Elevated Plasma YKL-40, Lipids and Lipoproteins, and Ischemic Vascular Disease in the General Population

Alisa D. Kjaergaard, MD, PhD; Julia S. Johansen, MD, DMSc; Stig E. Bojesen, MD, PhD, DMSc; Børge G. Nordestgaard, MD, DMSc

Background and Purpose—We tested the hypothesis that observationally and genetically elevated YKL-40 is associated with elevated lipids and lipoproteins and with increased risk of ischemic vascular disease.

Methods—We conducted cohort and Mendelian randomization studies in 96,110 individuals from the Danish general population, with measured plasma levels of YKL-40 (n=21,647), plasma lipids and lipoproteins (n=94,461), and CHI3L1 rs4950928 genotype (n=94,579).

Results—From 1977 to 2013, 3,256 individuals developed ischemic stroke, 5,629 ischemic cerebrovascular disease, 4,183 myocardial infarction, and 10,271 developed ischemic heart disease. The 91% to 100% versus 0% to 33% YKL-40 percentile category was associated with a 34% increase in triglycerides, but only with minor changes in other lipids and lipoproteins. For these categories, the multifactorially adjusted hazard ratio was 1.99 (95% confidence interval, 1.49–2.67) for ischemic stroke, 1.85 (1.44–2.37) for ischemic cerebrovascular disease, 1.28 (0.95–1.73) for myocardial infarction, and 1.23 (1.01–1.51) for ischemic heart disease. When compared with rs4950928 CC homozygosity, the presence of G-allele was associated with a doubling (GC) or tripling (GG) in YKL-40 levels, but not with triglyceride levels or with risk of ischemic vascular disease. A doubling in YKL-40 was associated with a multifactorially adjusted observational hazard ratio for ischemic stroke of 1.18 (1.11–1.27), and a genetic odds ratio of 1.04 (0.95–1.15). Corresponding risk estimates were 1.15 (1.09–1.22) observationally and 1.06 (0.99–1.14) genetically for ischemic cerebrovascular disease, 1.08 (1.00–1.15) observationally and 1.04 (0.96–1.13) genetically for myocardial infarction, and 1.07 (1.02–1.12) observationally and 1.01 (0.96–1.07) genetically for ischemic heart disease.

Conclusions—Elevated YKL-40 was associated with a 34% increase in triglyceride levels and with a 2-fold increased risk of ischemic stroke, whereas genetically elevated YKL-40 were not. (Stroke. 2015;46:329-335. DOI: 10.1161/STROKEAHA.114.007657.)

Key Words: brain ischemia ▪ molecular epidemiology ▪ myocardial ischemia ▪ triglycerides ▪ YKL-40 protein, human

YKL-40, a 40-kDa plasma protein with 3 N-terminal amino acids being Y (tyrosine), K (lysine), and L (leucine), is produced by lipid-laden macrophages inside the vessel wall1 and is elevated in patients with ischemic cardiovascular disease.2 We have previously shown that baseline-elevated plasma YKL-40 is associated with later development of ischemic stroke, but not with myocardial infarction.3 This discrepancy is possibly explained by differences in pathogenesis of atherosclerosis in cerebral and cardiac vasculature and might also be because of differences in levels of lipids and lipoproteins. It is also possible that elevated YKL-40 is a direct cause of ischemic stroke, but not of myocardial infarction.

These uncertainties can possibly be clarified in a cohort study of the general population and in a Mendelian randomization study. The latter design uses genetic variants associated with lifelong elevated YKL-40 levels that are largely unconfounded and not prone to reverse causation.4,5 Thus, the method mimics a controlled double-blind randomized trial, but it uses genetic variants rather than a drug and placebo, and like a blinded trial it allows inference about causality. YKL-40 levels are partly heritable and rs4950928, a single nucleotide polymorphism in the proximal promoter of CHI3L1, accounts for an estimated 9% of its variance.6 Mechanistically, MYC/MAK transcriptional factors bind well to the major allele of rs4950928,7 which increases CHI3L1 transcription,8 and consequently elevates YKL-40 levels.9–11

We tested the hypothesis that observationally and genetically elevated YKL-40 is associated with elevated levels of...
lipids and lipoproteins and with increased risk of ischemic vascular disease. For this purpose, we performed cohort and Mendelian randomization studies in 96,110 individuals from the Danish general population. First, in the cohort study, we examined the association of elevated YKL-40 levels with levels of lipids and lipoproteins and with risk of ischemic vascular disease. Second, in the Mendelian randomization study, we examined whether CHI3L1 rs4950928 genotype was associated with YKL-40 levels, with levels of lipids and lipoproteins, and with risk of ischemic vascular disease. Third, in cohort and Mendelian randomization studies combined, we compared the association of observationally and genetically elevated YKL-40 levels with levels of lipids and lipoproteins and with risk of ischemic vascular disease. Finally, because elevated YKL-40 levels were strongly associated with elevated triglyceride levels, in a post hoc analysis, we compared the association of observationally and genetically elevated triglyceride levels with YKL-40 levels.

Methods

Participants
We used 2 independent prospective studies with individuals selected to reflect the adult Danish population aged ≥20 years, the Copenhagen City Heart Study (CCHS) and the Copenhagen General Population Study (CGPS; Methods in the online-only Data Supplement). We included following covariates in the multifactorially adjusted model: age, sex, ever smoking (yes/no), cumulative smoking (pack-years), alcohol consumption (number of drinks per week, 1 drink = 12 g alcohol), body mass index (kg/m²), diabetes mellitus (yes/no), hypertension (yes/no), atrial fibrillation (yes/no for ischemic stroke and ischemic cerebrovascular disease analyses only), lipid-lowering treatment (yes/no), total cholesterol (mmol/L), high-density lipoprotein cholesterol (mmol/L), low-density lipoprotein cholesterol (mmol/L), triglycerides (mmol/L), C-reactive protein (mg/L), and study population (CCHS or CGPS).

Genotyping
We genotyped for CHI3L1 rs4950928 in 94,579 participants with DNA available (as we performed reruns, >99.8% were genotyped). Genotyping was performed using TaqMan (Applied Biosystems by Life Technologies Corporation, Carlsbad, CA) assays. We used a forward (AGTCCCATATAAGGGCTGTTTT) and a reverse (CAGAGGCCCTGTACTTCCTTTATAT) primer for the polymerase

<table>
<thead>
<tr>
<th>Table. Characteristics of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma YKL-40 Percentile Categories*</td>
</tr>
<tr>
<td>No. of participants, n</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>0%–33%</td>
</tr>
<tr>
<td>Women, %</td>
</tr>
<tr>
<td>Age, y</td>
</tr>
<tr>
<td>Alcohol consumption, weekly drinks</td>
</tr>
<tr>
<td>Eversmoker, %</td>
</tr>
<tr>
<td>Cumulative smoking, pack-years</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
</tr>
<tr>
<td>Diabetes mellitus, %</td>
</tr>
<tr>
<td>Hypertension, %</td>
</tr>
<tr>
<td>Atrial fibrillation, %</td>
</tr>
<tr>
<td>Lipid-lowering therapy, %</td>
</tr>
<tr>
<td>C-reactive protein, mg/L</td>
</tr>
</tbody>
</table>
| Values are baseline values collected at the 1991–1994 examination of the Copenhagen City Heart Study, and at the 2003 to 2013 examination of the Copenhagen General Population Study and expressed as numbers of participants, frequencies, or medians (interquartile ranges). Number of participants varies slightly because of availability of data. ICVD indicates ischemic cerebrovascular disease; IHBD, ischemic heart disease; IS, ischemic stroke; and MI, myocardial infarction. P values are from Cuzick nonparametric test for trend and represent comparisons *across YKL-40 percentile categories, †between participants with events vs controls (Table I in the online-only Data Supplement), and ‡across genotype coded as 0, 1, and 2 for number of G-alleles (Table II in the online-only Data Supplement).
chain reaction amplification and a common (CTCCCCACGCGGC) and a variant (ACTCCCCAGCGCGGC) probe to determine genotype. We also genotyped for 4 other single nucleotide polymorphisms in CHI3L1 (rs1538372, rs2071579, rs880633, and rs946259) using TaqMan assays (Figure II in the online-only Data Supplement).

Genetic variants in lipoprotein lipase, the most important enzyme in the metabolism of plasma triglycerides, are probably the best genetic instruments reflecting lifelong elevated triglyceride levels. In the CCHS, participants were genotyped for 4 genetic variants (rs328, rs1801177, rs268, and rs118204057) in the lipoprotein lipase gene as described. In the CGPS, we used TaqMan assays. We genotyped 600,048 individuals for all 4 variants.

Statistical Analysis
We used STATA version 13. We performed χ² test and Cuzick nonparametric test for trend. Hardy–Weinberg equilibrium was investigated using a χ² test. For power calculations, NCSS PASS was used. With the assumption of 2-sided P<0.05 and the observed number of events/nonevents, we calculated hazard ratios that could be detected in the genetic studies at 90% power (Methods in the online-only Data Supplement).

Results
The combined studies included 96,110 individuals from the Danish general population, of which 21,646 had plasma YKL-40 measured, 94,461 had lipids and lipoproteins measured, and 94,579 were genotyped for rs4950928; we only show combined results because the results for the 2 studies separately were similar. From 1977 to 2013, with information on all participants, 32,56 individuals developed ischemic stroke, 56,29 ischemic cerebrovascular disease, 41,83 myocardial infarction, and 10,271 developed ischemic heart disease (Table I in the online-only Data Supplement). CHI3L1 rs4950928 genotype was in Hardy–Weinberg equilibrium (Table II in the online-only Data Supplement; P=0.23) and had a minor allele frequency of 21%. Baseline characteristics as a function of YKL-40 percentile categories, ischemic vascular disease, and CHI3L1 rs4950928 genotype are shown in the Table and Tables I and II in the online-only Data Supplement. Collectively, these data illustrate that YKL-40 levels and ischemic vascular disease were strongly associated with many covariates and potential confounders, whereas genotype was associated with none, and therefore can be used as an unconfounded instrument to study the importance of lifelong elevated YKL-40 levels not prone to reverse causation.

Plasma YKL-40, Lipids and Lipoproteins, and Ischemic Vascular Disease: Cohort Study
Increased percentile of plasma YKL-40 was associated with levels of lipids and lipoproteins, that is, markedly increased levels of triglycerides, slightly increased levels of total cholesterol, and modestly decreased levels of low-density lipoprotein and high-density lipoprotein cholesterol (Figure 1).
For triglycerides; there was a 34% increase for 91% to 100% versus 0% to 33% YKL-40 percentile category.

Risk of ischemic stroke and ischemic cerebrovascular disease increased with increased plasma YKL-40 percentile category (Figure 2; Figure III in the online-only Data Supplement). The multifactorially adjusted hazard ratio for ischemic stroke was 1.99 (95% confidence interval, 1.49–2.67) for ischemic stroke, 1.85 (1.44–2.37) for ischemic cerebrovascular disease, 1.28 (0.95–1.73) for myocardial infarction, and 1.23 (1.01–1.51) for ischemic heart disease.

CHI3L1 rs4950928 Genotype, Plasma YKL-40, Lipids and Lipoproteins, and Ischemic Vascular Disease: Mendelian Randomization Study

There was a doubling (99% increase) in geometric mean of plasma YKL-40 levels from CC (4.4% of individuals) to GC (32.9% of individuals) genotype and a tripling (196% increase) in levels from CC to GG (62.7% of individuals) genotype (Figure IV in the online-only Data Supplement; \( P \)-trend<1×10−300); genotype explained 14% of the variation in plasma YKL-40 levels and the \( F \) statistic was 1690, indicating that rs4950928 is a strong instrument to examine the effect of lifelong elevated plasma YKL-40 levels. Indeed, CHI3L1 single nucleotide polymorphisms rs1538372, rs2071579, rs880633, and rs946259 were only associated with plasma YKL-40 levels through their modest to strong linkage disequilibrium with rs4950928 (Figure II in the online-only Data Supplement).

Geometric mean for plasma YKL-40 was 24 ng/mL for individuals with the CC genotype, 48 ng/mL for the GC genotype, and 71 ng/mL for the GG genotype. CHI3L1 rs4950928 genotype was not associated with levels of lipids or lipoproteins (Figure V in the online-only Data Supplement), or with risk of any ischemic vascular disease (Figure 3; Figure VI in the online-only Data Supplement); the higher risk of myocardial infarction in heterozygous GC individuals versus homozygous CC individuals is likely because of chance because this was not observed for homozygous GG versus homozygous CC or for the end point ischemic heart disease with more total events. We had 90% power to detect hazard ratios in GG and GC versus CC genotypes of 1.29 and 1.31 for ischemic stroke, of 1.21 and 1.23 for ischemic cerebrovascular disease, of 1.25 and 1.26 for myocardial infarction, and of 1.16 and 1.16 for ischemic heart disease.

Observationally Versus Genetically Elevated Plasma YKL-40 Levels, Lipids and Lipoproteins, and Ischemic Vascular Disease: Cohort and Mendelian Randomization Studies

A doubling in plasma YKL-40 levels was associated with a multifactorially adjusted increase in triglycerides of 8.8% (Figure VII in the online-only Data Supplement). In contrast, a doubling in YKL-40 was not associated with total cholesterol and was associated modestly with low-density lipoprotein and high-density lipoprotein cholesterol levels (ie, multifactorially adjusted decreases of 1.9% and 0.9%, respectively). A doubling in plasma YKL-40 because of genotype did not change plasma levels of any of the lipids and lipoproteins.

A doubling in YKL-40 was associated with a multifactorially adjusted observational hazard ratio for ischemic stroke of 1.18 (1.11–1.27) and a corresponding genetic odds ratio of 1.04 (0.95–1.15; Figure 4; Figure VIII in the online-only Data Supplement). Corresponding risk estimates were 1.15 (1.09–1.22) observationally and 1.06 (0.99–1.14) genetically for ischemic cerebrovascular disease, 1.08 (1.00–1.15) observationally and 1.04 (0.96–1.13) genetically for myocardial infarction, and 1.07 (1.02–1.12) observationally and 1.01 (0.96–1.07) genetically for ischemic heart disease.

Observationally Versus Genetically Elevated Plasma Triglyceride Levels and Plasma YKL-40: Cohort and Mendelian Randomization Studies

Because elevated YKL-40 levels were strongly associated with elevated triglyceride levels, in a post hoc analysis, we
compared the association of observationally and genetically elevated triglyceride levels with YKL-40 levels.

All 4 genetic variants (rs328, rs1801177, rs268, and rs118204057) in the lipoprotein lipase gene were in Hardy–Weinberg equilibrium. There was a 9% increase in geometric mean of plasma triglyceride levels from lipoprotein lipase allele score 0 (18.0% of individuals) to 1 (75.1% of individuals) and a 24% increase in triglyceride levels from allele score 0 to 2 (6.9% of individuals; Figure IX in the online-only Data Supplement; \(P\)-trend ≤1×10−500); allele score explained 0.7% of the variation in plasma triglyceride levels and the \(F\) statistic was 242, indicating that lipoprotein lipase allele score is a valid instrument to examine the effect of lifelong elevated plasma triglyceride levels.

A doubling in triglyceride levels was associated with a multifactorially adjusted observational increase for YKL-40 of 13% (11%–16%), whereas the corresponding genetic doubling in triglycerides was not associated with elevated YKL-40 levels (Figure 5).

**Discussion**

Studying 96,110 individuals from the general population, we found that elevated YKL-40 was associated with a 34% increase in triglyceride levels and with a 2-fold increased risk of ischemic stroke, whereas genetically elevated YKL-40 was not. Taken together, these are novel findings.

Mechanistically, these findings may be explained by another common factor that causes both elevated plasma YKL-40 levels and ischemic stroke. First, elevated plasma YKL-40 levels are strongly associated with elevated triglyceride levels. Second, elevated triglycerides are strongly associated with increased risk of ischemic stroke.\(^{23,24}\) Third, statin treatment lowers both plasma YKL-40\(^{25}\) and triglyceride levels. Thus, it is tempting to speculate that elevated triglycerides might be a common cause of both elevated plasma YKL-40 levels and ischemic stroke. The present study does, however, not support this hypothesis because genetically elevated triglycerides were not associated with elevated plasma YKL-40 levels.

We have previously shown that elevated YKL-40 is associated with increased risk of ischemic stroke, but not with myocardial infarction in the Copenhagen City Heart Study,\(^{26}\) which has been confirmed in an independent prospective study,\(^{26}\) as well as the Women’s Health Study.\(^{27}\) In our present compared with former study, we had a longer follow-up time with more cases in the Copenhagen City Heart Study, and we also included an even larger prospective study of the Danish general population, the Copenhagen General Population Study. In addition, we explored whether observationally and genetically elevated YKL-40 is associated with levels of lipids and lipoproteins and whether genetically elevated YKL-40 is associated with ischemic vascular disease. Finally, because elevated YKL-40 levels were strongly associated with elevated triglyceride levels, in a post hoc analysis, we compared the association of observationally and genetically elevated triglyceride levels with YKL-40 levels.

Because baseline elevated plasma YKL-40 is associated with later development of ischemic stroke, but not with myocardial infarction,\(^{2,27,28}\) it is possible that YKL-40 plays a role in the formation of embolisms rather than in development of atherosclerosis per se. Another explanation could come from the fact that YKL-40 expression in ischemic stroke was more astrocytic than macrophage based.\(^{29}\) Therefore, YKL-40 could be a sensitive marker of small cerebral infarctions and covert ischemic cerebrovascular disease, which is associated with increased risk of later development of ischemic stroke.

Our findings that elevated YKL-40 was not associated with total cholesterol levels, and only modestly with high-density lipoprotein and low-density lipoprotein cholesterol levels, differs slightly from previous studies that showed varied results.\(^{2,25,27,30,31}\) However, all studies, including the present study, showed that elevated YKL-40 is strongly associated with elevated triglyceride levels.

Limitations of the present study include survival bias and selective nonresponse to the invitation by individuals being sick because of ischemic vascular disease. Both would, however, lead to under-representation of these individuals in our sample of the general population and likely would bias the results toward the null hypothesis, and therefore are unlikely to explain our findings for ischemic stroke and ischemic cerebrovascular disease. Furthermore, because myocardial infarction and ischemic heart disease are more frequent diseases than ischemic stroke and ischemic cerebrovascular disease, but with similar mortality, we would not expect the true risk of myocardial infarction and ischemic heart disease by plasma YKL-40 levels to be considerably higher than the modest point estimates in the present study. Another limitation is that ischemic stroke events were not classified according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification. Consequently, we were not able to evaluate subtype predisposition and infer on mechanism.

**Figure 5.** Change in YKL-40 for a doubling in triglycerides, expressed as % increase. The genetic model is based on a 2-stage least squares regression with a linear regression at second stage. \(P\) values are from \(t\) tests in linear regression analyses, testing whether change in YKL-40 levels for a doubling in plasma triglycerides are significantly different from 0. CI indicates confidence interval.
Also, we measured nonfasting plasma triglycerides from blood samples drawn at random times after a meal. However, previous studies have shown that levels of triglycerides do not change very much after the intake of a normal meal. Furthermore, most people are in the nonfasting state most of the day, so nonfasting triglycerides likely represent average daily triglyceride levels better than fasting triglyceride levels. Our negative results on rs4950928 are likely accurate because the distribution of genotypes does not deviate from Hardy–Weinberg equilibrium, and because the instrument used explained as much as 14% of plasma YKL-40 levels and had an extremely high F value of 1690. Furthermore, we confirm the results from the Women’s Genome Health Study that rs4950928 was associated with plasma YKL-40 levels, but not with risk of ischemic vascular disease. Finally, as we only studied white individuals of Danish descent, our results are not necessarily directly applicable to other ethnic groups; however, we are not aware of data to suggest that the present results should not be applicable to most ethnic groups.

In conclusion, in this large prospective study of subjects from the general population, we found that elevated YKL-40 was associated with a 34% increase in triglyceride levels and with a 2-fold increased risk of ischemic stroke, whereas genetically elevated YKL-40 was not.

Acknowledgments

We thank Birgit Hertz, Nina Dahl Kjersgaard, Hanne Damm, Anja Jochumsen, Ulla Kjørrulf-Hansen, Annette Lundgren-Beck, and Debbie Nadelmann, all Herlev Hospital, for excellent technical assistance. The participants of the Copenhagen City Heart Study and the Copenhagen General Population Study are thanked for their willingness to participate.

Sources of Funding

The work was supported by University of Copenhagen and by grants from Chief Physician Johan Boserup and Lise Boserups Foundation, Vera and Carl Johan Michaelens’s Foundation, Toyota-Fonden, Denmark, The Research Council at Herlev Hospital, The Danish Heart Foundation and The Danish Medical Research Council. Quidel provided the study with YKL-40 ELISA kits.

Disclosures

None.

References


Elevated Plasma YKL-40, Lipids and Lipoproteins, and Ischemic Vascular Disease in the General Population
Alisa D. Kjaergaard, Julia S. Johansen, Stig E. Bojesen and Børge G. Nordestgaard

Stroke. 2015;46:329-335
doi: 10.1161/STROKEAHA.114.007657
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/46/2/329

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2015/01/23/46.2.329.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/
SUPPLEMENTAL MATERIAL

Elevated plasma YKL-40, lipids and lipoproteins, and ischemic vascular disease

Alisa D. Kjaergaard, M.D., Ph.D.,1,2,3, Julia S. Johansen, M.D., D.M.Sc.2,4, Stig E. Bojesen, M.D., Ph.D., D.M.Sc.1,2,3,5, Børge G. Nordestgaard, M.D., D.M.Sc.1,2,3,5

1 Department of Clinical Biochemistry, Herlev University Hospital, Denmark, 2 Faculty of Health and Medical Sciences, University of Copenhagen, Denmark, 3 The Copenhagen General Population Study, Herlev University Hospital, Denmark, 4 Department of Medicine and Oncology, Herlev University Hospital, Denmark, 5 The Copenhagen City Heart Study, Frederiksberg University Hospital, Denmark

Table of contents:

Methods
Table I. Characteristics of participants according to ischemic vascular disease.
Table II. Characteristics of participants according to CHI3L1 rs4950928 genotype.
Figure I. Plasma YKL-40 percentile categories corrected for age and sex.
Figure II. Plasma YKL-40 as geometric mean and linkage disequilibrium in 5 single nucleotide polymorphisms in CHI3L1.
Figure III. Hazard ratios of ischemic vascular disease according to age corrected plasma YKL-40 percentile category.
Figure IV. Geometric mean of plasma YKL-40 and change in percent by CHI3L1 rs4950928 genotype.
Figure V. Plasma levels of lipids and lipoproteins according to CHI3L1 rs4950928 genotype.
Figure VI. Hazard ratios (HR) of ischemic vascular disease according to CHI3L1 rs4950928 genotype.
Figure VII. Change in lipids and lipoproteins for a doubling in YKL-40, expressed as % increase.
Figure VIII. Observational HR and genetic OR of ischemic vascular disease for a doubling in plasma YKL-40 levels.
Figure IX. Geometric mean of triglycerides and YKL-40 by lipoprotein lipase allele score.
METHODS

Participants
All participants were whites of Danish descent, and there was no overlap between the two studies. Prior to examination in both studies, participants filled out a self-administered questionnaire which was completed together with an investigator on the day of attendance, where a physical examination was performed and blood was drawn for immediate biochemical analyses, as well as for storage at -80°C for later DNA and biochemical analyses.

The CCHS Study
We used the 1991-1994 examination of the CCHS. Of 9,601 participants, 8,773 were genotyped for rs4950928. Plasma YKL-40 levels were measured in samples frozen for 12-15 years in 8,444 of these participants.

The CGPS Study
The CGPS was initiated in 2003 with ongoing enrolment. Of the first 86,509 participants, 85,806 were genotyped for rs4950928. Plasma YKL-40 levels were measured in fresh samples in 3,211 participants and in samples frozen for 7 years in 9,991 participants.

Endpoints
Diagnoses of ischemic stroke, ischemic cerebrovascular disease, myocardial infarction and ischemic heart disease were based on World Health Organization, International Classification of Diseases 8th revision (ICD-8) codes 432-434, 432-435, 410, and 410-414 until end of 1993, and thereafter based on 10th revision (ICD-10) codes I63, I63-I64 and G45, I21-I22, and I20-I25, respectively. Ischemic stroke was focal neurological symptoms lasting more than 24 hours. Ischemic cerebrovascular disease was determined on the basis of sudden onset of focal neurological symptoms (ischemic stroke, transient ischemic attack or amaurosis fugax). Hemorrhagic stroke and subarachnoidal hemorrhage were excluded from the ischemic stroke and ischemic cerebrovascular disease groups. The diagnosis of myocardial infarction required the presence of characteristic chest pain, elevated cardiac enzymes and/or electrocardiographic changes indicative of myocardial infarction. Ischemic heart disease was defined as myocardial infarction and/or characteristic symptoms of angina pectoris.

Plasma YKL-40 levels
Plasma YKL-40 levels were determined in duplicates by a two-site, sandwich-type enzyme-linked immunosorbent assay (ELISA) (Quidel Corporation, San Diego, CA, USA), using streptavidin-coated microplate wells, a biotinylated-Fab monoclonal capture antibody, and an alkaline phosphatase-labeled polyclonal detection antibody. The recovery of the ELISA was 102% and the detection limit was 10 µg/L. The intra-assay coefficients of variations were 5% at 40 µg/L and 4% at 104 µg/L and 155 µg/L. The inter-assay coefficient of variation was <6%.

Covariates
Participants reported on present and past smoking habits, weekly alcohol consumption, and on lipid-lowering therapy in the questionnaire. We subdivided participants into never and eversmokers, including both previous and current smokers. For eversmokers, a pack-year was defined as 20 g tobacco smoked per day for 1 year. Body mass index was calculated as measured weight in kilograms divided by measured height in meters squared. Diabetes mellitus was defined as self-reported disease, current use of insulin or other medication for treatment of diabetes, and/or a nonfasting plasma glucose >11 mmol/L. Hypertension was defined as use of antihypertensive medication, a systolic blood pressure ≥ 140 mmHg, and/or a diastolic blood pressure ≥ 90 mmHg. Levels of nonfasting total cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, and CRP were measured with use of standard hospital assays. Low-density lipoprotein (LDL)
cholesterol was calculated using the Friedewald equation when plasma triglycerides were ≤4.0 mmol/l and measured directly using a standard hospital assay when plasma triglycerides were >4.0 mmol/l. Missing values for covariates (<1%) used for adjustments were imputed using multivariable normal regression (mi impute mvn), based on age, sex and study population (CCHS or CGPS).

**Statistical analysis**

Risk estimates and corresponding 95% confidence intervals for ischemic vascular disease were computed as hazard ratios using Cox regression analysis. We assessed the assumption of proportional hazards graphically by plotting –ln(survival probability) against ln(analysis time) as well as by using Schoenfeld residuals, and did not detect any major violations of the proportional hazard assumption. In the Cox regression model, we used age as time scale, and participants were included at the day of blood sampling, i.e. follow-up time was left-truncated. Participants with events prior to blood sampling were excluded from analyses of plasma YKL-40 levels. We followed individuals until 10th of April 2013, event, or death, whichever came first. 375 participants emigrated and were censored at the date of emigration. Hazard ratios were corrected for regression dilution bias. Plasma YKL-40 levels from 884 individuals attending both the 1991-1994 baseline examination and the 2001-2003 follow-up examination of the CCHS were used for this correction: since YKL-40 increases with increasing age, we used the individual residuals of the linear regression of base 2 logarithm of plasma YKL-40 as a function of age at time of blood sampling to calculate a regression dilution ratio of 0.76. Recurrent events were not considered.

In the Mendelian randomization study for analyses of CHI3L1 rs4950928 and lipoprotein lipase genotype, all participants were included, also those with events prior to blood sampling for genotype analyses, as all genetic variants were in Hardy-Weinberg equilibrium and preceded events. Otherwise, analyses were like those described above for the cohort study. Genotypes were coded 0, 1 and 2 according to the number of YKL-40 or triglyceride elevating alleles. Lipoprotein lipase allele score was calculated as the sum of all 4 genotypes: 0 coded for no more than one triglyceride elevating allele, 1 coded for two elevating alleles, and 2 coded for three or more triglyceride elevating alleles.

For cohort and Mendelian randomization studies combined, observational hazard ratios and genetic odds ratios for a doubling in YKL-40 levels were assessed by calculating these ratios for an increase of 1U of base 2 logarithm of YKL-40. In individuals without a YKL-40 measurement and for genetically elevated levels, the well-known association between genotype and YKL-40 levels in other participants were applied adjusting for age, sex and study population. In the instrumental variable analyses we used a 2-stage least squares regression, where rs4950928 genotype examines the effect of a genetically doubling in plasma YKL-40 on the levels of lipids and lipoproteins. The first stage of the 2-stage least squares estimator is a linear regression of rs4950928 genotype on YKL-40, which generates predicted values for YKL-40. The second stage is a linear regression of outcome on predicted YKL-40. We used the same procedure to compare the observational and genetic increase in YKL-40 for a doubling in triglyceride levels. To estimate the effect of a genetically doubling in YKL-40 on ischemic vascular disease, we used a control function estimator with a logistic regression at second stage. The control function estimator follows the same principle as the 2-stage estimator, but additionally includes the estimated residuals from the first stage regression. These residuals may be correlated to unmeasured confounders, in which case they will help to control the effect of unmeasured confounders on outcome. The strengths of rs4950928 and lipoprotein lipase genotypes as instruments were assessed from the first-stage regression, where an F statistic >10 usually indicates sufficient statistical strength. An R² value as a percentage was used as a measure of the contribution of rs4950928 and lipoprotein lipase genotypes to the variation in plasma YKL-40 and triglyceride levels, respectively.
REFERENCES

### Table I. Characteristics of participants according to ischemic vascular disease.

<table>
<thead>
<tr>
<th></th>
<th>Ischemic stroke</th>
<th>Ischemic cerebrovascular disease</th>
<th>Myocardial infarction</th>
<th>Ischemic heart disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Controls</td>
<td>P-value</td>
<td>Events</td>
</tr>
<tr>
<td>Number of participants, N</td>
<td>3,256</td>
<td>92,854</td>
<td>NA</td>
<td>5,629</td>
</tr>
<tr>
<td>Women, %</td>
<td>56</td>
<td>46</td>
<td>1x10⁻²⁷</td>
<td>48</td>
</tr>
<tr>
<td>Age, years</td>
<td>70 (62-76)</td>
<td>69 (61-76)</td>
<td>&lt;1x10⁻¹⁰⁰</td>
<td>57 (47-66)</td>
</tr>
<tr>
<td>Alcohol consumption, weekly drinks</td>
<td>8 (3-16)</td>
<td>9 (3-16)</td>
<td>0.61</td>
<td>8 (3-16)</td>
</tr>
<tr>
<td>Eversmoker, %</td>
<td>74</td>
<td>60</td>
<td>1x10⁻⁵²</td>
<td>71</td>
</tr>
<tr>
<td>Cumulative smoking, packyears</td>
<td>19 (0-37)</td>
<td>15 (0-34)</td>
<td>4x10⁻¹³⁶</td>
<td>4 (0-21)</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>26.5 (23.8-29.4)</td>
<td>26.3 (23.7-29.1)</td>
<td>2x10⁻²³</td>
<td>26.5 (23.2-28.5)</td>
</tr>
<tr>
<td>Diabetes, %</td>
<td>15</td>
<td>4.3</td>
<td>3x10⁻¹⁸¹</td>
<td>13</td>
</tr>
<tr>
<td>Hypertension, %</td>
<td>49</td>
<td>13</td>
<td>&lt;1x10⁻¹⁰⁰</td>
<td>45</td>
</tr>
<tr>
<td>Atrial fibrillation, %</td>
<td>26</td>
<td>5.8</td>
<td>&lt;1x10⁻¹⁰⁰</td>
<td>22</td>
</tr>
<tr>
<td>Lipid-lowering therapy, %</td>
<td>30</td>
<td>10</td>
<td>3x10⁻²⁷¹</td>
<td>30</td>
</tr>
<tr>
<td>C-reactive protein, mg/L</td>
<td>1.9 (1.3-3.6)</td>
<td>1.4 (1.0-2.4)</td>
<td>2x10⁻⁹⁶</td>
<td>1.8 (1.2-3.3)</td>
</tr>
</tbody>
</table>

Baseline values expressed as numbers of participants, frequencies, or medians (interquartile ranges).
Disease is prevalent disease at the end of follow-up. Some of the participants had more than one disease. NA=not applicable.
P-values are from Cuzick's nonparametric test for trend between participants with events versus controls.
Table II. Characteristics of participants according to *CHI3L1* rs4950928 genotype.

<table>
<thead>
<tr>
<th></th>
<th><em>CHI3L1</em> rs4950928 Genotype</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>GC</td>
<td>GG</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>Number of participants, N</td>
<td>4,180</td>
<td>31,115</td>
<td>59,284</td>
<td>0.23*</td>
<td></td>
</tr>
<tr>
<td>Women, %</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Age, years</td>
<td>58 (47-67)</td>
<td>58 (47-67)</td>
<td>58 (47-67)</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption, weekly drinks</td>
<td>8 (3-15)</td>
<td>8 (3-15)</td>
<td>8 (3-15)</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Eversmoker, %</td>
<td>60</td>
<td>61</td>
<td>61</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>Cumulative smoking, packyears</td>
<td>4 (0-22)</td>
<td>4 (0-22)</td>
<td>4 (0-22)</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>25.6 (23.1-28.4)</td>
<td>25.6 (23.2-28.5)</td>
<td>25.6 (23.2-28.5)</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Diabetes, %</td>
<td>4.8</td>
<td>4.7</td>
<td>4.5</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Hypertension, %</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation, %</td>
<td>6.4</td>
<td>6.4</td>
<td>6.4</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Lipid-lowering therapy, %</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>C-reactive protein, mg/L</td>
<td>1.4 (1.0-2.4)</td>
<td>1.5 (1.0-2.4)</td>
<td>1.5 (1.0-2.4)</td>
<td>0.39</td>
<td></td>
</tr>
</tbody>
</table>

Baseline values expressed as numbers of participants, frequencies, or medians (interquartile ranges). Number of participants varies slightly due to the availability of data. P-values are from Cuzick's nonparametric test for trend across genotype coded as 0, 1, and 2 for number of G-alleles, except *P-value for Hardy-Weinberg equilibrium calculated by χ²-test.*
Figure I.
Plasma YKL-40 percentile categories corrected for age and sex.
Range of plasma YKL-40 levels across mean age in 20 equally sized age and sex groups.
X-axis and Y-axis are truncated at either end to improve visualization of the data.
**Figure II.**

Plasma YKL-40 as geometric mean and linkage disequilibrium in 5 single nucleotide polymorphisms in *CHI3L1*. All genotypes were in Hardy-Weinberg equilibrium.
<table>
<thead>
<tr>
<th>YKL-40 category</th>
<th>Number of events / all</th>
<th>Age and sex adjusted</th>
<th>Multifactorially adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ischemic stroke</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91-100%</td>
<td>127 / 2,103</td>
<td>2.33 (1.77-3.07)</td>
<td>1.99 (1.49-2.67)</td>
</tr>
<tr>
<td>67-90%</td>
<td>240 / 5,092</td>
<td>1.50 (1.20-1.88)</td>
<td>1.40 (1.12-1.77)</td>
</tr>
<tr>
<td>34-66%</td>
<td>303 / 7,044</td>
<td>1.16 (0.94-1.44)</td>
<td>1.15 (0.93-1.42)</td>
</tr>
<tr>
<td>0-33%</td>
<td>303 / 7,095</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trend: p=1x10^-5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Ischemic cerebrovascular disease** |                        |                      |                           |
| 91-100%           | 171 / 2,066            | 2.15 (1.70-2.72)     | 1.85 (1.44-2.37)          |
| 67-90%            | 315 / 5,037            | 1.34 (1.10-1.62)     | 1.23 (1.01-1.50)          |
| 34-66%            | 425 / 6,968            | 1.14 (0.95-1.36)     | 1.11 (0.93-1.32)          |
| 0-33%             | 427 / 7,003            |                      |                           |
|                   | Trend: p=2x10^-6       |                      |                            |

| **Myocardial infarction** |                        |                      |                           |
| 91-100%           | 113 / 2,104            | 1.63 (1.23-2.17)     | 1.28 (0.95-1.73)          |
| 67-90%            | 241 / 5,007            | 1.30 (1.05-1.62)     | 1.12 (0.89-1.40)          |
| 34-66%            | 333 / 6,934            | 1.16 (0.95-1.42)     | 1.06 (0.86-1.29)          |
| 0-33%             | 331 / 7,023            |                      |                           |
|                   | Trend: p=3x10^-5       |                      |                            |

| **Ischemic heart disease** |                        |                      |                           |
| 91-100%           | 255 / 2,025            | 1.65 (1.37-1.99)     | 1.23 (1.01-1.51)          |
| 67-90%            | 553 / 4,836            | 1.35 (1.17-1.56)     | 1.14 (0.98-1.32)          |
| 34-66%            | 752 / 6,743            | 1.16 (1.01-1.32)     | 1.07 (0.93-1.22)          |
| 0-33%             | 751 / 6,840            |                      |                           |
|                   | Trend: p=8x10^-9       |                      |                            |

Hazard ratio (95%CI)

**Figure III.**
Figure III.

Hazard ratios of ischemic stroke, ischemic cerebrovascular disease, myocardial infarction, and ischemic heart disease according to age corrected plasma YKL-40 percentile category.

Individuals with event prior to blood sampling were excluded from analyses of the total of 21,646 participants with measured plasma YKL-40 levels; therefore, the numbers of participants vary slightly between the different disease endpoints. P-values are for Cuzick’s nonparametric test for trend.

Correction for multiple comparisons (4 endpoints) only affected the p-value for ischemic heart disease, which was no longer statistically significant in the multifactorially adjusted model (required p-value i> 0.05/4=0.01).
Figure IV.

Geometric mean of plasma YKL-40 and change in percent by \textit{CHI3L1} rs4950928 genotype. P-values are for Cuzick's nonparametric test for trend.
**CHI3L1 rs4950928 genotype**

Figure V. Plasma levels of lipids and lipoproteins according to CHI3L1 rs4950928 genotype.

Trend test was by Cuzick's nonparametric test for trend across the groups.
rs4950928  ΔYKL-40 events/total  Age and sex adjusted  Multifactorially adjusted

**Ischemic stroke**
- GG: 1,978/259,284 (+196%)
  - Hazard ratio: 1.13 (0.90-1.42)
  - Multifactorial adjusted: 1.13 (0.89-1.42)
- GC: 995/31,115 (+99%)
  - Hazard ratio: 1.10 (0.87-1.38)
  - Multifactorial adjusted: 1.10 (0.86-1.40)
- CC: 137/4,180 (0%)
  - Hazard ratio: 1.75
  - Multifactorial adjusted: 1.75

**Ischemic cerebrovascular disease**
- GG: 3,460/59,284 (+196%)
  - Hazard ratio: 1.07 (0.89-1.28)
  - Multifactorial adjusted: 1.07 (0.89-1.28)
- GC: 1,740/31,115 (+99%)
  - Hazard ratio: 1.00 (0.83-1.20)
  - Multifactorial adjusted: 1.00 (0.83-1.20)
- CC: 235/4,180 (0%)
  - Hazard ratio: 1.75
  - Multifactorial adjusted: 1.75

**Myocardial infarction**
- GG: 2,556/59,284 (+196%)
  - Hazard ratio: 1.27 (0.99-1.62)
  - Multifactorial adjusted: 1.28 (1.00-1.65)
- GC: 1,346/31,115 (+99%)
  - Hazard ratio: 1.34 (1.05-1.73)
  - Multifactorial adjusted: 1.38 (1.07-1.79)
- CC: 157/4,180 (0%)
  - Hazard ratio: 1.75
  - Multifactorial adjusted: 1.75

**Ischemic heart disease**
- GG: 6,246/59,284 (+196%)
  - Hazard ratio: 1.12 (0.98-1.30)
  - Multifactorial adjusted: 1.12 (0.96-1.31)
- GC: 3,021/31,115 (+99%)
  - Hazard ratio: 1.14 (0.98-1.33)
  - Multifactorial adjusted: 1.14 (0.98-1.34)
- CC: 411/4,180 (0%)
  - Hazard ratio: 1.75
  - Multifactorial adjusted: 1.75

Figure VI.
Hazard ratios (HR) of ischemic stroke, ischemic cerebrovascular disease, myocardial infarction, and ischemic heart disease according to CHI3L1 rs4950928 genotype.

The rare homozygote (CC) is the reference group. Individuals with events prior to blood sampling were not excluded from analyses of the total of 94,579 genotyped participants, because genotype precedes events. ∆YKL-40 is the change as in Figure IV.
Total cholesterol, mmol/L

- Observational
  - Age and sex adjusted
  - Multifactorially adjusted
- Genetic
  - rs4950928

% increase (95% CI) P-value
-0.1 (-0.3 to 0.2) 0.73
-0.1 (-0.4 to 0.2) 0.66
0.6 (-0.1 to 1.3) 0.10

LDL cholesterol, mmol/L

- Observational
  - Age and sex adjusted
  - Multifactorially adjusted
- Genetic
  - rs4950928

% increase (95% CI) P-value
-2.4 (-2.8 to -1.9) 5x10^{-24}
-1.9 (-2.3 to -1.4) 4x10^{-17}
1.0 (-0.2 to 2.2) 0.09

HDL cholesterol, mmol/L

- Observational
  - Age and sex adjusted
  - Multifactorially adjusted
- Genetic
  - rs4950928

% increase (95% CI) P-value
-0.6 (-1.1 to -0.2) 0.003
-0.9 (-1.3 to -0.5) 1x10^{-5}
0.1 (-1.0 to 1.2) 0.85

Triglycerides, mmol/L

- Observational
  - Age and sex adjusted
  - Multifactorially adjusted
- Genetic
  - rs4950928

% increase (95% CI) P-value
10 (9.6 to 11) 1x10^{-143}
8.9 (8.0 to 9.7) 1x10^{-108}
-0.2 (-2.1 to 1.8) 0.86

Percent increase in lipids and lipoproteins for a doubling in plasma YKL-40 (95% CI)

Figure VII.

Change in lipids and lipoproteins for a doubling in YKL-40, expressed as % increase.
The genetic model is based on a 2-stage least squares regression with a linear regression at second stage.
P-values are from t-tests in linear regression analyses, testing whether change in lipids and lipoproteins for a doubling in plasma YKL-40 are significantly different from 0.
Ischemic stroke
Observational
Age and sex adjusted 1.22 (1.15-1.30) 4x10^{-10}
Multifactorially adjusted 1.18 (1.11-1.27) 1x10^{-6}
Genetic rs4950928 1.04 (0.95-1.15) 0.37

Ischemic cerebrovascular disease
Observational
Age and sex adjusted 1.19 (1.13-1.26) 2x10^{-10}
Multifactorially adjusted 1.15 (1.09-1.22) 2x10^{-6}
Genetic rs4950928 1.06 (0.99-1.14) 0.11

Myocardial infarction
Observational
Age and sex adjusted 1.12 (1.05-1.19) 4x10^{-4}
Multifactorially adjusted 1.08 (1.00-1.15) 0.04
Genetic rs4950928 1.04 (0.96-1.13) 0.34

Ischemic heart disease
Observational
Age and sex adjusted 1.13 (1.08-1.18) 1x10^{-6}
Multifactorially adjusted 1.07 (1.02-1.12) 0.004
Genetic rs4950928 1.01 (0.96-1.07) 0.72

Observational hazard ratio or genetic odds ratio (95% CI) for a doubling in plasma YKL-40 levels

Figure VIII.
Figure VIII.

Observational hazard ratios (HR) and genetic odds ratios (OR) of ischemic stroke, ischemic cerebrovascular disease, myocardial infarction, and ischemic heart disease for a doubling in plasma YKL-40 levels.

The genetic model is based on a control function estimator with a logistic regression at second stage. P-values are from z-scores in Cox regression analyses, testing whether hazard ratios significantly differ from 1.

Correction for multiple comparisons (4 endpoints) only affected the p-value for myocardial infarction, which was no longer statistically significant in the multifactorially adjusted model (required p-value is 0.05/4=0.01).
Figure IX.
Geometric mean of triglycerides and YKL-40 by lipoprotein lipase allele score.
Trend test was by Cuzick’s nonparametric test for trend across the groups.