Novel Risk Factors for Type II Diabetes Mellitus and Coronary Heart Disease

Abbas Dehghan

Novel Risk Factors for Type II Diabetes Mellitus and Coronary Heart Disease

Nieuwe risicofactoren voor type II diabetes en coronaire hartziekten

Proefschrift

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

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdiging zal plaatsvinden op Woensdag 31 maart 2010 om 15:30 uur

door

Abbas Dehghan

geboren te Shiraz, Iran

Promotiecommissie

Promotor: prof.dr. J.C.M. Witteman

Overige leden: prof.dr. E. Boersma

                     prof.dr. J.M. Dekker

                     prof.dr. C.M. van Duijn

Copromotor: dr. E.J. G. Sijbrands

"Let the beauty of what you love be what you do."

Rumi - Persian poet (1207 - 1273)

تقدیم به

رها،

مادر عزیزم،

و به یاد پدر بزرگوارم

To Raha;
my dear mother;
and in loving memory of my father

Publications and manuscripts based on the studies described in this thesis

Chapter 2.1

Dehghan A, Kardys I, de Maat MP, Uitterlinden AG, Sijbrands EJ, Bootsma AH, Stijnen T, Hofman A, Schram MT, Witteman JC. Genetic variation, C-reactive protein levels, and incidence of diabetes. Diabetes. 2007;56:872-8.

Chapter 2.2

Dehghan A, van Hoek M, Sijbrands EJ, Stijnen T, Hofman A, Witteman JC. Risk of type II diabetes attributable to C-reactive protein and other risk factors. Diabetes Care. 2007;30:2695-9.

Chapter 2.3

Dehghan A, van Hoek M, Sijbrands EJ, Hofman A, Witteman JC. High serum uric acid as a novel risk factor for type 2 diabetes. Diabetes Care. 2008;31:361-2.

Chapter 3.1

Dehghan A, van Hoek M, Sijbrands EJ, Hofman A, Witteman JC. Risk of coronary heart diseases related to high serum C-reactive protein and traditional risk factors. Eur J Cardiovasc Prev Rehabil. 2010 (accepted for publication)

Chapter 3.2

Dehghan A, van Hoek M, Sijbrands EJ, Oostra BA, Hofman A, van Duijn CM, Witteman JC. Lack of association of two common polymorphisms on 9p21 with risk of coronary heart disease and myocardial infarction; results from a prospective cohort study. BMC Med. 2008;6:30.

Chapter 4.1

Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Baumert J, Schnabel RB, Nambi V, Kavousi M, Smith NL, Smith AV, Larson MG, Hoogeveen RCM, Tracy R, Launer LJ, Yamamoto JF, Folsom A, Sijbrands EG, Pankow J, Keaney JF, Sun W, Fontes JD, Lu C, Astor B, Hofman A, Aspelund T, Siscovick D, Guo X, Rotter J, Thorand B, Gieger C, Coresh J, Uitterlinden A, Harris TB, Rice K, Boerwinkle E, Vasan RS, Van Duijn CM, Psaty BM, Gudnason V, Koenig W, Ballantyne C, Benjamin EJ, Witteman JCM. Meta-analysis of genome-wide association study in 26,967 subjects confirms six loci for C-reactive protein serum levels. 2010 (submitted)

Chapter 4.2

Dehghan A, Kottgen A, Yang Q, Hwang SJ, Kao WL, Rivadeneira F, Boerwinkle E, Levy D, Hofman A, Astor BC, Benjamin EJ, van Duijn CM, Witteman JC, Coresh J, Fox CS. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet. 2008;372:1953-61.

Chapter 4.3

Dehghan A, Yang Q, Peters A, Basu S, Bis JC, Rudnicka AR, Kavousi M, Chen MH, Baumert J, Lowe GD, McKnight B, Tang W, de Maat M, Larson MG, Eyhermendy S, McArdle WL, Lumley T, Pankow JS, Hofman A, Massaro JM, Rivadeneira F, Kolz M, Taylor KD, van Duijn CM, Kathiresan S, Illig T, Aulchenko YS, Volcik KA, Johnson AD, Uitterlinden AG, Tofler GH, Gieger C, Psaty BM, Couper DJ, Boerwinkle E, Koenig W, O'Donnell CJ, Witteman JC, Strachan DP, Smith NL, Folsom AR. Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. Circ Cardiovasc Genet. 2009;2:125-33.
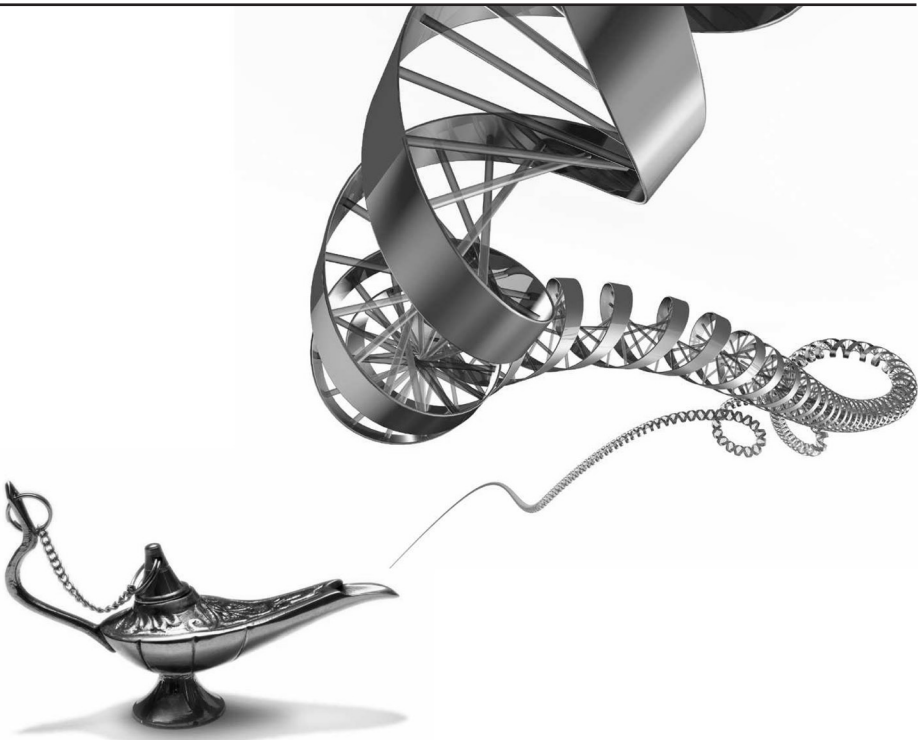
Table of contents

Chapter 1

Introduction

$\mathcal{D}$espite the huge advances made in the understanding of type II diabetes and coronary heart disease (CHD), these diseases still constitute a major health problem. Since the 1950s, epidemiologists focused on chronic disorders, including type II diabetes and CHD. Major aims of their research were to find predisposing factors and to reveal their pathophysiology. In the following decades, multiple traits and life-style behavioral factors were introduced and referred to as "risk factors". The so called traditional risk factors could explain part of the diseased cases, but a proportion of cases remained unexplained. For instance, obesity was identified as a major risk factor for type II diabetes, but not all patients were overweight. Similarly, it was estimated that at least 50% of CHD events were not caused by the traditional CHD risk factors[1]. These observations together with the needs for widening our knowledge on the pathogenesis of type II diabetes and CHD and better accuracy of disease prediction, called for moving beyond the known risk factors. In this thesis, we made an attempt to further study two novel risk factors.

C-reactive protein (CRP)

We studied inflammation as a novel risk factor both for type II diabetes and CHD. Chronic alteration of inflammatory function is related to insulin resistance and predisposes people to atherosclerosis. Inflammatory markers, likewise, predict the risk of type II diabetes[3] and CHD[4]. The causal role of CRP, a general marker of inflammation, in developing CHD has been highly discussed in recent years.

CRP was originally discovered by William Tillett and Thomas Francis, investigators from the Rockefeller University. They described a third fraction or "Fraction C", that could be isolated from patients infected with pneumococcus[5]. CRP was described as an "acute phase reactant" in 1947[6] and was found to be related to myocardial infarction in 1954[7]. Despite these early findings, it was not until the 1990s that cardiovascular interest in CRP was revitalized. In 1997, Ridker et al showed in a prospective study that baseline CRP levels were significantly higher among those who subsequently experienced myocardial infarction or stroke. Soon after that, prospective studies showed that CRP is also associated with the risk of type II diabetes[3].

Uric acid

Uric acid was initially discovered in the urine by Scheele in 1776[12]. A few years later, Wollastone extracted uric acid from a gouty tophus and speculated

on the relation between uric acid and gout[12]. In 1848, Garrod established the relation by showing that uric acid levels are high in gout patients[13].

Uric acid is the final product of purine metabolism and is mainly excreted by the kidneys. It has historically been viewed as a waste product, however, recent studies show that uric acid may have a wide range of actions, including being both a pro- and anti-oxidant[14], a neurostimulant[15], an inducer of inflammation[16] and an activator of the innate immune response[17].

At the end of 1980s, a new entity, the "metabolic syndrome", was incorporated into the medical terms and substituted the "X syndrome" which was described by Reaven. The metabolic syndrome was supposed to be the common root for type II diabetes and CHD. Hyperuricemia has been debated for a while as a potential component of the metabolic syndrome. Several studies showed that high serum uric acid is associated with components of the metabolic syndrome[8-11].

The impact of a risk factor on disease risk

How much of the incidence of a disease is due to a certain risk factor, and how many cases could be prevented if a risk factor is eliminated? The answer to these questions would reveal the impact of a risk factor on disease risk. Knowing the impact of most or all risk factors would help public health authorities to prioritize their actions[18]. This might also be of interest from an etiologic point of view. High impact risk factors are more likely to pinpoint to a crucial pathway in the pathogenesis of the disease.

The impact of risk factors on a disease could be estimated by calculating the population attributable risk (PAR). The PAR was originally calculated based on the incidence of the disease in the presence and the absence of the risk factor. Since such information is scarce, alternative formulas are developed based on the relative risk and the prevalence of the risk factor. The stronger and the more prevalent the risk factor, the larger is the PAR[19,20].

Genome-wide association studies

Connecting diseases with genes and discovering a new mechanism is the fundamental goal of human genetic studies. Genetic studies traditionally used candidate gene or family-based linkage studies to search for novel genes. Candidate gene studies rely on our partial understanding of the biologic pathways that relate genes to phenotypes. Though many studies can be found in the literature that used this approach, many of the reported findings fail to be replicated in a subsequent study[21]. Family-based linkage studies were successful in identifying genes for Mendelian diseases with large effects, however, achieved only little success in identifying genes for common diseases such as

type II diabetes and CHD[22]. Low power and lack of replication is also a problem with linkage studies.

In recent years, two fundamental breakthroughs helped the advent of a revolutionary and powerful approach in genetic epidemiology, the genome-wide association (GWA) study. The completion of the Human Genome Project and the International HapMap Project provided the possibility to select a set of genetic variants that nearly cover the whole variation in the human genome[23]. Furthermore, recent developments in genotyping technologies made it possible to assay hundreds of thousands of single-nucleotide polymorphisms (SNPs) in a short time for a reasonable price.

The strength of the genome-wide association study approach is the ability to investigate the genetic component of common diseases without relying on prior knowledge. This "hypothesis free" search of the whole genome for evidence of association with common variants promised to open up new avenues of research, through discovery of new genes for diseases[24].

Scope of the thesis

In this thesis, we aimed at expanding the knowledge on novel risk factors for type II diabetes and CHD. We focused on two risk factors that received attention in recent years, serum CRP and serum uric acid. We examined the contribution of these risk factors to the risks of type II diabetes and CHD. These studies were conducted in the Rotterdam Study. Moreover, we used the novel approach of GWA analysis to seek genetic factors that affect their levels. The GWA studies were performed within the framework of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, a consortium of population-based follow-up studies with genome-wide scan data.

In chapter 2, we focused on novel risk factors for type II diabetes. In chapter 2.1, we studied the association of serum CRP with risk of diabetes. We conducted a meta-analysis to clarify the role of obesity in the association between serum CRP and type II diabetes. Furthermore, we studied the association of variations in the CRP gene with the risk of diabetes. In chapter 2.2, we estimated the PAR of high CRP for type II diabetes. In chapter 2.3, we investigated the association of serum uric acid with type II diabetes.

Chapter 3 focuses on risk of CHD. In chapter 3.1, we estimated the proportion of CHD risk attributable to high serum CRP and traditional risk factors. In chapter 3.2, we examined the association of a genetic variation that was found to increase the risk of CHD by previous GWA studies, with risk of CHD in the Rotterdam Study.

In chapter 4, we report three large meta-analyses using the powerful approach of GWA study to identify genetic loci related to levels of CRP, uric acid and

fibrinogen. In chapter 4.1, we aimed to identify genetic loci related to serum CRP using GWA scans from six population-based cohort studies. In chapter 4.2, we conducted GWA studies for uric acid and gout in 11,847 individuals from the Framingham Heart Study and the Rotterdam Study and replicated the findings in 14,867 individuals from the Atherosclerosis Risk in Communities (ARIC) Study. In chapter 4.3, we performed a meta-analysis of GWA studies for levels of fibrinogen in 22,096 individuals from six population-based studies.

   In chapter 5, we discuss the main findings of this thesis are placed in a broader context. We also address the methodological considerations, potential clinical implications and directions for future research.

References

1.   Braunwald E. Shattuck lecture--cardiovascular medicine at the turn of the millennium: triumphs, concerns, and opportunities. N Engl J Med. 1997;337:1360-9.

2.   Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Brener SJ, Ellis SG, Lincoff AM, Topol EJ. Prevalence of conventional risk factors in patients with coronary heart disease. Jama. 2003;290:898-904.

3.   Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. Jama. 2001;286:327-34.

4.   Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N Engl J Med. 2004;350:1387-97.

5.   Tillett W, JT F. Serological reactions in pneumonia with a non-protein somatic fraction of pneumococcus. The Journal of Experimental Medicine. 1930;52:561-571.

6.   McCarty M. The Occurrence During Acute Infections of a Protein Not Normally Present in the Blood : Iv. Crystallization of the C-Reactive Protein. J Exp Med. 1947;85:491-498.

7.   Kroop IG, Shackman NH. Level of C-reactive protein as a measure of acute myocardial infarction. Proc Soc Exp Biol Med. 1954;86:95-7.

8.   Lee J, Sparrow D, Vokonas PS, Landsberg L, Weiss ST. Uric acid and coronary heart disease risk: evidence for a role of uric acid in the obesity-insulin resistance syndrome. The Normative Aging Study. Am J Epidemiol. 1995;142:288-94.

9.   Rathmann W, Funkhouser E, Dyer AR, Roseman JM. Relations of hyperuricemia with the various components of the insulin resistance syndrome in young black and white adults: the CARDIA study. Coronary Artery Risk Development in Young Adults. Ann Epidemiol. 1998;8:250-61.

10.   Emmerson B. Hyperlipidaemia in hyperuricaemia and gout. Ann Rheum Dis. 1998;57:509-10.

11.   Fam AG. Gout, diet, and the insulin resistance syndrome. J Rheumatol. 2002;29:1350-5.

12.   Richet G. The chemistry of urinary stones around 1800: a first in clinical chemistry. Kidney Int. 1995;48:876-86.

13.   Storey GD. Alfred Baring Garrod (1819-1907). Rheumatology (Oxford). 2001;40:1189-90.

14.   Sautin YY, Johnson RJ. Uric acid: the oxidant-antioxidant paradox. Nucleosides Nucleotides Nucleic Acids. 2008;27:608-19.

15.   Shi Y, Evans JE, Rock KL. Molecular identification of a danger signal that alerts the immune system to dying cells. Nature. 2003;425:516-21.

16.   Gersch MS, Johnson RJ. Uric acid and the immune response. Nephrol Dial Transplant. 2006;21:3046-7.

17.   Martinon F, Petrilli V, Mayor A, Tardivel A, Tschopp J. Gout-associated uric acid crystals activate the NALP3 inflammasome. Nature. 2006;440:237-41.

18.   Northridge ME. Public health methods--attributable risk as a link between causality and public health action. Am J Public Health. 1995;85:1202-4.

19.   Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL,

Schoenberg JB, Mayne ST, Dubrow R, Rotterdam H, West AB, Blaser M, Blot WJ, Gail MH, Fraumeni JF, Jr. Population attributable risks of esophageal and gastric cancers. J Natl Cancer Inst. 2003;95:1404-13.

20. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985;122:904-14.

21. Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. Jama. 2007;297:1551-61.

22. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. Am J Hum Genet. 2001;69:936-50.

23. Phimister EG. Genomic cartography--presenting the HapMap. N Engl J Med. 2005;353:1766-8.

24. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005;6:95-108.

Chapter 2

Novel Risk Factors for Type II Diabetes Mellitus
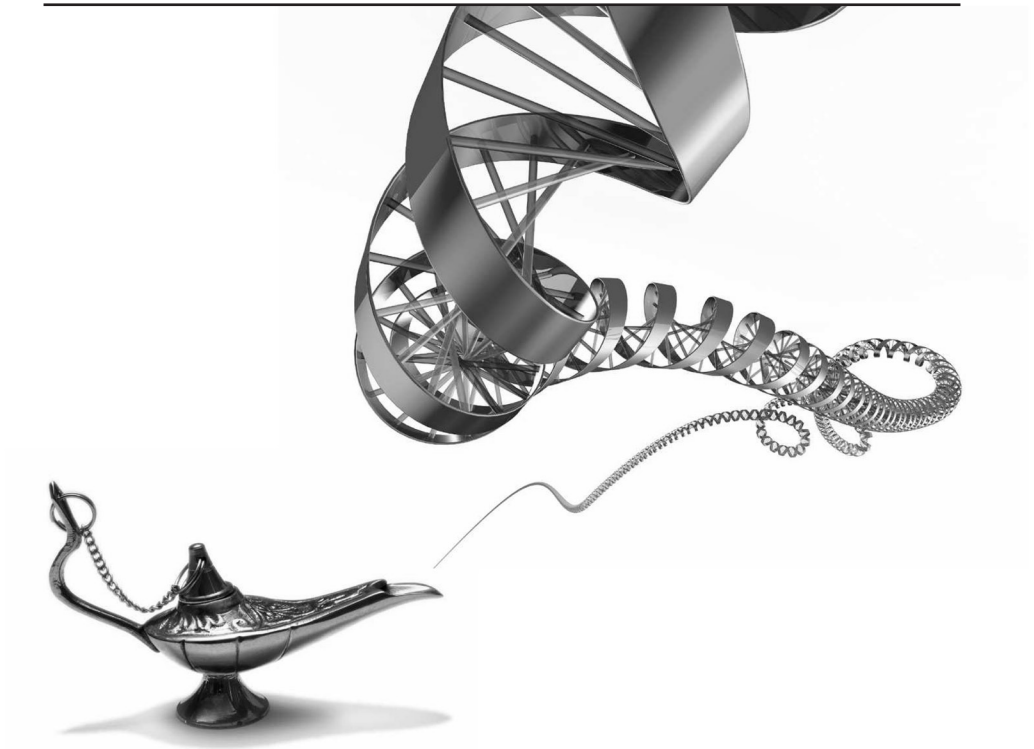
Chapter 2.1

C-reactive protein levels and incidence of type II diabetes

Abstract

Background

C-reactive protein (CRP) has been shown to be associated with type II diabetes but whether CRP has a causal role is not yet clear. We examined the association in the Rotterdam Study, a population-based, prospective cohort study.

Method

The association of baseline serum CRP and incident diabetes during follow-up was investigated and a meta-analysis was conducted on the BMI adjusted relation of CRP and diabetes. Furthermore, the association of CRP haplotypes with serum CRP and risk of diabetes was assessed.

Results

The age and sex adjusted hazard ratio (HR) for diabetes was 1.41 (95%CI=1.29-1.54) per 1 standard deviation increase in Ln CRP, and 1.88, 2.16 and 2.83 for the second, third and fourth quartiles of CRP respectively compared to the first quartile. The risk estimates attenuated but remained statistically significant after additional adjustment for obesity indexes, which agreed with the results of the meta-analysis. The most common genetic haplotype was associated with a significantly lower CRP level compared to the three other haplotypes. The risk of diabetes was significantly higher in the haplotype with the highest serum CRP level compared to the most common haplotype (OR=1.45; 95%CI=1.08-1.96).

Conclusion

These findings support the hypothesis that serum CRP enhances the development of diabetes.

Introduction

Prospective studies have shown that C-reactive protein (CRP), which is a general marker of systemic inflammation, is associated with the risk of diabetes mellitus[1-9]. CRP is produced by hepatocytes and its gene expression is regulated by tumor necrosis factor alpha (TNFα) and interleukin-6, which are secreted by adipocytes[10]. As a result, obese individuals who have more and larger adipocytes, have higher baseline serum CRP. Since diabetes is more common in obese individuals, an association is expected between serum CRP and diabetes. However, some of studies found that obesity does not explain the association of CRP with diabetes completely and suggested an independent role for CRP in development of diabetes[1,5,9].

Twin and familial studies have shown a substantial hereditability for CRP level [11] and a recent study found a strong association of serum CRP with genetic variations in the CRP promoter region[12]. Four haplotypes broadly represent the CRP gene variation in the European population[13]. Therefore, an association of CRP haplotypes with serum CRP and the incidence of diabetes may point at a contribution of CRP in the development of diabetes.

We studied the association of serum CRP with risk of diabetes in the Rotterdam Study, a prospective population-based cohort study among participants aged 55 years and older. Furthermore, we conducted a meta-analysis, which included our own study, to clarify whether CRP serum level is associated with diabetes, independent of obesity indexes. Finally, to investigate a potential role of CRP in the development of diabetes, we studied the association of genetic CRP haplotypes with the risk of diabetes.

Methods

Study Population

The study was conducted within the framework of the Rotterdam Study, an ongoing prospective, population-based cohort study on determinants of several chronic diseases. The Rotterdam Study has been described in detail elsewhere[14]. In brief, all inhabitants of Ommoord, a district of Rotterdam in the Netherlands, who were 55 years or over, were invited to participate in this study. Of all 10275 eligible individuals, 7983 agreed to participate (78%).

Participants were visited at home for an interview. Subsequently, they came to the research center for blood sampling and further examinations. Follow-up started at baseline and examinations were carried out periodically. In addition, participants were continuously monitored for major events through automated linkage with files from general practitioners and pharmacies working in the

study district of Ommoord. Information on vital status was obtained regularly from the municipal health authorities in Rotterdam. For the present study, follow-up data were present until October 1, 2005. Written informed consent was obtained from all participants and the Medical Ethics Committee of Erasmus University Rotterdam approved the study.

Diabetes

At baseline, prevalent cases of diabetes were diagnosed and excluded. Prevalent diabetes was defined as use of anti-diabetic medication, and/or abnormal non-fasting glucose, and/or an abnormal oral glucose tolerance test (OGTT). A non-fasting or post-load glucose level of 11.1 mmol/l or over was considered abnormal.15

During follow-up, incident cases of diabetes were diagnosed by use of information from the general practitioners, the pharmacies' databases, and our follow-up examinations. Based on guidelines of the American Diabetes Association16 and WHO17 we defined incident diabetes as follows: fasting plasma glucose level >= 7.0 mmol/l and/or random (non-fasting) plasma glucose level >= 11.1 mmol/l and/or use of oral anti-diabetic medication and/or use of insulin and/or treatment by diet and registered by a general practitioner as having diabetes.

Measurement of CRP serum levels

At baseline, serum levels of CRP were measured in 6658 out of 7129 participants who visited the research center. Non-fasting serum samples were collected. The samples were immediately put on ice and were processed within 30 minutes, after which they were kept frozen at -20 °C until measurement of CRP in 2003-2004. High-sensitivity CRP was measured using Rate Near Infrared Particle Immunoassay (Immage® Immunochemistry System, Beckman Coulter, USA). This system measures concentrations from 0.2 to 1440 mg/l, with a within-run precision <5.0%, a total precision <7.5% and a reliability coefficient of 0.995.

Genotyping

The Seattle SNPs Program for Genomic Applications has identified four haplotypes covering the CRP gene based on 18 SNPs that had a frequency of more than 5%, in 23 unrelated individuals of European descent from the CEPH pedigrees[13]. Results in these 23 individuals showed that each of these four haplotypes could be identified by a single tagging SNP. Hence, by determining three non-overlapping tagging SNPs we were able to infer all four haplotypes.

Genotyping for the three tagging SNPs 1184C/T, 2042C/T, and 2911C/G

was done for 5460 out of 7059 participants, whose blood was sampled during the baseline visit. The polymorphisms are described in relation to the start of the coding sequence of exon 1 using the Human May 2004 assembly (http://genome.ucsc.edu). These polymorphisms have also been described at http://www.ncbi.nlm.nih.gov/projects/SNP/ under identification numbers rs1130864 (1184C/T), rs1205 (2042C/T) and rs3093068 (2911C/G). CRP genotypes were determined in our study population in 2-ng genomic DNA with the Taqman allelic discrimination assay (Applied Biosystems, Foster City, California). We used the SNP assay-by-design service of Applied Biosystems to optimize the Primer and probe sequences (for details, see http://store.appliedbiosystems.com). Reactions were performed with the Taqman Prism 7900HT 384 wells format.

Statistical Analysis

A Cox regression analysis was used to assess the association of CRP with incident diabetes. We tested the proportional hazards assumption by using "log-minus-log" plots. A log transformation of serum CRP (LnCRP) was used since serum CRP had a right-skewed distribution. To examine the relation between CRP and incident diabetes, we computed the increase in hazard ratio (HR) per one standard deviation (SD) increase in Ln (CRP) level. We defined four quartiles of CRP level based on the population distribution (< 0.88, 0.88 – 1.80, 1.81 – 3.5, > 3.5 mg/l). In addition, a sensitivity analysis was done to assess the effect of adjustment for fasting blood sugar. Fasting blood sugar is measured in the third periodical examination of the Rotterdam Study. For this analysis, we used the third examination as the baseline measurement.

For the meta-analysis, previously published data were obtained by MEDLINE searches and scanning the reference lists until December 2005. Nine studies were found to be relevant and we added our results of the Rotterdam Study. No scoring system was used to qualify the studies. The effect estimates extracted were mostly HRs and incidence rate ratios (IRR). In one study, we used the odds ratio (OR) as an acceptable estimate of the risk ratio[4]. The effect estimates extracted were based on CRP tertiles, quartiles or quintiles, each stating the risk of diabetes in a specific range of serum CRP levels, compared to a reference group. To group the most relevant risk ratios for a corresponding serum CRP range, we defined three CRP level intervals. An effect estimate was allocated to an interval when its accredited CRP range was completely covered by that interval (0.5-1.8, 1-3.7, >2.6). The intervals overlap to allow more risk estimates to be allocated. Some effect estimates were not used because the CRP range they were based on did not fit in any of the defined intervals. Packages "meta"[18] and "rmeta"[19], running for "R"[20] was used to analyze the data. A random effects

model was used to weight the HRs. A weighted HR was calculated for each interval. Firstly, age and sex adjusted HRs and secondly, age, sex, and body mass index (BMI) adjusted HRs were used. We investigated whether race is a source of heterogeneity in the association of serum CRP and diabetes by performing a meta-regression using the SAS PROC MIXED[21].

We used the Haplo.Stats package running for R[20], to estimate the CRP haplotypes and to investigate the association of inferred haplotypes with serum CRP and risk of diabetes[22]. This method assigns the probability for each haplotype pair in each individual and then models an individual's phenotype as a function of each inferred haplotype pair, weighted by their estimated posterior probability so as to account for haplotype ambiguity. Haplo.glm function was used to calculate adjusted odds ratios for each haplotype. Haplo.glm is based on a generalized linear model and can be modeled for additive, dominant, or recessive effect of haplotypes[23,24]. We restricted the analysis to haplotypes with an inferred frequency of more than 0.02. The first haplotype, which was most frequent and was associated with the lowest serum CRP, was used as the reference group. To find the inheritance model (dominant, or recessive) that fits

Table 1 – Baseline characteristics of all participants and cases with incident type II diabetes

Characteristics	All participants*	Incident type II diabetic	P value†
Number	5901	544	-
Age, mean (SD), y	69.1 (9.1)	68.7 (8.1)	0.01
Men, (%)	2388 (40.5)	274 (43.5)	0.13
Body mass index, mean (SD), kg/m2	26.2 (3.7)	28.2 (3.8)	< 0.001
Waist circumference, mean (SD), cm	90.2 (11.1)	95.0 (11.0)	< 0.001
Systolic blood pressure, mean (SD), mm Hg	138.5 (22.1)	143.2 (20.6)	< 0.001
Diastolic blood pressure, mean (SD), mm Hg	73.8 (11.6)	75.6 (11.2)	< 0.001
Hypertension, (%)	1872 (32.3)	253 (46.4)	< 0.001
History of coronary artery disease, (%)	669 (12.1)	64 (13.7)	0.28
Total cholesterol, mean (SD), mmol/L	6.6 (1.2)	6.6 (1.2)	0.80
HDL cholesterol, mean (SD), mmol/L	1.4 (0.36)	1.3 (0.33)	< 0.001
Alcohol Intake, median (Interquartile range), g/d**	7.1 (1.5 – 19.3)	6.2 (1.4 – 17.7)	0.47
Current smoker, (%)	1320 (22.9)	131 (24.4)	0.22
Former smoker, (%)	2425 (42.1)	235 (43.8)	
Postmenopausal hormone therapy, (%)‡	1187 (20.1%)	102 (18.9)	0.47
C-reactive protein, median (Interquartile range), mg/L	1.8 (0.9 – 3.5)	2.3 (1.3 – 4.2)	< 0.001

* Contains both healthy participants and incident type 2 diabetic patients

†Cases are compared wit non-diabetic participants. T test for normally distributed continuous variables, Mann-Whitney for non-normally distributed covariates and Chi-square test for categorical variables

‡ History of use in women

**In drinkers

Table 2 – Hazard ratios for diabetes according to level of C-reactive protein

Quartile (level in mg/l)	Participants (cases)	HR (95% CI)		
		Model 1	Model 2	Model 3
1 (< 0.88)	1463 (77)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
2 (0.88 – 1.8)	1485 (141)	1.88 (1.42 – 2.48)	1.59 (1.19 – 2.11)	1.51 (1.13 – 2.03)
3 (1.81 – 3.5)	1475 (153)	2.16 (1.64 – 2.84)	1.70 (1.27 – 2.27)	1.54 (1.15 – 2.06)
4 (> 3.5)	1478 (173)	2.83 (2.16 – 3.70)	1.94 (1.45 – 2.59)	1.73 (1.29 – 2.33)
P for trend		< 0.001	< 0.001	< 0.001
Per SD Ln (CRP)	5901 (544)	1.41 (1.29 – 1.54)	1.24 (1.12 – 1.37)	1.19 (1.07 – 1.31)

Model 1: adjusted for age and sex

Model 2: adjusted for age, sex, BMI and waist circumference

Model 3: adjusted for age, sex, BMI, waist circumference, systolic and diastolic blood pressure and HDL-cholesterol

HR = Hazard ratio, SD = Standard deviation, CRP = C-reactive protein

best to the data, we used the Likelihood Ratio test (LR test) carried out on the variation of the log likelihood between two models.

Results

Serum CRP and diabetes in the Rotterdam Study

We compared 5901 participants who had CRP measurements with those 1034 participants whose serum CRP measurements were missing. Compared with the population used for analysis, participants with missing values were significantly older and more frequently female but had similar mean values for BMI, weight, waist circumference, cholesterol level, systolic and diastolic blood pressure, and daily alcohol consumption (data not shown). Table 1 shows the baseline characteristics of 5901 participants and individuals with incident diabetes. CRP ranged from 0.2 mg/l to 247 mg/l with a right skewed distribution. Median CRP was 1.86 mg/l in men and 1.78 mg/l in women (P<0.001). Age, BMI, weight, waist circumference, systolic and diastolic blood pressure, and HDL-cholesterol were significantly correlated with serum CRP. Except for HDL-cholesterol, the correlations were positive. The highest correlation coefficient was 0.27 for waist circumference.

During a mean follow-up of 9.8 years in 5901 participants, diabetes developed in 544 (incidence = 9.4 per 1000 person years). The age and sex adjusted HR for diabetes per 1 SD increase in Ln (CRP) was 1.41 (95% CI1.29 – 1.54). The HR attenuated to 1.24 (95%CI = 1.12 - 1.37) after adjustment for BMI and waist circumference. After further adjustment for systolic blood pressure, diastolic blood pressure, and HDL cholesterol, the HR slightly decreased to 1.19 (95%CI = 1.07 - 1.31) (Table 2). Considering the lowest CRP quartile as the reference

group, age and sex adjusted HRs were 1.88, 2.16 and 2.83 for the second, third and fourth quartiles of CRP respectively (P for trend < 0.001). Further adjustment for the above-mentioned covariates attenuated the HRs in model 2 and 3 (Table 2).

Sensitivity analysis

Exclusion of subjects with a history of hormone therapy had a minimal effect on the association. In the age and sex adjusted model, the HR for 1 SD increase in CRP attenuated to 1.37 (95% CI = 1.25 – 1.51). To evaluate the effect of adjustment for fasting blood sugar, we selected the third periodical examination of the Rotterdam Study as the baseline measurement. The mean follow-up time reduced to 5.34 years and the number of incident diabetes cases to 319 individuals. Serum CRP was significantly associated with diabetes in an age and sex adjusted model (HR for 1 SD Ln (CRP)=1.53; 95% CI=1.26-1.86). When we additionally adjusted the model for BMI, waist circumference, and fasting blood sugar the association attenuated (HR for 1 SD Ln (CRP)=1.39; 95% CI=1.13-1.71).

Meta-analysis

Ten studies were included in the meta-analysis. The studies had been conducted in the US, Europe, and Mexico. All studies showed a positive association between serum levels of CRP and incident diabetes. For the CRP intervals 0.5 - 1.8 mg/l, 1 - 3.7 mg/l, and 2.6 mg/l or higher, weighted age and sex adjusted risk ratios (95%CI) were 1.63 (1.35 - 1.98), 2.16 (1.81 - 2.57), and 4.00 (2.83 - 5.65), respectively. After additional adjustment for BMI, the weighted risks decreased to 1.44 (1.16 - 1.78), 1.72 (1.42 - 2.08), and 2.37 (1.57 -

Table 3 – Published studies on serum CRP and risk of diabetes

Study name	Study design	Diabetes / Non-diabetes[*]
Cardiovascular Heart Study (CHS)	Follow-up	45 / 4436
Women's Health Study (WHS)	Nested Case Control	188 / 362
West of Scotland Coronary Prevention Study (WOSCOPS)	Follow-up	127 / 5118
Japanese Americans Study	Follow-up	122 / 825
Nurses' Health Study (NHS)	Nested case control	737 / 785
Mexico City Diabetes Study	Follow-up	86 / 1158
Insulin Resistance Atherosclerosis Study (IRAS)	Follow-up	144 / 903
Atherosclerosis Risk in Communities Study (ARIC)	Case Cohort	581 / 572
Monica Augsburg Cohort Study (MONICA)	Follow-up	101 / 1951

Non-diabetes states the number of controls in the case-control studies. In cohort studies, it does not contain the diabetic cases

Table 4 – CRP haplotypes and their frequencies in the Rotterdam study and published studies

The Rotterdam Study		Miller et al.[25]		Carlson et al.[12]		Timpson et al.[26]	
Haplotype	Freq.	Haplotype	Freq.	Haplotype	Freq.	Haplotype	Freq.
H1 (CTC)*	0.33	H4	0.07	H1	0.06	GGT	0.07
		H3	0.27	H2	0.28	CGT	0.26
H2 (TCC)*	0.32	H2	0.29	H4	0.28	CAC	0.30
H3 (CCC)*	0.29	H1	0.27	H5	0.29	CGC	0.37
H4 (CCG)*	0.06	H5	0.06	H7	0.06		

* Coding from 1184C/T, 2042C/T and 2911C/G respectively

Freq. = Frequwency

3.58) (Figure 2, 3). Figure 3 shows that the association was more pronounced in Caucasians than other ethnic groups, although the slope was not significant in our regression (b= 0.01, 95%CI= -0.05 – 0.03).

CRP gene haplotypes, serum CRP and diabetes in the Rotterdam Study

We compared 6157 participants who had CRP genotypes with the 1826 participants whose CRP gene information were missing. Participants with missing values were older (5.3 years) and more frequently female. They had a lower weight (1.5 kg) and higher HDL cholesterol level (0.03 mmol/l). However, compared with the population used for analysis, there was no difference in their BMI, waist circumference, systolic and diastolic blood pressure, daily alcohol consumption and serum CRP level.

Genotype distributions of the three tagging SNPs were found to be in Hardy-Weinberg equilibrium. We estimated six allele-specific haplotypes with the Haplo.Stats program. Two of the haplotypes were present in <0.001% of the chromosomes and were therefore not used in our analyses. We coded the other 4 haplotypes from 1 to 4 according to decreasing population frequency. (Table 4)

Geometric mean of serum CRP level was 1.50 mg/l for carriers of haplotype 1. Mean serum CRP increased per copy of other haplotypes relative to haplotype 1. This increase was 0.3 mg/l for haplotype 2, 0.15 mg/l for haplotype 3 and 0.46 mg/l for haplotype 4. Serum CRP level was significantly higher in participants with haplotype 4 compared to the carriers of haplotype 1 (P value < 0.001). Elimination of the CRP levels over 10 mg/l did not materially change the results (data not shown). A nearly significant higher risk for diabetes was found for carriers of haplotype 4 compared to carriers of haplotype 1. In an additive model the odds ratio was 1.30 (95%CI = 0.99 – 1.71). Based on the log likelihood ratio test, none of the inheritance models (dominant or recessive) improved the fit of the model significantly. Other haplotypes did not change the risk of diabetes significantly (Table 5).

After adjustment for BMI, waist circumference, systolic and diastolic blood pressure, and HDL-cholesterol, the association increased in strength (OR = 1.45; 95%CI = 1.08 – 1.96). When the age and sex adjusted estimate was further adjusted for serum CRP, the odds ratio decreased (OR = 1.20; 95%CI = 0.89 – 1.60). Using the Mendelian randomization approach, the expected odds ratio for carriers of haplotype 4 compared to carriers of haplotype 1was 1.09 and was not significantly different from the observed odds ratio.

Discussion

The results of our population based cohort study and the meta-analysis showed that serum CRP is associated with risk of diabetes independently of obesity. We identified a genetic variant in the human CRP locus that associates with a high serum CRP and an increased risk of diabetes. The latter association reduced after adjustment for serum CRP. These findings support the hypothesis that CRP is etiologically involved in the pathogenesis of diabetes.

In our study including 544 cases with incident diabetes, we showed that the association of serum CRP with diabetes remains significant not only after adjustment for obesity indexes but also after adjustment for blood pressure and cholesterol. We adjusted for the latter variables as surrogates of the metabolic syndrome, which could confound the association. Adjustment for obesity indexes in other studies has provided controversial results. The Cardiovascular Heart study (CHS), the Women's Health Study, the West of Scotland Coronary Prevention Study, and the Nurses' Health Study (NHS) showed significant associations between CRP and incident diabetes even after adjustment for obesity indexes[1,5,7,9]. In contrast, the Atherosclerosis Risk in Communities Study (ARIC), the Monica Augsburg Cohort Study, and the Insulin Resistance Atherosclerosis Study (IRAS) showed non-significant associations after adjustment for obesity indexes[3,8,9]. To summarize these controversial results, we

Table 5 – Odds ratios for diabetes in different CRP haplotypes

CRP haplotypes	Number of alleles (diabetic alleles)	Age and sex adjusted	Multivariate adjusted *	Age, sex and CRP adjusted
Haplotype 1 (ref)	3490 (317)	1	1	1
Haplotype 2	3340 (327)	1.09 (0.92 – 1.28)	1.10 (0.92 – 1.32)	1.07 (0.90 – 1.27)
Haplotype 3	3198 (308)	1.06 (0.90 – 1.25)	0.99 (0.82 – 1.19)	1.03 (0.86 – 1.23)
Haplotype 4	624 (72)	1.30 (0.99 – 1.71)	1.45 (1.08 – 1.96)	1.20 (0.89 – 1.60)

Adjustments were done for age, sex, BMI, waist circumference, systolic and diastolic blood pressure, and HDL-cholesterol
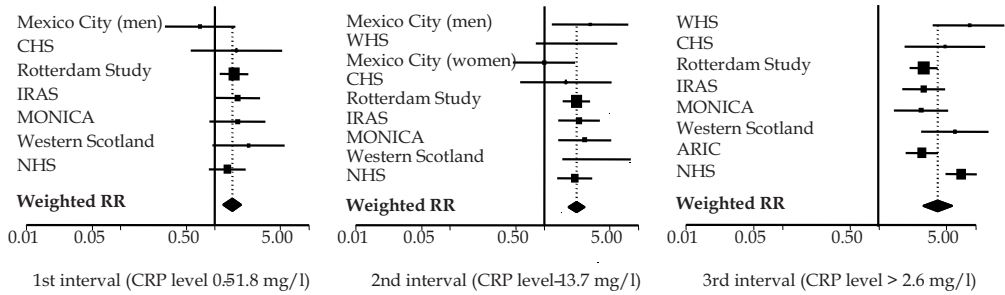
Figure 1.

Age and sex adjusted risk ratios for diabetes in different categories of CRP levels compared to the reference category (<0.5mg/l) - Weighted RR (95% CI) was 1.64 (1.35 - 1.98) for the first interval, 2.16 (1.81 - 2.57) for the second interval, and 4.00 (2.83 - 5.65) for the third interval.
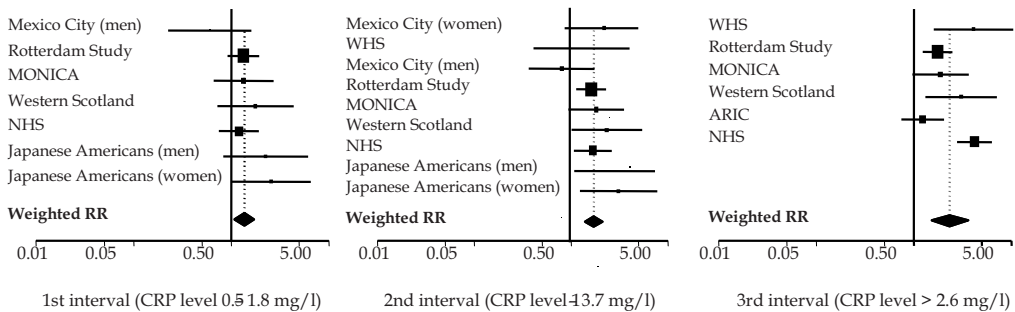
Figure 2.

BMI, age and sex adjusted risk ratios for diabetes in different categories of CRP levels compared to the reference category (<0.5mg/l) - Weighted RR (95% CI) was 1.44 (1.16 - 1.78) for the first interval, 1.72 (1.42 - 2.08) for the second interval, and 2.37 (1.57 - 3.58) for the third interval.

performed a meta-analysis. Weighted risk ratios showed a significant obesity-adjusted association between serum CRP and diabetes. We believe that the negative result in the three latter papers could be explained by several factors. The IRAS and the Monica Study had fewer diabetic cases and consequently had less power. In addition, in the ARIC Study, nearly one third of the participants were non-Caucasian. We observed in our meta-analysis that, although not significant, the association was more pronounced in Caucasians than other ethnic groups (figure 3).

We showed that serum CRP is significantly different in carriers of different haplotypes. Several studies have used haplotypes describing the total variation of the CRP gene to examine the issue. Miller et al resequenced 192 individuals

to identify a comprehensive set of common SNP variants. Later, they studied the association of the gene variation with serum CRP level in subsets of three cohorts. The haplotypes that Miller et al found is comparable to ours[25]. In our study, three of the haplotypes are the same. Haplotype 3 and 4 constitute our haplotype 1 and we have not identified their haplotype 6, which was present in 2.1% of their population. Miller et al found their haplotype 2 and 5 to be associated with higher CRP and their haplotype 1, 3 and 4 to be associated with lower CRP. This is in agreement with our findings. Carlson et al also defined all common genetic variation across the human CRP locus by resequencing the region in a multiethnic variation discovery panel, and selected haplotype tagging SNPs for genotyping in a larger panel (CARDIA study). Furthermore, they investigated the associations between common haplotypes and serum CRP, rendering significant results[12]. Although they used an approach similar to our study, Carlson et al investigated a mixed population, partly European and partly African-American, while the Rotterdam Study cohort is of almost exclusively European descent. As a result, Carlson et al found three haplotypes, which were not present in the European descent populations. The other haplotypes identified by Carlson et al were similar to the haplotypes we defined in the Rotterdam Study[12]. Timpson et al selected four SNPs based on published reports. They constructed 4 haplotypes by use of a genetic data analysis program, named SIMHAP. Their haplotypes (CGT, GGT, CAC, CGC) were close to the haplotypes that we used in our study. (Table 4) Timpson et al found a significant association between haplotypes and serum CRP. Their results were comparable to those of our study[26]. Several other studies observed a relation between genetic variation in the human CRP locus and serum CRP. For instance, Szalai et al constructed their haplotypes based on the bi-allelic –409G/A (rs3093032) and tri-allelic –390C/T/A (rs3091244) CRP gene promoter polymorphisms. Interesting to note is that these haplotypes affect transcription

Figure 3.
The percent of population comprised of white ethnicity and the risk ratio comparing highest with the lowest category of CRP in the meta-analysis
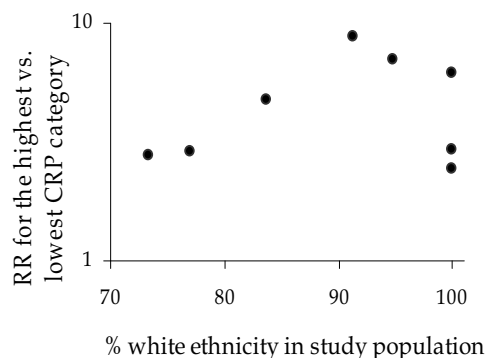
factor binding, alter transcription activity, and influence the variation of the serum CRP[27]. The T and A allele of the tri-allelic –390C/T/A polymorphism is present in participants with haplotype 2 and 4 in our study[13]. Thus, these functional SNPs may partly explain the higher serum CRP in carriers of these two haplotypes in our study. Other studies that examined the association of CRP gene haplotypes and serum CRP are not comparable with our study since they used different SNPs to build up their haplotypes, or used different populations in terms of health status, race or age[28-30].

We showed that participants carrying haplotype 4 have a significantly higher risk of diabetes. Thus, genetic susceptibility to high serum CRP increases the risk of diabetes. Our findings show that the association of genetic susceptibility with diabetes was independent of BMI and waist circumference, suggesting that obesity indexes do not explain the genetic susceptibility. Moreover, in a model adjusted for age, sex, and CRP, the association diminished suggesting that the effect of haplotype 4 is likely to be explained by the variation in serum CRP level.

We found no significant difference between the observed odds ratio and the one calculated based on Mendelian randomization. This provides evidence on an independent role of CRP in developing diabetes. The Mendelian randomization approach is a tool to assess the nature of associations. Since gene alleles that influence the intermediate phenotype are inherited at random, potential confounders for the association will be evenly distributed in those who do, and those who do not have the alleles. Consequently, any difference between these two groups should be free of confounding by environmental factors. Furthermore, regression dilution and reverse causation are not probable to occur since the genotype is constant over time and is determined before the onset of disease[31]. However, this approach has certain limitations such as the potential for confounding the gene-phenotype association by linkage disequilibrium with other genes. Similarly, population origin can confound the gene-disease association[32]. In addition, pleiotropic effects of the SNP in more than one biological pathway can violate one of the assumptions of Mendelian randomization.

Studies that found independent association between CRP and diabetes suggested various pathways. Many studies argued that the association reflects the effects of cytokines, such as IL-6 and TNFa on insulin resistance[2,5,8,33,34]. Some others explained the association through oxidative stress or innate immune system[6,7]. Nevertheless, none of the proposed mechanisms provide a causal role for CRP. A recent study, which investigated the association of CRP and metabolic syndrome pointed to the direct harmful effects of CRP on vessel walls, which may alter endothelial permeability and eventually lead to

insulin resistance[35]. However, further studies are necessary to find a reasonable mechanism.

To our knowledge, this is the first study in which the association between diabetes and serum CRP was partly explained by variations in the CRP gene. Wolford et al. showed that variation within the CRP locus might play a role in diabetes susceptibility in Pima Indians[36]. Their findings, in line with our study, were consistent with the hypothesis that CRP may play an etiologic role in the development of diabetes. In a recent study, Timpson et al. investigated the causal role of CRP in development of metabolic syndrome and reported that their findings provide evidence for CRP not to be causally involved in the pathogenesis of the metabolic syndrome. We believe that their study is not in contrast with our study. Timpson et al. combined haplotype 4 and haplotype 3 as the CGC haplotype, while we found an association exclusively with haplotype 4. Haplotype 4 is a rare allele and grouping it with another common haplotype, dilutes the effect[26]. Furthermore, diabetes has a well-defined nature compared to the metabolic syndrome, which may have resulted a stronger association. However, replication of our finding is necessary to establish the relationship.

The strengths of our study include a large sample size, a long follow-up period, a considerable number of incident diabetes cases, and the availability of detailed genotype information. Haplotypes provide more information on genetic variation compared to single SNPs. In addition, we used information on the gene and the protein together in one study. However, several limitations need to be discussed. In the Rotterdam Study, we screened the cohort for prevalent diabetes at baseline by use of a non-fasting glucose level and OGTT. Our baseline measurements revealed 10.8 percent of prevalent diabetes that is similar to the expected prevalence of diabetes in our population[37]. The studies included in the meta-analysis all used different categories of CRP and the geometric mean of the CRP level in the reference category ranged from 0.23 to 1.41 mg/l. To group the risk ratios, we allocated the risk estimates to 3 different CRP intervals but variation in mean CRP level within categories remained. However, we believe that this minor variation did not result in a sizeable under or overestimation of our results.

In conclusion, our meta-analysis showed that serum CRP is a risk factor for diabetes, independent of obesity. Furthermore, genetic variation in CRP was associated with the level of CRP and the risk of diabetes. These results support the hypothesis that CRP plays a role in the pathogenesis of diabetes.

References

1. Barzilay JI, Abraham L, Heckbert SR, Cushman M, Kuller LH, Resnick HE, Tracy RP. The Relation of Markers of Inflammation to the Development of Glucose Disorders in the Elderly: The Cardiovascular Health Study. Diabetes. 2001;50:2384-2389.
2. Duncan BB, Schmidt MI, Pankow JS, Ballantyne CM, Couper D, Vigo A, Hoogeveen R, Folsom AR, Heiss G. Low-Grade Systemic Inflammation and the Development of Type 2 Diabetes: The Atherosclerosis Risk in Communities Study. Diabetes. 2003;52:1799-1805.
3. Festa A, D'Agostino R, Jr., Tracy RP, Haffner SM. Elevated levels of acute-phase proteins and plasminogen activator inhibitor-1 predict the development of type 2 diabetes: the insulin resistance atherosclerosis study. Diabetes. 2002;51:1131-7.
4. Han TS, Sattar N, Williams K, Gonzalez-Villalpando C, Lean ME, Haffner SM. Prospective study of C-reactive protein in relation to the development of diabetes and metabolic syndrome in the Mexico City Diabetes Study. Diabetes Care. 2002;25:2016-21.
5. Hu FB, Meigs JB, Li TY, Rifai N, Manson JE. Inflammatory Markers and Risk of Developing Type 2 Diabetes in Women. Diabetes. 2004;53:693-700.
6. Nakanishi S, Yamane K, Kamei N, Okubo M, Kohno N. Elevated C-Reactive Protein Is a Risk Factor for the Development of Type 2 Diabetes in Japanese Americans. Diabetes Care. 2003;26:2754-2757.
7. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. Jama. 2001;286:327-34.
8. Thorand B, Lowel H, Schneider A, Kolb H, Meisinger C, Frohlich M, Koenig W. C-Reactive Protein as a Predictor for Incident Diabetes Mellitus Among Middle-aged Men: Results From the MONICA Augsburg Cohort Study, 1984-1998. Arch Intern Med. 2003;163:93-99.
9. Freeman DJ, Norrie J, Caslake MJ, Gaw A, Ford I, Lowe GDO, O'Reilly DSJ, Packard CJ, Sattar N. C-Reactive Protein Is an Independent Predictor of Risk for the Development of Diabetes in the West of Scotland Coronary Prevention Study Diabetes. 2002;51:1596-1600.
10. Trayhurn P, Beattie JH. Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ. Proc Nutr Soc. 2001;60:329-39.
11. Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH, Tracy RP. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis. 2001;154:681-9.
12. Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, Liu K, Williams OD, Iribarren C, Lewis EC, Fornage M, Boerwinkle E, Gross M, Jaquish C, Nickerson DA, Myers RM, Siscovick DS, Reiner AP. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. Am J Hum Genet. 2005;77:64-77.
13. The Seattle SNPs Program for Genomic Applications [online]. Available from http://pga.gs.washington.edu. Accessed 1 October 2005
14. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991;7:403-22.

15. Stolk RP, Hoes AW, Pols HA, Hofman A, de Jong PT, Lamberts SW, Grobbee DE. Insulin, hypertension and antihypertensive drugs in elderly patients: the Rotterdam Study. J Hypertens. 1996;14:237-42.

16. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care. 1997;20:1183-97.

17. WHO. Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications: Part 1: Diagnosis and Classification of Diabetes Mellitus. In. Geneva, Switzerland: World Health Organization; 1999.

18. Schwarzer G: The Meta Package. 8 March2005 . Available from http://www.cran.r-project.org/src/contrib/Descriptions/rmeta.html. Accessed May 2006

19. Lumley T: The rmeta package. 23 November2004 . Available from http://www.cran.r-project.org/src/contrib/Descriptions/rmeta.html. Accessed May 2006

20. Ihaka R GR. R: a language for data analysis and graphics. J Comput Graph Stat. 1996;5:299–314.

21. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002;21:589-624.

22. Schaid DJ: HaploStats. Rochester, MN, Mayo Clinic/Foundation,2005 . Available from http://mayoresearch.mayo.edu/mayo/research/biostat/schaid.cfm. Accessed January 2006

23. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet. 2002;70:425-34.

24. Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, Schaid DJ. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. Hum Hered. 2003;55:56-65.

25. Miller DT, Zee RY, Suk Danik J, Kozlowski P, Chasman DI, Lazarus R, Cook NR, Ridker PM, Kwiatkowski DJ. Association of common CRP gene variants with CRP levels and cardiovascular events. Ann Hum Genet. 2005;69:623-38.

26. Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ, Hattersley AT, Ebrahim S, Lowe GD, Rumley A, Davey Smith G. C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. Lancet. 2005;366:1954-9.

27. Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A, Zakharkin SO, George V, Allison DB, Cooper GS, Xie F, Fan Z, Edberg JC, Kimberly RP. Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. J Mol Med. 2005;83:440-7.

28. Chen J, Zhao J, Huang J, Su S, Qiang B, Gu D. -717A>G polymorphism of human C-reactive protein gene associated with coronary heart disease in ethnic Han Chinese: the Beijing atherosclerosis study. J Mol Med. 2005;83:72-8.

29. Russell AI, Cunninghame Graham DS, Shepherd C, Roberton CA, Whittaker J, Meeks J, Powell RJ, Isenberg DA, Walport MJ, Vyse TJ. Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. Hum Mol Genet. 2004;13:137-47.

30. Kovacs A, Green F, Hansson LO, Lundman P, Samnegard A, Boquist S, Ericsson CG, Watkins H, Hamsten A, Tornvall P. A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. Atherosclerosis. 2005;178:193-8.

31. Hingorani A, Humphries S. Nature's randomised trials. Lancet. 2005;366:1906-8.
32. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. Am J Epidemiol. 2006;163:397-403.
33. Ford E. Body mass index, diabetes, and C-reactive protein among U.S. adults. Diabetes Care. 1999;22:1971-1977.
34. Krakoff J, Funahashi T, Stehouwer CDA, Schalkwijk CG, Tanaka S, Matsuzawa Y, Kobes S, Tataranni PA, Hanson RL, Knowler WC, Lindsay RS. Inflammatory Markers, Adiponectin, and Risk of Type 2 Diabetes in the Pima Indian. Diabetes Care. 2003;26:1745-1751.
35. Lim S, Lee HK, Kimm KC, Park C, Shin C, Cho NH. C-reactive protein level as an independent risk factor of metabolic syndrome in the Korean population. CRP as risk factor of metabolic syndrome. Diabetes Res Clin Pract. 2005;70:126-33.
36. Wolford JK, Gruber JD, Ossowski VM, Vozarova B, Antonio Tataranni P, Bogardus C, Hanson RL. A C-reactive protein promoter polymorphism is associated with type 2 diabetes mellitus in Pima Indians. Molecular Genetics and Metabolism. 2003;78:136-144.
37. Harris MI, Eastman RC, Cowie CC, Flegal KM, Eberhardt MS. Comparison of diabetes diagnostic categories in the U.S. population according to the 1997 American Diabetes Association and 1980-1985 World Health Organization diagnostic criteria. Diabetes Care. 1997;20:1859-62.

Chapter 2.2

Risk of type II diabetes mellitus attributable to C-reactive protein and other risk factors

Abstract

Objective

To calculate the population attributable risk (PAR) of C-reactive protein (CRP) and other risk factors for type II diabetes.

Research design and methods

The Rotterdam Study is a population-based, prospective follow-up study among 7983 participants aged 55 years and older. Risk factors including serum CRP were determined at baseline. Participants with diabetes at baseline were excluded and the cohort was followed for a mean of 10.8 years. The hazard ratio (HR) and the PAR for diabetes were computed for all studied risk factors.

Results

Serum CRP >1mg/l (HR=1.67, PAR=0.33), body mass index > 25 kg/m^2 (HR=2.51, PAR=0.51), waist circumference > 102 for men and > 88 for women (HR=1.36, PAR=0.14), current smoking (HR=1.16, PAR=0.03), age > 65 years (HR=1.35, PAR=0.15), and family history of diabetes (HR=1.87, PAR=0.16), were related to diabetes and contributed to the risk of the disease. Serum CRP was a greater contributor to the risk of diabetes in women than in men (PAR values of 0.37 versus. 0.28, respectively). Age, and current smoking PARs were not statistically significantly contributing to the risk of diabetes in women. Combined PAR was 0.80 (95% CI: 0.74, 0.85) for all six studied risk factors and 0.71 (95% CI: 0.64, 0.78) for modifiable risk factors (serum CRP, BMI, waist circumference, and current smoking).

Conclusion

High CRP is one of the major contributors to the risk of type II diabetes. The contribution of modifiable risk factors to the risk of diabetes is considerable.

Introduction

There is a growing body of evidence that low-grade systemic inflammation enhances the risk of type II diabetes mellitus[1]. Furthermore, anti-inflammatory medication may prevent diabetes or delay the onset of the disease[2]. Whether inflammation is a major contributor to the risk of diabetes is not yet clear.

To judge the public health impact of different risk factors, the population attributable risk (PAR) is a relevant measure[3]. The PAR of a risk factor for a disease is the proportion of those with the disease that is due to that risk factor. The PAR depends on both the relative risk estimate and the prevalence of the risk factor.

C-reactive protein (CRP), a marker of inflammation, is independently associated with the development of diabetes[1,4,5] and can be reduced by the use of anti-inflammatory medications[6]. Therefore, the PAR of serum CRP for diabetes can be used to estimate the contribution of inflammation to the risk of diabetes.

To our knowledge there is no previously published study on the PAR of high serum CRP for diabetes. We sought to quantify the contribution of a number of risk factors including serum CRP to the risk of diabetes in the Rotterdam Study, a large population-based prospective cohort study in Caucasians 55 years or over.

Methods

Study Population

The study was conducted within the framework of the Rotterdam Study, an ongoing prospective, population-based cohort study on determinants of a number of chronic diseases. The Rotterdam Study has been described in detail elsewhere[7]. In brief, all inhabitants of Ommoord, a district of Rotterdam in the Netherlands who were 55 years or over, were invited to participate in this study. Of all 10275 eligible individuals, 7983 agreed to participate (78%).

The baseline examinations took place from 1990-1993. Follow-up for clinical events started at baseline and follow-up examinations were carried out periodically in 1995–1996, 1997–1999, and 2000-2005. In addition, participants were continuously monitored for major events through automated linkage with files from general practitioners and pharmacies working in the study district of Ommoord. Information on vital status was obtained regularly from municipal health authorities in Rotterdam. For the present study, follow-up data were available until October 1, 2005. Written informed consent was obtained from all participants and the Medical Ethics Committee of the Erasmus Medical Center approved the study.

Serum CRP

High sensitivity CRP was measured in non-fasting serum samples kept frozen at -20 °C by use of Rate Near Infrared Particle Immunoassay (Immage® Immunochemistry System, Beckman Coulter, USA). This method has been described in more detail elsewhere[1]. Serum samples were stored for approximately 10 years at −20°C until the measurements were carried out in 2003-2004. We compared these CRP measurements with CRP measurements in the serum samples stored at −80°C in a random sample of 29 participants. The Spearman correlation coefficient was 0.99 between the CRP serum level measurements carried out on samples kept in −20°C and −80°C (P<0.001).

Diabetes

At baseline, participants were defined as prevalent cases with type II diabetes, when they had non-fasting glucose > 11.1 mmol/l, oral glucose tolerance test (OGTT) > 11.1 mmol/l, or when they were using anti-diabetic medication.

Incident cases of type II diabetes were diagnosed based on fasting plasma glucose level >= 7.0 mmol/l or random (non-fasting) plasma glucose level >= 11.1 mmol/l or use of oral anti-diabetic medication or use of insulin or treatment by diet and registered by a general practitioner as having diabetes [1].

Population for analysis

We excluded 861 prevalent diabetic participants and 187 participants who did not provide any information on their glucose levels at baseline. The population for analysis consisted of 6935 participants. Of these, serum CRP level was available in 5901, BMI in 6136, waist circumference in 5837, and smoking status in 6765 of participants.

Statistical Analysis

High serum CRP[1,4], overweight[8], truncal fat distribution[9,10], physical inactivity[11], smoking[12-14], aging, and family history of diabetes[15] have been reported as risk factors for diabetes. Established cutoff points were used to dichotomize continuous covariates into normal and elevated levels. On this basis, serum CRP >= 1 mg/l, BMI >= 25 kg/m^2, waist circumference >= 102 cm for men and >= 88 cm for women, and age >= 65 years, were considered as risk factor for diabetes. Smoking was assessed as current smoking versus non-smoking, and family history of diabetes was considered positive in the presence of diabetes in parents, children or any of the siblings. A Cox regression analysis was used to investigate the association of risk factors with incidence of diabetes.

Population attributable risks (PAR) and 95% confidence intervals were calculated by the use of Interactive Risk Assessment Program (IRAP) developed

by Dr Mitchell Gail (US national Cancer Institute 2002)[16-19]. A PAR adjusted for confounding is estimated by

$$PAR = 1 - \sum_{i=1}^{I} \sum_{j=1}^{J} \rho_{ij} R_{i|j}^{-1}$$

where the relative risk is

$$R_{iJ} = \frac{PR(D = 1 | X = \chi_i, C = c_j)}{Pr(D = 1 | X = \chi_1, C = c_j)}$$

and

$$\rho_{ij} = Pr(X = \chi_i, C = c_j | D = 1)$$

given D=1 denoting presence of disease, X denoting exposure with I levels , and C denoting a confounder C with J levels . The relative risk is estimated from a multivariate Poisson model[18]. The bootstrap procedure was used to estimate the variance and 95% confidence interval of the PAR[19].

The PAR for a combination of risk factors corresponds with the proportion of the disease that can be attributed to any of the studied risk factors. The combined PAR is not a simple product of summing up the single PARs. A diseased case can simultaneously be attributed to more than one risk factor. As a result, the fraction of the population that is attributed to or prevented by each risk factor overlaps with other risk factors. Hence, the combined PAR is usually lower than the sum of individual PARs.

To estimate the proportion of the disease that is exclusively attributed to a specific risk factor, we calculated the combined PAR in the presence and absence of this risk factor. The difference is the so-called "extra attributable risk" which indicates the proportion of the disease that can be attributed exclusively to this specific risk factor[20].

To provide a similar study population for different analysis and to increase the statistical power, we imputed missing data using the expectation maximization method in SPSS 11.0, which is based on the correlations between each variable with missing values and all other variables.

Results

Table 1 shows the baseline characteristics of the studied population in tertiles of serum CRP.

During a mean follow-up time of 9.9 years (Interquartile range 6.5–13.2 years), diabetes developed in 645 persons (incidence rate 9.4 per 1000 person

years). Table 2 shows the proportion of the participants who were exposed to each risk factor and their association with risk of type II diabetes. BMI (> 25 kg/m^2), and family history of diabetes, were the strongest risk factors. High serum CRP (>1 mg/l) had a greater HR in women (1.77) than in men (1.42) and current smoking had a greater HR in men (1.37) than women (1.10). However, the differences between HRs were not significant. The association between age (>65years) and diabetes was stronger in men (HR = 1.64) than in women (HR = 1.15) and the difference between HRs was significant (p for interaction <0.05).

Multivariate adjusted PAR was 0.33 (95%CI: 0.21-0.46) for high serum CRP. The PAR of high serum CRP for diabetes was 0.17 (95% CI: 0.08-0.25) and 0.08 (95%CI: 0.02-0.15) when cutoff points of 2 mg/l, and 3 mg/l were used, respectively. Moreover, the PAR was 0.17 for the highest vs. the lowest and 0.32 for the top two tertiles vs. the lowest tertile of serum CRP. High BMI (>25 kg/m^2) was the main contributor to the risk of diabetes (PAR=0.51; 95%CI: 0.41-0.60) (table 3).

Collectively, studied risk factors contributed to 80% (95% CI: 74%-85%) of the risk of diabetes. Modifiable risk factors (serum CRP, BMI, waist circumference, and current smoking) contributed to 71% of the risk, suggesting that more than two third of incident diabetes cases might have been prevented if all the above risk factors were eliminated (table 4). Moreover, we estimated the combined PAR for modifiable risk factors in the absence of each risk factor to estimate the extra attributable risk. Exclusion of serum CRP decreased the combined PAR from 0.71 to 0.58 indicating that the extra attributable risk was 0.13 for high serum CRP (table 4).

Table 1 - Baseline characteristics of participants in different categories of serum CRP

Risk factor	Serum CRP			P value
	< 1mg/l	1 – 3 mg/l	> 3 mg/l	
Number	1717	2702	2516	-
Men (%)	40.5	64.5	57.2	< 0.001
Body mass index, means (SD), kg/m^2	24.9 (3.2)	26.5 (3.4)	26.9 (3.6)	< 0.001
Waist circumference, means (SD), cm	85.9 (10.4)	90.1 (10.2)	93.4 (9.5)	< 0.001
Current smoking (%)	16.6	19.5	28.7	< 0.001
Age, means (SD), years	67.3 (8.5)	68.5 (8.8)	72.9 (9.9)	< 0.001
Family history of diabetes (%)	21.3	21.3	19.8	0.37
HDL cholesterol, means (SD), mmol/L	1.44 (0.39)	1.36 (0.36)	1.29 (0.35)	< 0.001
Systolic blood pressure, means (SD), mm Hg	134.0 (21.4)	139.0 (21.7)	142.0 (22.3)	< 0.001
Diastolic blood pressure, means (SD), mm Hg	72.8 (11.3)	74.2 (11.3)	74.1 (11.9)	< 0.001
Hypertension (%)	23.5	33.7	42.7	< 0.001

Data are means±SD, and n (%)

Table 2 - Percent exposed and multivariate adjusted* hazard ratio (HR) of diabetes associated with risk factors

Risk factor	Exposed (%)		Hazard ratio (95% CI) for diabetes		
	Men	Women	All participants	Men	Women
CRP > 1 mg/L	74.8	75.9	1.67 (1.34 - 2.09)	1.42 (1.11 - 2.12)	1.77 (1.30 - 2.40)
BMI > 25 kg/m²	61.2	67.0	2.51 (2.00 - 3.16)	2.57 (1.86 - 3.56)	2.44 (1.76 - 3.39)
High waist circumference †	17.6	53.4	1.36 (1.14 - 1.63)	1.43 (0.97 - 1.68)	1.47 (1.14 - 1.88)
Current smoking	29.7	17.2	1.16 (0.96 - 1.40)	1.37 (0.93 - 1.56)	1.10 (0.84 - 1.44)
Age > 65 years	60.7	67.9	1.35 (1.14 - 1.59)	1.64 (1.26 - 2.11)	1.15 (0.92 - 1.42)
Family history of diabetes	18.8	22.0	1.87 (1.59 - 2.20)	1.86 (1.44 - 2.40)	1.88 (1.52 - 2.33)

* Multivariate adjusted model is adjusted for: C-reactive protein, body mass index, waist circumference, current smoking, age, and family history

† Waist circumference > 102 cm for men and > 88 cm for women

CRP: C-reactive protein; BMI: Body mass index

Discussion

In this study, we found that high serum CRP is a major contributor to the risk of type II diabetes independent of the other established risk factors. In addition, we observed that established risk factors account for a large proportion (80%) of the risk of type II diabetes in the general population more than 55 years old.

Our study underscores chronic inflammation, as a major contributor to the risk of diabetes, by showing that one third of the cases with diabetes are attributed to high serum CRP. Serum CRP, a marker of chronic low-grade inflammation, is a novel risk factor for diabetes. PAR is mostly estimated for

Table 3 - Multivariate adjusted* population attributable risks (PAR) and 95% confidence interval of different risk factors for diabetes

	All participants	Men	Women
CRP (3rd vs. 1st tertile)	0.17 (0.11 - 0.23)	0.16 (0.06 - 0.26)	0.18 (0.10 - 0.25)
CRP (2nd & 3rd vs. 1st tertile)	0.32 (0.22 - 0.42)	0.23 (0.8 - 0.39)	0.39 (0.26 - 0.530
C-reactive protein > 1 mg/l	0.33 (0.21 - 0.46)	0.28 (0.10 - 0.47)	0.37 9 0.20 - 0.53)
Body mass index > 25	0.51 (0.41 - 0.60)	0.50 (0.37 - 0.63)	0.51 (0.37 - 0.64)
High waist circumference †	0.14 (0.06 - 0.22)	0.07 (- 0.01 - 0.14)	0.22 (0.08 - 0.35)
Current smoking	0.03 (-0.01 - 0.07)	0.05 (- 0.02 - 0.13)	0.02 (-0.03 - 0.06)
Age > 65 years	0.15 (0.06 - 0.24)	0.25 (0.13 - 0.37)	0.06 (-0.07 - 0.19)
Family history of diabetes	0.16 (0.11 - 0.20)	0.15 (0.08 - 0.21)	0.16 (0.10 - 0.23)

* The model is adjusted for all present covariates: C-reactive protein, body mass index, waist circumference, current smoking, age, and family history

† Waist circumference > 102 cm for men and > 88 cm for women

the risk factors of which a causal role is evidenced. High serum CRP predicts diabetes and a growing body of evidence supports the causal role of CRP[1,2,4]. Hence, it would be logic to attribute a part of the risk of diabetes to chronic low-grade inflammation. However, estimation of PAR for a new risk factor when the causal role is not yet widely accepted illustrates the potential impact of the risk factor, were it later accepted to be causal[20].

Serum CRP is a marker of inflammation but is also closely related to adiposity. This may raise doubt whether CRP is a marker of inflammation or adiposity. We believe that even the variation of serum CRP, correlated with obesity, indicates an inflammatory state. The increased level of serum CRP in obese individuals is due to increased secretion of Il-6 and TNF-alpha in adipocytes, which regulate CRP production in hepatocytes and induce a chronic inflammatory state[21].

We adjusted the association for age, BMI and waist circumference as potential confounders. However, the covariates were dichotomized and dichotomization increases the likelihood of residual confounding. To estimate the magnitude of the residual confounding we introduced age, BMI and waist circumference as covariates with 10 categories to the model. Estimated PAR for high serum CRP slightly attenuated to 0.32 (95% CI = 0.20–0.45). Therefore, residual confounding by age and obesity in our findings should be trivial.

To obtain a reasonable estimate of the PAR, one should use a cutoff point that could be achieved in practice[22]. For serum CRP, however, no cutoff point has been recommended in relation to the risk of diabetes. The American Heart Association (AHA) suggests two cutoff points of 1 mg/l and 3 mg/l in relation to cardiovascular risk[23]. When we used the cutoff point of 1 mg/l to dichotomize serum CRP, 75% of our population was exposed, which may seem to be overestimating. However, where more than 61% of men and 67% of

Table 4 - Combined PAR (95% CI) of all modifiable risk factors* and combined of all risk factors with one of them deleted

Risk factor		All participants	Men	Women
Combined PAR		0.80 (0.74 - 0.85)	0.79 (0.71 - 0.87)	0.80 (0.73 - 0.87)
Modifiable risk factors*		0.71 (0.64 - 0.78)	0.65 (0.57 - 0.79)	0.75 (0.66 - 0.83)
Deleted factor	C-reactive protein > 1 mg/l	0.58 (0.50 - 0.66)	0.55 (0.43 - 0.67)	0.61 (0.51 - 0.71)
	Body mass index > 25	0.44 (0.33 - 0.55)	0.39 (0.22 - 0.55)	0.51 (0.37 - 0.66)
	High waist circumference †	0.68 (0.59 - 0.76)	0.66 (0.54 - 0.78)	0.69 (0.58 - 0.80)
	Current smoking	0.70 (0.63 - 0.84)	0.67 (0.55 - 0.78)	0.74 (0.66 - 0.83)

* CRP, BMI, waist circumference, and current smoking are entered to the model. The results are also adjusted for age and family history of diabetes.

† Waist circumference > 102 cm for men and > 88 cm for women

women were overweight or obese, it is not too far to consider serum CRP, which is highly correlated with BMI, to be high in 75% of our population in regard to diabetes.

A disease can simultaneously be attributed to or prevented by more than one risk factor. Therefore, the fractions of the disease, which are attributed to different risk factors, overlap with each other and cannot be simply summed up. To estimate the proportion of the disease that is attributed to a certain number of risk factors, combined PAR should be estimated. Our combined PAR showed that the majority of diabetes cases are preventable. This finding is in agreement with other studies. Hu and colleagues reported that 91% of diabetes cases in women can be attributed to overweight, poor diet, lack of exercise, smoking, and abstinence from alcohol[24]. Hu and colleagues studied diet and physical activity, which were not present in our study and their study was restricted to women. These may explain why they found a slightly higher estimate for the combined PAR. However, they did not study any marker of inflammation.

Extra attributable risk was 0.13 for high serum CRP. This should not be confused with the single adjusted PAR, which was 0.33 for high serum CRP. Single PAR indicates the fraction of cases that can be prevented by lowering serum CRP, assuming that the other risk factors remain unchanged. However, extra attributable risk suggests that if a hypothetical prevention program has eliminated all other studied risk factors, lowering serum CRP still can prevent 13% of incident diabetes cases. The difference between the single PAR and the extra attributable risk is due to those cases that were alternatively attributed to high serum CRP and other risk factors. These risk factors may act in the same pathway with CRP, leading to the development of diabetes. For instance, recent studies suggest that at least a part of the association of obesity[4] and smoking[25,26] with diabetes may be through low-grade chronic inflammation.

Caution should be taken in interpreting the PAR in practice. In computing PAR, we assume that all participants who are labeled as exposed will shift to the non-exposed group without causing any change in the risk factor distribution in the non-exposed group. Moreover, we assume that the risk of the disease decreases instantly after the intervention. In practice, however, the effect of an intervention is likely to be different. Firstly, a part of the population succeeds to modify the risk factor but cannot avoid it. Secondly, the risk factor distribution will change in the non-exposed population. Thirdly the risk of the disease does not decrease instantly after removing the risk factor. Therefore, one should be careful in translating the PAR from such studies into practice. Furthermore, a high combined PAR does not mean that no additional risk factors can be detected for diabetes. The diabetes cases that are attributed to the current risk factors can alternatively be attributed to a novel risk factor, when the novel risk

factor interacts with the currently known risk factors.

Our study has the advantage of having a large sample size, a long follow-up period, and a considerable number of incident diabetes cases. However, a limitation is that physical activity was not measured in our study at baseline. Inclusion of physical activity in the models will probably modify the hazard ratio and the PAR of other risk factors. One other limitation was that our study population was over 55 years old, which may raise a debate on the generalizability of our results. To examine the issue we divided the population to subgroups of < 65 and > 65 years old. The PAR estimates were nearly the same for both groups (32.3% vs. 32.9%). This is not surprising since the association between serum CRP and diabetes was stronger in subjects <65 years old, and high serum CRP (> 1 mg/L) was more prevalent in > 65 years old subjects. This shows that PAR estimates are not modified by age and our findings can be extrapolated to other age groups.

In conclusion, high CRP is a major contributor to the risk of type II diabetes. The modifiable risk factors studied contribute to two thirds of the risk of diabetes. A large part of the diabetes cases can be prevented if the modifiable risk factors were eliminated.

References

1. Dehghan A, Kardys I, de Maat MPM, Uitterlinden AG, Sijbrands EJG, Bootsma AH, Stijnen T, Hofman A, Schram MT, Witteman JCM. Genetic Variation, C-Reactive Protein Levels, and Incidence of Diabetes. Diabetes. 2007;56:872-878.
2. Deans KA, Sattar N. "Anti-inflammatory" drugs and their effects on type 2 diabetes. Diabetes Technol Ther. 2006;8:18-27.
3. Northridge ME. Public health methods--attributable risk as a link between causality and public health action. Am J Public Health. 1995;85:1202-4.
4. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. Jama. 2001;286:327-34.
5. Hu FB, Meigs JB, Li TY, Rifai N, Manson JE. Inflammatory Markers and Risk of Developing Type 2 Diabetes in Women. Diabetes. 2004;53:693-700.
6. Albert MA, Danielson E, Rifai N, Ridker PM. Effect of statin therapy on C-reactive protein levels: the pravastatin inflammation/CRP evaluation (PRINCE): a randomized trial and cohort study. Jama. 2001;286:64-70.
7. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991;7:403-22.
8. Colditz GA, Willett WC, Stampfer MJ, Manson JE, Hennekens CH, Arky RA, Speizer FE. Weight as a risk factor for clinical diabetes in women. Am J Epidemiol. 1990;132:501-13.
9. Carey VJ, Walters EE, Colditz GA, Solomon CG, Willett WC, Rosner BA, Speizer FE, Manson JE. Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women. The Nurses' Health Study. Am J Epidemiol. 1997;145:614-9.
10. Chan JM, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. Diabetes Care. 1994;17:961-9.
11. Manson JE, Rimm EB, Stampfer MJ, Colditz GA, Willett WC, Krolewski AS, Rosner B, Hennekens CH, Speizer FE. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. Lancet. 1991;338:774-8.
12. Will JC, Galuska DA, Ford ES, Mokdad A, Calle EE. Cigarette smoking and diabetes mellitus: evidence of a positive association from a large prospective cohort study. Int. J. Epidemiol. 2001;30:540-546.
13. Wannamethee SG, Shaper AG, Perry IJ. Smoking as a Modifiable Risk Factor for Type 2 Diabetes in Middle-Aged Men. Diabetes Care. 2001;24:1590-1595.
14. Foy CG, Bell RA, Farmer DF, Goff DC, Jr., Wagenknecht LE. Smoking and Incidence of Diabetes Among U.S. Adults: Findings from the Insulin Resistance Atherosclerosis Study Diabetes Care. 2005;28:2501-2507.
15. Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes. 2000;49:2201-7.
16. Benichou J, Gail MH. Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. Biometrics. 1990;46:991-1003.
17. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985;122:904-14.
18. Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL,

       Schoenberg JB, Mayne ST, Dubrow R, Rotterdam H, West AB, Blaser M, Blot WJ, Gail MH, Fraumeni JF, Jr. Population attributable risks of esophageal and gastric cancers. J Natl Cancer Inst. 2003;95:1404-13.

19.   Gail DM. Interactive Risk Assessment Program. In. 2.2 ed; 2002.

20.   Walter SD. Attributable Risk in Practice. Am. J. Epidemiol. 1998;148:414-.

21.   Trayhurn P, Beattie JH. Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ. Proc Nutr Soc. 2001;60:329-39.

22.   Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health. 1998;88:15-9.

23.   Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, 3rd, Criqui M, Fadl YY, Fortmann SP, Hong Y, Myers GL, Rifai N, Smith SC, Jr., Taubert K, Tracy RP, Vinicor F. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation. 2003;107:499-511.

24.   Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, Willett WC. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. N Engl J Med. 2001;345:790-7.

25.   Bazzano LA, He J, Muntner P, Vupputuri S, Whelton PK. Relationship between cigarette smoking and novel risk factors for cardiovascular disease in the United States. Ann Intern Med. 2003;138:891-7.

26.   Mascitelli L, Pezzetta F. Tobacco smoke, systemic inflammation and the risk of type 2 diabetes. J Intern Med. 2006;259:332; author reply 333.

Chapter 2.3

High serum uric acid as a novel risk factor for type II diabetes mellitus

Abstract

Objective

To investigate the association between serum uric acid level and risk of type II diabetes mellitus.

Research design and methods

The population for analysis consisted of 4536 subjects free from diabetes at baseline. During a mean of 10.1 years of follow up, 462 subjects developed diabetes.

Results

The age- and sex-adjusted hazard ratios (95% confidence interval) for diabetes were 1.30 (0.96 – 1.76) for the second, 1.63 (1.21 – 2.19) for the third, and 2.83 (2.13 – 3.76) for the fourth quartile of serum uric acid, compared to the first quartile. After adjustment for BMI, waist circumference, systolic and diastolic blood pressure, and HDL-cholesterol, the hazard ratios decreased to 1.08 (0.78 – 1.49), 1.12 (0.81 – 1.53), and 1.68 (1.22 – 2.30), respectively.

Conclusion

The results of this population-based study suggest that serum uric acid is a strong and independent risk factor for diabetes.

Introduction

Serum uric acid is positively associated with serum glucose in healthy subjects[1,2]. However, this association is not consistent in healthy and diabetic individuals[3-5] as a low serum level of uric acid is reported in the hyperglycemic state [6]. Since most individuals experience a phase of impaired glucose tolerance before progression to diabetes, it is not clear whether or not raised serum uric acid predicts the risk of type II diabetes[4,5]. We investigated the association between serum uric acid and risk of diabetes in the Rotterdam Study, a large population-based, prospective cohort study among subjects aged 55 years and older.

Methods

The Rotterdam Study has been described in detail elsewhere[7]. Written informed consent was obtained from all participants and the Medical Ethics Committee of Erasmus Medical Center approved the study. Serum uric acid was measured at baseline with a Kone Diagnostica reagent kit and a Kone autoanalyzer[8,9]. Prevalent diabetes cases were excluded at baseline[10]. Incident cases of diabetes were diagnosed during follow up based on the guidelines of the American Diabetes Association[11] and WHO[12] using information from the general practitioners, the pharmacies' databases, and fasting blood samples that were taken during follow-up examinations[10].

Cox regression analysis was used to investigate the association of serum uric acid and risk of type II diabetes. The PAR and 95% confidence intervals were calculated with the use of Interactive Risk Assessment Program (IRAP) developed by Dr Mitchell Gail (US national Cancer Institute 2002)[13].

Results

Serum uric acid was ranged from 107 μmol/l to 756 μmol/l with a mean (± standard deviation) of 323.7 (± 82.2) μmol/l. Age, BMI, waist circumference, systolic and diastolic blood pressure, and HDL cholesterol were significantly correlated with serum uric acid. The correlation coefficient ranged from 0.03 for diastolic blood pressure to 0.35 for waist circumference. Except for HDL-cholesterol, the correlations were positive.

During a mean follow-up time of 10.1 years, 462 subjects out of 4536 participants developed diabetes (incidence rate = 10.1 per 1000 person years). The age- and sex-adjusted hazard ratios (95% confidence interval) for diabetes were 1.30 (0.96 – 1.76) for the second, 1.63 (1.21 – 2.19) for the third, and 2.83

(2.13 – 3.76) for the fourth quartile of serum uric acid, compared to the first quartile. After adjustment for BMI, waist circumference, systolic and diastolic blood pressure, and HDL-cholesterol, the hazard ratios decreased to 1.08 (0.78 – 1.49), 1.12 (0.81 – 1.53), and 1.68 (1.22 – 2.30), respectively. The PAR of high serum uric acid for diabetes was 0.24 (0.17 – 0.30) for the fourth quartile, 0.09 (0.3 – 0.15) for the third quartile, and 0.04 (-0.01 – 0.10) for the second quartile.

Discussion

We showed that the subjects with higher levels of serum uric acid are more at risk of developing type II diabetes. We also found that one quarter of diabetes cases can be attributed to a high serum uric acid level.

Our finding is in agreement with previous studies. At least two studies in the 1980s reported on the association of serum uric acid with the risk of diabetes[14,15], however, the association was not adjusted for any potential confounder. Lately, it has been shown in middle-aged Japanese men[4] that serum uric acid level is significantly associated with the risk of diabetes even after adjustment for potential confounders. However, another study on middle-aged Japanese men in Osaka, showed that the association was not significant after adjustment for BMI, alcohol consumption, smoking, physical activity, fasting blood sugar and parental history of diabetes. The absence of an independent effect in this study could be explained by the fact that their study population only consisted of men. We observed in our study that, although not significantly, the association was weaker in men than in women (data not shown).

Recognition of high serum uric acid as a risk factor for diabetes has been a matter of debate for a few decades since hyperuricemia has been presumed

Table 1 – Hazard ratios for diabetes according to level of serum uric acid

Uric acid Quartile (level in µmol/l)	Participants (cases)	HR (95% CI)		
		Model 1	Model 2	Model 3
1 (≤ 267)	1153 (77)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
2 (268 – 310)	1141 (94)	1.30 (0.96 – 1.76)	1.14 (0.83 – 1.57)	1.08 (0.78 – 1.49)
3 (311 – 370)	1175 (120)	1.63 (1.21 – 2.19)	1.23 (0.89 – 1.67)	1.12 (0.81 – 1.53)
4 (> 370)	1067 (171)	2.83 (2.13 – 3.76)	1.92 (1.41 – 2.62)	1.68 (1.22 – 2.30)
P for Trend		< 0.001	< 0.001	< 0.001
One SD increment	4536 (462)	1.53 (1.39 – 1.67)	1.37 (1.23 – 1.52)	1.31 (1.18 – 1.46)

Model 1: adjusted for age and sex.

Model 2: model 1 + BMI, and waist circumference

Model 3: model 2 + systolic and diastolic blood pressure, and HDL-cholesterol

to be a consequence of insulin resistance rather than its precursor. However, recent findings suggest that uric acid could be related to the development of diabetes. Serum uric acid has been shown to be associated with oxidative stress[16] and production of TNF-alpha[16] which are both related to the development of diabetes. In addition, a recent study in rats showed that fructose-induced hyperuricemia plays a pathogenic role in the metabolic syndrome[17]. These findings support high serum uric acid as a precursor of type II diabetes.

Currently, gout and renal disorders are the only consequences considered for hyperuricemia. Recent studies have introduced serum uric acid as a potential risk factor for hypertension[18], stroke[8], and cardiovascular diseases[19]. Our findings, suggest that type II diabetes is another consequence of hyperuricemia. The importance of this finding is even clearer when considering that lowering serum uric acid in subjects in the highest quartile may decrease the incidence of diabetes by 24%, if the relationship is causal. Hence, the public health impact of high serum uric acid may be larger than currently thought. Even so, uric acid is neither a target for treatment in asymptomatic hyperuricemia nor a risk marker in clinical practice[20], but methods for assessment of serum uric acid are widely available and inexpensive. Moreover, xanthine oxidase inhibitors, which are currently used to decrease serum uric acid, are safe and inexpensive.

In conclusion, our findings together with those from previous literature indicate that lowering uric acid may be a novel treatment target for preventing diabetes and justify a prospective clinical trial on the possible benefits of the measurement and lowering serum uric acid on multiple chronic disease endpoints.

References

1. Facchini F, Chen YD, Hollenbeck CB, Reaven GM. Relationship between resistance to insulin-mediated glucose uptake, urinary uric acid clearance, and plasma uric acid concentration. Jama. 1991;266:3008-11.
2. Modan M, Halkin H, Karasik A, Lusky A. Elevated serum uric acid — a facet of hyperinsulinaemia. Diabetologia. 1987;30:713-718.
3. Wun YT, Chan CSY, Lui CS. Hyperuricaemia in type 2 diabetes mellitus. Diabetes, Nutrition and Metabolism - Clinical and Experimental. 1999;12:286-291.
4. Nakanishi N, Okamoto M, Yoshida H, Matsuo Y, Suzuki K, Tatara K. Serum uric acid and risk for development of hypertension and impaired fasting glucose or Type II diabetes in Japanese male office workers. Eur J Epidemiol. 2003;18:523-30.
5. Taniguchi Y, Hayashi T, Tsumura K, Endo G, Fujii S, Okada K. Serum uric acid and the risk for hypertension and Type 2 diabetes in Japanese men: The Osaka Health Survey. J Hypertens. 2001;19:1209-15.
6. Nan H, Dong Y, Gao W, Tuomilehto J, Qiao Q. Diabetes associated with a low serum uric acid level in a general Chinese population. Diabetes Res Clin Pract. 2006.
7. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991;7:403-22.
8. Bos MJ, Koudstaal PJ, Hofman A, Witteman JCM, Breteler MMB. Uric Acid Is a Risk Factor for Myocardial Infarction and Stroke: The Rotterdam Study. Stroke. 2006;37:1503-1507.
9. Trivedi R, Rebar L, Berta E, Stong L. New enzymatic method for serum uric acid at 500 nm. Clin Chem. 1978;24:1908-1911.
10. Dehghan A, Kardys I, de Maat MP, Uitterlinden AG, Sijbrands EJ, Bootsma AH, Stijnen T, Hofman A, Schram MT, Witteman JC. Genetic variation, C-reactive protein levels, and incidence of diabetes. Diabetes. 2007;56:872-8.
11. Mellitus TECotDaCoD. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care. 1997;20:1183-97.
12. WHO. Diabetes Mellitus fact sheet 138. Geneva; 2002.
13. Gail DM. Interactive Risk Assessment Program. In. 2.2 ed; 2002.
14. Brand FN, McGee DL, Kannel WB, Stokes J, 3rd, Castelli WP. Hyperuricemia as a risk factor of coronary heart disease: The Framingham Study. Am J Epidemiol. 1985;121:11-8.
15. Herman JB, Goldbourt U. Uric acid and diabetes: observations in a population study. Lancet. 1982;2:240-3.
16. Butler R, Morris AD, Belch JJ, Hill A, Struthers AD. Allopurinol normalizes endothelial dysfunction in type 2 diabetics with mild hypertension. Hypertension. 2000;35:746-51.
17. Nakagawa T, Hu H, Zharikov S, Tuttle KR, Short RA, Glushakova O, Ouyang X, Feig DI, Block ER, Herrera-Acosta J, Patel JM, Johnson RJ. A causal role for uric acid in fructose-induced metabolic syndrome. Am J Physiol Renal Physiol. 2006;290:F625-631.
18. Johnson RJ, Feig DI, Herrera-Acosta J, Kang DH. Resurrection of uric acid as a causal risk factor in essential hypertension. Hypertension. 2005;45:18-20.
19. Baker JF, Krishnan E, Chen L, Schumacher HR. Serum uric acid and cardiovascular

disease: Recent developments, and where do they leave us? Am J Med. 2005;118:816-826.

20.   Duffy WB, Senekjian HO, Knight TF, Weinman EJ. Management of asymptomatic hyperuricemia. Jama. 1981;246:2215-6.

Chapter 3

Novel Risk Factors for Coronary Heart Disease

Chapter 3.1

Risk of coronary heart diseases related to high serum C-reactive protein and traditional risk factors

Abstract

Background
We aimed to estimate the proportion of coronary heart diseases (CHD) that is attributable to high serum levels of C-reactive protein (CRP).

Methods
The Rotterdam Study is a population-based, prospective follow-up study among 7983 participants aged 55 years and older. Traditional risk factors and serum CRP were determined at baseline (1990-1993). We estimated the population attributable risk (PAR) to estimate the proportion of myocardial infarction (MI) and CHD that is attributable to each risk factor including high serum CRP.

Results
During a mean follow up time of 9.8 years, 1008 participants developed CHD, of which 396 were cases ofMI. The PAR (95% confidence interval) estimates showed that 0.34 (0.20, 0.51) of MI, and 0.25 (0.10, 0.51) of CHD cases were attributable to high serum CRP (> 1 mg/l). High serum cholesterol (> 5.2 mmol/l) was the only risk factor that contributed more to risk of MI and CHD than high serum CRP (0.30 for MI and 0.58 for CHD). Collectively, the studied risk factors contributed to 0.81 (0.69, 0.90) of the risk of MI, and 0.64 (0.52, 0.75) of the risk of CHD.

Conclusions
High serum CRP is one of the major contributors to the risk of CHD.

Introduction

Inflammation contributes to atherogenesis in all phases, from fatty streak initiation to cardiovascular events[1] and inflammatory markers are shown to predict coronary heart disease (CHD). So far, serum C-reactive protein (CRP), a marker of inflammation, has more consistently been associated with the risk of CHD than other studied inflammatory markers[2]. It has even been suggested that CRP is a better predictor of cardiovascular events than LDL cholesterol (LDL-C)[3]. Recent studies have provided some evidence that CRP plays a role in the pathogenesis of myocardial infarction (MI)[4]. However, it is unknown whether high serum CRP is a major contributor to the risk of CHD or not.

The proportion of diseased subjects that are attributable to a certain risk factor can be estimated by the population attributable risk (PAR). The PAR, which is a relevant measure to judge the public health impact of risk factors, depends on both the strength of the association and the prevalence of the risk factor[5,6].

In this study we estimated the proportion of the risk of CHD, which is attributable to high serum CRP. For this purpose, we estimated the PARs of serum CRP and established traditional risk factors for incident MI and CHD in the Rotterdam Study, a large population-based prospective cohort study in Caucasians aged 55 years or over.

Methods

Study Population

The study was conducted within the framework of the Rotterdam Study, an ongoing prospective, population-based cohort study on determinants of a number of chronic diseases. The Rotterdam Study has been described in detail elsewhere[7,8]. In brief, all inhabitants of Ommoord, a district of Rotterdam in the Netherlands, who were 55 years or over, were invited to participate in this study. Of all 10 275 eligible individuals, 7983 agreed to participate (78%). Written informed consent was obtained from all participants and the Medical Ethics Committee of the Erasmus Medical Center approved the study.

Baseline measurements

The baseline examinations took place from 1990 to 1993. Participants were visited at home for an interview. The information on current health status, medical history, use of medication, and smoking status were obtained during the interview. The interview was followed by two visits at the research center for blood sampling and further examinations. At the research center, height, and weight were measured, and the body mass index was calculated (kg/

m^2). Blood pressure was measured at the right brachial artery using a random-zero sphygmomanometer with the participant in sitting position. The first and fifth Korotkoff sounds were recorded twice and averaged to calculate systolic and diastolic blood pressure. Serum total cholesterol was determined by an automated enzymatic procedure in a non-fasting blood sample. HDL-C was measured after precipitation of the non-HDL fraction with phosphotungstate-magnesium. Serum glucose was measured in a non-fasting and a post load sample using a glucose hexokinase method. High sensitivity CRP was measured by use of Rate Near Infrared Particle Immunoassay (Immage® Immunochemistry System, Beckman Coulter, USA). This system measures concentrations from 0.2 to 1440 mg/l, with a within-run precision < 5.0%, a total precision < 7.5% and a reliability coefficient of 0.995.

Myocardial infarction

At baseline, participants were asked whether they have ever experienced a heart attack. Later, a 12-lead electrocardiogram (ECG) was recorded and stored digitally. All the ECGs were processed by the use of Modular ECG Analysis System (MEANS) to obtain the measurements and interpretations[9,10]. MEANS uses a comprehensive set of criteria, partly derived from the Minnesota codes, to determine MI. Additional information was collected in subjects who reported a previous MI but had no evidence in their ECG. This additional information was used to distinguish those who actually experienced MI (either experienced a non–Q-wave MI or the Q wave disappeared over time) from those who mistook other symptoms for MI. By use of the additional information, participants were classified into recognized MI (subjects with self-reported MI confirmed by matching EKG characteristics or clinical data), unrecognized MI (subjects without documented or self-reported MI, but with ECG characteristics matching an MI), and non-MI (subjects without indication of MI on ECG and no self-report or medical documentation of an earlier MI).

Follow up information

Follow up for clinical events started at baseline and follow up examinations were carried out periodically in 1995-1996, 1997-1999, and 2002-2004. Participants were continuously monitored for fatal and nonfatal cardiovascular events through automated linkage with files from general practitioners and pharmacies working in the study district of Ommoord. In addition, all medical records of the participants under the care of general practitioners outside the study area were checked annually.

Two research physicians independently coded all reported events according to the International Classification of Diseases, 10th edition (ICD-10). Codes on

which the research physicians disagreed were discussed to reach consensus. Finally, a medical expert in cardiovascular disease, whose judgment was considered final, reviewed all events. Information on vital status was obtained regularly from municipal health authorities in Rotterdam. For the present study, follow-up data were available until October 1, 2005. In identifying incident myocardial infarctions (ICD-10 code I21), all available information, which included ECG, cardiac enzyme levels, and the clinical judgment of the treating specialist, was used.

We defined incident CHD as fatal or nonfatal MI (ICD-10 code I21), coronary artery bypass grafting (CABG), percutaneous transluminal coronary angioplasty (PTCA), other forms of acute (I24) or chronic ischemic heart disease (I25), sudden cardiac death (I46), sudden death undefined (R96), and death due to ventricular fibrillation (I49) and congestive heart failure (I50).

Population for analysis

Serum level of CRP was measured in 6658 out of 7129 participants who visited the research center at baseline. We excluded those who had a history of MI, CABG, or PTCA. Hence, the population for analysis consisted of 5236 subjects.

Statistical Analysis

Established cutoff points were used to categorize continuous covariates. On this basis, cutoff points of 1 and 3 mg/l for serum CRP, 25 and 30 kg/m^2 for BMI, 5.2 and 6.2 mmol/liter for total cholesterol, and 1.0 and 1.6 unit for HDL cholesterol were used. Smoking was categorized as current smoking and former smoking versus never smoking. Hypertension was defined as systolic blood pressure >= 140 mm Hg, diastolic blood pressure >= 90 mm Hg, or use of antihypertensive medication. Diabetes was defined as non-fasting glucose > 11.1 mmol/liter, an abnormal oral glucose tolerance test (OGTT) > 11.1 mmol/liter, or use of anti-diabetic medication.

A Poisson regression analysis was used to investigate the association of the traditional risk factors and serum CRP with the incidence of MI and CHD. We estimated the PAR for variables that were significantly associated with the risk of CHD including serum CRP, total cholesterol, HDL cholesterol, hypertension and diabetes. PAR and 95% confidence intervals were calculated by the use of Interactive Risk Assessment Program (IRAP) developed by Dr Mitchell Gail (US national Cancer Institute 2002)[6,11,12]. Details of the statistical analysis have been described in previously published reports[13].

We estimated the PAR for individual and combined risk factors. The PAR for a combination of risk factors corresponds with the proportion of the disease that can be attributed to any of the studied risk factors. The combined PAR cannot be

computed by summing up the single PARs. A diseased case can simultaneously be attributed to more than one risk factor. As a result, the fraction of the population that is attributed to or prevented by each risk factor overlaps with other risk factors. Hence, the combined PAR is usually smaller than the sum of individual PARs and never exceeds 1.0.

To estimate the proportion of the disease that is exclusively attributed to a specific risk factor, we calculated the "extra attributable risk". Extra attributable risk indicates the proportion of the disease that can be attributed exclusively to the studied risk factor[14]. To calculate the extra attributable risk, we estimated the combined PAR in the presence and absence of the risk factor. The difference is the extra attributable risk for that risk factor.

All statistical analyses except for PAR calculations were performed with the use of SAS, version 8 (SAS Institute).

Results

Table 1 shows the baseline characteristics of the studied population in tertiles of serum CRP. Serum CRP was positively and significantly correlated with age, BMI, diastolic and systolic blood pressure. HDL cholesterol was significantly and negatively correlated with serum CRP. Serum CRP was also significantly higher in subjects with hypertension and diabetes.

During follow-up, 1008 participants developed CHD, of which 396 were cases of MI. The mean follow-up time was 10.0 years (incidence rate 15.5 per 1000 person years) for MI and 9.8 years (incidence rate 6.0 per 1000 person years)

Table 1 - Baseline characteristics of participants in different categories of serum CRP

Risk factor	Serum CRP			P value
	< 1mg/L	1 – 3 mg/L	> 3 mg/L	
Men (%)	40.6	37.4	44.0	< 0.001
Body mass index, mean (SD), kg/m2	24.9 (3.2)	26.5 (3.5)	27.1 (4.1)	< 0.001
Total cholesterol, mean (SD), mmol/liter	6.5 (1.2)	6.7 (1.2)	6.5 (1.3)	0.59
HDL cholesterol, mean (SD), mmol/liter	1.4 (0.4)	1.3 (0.4)	1.3 (0.3)	< 0.001
Never smoking (%)	37.5	37.2	31.1	< 0.001
Current smoking (%)	17.2	21.5	29.6	
Age, means±SD, years	67.8 (8.7)	69.4 (9.1)	71.4 (9.5)	< 0.001
Systolic blood pressure, mean (SD), mm Hg	134.9 (21.7)	139.9 (22.2)	142.7 (22.7)	< 0.001
Diastolic blood pressure, mean (SD), mm Hg	72.9 (11.4)	74.1 (11.5)	73.9 (12.2)	< 0.001
Hypertension (%)	25.2	34.5	42.4	< 0.001
Diabetes (%)	7.2	9.5	15.8	< 0.001

CRP = C-reactive protein

Table 2 - Multivariate adjusted* odds ratio (95% CI) of different risk factors for MI, and CHD

Risk Factor	% Exposed	Myocardial infarction	Coronary heart disease
C-reactive protein (2 vs. 1)	40.3	1.85 (1.36 - 2.51)	1.36 (1.10 - 1.69)
C-reactive protein (3 vs. 1)	32.0	1.54 (1.09 - 2.17)	1.59 (1.27 - 2.00)
Body mass index (2 vs. 1)	46.5	0.88 (0.69 - 1.13)	0.89 (0.75 - 1.06)
Body mass index (3 vs. 1)	14.5	0.83 (0.56 - 1.21)	0.89 (0.69 - 1.16)
Total cholesterol (2 vs. 1)	28.1	2.11 (1.25 - 3.57)	1.35 (1.00 - 1.81)
Total cholesterol (3 vs. 1)	59.4	2.80 (1.69 - 4.63)	1.57 (1.18 - 2.07)
HDL cholesterol (2 vs. 1)	53.8	1.23 (0.90 - 1.68)	1.15 (0.93 - 1.42)
HDL cholesterol (3 vs. 1)	21.3	1.70 (1.19 - 2.42)	1.70 (1.33 - 2.16)
Former smoking	40.8	1.16 (0.85 - 1.59)	1.13 (0.91 - 1.40)
Current Smoking	22.6	1.21 (0.85 - 1.72)	1.25 (0.99 - 1.60)
Hypertension	36.1	1.34 (1.05 - 1.70)	1.39 (1.18 - 1.64)
Diabetes	11.0	1.69 (1.23 - 2.33)	1.75 (1.41 - 2.16)

* Multivariate adjusted model is adjusted for: age, sex, C-reactive protein, body mass index, total cholesterol, HDL cholesterol, smoking, hypertension, and diabetes

MI = Myocardial Infarction, CHD = Coronary Heart Disease

for CHD. Table 2 shows the proportion of the participants who were exposed to different categories of high serum CRP and the traditional risk factors. It also shows the multivariate adjusted odds ratio of the studied risk factors for MI and CHD. High serum CRP, high serum cholesterol, low HDL cholesterol, hypertension, and diabetes were significantly associated with the risk of MI and CHD.

Multivariate adjusted PAR of high serum CRP (> 1 mg/l) was 0.34 (95%CI: 0.20-0.51) for myocardial infarction, and 0.25 (95% CI: 0.10, 0.51) for CHD. The PAR was 0.29 (0.16, 0.48) when the 2nd and 3rd tertile were compared to the first tertile (table 3). High serum cholesterol (> 5.2 mmol/liter) was the main

Table 3 - Multivariate adjusted* population attributable risk (PAR) of different risk factors for MI and CHD

Risk Factor	Myocardial infarction	Coronary heart disease
C-reactive protein (3 vs. 1)	0.08 (0.03 - 0.22)	0.12 (0.06 - 0.20)
C-reactive protein (2&3 vs. 1)	0.29 (0.16 - 0.48)	0.19 (0.09 - 0.34)
Total cholesterol (3 vs. 1)	0.46 (0.33 - 0.59)	0.23 (0.13 - 0.37)
Total cholesterol (2&3 vs. 1)	0.61 (0.41 - 0.78)	0.30 (0.16 - 0.49)
HDL cholesterol (3 vs. 1)	0.10 (0.04 - 0.22)	0.10 (0.06 - 0.17)
HDL cholesterol (2&3 vs. 1)	0.17 (0.04 - 0.48)	0.16 (0.06 - 0.35)
Hypertension	0.08 (0.02 - 0.23)	0.10 (0.05 - 0.18)
Diabetes	0.06 (0.02 - 0.13)	0.07 (0.04 - 0.11)

* Multivariate adjusted model is adjusted for: age, sex, C-reactive protein, body mass index, total cholesterol, HDL cholesterol, smoking, hypertension, and diabetes

PAR= Population attributable risk, MI = Myocardial Infarction, CHD = Coronary Heart Disease

contributor to the risk of MI (PAR=0.58; 95%CI: 0.38, 0.75), and CHD (PAR=0.30; 95%CI: 0.16, 0.49) (Table 3).

Collectively, these risk factors contributed to 80% (95% CI: 66%, 89%) of the risk of MI, and 60% (95% CI: 47%, 72%) of the risk of CHD. Table 5 shows the combined PAR in the presence and absence of each risk factor. Exclusion of serum CRP decreased the combined PAR from 0.80 to 0.73 for MI and from 0.64 to 0.51 for CHD. Therefore, the extra attributable risk of CRP was 0.07 for MI and 0.13 for CHD.

Discussion

In this study, we found that high serum CRP contributes substantially to the risk of MI and CHD independent of the traditional risk factors. In addition, we observed that traditional risk factors account for the majority of the risk of MI (72%) and more than half of the risk of CHD (53%) in subjects 55 years and older.

The key finding of our study is that 34% of MI and 25% of CHD cases are attributable to high serum CRP (> 1mg/l). Using a higher cutoff point of 3 mg/l for high serum CRP, we showed that this proportion was 8% for MI and 12% for CHD. This finding is in agreement with Cushman et al[15] who reported 11% of incident CHD is attributable to high serum CRP (> 3 mg/l) in an elderly population. However, in the study by Cushman, the PAR was neither estimated for other cut-off points nor in combination with the traditional risk factors.

It is worthy to note that our study may have underestimated the contribution of serum CRP to the risk of CHD. Firstly, our study evaluates serum CRP at a single point in time. If chronic inflammation plays a causal role in the development of cardiovascular events, a single measurement of serum CRP

Table 4 - Population attributable risk (PAR) of all risk factors combined, and all risk factors combined with one of them excluded for MI and CHD

Risk Factor		Myocardial infarction	Coronary heart disease
		PAR	PAR
Combined PAR		0.80 (0.66 - 0.89)	0.60 (0.47 - 0.72)
Deleted risk factor	C-reactive protein	0.73 (0.55 - 0.85)	0.51 (0.36 - 0.65)
	Total cholesterol	0.50 (0.34 - 0.66)	0.43 (0.31 - 0.56)
	HDL cholesterol	0.76 (0.60 - 0.87)	0.52 (0.39 - 0.65)
	Hypertension	0.78 (0.63 - 0.89)	0.56 (0.41 - 0.69)
	Diabetes	0.79 (0.64 - 0.89)	0.57 (0.43 - 0.70)

PAR= Population attributable risk , MI = Myocardial Infarction, CHD = Coronary Heart Disease

may not adequately reflect the cumulative inflammatory burden[16]. Secondly, the role of CRP may be more complex than previously appreciated[17]. We adjusted the association for traditional risk factors, considering them as confounders. However, CRP may intermediate the effect of traditional risk factors[18]. If so, the adjustment has attenuated the real effect of CRP.

Nearly 40% of individuals who develop cardiovascular disease have only one traditional risk factor and more than 20% have none[19]. Moreover, attempts to modify risk factors frequently fail and many people find it difficult to reduce their risk of developing cardiovascular diseases. Hence, there is an unmet need for novel risk factors that contribute to a large proportion of cases and provide worthy alternatives in risk modification. Serum CRP, a marker of chronic systematic inflammation, has consistently been shown to associate with the risk of cardiovascular diseases. In this study, we investigated the proportion of MI and CHD cases that are attributable to high serum CRP and found that a greater proportion of MI and CHD cases could be attributed to high serum CRP compared to other traditional risk factors, except for high serum cholesterol. Our finding supports that serum CRP, and in more general terms inflammation, is a major contributor to the cardiovascular risk. This stimulates interest in further investigation of the possibilities of anti-inflammatory interventions in primary prevention of cardiovascular diseases beyond what has been obtained from the traditional risk factors.

Inflammation, as reflected by the serum level of CRP, has been shown to strongly predict cardiovascular events. Many lifestyle interventions known to reduce cardiovascular risk also decrease serum CRP levels. Moreover, statins, which reduce vascular risk, lower serum CRP levels. Even so, the causal role of CRP in cardiovascular events is still a matter of debate[17]. PAR is mostly estimated for risk factors with a widely accepted causal role. Nevertheless, estimating the PAR for high serum CRP illustrates the potential impact of a novel risk factor[14].

A diseased subject can simultaneously be attributed to or prevented by more than one risk factor. Therefore, the fractions of the disease attributed to different risk factors overlap each other and cannot be simply summed up. To estimate the proportion of the disease that is attributable to a group of risk factors, one should estimate the combined PAR[20]. Based on the combined PAR estimation we found that 73% of MI cases and 51% of CHD cases are attributable to the traditional risk factors. The combined PAR of traditional risk factors has been estimated previously. The INTERHEART study showed that nine modifiable risk factors account for over 90% of the risk of first MI.[21] Stampfer et al showed that lifestyle factors including diet, exercise, and smoking are responsible for more than 80% of the coronary events[22]. The differences in estimates could

be due to the difference in the variables considered for the study. Moreover, estimates are likely to be different in elderly populations since the exposure rate and the strength of association alter by age.

There are certain assumptions in PAR estimation that should be taken into account when interpreting them[20]. In estimating PAR, we assume that all participants who are labeled as exposed will shift into the non-exposed group without changing the risk factor distribution in the non-exposed group. Moreover, we assume that the risk of the disease diminishes instantly after risk factor elimination. In practice, however, the effect of an intervention is likely to be different for several reasons. First, not all the people succeed to eliminate the risk factor. Second, the risk factor distribution is likely to change in the non-exposed group. Third, the risk of the disease is likely to stay high for years even after removing the risk factor. Therefore, one should be careful with interpreting the PAR from such studies for practical purposes. Our study has the advantage of having a large sample size, a long follow-up period and a considerable number of incident MI and CHD cases.

In conclusion, high serum CRP is a major contributor to the risk of MI and CHD. Compared to the traditional risk factors for CHD, a larger part of MI and CHD cases can be prevented by lowering serum CRP compared to any other of traditional risk factors, except for serum cholesterol.

References:

1. Ross R. Atherosclerosis--an inflammatory disease. N Engl J Med. 1999;340:115-26.
2. Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N Engl J Med. 2004;350:1387-97.
3. Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. N Engl J Med. 2002;347:1557-65.
4. Lange LA, Carlson CS, Hindorff LA, Lange EM, Walston J, Durda JP, Cushman M, Bis JC, Zeng D, Lin D, Kuller LH, Nickerson DA, Psaty BM, Tracy RP, Reiner AP. Association of polymorphisms in the CRP gene with circulating C-reactive protein levels and cardiovascular events. Jama. 2006;296:2703-11.
5. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985;122:904-14.
6. Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL, Schoenberg JB, Mayne ST, Dubrow R, Rotterdam H, West AB, Blaser M, Blot WJ, Gail MH, Fraumeni JF, Jr. Population attributable risks of esophageal and gastric cancers. J Natl Cancer Inst. 2003;95:1404-13.
7. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819-29.
8. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991;7:403-22.
9. van Bemmel JH, Kors JA, van Herpen G. Methodology of the modular ECG analysis system MEANS. Methods Inf Med. 1990;29:346-53.
10. de Bruyne MC, Mosterd A, Hoes AW, Kors JA, Kruijssen DA, van Bemmel JH, Hofman A, Grobbee DE. Prevalence, determinants, and misclassification of myocardial infarction in the elderly. Epidemiology. 1997;8:495-500.
11. Benichou J, Gail MH. Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. Biometrics. 1990;46:991-1003.
12. Gail DM. Interactive Risk Assessment Program. In. 2.2 ed; 2002.
13. Dehghan A, Kardys I, de Maat MPM, Uitterlinden AG, Sijbrands EJG, Bootsma AH, Stijnen T, Hofman A, Schram MT, Witteman JCM. Genetic Variation, C-Reactive Protein Levels, and Incidence of Diabetes. Diabetes. 2007;56:872-878.
14. Walter SD. Attributable Risk in Practice. Am. J. Epidemiol. 1998;148:414-.
15. Cushman M, Arnold AM, Psaty BM, Manolio TA, Kuller LH, Burke GL, Polak JF, Tracy RP. C-Reactive Protein and the 10-Year Incidence of Coronary Heart Disease in Older Men and Women: The Cardiovascular Health Study. Circulation. 2005;112:25-31.
16. Ledue TB, Rifai N. Preanalytic and analytic sources of variations in C-reactive protein measurement: implications for cardiovascular disease risk assessment. Clin Chem. 2003;49:1258-71.
17. Pepys MB, Hirschfield GM, Tennent GA, Gallimore JR, Kahan MC, Bellotti V,

Hawkins PN, Myers RM, Smith MD, Polara A, Cobb AJ, Ley SV, Aquilina JA, Robinson CV, Sharif I, Gray GA, Sabin CA, Jenvey MC, Kolstoe SE, Thompson D, Wood SP. Targeting C-reactive protein for the treatment of cardiovascular disease. Nature. 2006;440:1217-21.

18. Miller M, Zhan M, Havas S. High Attributable Risk of Elevated C-Reactive Protein Level to Conventional Coronary Heart Disease Risk Factors: The Third National Health and Nutrition Examination Survey. Arch Intern Med. 2005;165:2063-2068.

19. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Brener SJ, Ellis SG, Lincoff AM, Topol EJ. Prevalence of Conventional Risk Factors in Patients With Coronary Heart Disease. JAMA. 2003;290:898-904.

20. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health. 1998;88:15-9.

21. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. The Lancet. 2004;364:937-952.

22. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary Prevention of Coronary Heart Disease in Women through Diet and Lifestyle. N Engl J Med. 2000;343:16-22.

Chapter 3.2

Lack of association of two common polymorphisms on 9p21 with risk of coronary heart disease and myocardial infarction

Abstract

Background

Recent genome wide association (GWA) studies identified two Single Nucleotide Polymorphisms (SNP) (rs10757278 and rs10757274) in the region of the CDK2NA and CDK2NB genes to be consistently associated with the risks of coronary heart disease (CHD) and myocardial infarction (MI). We examined the SNPs in relation to the risk of CHD and MI in a large population based study of elderly population.

Methods

The Rotterdam Study is a population-based, prospective cohort study among 7983 participants aged 55 years and older. Associations of the polymorphisms with CHD and MI were assessed by use of Cox proportional hazards analyses.

Results

In an additive model, the age and sex adjusted hazard ratios (HRs) (95% confidence interval) for CHD and MI were 1.03 (0.90, 1.18) and 0.94 (0.82, 1.08) per copy of the G allele of rs10757274. The corresponding HRs were 1.03 (0.90, 1.18) and 0.93 (0.81, 1.06) for the G allele of rs10757278. The association of the SNPs with CHD and MI was not significant in any of the subgroups of CHD risk factors.

Conclusion

We were not able to show an association of the studied SNPs with risks of CHD and MI. This may be due to differences in genes involved in the occurrence of CHD in young and older people.

Introduction

It has been considered for long that genes play a substantial role in susceptibility to coronary heart disease (CHD)[1]. Up to now, a limited number of these genes have been identified through the candidate gene approach and genome wide linkage studies. Recently a number of genome wide association (GWA) studies have identified several genetic variants on chromosome 9p21 associated with the risk of CHD. McPherson et al. found a Single Nucleotide Polymorphism (SNP), rs10757274, on chromosome 9p21 associated with the risk of CHD[2]. Helgadottir et al. found a close-by SNP, rs10757278, in the same 9p21 region associated with the risk of myocardial infarction (MI)[3]. These findings were followed by another GWA study by Samani et al[4], which found rs1333049 to be associated with the risk of coronary artery disease[4]. All three SNPs are located within the same Linkage Disequilibrium (LD) block on chromosome 9 approximately 22 million base pairs from the 9p telomere, adjacent to two tumor suppressor genes, CDKN2A and CDKN2B. These genes are involved in regulation of cell proliferation. Abnormal proliferation is one of the characteristics of atherosclerosis, one of the pathological features of CHD and MI[5].

To date, the findings are replicated in several case-control studies comprising 12285 cases and 23184 controls and two cohort studies comprising 22056 subjects[6-10]. These replications have made this locus one of the best replicated findings for genetic susceptibility to cardiovascular diseases. Though these findings are promising, they will be of more clinical worth if translated to older patients who constitute a larger part of the patients. We chose to study rs10757278 and rs10757274 because they were most strongly and consistently associated with CHD and MI in GWA studies. The leading SNP of the study by Samani et al, rs1333049, is in the same LD block with rs10757278 and contributes to the same haplotype alleles. We attempted to replicate the association in the Rotterdam Study, a population-based cohort study among older subjects, but found no association.

Methods

Study Population

The study was conducted within the framework of the Rotterdam Study, an ongoing prospective, population-based cohort study on determinants of a number of chronic diseases. The Rotterdam Study has been described in detail elsewhere11. In brief, all inhabitants of Ommoord, a district of Rotterdam in

the Netherlands, who were 55 years or over, were invited to participate in this study. Of all 10275 eligible individuals, 7983 agreed to participate (78%). Written informed consent was obtained from all participants and the Medical Ethics Committee of the Erasmus Medical Center approved the study.

Baseline measurements

The baseline examinations took place from 1990-1993. Participants were visited at home for an interview. Information on current health status, medical history, use of medication, and smoking status were obtained during the interview. The interview was followed by two visits at the research center for blood sampling and further examinations.

At baseline, participants were asked whether they have ever experienced a heart attack. A 12-lead electrocardiogram (ECG) was stored digitally and analyzed by using the Modular ECG Analysis System (MEANS). Myocardial infarction found on ECG was based on criteria partly derived from the Minnesota code. A history of myocardial infarction was considered present in case of a self-report of myocardial infarction confirmed by ECG or additional clinical information, or the presence of an ECG characteristic of prior myocardial infarction[12,13].

Genotyping

Genomic DNA was extracted from leucocytes following standard procedures. Participants were genotyped for rs10757274 and rs10757278. Genotypes were determined in our study population in 2-ng genomic DNA by use of pre-designed TaqMan SNP genotyping assay (Assay ID C__26505812_10 and C__11841860_10, respectively; Applied Biosystems, Foster City, CA). Reactions were performed with the Taqman Prism 7900HT 384 wells format.

Follow up

Follow up for clinical events started at baseline and follow up examinations were carried out periodically in 1995-1996, 1997-1999, and 2002-2004. Participants were continuously monitored for fatal and nonfatal cardiovascular events through automated linkage with files from general practitioners and pharmacies working in the study district of Ommoord. In addition, all medical records of the participants under the care of general practitioners outside the study area were checked annually. Two research physicians independently coded all reported events according to the International Classification of Diseases, 10th edition (ICD-10). Codes on which the research physicians disagreed were discussed to reach consensus. Finally, a medical expert in cardiovascular disease, whose judgment was considered final, reviewed all

Table 1 – Baseline characteristics of all participants, and of cases with incident MI and incident CHD

Characteristics	All participants	MI cases	P value	CHD cases	P value
Number	6251	412	-	588	-
Age, y	69.5 (9.1)	70.3 (7.8)	0.02	68.6 (7.4)	0.23
Men, (%)	40.4	58.7	<0.001	61.2	<0.001
Body mass index, kg/m^2	26.3 (3.7)	26.3 (3.5)	0.29	26.3 (3.4)	0.07
Waist circumference, cm	90.5 (11.2)	91.9 (10.0)	0.89	92.2 (10.2)	0.33
Systolic blood pressure, mm Hg	139.3 (22.2)	142.3 (21.9)	0.01	141.3 (21.7)	0.004
Diastolic blood pressure, mm Hg	73.7 (11.5)	73.7 (11.5)	0.77	73.7 (11.3)	0.38
Hypertension, (%)	34.3	40.6	0.006	40.2	0.002
Diabetes, (%)	10.5	17.0	<0.001	15.8	<0.001
Total cholesterol, mmol/l	6.6 (1.2)	6.9 (1.2)	<0.001	6.9 (1.2)	<0.001
HDL cholesterol, mmol/l	1.3 (0.4)	1.2 (0.3)	<0.001	1.2 (0.3)	<0.001
Current smoker, (%)	22.8	26.0	<0.001	27.1	<0.001
Former smoker, (%)	41.5	49.5		51.5	

Data shown are Mean (standard deviation) unless otherwise indicated.

events. Information on vital status was obtained regularly from municipal health authorities in Rotterdam. For the present study, follow up data were available until October 1, 2005.

Incident coronary heart disease and myocardial infarction

To identifying incident myocardial infarction and coronary heart disease, we collected information from baseline (1990 - 1993) until January 1, 2005. Fatal or non-fatal MI reported by general practitioners in the research area, letters from medical specialists and discharge reports for hospitalized patients were the sources of information used. Two research physicians coded the events independently and in case of disagreement the consensus was made ina separate session. Finally a specialist whose judgment was considered final verified the coding. We defined incident MI as fatal or non fatal MI (ICD-10 code I21). Incident CHD was defined as fatal or nonfatal myocardial infarction (ICD-10 code I21), coronary artery bypass grafting (CABG), and percutaneous transluminal coronary angioplasty (PTCA).

Population for analysis

The SNPs, rs10757274 and rs10757278 were genotyped in 6251 and 6265 out of 7129 participants who visited the research center at baseline.

Statistical analysis

Genotype frequencies were tested for Hardy-Weinberg equilibrium with a chi-square test using The Hardy-Weinberg package for R (http://www.r-

project.org). To compare the baseline characteristics between healthy subjects and those who experienced CHD or MI, we used chi-square for categorical variables and ANOVA for continuous variables.

A Cox regression analysis was used to assess the association of SNPs with incident CHD and MI. The proportional hazards assumption was validated by the use of a time-dependent variable to check increasing or decreasing trends in the hazard ratio (HR) over time. The basic model was adjusted for age and sex. The multivariate adjusted model was additionally adjusted for BMI, systolic and diastolic blood pressure, total cholesterol, HDL cholesterol, smoking, and diabetes.

To examine whether the effect of SNPs vary by the level of other risk factors we performed the analysis stratified by age, sex, family history of cardiovascular disease, HDL cholesterol, diabetes, hypertension, smoking, and history of CHD. For smoking, participants were categorized to never, former, and current smokers. For hypertension and diabetes, participants were categorized into those with and without the condition. History of CHD was defined as a history of MI, percutaneous transluminal coronary angioplasty (PTCA) or coronary artery bypass grafting (CABG). For other risk factors, the population was divided into two equal subgroups by use of the median. All statistical analyses were performed with the use of SAS, version 8.

Results

Table 1 shows the baseline characteristics of the studied population. CHD and MI cases were significantly older, and more often male, hypertensive, diabetic, and smoker than subjects without these conditions. Moreover, systolic blood pressure, and total cholesterol were significantly higher and HDL cholesterol was significantly lower in CHD and MI cases. We found none of these

Table 2 - The age and sex adjusted and multivariate adjusted association of the SNPs with incident CHD and MI

	SNP	Allele	All participants	Incident cases	Age and sex adjusted	Multivariate adjusted HR*
CHD	rs10757274	G	45.8	45.5	1.03 (0.90 - 1.18)	1.00 (0.87 - 1.15)
	rs10757278	G	44.9	42.7	1.03 (0.90 - 1.18)	1.00 (0.87 - 1.15)
MI	rs10757274	G	45.8	43.8	0.94 (0.82 - 1.08)	0.97 (0.84 - 1.11)
	rs10757278	G	44.9	42.7	0.93 (0.81 - 1.06)	0.95 (0.83 - 1.10)

* adjusted for age, sex, BMI, serum total and HDL cholesterol, smoking, diabetes, systolic and diastolic blood pressure

CI = Confidence interval, CHD=Coronary heart disease, MI = Myocardial infarction

Table 3 – The age and sex adjusted and multivariate adjusted association of the SNPs with incident CHD

		Participants (cases)	Age and sex adjusted	Multivariate adjusted *
rs10757274	AA	1834 (184)	Reference	Reference
	AG	3107 (273)	0.89 (0.74 - 1.07)	0.90 (0.74 - 1.09)
	GG	1310 (131)	1.00 (0.80 - 1.26)	0.99 (0.78 - 1.25)
rs10757278	AA	1909 (188)	Reference	Reference
	AG	3097 (269)	0.90 (0.74 - 1.08)	0.90 (0.74 - 1.09)
	GG	1264 (127)	1.01 (0.81 - 1.27)	1.00 (0.79 - 1.26)

adjusted for age, sex, BMI, serum total and HDL cholesterol, smoking, diabetes, systolic and diastolic blood pressure

CI = Confidence interval, CHD=Coronary heart disease, MI = Myocardial infarction

characteristics to be significantly associated with the studied SNPs.

During follow-up, 588 participants developed CHD, of which 412 had experienced an MI. The mean follow-up time was 9.5 years for CHD (incidence rate 9.9 per 1000 person years) and 9.5 years for MI (incidence rate 6.9 per 1000 person years). Mean (standard deviation) age of onset was 68.6 (7.4) for CHD and 70.3 (7.8) years for MI. We examined the associations of rs10757274 and rs10757278 with risks of CHD and MI (table 2, 3 and 4). None of the SNPs were significantly associated with the risk of CHD or MI. The age and sex adjusted HR (95% confidence interval [CI]) for CHD and MI were 1.03 (0.90, 1.18) and 0.94 (0.82, 1.08) per copy of G allele of rs10757274, respectively. The corresponding HRs were 1.03 (0.90, 1.18) and 0.93 (0.81, 1.06) per copy of G allele of rs10757278.

Table 4 – The age and sex adjusted and multivariate adjusted association of the SNPs with incident MI

		Participants (cases)	Age and sex adjusted	Multivariate adjusted *
rs10757274	AA	1832 (133)	Reference	Reference
	AG	3106 (197)	0.90 (0.72 - 1.12)	0.94 (0.75 - 1.19)
	GG	1309 (82)	0.89 (0.67 - 1.17)	0.94 (0.70 - 1.25)
rs10757278	AA	1907 (139)	Reference	Reference
	AG	3096 (187)	0.85 (0.68 - 1.06)	0.89 (0.71 - 1.12)
	GG	1263 (80)	0.88 (0.67 - 1.16)	0.92 (0.70 - 1.23)

adjusted for age, sex, BMI, serum total and HDL cholesterol, smoking, diabetes, systolic and diastolic blood pressure

CI = Confidence interval, CHD=Coronary heart disease, MI = Myocardial infarction

We repeated the analysis with incident cases limited to those occurred before age 70. Age and sex adjusted HR (95% CI) for CHD and MI were 1.00 (0.97, 1.04) and 0.90 (0.74, 1.09) per copy of G allele of rs10757274. The corresponding HRs for MI were 1.00 (0.96, 1.04), and 0.91 (0.75, 1.10) per copy of G allele of rs10757278.

To investigate whether any of the covariates affect the relation of SNPs with CHD and MI, we repeated the analysis in subgroups of age, sex, family history of cardiovascular disease, HDL cholesterol, diabetes, hypertension, smoking, and history of CHD (figure 1). The association was not significant in any of the studied subgroups and no significant interaction was found.

Discussion

Our main goal was to replicate the results of recent GWA studies on CHD and MI in a population based study. Two SNPs, rs10757274 and rs10757278, which were most consistently and strongly associated with the risk of CHD in GWA studies were studied. We not only found no significant association between these two SNPs and the risks of CHD and MI, but also found, however non-significant, an inverse direction for the risk. We also did not find an association in subgroups of cardiovascular risk factors.

Different approaches have been used in recent decades to discover causal genes for cardiovascular diseases. A novel approach is the GWA study which searches large part of the genome for predisposing variants. Contrary to the formerly common approach, the candidate gene study, the GWA study is a hypothesis free approach, i.e. it holds no prior assumption on the location of predisposing genes. As an advantage, this approach promises a more comprehensive understanding of the causal genes. However, this method is liable to false positive findings. Hence, GWA studies always need to be replicated in independent samples to confirm their findings.

Our study had sufficient power to detect effect sizes as shown in the published studies. In an additive model and for a SNP with a minor allele frequency of 0.45 (lowest minor allele frequency: 0.43, shown in Icelandic population A)[3], our study had more than 80% power to detect a relative risk of 1.15 for CHD (the lowest effect: 1.18, shown in ARIC study [2]) and 1.23 for MI (the lowest effect: 1.25, shown in the Iceland population A) [14].

We did not find any of the SNPs to be associated with the risk of CHD in our study. One legitimate conjecture for the inconsistency of our results with former studies may be heterogeneity of the effect. Compared to the Rotterdam Study, most of the studied populations comprised young CHD or MI patients. If the risk allele on chromosome 9p21 invokes only early onset of CHD, the effect in
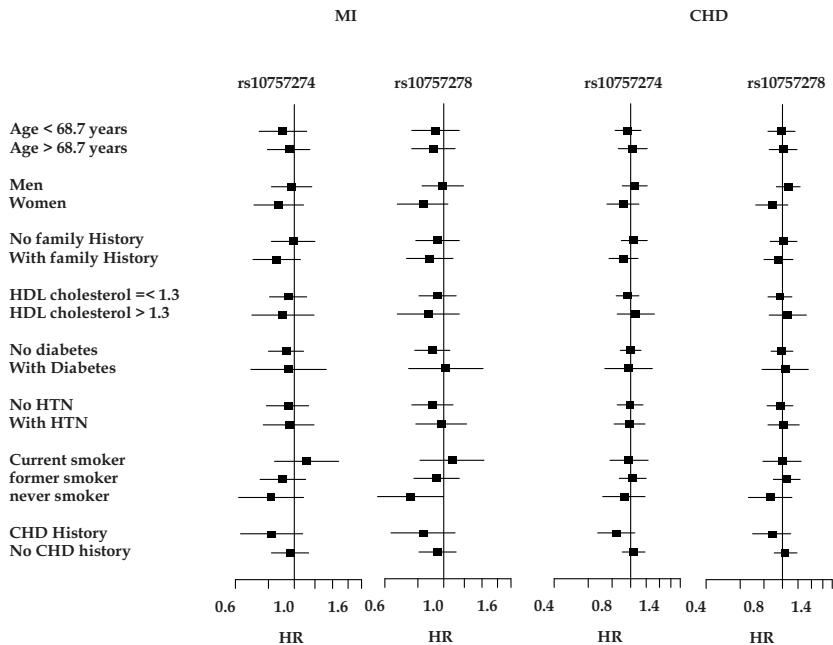
Figure 1.

Association of rs10757274 and rs10757278 with CHD and MI in subgroups of cardiovascular risk factors. The squares centered are on the hazard ratio estimate and scaled proportional to the sample size. Horizontal bars show the 95% confidence interval.

older subjects may not be large enough to be found in our study. Therefore, our negative finding may point to a heterogeneity of effect by age. In agreement with this conjecture, Helgadottir et al. showed that the association was stronger when only those with early onset MI were considered[3]. However, we failed to find an evidence of age affecting the association in our data. The strength of the association did not change materially when we limited the incident cases to those developed CHD or MI before age 70. Moreover, the strength of the association was not significantly different in age subgroups (figure1). We emphasize that our study may be underpowered to detect the effect of age on the association. It is noteworthy that the heterogeneity of effect has a particular clinical and public health impact. In Western countries the majority of morbidity and mortality from CHD occurs in elderly people. In the Netherlands, 74% of men and 91% women who experience fatal MI are older than 65 years (http://statline.cbs.nl/statweb/). The fact that CHD is less common in younger population implies that these subjects are not a good representative of the general population of CHD patients. Therefore, caution should be taken in generalizing the results of the published studies to an elderly population.

One may also speculate that those carrying the risk allele had developed CHD

at early age and were excluded at the baseline in our study. If this is true, the prevalence of risk allele should be higher in those who had a history of CHD at baseline i.e. prevalent cases of CHD. To examine this issue, we studied the association of the SNPs with prevalent CHD cases but found no association (data not shown). Moreover, the frequencies of the alleles in our population were high and comparable to former studies making selection bias unlikely.

Previous studies mainly employed standard case-control association studies. Our study has the advantage of employing a different approach, the prospective study in a large population based sample. One potential limitation of our study is that the participants were not fully followed and healthy subjects who carry the risk allele may later develop the disease.

In conclusion, we showed that the studied SNPs are not major players in development of CHD in the elderly population. Our negative finding offers a new perspective on 9p21 SNPs and shows that the association does not hold for all CHD cases. The lack of association may be due to the difference in genes involved in the development of CHD in young and older people. Individualized preventive measures and therapies constitute a major long-term goal of GWA studies. Heterogeneity of the effect, therefore, has substantial public health impact and needs to be acknowledged.

Reference

1.   Slack J, Evans KA. The increased risk of death from ischaemic heart disease in first degree relatives of 121 men and 96 women with ischaemic heart disease. J Med Genet. 1966;3:239-57.
2.   McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. Science. 2007;316:1488-91.
3.   Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007;316:1491-3.
4.   Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007;357:443-53.
5.   Rodriguez I, Coto E, Reguero JR, Gonzalez P, Andres V, Lozano I, Martin M, Alvarez V, Moris C. Role of the CDKN1A/p21, CDKN1C/p57, and CDKN2A/p16 genes in the risk of atherosclerosis and myocardial infarction. Cell Cycle. 2007;6:620-5.
6.   Abdullah KG, Li L, Shen GQ, Hu Y, Yang Y, Mackinlay KG, Topol EJ, Wang QK. Four SNPS on Chromosome 9p21 Confer Risk to Premature, Familial CAD and MI in an American Caucasian Population (GeneQuest). Ann Hum Genet. 2008.
7.   Chen Z, Qian Q, Ma G, Wang J, Zhang X, Feng Y, Shen C, Yao Y. A common variant on chromosome 9p21 affects the risk of early-onset coronary artery disease. Mol Biol Rep. 2008.
8.   Shen GQ, Rao S, Martinelli N, Li L, Olivieri O, Corrocher R, Abdullah KG, Hazen SL, Smith J, Barnard J, Plow EF, Girelli D, Wang QK. Association between four SNPs on chromosome 9p21 and myocardial infarction is replicated in an Italian population. J Hum Genet. 2008;53:144-50.
9.   Shen GQ, Li L, Rao S, Abdullah KG, Ban JM, Lee BS, Park JE, Wang QK. Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. Arterioscler Thromb Vasc Biol. 2008;28:360-5.
10. Schunkert H, Gotz A, Braund P, McGinnis R, Tregouet DA, Mangino M, Linsel-Nitschke P, Cambien F, Hengstenberg C, Stark K, Blankenberg S, Tiret L, Ducimetiere P, Keniry A, Ghori MJ, Schreiber S, El Mokhtari NE, Hall AS, Dixon RJ, Goodall AH, Liptau H, Pollard H, Schwarz DF, Hothorn LA, Wichmann HE, Konig IR, Fischer M, Meisinger C, Ouwehand W, Deloukas P, Thompson JR, Erdmann J, Ziegler A, Samani NJ. Repeated replication and a prospective meta-analysis of the

association between chromosome 9p21.3 and coronary artery disease. Circulation. 2008;117:1675-84.

11. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819-29.

12. van der Meer IM, Bots ML, Hofman A, del Sol AI, van der Kuip DA, Witteman JC. Predictive value of noninvasive measures of atherosclerosis for incident myocardial infarction: the Rotterdam Study. Circulation. 2004;109:1089-94.

13. de Bruyne MC, Mosterd A, Hoes AW, Kors JA, Kruijssen DA, van Bemmel JH, Hofman A, Grobbee DE. Prevalence, determinants, and misclassification of myocardial infarction in the elderly. Epidemiology. 1997;8:495-500.

14. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med. 2002;21:35-50.

Chapter 4

Genome-wide association studies of novel risk factors
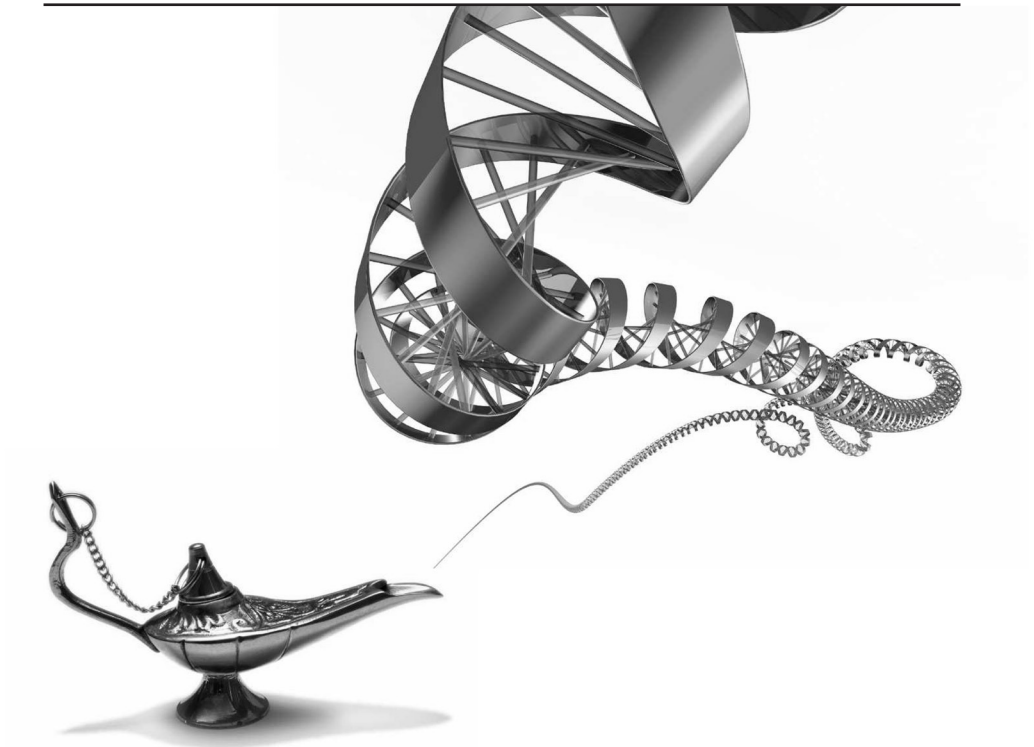
Chapter 4.1

Meta-analysis of genome-wide association studies in 26,967 subjects confirms six loci for C-reactive protein serum levels

Abstract

Background

C-reactive protein (CRP) is a heritable marker of chronic inflammation that is strongly associated with the risks of diabetes and cardiovascular disease.

Method

We performed a genome-wide association (GWA) analysis in 26,967 participants from six population-based studies to identify genetic variants that are associated with CRP levels.

Results

We found associations at six genomic loci marked by one or more SNPs that showed genome-wide significant p-values, ranging from 1.9×10^{-50} to 4.4×10^{-15} for the top SNP within each locus. These loci included the CRP gene and five other loci which were located in or close to genes encoding the leptin receptor (LEPR), interleukin 6 receptor (IL6R), glucokinase regulatory protein (GCKR), hepatic transcription factor 1 (HNF1A), and apolipoprotein E (APOE). The weighted genetic risk score based on the six top SNPs explained nearly 2% of the variance in CRP.

Conclusions

Our study confirms that CRP, APOC1, HNF1A, IL6A, LEPR, and GCKR affect CRP levels.

Introduction

C-reactive protein (CRP) is a general marker of systemic inflammation. High CRP levels are associated with increased risks of several diseases, including diabetes mellitus[1], hypertension[2], atrial fibrillation[3], coronary heart disease[4], stroke[5], and peripheral arterial disease[6], and with excess mortality[7]. The heritability of CRP levels is estimated to be 25 - 40% [8-10], suggesting that genetic variation is a major determinant of CRP levels in addition to known environmental factors. Recent genome-wide association (GWA) studies found six genes and a gene-desert region on 12q23.2 determining CRP levels[11,12]. The genes involved included the gene encoding CRP (*CRP*), leptin receptor (*LEPR*), interleukin 6 receptor (*IL6R*), glucokinase regulator (GCKR), hepatic nuclear factor 1 alpha (*HNF1A*), and apolipoprotein E (*APOE*). These genes are mainly involved in inflammatory regulation (*IL6R, LEPR*) and diabetes (*GCKR*, and *HNF1A*)[12]. In this study, we set out to discover novel genes related to CRP levels using GWA scans in 26,967 participants of six population-based cohort studies.

Methods

Participants were of European ancestry. All studies had protocols approved by local institutional review boards. Participants provided written informed consent and gave permission to use their DNA for research purposes.

The Age, Gene/Environment Susceptibility study (AGES)

The AGES Reykjavik Study cohort originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967[13]. A total of 19,381 people attended, resulting in a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within one month. One group was designated for longitudinal follow-up and was examined at all stages. One group was designated as a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the AGES-Reykjavik study re-examined 5764 survivors of the original cohort who had participated before in the Reykjavik Study. Participants came in a fasting state to the clinic. The AGES Reykjavik Study GWA study was approved by the National Bioethics Committee (VSN: 00-063) and the Data Protection Authority.

Genotyping was performed using the Illumina 370CNV BeadChip array on 3,664 participants. Sample exclusion criteria included sample failure, genotype mismatch with reference panel, and sex mismatch, resulting in cleaned genotype data on 3,219 individuals. Standard protocols for working with Illumina data

were followed with a clustering score greater than 0.4. From a total of 353,202 SNPS, 325,094 were used for imputation after exclusion of SNPs with a call rate < 97%, HWE deviation < 1×10^{-6}, mishap (PLINK haplotype-based test for non-random missing genotype data[14]) $p < 1 \times 10^{-9}$, and mismatched positions between Illumina, dbSNP and/or HapMap.

High sensitivity CRP was measured in serum on a Hitachi 912, using reagents from Roche Diagnostics and following the manufacturer's instructions. Both within- and between- assay quality control procedures were used and the coefficient of variation of the method was 1.3% to 3.4%, respectively, through the period of data collection. The assay could detect a minimal CRP concentraton of 0.1 mg/l and values below this level were classified as undetectable. All participants in this study had detectable CRP levels.

The Atherosclerosis Risk in Communities (ARIC) Study

The ARIC study is a longitudinal cohort study of atherosclerosis and its clinical sequelae. It recruited a population-based sample of 15,792 men and women aged 45-64 years from four US communities in 1987-89[15]. For this study, the analysis was restricted to subjects of European decent. Affymetrix 6.0 array genotypes were obtained in 8,861 self-identified whites: 734 individuals were excluded for the following reasons: 1) discordant with previous genotype data, 2) genotypic sex did not match phenotypic sex, 3) suspected first-degree relative of an included individual based on genome-wide genotype data, 4) genetic outlier (as assessed by average Identity by State (IBS) using PLINK and > 8 standard deviations along any of first 10 principal components in EIGENSTRAT after 5 iterations. SNPs without chromosomal location, monomorphic SNPs, SNPs whose genotype frequencies between two freezes differed by $p < 10^{-6}$, SNPs with HWE $p < 10^{-6}$ or call rate < 90% were excluded from analysis. Imputation of ~ 2.5 million autosomal SNPs in HapMap with reference to release 22 of the CEU sample was conducted using the algorithm implemented in MACH.

CRP was assessed using the immunoturbidimetric CRP-Latex (II) high-sensitivity assay from Denka Seiken (Tokyo, Japan). This assay, which has been validated against the Dade Behring method (Deerfield, Ill)[16], was performed according to the manufacturer's protocol and using a BN2 analyzer (Dade Behring (Deerfield, Il). To assess repeatability of measurements, 421 blinded replicates were measured on different dates. The reliability coefficient was 0.99.

Cardiovascular Health Study (CHS)

The CHS is a population-based cohort study of risk factors for CHD and stroke in adults ≥ 65 years conducted across four field centers[17]. The original

predominantly Caucasian cohort of 5201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists and an additional 687 African-Americans were enrolled subsequently for a total sample of 5888. DNA was extracted from blood samples drawn on all participants at their baseline examination in 1989-90.

In 2007-2008, genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV BeadChip system on 3980 CHS participants who were free of CVD at baseline, consented to genetic testing, and had DNA available for genotyping. Because the other populations in this meta-analysis were primarily of European ancestry, the self-described CHS African-Americans were excluded from this analysis to reduce the possibility of confounding due to population stratification. Genotyping has been attempted to date in 3,397 white participants, and was successful in 3,291 persons with a sample call rate > 95%. In CHS, the following exclusions were applied to identify a final set of 306,655 autosomal SNPs: call rate < 97%, HWE P < 10^{-5}, > 1 duplicate error or Mendelian inconsistency (for reference CEPH trios), heterozygote frequency = 0, SNP not found in HapMap's CEPH panel. After limiting the sample to those with successful genotypes and CRP measurements, the final dataset for this analysis comprised data on 306,655 SNPs in 3,265 CHS participants. Imputation was performed using BIMBAM v0.99 (n.b., 0.91 for QT) with reference to HapMap CEU using release 22, build 36 using one round of imputations and the default expectation-maximization warm-ups and runs. SNPs were excluded for variance on the allele dosage ≤ 0.01.

Blood was drawn in the morning after an overnight fast. Samples were promptly centrifuged at 3000g for 10 minutes at 4°C. Aliquots of plasma were stored in a central laboratory at -70°C. CRP was measured in all stored baseline plasma samples by a high sensitivity immunoassay, with an interassay coefficient of variation of 6.25%[18].

Framingham Heart Study (FHS)

The FHS (Framingham Heart Study) is a collection of three cohorts recruited to investigate cardiovascular disease and its risk factors[19]. Serum CRP measurements were available from the seventh exam (1998-2001) of the Framingham Offspring cohort (children of the original cohort and their spouses enrolled in 1971)[20,21], and the Third Generation cohort (Offspring adult children)[22] first examination, recruited from 2002-2005.

Genotyping was performed using Affymetrix 500K SNP arrays, supplemented with the MIPS 50K array. Genotypes of 8481 individuals passed our QC criteria that included call rate ≥ 97%, no excess Mendelian errors (< 1000) and average

heterozygosity within 5 SD of mean (between 25.758% and 29.958%). SNPs analyzed had minor allele frequency ≥ 1%, call rate ≥ 95% and HWE p value ≥ 10^{-6}. The analysis reported included data on 425,593 SNPs in 6,899 Framingham participants with measured CRP levels. The MACH software was used to perform imputation based on the haplotypes of the HapMap CEU trios, and 2,046,740 SNPs with good imputation quality, as measured by the ratio of observed to expected variance ≥ 0.3, were analyzed for association with CRP levels. The SNPs used for imputation passed more stringent filters including a minor allele frequency ≥ 0.01, SNP call rate ≥ 0.97, HWE p-value ≥ $1 × 10^{-6}$, differential missingness p-value ≥ $1 × 10^{-9}$ and <100 Mendelian errors; 378,163 SNPs passed these quality control criteria.

CRP was measured in fasting serum samples using a high-sensitivity assay (Dade Behring BN100). The minimum detectable dose of this assay is 0.16 mg/l, with a standard curve range of 0.16-1000 mg/l. The intra-assay coefficient of variation was 3.2%, while the inter-assay coefficient of variation was 5.3%. The final population for this analysis included 6899 individuals (Offspring n=3852, Third Generation n=3047).

The MONICA/KORA Augsburg Study (KORA)

The presented data were derived from the third population-based Monitoring of Trends and Determinants in Cardiovascular Disease (MONICA)/ Cooperative Health Research in the Region of Augsburg (KORA) survey S3[23]. This cross-sectional survey covering the city of Augsburg (Germany) and two adjacent counties was conducted in 1994/95 to estimate the prevalence and distribution of cardiovascular risk factors among individuals aged 25 to 74 years as part of the WHO MONICA study. The MONICA/KORA S3 study comprises 4,856 subjects. Among them, 3,006 subjects participated in a follow-up examination of S3 in 2004/05 (MONICA/KORA F3). All participants underwent standardized examinations including blood withdrawals for plasma and DNA. For the KORA genome-wide association study, 1,644 subjects, aged 45 to 69 years were selected from the KORA S3/F3 samples.

Genotyping for KORA F3 was performed using Affymetrix 500K Array Set consisting of two chips (Sty I and Nsp I). Hybridisation of genomic DNA was done in accordance with the manufacturer's standard recommendations. Genotypes were determined using BRLMM clustering algorithm (Affymetrix 500K Array Set). For quality control purposes, we applied a positive control and a negative control DNA every 48 samples. On chip level only subjects with overall genotyping efficiencies of at least 93% were included. In addition the called gender had to agree with the gender in the KORA study database. SNPs were excluded from analysis when monomorphic (MAF < 0.01). Imputation of

genotypes was performed using maximum likelihood method with the software MACH v1.0.9. After exclusion of subjects with missing information of CRP concentrations, the final population available for this analysis included 1,587 individuals.

CRP was measured in EDTA plasma by a high sensitivity in-house immunoradiometric assay (IRMA) in MONICA/KORA S3, using a five-point calibration with WHO International Reference Standard 85/506. The assay range was 0.05-10 mg/l. Samples with concentrations > 10 mg/l were remeasured at higher dilutions. CRP concentrations were determined in triplicate, and the mean was used for analysis. The interassay CV for CRP over all ranges was 12%[24].

Rotterdam Study (RS)

The Rotterdam Study is a prospective population-based cohort study to investigate the determinants of chronic diseases among participants aged 55 years and older[25-27]. In brief, residents of Ommoord, a district of Rotterdam (Netherlands), 55 years of age or older, were asked to participate, of whom 7983 participated. The baseline examination took place between 1990-1993.

The version 3 Illumina Infinium II HumanHap550 SNP chip array was used to conduct genotyping among self-reported Caucasians. Genotyping was successful in 6240 individuals with a sample call rate > 97.5%. SNPs with call rate < 95% and HWE pvalue < 10^{-6} were excluded. The final dataset for this analysis comprised data on 530,683 SNPs in 5974 RS participants. Imputation was conducted using the algorithm implemented in MACH. SNP filters including a minor allele frequency > 0.01, SNP Call Rate > 0.98, and HWE p-value > 1×10^{-6} were applied and 491,875 SNPs passed these filters. In total, 2,586,725 SNPs were imputed using phased haplotypes of HapMap CEU trios.

Non-fasting serum samples were collected. The samples were immediately put on ice and were processed within 30 minutes, after which they were kept frozen at -20 °C, until measurement of CRP in 2003-2004. High sensitivity CRP was measured by use of Rate Near Infrared Particle Immunoassay (Immage® Immunochemistry System, Beckman Coulter, USA). This system measures concentrations from 0.2 to 1440 mg/l, with a within-run precision < 5.0%, a total precision < 7.5% and a reliability coefficient of 0.995.

GWA analysis

Genome-wide scans were made independently in each cohort using various genotyping technologies. Quality control and data cleaning were conducted independently by each study. Each study carried out an association analysis using the genotype-phenotype data within their cohort using natural log

Table 1 - Characteristics of the Study Participants

	AGES	ARIC	CHS	FHS	KORA	RS
Number	3218	6431	3265	6899	1587	5567
Age, mean (SD), y	76.4 (5.5)	54.2 (5.7)	72.3 (5.4)	49.4 (13.8)	52.4 (10.1)	69.4 (9.1)
Men, (%)	42	46.8	39.0	46.7	50.1	58.9
Current smoker, (%)	12.7	21.7	11.3	15.6	18.0	22.5
Body mass index, mean (SD), kg/m2	27.1 (4.4)	27.0 (4.8)	26.3 (4.5)	27.4 (5.5)	27.3 (4.1)	26.3 (3.7)
Waist circumference, mean (SD), cm	100.8 (12.2)	96.0 (13.3)	93.1 (12.8)	96.0 (15.1)	90.3 (12.4)	90.6 (11.1)
Systolic blood pressure, mean (SD), mm Hg	142.6 (20.3)	118.0 (16.5)	135.2 (21.1)	121.2 (17.1)	133.2 (18.2)	139.2 (22.3)
Total/HDL cholesterol, mean (SD), ratio	3.8 (1.1)	4.6 (1.7)	0.3 (0.1)	3.9 (1.4)	4.8 (1.8)	5.2 (1.6)
Diabetes, (%)	11.6	7.4	11.7	7.5	4.4	10.7
Hypertension, (%)	80.6	25.6	36.7	19.4	25.2	32.4
Prevalent cardiovascular disease, (%)	23.5	5.4	NA	6.2	1.6	8.2
CRP, mean (SD), mg/l	3.7 (6.3)	4.2 (6.3)	0.8 (1.0)	1.5 (3.4)	1.3 (2.9)	3.4 (6.8)

AGES: The Age, Gene/Environment Susceptibility study; ARIC: The Atherosclerosis Risk in Communities Study; CHS: The Cardiovascular Health Study; FHS: The Framingham Heart Study; KORA: The MONICA/KORA Augsburg Study; RS: The Rotterdam Study
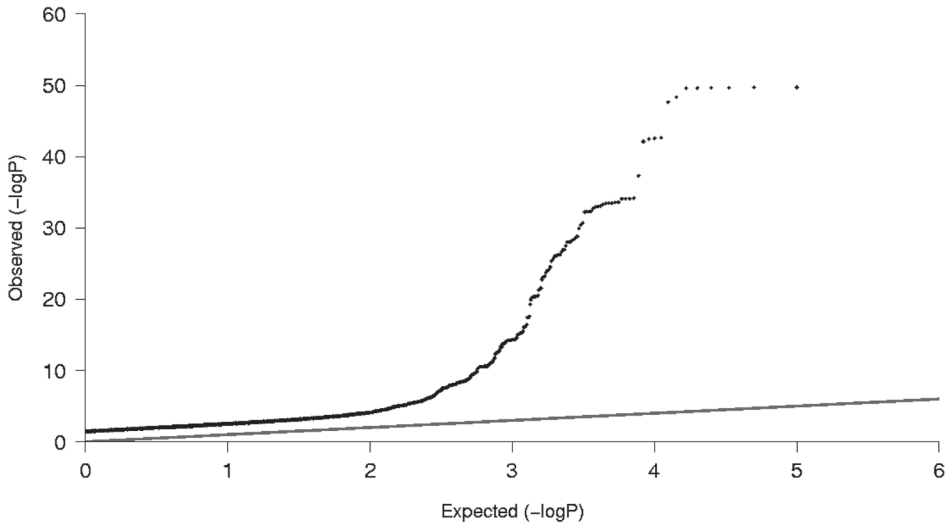
NA = Not applicable

Figure 1.
QQ-plot for the meta-analysis results. The distribution of observed p-values in the meta-analysis is compared to the expected p-values under the null hypothesis.

transformed CRP. Except for FHS, all studies conducted a linear regression analysis adjusted for age, sex, and site of recruitment (if necessary) for all SNPs based on an additive genetic model. In FHS, a linear mixed effects model was employed using the lmekin function of the kinship package in R with a fixed additive effect for the SNP genotype, fixed covariate effects, and a random environment effect. In each study, we estimated the genomic inflation rate, stated as lambda (λgc), by comparing the median chi-square to 0.4549, the median for null distribution values[28]. P values were adjusted for underlying population structure using the genomic inflation coefficient.

Meta-analysis

To calculate the combined p values and beta coefficients we used inverse-variance weighted meta-analyses with fixed-effects models. We used METAL, software designed to perform meta-analysis on GWA datasets. We applied an a priori threshold of genome-wide significance at 5.0×10^{-8} [29]. When more than one SNP clustered at a locus, we used the SNP with the smallest p-value, the 'top SNP'.

Results

Subjects and Measurements

Baseline and demographic characteristics obtained at intake are in presented in Table 1. The characteristics of the FHS shown in table 1 were not obtained at intake in the study but at the time of CRP measurement.

Genome-wide significant findings

Figure 1 is a quantile-quantile plot (Q-Q plot) which shows a large excess of small p-values on the right tail of the p-value distribution, indicating the presence of true associations of genetic variants with the variation in serum CRP levels. The primary findings from the meta-analysis are illustrated in figure 2. The figure presents the p-values for 2,624,803 SNPs across 22 autosomal chromosomes. A total of 204 SNPs exceeded the threshold of genome-wide significance ($p < 5 \times 10^{-8}$). These SNPs clustered around six loci on 1q23.2, 19q13.32, 12q24.31, 1q21.3, 1p31.3, and 2q13.

As presented in table 2, the strongest statistical evidence for an association was found for rs1205 (effective allele: T) on the C-reactive protein (*CRP*) gene (minor
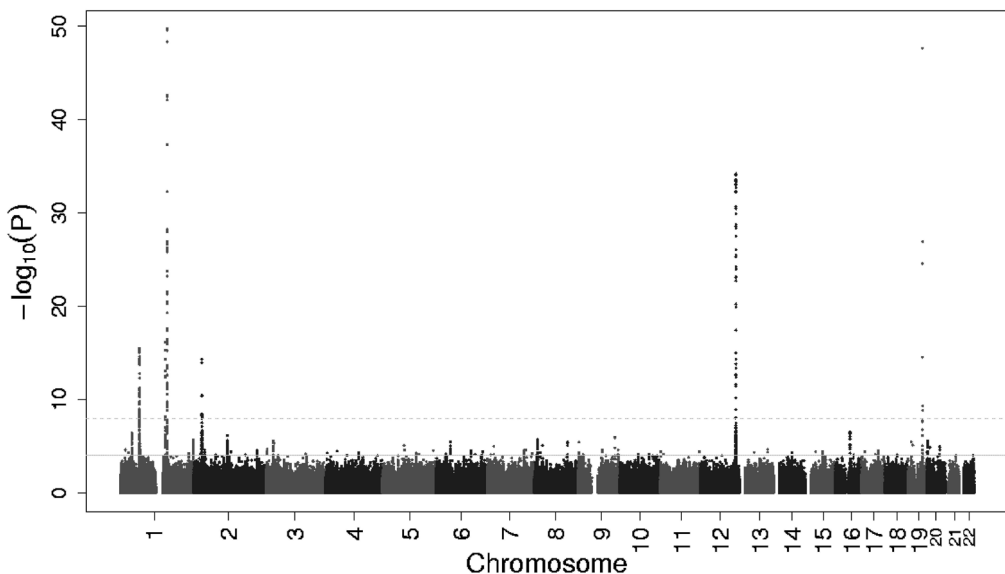

Figure 2.
Log plot for the meta-analysis. -Log p-values for all tests performed for the association of the SNPs with serum CRP levels. The grey dashed horizontal lines correspond to the p-value threshold of 5×10^{-8} and the grey solid line corresponds to the p-value threshold of 5×10^{-6}.

Table 2 - Meta-analysis results for the genome-wide significant loci associated with serum CRP levels

SNP	Chr	Band	Effective allele	MAF	Beta (SE)	P-value	Gene**
rs1205	1	1q23.2	T	0.34	-0.10 (0.007)	1.9×10^{-50}	CRP
rs4420638	19	19q13.32	G	0.80	0.12 (0.009)	2.2×10^{-48}	APOC1
rs2464195	12	12q24.31	A	0.38	-0.08 (0.007)	6.4×10^{-35}	HNF1A
rs4129267	1	1q21.3	T	0.60	-0.05 (0.006)	7.0×10^{-17}	IL6R
rs7516341	1	1p31.3	C	0.32	0.05 (0.006)	3.1×10^{-16}	LEPR
rs1260326	2	2q13	T	0.41	0.05 (0.006)	4.4×10^{-15}	GCKR

MAF: Minor allele frequency, SE: Standard error

** *Closest gene within 60kb*

Chr: Chromosome

allele frequency [MAF]: 0.31-0.34, meta-analysis p-value = 1.9×10^{-50}, serum CRP level change per minor allele [Δ]: -0.1 mg/L). The other significant loci were marked by rs4420638 (G) at 19q13.32, on the apolipoprotein C-I (*APOC1*) gene (MAF: 0.15-0.20, p = 1.2×10^{-36}, Δ: 0.12 mg/L), rs2464195 (A) at 12q24.31, on the HNF1 homeobox A (*HNF1A*) gene (MAF: 0.32-0.38, p = 6.4×10^{-35}, Δ: -0.08 mg/L), rs4129267 (T) at 1q21.3, on the interleukin 6 receptor (*IL6R*) gene (MAF: 0.39-0.42, p = 7.0×10^{-17}, Δ: -0.05 mg/L), rs7516341 (C) at 1p31.3, on the leptin receptor (*LEPR*) gene (MAF: 0.34-0.40, p = 3.1×10^{-16}, Δ: 0.05 mg/L), rs1260326 (T), at 2p23.3, on the glucokinase (hexokinase 4) regulator (*GCKR*) gene (MAF: 0.35-0.45, p = 4.4×10^{-15}, Δ: 0.05 mg/L). Table 3 shows the cohort-specific results for the association of these loci with CRP levels.

The genetic variants identified in our study explained nearly 2% of the overall variance in serum CRP in all studies except one.

Discussion

We identified six loci associated with circulating CRP levels through a meta-analysis of GWA scans from six cohort studies comprising 26,967 subjects. Our results confirm six gene annotated loci reported by Ridker et al[12] and failed to replicate their finding with a gene-desert region on 12q23.2.

CRP, and *IL6R* are associated with CRP levels at least partly through pathways related to innate and adapted immune response. The effect of the *IL6R* gene on CRP levels has been established in earlier studies. It is postulated that *IL6R* is a determinant of CRP levels since it mediates the effect of IL-6, IL-6 can induce an increase in CRP synthesis up to several 100-fold during an inflammatory state[30]. Several of the genes that were found to be associated with CRP levels are directly or indirectly related to

Table 3 – Association of six genome-wide significant loci with CRP levels in participating studies

		AGES	ARIC	CHS	FHS	KORA	RS
rs1205 (T)	P-value	3.1×10^{-10}	3.4×10^{-17}	9.7×10^{-12}	6.5×10^{-17}	3.7×10^{-9}	1.3×10^{-12}
	Beta	-0.05	-0.17	-0.19	-0.19	-0.24	-0.15
	(SE)	(0.01)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)
rs4420638 (G)	P-value	9.7×10^{-10}	4.4×10^{-19}	4.7×10^{-18}	6.9×10^{-17}	7.9×10^{-5}	5.2×10^{-11}
	Beta	-0.07	-0.22	-0.43	-0.24	-0.20	-0.22
	(SE)	(0.01)	(0.02)	(0.05)	(0.03)	(0.05)	(0.03)
rs2464195 (A)	P-value	2.3×10^{-9}	1.9×10^{-9}	6.2×10^{-8}	5.3×10^{-8}	3.6×10^{-3}	4.2×10^{-13}
	Beta	-0.05	-0.12	-0.13	-0.12	-0.12	-0.15
	(SE)	(0.01)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)
rs4129267 (T)	P-value	5.4×10^{-4}	1.7×10^{-8}	4.5×10^{-3}	6.1×10^{-5}	5.0×10^{-3}	1.0×10^{-6}
	Beta	-0.03	-0.11	-0.07	-0.09	-0.11	-0.10
	(SE)	(0.01)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)
rs7516341 (C)	P-value	8.9×10^{-4}	3.1×10^{-4}	2.0×10^{-4}	1.0×10^{-8}	2.5×10^{-5}	3.3×10^{-5}
	Beta	-0.03	-0.07	-0.11	-0.13	-0.17	-0.09
	(SE)	(0.01)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)
rs1260326 (T)	P-value	1.8×10^{-2}	2.8×10^{-14}	4.2×10^{-3}	6.4×10^{-6}	1.7×10^{-4}	6.9×10^{-1}
	Beta	0.02	0.15	0.07	0.10	0.08	0.01
	(SE)	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)

AGES: The Age, Gene/Environment Susceptibility study; ARIC, The Atherosclerosis Risk in Communities Study; CHS: The Cardiovascular Health Study; FHS: The Framingham Heart Study; KORA: The MONICA/KORA Augsburg Study; RS: The Rotterdam Study

metabolic regulatory pathways involved in diabetes. *HNF1A*, also known as T-cell factor 1 (*TCF1)*, is part of a complex network that regulates a number of genes in beta cells and hepatocytes. Common variation in this gene has been associated with impaired insulin secretion and maturity onset diabetes of the young (MODY) type 3[31]. The *GCKR* gene encodes the glucokinase regulatory protein (GCRP). Defects in the expression of glucokinase, which phosphorylates glucose to glucose-6-phosphate and probably acts as a glucose sensor, result in deficient insulin secretion[32]. Like the APOE/C region, GCKR has been implicated in the metabolism of lipids and triglyceride.

Our study has the benefit of a large sample size of 26,967 subjects. A limitation to this study is that we did not fine map the identified loci. Further fine mapping and functional studies on these loci may provide further insight into the common pathways that genetically affect CRP levels. All participants were of white European descent; hence the generalizability of our findings to individuals of other ancestry is unknown. We report SNPs associated with CRP concentrations, but as described above we acknowledge that the identified SNPs may be in linkage disequilibrium with variants causally related to CRP levels. Although we identified six loci associated with CRP levels, our GWA study does not exclude that other genetic loci influence CRP concentrations.

In conclusion, in this study among 26,967 subjects with genome wide scan data, we confirmed six genomic regions in which common variation influences CRP levels. Our study highlights immune response and metabolic regulatory pathways to be of importance in determining CRP levels.

References

1. Dehghan A, Kardys I, de Maat MP, Uitterlinden AG, Sijbrands EJ, Bootsma AH, Stijnen T, Hofman A, Schram MT, Witteman JC. Genetic variation, C-reactive protein levels, and incidence of diabetes. Diabetes. 2007;56:872-8.

2. Sesso HD, Buring JE, Rifai N, Blake GJ, Gaziano JM, Ridker PM. C-reactive protein and the risk of developing hypertension. Jama. 2003;290:2945-51.

3. Aviles RJ, Martin DO, Apperson-Hansen C, Houghtaling PL, Rautaharju P, Kronmal RA, Tracy RP, Van Wagoner DR, Psaty BM, Lauer MS, Chung MK. Inflammation as a risk factor for atrial fibrillation. Circulation. 2003;108:3006-10.

4. Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N Engl J Med. 2004;350:1387-97.

5. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. N Engl J Med. 1997;336:973-9.

6. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, Hennekens CH. Plasma concentration of C-reactive protein and risk of developing peripheral vascular disease. Circulation. 1998;97:425-8.

7. Hindorff LA, Rice KM, Lange LA, Diehr P, Halder I, Walston J, Kwok P, Ziv E, Nievergelt C, Cummings SR, Newman AB, Tracy RP, Psaty BM, Reiner AP. Common variants in the CRP gene in relation to longevity and cause-specific mortality in older adults: the Cardiovascular Health Study. Atherosclerosis. 2008;197:922-30.

8. Retterstol L, Eikvar L, Berg K. A twin study of C-Reactive Protein compared to other risk factors for coronary heart disease. Atherosclerosis. 2003;169:279-82.

9. Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH, Tracy RP. Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis. 2001;154:681-9.

10. Dupuis J, Larson MG, Vasan RS, Massaro JM, Wilson PW, Lipinska I, Corey D, Vita JA, Keaney JF, Jr., Benjamin EJ. Genome scan of systemic biomarkers of vascular inflammation in the Framingham Heart Study: evidence for susceptibility loci on 1q. Atherosclerosis. 2005;182:307-14.

11. Reiner AP, Barber MJ, Guan Y, Ridker PM, Lange LA, Chasman DI, Walston JD, Cooper GM, Jenny NS, Rieder MJ, Durda JP, Smith JD, Novembre J, Tracy RP, Rotter JI, Stephens M, Nickerson DA, Krauss RM. Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. Am J Hum Genet. 2008;82:1193-201.

12. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. Am J Hum Genet. 2008;82:1185-92.

13. Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G, Thorgeirsson G, Aspelund T, Garcia ME, Cotch MF, Hoffman HJ, Gudnason V. Age, Gene/ Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. Am J Epidemiol. 2007;165:1076-87.

14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75.

15.   The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989;129:687-702.

16.   Roberts WL, Moulton L, Law TC, Farrow G, Cooper-Anderson M, Savory J, Rifai N. Evaluation of nine automated high-sensitivity C-reactive protein methods: implications for clinical and epidemiological applications. Part 2. Clin Chem. 2001;47:418-25.

17.   Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991;1:263-76.

18.   Macy EM, Hayes TE, Tracy RP. Variability in the measurement of C-reactive protein in healthy subjects: implications for reference intervals and epidemiological applications. Clin Chem. 1997;43:52-8.

19.   Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci. 1963;107:539-56.

20.   Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med. 1975;4:518-25.

21.   Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol. 1979;110:281-90.

22.   Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol. 2007;165:1328-35.

23.   Wichmann HE, Gieger C, Illig T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005;67 Suppl 1:S26-30.

24.   Khuseyinova N, Imhof A, Trischler G, Rothenbacher D, Hutchinson WL, Pepys MB, Koenig W. Determination of C-reactive protein: comparison of three high-sensitivity immunoassays. Clin Chem. 2003;49:1691-5.

25.   Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991;7:403-22.

26.   Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819-829.

27.   Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol. 2009.

28.   Devlin B, Roeder K, Wasserman L. Genomic Control, a New Approach to Genetic-Based Association Studies. Theoretical Population Biology. 2001;60:155-166.

29.   Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic Epidemiology. 2008;32:381-385.

30.   Gabay C, Kushner I. Acute-phase proteins and other systemic responses to inflammation. N Engl J Med. 1999;340:448-54.

31.   Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, Southam L, Cox RD, Lathrop GM, Boriraj VV, Chen X, Cox NJ, Oda Y, Yano H, Le Beau MM, Yamada S, Nishigori H, Takeda J, Fajans SS, Hattersley AT, Iwasaki N, Hansen T, Pedersen O, Polonsky KS, Bell GI, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset

diabetes of the young (MODY3). Nature. 1996;384:455-8.

32. Kesavan P, Wang L, Davis E, Cuesta A, Sweet I, Niswender K, Magnuson MA, Matschinsky FM. Structural instability of mutant beta-cell glucokinase: implications for the molecular pathogenesis of maturity-onset diabetes of the young (type-2). Biochem J. 1997;322 (Pt 1):57-63.

.

Chapter 4.2

Association of three genetic loci with uric acid levels and gout risk

Abstract

Background

Hyperuricemia, a highly heritable trait, is a key risk factor for gout. We aimed to identify novel genes related to serum uric acid and gout.

Methods

Genome-wide association (GWA) studies were conducted for serum uric acid in the Framingham Heart Study (FHS; n=7699) and the Rotterdam Study (RS; n=4148). Genome-wide significant SNPs were replicated among white (n=11024) and black (n=3843) Atherosclerosis Risk in Communities (ARIC) Study participants. The association of these SNPs was evaluated with gout; results in whites were combined using meta-analysis.

Results

Three loci in FHS and two in the RS showed genome-wide significance with uric acid. Top SNPs in each locus were: missense SNP rs16890979 in SLC2A9 (p=7.0×10^{-168} [whites]; 2.9×10^{-18} [blacks]), missense SNP rs2231142 in ABCG2 (p=2.5×10^{-60} [whites]; 9.8×10^{-4} [blacks]), and rs1165205 in SLC17A3 (p=3.3×10^{-26} [whites]; 0.33 [blacks]). All SNPs showed direction-consistent association with gout in whites: rs16890979 (OR 0.58 per T allele, 95% CI 0.53-0.63, p=1.2×10^{-31}), rs2231142 (OR=1.74 per T allele, 1.51-1.99, p=3.3×10^{-15}), and rs1165205 (OR=0.85 per T allele, 0.77-0.94, p=0.002). In ARIC blacks, rs2231142 showed a direction-consistent association with gout (OR=1.71, 1.06-2.77, p=0.028). An additive genetic risk score (0-6) comprised of high risk alleles at the three loci showed graded associations in each study across scores with uric acid (from 272-351 μmol/l [FHS], 269-386 μmol/l [RS], and 303-426 μmol/l [ARIC whites]) and gout (prevalence 2-13% [FHS], 2-8% [RS], 1-18% [ARIC whites]).

Conclusions

We identified three genetic loci (two novel including a candidate functional variant Q141K in ABCG2) related to uric acid and gout. A score based on genes with a putative role in renal urate handling showed a substantial risk gradient for gout.

Introduction

Gout is one of the most common forms of arthritis[1,2]. Gout currently affects over 700,000 adults in the United Kingdom[2] and nearly three million adults in the United States[3], accounting for almost four million annual outpatient visits[4], with a substantial economic burden[5]. Epidemiological studies from a range of countries suggest that the prevalence and incidence of gout are increasing[6]. Gout is characterized by joint pain, inflammation, and painful tophi, and can result in joint destruction and disability if untreated[7].

Uric acid is the end product of purine metabolism in humans, and levels are primarily determined by endogenous metabolism (synthesis and cell turnover), and the rate of excretion and reabsorption in the kidney[1]. Humans lack uricase, the enzyme responsible for converting uric acid into its more soluble and excretable form. Renal excretion of urate is responsible for the majority of hyperuricemia and gout[8]. Thus, understanding the molecular mechanisms of urate transport in the kidney has potential research and clinical implications.

Known risk factors for gout include hyperuricemia, obesity, hypertension, diuretic use, and alcohol consumption[9]. Despite extensive research in the area of renal urate transport, the mechanisms influencing serum uric acid levels in humans by contributing to either secretion or reabsorption of urate in the proximal renal tubules have not been fully elucidated[10]. We have previously shown that the heritability of serum uric acid levels is 63%[11], suggesting that genetic variation may contribute to uric acid levels through regulation of uric acid synthesis, excretion, or reabsorption. Several recent genome-wide association (GWA) studies identified significant associations between single nucleotide polymorphisms (SNPs) in the gene SLC2A9 with uric acid levels and gout[12-16]. The gene product of SLC2A9 had not previously been implicated in uric acid metabolism, highlighting the power of GWA studies to identify unknown physiologic mechanisms contributing to disease.

The objective of this study was to identify genetic loci related to uric acid using GWA studies in two population-based studies (11847 participants) and subsequently replicate them in a third population-based study (14867 participants). Moreover, a meta-analysis of replicated SNPs was performed for uric acid and gout across studies to combine the results in whites. Finally the association of a genetic risk score summarizing the number of risk alleles was tested with both uric acid levels and gout risk.

Methods

Participants

The Framingham Heart Study started in 1948 when 5,209 participants began undergoing biannual examinations to identify cardiovascular disease and its risk factors[17,18]. In 1971, 5,124 participants were enrolled into the Framingham Offspring Study. Offspring subjects underwent examinations approximately every 4 years[19,20]. In 2002, the third generation, representing the children of the offspring cohort, was recruited (n=4095)[21]. Almost all participants in Framingham cohort are self-identified white (of European descent). The original cohort consisted of 1,644 spouse pairs, and the offspring cohort of 2,632 individuals with two parents in the original cohort (916 with at least one parent in the original cohort) and 1576 spouse pairs. The third generation consisted of 2,944 individuals with both parents in the offspring cohort, 1,148 individuals with at least one parent in the offspring cohort, and threevwith none of the parents in the offspring cohort. A broad range of phenotypes have been described and is publicly available (http://www.ncbi.nlm.nih.gov/projects/ gap/cgi-bin/study.cgi?study_id=phs000007.v2.p1). The study was approved by the institutional review board of the Boston University Medical Center (Boston, MA, USA).

The Rotterdam Study is a prospective, population-based cohort study on determinants of several chronic diseases among subjects aged ≥ 55 years[22,23]. All inhabitants of Ommoord, a district of Rotterdam (Netherlands), who were 55 years or older, were invited to take part in the study. Of 10,275 eligible individuals, 7,983 (78%) agreed to participate. At baseline (1990 - 1993), participants completed an interview at home and visited the research center for blood sampling and examination. Follow-up started at baseline and examinations were done periodically. In addition, participants were continuously monitored for major events through automated linkage with files from general practitioners and pharmacies working in the district of Ommoord. The medical ethics Committee of Erasmus Medical Center (Rotterdam, Netherlands) approved this study.

The ARIC Study is a continuing, population-based, prospective study in four US communities. From 1987 to 1989, 15792 white and African American participants aged 45-64 years were recruited by probability sampling and underwent baseline examination (visit 1), and were examined three more times, roughly every three years[24]. For this study, participants were excluded for non-consent to genetic research (n=53) or if they did not self-identify as black or white (n=47). Thus, 8923 of 11,440 white and 2,650 of 4,252 black participants attended visit 4. Further exclusions to samples were made for genotyping failure of all SNPs, and missing outcomes or covariates. The final sample for association analyses, therefore, consisted of 11,024 white and 3,843 black participants at visit 1, and 8,599 white and 2,392 black participants at visit 4. Institutional Review

Boards of the participating institutions (Johns Hopkins University, University of Minnesota, Wake Forest University, University of Mississipi, Baylor University, University of Texas and University of North Carolina) approved the study protocols. Collectively, in all three studies' participants provided written informed consent.

Genotyping

In the Framingham cohort, the SNP Health Association Resource (SHARe) project genotyped 9,274 participants with the Affymetrix 500K mapping array and the Affymetrix 50K supplemental array (Affymetrix, Santa Clara, CA, USA). 8,508 samples from these participants were genotyped successfully (sample call rate ≥97%). Exclusion of individuals with missing uric acid measurements (n=623) or covariates (n=186) resulted in a final sample size of 7699 (original cohort n=572; offspring n=3377; third generation n=3750). SNPs were excluded when SNP call rate was less than 95% or Hardy–Weinberg equilibrium $p<10^{-6}$, making the total final number of 503 551 SNPs.

In the Rotterdam cohort, plated DNA was available for 6680 of 7129 (94%) participants who visited the research centre. Genotyping was done with the Illumina 550K array (Illumina, San Diego, CA, USA) in self-reported white individuals, and succeeded in 6,240 individuals (sample call rate ≥97.5%). The final population for analysis consisted of 5,974 individuals. SNPs were excluded when minor allele frequency was 1% or less, Hardy-Weinberg equilibrium $p<10^{-5}$, or SNP call rate 90% or less, resulting in 530 683 SNPs. We imputed two SNPs, rs16890979 on chromosome 4 and rs1165205 on chromosome 6, that were not on the Illumina Infinium II HumanHap550 SNP chip (San Diego, CA, USA). Imputation was done with the maximum likelihood method implemented Markov Chain based Haplotyper (MACH) version 1.0.28 HapMap release 22 CEU phased genotypes were used as a reference. The R2 estimate of MACH (the ratio of the observed variation to the expected variation under Hardy–Weinberg equilibrium) was 0.96 for rs16890979 and 0.99 for rs1165205.

In the ARIC study, the central DNA laboratory genotyped SNPs rs16890979, rs2231142, and rs1165205 individually with TaqMan assays (Applied Biosystems, Foster City, CA, USA). Percent agreement of 315 blind duplicate samples was more than 98% for all genotyped SNPs.

Outcome measures

In the Framingham cohort, uric acid concentration was measured at the first examination cycle in every cohort with an autoanalyzer with a phosphotungstic acid reagent[26]. Gout was self-reported in the offspring cohort during examination cycles three to seven, and in the third-generation group during first

examination. Information on uric acid concentration and gout was available for 7699 and 7386 individuals, respectively.

In the Rotterdam cohort, uric acid concentration was measured at baseline with a Kone Diagnostica reagent kit and a Kone autoanalyzer[27]. Data for drug prescription were obtained from a computer network of pharmacies in the study area in which all prescriptions have been registered from Jan 1, 1991. Individuals treated with drugs exclusively prescribed for gout (allopurinol, probenecid, benzbromarone, and colchicine) were regarded as gout patients. Information about uric acid concentration was available for 4,148 individuals

Table 1 - Characteristics of 26714 study participants from the Framingham Heart Study, Rotterdam Study, and ARIC Study.

		FHS	Rotterdam	ARIC white	ARIC black
Samples for serum uric acid	Visit	Exam 1 (1948, 1971, 2002)§	visit 1 (1990 - 1993)	visit 1 (1987-89)	
	Number	7699	4148	11024	3843
	Uric Acid (µmol/l)*	315.2 (89.2)	321.2 (80.9)	350.9 (89.8)	374.7 (99.3)
	Female (%)	53	61	53	62
	Body Mass Index (kg/m²)	25.9 (4.9)	26.3 (3.7)	27.0 (4.9)	29.6 (6.1)
	Age (y)	37.9 (9.4)	69.7 (9.0)	54.3 (5.7)	53.5 (5.8)
	Alcohol drinking (g/week)**	48 (12, 120)	43 (3,85)	30 (0, 93)	53 (15, 125)
	Current Drinker (%)	81	79	65	32
	Hypertension treatment (%)	8	33	26	43
Samples for gout	Visit	Offspring and Third Generation£	During follow-up (1990 - 2006)	visit 4 (1996-98)	
	Number	7386	5741	8599	2392
	Gout (%,n)	2.7 (197)	3.3 (190)	5.4 (467)	8.8 (210)
	Female (%)	53	59	54	64
	Body Mass Index (kg/m²)	27.0 (6.5)	26.3 (3.7)	26.9 (4.7)	29.6 (5.9)
	Age (y)	50.0 (13.9)	69.0 (8.8)	54.1 (5.7)	52.9 (5.7)
	Alcohol drinking (g/week)**	33 (12, 60)	47 (3, 86)	28 (0, 91)	26 (0, 76)
	Current Drinker (%)	51	79	56	26
	Hypertension treatment (%)	21	32	39	59

Data shown are mean (standard deviation) unless otherwise indicated.

*** median (1st and 3rd quartile), among current drinkers*

§ Exam 1 of the Original Cohort (1948), Offspring (1971), and Third Generation (2002-2005)

£Exams 3-7 of Offspring (1987-1999) and exam 1 of the Third Generation (2002-2005)

Table 2 - Association of four SNPs in three loci with uric acid and gout

SNP Information	Trait		FHS §	RS	ARIC white	ARIC black	All Whites
rs16890979		MAF	0.23	0.21	0.23	0.42	
Chr 4: 9531265	Uric acid	p-value	1.6×10^{-76}	4.7×10^{-27}	2.3×10^{-105}	2.9×10^{-18}	7.0×10^{-168}
Gene: SLC2A9		beta*(se)	-0.36 (0.02)	-0.29 (0.03)	-0.34 (0.02)	-0.20 (0.02)	-0.34 (0.02)
Alleles: C/T		R^2	5.3%	2.8%	4.3%	2.0%	NA
V253I	Gout	p-value	1.3×10^{-3}	6.0×10^{-3}	1.8×10^{-9}	0.14	1.2×10^{-31}
		OR**	0.63	0.67	0.56	0.85	0.58
		95% C.I.	0.47-0.84	0.50-0.89	0.47-0.68	0.69-1.05	0.53-0.63
rs6449213		MAF	0.19	0.18	NA	NA	NA
Chr 4: 9603313	Uric acid	p-value	2.9×10^{-68}	1.15×10^{-29}	NA	NA	2.2×10^{-104}
Gene: SLC2A9		beta*(se)	-0.37 (0.02)	-0.32 (0.03)	NA	NA	-0.35 (0.02)
Alleles: T/C		R^2	4.5%	3.0%	NA	NA	NA
	Gout	p-value	1.1×10^{-2}	0.06	NA	NA	0.001
		OR**	0.66	0.75	NA	NA	0.69
		95% C.I.	0.49-0.91	0.55-1.01	NA	NA	0.55-0.86
rs2231142		MAF	0.11	0.12	0.11	0.03	NA
Chr 4: 89271347	Uric acid	p-value	9.0×10^{-20}	3.3×10^{-9}	9.7×10^{-30}	9.8×10^{-4}	2.5×10^{-60}
Gene: ABCG2		beta*(se)	0.25 (0.03)	0.20 (0.03)	0.25 (0.02)	0.22 (0.07)	0.24 (0.02)
Alleles: G/T		R^2	1.3%	0.8%	1.2%	0.3%	NA
Q141K	Gout	p-value	1.5×10^{-6}	1.5×10^{-4}	2.0×10^{-7}	0.03	3.3×10^{-15}
		OR**	1.97	1.71	1.68	1.71	1.74
		95% C.I.	1.49-2.59	1.30-2.25	1.38-2.04	1.06-2.77	1.51-1.99
rs1165205		MAF	0.46	0.47	0.47	0.13	NA
Chr 6: 25978521	Uric acid	p-value	5.6×10^{-10}	0.01	8.4×10^{-11}	0.33	3.8×10^{-29}
Gene: SLC17A3£		beta*(se)	-0.11 (0.02)	-0.06 (0.02)	-0.09 (0.01)	-0.03 (0.03)	-0.09 (0.01)
Alleles: A/T		R^2	0.7%	0.2%	0.4%	<0.1%	NA
	Gout	p-value	0.10	0.86	3.0×10^{-3}	0.33	2.0×10^{-3}
		OR**	0.83	0.98	0.81	1.16	0.85
		95% C.I.	0.67-1.04	0.80-1.21	0.71-0.93	0.86-1.56	0.77-0.94

*Beta coefficient represents 1 standard deviation change in the standardized residual of uric acid per copy increment in the minor allele, adjusting for age, sex, BMI, alcohol consumption, hypertension treatment, (cohort status in FHS, study center in ARIC). se is the standard error of the beta coefficient

**OR is the odds ratio for gout per per copy increment of the allele modeled adjusting for age, sex, BMI, alcohol consumption, hypertension treatment, (cohort status in FHS, study center in ARIC)

§FHS results generated using linear mixed effects models; FBAT results in FHS: rs16890979 (p=8.3×10^{-23}), rs6449213 (p=1.9×10^{-24}), rs2231142 (p=5.6×10^{-11}), rs1165205 (7.1×10^{-3})

£SLC17A3 refers to the entire SLC17A4/ SLC17A1/ SLC17A3 gene cluster
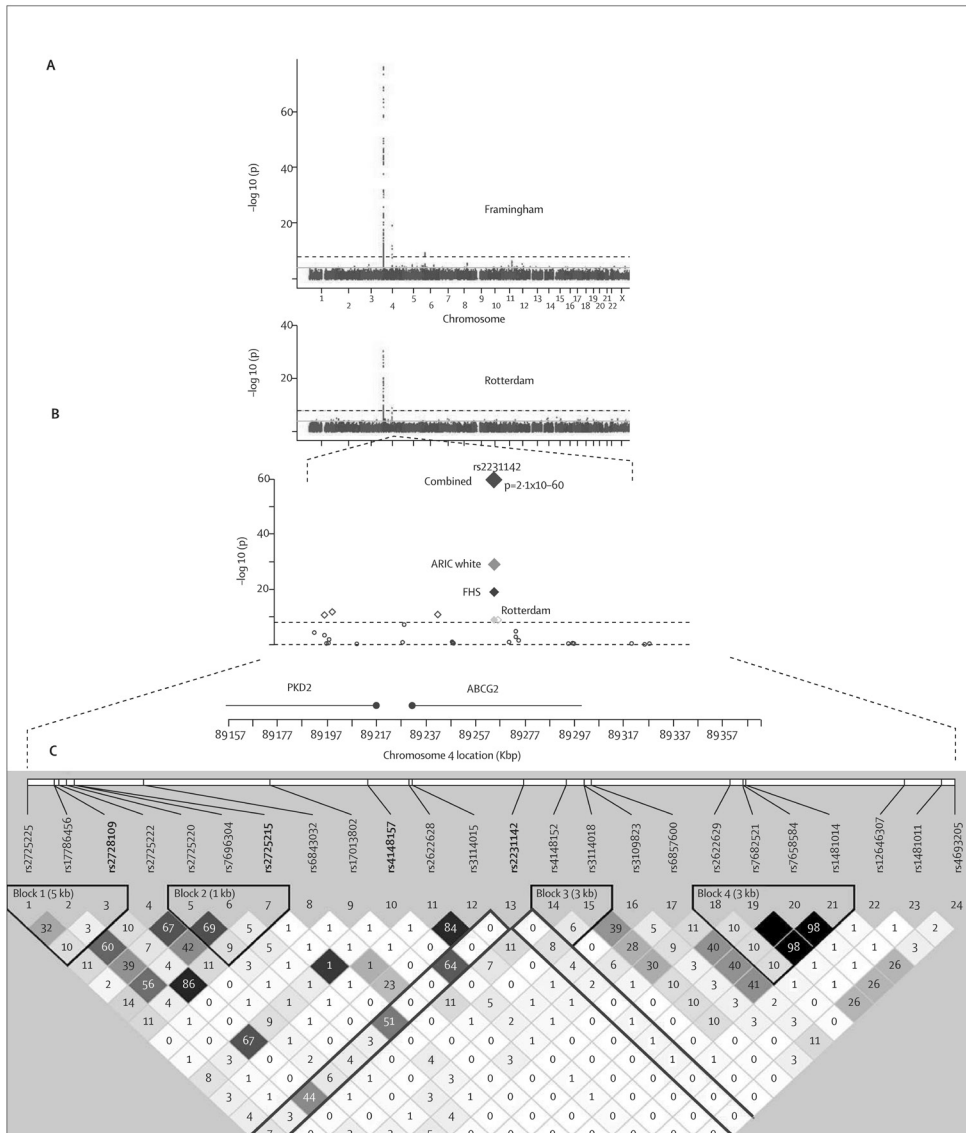
MAF = Minor Allele Frequency

Chapter 4.2

Figure 1.

ABCG2 locus (Panel A): −log10 p-values of genome-wide association analysis of uric acid for FHS and Rotterdam; (Panel B): −log10 p-values vs. physical position based on NCBI build 36.2 for SNPs (minor allele frequency >0.01) within 60Kb of ABCG2 (open diamonds) for uric acid association analysis for FHS, Rotterdam (only SNPs with p<10⁻⁷ included), and ARIC whites (only SNPs with p<10⁻⁷ included). The top associated SNP rs2231142 is plotted with solid diamonds for the three studies, and the p-value from the meta-analysis combining the results of the three studies is plotted with a solid diamond; (Panel C): Plot of linkage disequilibrium pattern in the ABCG2 region with all minor allele frequency >0.01 of SNPs typed in FHS. Each diamond contains a pair-wise r² value (no value means r²=1) between two SNPs, with a darker shade representing higher correlation. The relative locations of the SNPs are marked on the top panel. SNPs with p-value <10⁻⁸ in FHS are in bold font, and the r² tracks with top associated SNP are outlined by solid lines.

and that for gout for 5,741 individuals.

In the ARIC study, uric acid concentration was measured with the uricase method[28] at visit 1. Repeated measurements of uric acid in 40 individuals, taken at least 1 week apart, yielded a reliability coefficient of 0.91, and a coefficient of variation of 7.2%.30 Gout was identified by self-report at visit 4.

In all three studies, alcohol consumption was referred as self-reported number of drinks per week and converted into grams per week, and antihypertensive treatment was defined as self-reported intake of antihypertensive drugs or drug reconciliation.

Statistical analysis

Genome-wide association study analyses used cohort-specific (Framingham cohort only) and sex-specific standardized residuals from a least squares regression model of uric acid, adjusting for age, body-mass index, alcohol consumption, and hypertension treatment. In the Framingham cohort, genome-wide association study analysis was done as linear mixed-effect models to account for familial correlation, or family-based association testing to reduce the chance of false positives caused by population stratification[29,30]. In the Rotterdam cohort, linear regression was done with PLINK version 1.01.27 Both studies used an additive genetic model.

The most significant SNP that reached genome-wide significant association with uric acid for every region in either the Framingham cohort ($p<5.0×10^{-8}$) or the Rotterdam cohort ($p<1.0×10^{-7}$) was selected a priori for follow-up genotyping in the ARIC study. This criterion was met by rs16890979, rs2231142, and rs1165205 (Framingham cohort), and rs6449213 and rs2231142 (Rotterdam cohort). rs16890979 and rs6449213 are located in the same genetic region and are in moderate linkage disequilibrium with each other ($r^2 = 0.66$ in HapMap CEU). Therefore, only rs16890979, rs2231142, and rs1165205 were genotyped in the ARIC study. In all studies, the association with gout was judged as significant at $p < 0.05$, because only SNPs consistently associated with uric acid across studies were examined.

We combined the multivariable adjusted measures of beta coefficients and OR of replicated SNPs with uric acid and gout across studies. We used the Cochran Q test to identify heterogeneity across studies. Because no significant heterogeneity was present with uric acid concentration and gout (all $p > 0.07$), a fixed-effect model was used for both traits. We used the meta[31] package running under R[32] to calculate the combined estimates and p values.

A genetic risk score was generated for every individual by counting the number of alleles associated with high uric acid concentration (rs16890979 C, rs2231142 T, and rs1165205 A; range 0–6).
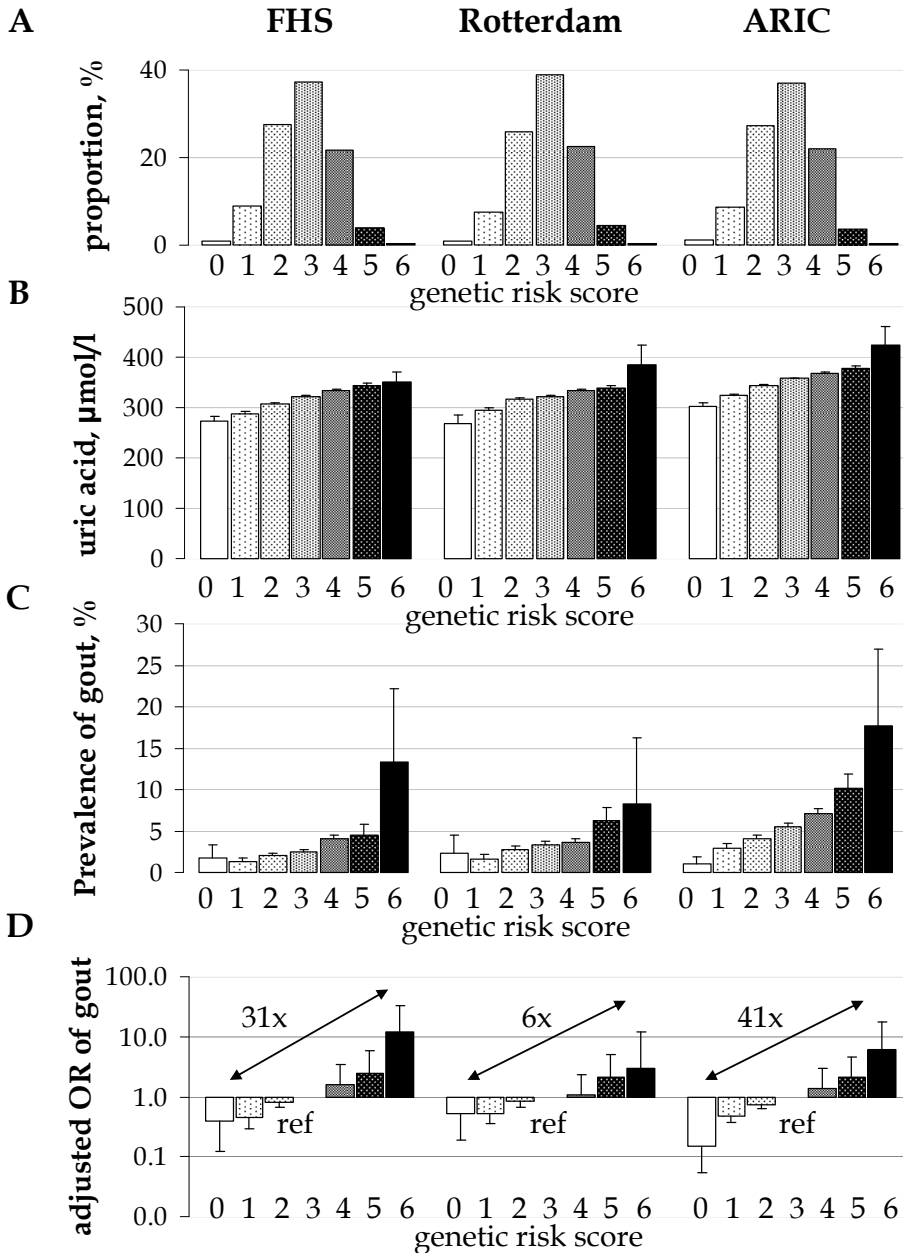
Figure 2. *Additive genetic risk score in the FHS, Rotterdam, and ARIC Studies. (A): prevalence of the genetic risk score; (B): mean serum uric acid, µmol/l; (C): prevalence of gout, %; (D) Odds ratio (OR) of gout, adjusted for age, sex, BMI, alcohol intake, antihypertensive medication, cohort in FHS and study center in ARIC. Results are presented for white ARIC participants only. Error bars present standard errors. Prevalence is period prevalence in the Rotterdam Study.*

Results

Table 1 shows the characteristics of 26,714 participants. SNPs genotyped in all three studies met quality control standards. Three loci had SNPs that reached genome-wide significance in the Framingham cohort. For each locus, the most significant SNPs were rs16890979 (a missense SNP in SLC2A9), rs2231142 (a missense SNP in ABCG2), and rs1165205 (intron 1 of SLC17A3) (table 2). Similarly, two loci showed genome-wide significance in the Rotterdam cohort: rs6449213 (intron 4 of SLC2A9) and rs2231142 (a missense SNP in ABCG2).

Figure 1 shows SNPs at the ABCG2 locus. The SLC17A3 locus shows linkage disequilibrium in white participants, extending downstream of SLC17A3 to include SLC17A1 and SLC17A4. The SLC17A3–SLC17A1–SLC17A4 region is referred to as SLC17A3 because of the location of the most associated variant.

Both rs16890979 and rs2231142 were strongly associated with uric acid concentration in white and black participants (table 2). rs1165205 was strongly associated with uric acid in white participants only (table 2).

All meta-analysis p values for uric acid reached genome-wide significance (table 2). rs16890979 explained the largest variation in uric acid concentration, ranging from 2.8% to 5.3% in white participants across studies. The total R^2 of all three SNPs (rs16890979, rs2231142, and rs1165205) in explaining uric acid concentration was 5.8% (white participants of in ARIC study), 2.4% (black participants in ARIC study), 7.1% (Framingham cohort), and 3.7% (Rotterdam cohort). Conditional on the top SNPs, only SNPs in the SLC2A9 region, one in the Framingham cohort and two in the Rotterdam cohort, remained significant.

Study-specific results for gout were direction-consistent with uric acid concentration (table 2). rs16890979 was associated with gout in white individuals from all three studies. Results showing significance were also seen for rs2231142 and rs1165205, and for rs6449213. In black individuals of the ARIC study, only rs2231142 in ABCG2 showed a marginal association with gout (table 2). Individually, the Framingham cohort and the Rotterdam cohort did not show any genome-wide significant findings for gout.

Secondary analyses further adjusted uric acid results for diabetes, systolic blood pressure, and estimated glomerular filtration rate. Results were not materially changed. After adjustment of gout results for uric acid, attenuation of ORs for gout was seen, although most loci retained significance. In the Framingham cohort, only rs2231142 remained associated with gout after adjustment for uric acid (OR 1.57, 1.14–2.16, p < 0.0053). All other loci did not retain significance. In the Rotterdam cohort, no SNP was significant after adjustment for serum uric acid. In the ARIC study, substantial attenuation of the genotypic effect for all three loci on gout risk was seen after adjustment for uric

acid. p values decreased from 1.8×10^{-9} to 2.4×10^{-4} for rs16890979; from 2.0×10^{-7} to 1.7×10^{-3} for rs2231142; and from 3.0×10^{-3} to 0.015 for rs1165205.

The genetic risk score (risk alleles 0–6) showed common variation of the population (figure 2). Mean uric acid concentration increased linearly with the number of risk alleles (figure 2). For individuals with no risk alleles, the crude prevalence of gout was 1–2% across studies and increased to 8–18% for those with six risk alleles (figure 2). The multivariable adjusted ORs of gout increased accordingly across the risk scores in the three studies (figure 2).

Substantial gene-by-sex interaction was seen for rs16890979 and rs2231142. rs16890979 had a stronger association with uric acid in women than it had in men in all three studies. Data from the ARIC study are presented in figure 3. For rs2231142, the T allele was associated with both higher uric acid concentration and ORs for gout in men than in women (figure 3). The sex-specific R^2 (proportion of variance explained [men to women]) was for rs16890979: 2.0% to 8.8% (Framingham cohort), 1.4% to 4.1% (Rotterdam cohort), 1.7% to 7.6% (ARIC study for white participants), 0.5% to 3.4% (ARIC study for black participants); for rs2231142: 2.1% to 0.8% (Framingham cohort), 1.6% to 0.5% (Rotterdam cohort), 2.0% to 0.6% (ARIC study for white participants), 0.4% to 0.3% (ARIC study for black participants).

We did not see any significant interaction of any tested SNP with age, body-mass index, alcohol intake, or hypertension treatment.
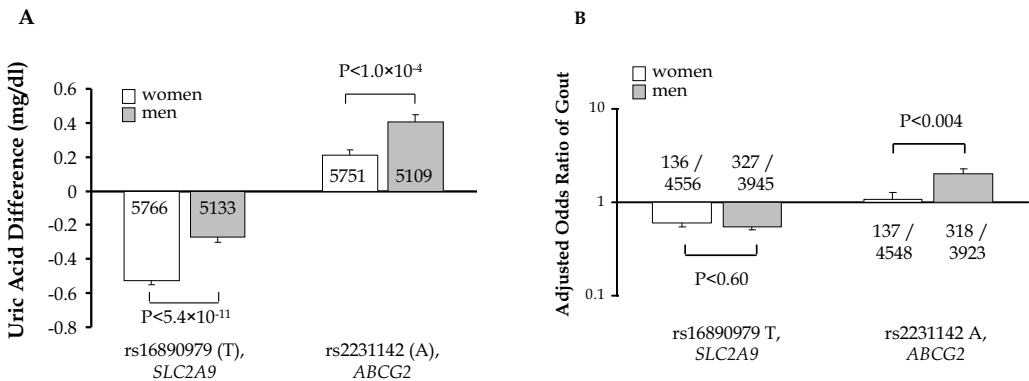

Figure 3.
Interaction of sex with SLC2A9 rs16890979 and ABCG2 rs2231142 on uric acid levels and gout risk. Multivariable adjusted (A) difference in mean uric acid levels; (B) odds ratio of gout. Results are presented for ARIC whites. Error bars represent standard errors. Numbers inside/next to bars present sample size (uric acid) and number of gout cases / sample size (gout).

Discussion

We identified two new loci, ABCG2 and SLC17A3, that show association with uric acid concentration and risk of gout. A missense SNP in ABCG2 (rs2231142; Q141K) was associated with uric acid concentration and gout in both white and black individuals and may be a causal candidate variant. Furthermore, we confirmed the previously reported association of variation in SLC2A9 with uric acid and gout in white individuals, and extended the findings to black individuals. Also, we described sex-specific effects of SNPs in ABCG2 and SLC2A9. Finally, we showed that an additive genetic risk score has strong and graded associations with uric acid concentration and gout in three population-based studies.

SNPs in SLC2A9 have been recently identified as being associated with uric acid concentration[12-15], and were connected with low renal fractional excretion of uric acid[15], which is the most common cause of hyperuricaemia[33]. We showed that the missense SNP rs16890979 in SLC2A9 has the strongest association with uric acid concentration and gout. This SNP leads to a valine-to-isoleucine aminoacid substitution (V253I). The valine residue is highly conserved across species. This association was also present in black participants in the ARIC study, in whom the linkage disequilibrium pattern differs. However, previous sequencing efforts with uric acid did not support the hypothesis that rs16890979 is the causal SNP in the region, because in 541 individuals from Sardinia this SNP was only slightly associated with uric acid concentrations (p < 0.02)[12]. Therefore, the potential causal role of this missense SNP remains unclear.

The apparent importance of renal urate transport influencing uric acid concentrations and subsequently gout is supported by the other two genetic loci we identified. ABCG2 encodes a transporter of the ATP-binding cassette (ABC) family[34]. Like SLC2A9, ABCG2 is expressed in the apical membrane of human kidney proximal tubule cells35, and transports purine nucleoside analogues, which resemble the molecular structure of uric acid36. We observed the strongest association with uric acid levels and gout in both white and black individuals with the ABCG2 missense SNP rs2231142. This SNP in exon 5 leads to a glutamine-to-lysine amino acid substitution (Q141K); the glutamine residue is highly conserved across species. Based on the FHS data, rs2231142 was not grouped into any LD block. Three other SNPs located downstream of and in disequilibrium with the Q141K variant were associated with uric acid, two of which are located in the PKD2 gene. However, neither these SNPs nor other SNPs in the region were independently associated with uric acid conditional on the Q141K variant in either FHS or Rotterdam. Combining this evidence with the relatively weak LD pattern in the ABCG2 region in the HapMap Yoruban

sample and the significant association in ARIC blacks despite the low minor allele frequency of 3%, suggests that the ABCG2 Q141K variant (rs2231142) could be causally related to uric acid levels.

SLC17A3 encodes a sodium phosphate (Na/Pi) transporter (NPT4), the rat homologue of which localizes to the apical membrane of renal proximal tubule cells[37]. Several prior studies have investigated the role of SLC17A1, located directly downstream of SLC17A3. SLC17A1 encodes NPT1, which is expressed in the human kidney and has been shown to transport urate in model systems38. In our study, the association of rs1165205 in SLC17A3 was weaker with uric acid levels compared to the other loci. In FHS, the missense SNP rs1165196 (T269I) in exon 7 of SLC17A1 also showed genome-wide significant association with uric acid levels ($p < 6.24 \times 10^{-10}$ in FHS; $p < 0.003$ in Rotterdam). This SNP was not in the same LD block as rs1165205, but both SNPs were in high pair-wise LD ($r^2 = 0.9$ in FHS). Additionally, the observed non-replication of rs1165205 with uric acid and gout among the black ARIC participants may allow for some degree of fine-mapping of the observed association. It is therefore conceivable that one or more causal genetic variants may be located downstream of SLC17A3, possibly in SLC17A1 or even further downstream in SLC17A4 due to the extensive LD in this region.

Although the gout risk conferred by the individual common genetic variants was modest, their combination resulted in a large effect on uric acid and gout prevalence. Further, the minor allele frequencies were common, suggesting that variants with low effect sizes will impact a large proportion of the population. Individual risk variants were associated with up to a 70% increased risk of developing gout, with effect sizes similar to that of known environmental risk factors[1]. Our genetic risk score was associated with up to a 40-fold increased risk of developing gout, substantially higher than environmental risk factors, suggesting that knowledge of genotype may help identify individuals at risk for developing gout long before the onset of clinical disease. This underscores the value of a one-time assessment of the genetic risk score, whereas the measurement of uric acid is subject to measurement error and physiologic variability over time.

In additional to risk prediction, knowledge of an individual's genotype or risk score could be used to help guide clinical decision making, especially with respect to the selection of medications known to increase uric acid levels and precipitate gout. Currently, gout prophylaxis for asymptomatic hyperuricemia is not recommended39, but it is conceivable that our genetic risk score could be used to identify individuals in which asymptomatic hyperuricemia should be treated. Since treatment decisions are best guided by randomized trials, stored specimens from existing trials should be tested to directly estimate

how this discovery of an easily determined strong genetic risk gradient can lead to personalized medicine. It is also possible that the genetic risk score, or certain genes that comprise it, differentially associate with gout complications, particularly joint destruction or poor response to medications.

Although novel agents for lowering uric acid such as febuxostat[40] are promising, allopurinol remains the mainstay of treatment for gout[41]. The efficacy of allopurinol can be limited by drug dosing and intolerance, drug-drug interactions, and treatment failure[7]. Errors are frequently made in allopurinol use[2], and only 21% of patients randomized to allopurinol in a clinical trial achieved optimal uric acid levels[40]. The genes identified here may provide the opportunity for the identification of novel proteins and molecular mechanisms influencing uric acid levels, and the opportunity for the discovery of needed novel drug targets in order to ultimately improve the treatment of gout. Limitations to our study include the self-reported ascertainment of gout in FHS and ARIC, which could lead to misclassification and underestimation of the true magnitude of the genotype-phenotype association. We used slightly different definitions of gout across studies. Nonetheless, the findings remain consistent, highlighting their robustness. Hyperuricemia may have influenced the diagnosis of gout in our sample. However, gout was not ascertained at the same time that uric acid was measured; therefore, this is unlikely to account for the joint association of the SNPs with uric acid levels and gout. We note that the association between the SNPs and gout was not completely attenuated by adjustment for uric acid levels, which may be due to the fact that uric acid levels were measured before the onset of gout in the majority of cases. Due to the limited power for GWA study for gout in this setting, we focused our genetic analyses on uric acid levels and only related SNPs for uric acid to gout. Therefore, there are likely to be additional loci for gout that we have not detected. Finally, we assigned identical risk to each allele in creating the genetic risk score for ease of interpretation, as done previously[42].

References

1. Choi HK, Mount DB, Reginato AM. Pathogenesis of gout. Ann Intern Med. 2005;143:499-516.

2. Mikuls TR, Farrar JT, Bilker WB, Fernandes S, Saag KG. Suboptimal physician adherence to quality indicators for the management of gout and asymptomatic hyperuricaemia: results from the UK General Practice Research Database (GPRD). Rheumatology (Oxford). 2005;44:1038-42.

3. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, Gabriel S, Hirsch R, Hochberg MC, Hunder GG, Jordan JM, Katz JN, Kremers HM, Wolfe F. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis Rheum. 2008;58:26-35.

4. Krishnan E, Lienesch D, Kwoh CK. Gout in ambulatory care settings in the United States. J Rheumatol. 2008;35:498-501.

5. Wu EQ, Patel PA, Yu AP, Mody RR, Cahill KE, Tang J, Krishnan E. Disease-related and all-cause health care costs of elderly patients with gout. J Manag Care Pharm. 2008;14:164-75.

6. Roddy E, Zhang W, Doherty M. The changing epidemiology of gout. Nat Clin Pract Rheumatol. 2007;3:443-9.

7. Sundy JS, Hershfield MS. Uricase and other novel agents for the management of patients with treatment-failure gout. Curr Rheumatol Rep. 2007;9:258-264.

8. Taniguchi A, Kamatani N. Control of renal uric acid excretion and gout. Curr Opin Rheumatol. 2008;20:192-7.

9. Saag KG, Choi H. Epidemiology, risk factors, and lifestyle modifications for gout. Arthritis Res Ther. 2006;8 Suppl 1:S2.

10. Eraly SA, Vallon V, Rieg T, Gangoiti JA, Wikoff WR, Siuzdak G, Barshop BA, Nigam SK. Multiple Organic Anion Transporters Contribute to Net Renal Excretion of Uric Acid. Physiol Genomics. 2008.

11. Yang Q, Guo CY, Cupples LA, Levy D, Wilson PW, Fox CS. Genome-wide search for genes affecting serum uric acid levels: the Framingham Heart Study. Metabolism. 2005;54:1435-1441.

12. Li S, Sanna S, Maschio A, Busonero F, Usala G, Mulas A, Lai S, Dei M, Orru M, Albai G, Bandinelli S, Schlessinger D, Lakatta E, Scuteri A, Najjar SS, Guralnik J, Naitza S, Crisponi L, Cao A, Abecasis G, Ferrucci L, Uda M, Chen WM, Nagaraja R. The GLUT9 Gene Is Associated with Serum Uric Acid Levels in Sardinia and Chianti Cohorts. PLoS Genet. 2007;3:e194.

13. Doring A, Gieger C, Mehta D, Gohlke H, Prokisch H, Coassin S, Fischer G, Henke K, Klopp N, Kronenberg F, Paulweber B, Pfeufer A, Rosskopf D, Volzke H, Illig T, Meitinger T, Wichmann HE, Meisinger C. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. Nat Genet. 2008.

14. Wallace C, Newhouse SJ, Braund P, Zhang F, Tobin M, Falchi M, Ahmadi K, Dobson RJ, Marcano AC, Hajat C, Burton P, Deloukas P, Brown M, Connell JM, Dominiczak A, Lathrop GM, Webster J, Farrall M, Spector T, Samani NJ, Caulfield MJ, Munroe PB. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. Am J Hum Genet. 2008;82:139-149.

15. Vitart V, Rudan I, Hayward C, Gray NK, Floyd J, Palmer CN, Knott SA, Kolcic I, Polasek O, Graessler J, Wilson JF, Marinaki A, Riches PL, Shu X, Janicijevic B, Smolej-Narancic N,

Gorgoni B, Morgan J, Campbell S, Biloglav Z, Barac-Lauc L, Pericic M, Klaric IM, Zgaga L, Skaric-Juric T, Wild SH, Richardson WA, Hohenstein P, Kimber CH, Tenesa A, Donnelly LA, Fairbanks LD, Aringer M, McKeigue PM, Ralston SH, Morris AD, Rudan P, Hastie ND, Campbell H, Wright AF. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. Nat Genet. 2008.

16. Stark K, Reinhard W, Neureuther K, Wiedmann S, Sedlacek K, Baessler A, Fischer M, Weber S, Kaess B, Erdmann J, Schunkert H, Hengstenberg C. Association of common polymorphisms in GLUT9 gene with gout but not with coronary artery disease in a large case-control study. PLoS One. 2008;3:e1948.

17. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. 1951;41:279-281.

18. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci. 1963;107:539-556.

19. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med. 1975;4:518-525.

20. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol. 1979;110:281-290.

21. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol. 2007;165:1328-1335.

22. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991;7:403-22.

23. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uiterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819-829.

24. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989;129:687-702.

25. SNP Health Association Resource. In.

26. Crowley LV. Determination of Uric Acid: An Automated Analysis Based on a Carbonate Method. Clin Chem. 1964;10:838-44.

27. Trivedi R, Rebar L, Berta E, Stong L. New enzymatic method for serum uric acid at 500 nm. Clin Chem. 1978;24:1908-1911.

28. Iribarren C, Folsom AR, Eckfeldt JH, McGovern PG, Nieto FJ. Correlates of uric acid and its association with asymptomatic carotid atherosclerosis: the ARIC Study. Atherosclerosis Risk in Communities. Ann Epidemiol. 1996;6:331-40.

29. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genet Epidemiol. 2000;19 Suppl 1:S36-42.

30. Rabinowitz D, Laird N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered. 2000;50:211-23.

31. Schwarzer G. The Meta package. In. 0.8-2 ed: CRAN; 2007.

32. Ihaka R. R: a language for data analysis and graphics. J Comput Graph Stat. 1996;5:299 - 314.

33. Terkeltaub R, Bushinsky DA, Becker MA. Recent developments in our understanding of the

renal basis of hyperuricemia and the development of novel antihyperuricemic therapeutics. Arthritis Res Ther. 2006;8 Suppl 1:S4.

34. Kusuhara H, Sugiyama Y. ATP-binding cassette, subfamily G (ABCG family). Pflugers Arch. 2007;453:735-744.

35. Huls M, Brown CD, Windass AS, Sayer R, van den Heuvel JJ, Heemskerk S, Russel FG, Masereeuw R. The breast cancer resistance protein transporter ABCG2 is expressed in the human kidney proximal tubule apical membrane. Kidney Int. 2008;73:220-225.

36. Takenaka K, Morgan JA, Scheffer GL, Adachi M, Stewart CF, Sun D, Leggas M, Ejendal KF, Hrycyna CA, Schuetz JD. Substrate overlap between Mrp4 and Abcg2/Bcrp affects purine analogue drug cytotoxicity and tissue distribution. Cancer Res. 2007;67:6965-6972.

37. Ishibashi K, Matsuzaki T, Takata K, Imai M. Identification of a new member of type I Na/phosphate co-transporter in the rat kidney. Nephron Physiol. 2003;94:p10-18.

38. Uchino H, Tamai I, Yamashita K, Minemoto Y, Sai Y, Yabuuchi H, Miyamoto K, Takeda E, Tsuji A. p-aminohippuric acid transport at renal apical membrane mediated by human inorganic phosphate transporter NPT1. Biochem Biophys Res Commun. 2000;270:254-259.

39. Campion EW, Glynn RJ, DeLabry LO. Asymptomatic hyperuricemia. Risks and consequences in the Normative Aging Study. Am J Med. 1987;82:421-6.

40. Becker MA, Schumacher HR, Jr., Wortmann RL, MacDonald PA, Eustace D, Palo WA, Streit J, Joseph-Ridge N. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. N Engl J Med. 2005;353:2450-2461.

41. Underwood M. Diagnosis and management of gout. Bmj. 2006;332:1315-1319.

42. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, Hirschhorn JN, Berglund G, Hedblad B, Groop L, Altshuler DM, Newton-Cheh C, Orho-Melander M. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008;358:1240-9.

Chapter 4.3

Association of novel genetic loci with circulating fibrinogen levels: a genome-wide association study in six population-based cohorts

Abstract

Background

Fibrinogen is both central to blood coagulation and an acute phase reactant. We aimed to identify common variants influencing circulation fibrinogen levels.

Methods

We conducted a genome-wide association analysis on six population-based studies, the Rotterdam Study, the Framingham Heart Study, the Cardiovascular Health Study, the Atherosclerosis Risk in Communities Study, the MONICA/KORA Augsburg Study, and the British 1958 Birth Cohort Study, including 22,096 participants of European ancestry.

Results

Four loci were marked by one or more single nucleotide polymorphisms (SNPs) that demonstrated genome-wide significance ($p<5.0\times10^{-8}$). These included a SNP located in the fibrinogen β chain (*FGB*) gene and three SNPs representing newly identified loci. The high-signal SNPs were rs1800789 in exon 7 of *FGB* ($p < 1.8\times10^{-30}$), rs2522056 downstream from the interferon regulatory factor 1 (*IRF1*) gene ($p < 1.3\times10^{-15}$), rs511154 within intron 1 of the propionyl coenzyme A carboxylase (*PCCB*) gene ($p < 5.9\times10^{-10}$), and rs1539019 on the NLR family, pyrin domain containing 3 isoforms (*NLRP3*) gene ($p < 1.04\times10^{-8}$).

Conclusions

Our findings highlight biological pathways that may be important in regulation of inflammation underlying cardiovascular disease.

Introduction

Elevated levels of fibrinogen within or above the normal range are consistently associated with an increased risk of cardiovascular disease[1]. Fibrinogen has a key role in blood coagulation but is also known as a marker of inflammation. Studies in persons of European ancestry have estimated the heritability of multivariable-adjusted fibrinogen levels from 24% in multiplex families[2] to more than 50% in twins[3]. The three genes encoding the three fibrinogen protein chains explain only a small part of the total estimated genetic variance of circulating levels of fibrinogen[4].

The objective of this study was to identify novel genetic loci related to plasma fibrinogen levels. A meta-analysis of genome-wide association (GWA) findings was conducted on six population-based studies. We analyzed GWA data of 2,661,766 SNPs from one or more studies from a total of 22,096 participants of European descent.

Methods

The setting for this meta-analysis is primarily the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium.[5] CHARGE includes the Rotterdam Study (RS), the Framingham Heart Study (FHS), the Cardiovascular Health Study (CHS), and the Atherosclerosis Risk in Communities (ARIC) Study. In addition, data from the British 1958 Birth Cohort (B58C) and the MONICA/KORA Augsburg Study (KORA) has been included.

Rotterdam Study (RS)
The RS is a prospective, population-based cohort study of determinants of several chronic diseases in older adults[6]. In brief, the study comprised 7,983 inhabitants of Ommoord, a district of Rotterdam in the Netherlands, who were 55 years or over. The baseline examination took place between 1990-1993.

Genotyping was conducted using the Illumina 550K array. SNPs were excluded for minor allele frequency ≤1%, Hardy-Weinberg equilibrium (HWE) $p<10^{-5}$, or SNP call rate ≤90% resulting in data on 530,683 SNPs. Imputation was done with reference to HapMap release 22 CEU using the maximum likelihood method implemented in MACH. The final population for this fibrinogen analysis comprised 2,068 individuals.

Framingham Heart Study (FHS)
The FHS started in 1948 with 5,209 randomly ascertained participants from Framingham, Massachusetts, US, who had undergone biannual examinations

to investigate cardiovascular disease and its risk factors[7]. In 1971, the Offspring cohort[8,9] (comprised of 5,124 children of the original cohort, and the children's spouses) and in 2002, the Third Generation (consisting of 4,095 children of the Offspring cohort), were recruited[10]. FHS participants in this study are of European ancestry.

Genotyping was carried-out as a part of the SHARe project using the Affymetrix 500K mapping array (250K Nsp and 250K Sty arrays) and the Affymetrix 50K supplemental gene focused array on 9,274 individuals. Genotyping resulted in 503,551 SNPs with successful call rate >95% and HWE $p>10^{-6}$ on 8481 individuals with call rate >97%. Imputation of ~2.5 million autosomal SNPs in HapMap with reference to release 22 CEU sample was conducted using the algorithm implemented in MACH. The final population for fibrinogen analysis included 7,022 individuals (Original Cohort n=383, Offspring n=2,806, Third Generation n=3,833).

Cardiovascular Health Study (CHS)

The CHS is a population-based, observational study of risk factors for clinical and subclinical cardiovascular diseases[11]. The study recruited participants 65 years of age and older from four US communities in two phases: 5,201 participants in 1989-1990, and 687 (primarily African American participants) in 1992-1993. A GWA study was conducted in a subset of CHS participants (n=3,980), all of whom were without clinical cardiovascular disease at their baseline clinical visit and provided consent to use their DNA for research. The study sample used in the fibrinogen analysis represented the first two of three rounds of genotyping, which was a stratified probability sample. Weights were assigned to each observation to reflect the likelihood of sampling from the 3,980 participants. The analysis was restricted to participants of European decent.

Genotyping was performed using the Illumina 370 CNV BeadChip system. Samples were excluded for sex mismatch, discordance with prior genotyping, or call rate <95%. SNPs were excluded from analysis when monomorphic, when HWE $p<10^{-5}$, and when call rates were <95%. Imputation was performed using BIMBAM v0.95 with reference to HapMap CEU using release 21A build. The population available for the fibrinogen analysis included 1,993 individuals.

The Atherosclerosis Risk in Communities (ARIC)

The ARIC study is a longitudinal cohort study of atherosclerosis and its clinical sequelae. It recruited a population-based sample of 15,792 men and women aged 45-64 years from four US communities in 1987-89[12]. The analysis was restricted to subjects of European decent.

Genotyping was performed using the Affymetrix Genome-Wide Human

SNP Array 6.0. SNPs were excluded for not being autosomal SNPs, not passing laboratory QC, no chromosome location, minor allele frequency ≤1%, SNP call rate <90%, or HWE p<10^{-6}. This resulted in data on 716,442 SNPs. Imputation to HapMap SNPs was performed using MACH. After excluding subjects who disallowed DNA use, subjects with a mismatch between called and phenotypic sex, with a mismatch on >10 of 47 previously analyzed SNPs in ARIC, all but one in sets of first degree relatives, and other individuals who were genetic outliers, the final population for fibrinogen analysis comprised 8,051 individuals.

The MONICA/KORA Augsburg Study (KORA)

The presented data were derived from the third population-based Monitoring of Trends and Determinants in Cardiovascular Disease (MONICA)/Cooperative Health Research in the Region of Augsburg (KORA) survey S3[13]. This cross-sectional survey covering the city of Augsburg (Germany) and two adjacent counties was conducted in 1994/95 to estimate the prevalence and distribution of cardiovascular risk factors among individuals aged 25 to 74 years as part of the WHO MONICA study. The MONICA/KORA S3 study comprises 4,856 subjects. Among them, 3,006 subjects participated in a follow-up examination of S3 in 2004/05 (MONICA/KORA F3). All participants underwent standardized examinations including blood withdrawals for plasma and DNA. For the KORA genome-wide association study, 1,644 subjects, aged 45 to 69 years were selected from the KORA S3/F3 samples.

Genotyping was performed using the Affymetrix 500K Array Set. Samples were excluded for sex mismatch, discordance with prior genotyping, or call rate <95%. SNPs were excluded from analysis when monomorphic (MAF<0.01), when call rates per SNP were <0.1 and per individual were <0.1. Imputation was done using maximum likelihood method implemented in MACH 1.0. The final population available for the fibrinogen analysis included 1,523 individuals.

British 1958 birth cohort (B58C)

The B58C is a national population sample followed periodically from birth. At age 44-45 years, 9,377 cohort members were examined by a research nurse in the home as described previously[14]. For this study we used a total of 1,480 cell-line-derived DNA samples from unrelated subjects of European ancestry, with nationwide geographic coverage, which were used as controls by the Wellcome Trust Case Control Consortium (WTCCC)[15].

Genotyping was performed using the Affymetrix 500K Mapping Array Set using the call algorithm CHIAMO as implemented by the WTCCC[15]. Genotypes at other loci were imputed by the program IMPUTE version 0.1.2, using 490,032

autosomal SNPs with CHIAMO calls and the linkage disequilibrium patterns in the HapMap CEU panel. Analysis of imputed genotypes used Marchini's SNPTEST version 1.1.3 and supplementary regression modeling used STATA version 10.0. A final sample size of 1,459 individuals was included in the fibrinogen analysis.

Measurement of fibrinogen

In the KORA study, fibrinogen was determined by an immunonephelometric method (Dade Behring Marburg GmbH, Marburg, Germany) on a Behring Nephelometer II analyzer. FHS study used the Clauss method[16] in the offspring and the third generation subjects, and a modified method of Ratnoff and Menzie[17] in the original cohort subjects. In the RS, fibrinogen levels were derived from the clotting curve of the prothrombin time assay using Thromborel S as a reagent on an automated coagulation laboratory 300 (ACL 300, Instrumentation Laboratory, Zaventem, Belgium). The other studies used the Clauss method for measuring plasma fibrinogen[16].

Statistical analysis

Each study independently analyzed their genotype-phenotype data. Except for FHS, which has a family structure, all studies conducted analyses of all directly genotyped and imputed SNPs using linear regression on untransformed fibrinogen measures using an additive genetic model adjusted for age, sex, and site of recruitment (if necessary). In FHS, a linear mixed effects model was employed with a fixed additive effect for the SNP genotype, fixed covariate effects, random family specific additive residual polygenic effects to account for within family correlations[18], and a random environment effect. In addition, FHS adjusted for population stratification using principal components of the directly measured SNPs which were computed using the Eigenstrat software.

To account for residual stratification, p-values were adjusted for genomic inflation. The inflation of the association test statistic, stated as inflation factor lambda (λ_{gc}), was small for all studies: 0.995 for RS, 1.016 for FHS, 1.031 for CHS, 1.024 for ARIC, 1.012 for KORA, and 1.008 for B58C. Using the study-specific results, we conducted a fixed effect model meta-analysis based on inverse-variance weighting. MetABEL, a package running under R was used to perform the meta-analysis. We used Bonferroni correction to deal with the problem of multiple testing. Simulation studies show that the effective number of independent tests in a GWA analysis is nearly one million[19]. Based on one million tests, we selected a p-value threshold of 5×10^{-8} as the level of genome-wide significance.

In addition, we estimated the effect of the top SNPs in strata of sex and

Table 1 – Characteristics of the study participants

	RS	FHS	CHS	ARIC	KORA	B58C
Number	2,068	7,022	1,993	8,051	1,523	1,459
Age, Mean (SD) (y)	70.8 (9.0)	46.6 (11.5)	73.2 (5.8)	54.3 (5.6)	52.1 (10.2)	44.9 (0.4)
Male (%)	36.8	46.1	44.4	47.1	49.3	50.0
Body mass index, Mean (SD) ((kg/m2)	26.5 (3.9)	27.0 (5.2)	26.4 (4.5)	27.0 (4.9)	27.3 (4.1)	27.4 (4.8)
Current smoker (%)	22.2	18.9	11.3	25.1	15.4	23.0
Alcohol drinker (%)	77.1	74.5	52.2	44.8	70	94.6
Systolic blood pressure, Mean (SD) (mmHg)	138.2 (21.4)	121.1 (17.1)	137.7 (21.7)	118.5 (17.0)	133.0 (18.3)	126.7 (15.3)
Diastolic blood pressure, Mean (SD) (mmHg)	72.2 (11.2)	75.0 (10.0)	70.6 (11.5)	71.7 (10.0)	81.9 (10.9)	79.1 (10.2)
Total cholesterol, Mean (SD) (mmol/l)	6.7 (1.3)	5.1 (1.0)	5.5 (1.0)	5.5 (1.1)	6.1 (1.1)	5.9 (1.0)
HDL cholesterol, Mean (SD) (mmol/l)	1.3 (0.4)	1.4 (0.4)	1.4 (0.4)	2.8 (0.9)	1.4 (0.4)	1.5 (0.4)
Prevalent cardiovascular disease (%)	7.9	12.6	0	6.7	0.7	4.3
Prevalent diabetes (%)	11.8	4.8	29.7	8.5	3.7	1.9
Hypertension (%)	34.8	11.5	37.9	19.8	23.5	21.4
Fibrinogen, Mean (SD) (g/L)	2.7 (0.7)	3.2 (0.7)	3.2 (0.6)	2.97 (0.6)	2.9 (0.7)	3.0 (0.6)

RS: The Rotterdam Study; FHS: The Framingham Heart Study; CHS: The Cardiovascular Health Study; ARIC, The Atherosclerosis Risk in Communities Study; KORA: The MONICA/KORA Augsburg Study; B58C: British 1958 birth cohort

Chapter 4.3

Table 2 - Association of the top SNPs in four loci with plasma fibrinogen levels

		RS	FHS	CHS	ARIC	KORA	B58C	Combined
rs1800789 (A/G)	Sample size	2,068	7,022	1,993	8,051	1,523	1,458	22,096
Chromosome: 4	P-value	2.6×10^{-05}	8.0×10^{-13}	1.5×10^{-02}	2.2×10^{-09}	1.0×10^{-04}	1.2×10^{-01}	1.8×10^{-30}
Location: 155702193	Beta *	0.108	0.103	0.179	0.068	0.097	0.042	0.087
Gene: FGB	Standard error	0.026	0.014	0.036	0.011	0.025	0.027	0.008
Modeled Allele: A	Obs./exp. variance	0.955	0.929	0.845	0.981	0.874	0.958	-
	R square (%)	0.8	0.8	2.5	0.4	0.8	0.2	-
	Coded allele frequency	0.20	0.21	0.21	0.22	0.24	0.20	-
rs2522056 (G/A)	Sample size	2,068	7,022	1,993	7,856	1,523	1,459	21,901
Chromosome: 5	P-value	2.4×10^{-01}	2.6×10^{-05}	6.2×10^{-04}	5.6×10^{-07}	4.8×10^{-03}	4.0×10^{-03}	1.3×10^{-15}
Location: 131829625	Beta*	-0.031	-0.058	-0.168	-0.060	-0.076	-0.082	-0.063
Gene: IRF1	Standard error	0.026	0.014	0.049	0.012	0.027	0.029	0.008
Modeled Allele: A	Obs./exp. variance	0.979	0.979	0.442	1.000	0.940	0.968	-
	R square (%)	0.1	0.3	1.6	0.4	0.5	0.6	-
	Coded allele frequency	0.19	0.21	0.17	0.209	0.21	0.18	-
rs511154 (A/G)	Sample size	2,068	7,022	1,993	8,051	1,523	1,459	22,096
Chromosome: 3	P-value	2.0×10^{-02}	2.1×10^{-03}	3.2×10^{-03}	1.7×10^{-04}	8.3×10^{-01}	1.3×10^{-02}	5.9×10^{-10}
Location: 137433611	Beta*	0.057	0.041	0.089	0.043	0.006	0.063	0.045
Gene: PCCB	Standard error	0.024	0.013	0.030	0.011	0.027	0.026	0.007
Modeled Allele: A	Obs./exp. variance	0.996	1.017	0.993	0.989	0.973	0.945	-
	R square (%)	0.3	0.1	0.7	0.2	< 0.1	0.4	-
	Coded allele frequency	0.22	0.24	0.23	0.22	0.21	0.24	-
rs1539019 (A/C)	Sample size	2,068	7,022	1,993	7,773	1,523	1,434	21,818
Chromosome: 1	P-value	4.2×10^{-02}	9.8×10^{-04}	9.9×10^{-01}	3.1×10^{-04}	6.5×10^{-03}	1.6×10^{-01}	1.0×10^{-08}
Location: 245666924	Beta*	-0.043	-0.041	0.000	-0.037	-0.061	-0.030	-0.038
Gene: NLRP3	Standard error	0.021	0.013	0.027	0.010	0.022	0.022	0.007
Modeled Allele: A	Obs./exp. variance	0.990	0.846	0.97	1.000	0.664	0.829	-
	R square (%)	0.2	0.2	< 0.1	0.3	0.4	0.1	-
	Coded allele frequency	0.39	0.40	0.41	0.37	0.42	0.41	-

* The beta coefficient is for an age and sex adjusted model

Obs./exp. variance : observed to expected variance; RS: The Rotterdam Study; FHS: The Framingham Heart Study; CHS: The Cardiovascular Health Study; ARIC, The Atherosclerosis Risk in Communities Study; KORA: The MONICA/KORA Augsburg Study; B58C: British 1958 birth cohort

smoking status. Gene-by-sex and gene-by-smoking interaction was tested in each study by introducing an interaction term into the linear model. We used a sample size weighted meta-analysis to combine the reported interaction p-values across studies for each of the top SNPs.

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

Replication in Women's Health Genome Study (WHGS)

We used the WHGS to replicate our genome wide significant findings and other loci for which our meta-analysis generated more modest evidence of an association (p-value of 5×10^{-7}). Participants in WGHS are derived from the genetic arm of the Women's Health Study and include American women with no prior history of cardiovascular disease, cancer, or other major chronic illness who provided a baseline blood sample during the enrollment phase of the Women's Health Study between 1992 and 1995[20]. Fibrinogen levels were measured using an immunoturbidimetric assay (Kamiya Biomedical, Seattle, Wash), which was standardized to a calibrator from the World Health Organization. Genotyping was done using the Illumina Infinium II assay to query a genome-wide set of 315,176 haplotype-tagging SNP markers (Human HAP300 panel) as well as a focused panel of 45,882 missense and haplotype
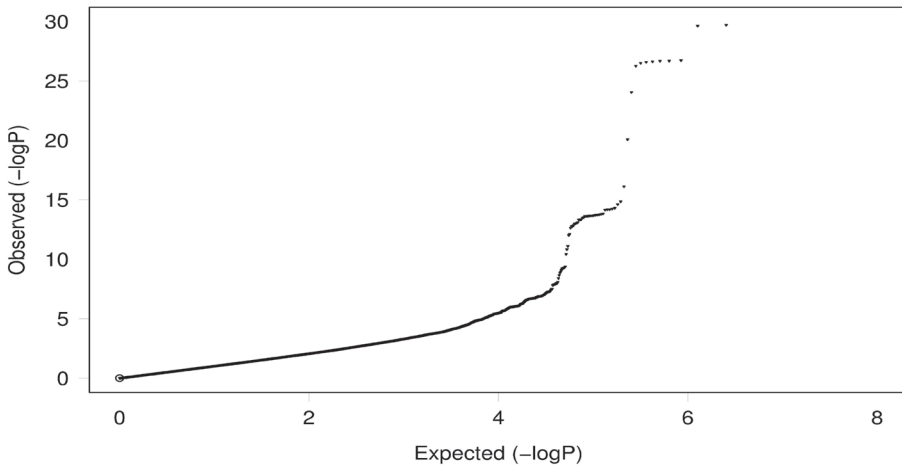
Figure 1.
QQ-plot for the meta-analysis results. Quantile-quantile plot of the observed and the expected distribution of p-values for all 2,661,766 SNPs and their association with fibrinogen levels based on meta-analyzed data.

tagging SNPs. For this analysis, the evaluation was performed on 17,686 non-diabetic individuals who were of Caucasian ancestry and were not taking lipid-lowering agents. The GWA results of the WHGS are reported in a companion manuscript.

Results

The sample size and participant characteristics from each study are shown in Table 1. A quantile-quantile plot (Q-Q plot) of the observed against expected p-value distribution is shown in Figure 1. Figure 2 illustrates the primary findings from the meta-analysis and presents p-values for each of the interrogated SNPs across the 22 autosomal chromosomes. A total of 73 SNPs (supplemental Table 1) exceeded the threshold of genome-wide significance and clustered around four loci on chromosomes 1 (2 SNPs), 3 (12 SNPs), 4 (23 SNPs), and 5 (36 SNPs) (Figure 3).

The strongest statistical evidence for an association was for rs1800789 which is located at 4q31.3 in exon 7 of the fibrinogen ß (FGB) gene (minor allele frequency [MAF]: 0.20-0.24, meta-analysis p-value = 1.75×10^{-30}, fibrinogen level change per minor allele [D]: 0.087 g/L). The other significant loci were marked by rs2522056, which is located at 5q23.3, 25 kb downstream of the interferon
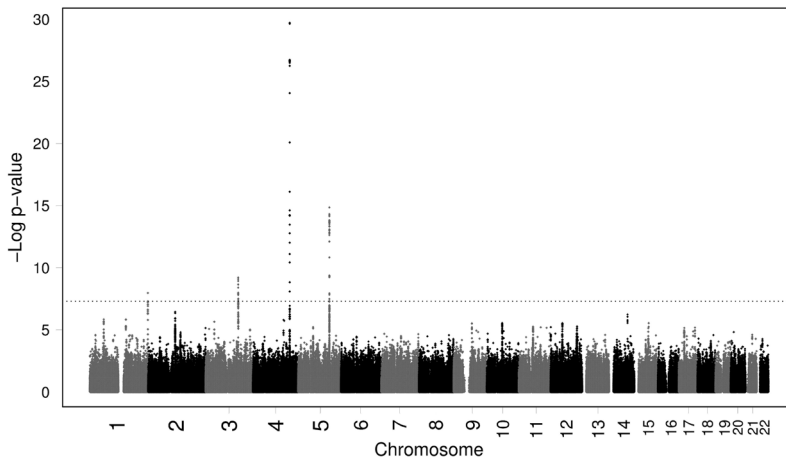

Figure 2.
Log plot for the meta-analysis. -Log p-values for each of the 2,661,766 tests performed as part of the genome-wide association analysis of fibrinogen levels. The grey dashed horizontal lines correspond to the p-value threshold of 5×10^{-8} and the grey solid line corresponds to 5×10^{-6}.

Table 3 – Mean(SD) plasma fibrinogen level (g/L) by genotype

SNP	Study	Sample size	Fibrinogen	Sample size	Fibrinogen	Sample size	Fibrinogen
			GG		AG		AA
rs1800789	RS	1,328	2.8 (0.69)	652	2.9 (0.66)	88	3.0 (0.69)
	FHS	4,489	3.2 (0.69)	2,250	3.3 (0.66)	283	3.4 (0.70)
	CHS	1,290	3.1 (0.62)	624	3.2 (0.63)	79	3.3 (0.61)
	ARIC	4,908	2.9 (0.62)	2,730	3.0 (0.59)	413	3.0 (0.67)
	KORA	884	2.8 (0.64)	533	3.0 (0.70)	106	3.0 (0.66)
	B58C	928	2.9 (0.61)	471	3.0 (0.55)	59	3.0 (0.60)
			GG		AG		AA
rs2522056	RS	1,349	2.8 (0.68)	641	2.8 (0.69)	78	2.7 (0.72)
	FHS	4,364	3.3 (0.68)	2353	3.2 (0.66)	305	3.1 (0.74)
	CHS	1,239	3.2 (0.61)	754	3.1 (0.63)	0	NA
	ARIC	4,946	2.9(0.62)	2,592	2.9 (0.61)	318	2.8(0.63)
	KORA	951	2.9 (1.01)	493	2.8 (1.01)	79	2.7 (1.02)
	B58C	971	3.0 (0.59)	438	2.9 (0.58)	50	2.9 (0.59)
			GG		AG		AA
rs511154	RS	1,253	2.8 (0.67)	706	2.8 (0.70)	109	2.9 (0.74)
	FHS	4047	3.2 (0.67)	2,536	3.3 (0.70)	439	3.2(0.66)
	CHS	1,206	3.2 (0.60)	687	3.2 (0.63)	100	3.3 (0.69)
	ARIC	4,925	2.9 (0.61)	2,712	2.9 (0.62)	414	3.0 (0.62)
	KORA	946	2.9 (0.68)	503	2.9 (0.63)	74	3.0 (0.62)
	B58C	834	2.9 (0.60)	541	3.0 (0.57)	84	3.1 (0.59)
			CC		CA		AA
rs1539019	RS	762	2.8 (0.69)	987	2.8 (0.69)	319	2.8 (0.65)
	FHS	2,419	3.3(0.68)	3,428	3.2(0.69)	1,175	3.2 (0.65)
	CHS	686	3.2 (0.63)	991	3.2 (0.61)	316	3.2 (0.61)
	ARIC	3,041	2.9 (0.63)	3,721	2.9 (0.60)	1,011	2.9 (0.60)
	KORA	530	3.0 (0.73)	715	2.9 (0.63)	278	2.8 (0.60)
	B58C	524	3.0 (0.62)	653	2.9 (0.57)	257	2.9 (0.59)

RS: The Rotterdam Study; FHS: The Framingham Heart Study; CHS: The Cardiovascular Health Study; ARIC, The Atherosclerosis Risk in Communities Study; KORA: The MONICA/KORA Augsburg Study; B58C: British 1958 birth cohort

regulatory factor 1 (IRF1) gene (MAF: 0.17-0.21, p = 1.3×10^{-15}, D: -0.063 g/L), rs511154, which is located at 3q22.3, in intron 1 of the propionyl coenzyme A carboxylase, beta polypeptide (PCCB) gene (MAF: 0.21-0.24, p = 5.94×10^{-10}, D: 0.045 g/L) and rs1539019 which is located at 1q44, on the NLR family, pyrin domain containing 3 isoforms (NLRP3) gene (MAF: 0.37-0.42, p = 1.04×10^{-8}, D: -0.038 g/L). Cohort-specific findings are presented for the top SNP within
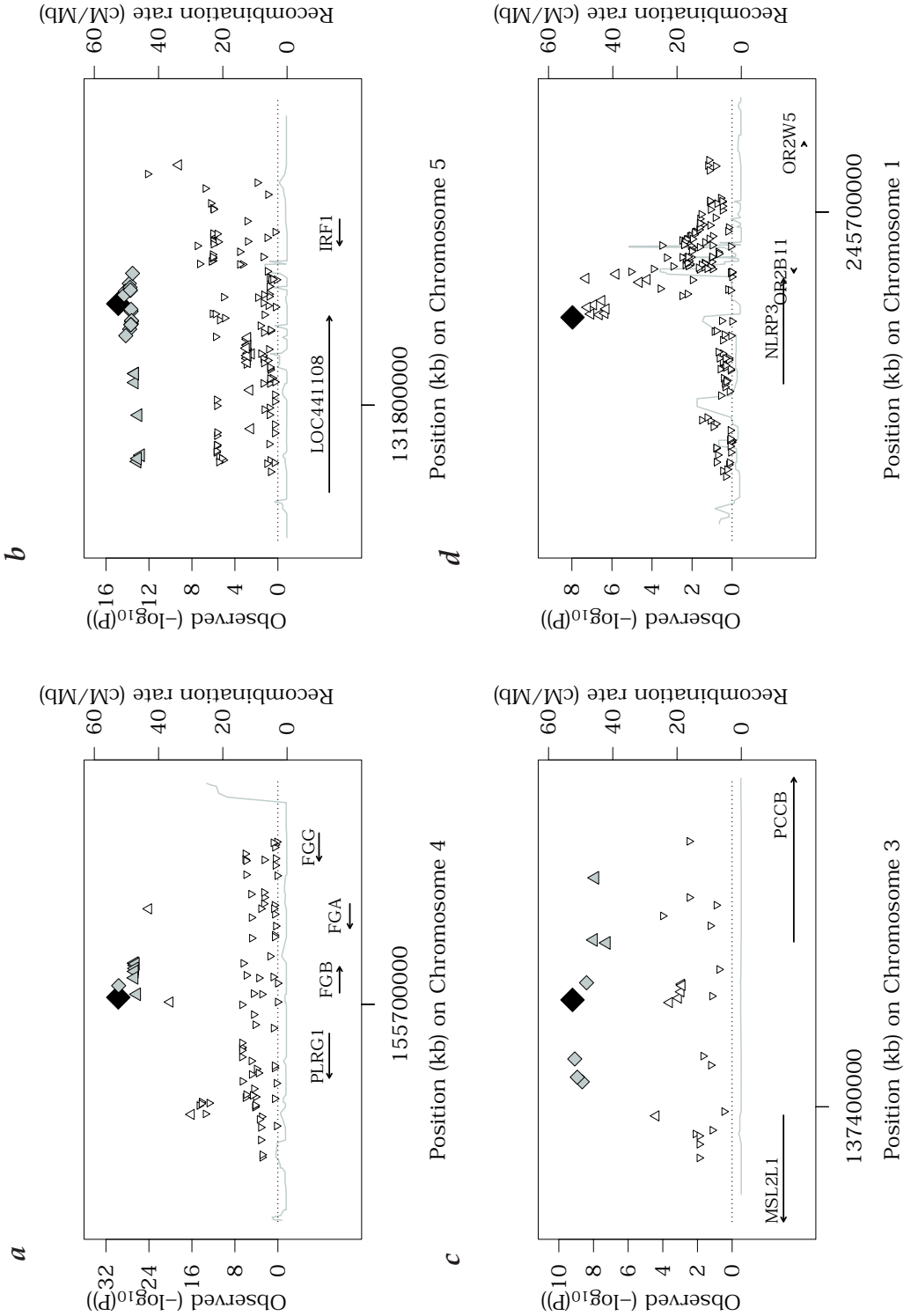
each locus in Table 2. Results did not change materially when we adjusted the model for other covariates (smoking, alcohol consumption, body mass index, systolic blood pressure, triglyceride, total- and HDL-cholesterol, diabetes, and cardiovascular disease) (data not shown). Table 3 shows the mean and standard deviations for fibrinogen levels by genotype for each of the four SNPs.

We estimated the association of the four SNPs by sex and smoking status separately but none of the SNPs showed a significantly different association between subgroups (Supplementary Table 2 and 3).

A combined risk alleles score summarizing the number of risk alleles was associated with a 15% increase in overall mean fibrinogen level comparing subjects with no risk allele (mean fibrinogen level 2.81 g/L) to subjects with six or more risk alleles (mean fibrinogen level 3.24 g/L). The genetic variants identified in our study explained less than 2% of the overall variance in plasma fibrinogen in all studies except one.

To investigate the validity of our findings, we sought replication of the four loci using WHGS data. Since WHGS did not genotype the identical SNPs as our six cohorts, the best proxy SNP was used for replication. For rs1800789, rs2522056, rs511154, and rs1539019, we used WHGS SNPs rs6056 (r^2=0.95; p=8.04×10^{-39}), rs1016988 (r^2=0.80; p=1.24×10^{-12}), rs684773 (r^2=1.0; p=1.92×10^{-5}), and rs1539019 (p=2.89×10^{-4}), respectively, as the proxy SNP. The direction of each association in WHGS was consistent with our findings.

In addition to our four genome-wide significant loci, two other loci demonstrated multiple-SNP hits with p-values <5×10^{-7}: one on chromosome 2 (rs4251961, p=3.5×10^{-7}) and one on chromosome 14 (rs8017049, p= 5.6×10^{-7}). When we examined the results for these two loci in the WHGS data, we found evidence for replication on chromosome 2 (rs4251961 in WHGS, p=8.5×10^{-3}).

Figure 3.
Regional plots of loci associated with fibrinogen. (a-d). The association p-values (-log10 transformed, indicated by the left y-axis) for SNPs in a 100kb region of each of the four loci (FGB, IRF1, PCCB, and NLRP3) are plotted against their chromosome positions (NCBI build 36) on x-axis. The black diamond represents the SNP with smallest p value. The linkage disequilibrium (estimated using HapMap CEU sample) between each SNP and the top associated SNP is illustrated by color and the shape of each SNP. The gray diamond indicates complete LD, gray triangle with point-up indicates LD between 0.8 and 1.0, a black diamond with point up indicates LD between 0.8 and 0.3, a white triangle indicates LD between zero and 0.3 and white point-down triangles indicate no LD. The light gray line shows the estimated recombination rates with values indicated by the right y-axis. The bottom panel displays the genes in the region based on the UCSC Genome Browser March 2006 assembly, with the arrow to right (left) indicate indicating +(-) strand.

Discussion

We identified four loci associated with circulating fibrinogen level through a meta-analysis of GWA data from six cohort studies comprising 22,096 subjects. We provide strong information of the previously reported associations with the FGB locus. Three of our findings are newly identified associations.

The most significant SNP in our study was rs1800789 which is located on the FGB gene. The FGB gene encodes the fibrinogen ß chain. A well-characterized SNP at this locus is rs1800787 (–148C/T) which resides 965 base pairs away from our top SNP (rs1800789) and is in high LD with it (D'~1.0, r^2=0.91). It is known that rs1800787 directly affects gene transcription in basal and IL6-stimulated conditions in luciferase expression studies[21]. Another well-characterized SNP in this region is rs1800790 (455G/A), which is also in strong LD with rs1800787, is known to be related to plasma fibrinogen22 and showed a strong association with fibrinogen levels in our study as well (p=5.04×10^{-27}, Supplementary Table 1).

The second locus is located 25 kb downstream from the IRF1 gene on chromosome 5. IRF1 is a member of the interferon regulatory transcription factor family and activates transcription of interferon α and ß. IRF1 also functions as a transcription activator of genes induced by interferon α, β and γ. Direct effects of interferons on fibrinogen have not previously been described, but it is known that they play a role in the regulation of acute phase proteins. Notably, the SNP is only 31 kb from a SNP strongly associated with Crohn's disease in a recent meta-analysis (rs2188962, p<2.32×10^{-18})[23]. Individuals with inflammatory bowel disease (IBD), including Crohn's disease, are at a threefold higher risk of venous thrombosis[24], accounting for substantial morbidity and mortality in this group[25]. Furthermore, multiple studies have indicated significantly elevated levels of fibrinogen in IBD patients[26]. This suggests that IRF1 or nearby genes may contribute to Crohn's disease via a mechanism mediated through an increase in acute phase responsiveness and fibrinogen levels.

The third locus on chromosome 3 is located in intron 1 of the PCCB gene. The PCCB gene is responsible for a particular step in the breakdown of the amino acids isoleucine, methionine, threonine, and valine. However, the available information about PCCB does not provide a strong hypothesis about the putative function of the gene in regulation of fibrinogen levels.

The fourth locus on chromosome 1 is located on the NLRP3 gene. The NLRP3 gene encodes a pyrin-like protein, which interacts with the apoptosis-associated speck-like protein PYCARD/ASC and is a member of the NALP3 inflammasome complex[27]. Activated NALP3 inflammasome drives processing of the pro-inflammatory cytokine pro-IL1ß to IL1ß. Recent data indicate that

the NALP3 inflammasome can be activated by endogenous 'danger signals' as well as compounds associated with pathogens and triggers an innate immune response[28].

The finding on chromosome 2 is located in the promoter region (1 kb upstream from the transcription start site) of the interleukin-1 receptor antagonist (IL1RN) gene. Fibrinogen is an acute phase protein that is regulated by cytokines, mainly IL1 and IL6, while the IL6-mediated transcription of the fibrinogen gene is inhibited by IL1ß[29]. This region has formerly reported to be associated with fibrinogen levels; rs2232354, which is in high LD with our top SNP, rs4251961, was associated with fibrinogen levels in an asymptomatic population[30].

Our findings were replicated in WGHS. Two of our four SNPs are reported by WGHS as genome-wide significant findings (rs6056 and rs1016988) and the other two have p-values which suggest non-chance findings in a replication (rs684773 and rs1539019). These results provide further credibility that our newly identified loci are valid.

We examined evidence for the top four fibrinogen loci among gene expression QTLs from recent GWA studies in human liver tissues[31] and lymphoblastoid cell lines32. In liver tissues, SNPs at the FGB locus were strongly associated with the expression of FGB (e.g., rs4508864, $p<1.20\times10^{-8}$) as well as with other trans-located mRNAs. Likewise, we observed that several SNPs in the region of the IRF1 locus were strongly associated with the expression of nearby genes (including IRF1, LOC441108, and SLC22A5) in both liver tissues and lymphoblastoid cell lines (e.g., rs2070729, $p=4.9\times10^{-10}$ for expression of the IRF1 gene). These results from independent genome-wide association studies strongly suggest a functional basis for the observed associations in the FGB and IRF1 loci.

Although heritability estimates for circulating fibrinogen are substantial, the genetic variants identified in our study explain only a small part of the overall variance. Therefore, our SNPs probably have limited value in prediction of cardiovascular disease. Rare variants, common variants with smaller effects, or variants which interact with other genetic and environmental factors may explain the remaining variation in plasma fibrinogen levels.

Fibrinogen was measured independently in the six cohorts. Though methods for measuring fibrinogen concentration were not standardized, they were all based on the Clauss method or another clotting assay, except for KORA which used nephelometry. Nonetheless, the effect estimates for the top SNPs were comparable between KORA and other studies.

Contributing studies used different genotyping platforms with different groups of SNPs. To enable the meta-analysis, each study imputed ~2.5 million SNPs in HapMap CEU samples. Imputation has previously been shown to be accurate and to increase the power. The power, of course, would have been

higher if all SNPs were genotyped in all studies.

In conclusion, we have identified four loci associated with fibrinogen levels through meta-analysis of GWA data from six cohort studies comprising 22,096 subjects. All four loci replicated in a seventh study. In addition, we replicated one of the two other loci which showed a close to significant association in our meta-analysis and is biologically plausible. Three of our findings (IRF1, PCCB, and NLRP3) represent newly identified associations. Among the genes in the novel loci implicated in our study are those that encode proteins playing a role in inflammation representing interesting targets for further research into biological pathways involved in cardiovascular disease and other chronic inflammatory conditions.

References

1. Danesh J, Lewington S, Thompson SG, Lowe GD, Collins R, Kostis JB, Wilson AC, Folsom AR, Wu K, Benderly M, Goldbourt U, Willeit J, Kiechl S, Yarnell JW, Sweetnam PM, Elwood PC, Cushman M, Psaty BM, Tracy RP, Tybjaerg-Hansen A, Haverkate F, de Maat MP, Fowkes FG, Lee AJ, Smith FB, Salomaa V, Harald K, Rasi R, Vahtera E, Jousilahti P, Pekkanen J, D'Agostino R, Kannel WB, Wilson PW, Tofler G, Arocha-Pinango CL, Rodriguez-Larralde A, Nagy E, Mijares M, Espinosa R, Rodriquez-Roa E, Ryder E, Diez-Ewald MP, Campos G, Fernandez V, Torres E, Marchioli R, Valagussa F, Rosengren A, Wilhelmsen L, Lappas G, Eriksson H, Cremer P, Nagel D, Curb JD, Rodriguez B, Yano K, Salonen JT, Nyyssonen K, Tuomainen TP, Hedblad B, Lind P, Loewel H, Koenig W, Meade TW, Cooper JA, De Stavola B, Knottenbelt C, Miller GJ, Bauer KA, Rosenberg RD, Sato S, Kitamura A, Naito Y, Palosuo T, Ducimetiere P, Amouyel P, Arveiler D, Evans AE, Ferrieres J, Juhan-Vague I, Bingham A, Schulte H, Assmann G, Cantin B, Lamarche B, Despres JP, Dagenais GR, Tunstall-Pedoe H, Woodward M, Ben-Shlomo Y, Davey Smith G, Palmieri V, Yeh JL, Rudnicka A, Ridker P, Rodeghiero F, Tosetto A, Shepherd J, Ford I, Robertson M, Brunner E, Shipley M, Feskens EJ, Kromhout D, Dickinson A, Ireland B, Juzwishin K, Kaptoge S, Memon A, Sarwar N, Walker M, Wheeler J, White I, Wood A. Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. Jama. 2005;294:1799-809.
2. Yang Q, Tofler GH, Cupples LA, Larson MG, Feng D, Lindpaintner K, Levy D, D'Agostino RB, O'Donnell CJ. A genome-wide search for genes affecting circulating fibrinogen levels in the Framingham Heart Study. Thromb Res. 2003;110:57-64.
3. de Lange M, Snieder H, Ariens RA, Spector TD, Grant PJ. The genetics of haemostasis: a twin study. Lancet. 2001;357:101-5.
4. Kathiresan S, Yang Q, Larson MG, Camargo AL, Tofler GH, Hirschhorn JN, Gabriel SB, O'Donnell CJ. Common genetic variation in five thrombosis genes and relations to plasma hemostatic protein level and cardiovascular disease risk. Arterioscler Thromb Vasc Biol. 2006;26:1405-12.
5. Psaty BM ODC, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, Uitterlinden AG, Harris TB, Witteman JCM, Boerwinkle E on behalf of the CHARGE Consortium. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. Circ Cardiovasc Genet; In Press.
6. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819-829.
7. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. Ann N Y Acad Sci. 1963;107:539-56.
8. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol. 1979;110:281-90.
9. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med. 1975;4:518-25.
10. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino

RB, Sr., Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol. 2007;165:1328-35.

11.   Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991;1:263-76.

12.   The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989;129:687-702.

13.   Wichmann HE, Gieger C, Illig T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005;67 Suppl 1:S26-30.

14.   Strachan DP, Rudnicka AR, Power C, Shepherd P, Fuller E, Davis A, Gibb I, Kumari M, Rumley A, Macfarlane GJ, Rahi J, Rodgers B, Stansfeld S. Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. Int J Epidemiol. 2007;36:522-31.

15.   Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661-78.

16.   Clauss A. [Rapid physiological coagulation method in determination of fibrinogen.] Gerinnungsphysiologische Schnellmethode zur Bestimmung des Fibrinogens. Acta Haematol. 1957;17:237-46.

17.   Kannel WB, Wolf PA, Castelli WP, D'Agostino RB. Fibrinogen and risk of cardiovascular disease. The Framingham Study. Jama. 1987;258:1183-6.

18.   Abecasis GR, Cardon LR, Cookson WO, Sham PC, Cherny SS. Association analysis in a variance components framework. Genet Epidemiol. 2001;21 Suppl 1:S341-6.

19.   Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008;32:381-5.

20.   Ridker PM, Chasman DI, Zee RY, Parker A, Rose L, Cook NR, Buring JE. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. Clin Chem. 2008;54:249-55.

21.   Verschuur M, de Jong M, Felida L, de Maat MP, Vos HL. A hepatocyte nuclear factor-3 site in the fibrinogen beta promoter is important for interleukin 6-induced expression, and its activity is influenced by the adjacent -148C/T polymorphism. J Biol Chem. 2005;280:16763-71.

22.   van der Bom JG, de Maat MP, Bots ML, Haverkate F, de Jong PT, Hofman A, Kluft C, Grobbee DE. Elevated plasma fibrinogen: cause or consequence of cardiovascular disease? Arterioscler Thromb Vasc Biol. 1998;18:621-5.

23.   Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield

J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008;40:955-62.

24. Bernstein CN, Blanchard JF, Houston DS, Wajda A. The incidence of deep venous thrombosis and pulmonary embolism among patients with inflammatory bowel disease: a population-based cohort study. Thromb Haemost. 2001;85:430-4.

25. Srirajaskanthan R, Winter M, Muller AF. Venous thrombosis in inflammatory bowel disease. Eur J Gastroenterol Hepatol. 2005;17:697-700.

26. Zilberman L, Rogowski O, Rozenblat M, Shapira I, Serov J, Halpern P, Dotan I, Arber N, Berliner S. Inflammation-related erythrocyte aggregation in patients with inflammatory bowel disease. Dig Dis Sci. 2005;50:677-83.

27. Petrilli V, Dostert C, Muruve DA, Tschopp J. The inflammasome: a danger sensing complex triggering innate immunity. Curr Opin Immunol. 2007;19:615-22.

28. Hornung V, Bauernfeind F, Halle A, Samstad EO, Kono H, Rock KL, Fitzgerald KA, Latz E. Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. Nat Immunol. 2008;9:847-56.

29. Zhang Z, Fuller GM. Interleukin 1beta inhibits interleukin 6-mediated rat gamma fibrinogen gene expression. Blood. 2000;96:3466-72.

30. Reiner AP, Wurfel MM, Lange LA, Carlson CS, Nord AS, Carty CL, Rieder MJ, Desmarais C, Jenny NS, Iribarren C, Walston JD, Williams OD, Nickerson DA, Jarvik GP. Polymorphisms of the IL1-receptor antagonist gene (IL1RN) are associated with multiple markers of systemic inflammation. Arterioscler Thromb Vasc Biol. 2008;28:1407-12.

31. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusis AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008;6:e107.

32. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WO. A genome-wide association study of global gene expression. Nat Genet. 2007;39:1202-7.

Chapter 4.2

Chapter 5

General discussion
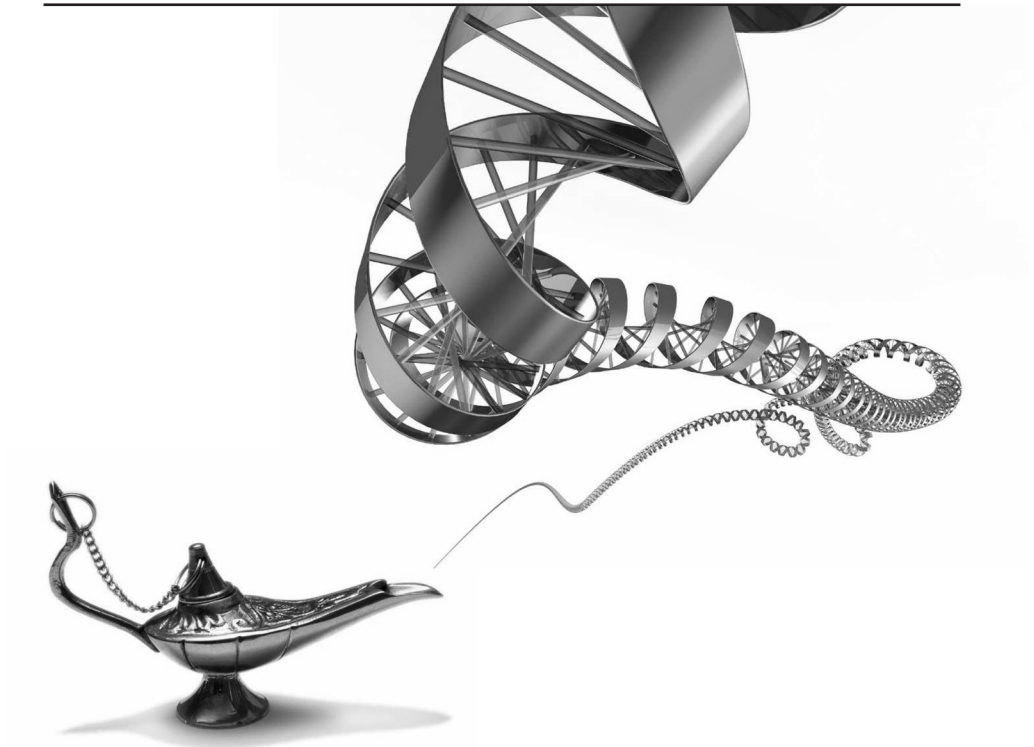
The main objective of the work described in this thesis was to expand our knowledge on novel risk factors for diabetes and coronary heart disease. We examined serum CRP and serum uric acid, as novel risk factors for type II diabetes and coronary heart disease (CHD). We further performed studies to identify genetic variants that affect CRP and uric acid and additionally also those affecting fibrinogen levels. We performed a number of studies in the frame-work of the Rotterdam study[1-3]. In genome-wide association (GWA) studies, we extended our work by participating in a consortium of population-based cohort studies, the CHARGE consortium.

In this discussion, the main findings are summarized. Furthermore, some methodological issues that came up in these studies will be addressed. Finally, potential clinical implications and views for future research are presented.

Review and interpretation of main findings

Novel risk factors for diabetes mellitus

In chapter 2.1, we provided support for an etiologic role of CRP in the pathogenesis of type II diabetes. We performed a meta-analysis to obtain evidence from the literature for an association between serum CRP and type II diabetes, independent of obesity. To further elaborate on the independent role of CRP in risk of diabetes, we showed that a genetically determined elevation of serum CRP is associated with a higher risk of type II diabetes. The method used to examine causality in observational genetic studies is Mendelian randomization. This approach will be discussed in detail in the section on methodological considerations. However, a recent study did not observe an increased risk of type II diabetes with genetically increased serum CRP[4]. Since the CCG haplotype which showed the association in our study was rather rare, one possibility is that our finding is due to chance. In chapter 2.2, we showed that a considerable proportion of type II diabetes cases could be attributed to high serum CRP.

Studies such as those described in chapters 2.1 and 2.2 provide evidence that serum CRP has a causal role in type II diabetes. To further elaborate this issue, we used Hill's criteria to categorize the present evidence (box1). These criteria are not used to judge on causality, but mere to evaluate the present evidence. The use of Hill's criteria in the process of judging about causation is further discussed in the paragraph on methodological considerations.

In chapter 2.3, we showed that higher levels of serum uric acid are associated with risk of developing type II diabetes. The association was strong and remained significant after adjustment for other risk factors for diabetes. In the same chapter, we estimated that one quarter of diabetes cases are attributable

to a high serum uric acid level. This finding, together with former studies on other diseases such as hypertension[5] and stroke[6], may point to serum uric acid as a novel treatment target for preventing diabetes and its complications. The finding may also make a case for randomized controlled clinical trials to assess the possible benefits of lowering serum uric acid on multiple chronic disease endpoints.

Novel risk factors for coronary heart diseases

Numerous epidemiologic studies have reported that CRP is an independent risk factor for cardiovascular disease[7]. In chapter 3.1, we examined the influence of high serum CRP and six traditional risk factors on risk of MI and CHD. Theoretically, a larger part of MI and CHD cases can be prevented by lowering serum CRP compared to any other traditional risk factor, except for serum cholesterol. Body mass index, total cholesterol, HDL-cholesterol, smoking, hypertension, and diabetes accounted for more than half of the risk of CHD in subjects 55 years and older.

In addition to these traditional risk factors, CHD has a familial basis; a positive family history increases the risk of CHD by a factor 2 to 4. Though candidate gene and linkage studies were successful in identifying causative mutations in monogenic cardiovascular diseases, they were not successful in identifying multiple causative genes that account for non-Mendelian diseases such as MI and CHD. GWA studies as the novel approach to investigate related genes are supposed to produce better replicable results. The first GWA studies on MI and CHD introduced a locus at 9p21.3 that was associated with risk of MI and CHD. This locus does not contain protein-coding genes, therefore it is likely that it affects the risk of MI and CHD through affecting a regulatory region of DNA. In chapter 3.2, we set out to replicate this association. We found no significant association between two SNPs from this locus and the risks of CHD and MI in the Rotterdam Study. Since this finding is by now replicated in many other populations, the lack of association in our study is of interest. The Rotterdam Study is a study in an older population. It is possible that genes involved in the development of CHD are different in young and older people. If that is the case, our finding is suggestive of a heterogeneity of the effect, which not only may be of interest for public health, but also may be of interest from an etiologic perspective.

Genome wide association studies of novel risk factors

In chapters 4.1, 4.2 and 4.3, we report results from three meta-analysis of GWA studies on novel risk factors including CRP, uric acid and fibrinogen levels. We confirmed six known loci[8] associated with circulating CRP levels (*CRP*, *APOE*

(*APOC1*), *HNF1A*, *LEPR*, *IL6R*, *GCKR*, and *ASCL*). We also identified three loci with uric acid level (*SLC2A9*, *ABCG2* and *SLC17A3*) and four loci with fibrinogen (*FGB*, *IRF1*, *PCCB*, and *NLRP3*).

The genes we found for CRP levels are related to immune response and metabolic regulation. Finding that genes related to metabolic pathways are involved in the regulation of CRP level is interesting, given that diet, exercise, and glucose control all lower CRP levels.

Interestingly, the genes we found for fibrinogen encode proteins that play a role in inflammation. Elevated levels of fibrinogen are a known sign of chronic inflammation and predict increased risks for MI and CHD. The identified genes may provide insight into novel biological pathways that connect chronic inflammation and CHD.

Uric acid levels are determined by its production (degradation of purine compounds) and its elimination. Elimination of uric acid occurs mainly through the kidneys. Our GWA study on uric acid revealed the role of renal transporters in the regulation of uric acid levels. Our finding provides support for the view that genetic differences in uric acid levels are more due to genetic differences in its elimination rate in the kidneys rather than to differences in its production.

Methodological considerations

Causation

Causation is one of the most difficult concepts to study. The concept is highly self-taught and our understanding of it has changed over decades[9]. There is now a philosophical agreement that causation cannot be proven. However, medical literature is full of epidemiologic studies that seek causal roles for risk factors. It seems that epidemiologists have agreed on a more pragmatic definition of causation. As a working definition, a cause could be defined as an antecedent event, condition, or characteristic without which the disease event either would not have occurred at all or would have occurred at a later time[10].

In spite of criticisms of inductive inference, inductive oriented considerations are still used in epidemiologic studies[9]. Bradford Hill's criteria are a set of requirements which could be used for this purpose. These criteria include strength of association, consistency of evidence (from different populations), specificity, temporal relationship, dose-response, biological plausibility, coherence, experimental evidence, and analogy[11]. Though these criteria do not prove a causal relation, its use helps epidemiologists to evaluate the available evidence for a causal role of a risk factor. Notably, Hill himself named these items as his "view points" and did not use the word "criteria". In this sense, causal criteria appear to function less like standards or principles and more like values[12].

Box 1

Hill's criteria for causality used for the example of CRP levels and risk of type II diabetes

Strength. In a meta-analysis of published studies, we showed that elevated CRP level predicts approximately a two to three fold increase in the risk of type II diabetes.

Consistency. An elevated serum CRP was found in most studies to be predictive for type II diabetes.

Specificity. We showed that a considerable proportion of type II diabetes cases could be attributed to high serum CRP. Moreover, serum CRP was higher than 1 mg/l in more than 80% of those who developed type II diabetes.

Temporality. An elevated serum CRP often precedes the development of type II diabetes.

Dose-dependent. The association of high serum CRP and risk of type II diabetes is continuous and dose-dependent.

Plausibility. Different biological mechanisms have been suggested by which elevated serum CRP may cause type II diabetes.

Coherence. The global increase in the prevalence of type II diabetes is associated with an increase in the prevalence of risk factors that increase the CRP levels such as obesity and smoking.

Experiment. The evidences for increased risk of type II diabetes associated with genetically elevated serum CRP is controversial.

Analogy. High serum CRP is also associated with risk of insulin resistance and metabolic syndrome.

Even as a set of "values" or "view points", Hill's criteria have certain pitfalls. Though these criteria were developed in the time that epidemiology had started studying chronic complex disorders, they are based on a mono-causal mental set that is possibly inherited from the infectious diseases era. As an

example, specificity would always fail when dealing with a multifactor causal model. Consistency may be missing due to lack of interacting factors in certain populations, i.e. heterogeneity of the effect. Idiosyncratic drug reactions and non-linear relations would also not match with a dose-response relationship. Despite these drawbacks, Hill's criteria are still in use[13,14].

Mendelian randomization

Observational studies, which comprise the majority of epidemiologic studies, may pose their own limitations when inferring a causal relation. Confounding effects and reverse causation are two major problems which may rise when investigating a causal relation in an observational study. Confounding bias happens when the effect estimate is distorted because it is mixed with or mistaken for the effect of an extraneous variable. Reverse causation is a process in which the disease occurs before the occurrence of the risk factor and the risk factor is present as a result of the disease.

A randomized controlled trial (RCT) design is the best design to overcome these drawbacks and to assess the causality of a risk factor. Confounding bias is solved with the help of randomization and reverse causation is not an issue due to the longitudinal design of RCTs. Though an RCT is a good choice for investigating causation, in many instances it is neither practical nor ethical to randomize human beings to certain interventions.

One alternative to an RCT is to perform an observational study with the so-called "Mendelian randomization" approach. In this approach, genetically elevated levels of the exposure are used to examine the non-confounded effect of the exposure on the outcome[15]. Though the Mendelian randomization approach is widely used, its pitfalls are sometimes not adequately acknowledged. Confounding by genetic factors (due to linkage disequilibrium) or population ancestry (population stratification), canalization, lack of power, and invalidity by pleiotropy are the most known limitations.

However, there are more considerations that should be taken into account.

To appropriately apply the Mendelian Randomization method, we need a functional genetic variant that associates with the exposure of interest. Recently, GWA studies have largely increased the number of genes known for each trait. One may think that these variants are all potential instruments for Mendelian randomization studies. However, the genetic variant should be specific for the exposure of interest. Non-specific genetic variants increase the chance that the variant is associated with the disease through pathways that do not involve the exposure. This would invalidate the results of a Mendelian randomization study. For instance, genetic variation in a leptin receptor gene, *LEPR*, is found to be associated with CRP levels[8]. This variant may seem to be an attractive candidate for a Mendelian randomization study on CRP and CHD, however, *LEPR* is possibly related to CHD through pathways that do not involve CRP. Moreover, GWA studies develop genetic risk scores that discriminate people with high and low

levels of the trait. Though these risk scores are very attractive due to their potential powerful effect on the exposure, the genetic variants are located on different genes and may confound the gene-disease association by altering the risk of the disease through other pathways (pleiotropy effect).

Most of the genetic variants used in Mendelian randomization studies only explain a very small proportion of the variance of the exposure. When using such variants as an instrument, researchers may try to achieve enough power by increasing the sample size. Nevertheless, recent studies on instrumental variables show that extreme weakness of the instrument could not be compensated by a larger sample size[16].

Population attributable risk

Population attributable risk (PAR), the proportion of the disease risk that can be attributed to a risk factor, is used to express the importance of a risk factor from an etiologic or public health point of view. Though PAR is widely used, there are several issues that should be considered when interpreting the estimates.

To estimate PAR, subjects should be labeled as exposed and unexposed. Dichotomous risk factors make this categorization by definition. When the risk factor is measured on a continuous scale, the exposed category is often defined using a cut-off point. Most of the time, the cut-off point is arbitrary and leaves the definition of the exposure to be broad or restricted. Such broad or restricted definitions may lead to incorrect conclusions about the importance of a risk factor. There are examples in the literature where a broad definition has spuriously overestimated the PAR[17]. This consideration also extends to genetic studies. Genetic studies, namely GWA studies, estimate the PAR for their findings to demonstrate the proportion of disease cases that theoretically could be prevented if the adverse effects of the genetic exposure were eliminated. For instance, one of the first GWA studies published on type II diabetes claimed that their findings together with a known gene, *TCF7L2*, would explain 70% of the incident diabetes cases[18]. One should note that the proportion of individuals exposed in genetic studies is determined by the allele frequency and the PAR estimate is affected by the frequency of the risk allele. When the risk allele is very common, the number of people who should receive an intervention to prevent the disease is high and may reach to nearly the whole population[19]. Moreover, misinterpreting a high PAR may lead to a misimpression that the majority of the disease genes are identified despite the fact that many (unknown) genetic variants may exist that interact with the currently known genes and environmental factors.

The PAR estimates may be misinterpreted as the etiologic fraction[20]. The PAR formula only allows to calculate the excess fraction, i.e. the proportion of the diseased cases that would not happen if the exposure was eliminated. Etiologic fraction, however, is the proportion of the diseased cases in which the exposure played a causal role. While excess fraction is more of interest from a

public health point of view, the etiologic fraction is important in an etiologic framework. One should note that all excess cases are etiologic cases, but not vice versa. Therefore, the excess fraction is always smaller or utmost equal to the etiologic fraction.

The combined PAR is the proportion of the diseased cases which would not happen if none of the risk factors were present. One important consideration is that PAR estimates for individual risk factors cannot be simply added together to calculate the combined PAR. The combined PAR is usually lower than the sum of individual PARs since a diseased case can simultaneously be attributed to more than one risk factor. Dual attribution happens since risk factors interact with each other. An exact combined PAR could only be estimated, if we had precise information on the magnitude of the interaction between the risk factors. Such information lacking is most of the time, therefore, we assume that the risk factors are only interacting with each other on a multiplicative scale. This is an important assumption which lies behind most of combined PAR estimates and should be considered when interpreting the results.

There are different computational formulas to estimate the PAR. Due to this diversity, errors happen in estimating PAR[21]. Two common errors are to use the hazard ratios instead of cumulative risk ratios[20] or use of the inappropriate formula for adjusted ratios[22]. PAR is the proportion of the disease cases that could be prevented within a distinct period of time. Hazard ratios express the point effect of a risk factor and lack information on the time period. Using adjusted ratios in the crude PAR formula is also problematic since the fraction of the outcome that is attributable to the confounder is not adequately accounted for[23].

Risk prediction

One effective way of preventing chronic diseases is to target preventive actions towards high-risk groups. To this end, it is essential to identify high-risk groups in the most accurate way. In recent decades, a number of novel risk factors, both environmental and genetic, are found for type II diabetes and CHD. Nevertheless, researchers have not succeeded to improve the performance of traditional prediction models by adding these novel risk factors[24,25], although some of them are strongly associated with risk of the disease. This discrepancy has led to extensive discussions about the evaluation methods of the prediction models. Here, we discuss a number of these considerations.

The c statistic is the most popular method of assessing the performance of a prediction model, however, it is argued that it does not give a comprehensive evaluation of its performance[26,27]. A prognostic model should be evaluated both for discrimination and calibration. Discrimination is the ability of a model to

separate individuals who will develop the disease from those who will not. Calibration indicates how close the predicted probabilities are to the observed cumulative incidence rates. The c statistic is only a measure of discrimination. Therefore, the c statistic should not be the only tool to judge whether novel risk factors offer additional value in prognostic models[28].

Several new statistical methods have been proposed to evaluate the contribution of a novel risk factor to a prognostic model. Cook et al introduced the reclassification table method. Reclassification examines how individuals are assigned to categories of risk and how this assignment is altered by the addition of a new risk marker[27]. Pepe et al proposed a new graphic, the predictiveness curve, which is claimed to complement the risk model[29]. Pencina et al introduced two measures to quantify improvement in the correct assignment to risk categories in reclassification tables. The Net Reclassification Improvement (NRI) requires an a priori risk category and the Integrated Discrimination Improvement (IDI) extends the idea to the case of "no cut-offs"[30]. One should note that these two concepts measure discrimination and not calibration.

Considerations about genome-wide association studies
Percentage of variance explained and the missing inheritability

Though GWA studies have identified many genetic variants for complex disorders, these variants only explain a small proportion of their heritability. Most speculation in the literature about the "missing" heritability focused on the possible contribution of rare variants. "Next generation" sequencing technologies are on the way and researchers envision that in the near future they will be able to shed more light on the unexplained heritability, which is nowadays termed as the 'dark matter' of GWA studies.

Sequencing studies are thought to increase the percentage of heritability explained in two ways. The first is to find novel genetic variants with minor allele frequencies (MAF) less than 5% and substantial effect sizes (risk ratios of two to three)[31]. The second way is to find the functional variants of the known loci which will have a larger effect than the currently identified variants. Though sequencing is likely to some extent to succeed in both ways, it is possible that the unexplained heritability still remains as an unsolved problem for several reasons.

Firstly, it is not clear how many unidentified rare variants with large effects exist. The power in GWA studies depends both on the MAF and the effect size. While increasing the sample size in GWA studies should increase the chance of discovery for both rare variants and the variations with modest or small effect, the largest meta-analysis GWA studies have more frequently identified variants with modest effect rather than rare variants.

Figure 1
Three hypothetical suficient causes of a disease
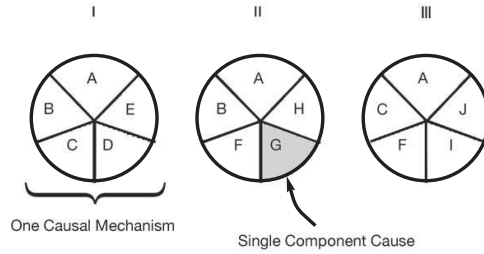
Secondly, rare variants with large effects could also be identified by current GWA studies. As an example, in a recent GWA study on lipid levels, 11 out of 30 loci which were found to carry common variants associated with lipid levels also carried rare variants known to cause Mendelian forms of dyslipidemia [32]. Therefore, it is not clear what proportion of the rare variants with large effects is left unidentified.

Thirdly, sequencing will provide detailed data on the order of DNA nucleotides; however, the majority of neighboring variants are in strong linkage disequilibrium (LD) with each other. The LD between genetic variants will make it difficult to disentangle the effect of these variants. To overcome this problem, we may need to study populations with shorter LD blocks such as Africans, or isolated populations who may have unique variants in their LD blocks. One other solution would be to pool the rare variants. Up to now, genetic variants are analyzed individually. If we can pool these variants based on biological or functional clusters, the carriers of the rare alleles are added together and the power of our study would be much larger[33].

The missing heritability is the difference between the heritability index and the combined variation explained by all identified genes. The variability that is due to the genes could either be due to the effect of genes or be due to the interaction between the genes and environment. In estimating the heritability index we assume that there is no gene-environment interaction and we calculate the maximum contribution of genetic factors to the trait. In other words, the heritability index also includes the proportion of variation that is caused by gene-environment interaction. Therefore, it could be speculated that the missing heritability is at least partly due to the contribution of gene-environment interaction.

Gene-environment interaction

GWA studies have identified hundreds of loci for complex diseases and their risk factors. However, hardly any genome-wide significant interaction has been reported that widely is replicated in different populations. Lack of robust

findings in gene-environment interaction analysis is usually attributed to lack of power in the identification study. Though most of these studies are indeed underpowered, other methodological and statistical considerations should be acknowledged.

Due to modest effects, gene-environment interaction could be investigated only with adequate information of the life-time exposure. To date, a great proportion of the GWA studies on complex disorders are comprised of case-control studies. Though it is easier to collect large sample sizes in a case-control design, adequate information on former exposure to environmental factors usually lacks in such studies.

A common approach to investigate gene-environment interaction is to examine the interaction of genome-wide significant or suggestive loci with known environmental risk factors. Although it seems common sense to examine the interaction between the most promising predictors of the disease, there could also be downsides to this approach as described below.

According to Rothman's sufficient cause model, a disease occurs when all the component causes are present (figure 1). The set of component causes in a sufficient cause that completes a causal factor is called the causal complement of that factor. It is known that the magnitude of the effect for a risk factor depends on the availability of its causal complement. When the causal complement is rare, the risk factor has a small effect and when it is abundant, the effect is large[10].

For many years, epidemiologic studies have searched for strong risk factors which consistently show the association in different populations. This means that most of the known risk factors have no genetic component in their complement factor or the necessary genetic variant is present in the majority of the population. Likewise, genetic variants which are found in GWA studies are probably not highly dependent on the existence of an environmental exposure since their effect is usually consistent in different age groups and populations. Therefore, choosing the genome wide significant genes may not be the best approach in searching for gene-environment interaction. Though the most powerful method should be figured out statistically, it could be speculated that genetic variations with modest but heterogeneous effect are more promising.

Replication and significance level

GWA studies are hypothesis free and their primary function is hypothesis generating. Replication is a critical step that improves the validity of the findings[34]. When the findings of a GWA study are replicated in several independent studies, the risk of a false-positive report diminishes and the validity of the finding is established. To provide a replication sample,

some GWA studies employ a two-stage design; a discovery panel and a replication panel. The discovery panel is normally comprised of studies with GWA information. The most promising findings of the discovery panel are genotyped in the replication panel. There are debates over a few methodological and statistical aspects of this approach. Here we review some of these considerations.

The first issue concerns the selection of SNPs taken to the next stage. Most studies take the genome-wide significant and a number of the suggestive SNPs to the next stage. The definition of suggestive loci is arbitrary. Using the p-value threshold where the Q-Q plot deviates from the diagonal line is reasonable. From this point onwards, true positive associations are expected. However, the definition may practically depend on biologic, logistic or financial considerations. Another issue is the number of SNPs that should be chosen from each locus. There is normally more than one significant or suggestive SNP per locus and the researcher should decide on the number of SNPs that should be replicated per locus. When the locus is too wide or the correlation between the SNPs is low, there is a need to choose more than one SNP, however, there is no definite guideline and the decision usually is made on arbitrary grounds.

Second, there is no consensus on how to interpret the results in the replication panel. Using a Bonferroni correction to adjust for the number of tests is criticized since the number of SNPs taken to the second phase is arbitrary. Some studies combine the results of both panels and use the genome-wide significance threshold for p-value level, i.e. 5×10^{-8}, or compare the discovery phase and the combined results and look for improved p-values. Though these latter approaches may be successful in discovering further loci, it should not be taken as replication.

In the GWA studies described in this thesis, we included all studies in the discovery panel. One practical reason was that all these studies had genome-wide data and this approach is more powerful[35]. We have sought replication in studies that could not be included in our discovery panel. In chapter 4.2, the Rotterdam Study and the Framingham Heart Study constructed the discovery panel and the ARIC study which did not have GWA scans at that point in time replicated the results. In chapter 4.3, we replicated our findings in the Women's Genomic Health Study, an independent population with data on the relevant phenotype. Chapter 4.1 describes a study that confirmed findings of a published GWA study and therefore did not need further replication.

Need for collaborative work in GWA studies

In a GWA study, we scan the majority of the variation in the genome of thousands of unrelated individuals and examine it in relation to the phenotype

of interest, free of any prior hypotheses. Though the approach is innovative and powerful, there is a need for large sample sizes - to prevent false negative findings - and replication - to prevent false positive findings. Since these two requirements are most of the time beyond the possibilities of a single study, collaborative work between the studies is necessary. As a successful example, the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium was formed to improve the statistical power of GWA studies and facilitate replication opportunities among multiple large population-based cohort studies[36]. The CHARGE consortium includes five prospective cohort studies from the United States and Europe: the Age, Gene/Environment Susceptibility—Reykjavik Study (AGES), the Atherosclerosis Risk in Communities Study (ARIC), the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), and the Rotterdam Study. In this thesis, we describe three studies in chapters 4.1, 4.2, and 4.3 that were performed within the framework of this consortium.

Since most of the variants with large effects are already identified, current studies are even performed with larger sample sizes. For instance, we have extended our GWA study on CRP levels to studies outside of the CHARGE consortium and collected data from more than 65,000 individuals from 15 studies.

Clinical implications

A first clinical implication of the finding of novel risk factors for type II diabetes and CHD is the use of these novel factors in prediction models. Among many inflammatory markers, CRP is the strongest and the most consistent novel risk factor that is associated with the risks of type II diabetes and CHD. There is evidence suggesting that adding CRP to risk prediction models among initially intermediate-risk individuals improves risk stratification[37-39]. Whether this added value would lead to clinically meaningful reclassification is doubted[40]. Methodological aspects of the correct method of assessment of additive value were discussed in the former section on "Risk prediction".

A second clinical implication of the search for novel risk factors is to find a target population for lifestyle and therapeutic interventions. In a recent trial, the JUPITER trial, rosuvastatin significantly reduced the incidence of major cardiovascular events in healthy individuals, who had no hyperlipidemia but had elevated CRP levels [41]. In general, targeting treatment only based on LDL-cholesterol has not provided optimal risk reduction for many individuals. The result of the JUPITER trial suggests that some of the benefits of rosuvastatin may relate to a reduction in inflammation. Nevertheless, the biological mechanism

does not need to be known when applying the result of trials like the JUPITER trial in clinical practice.

Third, GWA studies may increase our knowledge regarding the etiology and pathophysiology of the studied diseases. Like novel risk factors, GWA findings can contribute to the development of new therapies. In spite of the view that considers GWA findings far from clinical use, early clinical application of them is not unexpected. One example is the rapid development of a treatment strategy for age-related macular degeneration based on inhibitors of complement activation[42]. GWA findings may also help in risk prediction. Although the loci introduced up to now may not be enough to predict type II diabetes[43] or CHD events[44], they may together with the genes that will be discovered in the future - and possibly the information on gene-environment interaction - enable us to stratify people on their risk of different diseases.

Plans for future research

In the last two years, the work for this thesis focused on exploring the genetic structure of novel risk factors for type II diabetes and CHD. We have succeeded in identifying a number of loci for levels of CRP, uric acid, and fibrinogen through GWA studies. The main aim of future research should be the exploration of the unexplained variation of the traits. Moreover, we still have a long way to go from GWA findings to causal variants. Finally, all the efforts would be fruitless unless we find proper clinical applications for our findings.

In the coming years, GWA studies will probably remain as a suitable approach to identify novel loci for different traits. Novel findings, however, would be achievable only through enhanced statistical power. Mega meta-analysis or a large collection of individual studies will make it possible to achieve GWA information of tens of thousands or even hundreds of thousands of individuals. In line with this strategy, we have now extended our GWA studies for CRP to nearly 65,000 individuals and for uric acid to nearly 30,000 individuals. A large meta-analysis for fibrinogen is planned for the near future. Worthy to mention, the availability of a large sample size would also increase the chance of finding important gene-environment interactions.

The search for causal variant will not be an easy task. GWA studies have so far only marked wide loci in which one or more variants are associated with the trait. One approach to find the exact functional variant(s) would be targeted regional sequencing of the genome. Fortunately, sequencing technologies are becoming less expensive and more affordable. Moreover, the CHARGE consortium is planning to extend its collaborative work to sequencing efforts which will provide a unique substructure for such studies. Furthermore,

Chapter 5

functional work will help to extend our understanding of the mechanism by which the gene affects the trait.

We can further increase our etiologic knowledge by applying GWA findings to other traits. When conducted in a population independent from the discovery panel, testing GWA findings with other relevant traits is an effective strategy to extend the findings from risk factors to clinical events and vice versa. In chapter 4.2, we tested SNPs related to uric acid for their relation with gout. It would be of interest to examine the genetic variants that were found for serum CRP, uric acid, and fibrinogen levels with type II diabetes and CHD. Such studies would of course need very large sample sizes to obtain enough power.

Current studies have not succeeded in developing valid disease prediction models by the use of genetic variants identified by GWA studies. It is disputed whether an increase in the number of SNPs will eventually lead to an accurate disease prediction. Extra information such as pathway-oriented prediction models or use of the information that may become available on gene-environment interaction, however, may make the process of prediction more promising. We also need to use different evaluation measures - as previously discussed - to examine the performance of the prediction models with genetic variants.

References

1. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991;7:403-22.
2. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: objectives and design update. Eur J Epidemiol. 2007;22:819-829.
3. Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JC. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol. 2009.
4. Brunner EJ, Kivimaki M, Witte DR, Lawlor DA, Davey Smith G, Cooper JA, Miller M, Lowe GD, Rumley A, Casas JP, Shah T, Humphries SE, Hingorani AD, Marmot MG, Timpson NJ, Kumari M. Inflammation, insulin resistance, and diabetes--Mendelian randomization using CRP haplotypes points upstream. PLoS Med. 2008;5:e155.
5. Mene P, Punzo G. Uric acid: bystander or culprit in hypertension and progressive renal disease? J Hypertens. 2008;26:2085-92.
6. Kim SY, Guevara JP, Kim KM, Choi HK, Heitjan DF, Albert DA. Hyperuricemia and risk of stroke: a systematic review and meta-analysis. Arthritis Rheum. 2009;61:885-92.
7. Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N Engl J Med. 2004;350:1387-97.
8. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. Am J Hum Genet. 2008;82:1185-92.
9. Rothman KeJ. Modern Epidemiology. third edition ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
10. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95 Suppl 1:S144-50.
11. Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med. 1965;58:295-300.
12. Poole C. Causal values. Epidemiology. 2001;12:139-41.
13. Grant WB, Mohr SB. Ecological studies of ultraviolet B, vitamin D and cancer since 2000. Ann Epidemiol. 2009;19:446-54.
14. Johnson RJ, Feig DI, Herrera-Acosta J, Kang DH. Resurrection of uric acid as a causal risk factor in essential hypertension. Hypertension. 2005;45:18-20.
15. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol. 2004;33:30-42.
16. Bound J, Jaeger DA, Baker RM. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak. J Am Stat Assoc. 1995;90:443-450 CR.
17. Rockhill B, Weinberg CR, Newman B. Population attributable fraction estimation for established breast cancer risk factors: considering the issues of high prevalence and

unmodifiability. Am J Epidemiol. 1998;147:826-33.

18.	Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007;445:881-5.

19.	Janssens AC, Gwinn M, Valdez R, Narayan KM, Khoury MJ. Predictive genetic testing for type 2 diabetes. Bmj. 2006;333:509-10.

20.	Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. Am J Epidemiol. 1988;128:1185-97.

21.	Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health. 1998;88:15-9.

22.	Rockhill B, Weinberg C. Error in population attributable risk calculation. Epidemiology. 1996;7:453-4.

23.	Williamson DF. Population Attributable Risk of Incident Hypertension in Women. JAMA. 2009;302:2550-.

24.	Folsom AR, Chambless LE, Ballantyne CM, Coresh J, Heiss G, Wu KK, Boerwinkle E, Mosley TH, Jr., Sorlie P, Diao G, Sharrett AR. An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the atherosclerosis risk in communities study. Arch Intern Med. 2006;166:1368-73.

25.	Wang TJ, Gona P, Larson MG, Tofler GH, Levy D, Newton-Cheh C, Jacques PF, Rifai N, Selhub J, Robins SJ, Benjamin EJ, D'Agostino RB, Vasan RS. Multiple biomarkers for the prediction of first major cardiovascular events and death. N Engl J Med. 2006;355:2631-9.

26.	Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. Biostat. 2005;6:227-239.

27.	Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. Circulation. 2007;115:928-935.

28.	Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21:128-38.

29.	Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2008;167:362-8.

30.	Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27:157-72; discussion 207-12.

31.	Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106:9362-7.

32.	Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS,

Group L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA. Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet. 2009;41:56-65.

33. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet. 2008;83:311-21.

34. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet. 2001;29:306-9.

35. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006;38:209-13.

36. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, Uitterlinden AG, Harris TB, Witteman JCM, Boerwinkle E, on Behalf of the CHARGE Consortium. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts. Circ Cardiovasc Genet. 2009;2:73-80.

37. Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. N Engl J Med. 2002;347:1557-65.

38. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med. 2006;145:21-9.

39. Buckley DI, Fu R, Freeman M, Rogers K, Helfand M. C-reactive protein as a risk factor for coronary heart disease: a systematic review and meta-analyses for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151:483-95.

40. Wang TJ. New cardiovascular risk factors exist, but are they clinically useful? Eur Heart J. 2008;29:441-4.

41. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM, Jr., Kastelein JJ, Koenig W, Libby P, Lorenzatti AJ, MacFadyen JG, Nordestgaard BG, Shepherd J, Willerson JT, Glynn RJ. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med. 2008;359:2195-207.

42. Gehrs KM, Anderson DH, Johnson LV, Hageman GS. Age-related macular degeneration--emerging pathogenetic and therapeutic concepts. Ann Med. 2006;38:450-71.

43. Dehghan A, van Hoek M, Sijbrands EJ, Hofman A, Witteman JC. High serum uric acid as a novel risk factor for type 2 diabetes. Diabetes Care. 2008;31:361-2.

44. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, Roos C, Hirschhorn JN, Berglund G, Hedblad B, Groop L, Altshuler DM, Newton-Cheh C, Orho-Melander M. Polymorphisms associated with cholesterol and risk of cardiovascular events. N Engl J Med. 2008;358:1240-9.

Chapter 6

Summary
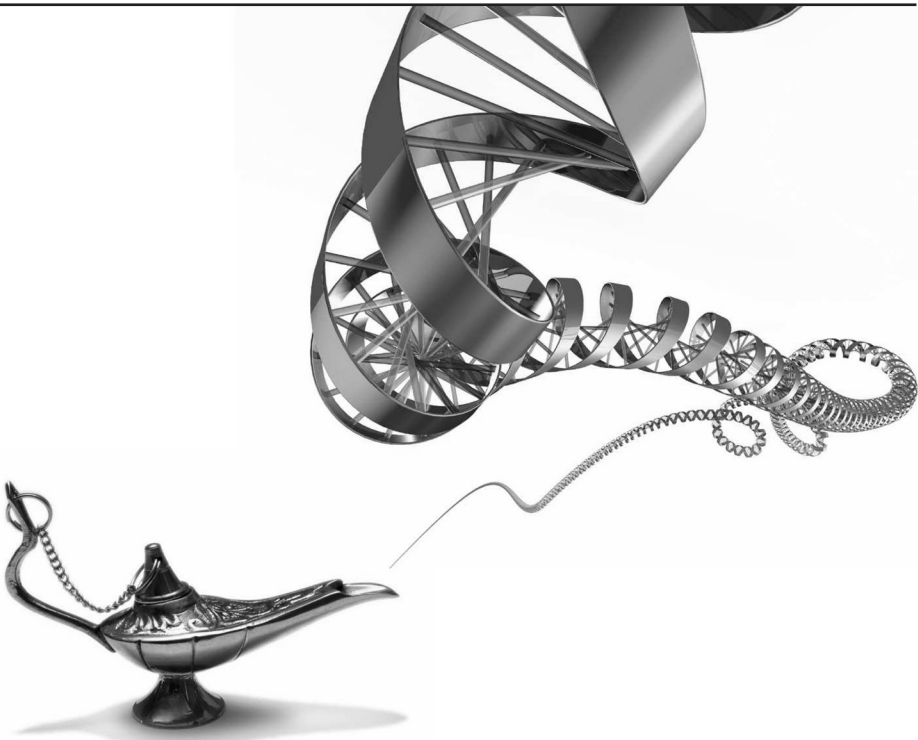
Summary

Since the 1950s, epidemiologic studies focused on chronic disorders including type II diabetes and coronary heart disease (CHD). Although several risk factors have been identified, there is a need to move beyond the known risk factors and to search for novel factors. C-reactive protein (CRP) and uric acid are two markers that have been studied as novel risk factors for type II diabetes and CHD in recent years.

The advent of hypothesis free approaches in genetic studies has made it possible to investigate the genetic structure of different traits even beyond our current biologic knowledge. These studies may extend our understanding of the determinants of disease and the pathways that are involved in the regulation of traits and open new avenues in etiologic and therapeutic research.

In this thesis, we studied the roles of serum CRP and serum uric acid as novel risk factors for type II diabetes and CHD. Furthermore, we investigated genetic variants that affect the levels of CRP, uric acid and fibrinogen.

Chapter 1 gives a brief introduction of CRP and uric acid as novel risk factors for type II diabetes and CHD. The chapter introduces population attributable risk (PAR) as a measure for the impact of risk factors on disease risk and it introduces the genome-wide association (GWA) study as a novel method to investigate genetic components of different traits.

Chapter 2 focuses on novel risk factors for type II diabetes. In chapter 2.1, we studied the association of serum CRP with type II diabetes. Our meta-analysis showed that serum CRP is associated with risk of type II diabetes independent of obesity. Moreover, we showed that a genetic variant in the human CRP locus was associated both with a high serum CRP, and an increased risk of type II diabetes. These findings provide support for CRP as an etiologic factor involved in the pathogenesis of type II diabetes. In chapter 2.2, we estimated the proportion of type II diabetes cases which is attributable to a high serum CRP level. The results showed that high serum CRP is one of the major contributors to the risk of type II diabetes. Furthermore, we found that the contribution of modifiable risk factors to the risk of type II diabetes is considerable. In chapter 2.3, we examined the association between serum uric acid and type II diabetes. The results showed that subjects with higher levels of serum uric acid are at higher risk of developing type II diabetes. We further found that one quarter of diabetes cases could be attributed to a high serum uric acid level.

In chapter 3 we examined novel risk factors for CHD. In chapter 3.1, we showed that a substantial proportion of the risks of myocardial infarction (MI) and CHD could be attributed to inflammation, independent of traditional risk factors. We also demonstrated that modifiable risk factors explain a substantial

proportion of the risks of MI and CHD. In chapter 3.2, we sought replication for the results of a GWA study on CHD and MI. We found, however, that two genetic variants in 9p21 that were reported to be associated with risk of MI and CHD do not replicate in an elderly population. The lack of association may be due to differences in the etiology of CHD in young and older people.

Chapter 4 describes several GWA studies. In chapter 4.1, we performed a meta-analysis of GWA scans on serum CRP in six studies comprising nearly 27,000 individuals. We confirmed six previously reported loci annotated by or close to genes including CRP, leptin receptor (LEPR), interleukin 6 receptor (IL6R), glucokinase regulatory protein (GCKR), hepatic transcription factor 1 (HNF1A), and apolipoprotein E (APOE). This study highlights immune response and metabolic regulatory pathways as important pathways in determining CRP levels and provides insights into the genetic architecture of a well-known inflammation marker. In chapter 4.2, we performed a GWA study in the Framingham Heart Study and the Rotterdam Study and replicated the results in the ARIC Study. The study confirmed the previously reported association of variation in SLC2A9 with uric acid and gout, and extended these findings to blacks. Moreover, two new loci, ABCG2 and SLC17A3, were identified for serum uric acid and gout. In chapter 4.3, we performed a GWA study for fibrinogen levels. The GWA data were provided by six cohort studies and comprised around 22,000 subjects. We confirmed the previously known association of the FGB locus with fibrinogen levels. Further, we identified three novel loci IRF1, PCCB, and NLRP3. Since these genes also play a role in inflammation, they may represent interesting targets for further research into biological pathways involved in cardiovascular disease and other chronic inflammatory conditions.

In chapter 5, we reviewed our findings in the context of a general discussion. Methodological considerations with regard to the studies in this thesis and similar studies are discussed. Also, potential clinical implications of our findings are addressed. Furthermore, future plans for research are discussed.

Samenvatting

Sinds de jaren 50 van de vorige eeuw richt epidemiologisch onderzoek zich op chronische aandoeningen als diabetes type II en coronaire hartziekten (CHD). Hoewel er meerdere risicofactoren gevonden zijn, is er de noodzaak om nieuwe factoren te identificeren. C-reactive proteïne (CRP) en urinezuur zijn twee markers die de afgelopen jaren het onderwerp van studie zijn geweest als nieuwe risicofactoren voor diabetes type II en CHD.

De opkomst van hypothese-vrije methoden in het genetisch onderzoek heeft het mogelijk gemaakt de genetische achtergrond van verschillende aandoeningen te onderzoeken zonder afhankelijk te zijn van bestaande kennis over biologie. Dit soort onderzoek kan ons inzicht in de determinanten van ziekte en de regulatie van normale fysiologische processen vergroten en zo nieuwe deuren openen voor etiologisch en therapeutisch onderzoek.

In dit proefschrift hebben wij de rol bestudeerd van CRP en urinezuur in het serum als nieuwe risicofactoren voor diabetes type II en CHD. Daarnaast hebben we genetische varianten onderzocht die de hoeveelheid CRP, urinezuur en fibrinogeen in het bloed beïnvloeden.

Hoofdstuk 1 geeft een korte introductie over CRP en urinezuur als nieuwe risicofactoren voor diabetes type II en CHD. In dit hoofdstuk wordt het populatie attributieve risico (PAR) geïntroduceerd, een maat voor het effect van risicofactoren op het ziekterisico. Ook wordt het genoom-wijde associatie (GWA) onderzoek geïntroduceerd als een nieuwe methode om de genetische componenten van ziektes en fysiologische processen te onderzoeken.

Hoofdstuk 2 richt zich op nieuwe risicofactoren voor diabetes type II. In hoofdstuk 2.1 beschrijven wij de associatie van serum CRP met diabetes type II. Onze meta-analyse laat zien dat serum CRP geassocieerd is met het risico op diabetes type II, onafhankelijk van obesitas. Daarnaast laten we zien dat een genetische variant in het humane CRP locus geassocieerd is met zowel een hoog serum CRP als met een verhoogd risico op diabetes type II. Deze bevindingen ondersteunen de rol van CRP als een etiologische factor in de pathogenese van diabetes type II. In hoofdstuk 2.2 maken we een schatting welk deel van van het voorkomen van diabetes type II toe te schrijven is aan een hoog serum CRP. Onze resultaten laten zien dat een hoog serum CRP een belangrijke bijdrage levert aan het risico op diabetes type II. Daarnaast hebben we gevonden dat de bijdrage van beïnvloedbare risicofactoren aan het risico op diabetes type II aanzienlijk is. In hoofdstuk 2.3 onderzoeken we de associatie tussen serum urinezuur en diabetes type II. De resultaten laten zijn dat personen met een hoger serum urinezuur ook een hoger risico hebben om diabetes type II te ontwikkelen. Verder hebben we gevonden dat in de populatie een kwart het voorkomen van met diabetes toe te schrijven is aan hoge serum urinezuur spiegels.

In hoofdstuk 3 onderzoeken we nieuwe risicofactoren voor CHD. In hoofdstuk 3.1 laten we zien dat een substantieel deel van het risico op myocardinfarct (MI) en CHD toe te schrijven is aan ontstekingsprocessen, onafhankelijk van de bekende risicofactoren. We laten ook zien dat beïnvloedbare risicofactoren een substantieel deel van het risico op MI en CHD verklaren. In hoofdstuk 3.2 hebben we geprobeerd de resultaten van een eerder uitgevoerd GWA onderzoek naar CHD en MI te repliceren. Wij hebben echter in onze oudere populatie twee genetische varianten op 9p21, waarvan eerder beschreven was dat ze een associatie zouden hebben met het risico op MI en CHD, niet kunnen repliceren. Het feit dat wij geen associatie gevonden hebben zou kunnen samenhangen met verschillen in de etiologie van CHD tussen jongere en oudere mensen.

In hoofdstuk 4 wordt een aantal GWA onderzoeken beschreven. In hoofdstuk 4.1 hebben we een meta-analyse gedaan van GWA onderzoeken van CRP spiegels van zes verschillende studies met een totaal van bijna 27.000 deelnemers. Wij konden zes eerder gevonden loci bevestigen, in of vlakbij de genen CRP, leptin receptor (LEPR), interleukin 6 receptor (IL6R), glucokinase regulatory protein (GCKR), hepatic transcription factor 1 (HNF1A) en apolipoproteïn E (APOE). Dit onderzoek laat zien dat de immuunrespons en metabole regulatoire pathways een belangrijke rol spelen in het bepalen van de hoeveelheid CRP in het bloed. Daarnaast geeft deze studie inzicht in de genetische achtergrond van deze bekende ontstekingsmarker. In hoofdstuk 4.2 hebben wij een GWA onderzoek gedaan in de Framingham Heart Study en de Rotterdam Study (ERGO onderzoek) en de resultaten gerepliceerd in het ARIC onderzoek. Dit onderzoek bevestigde de al eerder bekende associatie van variatie in SLC2A9 met urinezuur en jicht en liet zien dat deze bevindingen ook voor Afro-Amerikanen gelden. Daarnaast heeft dit onderzoek twee nieuwe loci, ABCG2 en SLC17A3, geïdentificeerd die het serum urinezuur niveau en risico op jicht beïnvloeden. In hoofdstuk 4.3 hebben wij een GWA onderzoek gedaan naar fibrinogeenspiegels. De GWA data kwamen van zes cohortstudies met een totaal van ongeveer 22.000 deelnemers. Wij hebben de al bekende associatie van het FGB locus met fibrinogeenspiegels bevestigd. Daarnaast hebben wij gevonden dat drie andere loci, IRF1, PCCB en NLRP3, ook geassocieerd zijn met fibrinogeenspiegels. Omdat deze genen ook een rol spelen in ontstekingsprocessen, zouden zij interessante aangrijpingspunten kunnen vormen voor verder onderzoek naar de biologische processen die betrokken zijn bij cardiovasculaire aandoeningen en andere ziektes die samengaan met chronische ontstekingsprocessen.

In hoofdstuk 5 bediscussiëren wij onze bevindingen en methodologische aspecten van de studies die in dit proefschrift beschreven worden en van soortgelijke studies. Ook bespreken wij mogelijke klinische implicaties van onze bevindingen en plannen voor toekomstig onderzoek.

Chapter 7

Acknowledgements
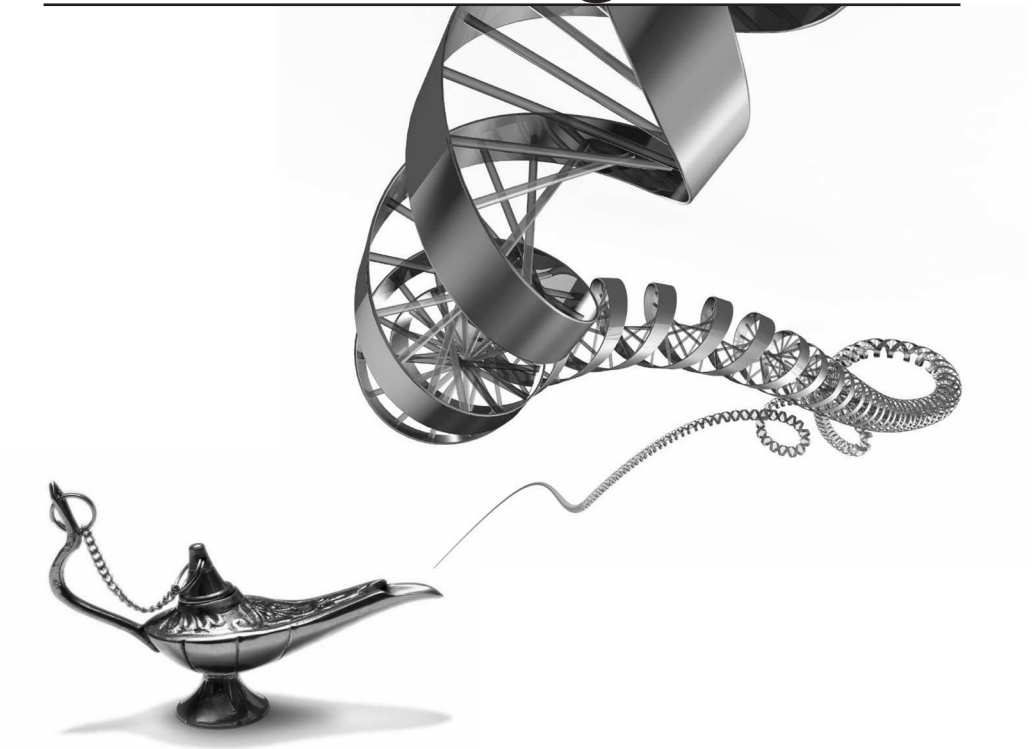
In the last couple of years I have been on my way to accomplish this thesis. This was an adventurous journey with its own challenges and obstacles. Completion of this thesis would have been impossible without the help of many people who assisted and encouraged me to follow this road. It is my sincere pleasure to take a moment to express my gratitude to those who directly or indirectly contributed to my PhD trajectory and helped me to grow as a researcher.

Above all, my deepest gratitude goes to my promotor, Professor Jacqueline Witteman. Dear Jacqueline, thanks for giving me the chance to pursue my PhD in your group. You taught me how to do research. Moreover, you taught me how important it is to take care of morality and ethics in research. Your guidance was with me throughout my PhD and your scientific input improved my thesis a lot. Your reviews of my manuscripts were always well detailed. You had insightful comments and suggestions. I have been inspired by your knowledge and talent of epidemiology. In these years I had the pleasure to discuss many methodological issues with you which taught me a lot. It was also an honor for me to be your assistant in your courses. This gave me the opportunity to learn teaching skills from you. Last but not least, your confidence in my abilities has always been encouraging and helped me to develop confidence in my career.

I am extremely thankful to my co-promotor, Dr. Eric Sijbrands. Dear Eric, without your assistance and advices, I would not have been able to write successful papers in the field of diabetes. Your clinical input in my papers was priceless. Your warm and friendly personality together with your hospitality encouraged me to drop by your office whenever I had something to discuss. You are an excellent group leader who inspires young researchers. Your research meetings were not only motivating for me, but also instructive. My memories with you do not restrict to work! I and other colleagues were honored - many times - to enjoy our time at your home and with your nice family. Your nice dinners and generous parties constitute some of my best memories in the Netherlands.

I am grateful to the members of the reading committee for their time and work: Professor Cornelia van Duijn, Professor Jacqueline Dekker and Professor Eric Boersma. Cornelia, I learned a lot about genetic epidemiology in your courses. Your comments and advices on my papers were always helpful.

My special thanks also go to the members of the plenary committee, Professor André Uitterlinden, Professor J Van Saase, Professor J. Laman, and Professor G. Pasterkamp for their presence in the plenary committee. Dear Andre, I owe you a special thanks for your efforts to open the field of GWAS to us.

I would like to extend my gratitude to Professor Albert Hofman, not only as the head of the department and the initiator of the Rotterdam Study but also as the head of the Netherlands Institute for Health Sciences (Nihes), the institute that

for providing me with the data of the Rotterdam Study. Frank, your constant smile makes you a pleasure to work with.

I would like to thank the secretaries of the department, Petra, Sandra, Marjolijn, Marti, Kabita, Hetty, Erica, and Marion for all their kindness and help.

My very deep thanks goes to my first tutor and colleague, Dr. Isabella Kardys. Isabella, you helped me so much - with a lot of patience - when I started to work in the department as an MSc student. You were a very pleasant colleague and a nice friend. You contributed to my thesis not only by critically reading my manuscripts, but also through daily discussions that we had in our office. I do not also forget that you were my first Dutch teacher! And you were always willing to help in any case with your divine phone calls!

My special thanks to my office mates, Janine and Maryam. Firstly, thanks to both of you for being my paranymphs! Janine, you are such a kind and energetic colleague who is always open to my questions and to spend as much time as necessary to share your thoughts with me. I have learned a lot from you in our daily discussions - especially the ones beside the coffee machine! Moreover, my experience with and knowledge of the Netherlands has improved a lot since you have joined our group. Maryam, it's a priceless asset to have a nice country mate such as you as a friend and colleague in the office. I very much appreciate your warm personality and your willingness to help. I have learned a lot from you not only on work related topics, but also on daily life.

My special thanks to Toos and Inge, two pleasant officemates whose incredible efforts with the measurements in the Rotterdam Study I appreciate.

I am thankful to Dr. Mandy van Hoek. Mandy, you were a nice colleague and a responsible and helpful co-author on my papers. Your warm and lively personality always gave me energy.

It is almost impossible to name all of the friends and colleagues in the department who helped me and supported me throughout my study period. My special gratitude extends to Arfan and Mark who were always willing to think about my questions and spend time to share their thoughts with me. Najaf, Karol, and Aaron, thanks for your help and advices especially on genetic studies. I thank Dr. Fakhredin Sayed Ali Tabatabaee for his help with my first meta-analysis. I am also thankful to my dear colleagues, Germaine, Charlotte, Caroline, Rachel, Suzette, Maarten and Quirijn. Maarten, thanks for teaching me Dutch - spreekwoord van de dag!

My gratitude goes to David and Rivka Alexander. David, you taught me a lot on how to write a readable and successful paper. Moreover, I have a lot of nice memories with you and Rivka.

I would also like to name Professor Mohsen Janghorbani, my first epidemiology teacher in medical school whose enthusiasm and wonderful teaching skills

inspired me so much that I started to love epidemiology and chose to become an epidemiologist.

Living alone in the Netherlands, I would have not survived these years if there had not been so many good friends around me. Dear Hadi, your help started early when we were both in Iran. Common ambitions, common aims; we chose the same path and came to the Netherlands together in 2004. You were a nice classmate and a great support in the first year of my stay in the Netherlands. Mohammad, Maryam, Ghazal, and Erfan, who are now far away and I miss them a lot. You were such a friendly and warm family that with you, I always felt at home. Siamak, I will never forget your help in the first months after I arrived in the Netherlands. Mojgan and Nahid, thanks for your help, and especially for your useful tips for my first years in the Netherlands. Jolanda, Erik, Folco and Nathan, I have a lot of nice memories with you. Jolanda, my special thanks to you for proof reading my thesis and your helpful comments. I would also like to name Azadeh, Behnaz, Dana, Mehlika, Millad, Mohammad (Riaz), Payam, Shuangjie and Sima for their friendship and their generous help.

Accomplishment of this thesis would have been impossible without the love and support of my family. My deepest appreciation goes to my parents, whom I owe all my success in life to. My special thanks to my late father, Abdolkarim Dehghan, for his belief in me and his lifetime of encouraging comments. Pedarjan (father), your valuable guidance and your helping hand made a big difference in who I am today. My sincere thanks to my mother for her unconditional love, support and understanding. My special thanks goes to my love, Raha. My dearest Raha, thanks for your love and accompaniment. Life is much nicer with you. My deep thanks to my brother, Koorosh, my sister in law, Marzieh, and their adorable kids: Taraneh and little Arad. Taraneh, your nice drawings were always pasted on the wall behind my computer to remind me of my family; those who no matter how far they are, are always in my heart. My sisters Shirin, Shiva, Shahrzad and my brother in law, Amin. Words cannot express how much I love you all. Shirin, all these years you were only one phone call away from me to help with everything. Shiva, thanks for your support and for helping me with the cover design. Shahrzad, thanks for your love and support. Last but not least, I would like to thank my parents in law for their generous love and support.

At the end, I would like to thank all those who in one way or another helped me get where I am now but for any reason I have missed to thank them by name in this text.

List of publications

Dehghan A*, Kottgen A*, Yang Q*, Hwang SJ, Kao WL, Rivadeneira F, Boerwinkle E, Levy D, Hofman A, Astor BC, Benjamin EJ, van Duijn CM, Witteman JC, Coresh J, Fox CS. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet. 2008;372:1953-61.

Dehghan A, Kardys I, de Maat MP, Uitterlinden AG, Sijbrands EJ, Bootsma AH, Stijnen T, Hofman A, Schram MT, Witteman JC. Genetic variation, C-reactive protein levels, and incidence of diabetes. Diabetes. 2007;56:872-8.

Dehghan A, van Hoek M, Sijbrands EJ, Stijnen T, Hofman A, Witteman JC. Risk of type 2 diabetes attributable to C-reactive protein and other risk factors. Diabetes Care. 2007;30:2695-9.

Dehghan A, van Hoek M, Sijbrands EJ, Hofman A, Witteman JC. High serum uric acid as a novel risk factor for type 2 diabetes. Diabetes Care. 2008;31:361-2.

Dehghan A*, Yang Q*, Peters A*, Basu S, Bis JC, Rudnicka AR, Kavousi M, Chen MH, Baumert J, Lowe GD, McKnight B, Tang W, de Maat M, Larson MG, Eyhermendy S, McArdle WL, Lumley T, Pankow JS, Hofman A, Massaro JM, Rivadeneira F, Kolz M, Taylor KD, van Duijn CM, Kathiresan S, Illig T, Aulchenko YS, Volcik KA, Johnson AD, Uitterlinden AG, Tofler GH, Gieger C, Psaty BM, Couper DJ, Boerwinkle E, Koenig W, O'Donnell CJ, Witteman JC, Strachan DP, Smith NL, Folsom AR. Association of novel genetic Loci with circulating fibrinogen levels: a genome-wide association study in 6 population-based cohorts. Circ Cardiovasc Genet. 2009;2:125-33.

Dehghan A, van Hoek M, Sijbrands EJ, Oostra BA, Hofman A, van Duijn CM, Witteman JC. Lack of association of two common polymorphisms on 9p21 with risk of coronary heart disease and myocardial infarction; results from a prospective cohort study. BMC Med. 2008;6:30.

Kottgen A*, Glazer NL*, **Dehghan A***, Hwang SJ, Katz R, Li M, Yang Q, Gudnason V, Launer LJ, Harris TB, Smith AV, Arking DE, Astor BC, Boerwinkle E, Ehret GB, Ruczinski I, Scharpf RB, Ida Chen YD, de Boer IH, Haritunians T, Lumley T, Sarnak M, Siscovick D, Benjamin EJ, Levy D, Upadhyay A, Aulchenko YS, Hofman A, Rivadeneira F, Uitterlinden AG, van Duijn CM, Chasman DI, Pare G, Ridker PM, Kao WH, Witteman JC, Coresh J, Shlipak MG, Fox CS. Multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet. 2009.

Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, Watzinger N, Larson MG, Smith NL, **Dehghan A**, Grosshennig A, Schillert A, Teumer A, Schmidt R, Kathiresan S, Lumley T, Aulchenko YS, Konig IR, Zeller T, Homuth G, Struchalin M, Aragam J, Bis JC, Rivadeneira F, Erdmann J, Schnabel RB, Dorr M, Zweiker R, Lind L, Rodeheffer RJ, Greiser KH, Levy D, Haritunians T, Deckers JW, Stritzke J, Lackner KJ, Volker U, Ingelsson E, Kullo I, Haerting J, O'Donnell CJ, Heckbert SR, Stricker BH, Ziegler A, Reffelmann T, Redfield MM, Werdan K, Mitchell GF, Rice K, Arnett DK, Hofman A, Gottdiener JS, Uitterlinden AG, Meitinger T, Blettner M, Friedrich N, Wang TJ, Psaty BM, van Duijn CM, Wichmann HE, Munzel TF, Kroemer HK, Benjamin EJ, Rotter JI, Witteman JC, Schunkert H, Schmidt H, Volzke H, Blankenberg S. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. Jama. 2009;302:168-78.

Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, **Dehghan A**, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Kottgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FU, Smith AV, Taylor K, Scharpf RB, Hwang SJ, Sijbrands EJ, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh

J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JC, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, van Duijn CM. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009.

Benjamin EJ, Rice KM, Arking DE, Pfeufer A, van Noord C, Smith AV, Schnabel RB, Bis JC, Boerwinkle E, Sinner MF, **Dehghan A**, Lubitz SA, D'Agostino RB, Sr., Lumley T, Ehret GB, Heeringa J, Aspelund T, Newton-Cheh C, Larson MG, Marciante KD, Soliman EZ, Rivadeneira F, Wang TJ, Eiriksdottir G, Levy D, Psaty BM, Li M, Chamberlain AM, Hofman A, Vasan RS, Harris TB, Rotter JI, Kao WH, Agarwal SK, Stricker BH, Wang K, Launer LJ, Smith NL, Chakravarti A, Uitterlinden AG, Wolf PA, Sotoodehnia N, Kottgen A, van Duijn CM, Meitinger T, Mueller M, Perz S, Steinbeck G, Wichmann HE, Lunetta KL, Heckbert SR, Gudnason V, Alonso A, Kaab S, Ellinor PT, Witteman JC. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. Nat Genet. 2009;41:879-81.

Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, **Dehghan A**, Kuhnel B, Aspelund T, Yang Q, Tanaka T, Jaffe A, Bis JC, Verwoert GC, Teumer A, Fox CS, Guralnik JM, Ehret GB, Rice K, Felix JF, Rendon A, Eiriksdottir G, Levy D, Patel KV, Boerwinkle E, Rotter JI, Hofman A, Sambrook JG, Hernandez DG, Zheng G, Bandinelli S, Singleton AB, Coresh J, Lumley T, Uitterlinden AG, Vangils JM, Launer LJ, Cupples LA, Oostra BA, Zwaginga JJ, Ouwehand WH, Thein SL, Meisinger C, Deloukas P, Nauck M, Spector TD, Gieger C, Gudnason V, van Duijn CM, Psaty BM, Ferrucci L, Chakravarti A, Greinacher A, O'Donnell CJ, Witteman JC, Furth S, Cushman M, Harris TB, Lin JP. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet. 2009;41:1191-8.

Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, Loos RJ, Manning AK, Jackson AU, Aulchenko Y, Potter SC, Erdos MR, Sanna S, Hottenga JJ, Wheeler E, Kaakinen M, Lyssenko V, Chen WM, Ahmadi K, Beckmann JS, Bergman RN, Bochud M, Bonnycastle LL, Buchanan TA, Cao A, Cervino A, Coin L, Collins FS, Crisponi L, de Geus EJ, **Dehghan A**, Deloukas P, Doney AS, Elliott P, Freimer N, Gateva V, Herder C, Hofman A, Hughes TE, Hunt S, Illig T, Inouye M, Isomaa B, Johnson T, Kong A, Krestyaninova M, Kuusisto J, Laakso M, Lim N, Lindblad U, Lindgren CM, McCann OT, Mohlke KL, Morris AD, Naitza S, Orru M, Palmer CN, Pouta A, Randall J, Rathmann W, Saramies J, Scheet P, Scott LJ, Scuteri A, Sharp S, Sijbrands E, Smit JH, Song K, Steinthorsdottir V, Stringham HM, Tuomi T, Tuomilehto J, Uitterlinden AG, Voight BF, Waterworth D, Wichmann HE, Willemsen G, Witteman JC, Yuan X, Zhao JH, Zeggini E, Schlessinger D, Sandhu M, Boomsma DI, Uda M, Spector TD, Penninx BW, Altshuler D, Vollenweider P, Jarvelin MR, Lakatta E, Waeber G, Fox CS, Peltonen L, Groop LC, Mooser V, Cupples LA, Thorsteinsdottir U, Boehnke M, Barroso I, Van Duijn C, Dupuis J, Watanabe RM, Stefansson K, McCarthy MI, Wareham NJ, Meigs JB, Abecasis GR. Variants in MTNR1B influence fasting glucose levels. Nat Genet. 2009;41:77-81.

van Hoek M, **Dehghan A**, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC. Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes. 2008;57:3122-8.

van Hoek M, **Dehghan A**, Zillikens MC, Hofman A, Witteman JC, Sijbrands EJ. An RBP4 promoter polymorphism increases risk of type 2 diabetes. Diabetologia. 2008;51:1423-8.

Johnson AD, Kavousi M, Smith AV, Chen MH, **Dehghan A**, Aspelund T, Lin JP, van Duijn CM, Harris TB, Cupples LA, Uitterlinden AG, Launer L, Hofman A, Rivadeneira F, Stricker B, Yang Q, O'Donnell CJ, Gudnason V, Witteman JC. Genome-wide association meta-analysis for total serum bilirubin levels. Hum Mol Genet. 2009;18:2700-10.

Riaz M, Elstrodt F, Hollestelle A, **Dehghan A**, Klijn JG, Schutte M. Low-risk susceptibility alleles in 40

human breast cancer cell lines. BMC Cancer. 2009;9:236.

Schnabel RB, Baumert J, Barbalic M, Dupuis J, Ellinor PT, Durda P, **Dehghan A**, Bis JC, Illig T, Morrison AC, Jenny NS, Keaney JF, Jr., Gieger C, Tilley C, Yamamoto JF, Khuseyinova N, Heiss G, Doyle M, Blankenberg S, Herder C, Walston JD, Zhu Y, Vasan RS, Klopp N, Boerwinkle E, Larson MG, Psaty BM, Peters A, Ballantyne CM, Witteman JC, Hoogeveen RC, Benjamin EJ, Koenig W, Tracy RP. Duffy antigen receptor for chemokines (Darc) polymorphism regulates circulating concentrations of monocyte chemoattractant protein-1 and other inflammatory mediators. Blood. 2009.

Maja Barbalic*,Dupuis J*, **Dehghan A***, Bis JC*, Hoogeveen RC*, Schnabel RB*, Nambi V, Bretler M, Smith NL, Peters A, Lu C, Tracy RP, Aleksic N, HeerigaJ, Keaney JF, Rice K, Lip GYH, Vasan RS, Glazer NL, Larson MG, Uitterlinden AG, Yamamoto J, Durda P, Haritunians T, Psaty BM, Boerwinkle E, Hofman A, Koenig W, Jenny NS, Witteman JCM, Ballantyne C, Benjamin EJ. Large-scale genomic studies reveal central role of ABO in sP-selectin and sICAM-1 levels. (Accepted in Human Molecular Genetics)

Kottgen A, Hwang SJ, Larson MG, Eyk JE, Fu Q, Benjamin EJ, **Dehghan A**, Glazer NL, Kao WH, Harris TB, Gudnason V, Shlipak MG, Yang Q, Coresh J, Levy D, Fox CS. Uromodulin Levels Associate with a Common UMOD Variant and Risk for Incident CKD. J Am Soc Nephrol. 2009.

Zillikens MC, van Meurs JB, Sijbrands EJ, Rivadeneira F, **Dehghan A**, van Leeuwen JP, Hofman A, van Duijn CM, Witteman JC, Uitterlinden AG, Pols HA. SIRT1 genetic variation and mortality in type 2 diabetes: interaction with smoking and dietary niacin. Free Radic Biol Med. 2009;46:836-41.

Sulo G, Burazeri G, **Dehghan A**, Kark JD. Partner's smoking status and acute coronary syndrome: population-based case-control study in Tirana, Albania. Croat Med J. 2008;49:751-6.

Smith N*, Chen M*, **Dehghan A***, Strachan D*, Basu S*, Bis J, de Maat M, Rumley A, Kong X, Yang Q, Wiggins K, Van Duijn C, McArdle W, Pankow J, Johnson A, McKnight B, Uitterlinden A, Aleksic N, Meigs J, Peters A, Koenig W, Cushman M, Kathiresan S, Rotter J, Bovill E, Hofman A, Boerwinkle E, Tofler G, Psaty B, Leebeek F, Folsom A, Larson M, Lumley T, Witteman J, Tang W, O'Donnell CJ. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE Consortium. (accepted in Circulation)

Dehghan A, van Hoek M, Sijbrands EJ, Hofman A, Witteman JC. Risk of coronary heart diseases related to high serum C-reactive protein and traditional risk factors. Eur J Cardiovasc Prev Rehabil. (submited)

Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Baumert J, Schnabel RB, Nambi V, Kavousi M, Smith NL, Smith AV, Larson MG, Hoogeveen RCM, Tracy R, Launer LJ, Yamamoto JF, Folsom A, Sijbrands EG, Pankow J, Keaney JF, Sun W, Fontes JD, Lu C, Astor B, Hofman A, Aspelund T, Siscovick D, Guo X, Rotter J, Thorand B, Gieger C, Coresh J, Uitterlinden A, Harris TB, Rice K, Boerwinkle E, Vasan RS, Van Duijn CM, Psaty BM, Gudnason V, Koenig W, Ballantyne C, Benjamin EJ, Witteman JCM. Meta-analysis of genome-wide association study in 26,967 subjects confirms six loci for C-reactive protein serum levels. 2010

PhD Portfolio

Summary of PhD training and teaching

Name: Abbas Dehghan
Erasmus MC Department: Epidemiology
Research School: Netherlands Institute for Health Sciences
PhD period: 2006 - 2010
Promotor: prof.dr. J.C.M. Witteman

	Year	Workload	
		Hours	ECTs
1. PhD training			
Research skills			
MSc in Epidemiology, Nihes, The Netherlands	2004 - 2005		30
DSc in Epidemiology, Nihes, The Netherlands	2005 - 2006		30
In depth courses			
How to Protect and utilize Your Scientific Discovery, Erasmus Postgraduate School Molecular Medicine, The Netherlands	2009	8	
Finding your way in biological information, Erasmus Postgraduate School Molecular Medicine, The Netherlands	2009	16	
Molecular Biology in Cardiovascular Research, COER, The Netherlands	2009	16	
Writing Successful Grant Proposals, Erasmus Postgraduate School Molecular Medicine, The Netherlands	2008	8	
The Role of DNA Polymorphism in Complex Traits and Diseases Erasmus Postgraduate School Molecular Medicine, The Netherlands	2008	32	
Basic Data Analysis on Gene Expression Arrays I , Erasmus Postgraduate School Molecular Medicine, The Netherlands	2008	8	
Repeated Measurements in Clinical Studies, Nihes, The Netherlands	2007	20	
Missing Values in Clinical Research, Nihes, The Netherlands	2007	20	
Bayesian Statistics, Nihes, The Netherlands	2007		0.9
Planning and Evaluation of Screening, Nihes, The Netherlands	2007		1.4
Diabetes Module, MRC, Cambridge, UK	2006	16	
The Role of DNA Polymorphisms in Complex Traits and Diseases, Erasmus Postgraduate School Molecular Medicine, The Netherlands	2006	32	
Biomedical English Writing and Communication, Erasmus MC, The Netherlands	2006	40	
Principles of Epidemiologic Data Analysis, Nihes, The Netherlands	2006	40	
Presentations at (Inter)national Conferences			
4th CHARGE Meeting in Washington DC, USA	2009	30	
3rd CHARGE Meeting in Rotterdam, The Netherlands	2009	30	

49th AHA Conference on Cardiovascular Disease Epidemiology and Prevention, Florida, USA	2009	30
42nd Annual Meeting of the European Diabetes Epidemiology Group of EASD, Robinson College, Cambridge, UK	2007	30
IEA EEF European Congress of Epidemiology, Utrecht, the Netherlands	2006	30
9th WHO/IDF Cambridge Seminar on the Epidemiology and Public health Aspects of Diabetes Mellitus, Cambridge, UK	2006	30
3rd National Epidemiology Congress, Kerman, Iran	2006	30

Long-term working visits

Framingham Heart Study, Boston, USA	2009	80

2. Teaching

Lecturing

Lecturer, Study Design, Nihes, The Netherlands	2009	60
Lecturer, Methodological Topics in Epidemiologic Research, Nihes, The Netherlands	2009	60
Teaching assistant, Study Design, Nihes, The Netherlands	2005-2008	20
Teaching assistant, Methodological Topics in Epidemiologic Research Nihes, The Netherlands	2005-2008	20
Teaching assistant, Principles of Epidemiology, Erasmus Summer Program, The Netherlands	2007	10

Supervising Master students

Raha Pazoki: HDL cholesterol and genetic variation in Estrogen Induced Gene121;: The Rotterdam study	2008		6
Gerhard Sulo: Partner's smoking status and acute coronary syndrome in a country in transition: a population-based case-control study in Tirana, Albania	2008		6
David K Amankwa: Population attributable risk or risk factors for coronary heart disease	2007		6
Millad Soluki: Unrecognised myocardial infarction and the risk of mortality: The Rotterdam Study	2007		6