, Meisinger C, Esko T, Mihailov E, Escher SA, Alver M, Moebus S, Morris AD, Müller-Nurasyid M, Nikpay M, Olivieri O, Perreault LPL, Qarawi A, Robertson NR, Akinsansya KO, Reilly DF, Vogt TF, Yin W, Asselbergs FW, Kooperberg C, Jackson R, Stahl E, Strauch K, **Varga TV**, Waldenberger M, Wellcome Trust Case Control Consortium, Zeng L, Kraja AT, Liu C, Ehret GB, Newton-Cheh C, Chasman DI, Chowdhury R, Ferrario M, Ford I, Jukema JW, Kee F, Kuulasmaa K, MORGAM Investigators, Nordestgaard BG, Perola M, Saleheen D, Sattar N, Surendran P, Tregouet D, Young R, Howson JMM, Butterworth AS, Danesh J, Ardissino D, Bottinger EP, Erbel R, Franks PW, Girelli D, Hall AS, Hovingh GK, Kastrati A, Lieb W, Meitinger T, Kraus WE, Shah SH, McPherson R, Orho-Melander M, Melander O, Metspalu A, Palmer CNA, Peters A, Rader DJ, Reilly MP, Loos RJF, Reiner AP, Roden DM, Tardif JC, Thompson JR, Wareham NJ, Watkins H, Willer CJ, Kathiresan S, Deloukas P, Samani NJ, Schunkert H. *Coding Variation in* ANGPTL4, LPL, *and* SVEP1 *and the Risk of Coronary Disease*. The New England Journal of Medicine. 2016; DOI: 10.1056/NEJMoa1507652

Abbreviations

AALC – average annual lipid change

AUC – area under the curve

BMI – body mass index

CAD – coronary artery disease

CARDIoGRAMplusC4D – Coronary ARtery DIsease Genome-wide Replication and Meta-analysis Consortium

CETP - cholesteryl ester transfer protein

CHD – coronary heart disease

CVD – cardiovascular disease

Δ (delta) – this symbol refers to trait change

DHS – DNaseI hypersensitive site

DIAGRAM – DIAbetes Genetics Replication And Meta-analysis Consortium

DPP – Diabetes Prevention Program

FFQ – food frequency questionnaire

GLACIER – Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk

GLGC – Global Lipids Genetics Consortium

GLM – generalized linear model

GRS – genetic risk score

GWAS – genome-wide association study

HDL-C – high-density lipoprotein cholesterol

HRPD – Human Protein Reference Database

ICBP – International Consortium for Blood Pressure

IDL – intermediate-density lipoprotein

ILI – intensive lifestyle intervention

LD – linkage disequilibrium

LDL-C – low-density lipoprotein cholesterol

LOCUS - LOngitudinal traits - ConsortiUm of prospective Studies

MAF – minor allele frequency

MDC/MDCS – Malmö Diet and Cancer Study

MI – myocardial infarction

MRC Ely – Medical Research Council Ely

OGTT – oral glucose tolerance test

PC – principal component

PCA – principal component analysis

PCSK9 - proprotein convertase subtilisin/kexin type 9

PIVUS – Prospective Investigation of the Vasculature in Uppsala Seniors

QC – quality control

RCT – randomized controlled trial

ROC – receiver operating curve

SNP – single nucleotide polymorphism

TC – total cholesterol

TG – triglyceride

ULSAM – Uppsala Longitudinal Study of Adult Men

VIP – Västerbotten Intervention Programme

VLDL – very high-density lipoprotein

wGRS – weighted genetic risk score

WHO – World Health Organization

Introduction

Dyslipidemia is a pathologic, chronic deviation from normal blood lipid levels and is a heterogeneous disease with many subclasses,[1] caused by the complex interplay of a plethora of genetic and environmental factors.[2] Dyslipidemia is on the causal pathway to a range of cardiovascular diseases (CVDs), including atherosclerosis, myocardial infarction (MI), coronary artery disease (CAD) and stroke [3-5] and is often associated with comorbidities such as obesity, type 2 diabetes, high blood pressure and non-alcoholic fatty liver disease.[6-11] CVDs – a group of diseases considered rarities 100 years ago – have increased to epidemic proportions and are now the leading causes of death worldwide.[12, 13] According to the World Health Organization (WHO), in 2012, approximately 30% of global deaths were attributable to CVDs and their comorbidities.[14] As the global prevalence of CVDs is increasing, these diseases continue to consume huge portions of national healthcare budgets.[14] As dyslipidemia is one of the leading causes of CVDs, huge efforts have been undertaken to study the environmental and genetic background of lipid levels and lifestyle interventions and pharmacological therapies to normalize lipid levels are in the frontline of research and development.

Blood lipid levels are heritable and notably modifiable phenotypes that vary considerably among people in general. Their levels are difficult to categorize as "beneficial" or "adverse for health". Just as obesity is defined based on arbitrary body mass index (BMI) categories [15] and diabetes is defined based on arbitrary fasting/post-challenge glucose concentrations [16], dyslipidemias are also defined by arbitrary cut-offs in lipid levels.[17, 18] The four most studied blood lipids traits are triglycerides (TG), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) and total cholesterol (TC) concentrations. Apart from these traits, a number of other lipid phenotypes have important biological functions. The most common dyslipidemias are hyperlipidemias – chronic elevations in blood lipid levels. The two most prevalent subtypes are hypertriglyceridemia (defined by high TG concentrations) and hypercholesterolemia (defined by high LDL-C or TC), diseases which can be further stratified into more defined subclasses based on etiology and phenotype.[1] Other dyslipidemias, such as hypocholesterolemia are characterized by lower-than-normal blood lipid concentrations.[19] For instance, hypoalphalipoproteinemia is a deficiency with low HDL-C levels often resulting in premature CAD.[20]

There are thousands of published observational studies on the etiology of dyslipidemias, many of which have examined genetic, environmental, and behavioral risk factors such as diet and physical activity. Many published reports explore the effects of lifestyle interventions and medications on dyslipidemia and their interactions with genetic predisposition to dyslipidemia. The following section

will provide an overview of this literature and highlight research questions that are addressed later in this thesis. Specifically, I will

i. define the most important lipid traits and their role in lipid metabolism;
ii. overview how various lifestyles and pharmacological interventions relate to dyslipidemias;
iii. overview the main genetic discoveries in lipid research and summarize what is known about prospective genetic associations and gene × environment interactions in lipid traits;
iv. give a brief overview on how known genetic loci relate to lipid metabolism.

The good, the bad and the ugly – trait definitions

Lipids are heterogeneous in size, function and chemical properties. The most studied lipid traits in genetic epidemiologic studies are TG, LDL-C, HDL-C and TC, as they are easier and cheaper to measure with standard laboratory equipment than other more refined lipid phenotypes.

To briefly define these phenotypes, TG is a glycerol ester of three fatty acids.[21] TG's most important role is energy storage. Lipids contain approximately double the energy (9.3 calories/g) as carbohydrates or proteins (4.1 calories/g) – large volumes of TG are stored in white adipose tissue waiting to be hydrolyzed and converted to energy.[21] Cholesterols are steroid alcohols and are very important membrane lipids; among other functions, they control membrane structure and fluidity. Cholesterols are also constituents of bile, hormones and vitamin D.[21, 22] External cholesterol from food sources is absorbed by the intestines in the form of free cholesterol. Internal cholesterol is synthesized mostly in the liver and enterocytes through the mevalonate pathway.[23, 24] LDL and HDL are lipoproteins that transport hydrophobic TG and cholesterol in the blood, which is a hydrophilic, aqueous fluid.[21] Lipoproteins are complex multi-molecular structures assembled from lipids and proteins, as their name *lipoprotein* indicates. They are structured as an amphipathic layer of lipids, which harbors cholesterol and proteins (termed apolipoproteins) and an inner, hydrophobic core where TG and cholesterol-esters are stored during transport.[21, 25] There is an inverse correlation between lipoprotein density and size. HDL has the highest density and the lowest size among lipoproteins, while LDL has a lower density and bigger size.[21] Each lipoprotein subclass can be characterized by size, density and a unique set of apolipoproteins in their membranes. When we study HDL-C and LDL-C, we study the amount of cholesterol carried by HDL and LDL, respectively. In the laboratory, lipoprotein classes are first separated from blood and then from each other; cholesterol is extracted and subsequently measured from each

lipoprotein class separately. In layman's terms, LDL-C is often termed "bad cholesterol" and HDL-C is often termed "good cholesterol" based on LDL's and HDL's biological functions [26] – LDL carries cholesterol to the periphery (adipose- and muscle tissue and the endothelium) and considered an atherogenic lipoprotein subclass,[21] while HDL carries cholesterol back from the periphery (termed reverse cholesterol transport) and is therefore considered atheroprotective lipoproteins.[27, 28]

Other lipoproteins such as chylomicrons, intermediate-density lipoprotein (IDL), very low-density lipoprotein (VLDL) and other refined subclasses of each lipoprotein classes have been identified based on size, density and biological function. The lipoprotein fractions mentioned above can be further categorized into small, medium and large HDL, LDL and VLDL subfractions.

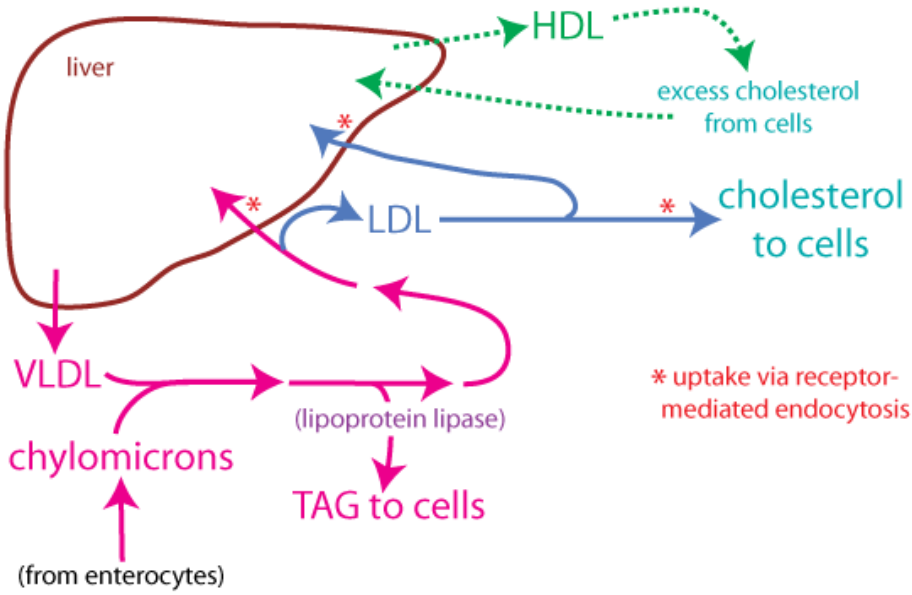

Figure 1. The endogenous and exogenous lipoprotein pathways.
HDL – high-density lipoprotein; LDL – low-density lipoprotein; TAG – triglycerides (triacylglycerol); VLDL – very-low-density lipoprotein. Ref: University of Washington, UW Courses Web Server: https://courses.washington.edu/conj/bess/cholesterol/liver.html (accessed: 15 March 2016)
A detailed figure of the endogenous and exogenous lipoprotein pathways can be found in Figure 1 of Lusis AJ *et al.* A treasure trove for lipoprotein biology. Nature Genetics. 2008;40:129-130.

To adequately understand lipoprotein metabolism, one needs to consider the exogenous and endogenous lipoprotein pathways (Figure 1).[29] In brief, the exogenous pathway involves chylomicrons carrying TGs (which were absorbed by the intestines) to the adipose tissue for storage, to the muscle tissue in need of energy or to the liver for further use (e.g. the creation of VLDL particles). The chylomicrons finish their journey at the liver where their constituents are reused.[21, 25] The

endogenous pathway starts at the liver with the assembly of VLDL particles. VLDL particles are similar in constitution and physiological role to chylomicrons, as they are large, very low in density and carry mostly TGs. VLDL particles are released from the liver and carry out similar functions in the endogenous pathway as the chylomicrons do in the exogenous pathway. As they release more and more TG in the periphery, they decrease in size, increase in density and become VLDL remnants or IDL particles. IDL is an intermediary lipoprotein between VLDL and LDL. As IDL restructures and recruits a unique set of apolipoproteins, it becomes an IDL-remnant or LDL, which carries less TG and more cholesterol than the particles mentioned above. LDL is subsequently degraded in the liver and its components are reused there.[21, 25]

HDL is centrally involved in both the exogenous and endogenous pathways. It is produced by the liver and its main function is reverse cholesterol transport from the periphery, most importantly, from the endothelium.[27, 28] HDL is also responsible for converting lipoproteins into their mature forms (e.g. HDL converts VLDL to IDL and subsequently, LDL).[21]

Atherosclerosis is a complex phenotype directly leading to a range of clinical outcomes, such as CAD and stroke.[5, 30, 31] While it is a chronic inflammatory state of the arterial wall with multiple factors contributing [32], one of the most important predictors of atherosclerosis is the imbalance in lipoprotein functions.[25] In brief, the majority of serum cholesterol is carried by LDL particles.[21] Although LDL transports a large fraction of its content to muscle- and adipose tissue, some of the cholesterol reaches the arterial endothelium. Lipid accumulation in the endothelium attracts macrophages and as macrophages accumulate in the endothelium to take up lipoproteins they become "foam-cells".[5] These structures become more and more extensive and turn into fibrous lesions, which trigger smooth muscle migration to the site. The lesions, together with the mass of smooth muscle tissue in the endothelium results in a fibrous cap. This can occlude the arteries and eventually rupture and create a thrombus which can subsequently obstruct the blood flow causing MI or stroke.[4] CAD is almost always a result of occluded blood vessels in the heart.[13, 33, 34] According to certain estimations, atherosclerosis causes approximately 50% of all deaths worldwide.[5]

Blood lipid levels and environmental susceptibility

The National Cholesterol Education Program suggests an increase in physical activity, nutritional and weight loss interventions to people with dyslipidemia based on results from observational and clinical trial studies.[18] Combined intensive lifestyle interventions proved to be successful in improving lipid and lipoprotein subfraction levels [35, 36] and other markers related to CVD.[37] There is moderate to reasonably strong evidence that physical activity interventions (e.g. aerobic exercise

training), some dietary interventions (e.g. the DASH diet) and lower intakes of saturated and trans-fats improve blood lipid profiles.[38] Although often assumed, there is little evidence that reducing dietary cholesterol intake lowers circulating blood lipid levels.[38] Smoking has a notable adverse effect on lipoprotein profiles, especially on HDL-C levels.[39] Blood lipid levels are also susceptible to pharmacologic interventions; for example, lipid-lowering agents, such as statins, fibrates, ezetimibe, and niacin are widely prescribed to people with dyslipidemia.[40]

In regards to primary prevention of coronary heart disease (CHD), apart from lipid-lowering medications, currently only lifestyle advice on smoking cessation, healthy diet and exercise is given.[41, 42] Some diets influence lipid levels, which subsequently exert CHD risk [43], but exactly how much diet contributes to dyslipidemia and CHD is debated – further studies are warranted to explore their relationships.

A brief history of lipid genetics

Blood lipid levels have a strong environmental underpinning. However, family and twin studies estimate that ~40-70% of the phenotypic variance in lipid traits may be explained by genetic factors.[44-46]

For decades, linkage studies (family-based genetic studies) have been extremely successful in identifying regions associated with single-gene Mendelian disorders. This method is based on the co-segregation of the disease-causing alleles with pre-selected micro-satellite marker alleles.[47, 48] Linkage analysis is a hypothesis generating approach, identifying multiple regions of the genome for further study. Linkage analysis, however, proved to be much less successful in the detection of causal variants in non-Mendelian diseases or traits with complex, polygenic etiologies due to the low penetrance of genes contributing to these diseases.[49] Subsequent candidate gene studies drew prior assumptions from linkage studies, animal studies or positional cloning for a given disease or trait.

The GWAS Era

Genome-wide association studies (GWAS) emerged from huge efforts invested in mapping the human genome in the 1990s and early 2000s [50, 51] and the large-scale sequencing efforts that followed, such as the International HapMap Project [52, 53] and the 1000 Genomes Project.[54, 55] Through the Human Genome Project, it was revealed that the entire human genome contains approximately 3 billion basepairs, 20,000-22,000 genes and these genes only occupy around 1-2% of the whole genome. Therefore, the remaining 98-99% of the DNA is considered non-coding. The shear length of the coding regions alone is 30 million basepairs (calculated as 1% of the 3 billion basepairs), and genotyping individuals for all these basepairs would be a prohibitively time-consuming and costly procedure. However, due to

recombination at certain spots in the genome at meiosis, larger regions of the genome are inherited together and therefore, markers in these regions are dependent on one another, which can be exploited to reduce the amount of the genome that needs to be mapped to define common variation. These regions are called haplotype blocks and the phenomenon when multiple basepairs are inherited together is termed linkage disequilibrium (LD).[56, 57] Thus, genotyping a few carefully selected genetic markers can facilitate the accurate imputation of all common markers in the given haplotype block. To do so requires a reference panel; such panels are publically available (such as HapMap and 1000 Genomes) and are often used for the imputation of common markers. In the late 1990s, Kruglyak postulated that at least 500,000 common variants (with minor allele frequency (MAF) > 5%) will be required to be genotyped to provide sufficient coverage for imputation in order to map all of the coding regions.[58] Although met with criticism (many were hoping for a much smaller number and 500,000 seemed overly pessimistic), his predictions were correct: current GWAS chips typically use >500,000 markers (e.g. Illumina HumanHap550 BeadChip or the Human1M BeadChip, which use ~550,000 and ~1 million markers, respectively). The first GWAS was published in 2005 [59] and the approach quickly became extremely popular. In contrast to candidate gene studies, GWAS were designed to scan the entire coding region of the genome in a so-called *hypothesis-free* manner. As with this approach, millions of markers are tested in statistical models for disease traits, test statistics need to be corrected for multiple testing. Assuming ~1 million independent markers, only results with $P<5\times10^{-8}$ (0.05/1,000,000) are usually considered statistically significant.[60] Since 2005, GWAS has been highly successful in identifying novel genetic loci associated with disease and other phenotypes.[49] By 2011, more than 1,200 GWAS have examined hundreds of thousands of participants and identified thousands of SNP-disease associations. Many of these SNPs later proved to be clinically relevant.[61, 62]

Common variants associated with lipids

Many GWAS have been undertaken in relation to lipids, some with comparatively low sample sizes. Although early GWAS by Saxena *et al* [63], Kooner *et al* [64], Kathiresan *et al* [65-67], Sandhu *et al* [68], Willer *et al* [69] and Heid *et al* [70] implicated a number of loci for individual lipid traits, the first comprehensive lipid-GWAS was published in 2008 by Aulchenko *et al*.[71] This study identified 22 loci in relation to TC, TG, LDL-C or HDL-C. A major breakthrough in lipid locus discovery came with a paper in 2010, from Teslovich *et al* (as part of the Global Lipids Genetics Consortium (GLGC)), within which 95 loci for the four core lipid traits were reported following a large-scale meta-analysis of GWAS data from 46 cohorts with a total sample size of >100,000 individuals.[72] In general, it became clear that by increasing sample size, many new genetic associations were detectable and many loci that had not met the conservative genome-wide significance threshold ($5\times10^{-8}<P<5\times10^{-7}$) in earlier studies[60] were replicated in this large meta-analysis.[72] The successes of the many GWAS consortia focused on cardiometabolic traits motivated

the design of a targeted genotyping array from Illumina named the CardioMetaboChip (in short, the MetaboChip array).[73] The MetaboChip is a custom array containing ~220,000 common variants and was designed for studies that would later **a)** follow-up on previous findings through fine-mapping of GWAS-established regions, **b)** seek replication of sub-significant GWAS loci, **c)** study disease pathways and facilitate other consortium-specific research questions. Over the following years, additional GWAS and MetaboChip data had accrued, prompting the formation of a new GLGC meta-analysis. The emerging results were reported in two papers by Willer *et al* and Do *et al* in which 62 novel variants for lipids were reported, extending the total number of established common lipid loci to 157 (Figure 2).[74, 75] Subsequent GWAS have discovered a few additional ethnic-specific variants; e.g. a novel variant in *ABCA6* was discovered in the Dutch population following a population-specific imputation of local GWAS studies.[76] Lu *et al* identified three Chinese-specific variants with genome-wide significance in already established lipid loci.[77]

Other lipid traits have been studied as well; in 2009, Chasman *et al* identified 43 loci in relation to lipoprotein subfractions.[78] Lemaitre *et al* discovered a number of loci, *FADS1*/*FADS2*, *ELOVL2* and *GCKR* in relation to circulating n-3 fatty acids in a GWAS study of ~9,000 adults of European ancestry.[79] In a subsequent study, Lemaitre *et al* studied circulating very long-chain saturated fatty acid concentrations and discovered two associated loci, *SPTLC3* and *CERS4*.[80] Moreover, Mozaffarian *et al* identified rs174548 at *FADS1*/*FADS2* in relation to cis/trans-18:2, a circulating trans-fatty acid biomarker.[81]

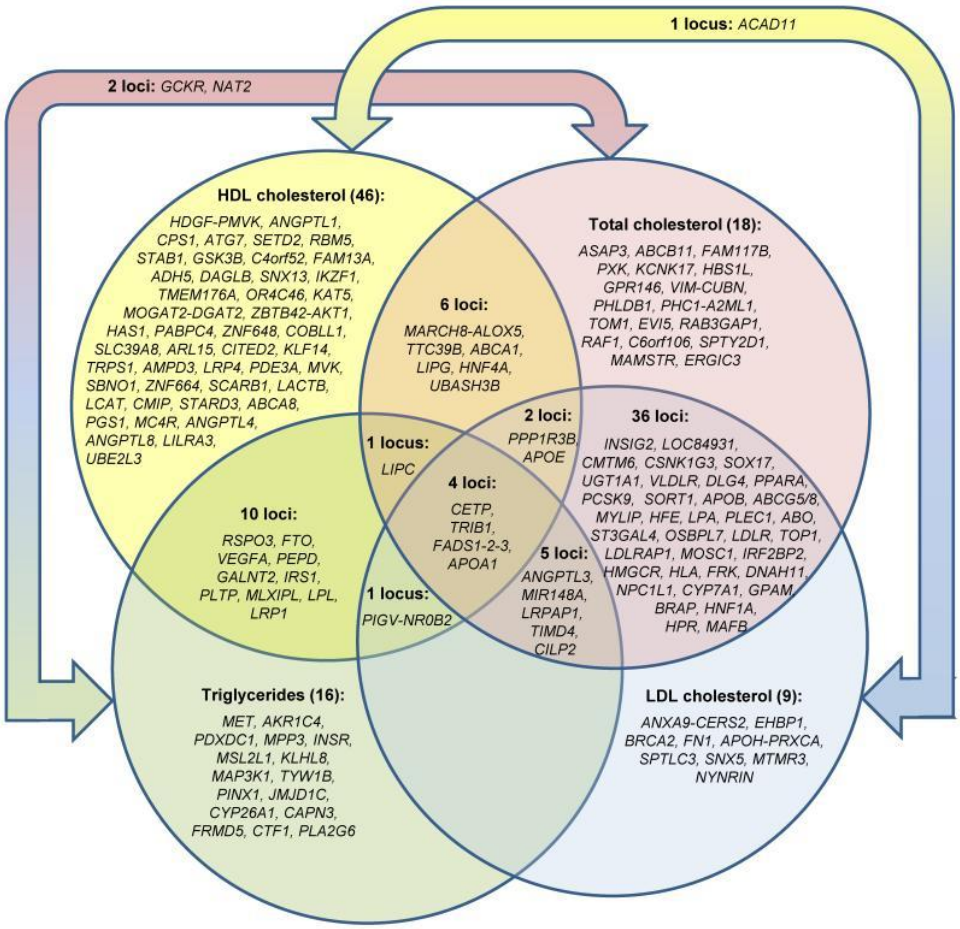

Figure 2. Dyslipidemia-associated loci by the Global Lipids Genetics Consortium.
HDL – high-density lipoprotein; LDL – low-density lipoprotein. Ref: Willer CJ *et al*. Discovery and refinement of loci associated with lipid levels. Nature Genetics. 2013;45:1274-1283.

As GWAS sample size increased, the more statistically powered these studies became to identify genetic loci with smaller and smaller effect sizes. These discoveries illustrate the complex, multifactorial nature of lipid traits, but also raise the question of whether common genetic variants with very small effect sizes are clinically relevant.

Indeed, as of today, almost 200 genetic loci are discovered in relation to lipid traits with an available sample size of ~200,000, but only a very small fraction of the postulated phenotypic trait variance (~10% of the genetic variance, ~5% of total trait variance) is explained by GWAS-established loci.[74] This is not unique for blood lipid traits, as loci discovered for other traits like BMI, for example, explain less than 5% of the BMI variance in the biggest GWAS study published to date.[82]

Rare variants associated with lipids

The question I encountered the most throughout my PhD studies was "Where is the missing heritability?" for common complex traits.[83] As rare genetic disorders are usually caused by highly penetrant rare variants (MAF<1%), another frequently heard question has been whether common disease would be caused by common or rare variants.[84] As the first waves of GWAS since 2005 and on have identified novel loci after novel loci, the scientific community leaned towards the *common disease – common variant* hypothesis, but as GWAS-identified variants were unable to account for the expected phenotypic variance explained by genetics, new approaches started to emerge to study the *common disease – rare variant* hypothesis.[85] These approaches involve whole-exome and whole-genome sequencing and new, array-based methods, such as the Exome Chip, which is a custom array designed to study rare variation in the exonic regions of genes.[86]

A whole-exome sequencing study by Albrechtsen *et al* identified a low-frequency (MAF=0.035) missense variant (a SNP causing the substitution of an amino acid in the translated protein) at *CD300LG* (R82C) for fasting HDL-C. In a subsequent meta-analysis of the discovery and replication cohorts in the same study, three SNPs were associated with lower levels of HDL-C, including the non-synonymous polymorphism at *CD300LG* discovered earlier.[87] Lange *et al* conducted an exome sequencing study in ~3,000 white and black adults, including ~800 individuals with extreme LDL-C concentrations. One SNP, rs1160983 at *TOMM40* was significantly associated with LDL-C.[88] Futema *et al* exome sequenced patients with familial hypercholesterolemia in pursuit of rare variants without any novel penetrant variant being discovered.[89] Peloso *et al* studied >50,000 white and black adults with Exome Chip data available. Four low-frequency variants, *ANGPTL8* rs145464906, *PAFAH1B2* rs186808413, *COL18A1* rs114139997 and *PCSK7* rs142953140, with large effect estimates associated with either HDL-C or TG were discovered (e.g. *ANGPTL8* rs145464906 and *PCSK7* rs142953140 were associated with 15% and 30% decrease in TG levels, respectively). None of these SNPs was associated with CHD in any of the additional ethnicities studied.[90] Using the Exome Chop in >12,000 Chinese adults, Tang *et al* successfully identified three Chinese-specific rare variants at *CETP*, *PCSK9* and *LDLR* (established lipid loci) for the four main lipid traits. The authors also reported two missense variants at *PNPLA3* and *PKD1L3* that influence both LDL-C and TG.[91] Timpson *et al* discovered a novel rare variant at *APOC3* for circulating plasma TG and VLDL concentrations through whole-genome sequencing of >3,200 individuals.[92] A study by the UK10K Consortium identified novel alleles at *APOB* associated with TG, *ADIPOQ* for adiponectin and *LDLR* and *RGAG1* for LDL-C concentrations by whole-genome and whole-exome sequencing >10,000 UK adults.[93] Johansen *et al* demonstrated that both GWAS-identified and non-GWAS-identified genetic loci are enriched in rare hyperlipidemia-associated variants and the consideration of these variants increase the explained phenotypic variance for hyperlipidemia.[94, 95]

Rationale for prospective trait genetics

As outlined above, a large number of cross-sectional genetic association studies have been undertaken to study the genetic component of dyslipidemias. However, only few genetic studies have been carried out in prospective studies or clinical trials in relation to lipids.

As suggested by Kurbasic *et al* [96], some loci might demonstrate age-dependent effects (Figure 3) and these loci might not be detectable in cross-sectional GWAS meta-analyses owing to the age heterogeneity of the cohorts. Genetic effects might differ by age due to a number of factors including changes in genetic penetrance (the proportion of individuals that carry a certain genetic variation that also express a given phenotype might change with age), age-related changes in gene expression and interactions with cumulative environmental or lifestyle factors (e.g. environmental pollutants or nutritional intakes).[97] This hypothesis is supported by Dumitrescu *et al*, who demonstrated significant heterogeneity in the association between variants at the *FADS1* locus and LDL-C in white males aged less vs. more than 50 years, with a significant association observed in older, but not younger participants.[98] A twin-study by Middelberg *et al* and earlier related studies from the same research group also conclude that different loci may affect various lipid traits at different ages and these studies also suggests that cross-sectional genetic association studies might underestimate heritability for lipid traits.[99-102]

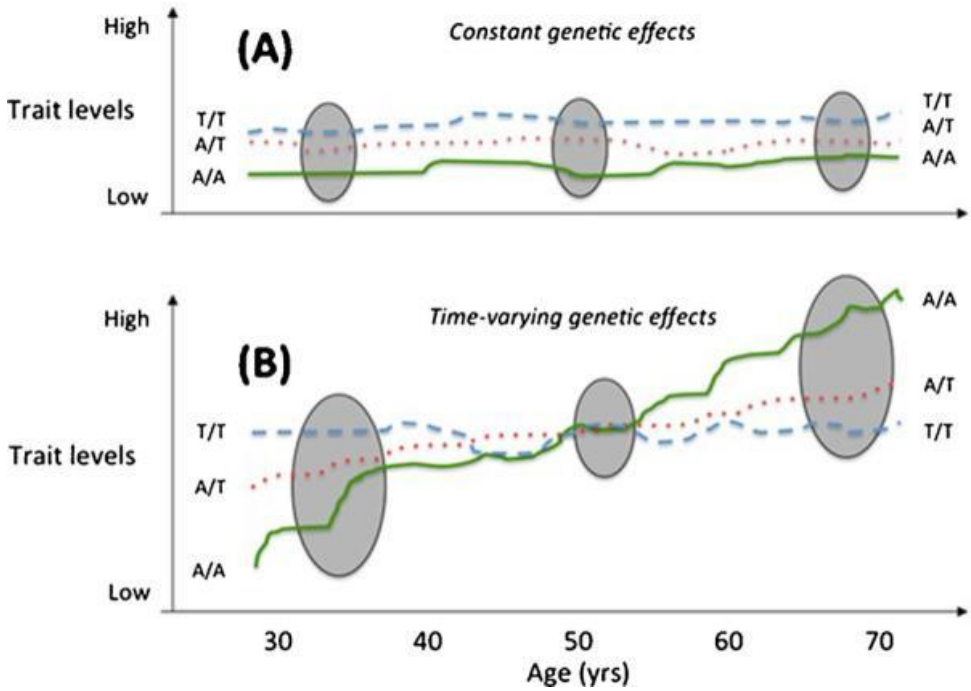

Figure 3. Constant and time-dependent genetic effects.
Time-varying genetic effects may be undetectable in a cross-sectional analysis. The locus shown in Panel A does not demonstrate time-dependent genetic effects (its effect is constant with age). For the locus presented in Panel B, cross-sectional studies in younger adults (age between 30 and 40) would conclude that the AA genotype is protective, even though the same genotype is associated with higher risk in the elderly (age around 70). Ref: Kurbasic A *et al*. Gene-lifestyle interactions in complex diseases: Design and description of the GLACIER and VIKING studies. Current Nutrition Reports. 2014;3:400-411.

Genetic associations for lipid changes

Lutsey *et al* studied the 95 GLGC-established lipid loci [72] in the prospective ARIC Study. In their study, a TG-specific genetic risk score (GRS) associated with increase in TG over a 9-yr follow-up period, but the TC-, HDL-C- and LDL-C-related GRSs were not correlated with respective trait changes.[97] Lu *et al* examined 243 genomic loci in cholesterol metabolism-related pathways in a prospective study with an 11-yr follow-up. In this study, a TC GRS was created from 23 TC-associated variants. This GRS was associated with change in TC levels from baseline to follow-up.[103] Constanza *et al* examined 20 SNPs in 13 lipid-associated loci in cross-sectional and longitudinal settings in ~2,000 Swiss adults with a 6-yr follow-up. While their hypothesis was that both the cross-sectional and the prospective analyses would yield similar findings, they observed only moderate consistency between cross-sectional and longitudinal findings. This result suggests different underlying biological mechanisms for lipid levels and lipid level changes for some lipid loci.[104] Huang *et al* studied >3,000 black and white adults from the CARDIA

Study over a 20-yr follow-up period. Although they report baseline LDL-C differences by *PCSK9* genotype, this study did not demonstrate evidence for longitudinal genetic associations.[105] Webster *et al* studied three candidate loci, *APOA5*, *LPL* and *GCK*, for lipid level changes in >2,500 white adults over an average follow-up of 17-yrs. Although the authors did not observe striking differences between baseline (cross-sectional) and prospective genetic associations, few associations tended to change with increasing age.[106] Lu *et al* examined changes in lipid traits in their latest GWAS in a Chinese population. All of their strong findings in relation to lipid changes also show cross-sectional associations for baseline lipid traits.[77]

Although statistical methods improve, sample sizes grow, lipid assays become more detailed and accurate, and more genetic studies are published on cardiometabolic traits than ever before, the whereabouts of the "missing heritability" remains unclear. One plausible explanation is that early family and twin studies might have overestimated the heritable basis of common diseases.[107] On the other hand, as mentioned above, some prospective studies suggest the opposite is true and state that cross-sectional studies, in fact, underestimate heritability.[101] GWAS-identified variants ($P<5\times10^{-8}$) explain less than 10% of the genetic variance for lipid levels. However, in a study by Yang *et al* in 2010, 50 GWAS-established variants explained only 5% of the phenotypic variance in height, but using 250,000 common variants explained 45%.[108] This observation shows that multiple yet undetected genetic loci are in fact associated with complex traits and can partially account for the unexplained trait variance. Fine-mapping established loci has also yielded significant findings, shedding further light on biological function and adding to the explained phenotypic variance.[109, 110]

Gene x environment interactions

Another phenomenon contributing to the missing heritability could be that certain genetic effects can also be modified by other genetic variants or environmental factors.[107] These are termed gene × gene and gene × environment/lifestyle interactions, respectively. Gene × environment interactions could be interpreted as a given genetic variant's effect differing across an environmental exposure or similarly, a given environmental or lifestyle exposure exerting a different effect in individuals with genotypic differences.[111] This concept is shown in Figure 4; in this hypothetical scenario, among carriers of the *aa* genotype the environmental exposure demonstrate no associations with the outcome trait, whereas among the carriers of the *AA* genotype, environment has a strong association with the outcome trait. To look at the same example from another perspective, among those who are less exposed to an environmental exposure (Environmental exposure = 1 or 2), the *aa* genotype is associated with increased levels of the outcome trait compared to the other two genotypes (*Aa* and *AA*). Among those who are more exposed to an environmental exposure (Environmental exposure = 4 or 5), the same *aa* carriers

have lower levels of the outcome trait than carriers of the other two genotypes (*Aa* and *AA*).

Examples exist in the literature where gene × environment interactions explain comparable phenotypic variance for common disease as genetic factors alone. Zheng *et al* reported genome-wide interaction studies where the interactions between environmental factors, such as carbohydrate and polyunsaturated fat intake and all the genetic loci on their array explained >20% of phenotypic variance in various quantitative glycemic traits and type 2 diabetes.[112, 113]

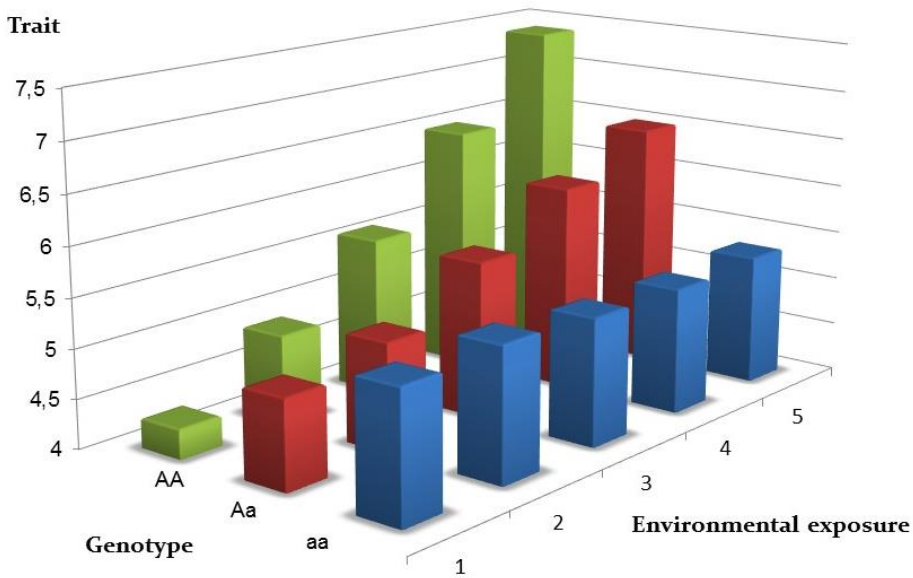

Figure 4. The concept of gene × environment interactions
Gene × environment interactions can be defined as the genetic effect on a trait that differ in magnitude across environmental exposures or environmental effects on traits that differ in magnitude across genotypes of a given genetic locus. Ref: Franks PW. Gene x environment interactions in type 2 diabetes. Curr Diab Rep. 2011;11:552-561.

Examples of gene x environment interactions in lipids

Justesen *et al* discovered an interaction between the TG-specific GRS and anthropometric traits, BMI and waist circumference on fasting TG levels in a meta-analysis of two Danish cohorts. In their study, the TG-specific GRS was more strongly associated with TG concentrations in metabolically unhealthy participants compared with metabolically healthy, normal weight participants. In addition to this, an interaction between the HDL-C-specific GRS and physical activity was

observed in relation to HDL-C levels. The GRS in this case exerted a smaller effect in more physically active participants than in those who reported being inactive.[114] Loria-Kohen *et al* report an interaction between the *PPARA* rs135549 SNP and milk consumption in relation to the reduction of TC/HDL-C and LDL-C/HDL-C ratios in an intervention study of 161 adults. Among skimmed milk drinkers, an improved lipid profile was apparent in the TT genotype carriers after the intervention, but no association was observed in the other genotypes.[115] In a project studying lipid levels in an Inuit population, Rudkowska *et al* found interactions between 11 GWAS-established SNPs and total fat / saturated fat intake on at least one plasma lipid biomarker. Genetic effects were generally less pronounced when self-reported intakes of total and saturated dietary fats were high.[116] Stojkovic *et al* identified a suggestive interaction between a missense mutation at *PNPLA3* with adiposity. At this locus, the G allele only associated with lower TG concentrations in overweight individuals.[117] Lu *et al* observed interactions between the rs174546 variant at the *FADS1*/*FADS2* gene cluster and polyunsaturated fat intake on non-HDL-C and HDL-C concentrations. The genotype only showed associations with non-HDL-C in those with high intakes of n-3 fatty acids and only showed associations with HDL-C in those with high intakes of n-6 fatty acids.[118] Takkunen *et al* also identified interactions between *FADS1*/*FADS2* genetic variants and polyunsaturated fatty acid intake on various circulating long chain fatty acids, although none of the interaction statistics survived correction for multiple testing.[119] Xu *et al* examined >700 overweight or obese adults in a diet intervention trial and found that dietary fat intake modified the effect of the *LIPC* rs2070895 variant on changes in TC, LDL-C and HDL-C. In individuals assigned to the low-fat diet, the A allele of the variant was associated with decreased TC and LDL-C levels, while the same allele was associated with increased levels of these traits in the high-fat diet group.[120] In a study by Smith *et al*, physical inactivity (measured by the time spent watching television or using a computer) modified the association of the *LIPG* variant rs6507931 on HDL-C levels; in sedentary C allele carriers HDL-related traits were higher and LDL-related traits were lower compared to TT homozygotes.[121] Abellán *et al* reported an interaction between rs4148102 at the *ABCG1* locus and polyunsaturated fat intake on LDL-C concentrations in two Spanish cohorts. In their study, LDL-C concentrations appeared to be elevated only in participants carrying the AA genotype consuming a high level of dietary fat.[122] Several studies examining *PCSK9*, *LIPG* and *PPARD* (established lipid loci) reported on gene × alcohol intake interactions in serum lipid levels.[123-125] Junyent *et al* identified multiple interactions between SNPs at the *ABCG5*/*ABCG8* gene complex and smoking habits on HDL-C levels in >800 Puerto Ricans.[126]

Corella *et al* identified an interaction between the rs13702 variant at *LPL* and monounsaturated and polyunsaturated fat intake on baseline TG levels in the PREDIMED clinical trial. After three years, the same variant was associated with a greater decrease in TG, but only in those who consumed the Mediterranean diet high in unsaturated fats.[127] In another clinical trial, the Diabetes Prevention Program (DPP), Pollin *et al* examined the associations of a global (non-trait-specific) lipid

GRS on blood lipid and lipoprotein subfraction concentrations and interaction effects across treatment arms. The GRS was associated with higher LDL-C and small LDL particle concentrations in the lifestyle intervention arm, but not in the placebo control arm, which suggests that those with high genetic burden respond to lifestyle interventions less than those with a lower genetic burden to higher lipid levels.[36]

Sex-heterogeneity in gene × environment interactions

In studies where sex modifies genetic effects, sex is the "environmental" component in the gene × environment interaction model. The study by Justesen *et al* mentioned above displayed strong sex heterogeneity; the interaction reported was only apparent in women and not men.[114] In another study by Brondani *et al*, the rs1746661 SNP at *FNDC5* was associated with higher levels of TC, LDL-C and lower levels of HDL-C in women but not men with type 2 diabetes.[128] Taylor *et al* examined 49 GWAS-established lipid SNPs and observed statistically significant sex heterogeneity in two SNPs, rs28927680 and rs3135506 at the *APOA1*/*APOC3*/*APOA4*/*APOA5*/*BUD13* gene complex for TG (larger effect estimates in males), rs7679 at *PLTP* for HDL-C (larger effect estimates in females) and rs12654264 at *HMGCR* for LDL-C (larger effect estimates in males).[129] Five of 17 loci previously demonstrating sex heterogeneity in large-scale meta-analyses [71, 72, 130] also replicated in this study.

Gene × lipid-lowering medication interactions in lipids

Examples where genetic variations modify responses to lipid-lowering medications have also been described. Barber *et al* identified variants at *APOE*, *CLMN* and *PCSK9* that are associated with LDL-C response to various statin therapies in a combined GWAS of >4,000 individuals.[131] Thompson *et al* identified SNPs at *APOE*, *PCSK9* and *HMGCR* that associated with LDL-C response to atorvastatin therapy in the TNT trial.[132] Deshmukh *et al* conducted a GWAS on LDL-C response to a 10-mg atorvastatin therapy. The rs10455872 at *LPA* and rs445925 and rs4420638 at *APOE* strongly associated with LDL-C response to medication, indicating that these variants modulate the efficacy of statin therapy.[133] Hu *et al* examined the genotypic factors that might influence response to niacin treatment in Chinese patients; *DGAT2* rs3060 showed suggestive evidence for modulating LDL-C response.[134] A similar interaction was observed in an earlier study in relation to TG response to niacin therapy.[135]

Genetic loci and their role in lipid metabolism

The landmark GLGC publications by Teslovich *et al* and Willer *et al* report on the biological and clinical relevance of the loci identified in their papers.[72, 74]

Several loci that were robustly associated with lipid levels in GWAS have well-known biological functions, but it is often unknown how specific SNPs at these loci impact lipid metabolism. In addition, several lead SNPs discovered by GWAS lie outside genes in noncoding regions or in intronic, non-expressed regions. This implies that many of the GWAS-identified variants influence gene regulation perhaps via epigenomic mechanisms.

The most studied genetic loci

An extensively researched locus is *LPL*, which codes for lipoprotein lipase and *ANGPTL3* and *ANGPTL4*, which are lipoprotein lipase inhibitors. Lipoprotein lipase hydrolyzes TG in the chylomicrons and VLDL particles, and also mediates uptake of lipoproteins in endothelial, muscle and fat cells.[136] *LDLR* encodes the LDL receptor and the protein product of *LDLRAP1* interacts with the LDL receptor to remove cholesterol from the bloodstream by internalizing LDL particles.[72] In the hepatic cells, lipoproteins are broken down and free cholesterol is released. Protein products of *LRP1*, *LRP2* and *LRP4* also mediate LDL uptake. *SCARB1* codes for a specific HDL receptor, which mediates the selective uptake of HDL particles in the liver. This uptake is crucial for the reverse cholesterol transport. *ABCA1* codes for a transmembrane protein called ATP-binding cassette transporter ABCA1. This transporter protein is a very important regulator of cellular cholesterol transport; it is responsible for transporting cholesterol to lipid-poor lipoproteins and thereby facilitating their transformation to nascent HDL.[137] Other loci coding for ATP-binding cassette proteins, such as ABCG5/8 and ABCA8, have been identified in GWAS for lipid levels.

Genetic variants at *APOA1*/*APOC3*/*APOA4*/*APOA5*/*BUD13*, *APOB*, and *APOE/APOC1* have shown very strong associations with multiple lipids traits.[72] These loci code for various apolipoproteins strongly associated with certain lipoprotein subclasses. *APOB* encodes apolipoprotein B, the primary apolipoprotein of chylomicrons, VLDL, IDL and LDL.[137] *APOE* encodes apolipoprotein E, which is found in chylomicrons and IDL and its main function is to transport cholesterol into the lymph system and the blood.[137] *APOC1* encodes apolipoprotein C1, which exchanges cholesterol between lipoproteins, removes cholesterol from tissues and is harbored mainly on HDL particles. *APOA1* encodes apolipoprotein A1, which is the main apolipoprotein of HDL particles. This protein, in conjunction with the protein product of *LCAT* (also a known lipid locus) promotes fat and cholesterol removal from adipocytes, myocytes and the endothelial cells and facilitates their excretion in the liver.[137] *APOC3* encodes apolipoprotein C3, which is the constituent of VLDL particles. It inhibits endothelial and hepatic lipases (encoded by *LIPG* and *LIPC*, respectively – both GWAS-established loci for lipids) and hepatic uptake of fats, thereby increasing blood TG concentrations.[137] *APOA4* codes for apolipoprotein A4, which is expressed on chylomicrons. Its function is not yet known, but intestinal fat absorption increases apolipoprotein A4 levels. *APOA5* encodes apolipoprotein A5, which can be found on chylomicrons, VLDL and HDL

particles and is an important determinant of TG levels. Lipoprotein(a) is a separate, highly atherogenic lipoprotein class containing an apolipoprotein(a) molecule bound to an apolipoprotein B-100. The *LPA* locus affects lipoprotein(a) concentrations and is associated with lipid levels and CHD.[137]

SORT1 encodes the sortilin 1 protein, which is responsible for the degradation of nascent VLDL in the liver, thereby regulating the level of their release in the bloodstream. *VLDLR* codes for the VLDL receptor, which is expressed in several tissues in the body, most importantly, in the brain, heart and skeletal muscle and adipose tissues. It is not expressed, however, in the liver. Its function is to facilitate TG uptake in these tissues from chylomicrons and VLDL particles.[74] *FTO* is a known locus for obesity and type 2 diabetes, and in the second GLGC meta-analysis, variants at this locus were independently associated with HDL-C and TG.[74, 82]

Lipid loci and CAD

Some established lipid loci are highly relevant from a clinical perspective, as in addition to lipid levels, they also showed robust associations with CAD. In the second GLGC GWAS, 40 lipid loci reached nominal significance for CAD.[74] These include *IRS1*, *C6orf106*, *KLF14*, *NAT2*, *RBM5* and *CMTM6*, some of which were originally only associated with HDL-C concentrations. Although circulating blood lipid levels (TG, LDL-C and HDL-C) had been consistently associated with risk of CHD and atherosclerosis, the causal nature of these observations have been debated. Recent Mendelian Randomization studies suggest that LDL-C [138], TG [139], TG-rich lipoprotein and lipoprotein(a) concentrations [140] are causally related to CHD and MI. On the other hand, Mendelian Randomization studies and pharmacological evidence from recent clinical trials suggest that HDL-C and some other previously implicated factors such as C-reactive protein are not on the causal pathway to CHD. [138, 141-145] Whilst it is possible that the loci mentioned above have pleiotropic effects on HDL-C and CAD (or related traits, such as type 2 diabetes or insulin resistance), it is possible that these loci affect CAD risk through their intermediate effects on HDL-C levels.

Genetic loci and lipid lowering medications

Dyslipidemias are usually treated with lipid-lowering medications. Statins are considered the frontline therapy in treating patients with high lipid levels.[146] Other classes of lipid medications, such as fibrates, niacin [40] and ezetimibe [147] are also often used in dyslipidemia. Recently, beneficial effect of gastric bypass surgery on lipid levels were reported.[148]

Some lipid medications, such as statins, ezetimibe, torcetrapib and PCSK9 inhibitors mimic mutations in known lipid loci, *HMGCR*, *NPC1L1*, *CETP* and *PCSK9*, respectively.

HMGCR encodes the HMG-CoA reductase enzyme, which controls a rate-limiting step in the mevalonate pathway [21] (converting HMG-CoA to mevalonic acid). As internally produced cholesterols are downstream from this step in the mevalonate pathway, inhibiting the HMG-CoA reductase enzyme by statin drugs results in decreased cholesterol synthesis and therefore, lower concentrations of circulating cholesterol.[149] Statins are the most prescribed lipid lowering agents as they have proven extremely effective in LDL-C lowering.[150]

NPC1L1 encodes the Niemann-Pick C1-like protein 1, which is expressed in the liver and in the gastrointestinal tract. The inhibition of this protein by ezetimibe decreases cholesterol uptake in the intestinal cells and thereby lowers blood cholesterol concentrations.[147] Recently, rare inactivating mutations have been found at the *NPC1L1* locus. These mutations proved to be associated with reduced plasma LDL-C levels and CHD.[151]

PCSK9 encodes the proprotein convertase subtilisin/kexin type 9 (PCSK9) enzyme, which binds to LDL receptors in the liver and degrades them before they are able to clear LDL particles from the circulating blood.[152] PCSK9 inhibitors are a novel class of drugs currently undergoing large randomized clinical trials to prove efficacy and safety.[152] PCSK9 inhibitors are extremely important, as their discovery was dependent on genetic discoveries; rare loss-of-function mutations at the *PCSK9* locus were detected and found to be associated with a decreased risk for CAD [153], These findings spurred subsequent interest for pharmacologic intervention on the protein product of the gene.

The *CETP* locus has one of the strongest associations (both in terms of effect size and *P* value) for lipids ever discovered; it encodes the cholesteryl ester transfer protein (CETP), which facilitates cholesterol exchange between various lipoproteins. As high HDL-C levels are associated with a decreased risk for atherosclerosis, CHD and MI, novel drugs aiming to increase HDL-C levels by CETP inhibition have emerged. The drug torcetrapib was developed and tested in a large randomized clinical trial. The trial was stopped due to unexpected adverse cardiovascular outcomes (raise in blood pressure) in the torcetrapib arm.[142, 154] Currently, novel classes of CETP inhibitors (e.g. anaceptrapib [40]) are undergoing clinical trials to discover whether increasing HDL-C is a futile strategy or not.[155, 156]

The first commercially available gene therapy in the world also relates to genetic discoveries in relation to lipid levels. The Glybera® gene therapy aims to treat patients with a rare mutation in the *LPL* gene that causes extreme high levels of TG and abnormally large lipid droplets in the blood, resulting in myocardial infarctions early in young adulthood. This ground-breaking new therapeutic approach will comprise of ~20 intramuscular injections with an adenovirus vector carrying *LPL*.[157]

As reviewed by Preiss *et al* and Rached *et al*, several new trials and clinical experiments of dyslipidemia drugs are currently underway [40, 156], some of which are inspired by genetic findings.

As more than 150 GWAS loci have been established so far, there are certainly many potential targets for pharmaceutical interventions. However, only a small fraction of these loci's etiology is fully unraveled and understood. In order to develop new targets for intervention, functional characterization of established lipid loci is much needed. Furthermore, new genetic studies with more detailed and accurate genotype and phenotype information are warranted to discover additional variation in relation to dyslipidemia.

Aims

Identifying causal environmental and genetic determinants and gene × environment interactions for lipids is important in the prevention and treatment of dyslipidemias and subsequently, atherosclerosis and CVD. This thesis focuses on genetic associations in prospective cohort studies and detecting gene × environment interactions in observational cohort studies and clinical trials. The goal of the papers presented here are primarily hypothesis generation; thus, further studies are needed to elucidate the mechanisms of these discoveries.

The overarching aims of this thesis are to

1. familiarize the reader with the concept of prospective trait genetics where the outcomes are trait changes, not cross-sectional measures of association
2. explore the influence of gene × environment interactions on lipid levels in population-based studies and randomized controlled trials

Paper I – this paper reports analyses of the 157 currently established lipid-associated loci (both as single variants and aggregated into trait-specific genetic risk scores) in relation to 10-yr changes in fasting total cholesterol and triglyceride levels.

Paper II – this paper extends Paper I by examining of a larger set of SNPs. Replication studies, *in silico* look-ups in international consortia and functional annotations were carried out to provide additional levels of evidence for our findings.

Paper III – in this paper, a previously reported gene × environment interaction between a triglyceride-associated genetic risk score and BMI on blood triglyceride concentrations was explicitly tested and replicated in a cross-sectional setting. A meta-analysis of four cohort studies was conducted and protein-protein interaction analysis was carried out.

Paper IV – in this paper, the analyses focused on assessing gene × treatment (intensive lifestyle intervention and metformin treatment) interactions for loci with established association signals for a range of lipids and lipoproteins within a randomized controlled trial (the Diabetes Prevention Program).

Materials and methods

Study populations

The analyses described herein were undertaken in several different study populations. We used the Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk (GLACIER) Study as the discovery cohort for **Papers I** and **II** and as a replication cohort for **Paper III**. In **Papers I-III** we used the Malmö Diet and Cancer (MDC) Study for replication. In **Paper II**, in addition to the MDC Study, we also utilized the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS), Uppsala Longitudinal Study of Adult Men (ULSAM), Medical Research Council (MRC) Ely studies for replication. In **Paper III**, we meta-analyzed the results from the GLACIER and MDC studies with the Inter99 and Health2006 studies. We used the Diabetes Prevention Program (DPP) randomized controlled clinical trial (RCT) for **Paper IV**. Materials and Methods describes the GLACIER Study and the DPP RCT in detail, as these two studies were central to this PhD work. The section under the subheading *Study populations* / *Other studies* contains a brief description of all the other studies involved in **Papers I-III**. An overview of the studies involved in **Papers I-IV** is presented in Table 1. Project flowcharts for **Papers I**, **II**, **III** and **IV** are shown in Figures 5, 6, 7 and 8, respectively. For a more detailed description, clinical characteristics, genotyping information and statistical methods of these studies please see the manuscripts or the references included in the respective sections.

Table 1. Studies in Papers I-IV.

	Paper I	**Paper II**	**Paper III**	**Paper IV**
Discovery phase	GLACIER*	GLACIER*	Inter99 Health2006	DPP*
Replication phase	MDC	MDC PIVUS ULSAM MRC Ely	GLACIER* MDC	-
Meta-analysis	GLACIER* MDC*	GLACIER* MDC* PIVUS* ULSAM* MRC Ely*	Inter99 Health2006 GLACIER MDC	-

*analysis performed primarily by Tibor V. Varga

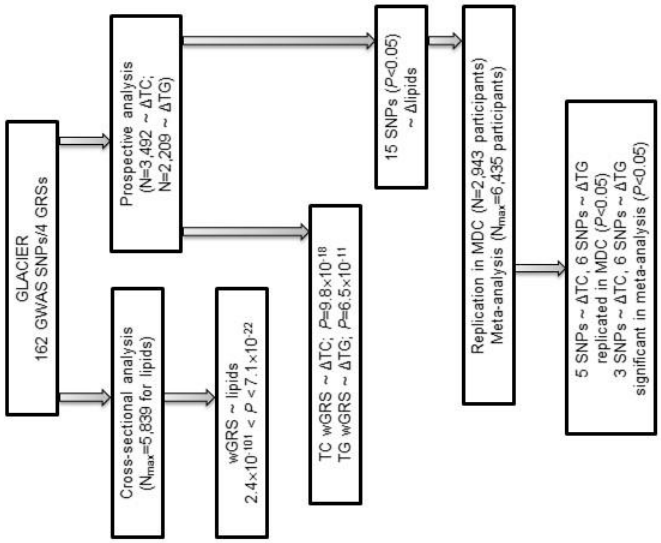

Figure 5. Project flowchart of Paper I.
Δ – trait change; GRS – genetic risk score; GWAS – genome-wide association study; SNP – single nucleotide polymorphism; TC – total cholesterol; TG – triglyceride, wGRS – weighted genetic risk score.

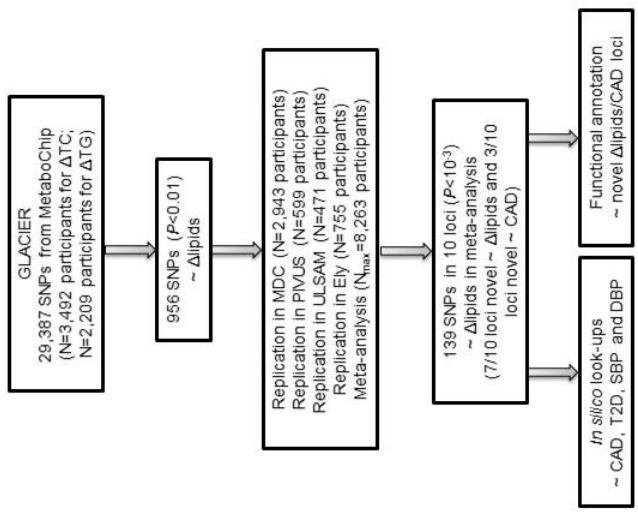

Figure 6. Project flowchart of Paper II.
CAD – coronary artery disease; Δ – trait change; DBP – diastolic blood pressure; SBP – systolic blood pressure; SNP – single nucleotide polymorphism; T2D – type 2 diabetes; TC – total cholesterol; TG – triglyceride.

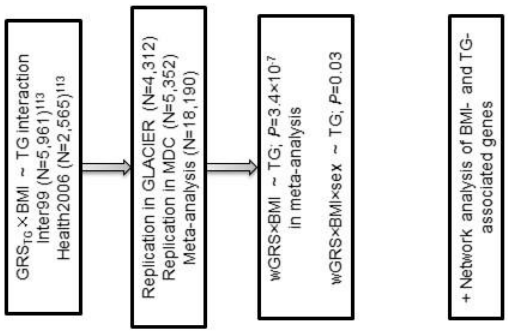

Figure 7. Project flowchart of Paper III.
BMI – body mass index; GRS – genetic risk score; TG – triglyceride; wGRS – weighted genetic risk score.

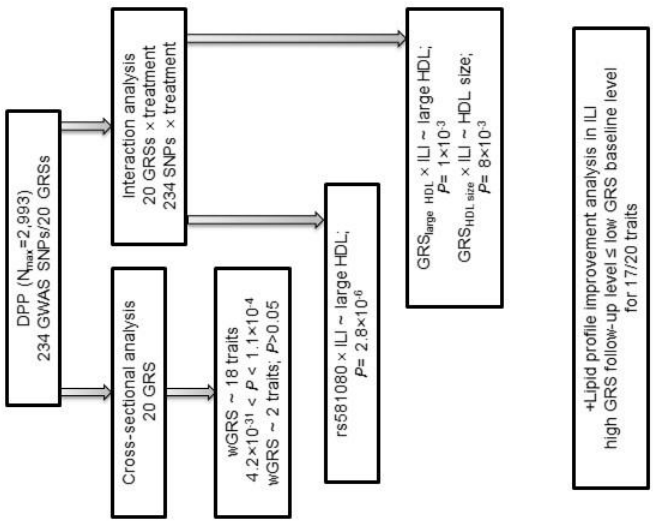

Figure 8. Project flowchart of Paper IV.
GWAS – genome-wide association study; GRS – genetic risk score; HDL – high-density lipoprotein; ILI – intensive lifestyle intervention; SNP – single nucleotide polymorphism; wGRS – weighted genetic risk score.

The following section provides an overview of each of the study cohorts.

GLACIER

The GLACIER Study is nested in the Västerbotten Health Survey (VHU, also known as the Västerbotten Intervention Programme (VIP)).[158] As mortality from CVD in the northern Swedish county of Västerbotten was the highest in Sweden in the 1980s, VHU was launched as a community-based intervention in 1985. Participants, first from Norsjö, then from several cities in Västerbotten, were invited to their primary health care facility for a detailed health screening and lifestyle counselling in their 40th, 50th and 60th year of birth. At each visit, blood samples were drawn after an overnight fast; these samples are stored at the Umeå University Medical Biobank. A wide-range of biomarkers is available, including basic anthropometric measures, systolic and diastolic blood pressure, fasting and post-challenge glucose levels, obtained by a 75g oral glucose tolerance test and fasting blood lipid levels. In addition to clinical characteristics , lifestyle was ascertained by detailed questionnaires, including an 86-item food frequency questionnaire (FFQ). Information about leisure-time and occupational physical activity, level of education and quality of life were also collected.

The GLACIER Study is nested within this background population and data resource. GLACIER is comprised of ~19,000 adults, from which we use smaller subsamples in our studies as a result of analysis-specific inclusion and exclusion criteria, partly due driven by to insufficient genotype or phenotype information.[96] Currently, ~5,000 GLACIER participants have one follow-up visit (with an average follow-up period of ~10 years). A subset of the GLACIER Study (~6,000 participants) is genotyped with the Illumina Cardio-MetaboChip custom array.[73] All GLACIER participants gave informed consents and the study was approved by the Regional Ethical Review Board in Umeå.

DPP

The DPP is a multi-centered, multi-ethnic (White, African American, Hispanic, Asian and American Indian by self-report) RCT based in the USA.[159, 160] The primary outcome of the trial is diabetes incidence confirmed by an oral glucose tolerance test (OGTT). A total of ~4,000 non-diabetic, overweight/obese participants at high risk of developing type 2 diabetes were recruited based on age, BMI and fasting and post-challenge glucose concentrations. Eligible study participants are characterized by:

- age $\geq$ 25 years
- BMI above ethnicity-specific thresholds (e.g. BMI $\geq$ 24 kg/m^2 in Whites)
- fasting plasma glucose < 7.0 mmol/l

- 7.8 mmol/l ≤ 2-h post-load plasma glucose < 11.1 mmol/l.

Participants were randomized into four arms: intensive lifestyle intervention (ILI), metformin, troglitazone and standard of care (placebo). In the ILI arm participants were recommended to conduct moderate-intensity physical activity for 150 minutes/week and to maintain a healthy diet to achieve a weight loss goal of 7% during the trial period. They were given intensive coaching in diet, physical activity and behavior modification.[161] Participants in the metformin arm were prescribed 850 mg metformin pills twice a day. Participants in the troglitazone arm were prescribed 400 mg of troglitazone pills once a day. All arms, including the comparison arm received standard of care including written information and yearly short individual sessions on physical activity and diet. After two years of the initiation of the trial, the troglitazone arm was discontinued due to liver toxicity concerns.[162] The remaining three arms comprise ~3,000 participants available for analyses. A wide range of biomarkers is available in the DPP including glycaemic traits and fasting blood lipid levels. The DPP is genotyped with the Illumina Cardio-MetaboChip custom array.[73] All participants gave written informed consent and ethical approvals were issued at the respective DPP study centers (approved by their respective Institutional Review Boards).

Other studies

The **MDC** Study is a population-based prospective cohort study from the southern Swedish county of Skåne.[163, 164] During 1991-1996, women born between 1923 and 1945 and men born between 1923 and 1950 living in the city of Malmö were invited to participate. A total of ~3,000 participants with no history of coronary events had relevant follow-up data available for replication analyses in the prospective trait genetics projects. Basic anthropometric measures were collected by trained nurses and fasting blood was obtained. Lipid concentrations were measured by a DAX 48 automatic analyzer (Bayer AB, Göteborg, Sweden). MDC study participants were genotyped using the Illumina HumanOmniExpress BeadChip v.1 and subsequently imputed using the 1000G (March 2012) reference sequences using IMPUTE2.[54, 165] Genotyping was performed at the Broad Institute, Cambridge, MA USA. All participants provided written informed consent and the study was approved by the Research Ethics Committee at Lund University.

The **PIVUS** study includes ~2,000 randomly selected 70-year-olds living in Uppsala County, Sweden, who were invited for a baseline visit between 2001 and 2004.[166] A follow-up examination was performed at the age of 75 years. A total of 599 participants had data available for replication analysis in our project. Blood samples were collected after an overnight fast. Participants were genotyped with the Illumina Cardio-Metabochip array.[73]

The **ULSAM** study comprises ~1,000 50-year-old men living in Uppsala County, Sweden. Baseline data were collected between 1970 and 1974.[167] Men were re-examined at ages 71 and 77. A total of 471 participants had complete baseline and

follow-up data for our analyses. Blood samples were collected after an overnight fast. Participants were genotyped with the Illumina Cardio-Metabochip custom genotyping array.[73]

The **MRC Ely** Study is a UK-based prospective cohort study focused on studying the etiology of type 2 diabetes.[168] Baseline data was collected between 1990 and 1992. Approximately 1,000 adults underwent a standard 75 g OGTT and detailed screening for related cardiovascular and metabolic risk factors. All participants were invited for identical screening measurements ~10 years later. A total of 755 participants had data available for the analyses described in this thesis (Paper II). Blood samples were collected after an overnight fast. Participants were genotyped with the Illumina Cardio-Metabochip array.[73]

The **Inter99** and **Health2006** studies served as a basis for the replication reported in Paper III. As we only used extracted baseline summary statistics and did not use raw data, we only provide a very brief description about these studies. Inter99 is a randomized nonpharmacological intervention study for the prevention of ischemic heart disease.[114, 169] At baseline, 30–60 years old participants were examined at the Research Centre for Prevention and Health in Glostrup, Denmark. The Inter99 study was approved by The Copenhagen County Ethical Committee and the National Board of Health. Health2006 is a population-based study examining general health and several cardiovascular and other metabolic diseases in 18-74 years old Danish adults.[114, 170] The Health2006 study was conducted at the Research Centre for Prevention and Health in Glostrup, Denmark. Both Inter99 and Health2006 participants were genotyped with the Illumina Cardio-Metabochip array.[73]

In Paper II, we conducted *in silico* look-ups for association summary statistics in publicly available datasets of international consortia. We looked-up effect estimates and *P* values in the Coronary ARtery DIsease Genome-wide Replication and Meta-analysis (CARDIoGRAMplusC4D) Consortium[171] (N_{max}~190,000) for SNP-coronary artery disease associations, the Global Lipids Genetics Consortium (GLGC)[74] (N_{max}~190,000) for SNP-lipid associations, the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium[172] (N_{max}~90,000) for SNP-type 2 diabetes associations, and the International Consortium for Blood Pressure (ICBP)[173] (N_{max}~200,000) for SNP-blood pressure associations.

Study-specific materials and methods

Clinical characteristics

In GLACIER (**Paper I-III**), capillary blood was drawn following an overnight fast. Approximately 5% of the study participants reported a shorter than ideal fasting period (<8 hours) before the blood draw, and information on fasting status was missing in a ~15% of the cohort; therefore, analyses were adjusted with three

dummy variables indicating fasting status (0/1 – fasting status unknown/known; 0/1 – fasted less than 4 hours; 0/1 – fasted between 4-8 hours), with participants fasted ≥8 hours (~80% of the the study participants) as reference. Serum lipid concentrations were measured from fresh capillary plasma using a Reflotron bench-top analyzer (Roche Diagnostics Scandinavia AB). Due to the sensitivity of the analyzer, all TG values below 0.8 mmol/l were set to missing. HDL-C was measured after precipitation of the other lipoproteins with sodium phosphowolframate-magnesium chloride. LDL-C concentrations were calculated with the Friedewald equation.[174] Approximately 1% of GLACIER participants reported using lipid-lowering medications (no information available on the specific type), which we controlled for in analyses using constants. At the time of the examinations the most common type of lipid lowering drugs in Västerbotten was HMG-CoA reductase inhibitors, also known as statins, used by ~96% of lipid lowering medication users.[175] Consequently, to correct blood lipid levels we used the statin constants proposed by Wu *et al*: [176] HDL-C: -0.059 mmol/l; LDL-C: +1.279 mmol/l, TC: +1.336 mmol/l, TG: +0.207 mmol/l. Using simulations and real-data validation, Tobin *et al* [177] reported that correction with constants among medication-users yields less biased estimates and is more powerful than adjusting with a binary variable or to exclude participants on medication. In cross-sectional analyses, TG values were transformed to the natural log scale prior to analyses to approximate a normal distribution in each cohort. BMI was calculated as weight (kg) / height (m^2).

In the DPP (**Paper IV**), fasting (>12 hours) venous blood was drawn. Measurements of TG, TC and HDL-C were made using enzymatic methods standardized to the Centers for Disease Control and Prevention reference methods.[178] HDL-C concentrations were obtained by precipitation of apolipoprotein B-containing lipoproteins by dextran sulfate Mg^{2+} treatment.[179] LDL-C concentrations were calculated with the Friedewald equation.[174] Where TG levels exceeded 4.5 mmol/l, lipoprotein fractions were separated using ultracentrifugation.[180] Nuclear magnetic resonance (NMR) spectroscopy (LipoScience Inc., Raleigh, NC) was used to quantify IDL-C and ApoB concentrations, VLDL particle numbers (total and small, medium and large subfractions), LDL particle numbers (total and small and large subfractions) and HDL particle numbers (total and small, medium and large subfractions) and their total particle sizes.[181] We conducted two parallel sets of analysis; first, traits were analyzed in their native distributions. Second, we inverse normalized all traits (mean=0, SD=1) and repeated the analyses to facilitate comparisons of effect estimates between traits. Individuals using lipid medications were excluded. Semiannual fasting plasma glucose measurements and annual 75g OGTT were performed. In our study we used baseline and 1-yr fasting and post-challenge glucose levels and incident type 2 diabetes status.

Lifestyle exposures

In **Papers I** and **III**, various lifestyle information was used. These were obtained using a validated questionnaire. In **Paper I**, alcohol intake was used as a percentage

of total energy intake and smoking was coded as 1, 2, 3 for current, former and never smokers, respectively. In **Paper III**, smoking was coded the same way as in **Paper I**. In **Paper III**, leisure time physical activity was coded as exercise performed with the following frequencies: 1 – never, 2 – sometimes, 3 - once per week, 4 - 2-3 times a week, 5 - more than 3 times a week; these categories were directly obtained from a single question related to leisure time physical activity in the FFQ. A diet variable was calculated by conducting principal component analysis (PCA). This method uses between food-intake correlation measures to identify underlying patterns in the data.[182] Intake of eight macronutrients (carbohydrate, sugar, protein, saturated fat, total fat, fiber, monounsaturated fatty acid and polyunsaturated fatty acid) was used to calculate the diet principal components. The first component principal component was used in the analysis.

Genotype data

In GLACIER (**Paper I-III**), DNA was extracted from peripheral white blood cells and genomic DNA samples were diluted to 4 ng/µl.[183, 184] Participants were genotyped by the Illlumina Cardio-MetaboChip array.[73] In **Paper I**, we studied previously reported genome-wide significant ($P<5\times10^{-8}$) SNPs for TC, TG, HDL-C and LDL-C levels. Teslovich *et al* reported 95 genomic loci,[72] and a later meta-analysis by Willer *et al* reported an additional 62 loci for these four traits.[74] We used highly correlated proxy SNPs ($r^2>0.8$) for 18 variants. We were unable to find suitable proxies for two SNPs; these were excluded from the analysis. We conducted analyses with 162 SNPs in total - 75 SNPs for TC; 43 SNPs for TG, 73 SNPs for HDL-C and 58 SNPs for LDL-C (including SNPs with multiple associations, e.g. rs1532085 in *LIPC* associates with TC, TG and HDL-C). The average genotyping success rate was 99.9%. None of the SNPs deviated significantly from Hardy-Weinberg equilibrium at a study-wise corrected type I error level ($\alpha=0.0001$). For this study, we also created trait-specific genetic risk scores (GRS) to study the combined effects of the individual SNPs. For each participant, we first created unweighted GRSs by summing the number of risk alleles (trait elevating alleles for TC, TG and LDL-C and trait decreasing alleles for HDL-C; e.g. the risk allele for rs1532085 in *LIPC* is A for TC, but G for HDL-C. For a participant with the GG genotype, this SNP contributes a score of 0 to the TC GRS, but 2 to the HDL-C GRS. A heterozygous genotype of AG will contribute 1 risk allele to both GRSs). The theoretical minimum values for the unweighted GRSs is 0 for all traits, while the theoretical maximum values are 150, 86, 146 and 116 for the TC-, TG-, HDL-C- and LDL-C-associated GRSs, respectively. In a subsequent step, we weighted the GRSs (wGRS) by multiplying effect allele counts with published effect sizes reported by Willer *et al*.[74] To make results easier to compare, we transformed the scale of the wGRSs (maximum values depend on the reported effect sizes and number of SNPs) to the original scale of the unweighted GRS (same maximum values as mentioned above).[185] The wGRSs are created as shown in Equation 1 –

SNPs denote the allele count for a given SNP and ESs denote published effect sizes for each SNPs.

Equation 1. $wGRS = [(ES_1 * SNP_1) + (ES_2 * SNP_2) + \cdots + (ES_n * SNP_n)] * (n/\sum_n^1 ES)$

For each SNP, missing genotype information was imputed by assigning average genotype values to missing values in order to keep allele frequencies constant.[186]

In **Paper II** we extended our analyses to a wider range of SNPs from the Illumina Cardio-MetaboChip array. We selected all the previously reported genome-wide significant SNPs (as in **Paper I**), *Replication SNPs*, ~5,000 of the most statistically significant, but sub-genome-wide significant SNPs ($P<5\times10^{-8}$), *Fine-mapping SNPs* around the genome-wide significant SNPs and *Wildcard SNPs* (put on the array for consortia-specific reasons, namely for deep sequencing efforts or from a particular pathway of interest) for the four lipid traits.[73] In total, we extracted 43,690 SNPs from the array for the discovery phase of the project. Rare variants (MAF<1%; n=14,280 SNPs) and SNPs not in Hardy-Weinberg equilibrium (α=0.0001; n=23 SNPs) were excluded. In total, 29,387 SNPs were analyzed in the discovery phase in GLACIER.

In **Paper III**, we analyzed the 43 TG-related SNPs and the TG-related GRS and wGRS, which are described in detail in **Paper I** above.

In the DPP (**Paper IV**), DNA was extracted from peripheral white blood cells. We selected previously reported genome-wide significant SNPs for TC (71/75 SNPs available), TG (37/43 SNPs available), HDL-C (68/73 SNPs available), LDL-C (54/58 SNPs available) at 157 genomic loci[74] and 91 lipoprotein subfraction associated SNPs at 43 genomic loci[78] from recent GWAS meta-analyses. All together, we used 234 SNPs in our analyses – these were extracted from the Illumina Cardio-MetaboChip genotyping array.[73] During quality control (QC) procedures, participants with failed genotyping, gender inconsistency, or cryptic familial relatedness were excluded. During QC, Hardy-Weinberg equilibrium ($\alpha=10^{-7}$) was assessed in each ethnic group separately. Where the index SNP was not available, we used a highly correlated proxy ($r^2>0.8$ in the respective HapMap population). The genotyping success rate for the 234 SNPs was 99.6%.

Statistical methods

Statistical analyses were undertaken predominantly using STATA (version 12.1 and subsequently 13.1, StataCorp LP, TX, USA) and PLINK (version 1.07).[187, 188] The GRSs for **Papers I** and **III** were created using SAS (version 9.3, SAS Institute Inc., NC, USA).[189] Analysis for **Paper III** was done using R (version 3.2.2, The R

Foundation for Statistical Computing).[190] We used a range of tools for meta-analysis (reported below).

Regression models

In the section below, I include equations to facilitate a better understanding of the regression models used. In these equations, the dependent variable is on the left side of the equal sign (=), while the independent variables are on the right side of the equal sign.

In **Paper I**, we conducted cross-sectional analyses investigating the GRSs' main effects on baseline lipid traits. In these models we used generalized linear models (GLM) by fitting the lipid measures as dependent variables and the GRSs as independent variables. In these models, we adjusted for age, age^2, sex, fasting variables as explained above, and population substructure using the first four genomic principal components (PC) (Equation 2).

Equation 2. $baseline\ lipid = \alpha + \beta_{GRS} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{fasting} + \beta_{PC1-4} + \varepsilon$

The main focus of **Papers I** and **II**, were the longitudinal models using prospective data. In these models we fitted the follow-up lipid measures as dependent variables and the SNPs/GRSs (these genetic components are denoted as G in Equation 3) as independent variables. We adjusted these models for age, age^2, follow-up age, sex, fasting time the first four PCs and the traits' respective baseline value (Equation 3). We included both age and follow-up age to account for inter-individual follow-up time differences.

Equation 3. $follow-up\ lipid = \alpha + \beta_G + \beta_{age} + \beta_{age^2} + \beta_{follow-up\ age} + \beta_{sex} + \beta_{fasting} + \beta_{PC1-4} + \beta_{baseline\ lipid} + \varepsilon$

In **Papers I** and **II**, we conducted meta-analyses to obtain summary effect estimates for the longitudinal models undertaken in the separate studies. As follow-up times varied between studies, we created a synchronized outcome for the meta-analysis, which we termed average annual lipid change (AALC). We computed this variable by substracting the baseline level from the follow-up level for each trait and dividing the delta by the follow-up period (in years). (Equation 4).

Equation 4. $AALC = \frac{(follow-up\ lipid)-(baseline\ lipid)}{follow-up\ period\ (yr)}$

The analyses in GLACIER and all replication analyses were repeated using AALC as the outcome in the GLMs (Equation 5) and summary statistics from these models were subsequently meta-analyzed. Note that only thing that differs between Equation 4 and Equation 5 is the definition of the dependent variable.

Equation 5. $AALC = \alpha + \beta_G + \beta_{age} + \beta_{age^2} + \beta_{follow-up\ age} + \beta_{sex} + \beta_{fasting} + \beta_{PC1-4} + \beta_{baseline\ lipid} + \varepsilon$

In **Paper III**, we used similar GLMs as in the cross-sectional models (Equation 2). In these models we fitted TG as the dependent variable and the wGRS as the independent variable. We adjusted the model for age, age^2, sex, fasting time, BMI and the first four PCs (Equation 6).

Equation 6. $TG = \alpha + \beta_{wGRS} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{fasting} + \beta_{BMI} + \beta_{PC1-4} + \varepsilon$

Models were repeated by including a BMI × wGRS interaction term (Equation 7).

Equation 7. $TG = \alpha + \beta_{BMI\times wGRS} + \beta_{wGRS} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{fasting} + \beta_{BMI} + \beta_{PC1-4} + \varepsilon$

In additional models, we adjusted the interaction model for lifestyle factors including education level, leisure time physical activity, smoking, and total energy intake (Equation 8).

Equation 8. $TG = \alpha + \beta_{BMI\times wGRS} + \beta_{wGRS} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{fasting} + \beta_{BMI} + \beta_{PC1-4} + \beta_{lifestyle} + \varepsilon$

To test whether these lifestyle factors interact with BMI, we also tested lifestyle × wGRS interactions (Equation 9).

Equation 9. $TG = \alpha + \beta_{BMI\times wGRS} + \beta_{wGRS} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{fasting} + \beta_{BMI} + \beta_{PC1-4} + \beta_{lifestyle} + \beta_{lifestyle\times wGRS} + \varepsilon$

We further tested three-way interactions between the wGRS, BMI and sex (Equation 10).

Equation 10. $TG = \alpha + \beta_{BMI \times wGRS \times sex} + \beta_{wGRS} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{fasting} + \beta_{BMI} + \beta_{PC1-4} + \varepsilon$

In **Paper IV**, we began by conducting cross-sectional analyses using baseline data. We assessed individual SNPs and trait-specific GRSs (these genetic components are denoted as G in Equation 11 and 12) in relation to their respective baseline traits. In these analyses we adjusted for age, age^2, sex and the first four genomic PCs (Equation 11).

Equation 11. $baseline\ lipid = \alpha + \beta_G + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{PC1-4} + \varepsilon$

We then tested whether the SNPs or the GRSs modify treatment effects at follow-up. We fitted SNP/GRS × intensive lifestyle intervention and SNP/GRS × metformin intervention interaction terms as independent variables with age, age^2, sex, dummy variables for ILI (1/0) and metformin interventions (1/0), the genetic component, the baseline lipid value and the first four genomic PCs (Equation 12).

Equation 12. $follow-up\ lipid = \alpha + \beta_{G \times ILI} + \beta_{G \times metformin} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{ILI} + \beta_{metformin} + \beta_G + \beta_{baseline\ lipid} + \beta_{PC1-4} + \varepsilon$

For all lipid-associated GRSs, we assessed genetic associations at baseline and ILI/metformin treatment interactions at follow-up in relation to glycemic traits (fasting glucose and post-challenge glucose) in parallel models to those conducted for the lipid traits (Equation 11 for baseline and Equation 12 for interactions at follow-up). We also conducted GRS analyses for type 2 diabetes incidence using Cox proportional hazards models (Equation 13).

Equation 13. $h(t|x) = h_0(t) \times \exp(\beta_{G \times ILI} + \beta_{G \times metformin} + \beta_{age} + \beta_{age^2} + \beta_{sex} + \beta_{ILI} + \beta_{metformin} + \beta_G + \beta_{PC1-4})$

Meta-analysis

In **Paper I**, we conducted random-effects meta-analysis. We used the *metan* module in STATA.[191]

In **Paper II** we conducted random-effects and fixed-effect meta-analysis using GWAMA.[192] We expected high heterogeneity between the studies, mainly due to different baseline age, follow-up time, fasting status, medication usage. However, heterogeneity (assessed by Cochran's *P* values and I^2 statistics as suggested by Higgins *et al*)[193] was low in the random-effects models, therefore we report estimates from the fixed-effect models.

In **Paper III**, cohort-specific estimates were meta-analyzed using the *metagen* and *forest* functions of the *meta* package in R.[194] Both random-effects and fixed-effect meta-analyses were conducted. Between-study effect estimates were very similar (low heterogeneity). Therefore, estimates were reported from the fixed-effect models.

Multiple testing correction

Multiple testing can result in alpha inflation and type I error. There are several methods to correct for multiple testing ranging from highly conservative methods (e.g. Bonferroni correction) to less conservative ones (e.g. permutations).[195, 196]

In **Paper I**, Benjamini-Hochberg false discovery rate (FDR) was used to correct for multiple testing.[197] In **Paper II**, instead of correcting for multiple testing, all results with $P<0.01$ in the discovery phase in the GLACIER Study were selected for replication. After the replication and the meta-analysis, we functionally annotated and further examined loci reaching $P<10^{-3}$ in the fixed-effect meta-analysis. In **Paper III**, Benjamini-Hochberg FDR was applied to correct for multiple testing.[197] In **Paper IV**, no multiple testing corrections were applied in the baseline analyses. In the SNP/GRS interaction analysis at follow-up, Bonferroni correction was applied, where the nominal $\alpha=0.05$ was divided by the number of statistical tests done.[198]

Other study-specific statistical methods

In **Papers I** and **II,** trait differences between baseline and follow-up were assessed by paired samples t-tests. In **Paper IV**, differences between subgroups were assessed by independent samples t-tests. In these analyses we compared participants at high and low genetic risk (defined by the GRS); To graphically illustrate results, we stratified the cohort into participants with high genetic risk vs. low genetic risk for a given trait based on median GRS values.

In **Papers I** and **IV**, we assessed the proportion of the variance in the trait explained by the GRSs. We obtained these estimates from linear regression models where we

fitted the lipid traits as dependent variables and the respective GRSs as independent variables. In these models, no other covariates were fitted.

In **Paper I**, we assessed predictive accuracy by using receiver operating curve (ROC) area under the curve (AUC) analyses.[199] For the these analyses, lipid levels were dichotomized (low/high) according to the American Heart Association criteria.[200] We compared 4 models:

- M1: age, age^2, sex and BMI
- M2: M1 + trait specific wGRS
- M3: M1 + traditional risk factors for hyperlipidemia reported by Lu *et al* (age, sex, BMI, smoking status, alcohol intake)[103]
- M4: M1 + trait specific wGRS + traditional risk factors for hyperlipidemia

In **Paper II**, we attempted to functionally annotate the loci with the strongest evidence for association in relation to lipid level changes and coronary artery disease. We characterized the regulatory potential of these loci by examining how they overlap with evolutionarily conserved elements determined using the GERP algorithm.[201] Using data on DNaseI hypersensitive sites (DHS) in 125 cell lines generated by The ENCODE Project[202] we determined the frequencies with which DHS-sites overlap with candidate loci. We used ChIP-seq data, also from the ENCODE Project to assess binding capacity of various regulatory factors to the loci in question. Furthermore, we used genome-wide chromatin state maps from nine cell lines on histone modification patterns and binding of CCCTC-binding factors generated by Ernst *et al*.[203] Overlaps between genomic annotation tracks and SNPs were calculated using the *GenomicRanges* package in R.[204]

In **Paper III**, we undertook protein-protein network analyses. We used the Human Protein Reference Database (HRPD)[205] to extract relevant protein-protein interactions networks.[206] We connected genomic loci linked to GWAS-reported BMI- and TG-associated loci to the HPRD protein-protein interaction network. We extracted neighboring nodes from the mapped genes to generate a TG-BMI protein interaction network comprised of the genes associated with BMI, TG and their interacting partner genes. The TG-BMI interaction network consists of genes that directly connect the two phenotypes (directly interacting nodes) and those that interact through one intermediate node. Pathway enrichment analyses were performed on this sub-network using a Cytoscape plug-in designed for pathway analyses based on the REACTOME database.[207] In **Paper IV**, we undertook pathway analysis using the REACTOME database.[208]

In **Paper IV**, we calculated Pearson correlation coefficients between traits. SNPs and GRSs for the four main lipid traits (TC, TG, HDL-C and LDL-C) were analyzed in relation to their respective traits and any lipoprotein subfraction concentration/size with a $|r| \geq 0.5$.

In **Paper IV**, we conducted subgroup analysis by repeating all GRS analyses in self-reported white participants only.

Results and Discussion

In this section I give a short summary of the main findings of Papers I-IV. Although methods are explained in detail above and in the attached manuscripts, the main aims and analytic strategy for Papers I-IV is also presented here in the first paragraphs of the respective sections. Apart from presenting the results, I also discuss Papers I-IV by explaining the implications of the studies, putting them into the context of the existing scientific evidence on the subject and explaining the main strengths and limitations of the projects.

Paper I

In Paper I, I examined the published TC-, TG-, HDL-C and LDL-C-related SNPs in the northern Swedish GLACIER Study in cross-sectional and longitudinal settings. I assessed associations between the four main lipid traits and SNPs identified by GWAS for lipid levels. I undertook further analysis to investigate these SNPs' combined effects on the same lipid traits. I created trait-specific GRSs and wGRSs. The variants used in these analyses were initially identified in large-scale cross-sectional meta-analyses. I also conducted prediction analysis for developing hyperlipidemia from baseline to follow-up. In these analyses I assessed how much traditional risk factors (including smoking and alcohol intake) and the wGRS add to a simple model using only age, age^2, sex and BMI information in predicting hyperlipidemia at follow-up. This study was the first comprehensive assessment of all published lipid loci in a prospective analysis.

The previously associated GWAS SNPs explained 8.8%, 4.9%, 4.8% and 9.1% of the variance in baseline TC, TG, HDL-C and LDL-C, respectively. The trait-specific wGRSs explained 7.0%, 3.9%, 2.6% and 6.9% variance for baseline TC, TG, HDL-C and LDL-C, respectively. In the first lipid GWAS reported by Teslovich *et al*, GRSs for the main lipid traits explained ~12% trait variance.[72] The second GLGC GWAS of lipid traits reported 62 novel loci, but made only a modest addition to the explained trait variances (~2% for each of the four traits).[74] Previous studies reported trait variances explained by GRSs comprised of only a handful of established lipid-associated loci in the range of 5-7%.[71, 103, 209] In Lutsey *et al*'s study (a study similar to ours), GRSs (comprised of the 95 established lipid loci by Teslovich *et al*) explained 6.8%, 6.0%, 6.0% and 1.6% trait variances for TC, TG, LDL-C and HDL-C.[97] These numbers correspond well with the numbers obtained from the GLACIER Study.

In baseline, cross-sectional analyses, 48% (36/75) of the TC SNPs, 30% (13/43) of the TG SNPs, 14% (10/73) of the HDL-C SNPs and 33% (19/58) of the LDL-C SNPs were nominally associated ($P<0.05$) with their corresponding traits. The trait-specific GRSs and wGRSs showed robust associations with their respective baseline traits. As weights were obtained from the latest cross-sectional GLGC GWAS,[74] we anticipated that using the wGRSs would strengthen the associations. Indeed, as shown in Table 2, using the wGRS strengthened the associations in all cases, although the effect sizes remained similar in magnitude.

Table 2. Trait specific GRS associations with lipid traits at baseline

Trait	β (mmol/allele)	95% CI (mmol/allele)	SE	P_{GRS}	P_{wGRS}
TC	0.05	0.04, 0.06	0.003	3.01×10^{-67}	2.43×10^{-101}
TG	0.03	0.02, 0.04	0.003	4.58×10^{-23}	4.24×10^{-41}
HDL-C	-0.01	-0.012, -0.008	0.001	1.35×10^{-11}	7.06×10^{-22}
LDL-C	0.05	0.04, 0.06	0.005	2.44×10^{-28}	1.17×10^{-50}

β – effect size; CI – confidence interval; SE – standard error. *P* values are obtained from multiple linear regression models.

The main focus of this paper was on longitudinal analyses. We hypothesized that previously established variants would associate differently with lipid level changes than with baseline levels. Similarly, as weights for the lipid-loci were extracted from a cross-sectional GWAS,[74] we did not expect that the use of wGRSs instead of the GRSs would necessarily strengthen results. Due to different measurement techniques of HDL-C (HDL-C is obtained by dextran-sulfate precipitation of LDL-C and VLD-C from TC) and the fact that LDL-C is derived from measures of TC, TG and HDL-C using the Fridewald formula,[174] we only had sufficient data for TC and TG level changes in GLACIER. TC concentrations decreased between baseline and follow-up (ΔTC=-0.18±1.12 mmol/l; $P<0.0001$). No change in TG levels was observed (ΔTG=0.02±1 mmol/l; $P=0.32$). When presenting the SNP/GRS results for TC and TG level changes, we use the ΔTC and ΔTG expressions. However, in our regression models, the follow-up lipid measures were fitted as dependent variables and analyses were adjusted for baseline levels. This means that the dependent variables were not TC and TG level changes (ΔTC and ΔTG) *per se*. This approach accounts for baseline trait differences and considered an appropriate way to model trait changes in a prospective setting with only two time points.

The TC- and TG-specific GRSs robustly associated with their respective trait level changes (Figures 9A and 9B) during follow-up and using the wGRSs increased the strength and magnitude of the results as shown in Table 3.

Table 3. Trait specific GRS associations with lipid trait changes

Trait	β (mmol/l/decade follow-up)	95% CI	SE	P_{GRS}	P_{wGRS}
ΔTC	0.02	0.01, 0.03	0.003	2.0×10^{-11}	9.8×10^{-18}
ΔTG	0.02	0.01, 0.03	0.005	0.0005	6.5×10^{-11}

β – effect size; CI – confidence interval; SE – standard error. 95% CI and SE have the same units as the β. *P* values are obtained from multiple linear regression models.

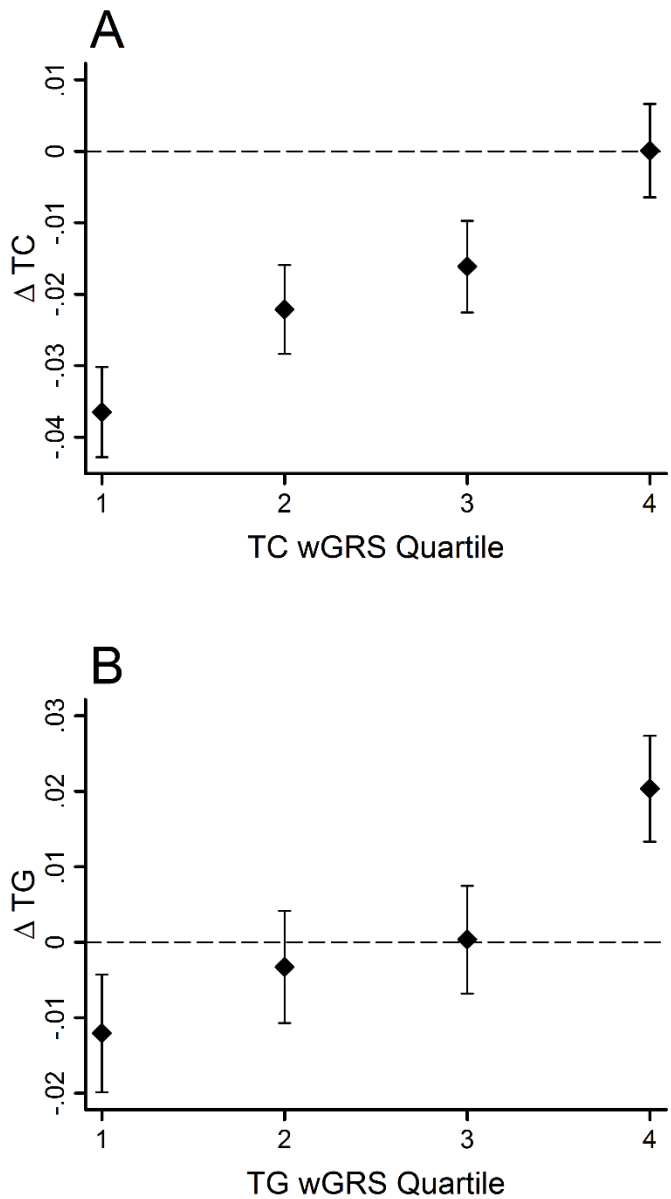

Figure 9. TC and TG level changes over 10-yr follow-up by wGRS quartiles.
Δ – trait change; TC – total cholesterol; TG – triglycerides; wGRS – weighted genetic risk score. Error bars represent 95% confidence intervals. Ref: Varga TV *et al*. Genetic determinants of long-term changes in blood lipid concentrations: 10-year follow-up of the glacier study. PLoS Genetics. 2014;10:e1004388.

In individual SNP analysis, FDR corrected statistically significant associations were observed for the rs6589564 in *APOA1* and ΔTG (β=0.31 mmol/l per allele per decade follow-up, 95% CI: 0.21, 0.41, SE=0.05, $P_{FDR}=6.6\times10^{-7}$), rs2954029 in *TRIB1* and ΔTG (β=0.09 mmol/l per allele per decade follow-up, 95% CI: 0.03, 0.15, SE=0.03, $P_{FDR}=0.009$) and rs4420638 in *APOE* and ΔTC (β=0.12 mmol/l per allele per decade follow-up, 95% CI: 0.06, 0.18, SE=0.03, $P_{FDR}=0.002$). In total, 15 variants associated with lipid trait changes at a level of nominal statistical significance (unadjusted $P<0.05$). These SNPs were carried forward for replication in the MDC Study. As mentioned above, we did not have data on HDL-C and LDL-C changes in GLACIER, but these variables were available in the MDC Study. Therefore, the 15 top hits in the discovery part of the analyses were tested in relation to these traits as well. Associations for five SNPs for ΔTC and six SNPs for ΔTG were nominally statistically significant ($P<0.05$) and directionally consistent with GLACIER results in MDC. Four SNPs also associated with ΔLDL-C. None of the SNPs associated with ΔHDL-C. All three variants robustly associated with lipid level changes in GLACIER in the discovery phase (rs6589564 in *APOA1* and rs2954029 in *TRIB1* for ΔTG; rs4420638 in *APOE* for ΔTC) replicated in MDC. In random-effects meta-analysis, three ΔTC associated variants and six ΔTG associated variants yielded statistically nominally significant ($P<0.05$) effect estimates. Our three most robust results, *TRIB1*, *APOA1* and *APOE* are all well-known dyslipidemia-related loci with many studies published on them. Apart from the associations with lipid levels, *TRIB1* also associates with CAD, ischemic heart disease and MI.[66, 69, 210, 211] *APOE* lies in the *APOE-APOC1-APOC2* gene cluster and it is also associated with familial dyslipoproteinemia, polygenic dyslipidemia, plasma C-reactive protein levels, CAD and MI.[67, 212-215] *APOA1* lies in the *APOA1-APOC3-APOA4-APOA5-BUD13* gene cluster and it is also associated with polygenic dyslipidemia, metabolic syndrome, CAD and MI.[67, 216-218]

Prediction algorithms may contain unmodifiable risk factors (e.g. age, sex, genetic predisposition, family history), and modifiable risk factors (e.g. dietary intakes, levels of exercise). It is important to denote that factors in risk prediction models do not need to be causal for the outcome. We assessed differences in predictive accuracy for developing hyperlipidemia by comparing four models. The simplest one contained basic information about age, age^2, sex and BMI (which is obtained by basic anthropometric measures) (M1). We compared this model to three other models including the basic information + wGRS (M2), basic information + traditional risk factors for hyperlipidemia (M3) and basic information + wGRS + traditional risk factors for hyperlipidemia (M4). The differences between M1 and M4 were 4% (62% for M1 - 66% for M4; $P=0.011$) for TC and 2% (65% for M1 - 67% for M4; $P=0.052$) for TG. We conclude that the predictive ability of these loci in the GLACIER Study is low, although comparable with an earlier study by Lu *et al* (3% AUC difference for TC), who used a GRS comprised of 12 TC-associated loci. Genetic factors have not yet proved to add significantly to existing risk prediction models.[219] More precise measurements and better models are warranted to more accurately predict dyslipidemia.

Although there are few studies reporting on genetic associations for prospective traits,[97, 103, 104, 138] none has comprehensively assessed all published dyslipidemia-associated loci in a longitudinal setting. In a previous study, Lutsey *et al* [97] studied 95 dyslipidemia-related loci identified by Teslovich *et al*.[72] They examined whether trait-specific genetic risk scores for TC, TG, LDL-C and HDL-C associate with 9-year changes of lipid levels in the ARIC Study. Although higher genetic risk scores (adverse genetic predisposition) for all traits were associated with a higher incidence of dyslipidemia at follow-up, only the TG-specific GRS was statistically significantly associated with lipid level change in their study. In our study (with a similar baseline age and follow-up time), GRSs associated both with TG- and TC level changes. An important strength of our study is the inclusion of replication data from the MDC Study. Our findings from the GLACIER Study were corroborated by results from the MDC Study and together they provide evidence for a small number of loci associating with lipid level changes.

Paper II

In Paper II, I extended the work done in Paper I [220] by assessing a wider range of loci. We hypothesized that SNPs in large-scale GWAS which almost reached a genome-wide level of statistical significance ($P=5\times10^{-8}$) might not have reached it due to time-dependent effects and the age-heterogeneity of the participating cohorts the results were obtained from.[96] The Illumina Cardio-MetaboChip array was constructed after the first wave of GWAS studies and a large number (~5000) of these "almost genome-wide significant" sets of SNPs (termed *Replication SNPs*) for the most studied cardio-metabolic traits were put on this array. Other SNPs on the array include *Fine-mapping SNPs* for established loci and *Wildcard SNPs* for consortium-specific purposes.[73] In total, we extracted 29,387 *Replication*/*Fine-mapping*/*Wildcard* SNPs for TC, TG, HDL-C and LDL-C and analyzed these in the discovery stage in GLACIER. This set of SNPs was not pruned for LD. We selected all SNPs having an association with TC or TG level changes with $P<0.01$ for replication in four European prospective studies and meta-analyzed estimates. We then selected all loci with $P<10^{-3}$ in the fixed-effect meta-analysis and further examined these loci by doing *in silico* look-ups in publicly available results from international consortia[74, 171-173] and conducted functional annotations.

In the discovery stage of the analysis, a few SNPs reached genome-wide significance in relation to ΔTG. These included a single low frequency (MAF<5%) SNP at *APOE* (chr19:50121999; β=0.64 mmol/l/10-yr; SE=0.11; $P=1.7\times10^{-8}$) and a number of SNPs in LD at the *APOA1-APOC3-APOA4-APOA5-BUD13* gene cluster (top SNP rs9326246, β=0.32 mmol/l/10-yr; SE=0.05; $P=4.4\times10^{-9}$). SNPs at *DOCK7* (chr1:62714800; β=0.59 mmol/l/10-yr; SE=0.11; $P=9.6\times10^{-8}$), *BRE* (chr2:28165690; β=0.59 mmol/l/10-yr; SE=0.12; $P=6.7\times10^{-7}$), *KCNIP1* (rs10041010; β=0.54 mmol/l/10-yr; SE=0.11; $P=8.7\times10^{-7}$) and *SYNE1* (rs594522;

β=0.55 mmol/l/10-yr; SE=0.11; $P=1.4\times10^{-6}$) reached study-wide significance ($\alpha=0.05/29{,}387=1.7\times10^{-6}$) for ΔTG. The rs7412 variant in *APOE* (β=-0.23 mmol/l/10-yr; SE=0.04; $P=1.8\times10^{-7}$) reached study-wide significance for ΔTC. In total, from the 29,387 tested SNPs, 478 and 490 SNPs associated ($P<0.01$) with ΔTC and ΔTG, respectively (in total, 956 variants, as 12 variants associated both with ΔTC and ΔTG). As the original set of SNPs was not pruned, these SNPs represent dependent signals (e.g. multiple associations of correlated SNPs in an LD block in relation to ΔTC or ΔTG). We looked-up association statistics of these 956 variants in relation to TC and TG levels in the GLGC publicly available cross-sectional data for lipid traits. Approximately 30% of the associations (273/956) had no cross-sectional associations with TC/TG levels (148 variants associated with TG, 183 variants associated with TC and 351 variants associated with both traits $P<5\times10^{-8}$). These results show that while many variants convey both cross-sectional and longitudinal associations in relation to lipid levels, most cross-sectional signals do not demonstrate prospective trait genetic associations, and a fraction of longitudinal signals do not associate with lipid traits cross-sectionally (although loci from this latter group have not been robustly replicated in the following replication analysis).

In the replication phase (meta-analysis of replication cohorts without GLACIER), the variant at rs7412 at *APOE* replicated for ΔTC ($P=1.\ 5\times10^{-7}$) and multiple signals at the *APOA1-APOC3-APOA4-APOA5-BUD13* gene cluster replicated for ΔTG ($P_{min}=6.5\times10^{-7}$). In these analyses, 28% of the SNPs replicated for ΔTC (135/478 SNPs), while 13% of the SNPs replicated for ΔTG (66/490 SNPs) at a nominal level of statistical significance ($P<0.05$).

In the meta-analysis including all cohorts (GLACIER and all replication cohorts) 88 SNPs at six loci for ΔTC and 51 SNPs at five loci for ΔTG (ten loci in total, as the *DOCK7-ANGPTL3* complex associated with both traits) replicated ($P<10^{-3}$); these loci are shown in Table 4.

Table 4. Lead SNPs from the meta-analysis of the five studies in Paper II. (N_{max}=8,263)

Trait	SNP	LOCUS	β (mmol/l/yr)	SE	*P*	I^2
ΔTC	chr1:62912318	*DOCK7/ANGPTL3*	-0.004	0.001	2.5×10^{-4}	0.43
ΔTC	rs13282247	*LPL/CSGALNACT1*	-0.004	0.001	1.7×10^{-4}	0.45
ΔTC	chr15:40469997	*CAPN3*	0.015	0.004	1.5×10^{-4}	0.08
ΔTC	chr16:70636491	*HP/HPR*	0.008	0.002	2.3×10^{-4}	0.56
ΔTC	rs17304534	*LDLR*	-0.004	0.001	1.0×10^{-3}	0.53
ΔTC	rs7412	*APOE/C1*	-0.014	0.002	2.0×10^{-12}	0.38
ΔTG	chr1:62954723	*DOCK7/ANGPTL3*	-0.002	0.001	5.2×10^{-4}	0.17
ΔTG	rs2954029	*TRIB1*	-0.003	0.001	8.1×10^{-6}	0.53
ΔTG	rs651821	*ZNF159/APOA5/A1*	0.009	0.001	1.4×10^{-10}	0.83
ΔTG	rs10406431	*GIPR*	-0.003	0.001	1.7×10^{-4}	0.27
ΔTG	rs11668847	*QPCTL*	0.003	0.001	2.7×10^{-5}	0.54

β – effect size; CI – confidence interval; SE – standard error; SNP – single nucleotide polymorphism. SE has the same units as the β. I^2 values denote heterogeneity estimates in the meta-analysis. *P* values are obtained from fixed-effect meta-analysis of the five studies.

All these associations remained nominally statistically significant ($P<0.05$) after the removal of the GLACIER Study from the meta-analysis. With the inclusion of all five cohorts, the three most significant loci from Varga *et al* (Paper I),[220] *TRIB1*, *APOE* and *APOA1* were replicated with a statistically more significant *P* value.

We further examined these ten loci by conducting *in silico* look-ups in publicly available databases (CARDIoGRAMplusC4D for CAD,[171] DIAGRAM for T2D[172] and ICBP for SBP and DBP[173]). We used published, publicly available data from international consortia presenting results from large-scale cross-sectional meta-analyses, identifying a locus with a *P* value below 5×10^{-8} would obviously mean that this locus is already published and not a novel finding. Hence, our study had the caveat of not being able to demonstrate novel, extremely robust evidence for associations with CAD, T2D, SBP or DBP. Therefore, we were primarily interested in sub-GWAS significant loci where we hypothesized that this level of evidence will be strengthened by the original association with lipid level changes and potentially, other *in silico* findings (e.g. results from a different consortium or results from the functional annotation). To be clearer, we hypothesized that a given SNP with a $P\sim10^{-4}$ for TG changes, $P\sim10^{-4}$ for CAD, $P\sim10^{-4}$ for SBP and further evidence from functional annotations could be equally interesting and worthy of follow-up as a results with a single $P<5\times10^{-8}$ for CAD.

Five of the ten loci associated with lipid changes were genome-wide significant for CAD ($P<5\times10^{-8}$) (*APOE-APOC1*, *TRIB1*, *ZNF159-APOA5-APOA1*, *LPL-CSGALNACT* and *LDLR*), two were not associated with CAD ($P>0.05$ for *DOCK7-ANGPTL3* and *GIPR*). Three loci showed promising suggestive evidence for CAD ($5\times10^{-8}<P<10^{-3}$) – these loci were *CAPN3*, *HP-HPR* and *QPCTL*. These results are demonstrated in Table 5.

Table 5. Lead SNPs from the meta-analysis of the five studies in Paper II. (N_{max}=8,263)

Trait	SNP	LOCUS	Δlipid *P*	SNP	CAD OR	CAD *P*
ΔTC	chr15:40469997	*CAPN3*	1.5×10^{-4}	rs2412710	1.14	2.4×10^{-4}
ΔTC	chr16:70636491	*HP/HPR*	2.3×10^{-4}	rs7197453	1.04	6.1×10^{-5}
ΔTG	rs11668847	*QPCTL*	2.7×10^{-5}	rs2341097	1.03	3.5×10^{-4}

CAD – coronary artery disease; SNP – single nucleotide polymorphism; OR – odds ratio; TC – total cholesterol; TG - triglycerides. Δlipid *P* values are obtained from fixed-effect meta-analysis of the five studies. CAD *P* values are extracted from the CARDIoGRAMplusC4D Consortium publicly available dataset.

Three loci (*APOE-APOC1*, *TRIB1* and *CAPN3*) showed suggestive statistically significant associations with T2D ($5\times10^{-8}<P<10^{-3}$), while two loci provided further suggestive evidence for systolic blood pressure (*CAPN3* and *DOCK7-ANGPTL3*). None of the ten loci was associated with diastolic blood pressure.

We functionally annotated our lead SNPs at the *LPL-CSGALNACT1*, *LDLR*, *DOCK7-ANGPTL3*, *CAPN3*, *HP-HPR*, *GIPR* and *QPCTL* loci. Two SNPs, rs1050362 near *HPR* and rs2341097 at *SIX5* (in close proximity to *QPCTL* and *GIPR*) overlapped with elements displaying excess constraint. Variants rs2412710 at *CAPN3*, rs1050362 near *HPR*, rs2341097 at *SIX5* and rs13282247 at *LPL-CSGALNACT1* disrupted CpG-dinucleotides, potentially altering the local balance of DNA methylation. In total, 6/11 SNPs exhibited DHSs in at least one ENCODE cell line, including rs2412710 at *CAPN3*. The rs2412710 variant at *CAPN3* was indicated by a "weak enhancer" state in HepG2 cells, the rs10406431 variant at *GIPR* was indicated by a "weak enhancer" state in K562 cells and a "repressed" state in GM12878 cells. The rs11668847 at *QPCTL* was indicated by a "strong enhancer" state in GM12878 cells. In total, 5/11 SNPs were in regions with ChIP-seq peaks. Variants rs11668847 at *QPCTL*, rs10406431 at *GIPR*, rs17304534 at *LDLR*, chr1:62954723 at *DOCK7-ANGPTL3* and rs13282247 at *LPL-CSGALNACT1* displayed evidence for binding of multiple enhancer and gene regulation-related factors.

One of the most interesting findings of this study was for the *CAPN3* locus, which was associated with TC level changes ($P=1.5\times10^{-4}$) and suggestively associated with CAD ($P=1.2\times10^{-4}$), T2D ($P=1.6\times10^{-3}$) and SBP ($P=6.9\times10^{-3}$). Functional annotation of the top SNP showed that it disrupts CpG-dinucleotides, displays DNaseI hypersensitivity sites in two cell lines and a "weak enhancer" status in HepG2 cells. Prior evidence demonstrated links between this locus and heart disease. In a murine study, targeted knock-in of calpain3 leads to increased circulating creatine kinase levels, a known marker for MI.[221] A Japanese autopsy study reported on two unrelated patients with calpainopathy (symmetric, progressive weakness of proximal muscles), who had died of ischemic cardiomyopathy and systemic circulatory failure, carrying mutations in *CAPN3*.[222]

Further interesting evidence was observed at the *QPCTL* and *GIPR* loci for TG change ($P=1.7\times10^{-4}$ and $P=2.7\times10^{-5}$, respectively). The top SNPs in these loci show evidence for functionality by displaying DNaseI hypersensitivity and regulatory

factor binding potential in multiple cell lines. The rs2341097 variant in the *SIX5* locus is in the close proximity of the *QPCTL* and *GIPR* loci and it shows suggestive evidence for association with CAD (P=1.9×10^{-4}). The *SIX5* rs16980013 variant has been associated with MI in Japanese adults.[223]

The *HP-HPR* locus was significantly associated with TC changes (P=2.3×10^{-4}) and showed suggestive evidence of association with CAD (P=3.5×10^{-5}). Although high haptoglobin level is a known risk factor for acute MI, stroke and heart failure,[224, 225] *HP-HPR* has not been associated with CAD in previous studies.

To understand prospective genetic associations with lipids, further studies examining HDL-C and LDL-C and other lipoprotein subfraction changes are warranted. In this study, we examined almost 30,000 variants from the Illumina CardioMetaboChip array, a genotyping chip enriched with sub-GWAS significant loci.[73] We assumed that these loci did not reach the required level of statistical significance due to various sources of heterogeneity (including heterogeneity in the average age of the study participants) in the original GWAS studies. Nevertheless, hypothesis-free studies might help to identify novel genetic variation in relation to lipid level changes. An important limitation of our study is the use of cohorts only from Northern Europe (four studies from Sweden, one from the UK), which results in a homogenous sample to study genetic associations. Further studies would benefit from the assessment of the genetic background of other ethnicities.

Paper III

In Paper III, we aimed to replicate a Danish study showing a statistically significant interaction between BMI and a GRS comprised of TG-associated variants on circulating TG levels. The Danish study by Justesen *et al* suggested that previously established TG-associated loci modify the association between BMI and TG, but to our knowledge, no independent replication studies had yet been published.[114] In addition, the original study showed heterogeneous interaction effects by sex but no formal three-way GRS × BMI × sex interactions were tested. In this study we independently replicated the GRS×BMI interaction in the Swedish GLACIER and MDC Studies and meta-analyzed our results with the findings from the original study. We tested for sex heterogeneity and we mapped whether these associations are mediated by various lifestyle factors including smoking, physical activity and diet. We also tested whether BMI × lifestyle interactions might underlie our findings. Last, we performed protein-protein interaction network analysis. In Paper III, I refer to the TG-associated weighted wGRS simply as wGRS as we did not work with any other traits' GRS in this project.

First, we conducted replication analysis in the northern Swedish GLACIER Study and the southern Swedish MDC Study. In GLACIER, both BMI and the wGRS robustly associated with TG levels (2.8% higher TG/BMI unit; P=8.4×10^{-84} and 2%

higher TG/wGRS unit; $P=7.6\times10^{-48}$). No association was observed between the wGRS and BMI ($P=0.14$). The association between the wGRS and TG was stronger in overweight/obese (defined as BMI ≥ 25) participants (2.4% higher TG/wGRS unit; 5.7% TG population variance explained by the wGRS) than in normal weight (defined as 18 ≤ BMI < 25) participants (1.5% higher TG/wGRS unit; 3.4% TG population variance explained by the wGRS). This heterogeneity indicates the presence of a gene × environment interaction, which was formally tested ($P_{interaction}=0.056$). Analysis in the MDC Study showed results directionally consistent and similar in magnitude to those observed in GLACIER. The interaction effect size in MDC was directionally consistent, albeit less statistically robust ($P_{interaction}=0.083$) than the one observed in GLACIER (Figure 10).

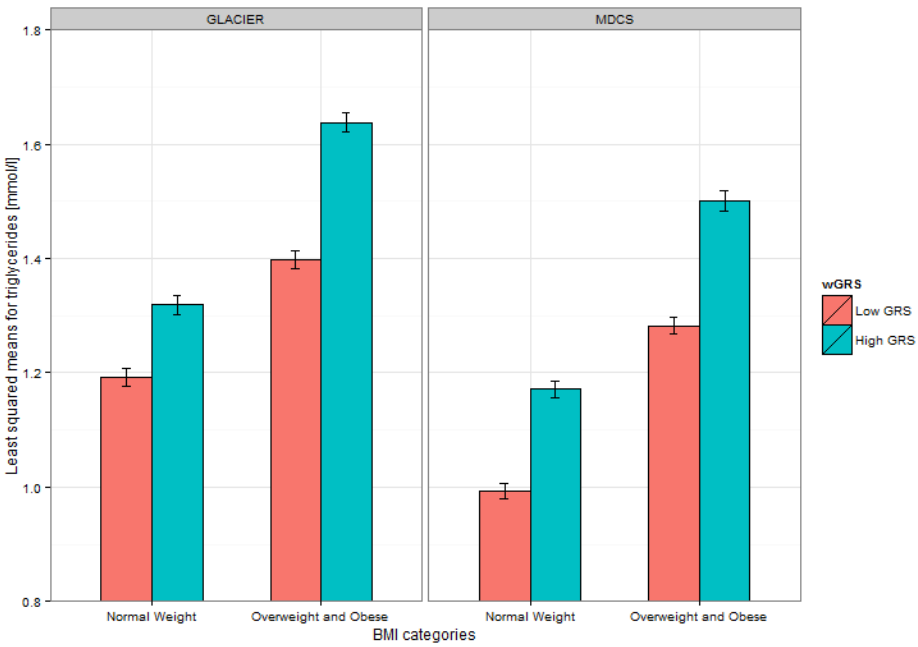

Figure 10. Bar plot indicating adjusted least square mean TG concentrations stratified by wGRS level.
BMI – body mass index; GRS – genetic risk score; wGRS – weighted genetic risk score. Black bars ≤ 44.6; grey bars > 44.6 units/alleles. Error bars indicate standard error of the least squared means.

We tested whether any of the individual SNPs from the wGRS underlie the observed interaction in GLACIER and MDC. None of the SNP × BMI interactions replicated between GLACIER and MDC (none of the SNP interactions have $P<0.05$ in both studies). In the literature, there are examples of robust interactions between established lipid-associated variants and lifestyle factors. A study by Zhang *et al* reports on a single variant in the *LIPC* locus interacting with dietary fat and BMI in relation to HDL-C levels in male participants of the Health Professionals Follow-

up Study.[226] Another study by Ahmad *et al* observed interactions between three lipid-associated loci, *LPL*, *CETP* and *LIPC* with physical activity on HDL-C levels and one, *LPL*, on MI in the Women's Health Initiative.[227]

In GLACIER, the wGRS × BMI interaction was only observed in females ($P_{interaction}$=0.0014), not in males ($P_{interaction}$=0.15). We formally tested a wGRS × BMI × sex three-way interaction, which proved to be statistically significant ($P_{interaction}$=0.0084). We conducted the same analysis in MDC, where the interaction was directionally consistent with GLACIER, but not statistically significant ($P_{interaction}$=0.11).

We tested whether lifestyle factors, such as dietary factors, physical activity, education or smoking mediate the interaction effects in GLACIER. Adjusting the interaction model with these factors strengthened the results ($P_{interaction}$=0.003 after diet score adjustment, $P_{interaction}$=0.01 after physical activity adjustment and $P_{interaction}$=0.01 after smoking adjustment). In separate models, we included wGRS × environmental factor interaction terms. None of these interactions were statistically significant ($P_{interaction}$=0.26 for the diet score; $P_{interaction}$=0.85 for leisure time physical activity; $P_{interaction}$=0.89 for smoking). It is possible that environmental factors correlated with BMI underlie the observed interactions - however, analyses testing this by further adjusting the models with lifestyle variables or including additional interaction terms did not materially change the results. Therefore, we conclude that the three major lifestyle factors, diet, physical activity and smoking are unlikely to underlie the observed wGRS × BMI interaction.

We meta-analyzed interaction effect estimates from GLACIER and MDC. We also extracted summary statistics from the original publication of the Danish studies. This way we were also able to meta-analyze the Swedish studies with the two Danish studies (Inter99 and Health2006). The meta-analysis of the two Swedish cohorts yielded a statistically significant wGRS × BMI interaction effect estimate ($P_{interaction}=6.0\times10^{-4}$), which was further strengthened with the inclusion of the two Danish studies in the meta-analysis ($P_{interaction}=6.5\times10^{-7}$). No heterogeneity was observed when meta-analyzing the four cohorts (I^2=0%, Cochran's P=0.70).

In the meta-analysis of the two Swedish cohorts, the wGRS × BMI × sex three-way interaction effect was nominally statistically significant ($P_{interaction}$=0.03). As the original study conducted no sex-specific interaction analyses, we were only able to meta-analyze their sex-stratified estimates with estimates from the Swedish studies. The sex-stratified meta-analysis demonstrated that the interaction effect size is much stronger in females ($P_{interaction}=6.8\times10^{-6}$) than in males ($P_{interaction}$=0.029). No heterogeneity was observed in the females-only meta-analysis (I^2= 0%, Cochran's P=0.75), but moderate-to-high heterogeneity was observed in the males-only meta-analysis (I^2=58.2%, Cochran's P=0.06).

In silico network analyses based on biological protein-protein interactions revealed that BMI-related protein products and TG-related protein products form two separate dense protein-protein interaction networks connected through interacting

nodes (Figure 11). This analysis suggests that there is cross-talk between BMI-associated and TG-associated gene networks. We identified three genomic loci, *low density lipoprotein receptor-related protein 1B* (*LRP1B*), *SH2B adaptor protein 1* (*SH2B1*) and *tubby bipartite transcription factor* (*TUB*) close to established BMI-associated loci that directly interact with genes *low density lipoprotein receptor-related protein associated protein 1* (*LRPAP1*), *insulin receptor* (*INSR*) and *insulin receptor substrate 1* (*IRS1*) close to established TG-associated loci. Six of the nine genes in this network are enriched in the *Lipid digestion, mobilization, and transport* Reactome pathway (P_{FDR}<0.001).

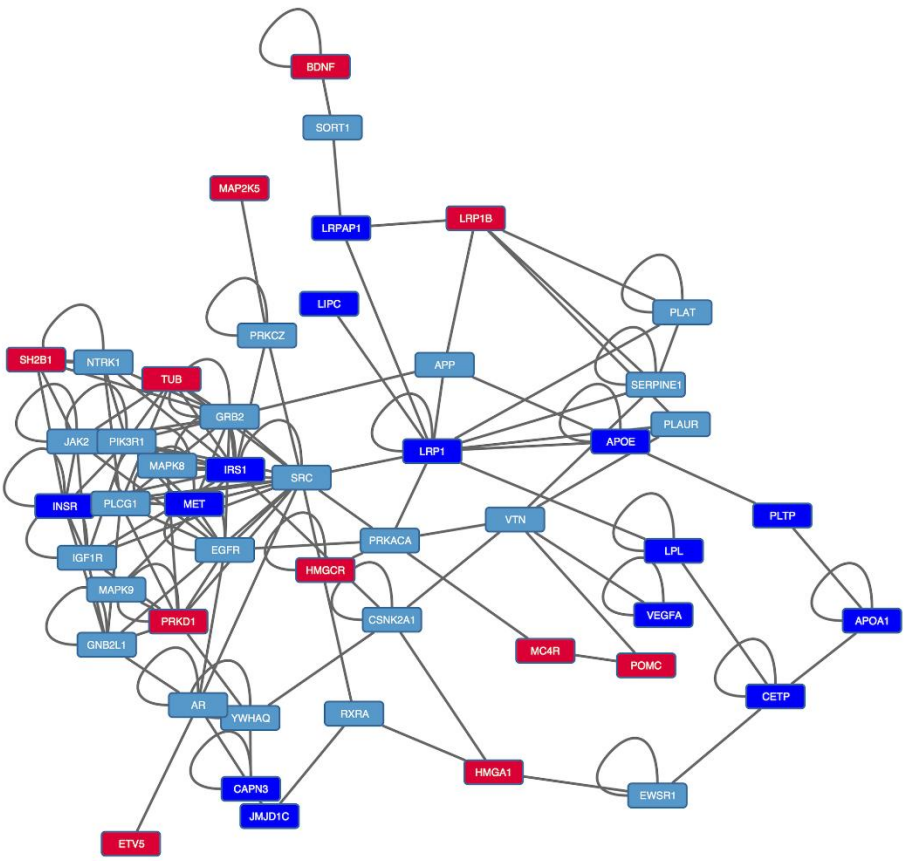

Figure 11. A sub-network of BMI- and TG-associated loci extracted from Human Protein Reference Database.
Genes close to BMI-associated loci (red) and genes close to TG-associated loci (blue) connected with only one intermediate gene (light blue) are displayed.

An important limitation of this study is the cross-sectional setting. In studies like ours, it is a difficult task to determine whether the observed findings are causal or not. Evidence from randomized controlled trials or studies with elaborate statistical

methods, such as well-powered Mendelian randomization might elucidate the causal of nature of the observations reported here.

Paper IV

In Paper IV, we aimed to comprehensively assess the established lipid-, lipoprotein subfraction- and lipoprotein size-associated genomic loci in relation to their respective traits in the DPP, a multi-ethic, multi-center randomized controlled trial. We examined the GWAS-established variants (the SNPs individually and in trait-specific GRSs) in relation to the lipoprotein traits at the baseline of the DPP. We assessed how much trait variance (in each lipoprotein trait) the GRSs and wGRSs account for. We studied SNP and GRS/wGRS × treatment arm (lifestyle intervention arm and metformin arm) interactions in relation to lipoprotein traits.

We tested 150 SNPs in relation to the four standard lipid traits (TG, TC, LDL-C and HDL-C) and 113 SNPs in relation to 16 other traits (small, large and total LDL particle numbers; small, medium, large and total VLDL particle numbers; small, medium, large and total HDL particle numbers; IDL-C, ApoB; LDL, VLDL and HDL particle sizes). Where any of these 16 traits correlated ($|r|\geq 0.5$) with a standard lipid trait (TG, TC, LDL-C and HDL-C), we tested associations between the correlated trait and the genetic component (SNP/GRS) for the standard trait (e.g. as LDL particle size correlated with TG, we tested associations between LDL particle size and TG-associated loci and the TG GRS. We also tested whether TG-associated loci or the TG GRS interact with the treatment arms on LDL particle size). Last, we assessed whether high genetic burden can be overcome with intensive lifestyle intervention for the traits in question by stratifying the DPP by the median GRS for all traits and comparing pre-intervention low GRS levels with post-intervention high GRS levels for each trait.

In SNP analyses at baseline, 59 (25.7%) of the associations between SNPs and standard lipid traits (TG, TC, HDL-C, LDL-C) replicated at the nominal $\alpha=0.05$ level. For the lipoprotein sub-fractions, of the 207 trait-specific associations and the 673 associations with correlated traits, 180 (20.5%) replicated at the nominal $\alpha=0.05$ level. Collectively, 227/1,110 (20.5%) of the association tests were statistically significant at a critical $\alpha=0.05$. In all baseline single SNP analyses, 28 SNPs (2.5%) reached the Bonferroni adjusted level of $\alpha=0.05/1110=4.5\times10^{-5}$.

In baseline GRS/wGRS analyses, in almost all the cases (32/34), the associations between the GRSs and their respective traits were statistically significant (*P* values ranging from 1.3×10^{-4} for total LDL to 2.4×10^{-16} for TC), with $P>0.05$ for tests of association for medium HDL and IDL-C with their respective GRSs. Using the wGRS strengthened the results for the majority of the traits (28/34 associations strengthened by using the wGRS instead of the GRS). The GRSs were positively correlated with baseline concentrations of TG, TC, LDL-C, small, large and total

LDL particle numbers, small, medium, large and total VLDL particle numbers, ApoB; LDL and VLDL particle sizes. The GRSs were negatively correlated with IDL-C, HDL-C, HDL particle size, and small, medium, large and total HDL particle numbers. The GRS – wGRS pairs (e.g. the GRS and wGRS for large VLDL particle concentration) were highly correlated; the average correlation coefficient between them was r^2=0.84.

The average trait variance explained by the trait-specific GRSs was 1.7% (minimum: 0.06% for IDL-C; maximum: 3.54% for small LDL particles). The trait-specific wGRSs explained on average 2.4% of the phenotypic variance (minimum: 0.02% for IDL-C; maximum: 4.66% for ApoB). We repeated these analyses by assessing how much trait variances the 20 GRSs explain cumulatively. All GRSs explained 5% of the phenotypic variance on average (minimum: 1.82% for HDL size; maximum: 8.88% for large VLDL particles). All wGRSs explained 6% of the phenotypic variance on average (minimum: 2.57% for medium HDL particles; maximum: 9.93% for large VLDL particles). The trait variances explained by the wGRS for the four main lipid traits ranged between 2% and 4.2%. Even though these wGRSs contain all dyslipidemia-associated loci, these explained variances are lower than those reported in Lutsey *et al* and Varga *et al*, other comprehensive assessments of lipid-related loci.[97, 220] As the DPP's sample size is smaller than these studies, the numbers reported here are probably less reliable.

In SNP × treatment arm interaction analyses, only one interaction test passed the Bonferroni corrected critical α level ($\alpha=0.05/1101=4.5\times10^{-5}$). The rs581080 variant in *tetratricopeptide repeat domain 39B* (*TTC39B*) showed evidence for interaction with the lifestyle treatment in relation to large HDL particle numbers ($P_{interaction}=2.8\times10^{-6}$ for lifestyle vs. placebo). The interaction between rs581080 and lifestyle intervention was not apparent in the metformin arm ($P_{interaction}$=0.19 for metformin vs. placebo). The interaction for rs581080 was no longer statistically significant when assessed only in European ancestry participants ($P_{interaction}$=0.12 for lifestyle vs. placebo). When visualizing our results, we concluded that the observed interaction between this SNP and the lifestyle intervention is likely to be a spurious result (interaction is driven by pre-randomization differences). The rs581080 variant was originally associated with HDL-C and TC concentrations in Willer *et al*. [74] Mouse experiments reported that *in vivo* knockdown of the *TTC39B* mouse homolog results in higher plasma HDL-C concentrations.[211, 219]

In GRS/wGRS × treatment arm interaction analyses, the metformin treatment arm and the lifestyle intervention arm modified the effect of the large HDL particle GRS, such that a higher GRS was associated with lower 1-year baseline-adjusted large HDL particle numbers in the metformin group (β=-0.08 µmol/l per GRS risk allele; 95%CI -0.141, -0.008; P=0.027; $P_{interaction}$=0.07 for metformin vs. placebo) and the lifestyle group (β=-0.11 µmol/l per GRS risk allele; 95%CI -0.188, -0.033; $P=5\times10^{-3}$; $P_{interaction}=1\times10^{-3}$ for lifestyle vs. placebo), but not the placebo group (β=-0.02 µmol/l per GRS risk allele; 95%CI -0.086, 0.042; P=0.50). Lifestyle intervention also modified the effect of the HDL size GRS, such that higher levels of the GRS

was associated with lower 1-year baseline-adjusted HDL particle size in the lifestyle group (β=-0.02 nm per GRS risk allele; 95%CI -0,036, -0,009; $P=1\times10^{-3}$; $P_{interaction}=8\times10^{-3}$ for lifestyle vs. placebo), but not in the placebo and metformin groups (β=-0.004 nm per GRS risk allele; 95%CI -0.014, 0.007; $P=0.50$ for placebo and β=-0.01 nm per GRS risk allele; 95%CI -0.021, 0.002; $P=0.11$; $P_{interaction}=0.28$ for metformin vs. placebo). This latter interaction in relation to HDL size did not survive our Bonferroni corrected critical α=0.05/34=0.0015. Using the wGRS attenuated these results, such that the interaction between lifestyle intervention and the large HDL particles wGRS on large HDL particle numbers ($P_{interaction}=6\times10^{-3}$ for lifestyle vs. placebo) and the interaction between lifestyle intervention and the HDL size wGRS on HDL particle size ($P_{interaction}=0.024$ for lifestyle vs. placebo) both became nominally significant. The interaction between large HDL particle concentrations and the lifestyle intervention arm is the most important finding of this study. The interaction means that even though lifestyle intervention was effective in the increase of large HDL particles on average, those with high genetic risk benefit less from lifestyle intervention than those with favorable genetic risk (Figure 12). In addition to this and as expected, those with the higher genetic risk have lower concentrations of large HDL particles from the outset.

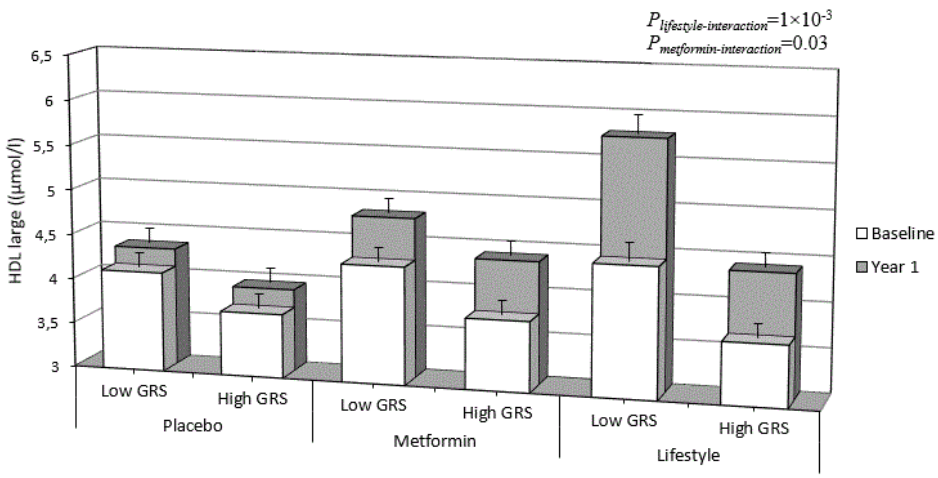

Figure 12. Large HDL particle numbers at baseline and 1-year later stratified by treatment group and high and low levels of the trait-specific genetic risk score (GRS).
GRS – genetic risk score; HDL – high-density lipoprotein. GRS by treatment interactions are shown for each active treatment group compared with the placebo group. Error bars represent standard deviations of the means.

Our study helps understand why some people are more responsive to drug and lifestyle therapies than others. It is important to note that we only observed interactions for lipoprotein subfractions and not for the four most studied lipid traits (TC, TG, LDL-C, HDL-C). Studies report that therapies targeting lipoprotein subfractions instead of the major fractions might be more clinically effective.[228, 229] Also, some studies report lipoprotein subfractions performing better in risk prediction algorithms focusing on hard outcomes such as type 2 diabetes and vascular diseases.[230, 231] Fibrate use resulting in the pharmacologic increase of small HDL particle concentrations reduced cardiovascular disease risk. [232]

A previous project in the DPP by Pollin *et al* [36] investigated treatment interactions with a GRS comprised of 32 lipid-associated variants reported by Kathiresan *et al*.[67] The GRS used in that study was not strait-specific; interactions were tested for each lipoprotein subfraction using the same risk score. In this study, two interactions were observed; the GRS was associated with higher baseline-adjusted LDL-C and small LDL particle concentrations in the lifestyle arm but not in the placebo arm. This means that high genetic risk attenuated the response to the intervention (those with high genetic risk were less able to decrease their LDL-C and small LDL particle levels in the lifestyle intervention arm), similarly as we observed with large HDL particles and HDL size in our study. Pollin *et al*'s and our study provide evidence that genetic burden cannot be completely overcome by interventions. On the other hand, both individuals in the low genetic risk group and the high genetic risk group benefited from the lifestyle intervention. It is important to note, that the results of these studies do not advocate against lifestyle interventions in high genetic risk groups, they merely provide evidence for how genetic burden is able to modify treatment responses. Goldberg *et al* examined the general effects of the metformin treatment and the intensive lifestyle intervention on lipid traits and lipoprotein subfractions in the DPP.[35] Lifestyle had a prominent favorable impact on VLDL particle numbers, mainly through large VLDL particle concentrations, a lipoprotein subclass very important in diabetic dyslipidemia.[233] Lifestyle intervention was also shown to be associated with increased large HDL particle numbers and size, LDL size, decreased total LDL particle numbers and most importantly, decreased small LDL particle numbers, a highly atherogenic lipoprotein subclass.[234] The metformin treatment arm elevated total and small HDL particle concentrations and lowered LDL subfraction concentrations.

We also undertook analysis to examine whether the dyslipidemia-associated GRSs/wGRSs modify treatment effects in relation to fasting and post-challenge glucose at follow-up and incident diabetes, the primary outcome of the DPP. None of the interactions survived multiple testing correction and we concluded that genetic risk for dyslipidemia are unlikely to modify treatment effects in relation to glycaemia. Previous evidence shows that many lipid and lipoprotein subfraction traits associate with glycemic traits (e.g. VLDL particle concentration is associated with diabetic dyslipidemia, LDL size correlates with fasting insulin).[233, 235]

Therefore, further analysis is warranted to identify whether the relationships between lipid and glycemic traits are modified by genetic loci.

In our analysis, lipid levels at randomization (baseline) in low genetic risk participants were compared with post- lifestyle intervention levels in high genetic risk participants. With this analysis, we aimed to assess whether high genetic burden can be overcome by intensive lifestyle intervention. For 17 out of the 20 traits, the post-intervention levels in the high GRS group were equal or more favorable than the pre-intervention levels in the low GRS group. This finding provides encouraging evidence that to a certain extent, high genetic risk can be overcome by intensive lifestyle interventions.

This is the most comprehensive assessment to date of dyslipidemia-associated loci in a longitudinal, randomized controlled trial setting. A major strength of this study is its design. Randomized controlled trials are considered to provide the highest level of evidence as they are the least prone to confounding, reverse causality and other sources of bias. Other strengths include the comprehensive assessment of all published dyslipidemia associated loci and the multi-ethnic nature of the analysis. An important limitation of the study is the low statistical power to detect associations and the lack of replication RCTs. As DPP is a unique source of data, we were unable to find an adequate replication study with a similar setting, availability of lipoprotein subfractions and genetic data. In the manuscript, we provided detailed power calculations and when presenting results, we always provided 95% CIs to inform the reader of the precision of our estimates. We acknowledge that the DPP has limited power to show gene × treatment interactions with low effect sizes; however, interactions with low effect sizes are unlikely to be clinically relevant. In conclusion, this study provides evidence for gene × treatment interactions in relation to large HDL particle concentrations. The remaining GWAS-derived variants, of which there are hundreds, did not modify treatment effects in the DPP to a clinically relevant degree.

Summary and conclusions

During my PhD, I conducted multiple analyses on the topic of genetics, gene × environment interactions and genetic epidemiology of complex traits. This thesis comprises the manuscripts/publications related to my work in lipid genetics, my primary interest during the past four years. In the two first projects, I aimed to assess whether genetic variants are associated with lipid level changes. Although there are a small number of genetic association studies which examine longitudinal genetic effects, our approach to meta-analyze estimates from multiple prospective studies can be considered a pioneering step. In Paper I, I examined previously published variants, while in Paper II, I extended this project by conducting a hypothesis-free study of a larger number of genetic variants. Apart from these longitudinal genetic association studies, which had little focus on environmental or lifestyle factors, I was also interested in whether there is evidence for gene × environment interactions in relation to lipid levels in observational studies or randomized clinical trials. In Paper IV, I studied established dyslipidemia loci in a clinical trial, where I assessed whether these genetic variants modify treatment effects. Paper III is the only study in this thesis that was conducted in a cross-sectional setting. Here I assessed whether obesity modifies the effects of established genetic risk factors for hypertriglyceridemia.

The most important conclusions of these papers are:

- **The majority of the established lipid-related loci, which are highly significantly associated with lipid levels in large-scale cross-sectional meta-analyses, do not associate with lipid level changes in prospective studies.** In Paper I, only three of the 157 established dyslipidemia loci associated with lipid level changes. In Paper II, multiple hits from the discovery phase of the analysis had no cross-sectional associations.
- **There are genetic variants that associate with lipid level changes.** Both Paper I (three novel variants) and Paper II (seven novel variants) report on novel genetic associations in relation to lipid level changes.
- **Established dyslipidemia-associated loci are poor predictors of future dyslipidemia.** Paper I demonstrated that genetic risk scores add very little to risk prediction models containing only basic information, such as age, sex and BMI.

- **Some genetic loci, which associate with lipid level changes show associations with clinical outcomes such as coronary artery disease, type 2 diabetes and hypertension.** In Paper II six variants showed evidence for associations for these traits in addition to the longitudinal genetic associations in relation to lipid levels.
- **Statistically significant gene × environment interactions can be detected in both observational, cross-sectional studies and randomized clinical trials.** Both Paper III and Paper IV report on statistically significant gene × environment interactions with the genetic factor being single variants and genetic risk scores and the environmental factors being adiposity and a complex lifestyle intervention. While the interactions we detected are unlikely to be of high clinical relevance, they clearly demonstrate the complex multifactorial nature of blood lipid levels.
- **Higher genetic risk for certain lipoprotein traits attenuates responses to lifestyle intervention.** Paper IV reports a statistically significant gene × treatment interaction in relation to large HDL particle concentrations. This interaction means that response to lifestyle intervention is attenuated by high genetic risk. Although both low and high genetic risk participants had improved large HDL particle concentrations in the lifestyle intervention arm at follow-up, those with the adverse genetic predisposition benefited less from the intervention.

Future Perspectives

To continue the efforts of Paper I and Paper II, we have recently established a larger consortium, The LOngitudinal traits - ConsortiUm of prospective Studies (LOCUS) to study genetic effects in relation to longitudinal changes in various quantitative anthropometric and cardiometabolic traits. These traits are TC, TG, HDL-C, LDL-C, height, waist circumference, waist-to-hip ratio, BMI, SBP, DPB, pulse pressure, mean arterial pressure, fasting glucose, 2-hr glucose, HbA1c and fasting insulin. LOCUS's overarching focus is on discovering genetic variants that convey time-dependent effects on these 16 anthropometric and cardiometabolic traits. In addition to the standard genetic association study in relation to trait changes, we are conducting heterogeneity of variance (an analytical method to infer gene × environment interactions) analysis and cross-phenotype association studies, as these may shed light on genetic effects and gene-environment interactions that are conditional on one or more additional traits. As many of the analyzed traits demonstrate sex heterogeneity, we study males and females separately.

GWAS has been extremely successful in detecting common genetic variation in relation to complex diseases. In order to validate findings from these studies, it is important to functionally characterize genetic associations and gene × environment interaction results. Current studies aim to refine these associations in pursuit of the causal loci and the causal variant(s) for cardiometabolic outcomes. Although we conducted functional studies in Papers II (functional annotation of top hits), III (protein-protein interaction network analysis) and IV (pathway enrichment analysis) using *in silico* resources, it is important that our findings and findings from GWAS studies are carried further and studied more so we fully understand their underlying physiology. In addition to this, emerging novel methods (e.g. Mendelian randomization studies) should be utilized to study the causal nature of genetic associations and gene × environment interactions.

Similarly, results from cross-sectional genetic association studies should be tested in prospective settings to gain further evidence of association/deeper understanding of etiology. These might include prospective cohort studies, randomized controlled trials, or new, cutting-edge study designs such as randomized controlled trials based on phenotype- or genotype-based recalls. In addition, as epidemiologic evidence has accumulated, results from genetic analyses should be carried further and studied in *in vitro* studies, such as knock-out/knock-down experiments or with CRISPR-Cas9/CRISPR-Cpf1 genome editing tools.

Genetic findings have proven extremely useful in developing novel drug targets (e.g. ongoing RCTs of PCSK9 inhibitors, LPL gene therapy). As our knowledge

about genetic associations grows, the greater our ability will be to identify regions that are good candidates for pharmacologic intervention.

Some of the studies in the literature and our study in the DPP show statistically significant interactions where those with adverse genetic predisposition benefit less from an intervention or have a relatively more unfavorable level of a certain trait after intervention. It would be potentially disastrous if one were to draw from these results the conclusion that lifestyle intervention should not be recommended to those with higher genetic risk. Our study only shows an example where genetic background modifies treatment effects and explains in part why certain individuals respond to various treatments differently. On the contrary to the detrimental conclusion mentioned above, those with high genetic risk should rather be prioritized for novel, more effective treatments. Further research, novel interventions and study designs are warranted to study and help these subgroups of individuals.

The study of clinical genetics has the potential to shift general medicine to precision medicine and eventually towards personalized medicine. Understanding how individual differences in the genome contribute to complex phenotypes may one day prove to be a powerful approach in designing more targeted clinical interventions, developing accurate prediction models, reducing side-effects of medications, improving cost-effectiveness of interventions and providing relief from the suffering associated with disease.

Swedish summary

Dyslipidemi betecknar ett tillstånd med kroniska avvikelser från normala blodfettsnivåer som ökar risken att drabbas av ateroskleros och andra kardiovaskulära sjukdomar. Dyslipidemi är resultatet av en kombination av ärftliga faktorer och miljö- och livsstilsrelaterade faktorer. Även om nivåerna av totalt kolesterol, HDL-kolesterol, LDL-kolesterol och triglycerider i blodet delvis är ärftliga är inte mycket känt om genetiska faktorer som kan bidra till försämring av blodfettsnivåer över tid. Genom det arbete som beskrivs i denna avhandling har jag försökt identifiera tidigare okända genetiska markörer för förändringar i blodfetter över tid. Jag har även undersökt om livsstilsfaktorer kan modifiera genetisk riskbenägenhet för dyslipidemi.

I **delarbete I** studeras de genetiska markörer som i tvärsnittstudier visat samband med nivåer av totalt kolesterol, HDL-kolesterol, LDL-kolesterol och triglycerider i storskaliga genomvida associationsstudier (GWAS). Samband mellan dessa genetiska markörer och blodfettsnivåer undersöktes med tvärtsnitts- och longitudinella analyser (10-års uppföljning) i GLACIER-kohorten från norra Sverige för att undersöka samband med både absoluta värden och förändringar i blodfetter över tid. Både den enskilda samt sammantagna effekten av de genetiska markörerna undersöktes. Detta arbete är den första studie där dessa genetiska markörer undersöks i longitudinella analyser. Jag utförde även riskprediktionsanalyser för risken att utveckla dyslipidemi under uppföljningstiden, detta för att undersöka hur mycket traditionella riskfaktorer samt genetiska riskfaktorer bidrar till en grundläggande riskprediktionsmodell som endast inkluderar ålder, kön och antropometriska data.

Delarbete II är en utvidgning av delarbete I där analyserna initialt utgår från ett mycket större antal genetiska markörer. Vår hypotes var att genetiska markörer som i genomvida associationsstudier inte riktigt (men nästan) uppvisar statistiskt säkerställda samband kanske inte gör det på grund av att sambandet mellan genetisk markör och blodfettsnivåer är tidsberoende och på olikheter i spridningen i ålder på deltagarna i de kohorter som analyserna genomförs i. I GLACIER-kohorten valde vi därför ut de totalt 29 387 genetiska markörer för blodfetter som inkluderats på MetaboChip array:en, dessa inkluderar även genetiska markörer som i tidigare studier visat tecken på samband som inte kunnat säkerställas statistiskt. I en första undersökande analys valde vi ut de genetiska markörer som uppvisade ett statistiskt samband med förändringar över tid i totalt kolesterol eller triglyceridnivåer i GLACIER-kohorten med ett p-värde <0.01. Sambandet mellan de utvalda genetiska markörerna och blodfetter undersöktes sedan i uppföljande analyser i fyra Europeiska prospektiva kohorter och slutligen meta-anlyserades resultaten från de totalt fem studierna.

Syftet med **delarbete III** var att upprepa (och därmed verifiera) resultaten från en dansk studie som visar att det observerade sambandet mellan vissa genetisk markörer och triglyceridnivåer i blodet modifieras av BMI. Med andra ord, sambandet mellan genetisk markör och triglyceridnivåer är inte konstant utan förändras beroende på forskningspersonens BMI. Detta fenomen kallas gen-miljöinteration. Interaktionsanalyserna genomfördes i GLACIER-kohorten och Malmö Kost Cancer studien och resultaten från de två studierna meta-analyserades tillsammans med resultatet från den ursprungliga danska studien. Vi studerade också om interaktionen var könsspecifik och hur den påverkas om man på olika sätt tar hänsyn till livsstilsfaktorer som rökning, fysiskt aktivetet och kost.

I **delarbete IV** var syftet att genomföra en omfattande analys av de genetiska markörer som identifierats för blodfetter, lipoproteinsubfraktioner och lipoproteinstorlek i DPP som är en multietnisk, multicenter, randomiserad, kontrollerad studie i USA. De genetiska markörer som identifierats i genomvida associationsstudier studerades i relation till blodfettsnivåer som mätts vid det basala undersökningstillfället. Både den enskilda samt sammantagna effekten av de genetiska markörerna som identifierats för respektive blodfettsprofil undersöktes. Vi studerade också hur mycket av den variation som fanns i uppmätta blodfettsvärden mellan forskningsdeltagarna som förklarades av de genetiska markörerna. Slutligen undersöktes om det fanns en gen-interventionsinteraktion; ifall sambandet mellan den sammantagna effekten av de genetiska markörerna för varje blodfettsprofil skiljde sig åt i de olika interventionsgrupperna (placebo, livsstil eller metformin-behandling).

De viktigaste resultaten och slutsatserna från delarbetena i min avhandling är:

- **Majoriteten av de genetiska markörer för blodfetter som identifierats i storskaliga meta-analyser av tvärsnittstudier är inte kopplade till förändringar i blodfetter över tid i prospektiva studier.** I delarbete I visade endast tre av 157 etablerade genetiska markörer för blodfetter på samband med förändring av blodfettsnivåer över tid. En del av de genetiska markörerna som i delarbete II uppvisade preliminära samband med förändringar i blodfetter över tid visade tvärtom inget samband med absoluta nivåer av blodfetter i tvärtsnittsanalyser.

- **Det finns genetiska variationer som fungerar som markörer för förändringar i blodfetter över tid.** Både delarbete I (tre genetiska variationer) och delarbete II (sju genetiska variationer) rapporterar tidigare okända samband mellan genetisk variation och förändring i blodfetter över tid.

- **Etablerade genetiska markörer för dyslipidemi är dåliga på att prediktera framtida risk för dyslipidemi.** Delarbete I demonstrerar att den sammantagna genetiska risken för dyslipidemi tillför väldigt lite till kliniska riskprediktionsmodeller som innefattar basal information som ålder, kön och BMI.

- **Vissa genetiska markörer som kopplas samman med förändring av blodfetter över tid visar också ett samband med ökad risk för kransartärsjukdom, typ 2-diabetes och högt blodtryck.** I delarbete II visade sex av de genetiska markörer som visade samband med förändringar i blodfetter över tid även samband med kliniska utfall som kransartärsjukdom, typ 2-diabetes och högt blodtryck.

- **Gen-miljöinteraktioner kan studeras i både observations-, tvärtsnitts- och randomiserade kliniska studier.** Både delarbete I och IV visar på statistiskt signifikanta gen-miljöinteraktioner där den genetiska komponenten är individuella genetiska variationer eller sammantagen genetisk risk och där miljökomponenten är fetma och livsstilsintervention. Även om de interaktioner som vi rapporterar i delarbete I och IV sannolikt inte har stor klinisk betydelse visar de tydligt på den komplexa och multifaktoriella regleringen av blodfetter.

- **Effekten av livsstilsintervention på lipoproteinprofiler kan reduceras till följd av en persons genetiska bakgrund.** Delarbete IV visar på en statistiskt signifikant gen-miljöinteraktion med avseende på koncentrationen av stora HDL-partiklar i blodet. Den observerade interaktionen innebär att även om personerna i gruppen som erhöll livsstilsintervention överlag uppvisade förbättrade värden med avseende på koncentrationen av stora HDL-partiklar i blodet vid det uppföljande besökstillfället så var effekten inte lika påtaglig bland forskningspersonerna med en hög genetisk predisposition för dyslipidemi.

Acknowledgements

Before the thank yous, I would like to share my brief personal story and how I ended up being who I am, where I am and what I am working with.

After gaining my BSc degree in Biology from the Eötvös Loránd University (Budapest, Hungary) I was lucky to get a five months scholarship to work with plant genetics / disease susceptibility at the University of Saskatchewan (Saskatoon, Canada). During my time in Canada I had ample time to read and I stumbled upon some bestsellers in the field of human nutrition. Among other books, *The Omnivore's Dilemma* and *In Defense of Food* from Michael Pollan had a great influence on me. Returning to Hungary I applied to an MSc program in Nutritional Sciences at the Semmelweis University (Budapest, Hungary). After completing the first year of the two-year program I realized that I miss working abroad and it would be great if I could do my MSc thesis work in a different country. I also knew that I would like to work with genetics and nutrition, thereby synthetize my BSc and MSc studies and my two main interests at that time. In the spring of 2011 I googled "*top100 universities*" and searched for departments and research groups where the main research focus was genetics and nutrition. I spent months researching work opportunities and writing e-mails to hundreds of professors, administrators, department heads with a very low response rate. Altogether, five professors showed interest, four of which did research in the United Kingdom, one – also from the United Kingdom – in Sweden. The professor in Sweden was straightforward and in his first e-mail asked me: "*Are you asking to come work with my group?*" After an interview with Paul and sorting out some details I started working with him on the 2nd September 2011. We were set out to work on an MSc thesis project for half year, but in December he offered me a PhD position in the Unit. I started my PhD in the spring of 2012. From the first days I was handling multiple projects – many individual ones and many in international consortia; by the end of my PhD this work resulted in ~25 manuscripts published or in submission / resubmission and an additional ~10 projects in the analysis stage. My interest in lipids and lipid genetics was triggered by a project I was leading (Paper I in this thesis) in 2013-2014. Since this undertake, I have been working more and more with lipid related projects and I decided to continue to work with complex trait genetics after finishing my PhD studies too.

The past four years represented a truly remarkable period in my life. I developed a great deal professionally, and I would like to believe – personally too. I would like to thank everybody who made this period in my life impossible to forget.

First and foremost, **Paul**. I am indebted for life for the opportunity to work with you. Your kindness, knowledge, experience and humor have helped me so much during the past five years. You are an amazing mentor and I will never forget the years I spent in your group.

Frida, thank you for being there for me and helping me. Your kindness and emotional intelligence were truly invaluable for me, you always recognized if something was up and were ready to help me.

Robert, we started our journey the same day and you have been an awesome friend, you know me well and you made the time I spent with the group so much fun. Thank you for all the chats, your great communication skills and the countless jokes, laughs and stories we share. This really has been something dude!

Angela, thank you for being my friend. I mentioned my personal development earlier – you played the most important role in this. You are my mentor and muse. Thank you for your warmth, honesty and trust; I would be lost in the fog without you.

Lili, my youngest sister – thank you for your unconditional love and innocence that brings so much light in my life. You have a special place in my heart.

Thanks to all my colleagues and friends, **Alaitz**, **Ashfaq**, **Azra**, **Britta**, **Dmitry**, **Jeremy**, **Jonathan**, **Johanna**, **Louise**, **Naeimeh**, **Pernilla**, **Shafqat**, **Simon**, **Sonja**, **Stefan**, **Yan**, **Åsa** for all your help and kindness. It was awesome to work with you. Thanks to my co-supervisor **Anders Rosengren** for all the valuable input and the help with future planning.

Thanks to my collaborators all around the world, most of all to those whom I could spend a research visit with - **Naveed Sattar** (University of Glasgow), **Sekar Kathiresan** and **Gina Peloso** (Broad Institute of MIT and Harvard), **Robert Scott** (University of Cambridge) and **Toni Pollin** (University of Maryland). The past few years were about learning and by spending time at your groups, I learned a lot.

A huge thanks to my friends in Hungary –**Gábor**, **Zsuzsa** and **Bálint** – as we get older we can always sit down and continue our conversations where we left off a while ago. Thank you for your love and support.

Last, I would like to thank my family, **Mom**, **Dad**, **Bali**, my sisters, **Zsófi**, **Saca**, **Fanni**, my grandmothers, **Kati mama**, **Ági mama**, **Amama**. You are my safe haven where I can always return to and I love you very much.

References

1. Fredrickson DS, et al. *A System for Phenotyping Hyperlipoproteinemia*. Circulation. 1965;**31**:321-327
2. Pollin TI, et al. *What We Know About Diet, Genes, and Dyslipidemia: Is There Potential for Translation?* Curr Nutr Rep. 2013;**2**:236-242
3. *Cardiovascular disease risk factors: new areas for research. Report of a WHO Scientific Group*. World Health Organ Tech Rep Ser. 1994;**841**:1-53
4. Grech ED. *Pathophysiology and investigation of coronary artery disease*. BMJ. 2003;**326**:1027-1030
5. Lusis AJ. *Atherosclerosis*. Nature. 2000;**407**:233-241
6. Sattar N. *Revisiting the links between glycaemia, diabetes and cardiovascular disease*. Diabetologia. 2013;**56**:686-695
7. Sattar N, et al. *Non-alcoholic fatty liver disease*. BMJ. 2014;**349**:g4596
8. Sattar N, et al. *Type 2 diabetes as a disease of ectopic fat?* BMC Med. 2014;**12**:123
9. Taylor R. *Pathogenesis of type 2 diabetes: tracing the reverse route from cure to cause*. Diabetologia. 2008;**51**:1781-1789
10. Fall T, et al. *Using Genetic Variants to Assess the Relationship Between Circulating Lipids and Type 2 Diabetes*. Diabetes. 2015;**64**:2676-2684
11. Gorden DL, et al. *Biomarkers of NAFLD progression: a lipidomics approach to an epidemic*. J Lipid Res. 2015;**56**:722-736
12. Nabel EG, et al. *A tale of coronary artery disease and myocardial infarction*. N Engl J Med. 2012;**366**:54-63
13. Goldstein JL, et al. *A century of cholesterol and coronaries: from plaques to genes to statins*. Cell. 2015;**161**:161-172
14. WHO. *Cardiovascular diseases (CVDs), Fact sheet N°317 http://www.who.int/mediacentre/factsheets/fs317/en/ . Accessed: 17 Feb 2016*.
15. WHO. *Obesity and overweight, Fact sheet N°311. http://www.who.int/mediacentre/factsheets/fs311/en/ . Accessed: 17 Feb 2016*.
16. American Diabetes A. *Diagnosis and classification of diabetes mellitus*. Diabetes Care. 2014;**37 Suppl 1**:S81-90
17. Anderson TJ, et al. *2012 update of the Canadian Cardiovascular Society guidelines for the diagnosis and treatment of dyslipidemia for the prevention of cardiovascular disease in the adult*. Can J Cardiol. 2013;**29**:151-167
18. National Cholesterol Education Program Expert Panel on Detection E, et al. *Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report*. Circulation. 2002;**106**:3143-3421
19. Vyroubal P, et al. *Hypocholesterolemia in clinically serious conditions--review*. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2008;**152**:181-189
20. Genest J, Jr., et al. *Familial hypoalphalipoproteinemia in premature coronary artery disease*. Arterioscler Thromb. 1993;**13**:1728-1737

21. Dashty M. *A Quick Look at Biochemistry: Lipid Metabolism*. J Diabetes Metab. 2014;**5**
22. Hanukoglu I. *Steroidogenic enzymes: structure, function, and role in regulation of steroid hormone biosynthesis*. J Steroid Biochem Mol Biol. 1992;**43**:779-804
23. Buhaescu I, et al. *Mevalonate pathway: a review of clinical and therapeutical implications*. Clin Biochem. 2007;**40**:575-584
24. Goldstein JL, et al. *Regulation of the mevalonate pathway*. Nature. 1990;**343**:425-430
25. Hegele RA. *Plasma lipoproteins: genetic influences and clinical implications*. Nat Rev Genet. 2009;**10**:109-121
26. AHA. *Good vs. Bad Cholesterol. http://www.heart.org/HEARTORG/Conditions/Cholesterol/AboutCholesterol/Good-vs-Bad-Cholesterol_UCM_305561_Article.jsp#.VsSHD_krJaQ . Accessed: 17 Feb 2016*.
27. van der Velde AE. *Reverse cholesterol transport revisited*. World J Gastroenterol. 2010;**16**:5907
28. Franceschini G, et al. *Reverse cholesterol transport: physiology and pharmacology*. Atherosclerosis. 1991;**88**:99-107
29. Lusis AJ, et al. *A treasure trove for lipoprotein biology*. Nat Genet. 2008;**40**:129-130
30. Glass CK, et al. *Atherosclerosis. the road ahead*. Cell. 2001;**104**:503-516
31. Lusis AJ. *Genetics of atherosclerosis*. Trends Genet. 2012;**28**:267-275
32. Moore KJ, et al. *Macrophages in atherosclerosis: a dynamic balance*. Nat Rev Immunol. 2013;**13**:709-721
33. Libby P. *Mechanisms of acute coronary syndromes and their implications for therapy*. N Engl J Med. 2013;**368**:2004-2013
34. Libby P, et al. *Pathophysiology of coronary artery disease*. Circulation. 2005;**111**:3481-3488
35. Goldberg R, et al. *Lifestyle and metformin treatment favorably influence lipoprotein subfraction distribution in the Diabetes Prevention Program*. J Clin Endocrinol Metab. 2013;**98**:3989-3998
36. Pollin TI, et al. *Genetic modulation of lipid profiles following lifestyle modification or metformin treatment: the Diabetes Prevention Program*. PLoS Genet. 2012;**8**:e1002895
37. Esposito K, et al. *Effect of weight loss and lifestyle changes on vascular inflammatory markers in obese women: a randomized trial*. JAMA. 2003;**289**:1799-1804
38. Eckel RH, et al. *2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines*. Circulation. 2014;**129**:S76-99
39. He BM, et al. *Effects of cigarette smoking on HDL quantity and function: implications for atherosclerosis*. J Cell Biochem. 2013;**114**:2431-2436
40. Preiss D, et al. *Emerging therapeutic approaches to treat dyslipidemia*. Curr Cardiol Rep. 2014;**16**:506
41. Ebrahim S, et al. *Statins for the primary prevention of cardiovascular disease*. BMJ. 2014;**348**:g280

42. Majeed A. *Statins for primary prevention of cardiovascular disease*. BMJ. 2014;**348**:g3491
43. Siri-Tarino PW, et al. *Saturated fatty acids and risk of coronary heart disease: modulation by replacement nutrients*. Curr Atheroscler Rep. 2010;**12**:384-390
44. Shah S, et al. *Influence of common genetic variation on blood lipid levels, cardiovascular risk, and coronary events in two British prospective cohort studies*. Eur Heart J. 2013;**34**:972-981
45. Mitchell BD, et al. *Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study*. Circulation. 1996;**94**:2159-2170
46. Pilia G, et al. *Heritability of cardiovascular and personality traits in 6,148 Sardinians*. PLoS Genet. 2006;**2**:e132
47. Ott J, et al. *Genetic linkage analysis in the age of whole-genome sequencing*. Nat Rev Genet. 2015;**16**:275-284
48. Pulst SM. *Genetic linkage analysis*. Arch Neurol. 1999;**56**:667-672
49. Visscher PM, et al. *Five years of GWAS discovery*. Am J Hum Genet. 2012;**90**:7-24
50. Lander ES. *Initial impact of the sequencing of the human genome*. Nature. 2011;**470**:187-197
51. Green ED, et al. *Human Genome Project: Twenty-five years of big biology*. Nature. 2015;**526**:29-31
52. The International HapMap Consortium. *The International HapMap Project*. Nature. 2003;**426**:789-796
53. The International HapMap Consortium. *A haplotype map of the human genome*. Nature. 2005;**437**:1299-1320
54. The 1000 Genomes Project Consortium. *An integrated map of genetic variation from 1,092 human genomes*. Nature. 2012;**491**:56-65
55. The 1000 Genomes Project Consortium. *A global reference for human genetic variation*. Nature. 2015;**526**:68-74
56. Ardlie KG, et al. *Patterns of linkage disequilibrium in the human genome*. Nat Rev Genet. 2002;**3**:299-309
57. Wall JD, et al. *Haplotype blocks and linkage disequilibrium in the human genome*. Nat Rev Genet. 2003;**4**:587-597
58. Kruglyak L. *Prospects for whole-genome linkage disequilibrium mapping of common disease genes*. Nat Genet. 1999;**22**:139-144
59. Klein RJ, et al. *Complement factor H polymorphism in age-related macular degeneration*. Science. 2005;**308**:385-389
60. Panagiotou OA, et al. *What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations*. Int J Epidemiol. 2012;**41**:273-286
61. Welter D, et al. *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations*. Nucleic Acids Res. 2014;**42**:D1001-1006
62. Hindorff LA, et al. *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits*. Proc Natl Acad Sci U S A. 2009;**106**:9362-9367
63. The Diabetes Genetics Initiative. *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels*. Science. 2007;**316**:1331-1336

64. Kooner JS, et al. *Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides*. Nat Genet. 2008;**40**:149-151
65. Kathiresan S, et al. *A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study*. BMC Med Genet. 2007;**8 Suppl 1**:S17
66. Kathiresan S, et al. *Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans*. Nat Genet. 2008;**40**:189-197
67. Kathiresan S, et al. *Common variants at 30 loci contribute to polygenic dyslipidemia*. Nat Genet. 2009;**41**:56-65
68. Sandhu MS, et al. *LDL-cholesterol concentrations: a genome-wide association study*. Lancet. 2008;**371**:483-491
69. Willer CJ, et al. *Newly identified loci that influence lipid concentrations and risk of coronary artery disease*. Nat Genet. 2008;**40**:161-169
70. Heid IM, et al. *Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions*. Circ Cardiovasc Genet. 2008;**1**:10-20
71. Aulchenko YS, et al. *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts*. Nat Genet. 2009;**41**:47-55
72. Teslovich TM, et al. *Biological, clinical and population relevance of 95 loci for blood lipids*. Nature. 2010;**466**:707-713
73. Voight BF, et al. *The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits*. PLoS Genet. 2012;**8**:e1002793
74. The Global Lipids Genetics Consortium. *Discovery and refinement of loci associated with lipid levels*. Nat Genet. 2013;**45**:1274-1283
75. Do R, et al. *Common variants associated with plasma triglycerides and risk for coronary artery disease*. Nat Genet. 2013
76. van Leeuwen EM, et al. *Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels*. Nat Commun. 2015;**6**:6065
77. Lu X, et al. *Genetic Susceptibility to Lipid Levels and Lipid Change Over Time and Risk of Incident Hyperlipidemia in Chinese Populations*. Circ Cardiovasc Genet. 2016;**9**:37-44
78. Chasman DI, et al. *Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis*. PLoS Genet. 2009;**5**:e1000730
79. Lemaitre RN, et al. *Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium*. PLoS Genet. 2011;**7**:e1002193
80. Lemaitre RN, et al. *Genetic loci associated with circulating levels of very long-chain saturated fatty acids*. J Lipid Res. 2015;**56**:176-184
81. Mozaffarian D, et al. *Genetic loci associated with circulating phospholipid trans fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium*. Am J Clin Nutr. 2015;**101**:398-406
82. Locke AE, et al. *Genetic studies of body mass index yield new insights for obesity biology*. Nature. 2015;**518**:197-206

83. Manolio TA, et al. *Finding the missing heritability of complex diseases*. Nature. 2009;**461**:747-753
84. Gibson G. *Rare and common variants: twenty arguments*. Nat Rev Genet. 2011;**13**:135-145
85. Willer CJ, et al. *Finding genes and variants for lipid levels after genome-wide association analysis*. Curr Opin Lipidol. 2012;**23**:98-103
86. Wessel J, et al. *Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility*. Nat Commun. 2015;**6**:5897
87. Albrechtsen A, et al. *Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes*. Diabetologia. 2013;**56**:298-310
88. Lange LA, et al. *Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol*. Am J Hum Genet. 2014;**94**:233-245
89. Futema M, et al. *Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations*. J Med Genet. 2014;**51**:537-544
90. Peloso GM, et al. *Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks*. Am J Hum Genet. 2014;**94**:223-232
91. Tang CS, et al. *Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese*. Nat Commun. 2015;**6**:10206
92. Timpson NJ, et al. *A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans*. Nat Commun. 2014;**5**:4871
93. The UK10K Consortium. *The UK10K project identifies rare variants in health and disease*. Nature. 2015;**526**:82-90
94. Johansen CT, et al. *Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia*. Nat Genet. 2010;**42**:684-687
95. Johansen CT, et al. *Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia*. Circ Cardiovasc Genet. 2012;**5**:66-72
96. Kurbasic A, et al. *Gene-Lifestyle Interactions in Complex Diseases: Design and Description of the GLACIER and VIKING Studies*. Curr Nutr Rep. 2014;**3**:400-411
97. Lutsey PL, et al. *Relation of lipid gene scores to longitudinal trends in lipid levels and incidence of abnormal lipid levels among individuals of European ancestry: the Atherosclerosis Risk in Communities (ARIC) study*. Circ Cardiovasc Genet. 2012;**5**:73-80
98. Dumitrescu L, et al. *Post-genome-wide association study challenges for lipid traits: describing age as a modifier of gene-lipid associations in the Population Architecture using Genomics and Epidemiology (PAGE) study*. Ann Hum Genet. 2013;**77**:416-425
99. Middelberg R, et al. *Evidence of age-dependent genetic influences on plasma total cholesterol*. Eur J Cardiovasc Prev Rehabil. 2005;**12**:380-386
100. Middelberg RP, et al. *Evidence of differential allelic effects between adolescents and adults for plasma high-density lipoprotein*. PLoS One. 2012;**7**:e35605
101. Middelberg RP, et al. *Longitudinal genetic analysis of plasma lipids*. Twin Res Hum Genet. 2006;**9**:550-557
102. Middelberg RP, et al. *A longitudinal genetic study of plasma lipids in adolescent twins*. Twin Res Hum Genet. 2007;**10**:127-135

103. Lu Y, et al. *Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study*. Atherosclerosis. 2010;**213**:200-205
104. Costanza MC, et al. *Consistency between cross-sectional and longitudinal SNP: blood lipid associations*. Eur J Epidemiol. 2012;**27**:131-138
105. Huang CC, et al. *Longitudinal association of PCSK9 sequence variations with low-density lipoprotein cholesterol levels: the Coronary Artery Risk Development in Young Adults Study*. Circ Cardiovasc Genet. 2009;**2**:354-361
106. Webster RJ, et al. *The association of common genetic variants in the APOA5, LPL and GCK genes with longitudinal changes in metabolic and cardiovascular traits*. Diabetologia. 2009;**52**:106-114
107. Zuk O, et al. *The mystery of missing heritability: Genetic interactions create phantom heritability*. Proc Natl Acad Sci U S A. 2012;**109**:1193-1198
108. Yang J, et al. *Common SNPs explain a large proportion of the heritability for human height*. Nat Genet. 2010;**42**:565-569
109. Sanna S, et al. *Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability*. PLoS Genet. 2011;**7**:e1002198
110. Edwards SL, et al. *Beyond GWASs: illuminating the dark road from association to function*. Am J Hum Genet. 2013;**93**:779-797
111. Franks PW. *Gene x environment interactions in type 2 diabetes*. Curr Diab Rep. 2011;**11**:552-561
112. Zheng JS, et al. *Genome-wide contribution of genotype by environment interaction to variation of diabetes-related traits*. PLoS One. 2013;**8**:e77442
113. Zheng JS, et al. *Genome-wide interaction of genotype by erythrocyte n-3 fatty acids contributes to phenotypic variance of diabetes-related traits*. BMC Genomics. 2014;**15**:781
114. Justesen JM, et al. *Interactions of Lipid Genetic Risk Scores with Estimates of Metabolic Health in a Danish Population*. Circ Cardiovasc Genet. 2015
115. Loria-Kohen V, et al. *A genetic variant of PPARA modulates cardiovascular risk biomarkers after milk consumption*. Nutrition. 2014;**30**:1144-1150
116. Rudkowska I, et al. *Gene-diet interactions on plasma lipid levels in the Inuit population*. Br J Nutr. 2012:1-9
117. Stojkovic IA, et al. *The PNPLA3 Ile148Met interacts with overweight and dietary intakes on fasting triglyceride levels*. Genes Nutr. 2014;**9**:388
118. Lu Y, et al. *Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem Cohort Study*. Am J Clin Nutr. 2010;**92**:258-265
119. Takkunen MJ, et al. *Gene-diet interaction of a common FADS1 variant with marine polyunsaturated fatty acids for fatty acid composition in plasma and erythrocytes among men*. Mol Nutr Food Res. 2016;**60**:381-389
120. Xu M, et al. *Dietary Fat Intake Modifies the Effect of a Common Variant in the LIPC Gene on Changes in Serum Lipid Concentrations during a Long-Term Weight-Loss Intervention Trial*. J Nutr. 2015;**145**:1289-1294
121. Smith CE, et al. *Physical inactivity interacts with an endothelial lipase polymorphism to modulate high density lipoprotein cholesterol in the GOLDN study*. Atherosclerosis. 2009;**206**:500-504

122. Abellan R, et al. *Dietary polyunsaturated fatty acids may increase plasma LDL-cholesterol and plasma cholesterol concentrations in carriers of an ABCG1 gene single nucleotide polymorphism: study in two Spanish populations*. Atherosclerosis. 2011;**219**:900-906
123. Aung LH, et al. *Proprotein convertase subtilisin/kexin type 9 gene E670G polymorphism interacts with alcohol consumption to modulate serum lipid levels*. Int J Med Sci. 2013;**10**:124-132
124. Liu WY, et al. *Interactions of the LIPG 584C>T polymorphism and alcohol consumption on serum lipid levels*. Alcohol. 2011;**45**:681-687
125. Wei XL, et al. *The peroxisome proliferator-activated receptor delta +294T > C polymorphism and alcohol consumption on serum lipid levels*. Lipids Health Dis. 2011;**10**:242
126. Junyent M, et al. *The effects of ABCG5/G8 polymorphisms on plasma HDL cholesterol concentrations depend on smoking habit in the Boston Puerto Rican Health Study*. J Lipid Res. 2009;**50**:565-573
127. Corella D, et al. *MicroRNA-410 regulated lipoprotein lipase variant rs13702 is associated with stroke incidence and modulated by diet in the randomized controlled PREDIMED trial*. Am J Clin Nutr. 2014;**100**:719-731
128. Brondani LA, et al. *Irisin-encoding gene (FNDC5) variant is associated with changes in blood pressure and lipid profile in type 2 diabetic women but not in men*. Metabolism. 2015;**64**:952-957
129. Taylor KC, et al. *Investigation of gene-by-sex interactions for lipid traits in diverse populations from the population architecture using genomics and epidemiology study*. BMC Genet. 2013;**14**:33
130. Asselbergs FW, et al. *Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci*. Am J Hum Genet. 2012;**91**:823-838
131. Barber MJ, et al. *Genome-wide association of lipid-lowering response to statins in combined study populations*. PLoS One. 2010;**5**:e9763
132. Thompson JF, et al. *Comprehensive whole-genome and candidate gene analysis for response to statin therapy in the Treating to New Targets (TNT) cohort*. Circ Cardiovasc Genet. 2009;**2**:173-181
133. Deshmukh HA, et al. *Genome-wide association study of genetic determinants of LDL-c response to atorvastatin therapy: importance of Lp(a)*. J Lipid Res. 2012;**53**:1000-1011
134. Hu M, et al. *Effects of phenotypic and genotypic factors on the lipid responses to niacin in Chinese patients with dyslipidemia*. Medicine (Baltimore). 2015;**94**:e881
135. Hu M, et al. *Liver fat reduction with niacin is influenced by DGAT-2 polymorphisms in hypertriglyceridemic patients*. J Lipid Res. 2012;**53**:802-809
136. Johansen CT, et al. *Genetic determinants of plasma triglycerides*. J Lipid Res. 2011;**52**:189-206
137. Gotto AM, et al. *Manual of lipid disorders : reducing the risk for coronary heart disease*. Philadelphia: Williams & Wilkins; 2003.
138. Voight BF, et al. *Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study*. Lancet. 2012;**380**:572-580
139. Holmes MV, et al. *Mendelian randomization of blood lipids for coronary heart disease*. Eur Heart J. 2015;**36**:539-550

140. Musunuru K, et al. *Surprises From Genetic Analyses of Lipid Risk Factors for Atherosclerosis*. Circ Res. 2016;**118**:579-585
141. *Learning lessons from Pfizer's $800 million failure*. Nat Rev Drug Discov. 2011;**10**:163-164
142. Joy TR, et al. *The failure of torcetrapib: what have we learned?* Br J Pharmacol. 2008;**154**:1379-1381
143. Miller NE. *CETP inhibitors and cardiovascular disease: Time to think again*. F1000Res. 2014;**3**:124
144. Tall AR, et al. *The failure of torcetrapib: was it the molecule or the mechanism?* Arterioscler Thromb Vasc Biol. 2007;**27**:257-260
145. Jansen H, et al. *Mendelian randomization studies in coronary artery disease*. Eur Heart J. 2014;**35**:1917-1924
146. Preiss D, et al. *The effect of statin therapy on heart failure events: a collaborative meta-analysis of unpublished data from major randomized trials*. Eur Heart J. 2015;**36**:1536-1546
147. Knopp RH, et al. *Effects of ezetimibe, a new cholesterol absorption inhibitor, on plasma lipids in patients with primary hypercholesterolemia*. Eur Heart J. 2003;**24**:729-741
148. Osto E, et al. *Rapid and body weight-independent improvement of endothelial and high-density lipoprotein function after Roux-en-Y gastric bypass: role of glucagon-like peptide-1*. Circulation. 2015;**131**:871-881
149. Baigent C, et al. *Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins*. Lancet. 2005;**366**:1267-1278
150. Pedersen TR. *The Success Story of LDL Cholesterol Lowering*. Circ Res. 2016;**118**:721-731
151. The Myocardial Infarction Genetics Consortium. *Inactivating mutations in NPC1L1 and protection from coronary heart disease*. N Engl J Med. 2014;**371**:2072-2082
152. Dadu RT, et al. *Lipid lowering with PCSK9 inhibitors*. Nat Rev Cardiol. 2014;**11**:563-575
153. Cohen JC, et al. *Sequence variations in PCSK9, low LDL, and protection against coronary heart disease*. N Engl J Med. 2006;**354**:1264-1272
154. Karalis I, et al. *Journey through cholesteryl ester transfer protein inhibition: from bench to bedside*. Circ Cardiovasc Qual Outcomes. 2013;**6**:360-366
155. Millar JS, et al. *Anacetrapib lowers LDL by increasing ApoB clearance in mildly hypercholesterolemic subjects*. J Clin Invest. 2015;**125**:2510-2522
156. Rached FH, et al. *An overview of the new frontiers in the treatment of atherogenic dyslipidemias*. Clin Pharmacol Ther. 2014;**96**:57-63
157. Morrison C. *$1-million price tag set for Glybera gene therapy*. Nat Biotechnol. 2015;**33**:217-218
158. Hallmans G, et al. *Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort - evaluation of risk factors and their interactions*. Scand J Public Health Suppl. 2003;**61**:18-24
159. The Diabetes Prevention Program. *Design and methods for a clinical trial in the prevention of type 2 diabetes*. Diabetes Care. 1999;**22**:623-634

160. Knowler WC, et al. *Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin*. N Engl J Med. 2002;**346**:393-403
161. The Diabetes Prevention Program. *The Diabetes Prevention Program (DPP): description of lifestyle intervention*. Diabetes Care. 2002;**25**:2165-2171
162. Knowler WC, et al. *Prevention of type 2 diabetes with troglitazone in the Diabetes Prevention Program*. Diabetes. 2005;**54**:1150-1156
163. Berglund G, et al. *The Malmo Diet and Cancer Study. Design and feasibility*. J Intern Med. 1993;**233**:45-51
164. Manjer J, et al. *The Malmo Diet and Cancer Study: representativity, cancer incidence and mortality in participants and non-participants*. Eur J Cancer Prev. 2001;**10**:489-499
165. Marchini J, et al. *Genotype imputation for genome-wide association studies*. Nat Rev Genet. 2010;**11**:499-511
166. Lind L, et al. *A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study*. Arterioscler Thromb Vasc Biol. 2005;**25**:2368-2375
167. Hedstrand H. *A study of middle-aged men with particular reference to risk factors for cardiovascular disease*. Ups J Med Sci Suppl. 1975;**19**:1-61
168. Williams DR, et al. *Undiagnosed glucose intolerance in the community: the Isle of Ely Diabetes Project*. Diabet Med. 1995;**12**:30-35
169. Jorgensen T, et al. *Effect of screening and lifestyle counselling on incidence of ischaemic heart disease in general population: Inter99 randomised trial*. BMJ. 2014;**348**:g3617
170. Thuesen BH, et al. *Cohort Profile: the Health2006 cohort, research centre for prevention and health*. Int J Epidemiol. 2014;**43**:568-575
171. The CARDIoGRAMplusC4D Consortium. *Large-scale association analysis identifies new risk loci for coronary artery disease*. Nat Genet. 2013;**45**:25-33
172. Morris AP, et al. *Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes*. Nat Genet. 2012;**44**:981-990
173. The International Consortium for Blood Pressure Genome-Wide Association Studies. *Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk*. Nature. 2011;**478**:103-109
174. Friedewald WT, et al. *Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge*. Clin Chem. 1972;**18**:499-502
175. Eliasson M, et al. *Time trends in population cholesterol levels 1986-2004: influence of lipid-lowering drugs, obesity, smoking and educational level. The northern Sweden MONICA study*. J Intern Med. 2006;**260**:551-559
176. Wu J, et al. *An investigation of the effects of lipid-lowering medications: genome-wide linkage analysis of lipids in the HyperGEN study*. BMC Genet. 2007;**8**:60
177. Tobin MD, et al. *Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure*. Stat Med. 2005;**24**:2911-2935
178. Warnick GR. *Enzymatic methods for quantification of lipoprotein lipids*. Methods Enzymol. 1986;**129**:101-123

179. Warnick GR, et al. *Dextran sulfate-Mg2+ precipitation procedure for quantitation of high-density-lipoprotein cholesterol*. Clin Chem. 1982;**28**:1379-1388
180. Hainline A, Jr, Karon J, Lippel K (eds.). *Manual of laboratory operations: Lipid and lipoprotein analysis (2nd ed.)* [HEW Pub. No. (NIH) 75-628 (rev.), U.S. Government Printing Office Publication No. 1982-361-132:678.] Bethesda, MD: National Heart, Lung and Blood Institute, Lipid Research Clinics Program.
181. Otvos JD. *Measurement of lipoprotein subclass profiles by nuclear magnetic resonance spectroscopy*. Clin Lab. 2002;**48**:171-180
182. Smith AD, et al. *Dietary patterns obtained through principal components analysis: the effect of input variable quantification*. Br J Nutr. 2013;**109**:1881-1891
183. Franks PW, et al. *Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations*. Diabetologia. 2008;**51**:458-463
184. Renstrom F, et al. *Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden*. Hum Mol Genet. 2009;**18**:1489-1496
185. Cornelis MC, et al. *Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry*. Ann Intern Med. 2009;**150**:541-550
186. Fontaine-Bisson B, et al. *Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population*. Diabetologia. 2010;**53**:2155-2162
187. Purcell S, et al. *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet. 2007;**81**:559-575
188. STATA 12.1 SL, TX, USA. *http://www.stata.com/*.
189. SAS 9.2 SII, Cary, NC, USA. *http://support.sas.com/documentation/92/index.html*.
190. R Core Team. *R: A Language and Environment for Statistical Computing*. 2013
191. Harris RJ, et al. *metan: fixed- and random-effects meta-analysis*. The Stata Journal. 2008;**8**:pp. 3-28
192. Magi R, et al. *GWAMA: software for genome-wide association meta-analysis*. BMC Bioinformatics. 2010;**11**:288
193. Higgins JP, et al. *Measuring inconsistency in meta-analyses*. BMJ. 2003;**327**:557-560
194. Baujat B, et al. *A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials*. Stat Med. 2002;**21**:2641-2652
195. Miller RG. *Simultaneous statistical inference. 2nd ed.*: Springer Verlag; 1981:6-8.
196. Bender R, et al. *Adjusting for multiple testing--when and how?* J Clin Epidemiol. 2001;**54**:343-349
197. Benjamini, Y; Hochberg, Y (1995). *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. Journal of the Royal Statistical Society, Series B **57** (1): 125–133.
198. Abdi, H. (2007). *Bonferroni and Šidák corrections for multiple comparisons.* In Salkind, N. J. Encyclopedia of Measurement and Statistics. Thousand Oaks, CA: Sage.
199. Zou KH, et al. *Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models*. Circulation. 2007;**115**:654-657
200. American Heart Association (AHA) - *What Your Cholesterol Levels Mean. http://www.heart.org/HEARTORG/Conditions/Cholesterol/AboutCholesterol/Wha*

t-Your-Cholesterol-Levels-Mean_UCM_305562_Article.jsp Accessed: 05 May 2013.
201. Cooper GM, et al. *Distribution and intensity of constraint in mammalian genomic sequence*. Genome Res. 2005;**15**:901-913
202. The ENCODE Project Consortium. *An integrated encyclopedia of DNA elements in the human genome*. Nature. 2012;**489**:57-74
203. Ernst J, et al. *Mapping and analysis of chromatin state dynamics in nine human cell types*. Nature. 2011;**473**:43-49
204. Alekseyenko AV, et al. *Nested Containment List (NCList): a new algorithm for accelerating interval query of genome alignment and interval databases*. Bioinformatics. 2007;**23**:1386-1393
205. Keshava Prasad TS, et al. *Human Protein Reference Database--2009 update*. Nucleic Acids Res. 2009;**37**:D767-772
206. Shannon P, et al. *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Res. 2003;**13**:2498-2504
207. Ligtenberg WP, et al. *Reconn: a cytoscape plug-in for exploring and visualizing reactome*. J Bioinform Comput Biol. 2013;**11**:1350004
208. Croft D. *Building models using Reactome pathways as templates*. Methods Mol Biol. 2013;**1021**:273-283
209. Sabatti C, et al. *Genome-wide association analysis of metabolic traits in a birth cohort from a founder population*. Nat Genet. 2009;**41**:35-46
210. Varbo A, et al. *TRIB1 and GCKR polymorphisms, lipid levels, and risk of ischemic heart disease in the general population*. Arterioscler Thromb Vasc Biol. 2011;**31**:451-457
211. Waterworth DM, et al. *Genetic variants influencing circulating lipid levels and risk of coronary artery disease*. Arterioscler Thromb Vasc Biol. 2010;**30**:2264-2276
212. Drenos F, et al. *Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk*. Hum Mol Genet. 2009;**18**:2305-2316
213. Elliott P, et al. *Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease*. JAMA. 2009;**302**:37-48
214. Golledge J, et al. *Apolipoprotein E genotype is associated with serum C-reactive protein but not abdominal aortic aneurysm*. Atherosclerosis. 2010;**209**:487-491
215. Smelt AH, et al. *Apolipoprotein E and familial dysbetalipoproteinemia: clinical, biochemical, and genetic aspects*. Semin Vasc Med. 2004;**4**:249-257
216. Kristiansson K, et al. *Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits*. Circ Cardiovasc Genet. 2012;**5**:242-249
217. Tragante V, et al. *The impact of susceptibility loci for coronary artery disease on other vascular domains and recurrence risk*. Eur Heart J. 2013;**34**:2896-2904
218. Xu L, et al. *An association study between genetic polymorphisms related to lipoprotein-associated phospholipase A(2) and coronary heart disease*. Exp Ther Med. 2013;**5**:742-750
219. Holmes MV, et al. *Utility of genetic determinants of lipids and cardiovascular events in assessing risk*. Nat Rev Cardiol. 2011;**8**:207-221

220. Varga TV, et al. *Genetic determinants of long-term changes in blood lipid concentrations: 10-year follow-up of the GLACIER study*. PLoS Genet. 2014;**10**:e1004388
221. Ojima K, et al. *Dynamic distribution of muscle-specific calpain in mice has a key role in physical-stress adaptation and is impaired in muscular dystrophy*. J Clin Invest. 2010;**120**:2672-2683
222. Hashiguchi S, et al. *[A Clinicopathological Investigation of Two Autopsy Cases of Calpainopathy (LGMD2A)]*. Brain Nerve. 2014;**66**:1097-1102
223. Fujimaki T, et al. *Association of genetic variants in SEMA3F, CLEC16A, LAMA3, and PCSK2 with myocardial infarction in Japanese individuals*. Atherosclerosis. 2010;**210**:468-473
224. Froguel P, et al. *A genome-wide association study identifies rs2000999 as a strong genetic determinant of circulating haptoglobin levels*. PLoS One. 2012;**7**:e32327
225. Holme I, et al. *Haptoglobin and risk of myocardial infarction, stroke, and congestive heart failure in 342,125 men and women in the Apolipoprotein MOrtality RISk study (AMORIS)*. Ann Med. 2009;**41**:522-532
226. Zhang C, et al. *Interactions between the -514C->T polymorphism of the hepatic lipase gene and lifestyle factors in relation to HDL concentrations among US diabetic men*. Am J Clin Nutr. 2005;**81**:1429-1435
227. Ahmad T, et al. *Physical activity modifies the effect of LPL, LIPC, and CETP polymorphisms on HDL-C levels and the risk of myocardial infarction in women of European ancestry*. Circ Cardiovasc Genet. 2011;**4**:74-80
228. Nikolic D, et al. *Lipoprotein subfractions in metabolic syndrome and obesity: clinical significance and therapeutic approaches*. Nutrients. 2013;**5**:928-948
229. Krauss RM. *Lipoprotein subfractions and cardiovascular disease risk*. Curr Opin Lipidol. 2010;**21**:305-311
230. Fizelova M, et al. *Associations of multiple lipoprotein and apolipoprotein measures with worsening of glycemia and incident type 2 diabetes in 6607 non-diabetic Finnish men*. Atherosclerosis. 2015;**240**:272-277
231. Sniderman AD, et al. *The strengths and limitations of the apoB/apoA-I ratio to predict the risk of vascular disease: a Hegelian analysis*. Curr Atheroscler Rep. 2007;**9**:261-265
232. Barter PJ, et al. *Cardioprotective properties of fibrates: which fibrate, which patients, what mechanism?* Circulation. 2006;**113**:1553-1555
233. Krauss RM. *Lipids and lipoproteins in patients with type 2 diabetes*. Diabetes Care. 2004;**27**:1496-1504
234. Grundy SM. *Small LDL, atherogenic dyslipidemia, and the metabolic syndrome*. Circulation. 1997;**95**:1-4
235. Lamarche B, et al. *Fasting insulin and apolipoprotein B levels and low-density lipoprotein particle size as risk factors for ischemic heart disease*. JAMA. 1998;**279**:1955-1961