

Risk Alleles of *USF1* Gene Predict Cardiovascular Disease of Women in Two Prospective Studies

Kati Komulainen¹, Mervi Alanne¹✉, Kirsi Auro¹✉, Riika Kilpikari¹, Päivi Pajukanta², Janna Saarela¹, Pekka Ellonen¹, Kaisa Salminen³, Sangita Kulathinal³, Kari Kuulasmaa³, Kaisa Silander¹, Veikko Salomaa³, Markus Perola^{1,4}, Leena Peltonen^{1,4,5*}

1 Department of Molecular Medicine, National Public Health Institute, Biomedicum, Helsinki, Finland, **2** Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California Los Angeles, Los Angeles, California, United States of America, **3** Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland, **4** Department of Medical Genetics, University of Helsinki, Helsinki, Finland, **5** The Broad Institute, Massachusetts Institute of Technology, Boston, Massachusetts, United States of America

Upstream transcription factor 1 (*USF1*) is a ubiquitously expressed transcription factor controlling several critical genes in lipid and glucose metabolism. Of some 40 genes regulated by *USF1*, several are involved in the molecular pathogenesis of cardiovascular disease (CVD). Although the *USF1* gene has been shown to have a critical role in the etiology of familial combined hyperlipidemia, which predisposes to early CVD, the gene's potential role as a risk factor for CVD events at the population level has not been established. Here we report the results from a prospective genetic-epidemiological study of the association between the *USF1* variants, CVD, and mortality in two large Finnish cohorts. Haplotype-tagging single nucleotide polymorphisms exposing all common allelic variants of *USF1* were genotyped in a prospective case-cohort design with two distinct cohorts followed up during 1992–2001 and 1997–2003. The total number of follow-up years was 112,435 in 14,140 individuals, of which 2,225 were selected for genotyping based on the case-cohort study strategy. After adjustment for conventional risk factors, we observed an association of *USF1* with CVD and mortality among females. In combined analysis of the two cohorts, female carriers of a *USF1* risk haplotype had a 2-fold risk of a CVD event (hazard ratio [HR] 2.02; 95% confidence interval [CI] 1.16–3.53; $p = 0.01$) and an increased risk of all-cause mortality (HR 2.52; 95% CI 1.46–4.35; $p = 0.0009$). A putative protective haplotype of *USF1* was also identified. Our study shows how a gene identified in exceptional families proves to be important also at the population level, implying that allelic variants of *USF1* significantly influence the prospective risk of CVD and even all-cause mortality in females.

Citation: Komulainen K, Alanne M, Auro K, Kilpikari R, Pajukanta P, et al. (2006) Risk alleles of *USF1* gene predict cardiovascular disease of women in two prospective studies. *PLoS Genet* 2(5): e69. DOI: 10.1371/journal.pgen.0020069

Introduction

The *upstream transcription factor 1 (USF1)* gene encoding *USF1*, a ubiquitously expressed transcription factor controlling some 40 genes [1], based on its function is an attractive candidate gene for cardiovascular disease (CVD). Initially this gene was identified as the first familial combined hyperlipidemia (FCHL) gene in rare Finnish pedigrees with multiple affected individuals having a greatly increased risk for CVD [2]. This finding was rapidly replicated in Mexican families [3], and since then *USF1* has also been associated with the metabolic syndrome and type II diabetes in study samples ascertained for these traits [4].

Thus, the combined evidence indicates a role for *USF1* in the molecular background of hyperlipidemias, yet the direct contribution of the *USF1* gene to CVD at the population level has not been addressed. To adequately evaluate the impact of a gene or a risk allele on the disease risk at the population level requires a prospective follow-up study, a golden standard in traditional epidemiology, in which risk factor(s) are measured at the beginning of the follow-up, and the diagnostic endpoints are registered as the study proceeds. We have in this study used two unique cohorts from Finland, the population in which the *USF1* gene was identified as an

FCHL gene, to address the general significance of the *USF1* gene as a risk factor for CVD in population-based, prospective manner. We evaluated allelic variants of the *USF1* gene by genotyping haplotype-tagging single-nucleotide polymorphisms (htSNPs) of the *USF1* locus and assessing their association with CVD events in the two prospective cohorts.

Editor: Jonathan Flint, University of Oxford, United Kingdom

Received: December 28, 2005; **Accepted:** March 23, 2006; **Published:** May 12, 2006

A previous version of this article appeared as an Early Online Release on March 23, 2006 (DOI: 10.1371/journal.pgen.0020069.eor).

DOI: 10.1371/journal.pgen.0020069

Copyright: © 2006 Komulainen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: BMI, body mass index; CHD, coronary heart disease; CI, confidence interval; CVD, cardiovascular disease; FCHL, familial combined hyperlipidemia; HDL, high-density lipoprotein; HR, hazard ratio; htSNP, haplotype-tagging single nucleotide polymorphism; ICD, International Classification of Diseases; LD, linkage disequilibrium; SD, standard deviation; SNP, single nucleotide polymorphism; *USF1*, upstream transcription factor 1

* To whom correspondence should be addressed. E-mail: leena.peltonen@ktl.fi

✉ These authors contributed equally to this work.

Synopsis

Better characterization of molecular events resulting in cardiovascular disease (CVD) requires elucidation of genetic background of CVD. After a CVD candidate gene is identified in family-based studies or case-control studies, population-based prospective studies are needed to demonstrate any potential impact of allelic variants on the CVD risk at the population level. This study addresses the role of different alleles of the *upstream transcription factor 1* (*USF1*) gene, encoding a transcription factor and originally associated with familial combined hyperlipidemia in rare families with multiple affected individuals. The product of *USF1* regulates numerous genes of lipid and glucose metabolism, and the authors show in large population cohorts that specific alleles of *USF1* are associated with the risk of CVD and all-cause mortality among females. The study implies an interesting female-specific risk effect, and should stimulate additional studies of the sex-specific CVD risk genes in different populations.

Specific alleles of the *USF1* gene proved to modify the CVD risk in women and to contribute both to CVD and mortality at the population level.

Results

Figure 1 shows characteristics of the two FINRISK [5] cohorts. The number of individuals recruited in the FINRISK-97 cohort was larger than in the FINRISK-92 at the baseline of the study. However, since the follow-up time was shorter for the FINRISK-97, the number of person-years was smaller for this younger cohort (54,577 person-years in FINRISK-97

versus 57,858 person-years in FINRISK-92). We selected incident cases and random subcohorts from the FINRISK cohorts for genotyping of six htSNPs as shown in Figure 1 and Tables 1 and 2.

The genotype frequencies of all six SNPs followed Hardy-Weinberg equilibrium at the 0.05 level among the subcohorts of the FINRISK study samples. The linkage disequilibrium (LD) structure of the *USF1* gene and haplotype tagging properties of the genotyped SNPs were examined combining the two subcohorts. The selected six htSNPs captured the common haplotypes of *USF1* well (all haplotypes with frequency > 4% in SeattleSNPs database (Figure 2) [6]). Due to the high LD across the *USF1* gene locus (D' 0.995–1.000 between all SNP pairs), only five haplotypes were observed. Importantly, based on the distribution of the SNP alleles into various haplotypes, it was evident that the SNP alleles unequivocally specified haplotypes; thus, the selected SNPs truly represented htSNPs for this gene.

The minor allele frequencies of the *USF1* SNPs varied from 11% to 41% in the subcohorts (Table 3). In both FINRISK cohorts, SNP *rs2073658* showed a difference in the minor allele frequency between the female incident CVD cases and the subcohort free of CVD ($p = 0.05$ for FINRISK-92 and FINRISK-97 combined). This SNP represents the variant previously associated with FCHL in Finnish pedigrees, and in both cohorts the minor allele seemed to be more prevalent among incident female CVD cases when compared to subcohort females free of CVD. For another SNP, *rs2774279*, located 6.8 kb 5' from *rs2073658*, the minor allele frequency was higher among the subcohort females free of CVD than among incident female CVD cases ($p = 0.03$ for FINRISK-92 and FINRISK-97 combined), especially in the FINRISK-92 ($p = 0.008$) study sample with a longer follow-up. The minor alleles of these two SNPs both tag one haplotype of *USF1*. The minor allele of *rs2073658* tags the most common *USF1* haplotype, and the minor allele of *rs2774279* tags the second most common haplotype of *USF1* (Figure 2).

Cox proportional hazards regression analysis measuring time-to-event was used to estimate the risk of an incident CVD event or mortality during the follow-up period in relation to the *USF1* SNPs. Allele-specific hazard ratios for the six SNPs tested are shown in Table 4 for analyses testing the two FINRISK cohorts separately and for analyses pooling the cohorts. Analysis of the previously analyzed SNP *rs2073658* [1–4,7,8] revealed an increased risk of incident CVD (HR 2.02; 95% confidence interval [CI] 1.16–3.53; $p = 0.01$) and all-cause mortality (HR 2.52; 95% CI 1.46–4.35; $p = 0.0009$) for carriers of T-allele among females when compared to noncarriers of the allele. The increased risk was evident especially among the FINRISK-97 females; however, the hazard ratio was increased also among the FINRISK-92 females, although it did not reach statistical significance. The second SNP showing a difference in the minor allele frequency between incident cases and subcohort free of CVD, *rs2774279*, was also associated with increased risk of incident CVD and mortality in time-to-event analysis. Risk of mortality was increased among the female carriers of the G-allele of *rs2774279* (HR 4.43; 95% CI 1.58–12.40; $p = 0.005$). Again, the increased risk was observed especially among the FINRISK-97 females. Also, the risk of CVD was higher among female carriers of the *rs2774279* G-allele (HR 4.01; 95% CI 1.30–12.39; $p = 0.02$), and the effect of the SNP was observed in

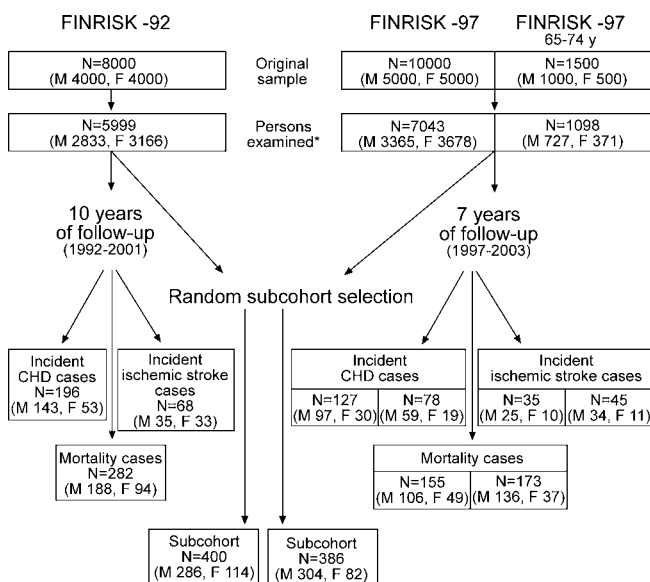


Figure 1. A Description of the FINRISK 1992 and 1997 Cohorts

Compared to FINRISK-92, the FINRISK-97 cohort includes an additional sample of individuals aged 65–74 y. Numbers for this additional sample are described at the right-hand side for each endpoint. Persons examined refers to cohort individuals for whom information on smoking, blood pressure, cholesterol, and DNA, as well as consent for the use of DNA to study CHD and stroke, were available. Subcohorts are stratified random samples of the original cohorts including also cases. Mortality cases show total mortality, including also those who died from CHD or stroke. Thus, numbers in the boxes of subcohorts and outcome events are not mutually exclusive (see Table 1). F, females; M, males.

DOI: 10.1371/journal.pgen.0020069.g001

Table 1. Numbers of Study Subjects in the Random Subcohort and with Different Endpoint Events in FINRISK-92 and FINRISK-97 Cohorts

Cohort	Sex	Incident CVD ^a	Death (Any Cause)	Subcohort	Incident CVD and Death	Incident CVD and Subcohort	Death and Subcohort	Incident CVD, Death, and Subcohort
FINRISK-92	Males	170	188	286	48	37	34	12
	Females	86	94	114	10	7	9	3
FINRISK-97	Males	204	242	304	54	23	35	7
	Females	68	86	72	24	5	8	3

^aIncident CHD and incident ischemic stroke combined; the first one (either CHD or stroke event) was used in the analyses. Altogether 21 out of the 528 CVD cases had both incident CHD and incident ischemic stroke during the follow-up period. Two were FR97 females, 11 were FR97 males, and eight were FR92 males.
DOI: 10.1371/journal.pgen.0020069.t001

both FINRISK cohorts. Among FINRISK-97 females, the effect of SNP *rs2774279* was visible when comparing carriers of the G-allele with others (HR 15.38; 95% CI 2.49–94.87; $p = 0.003$). Interestingly, among the FINRISK-92 females, the effect of SNP *rs2774279* was observed when comparing carriers of the A-allele with others (HR 0.33; 95% CI 0.13–0.85; $p = 0.02$).

No significant associations with the risk of an incident CVD event or mortality were observed among males in combined analysis of the two cohorts. Among FINRISK-92 males, a suggestive association was obtained for risk of mortality and the T-allele of SNP *rs2516839* (HR 1.83; 95% CI 1.00–3.36; $p = 0.05$), and including FINRISK-92 females to the analysis further strengthened the results (HR 1.86; 95% CI 1.13–3.05; $p = 0.01$). Otherwise, the only association seen in the pooled

analyses of males and females was a decreased risk of CVD (HR 0.53; 95% CI 0.30–0.95; $p = 0.03$) observed for carriers of the *rs2073658* C-allele when compared to noncarriers of the allele among the FINRISK-92 cohort.

We used permutation analysis to determine a critical p -value for 5% level of significance corrected for multiple testing for our results; we permuted the genotype while retaining the phenotype data and repeated the same analyses that were performed with the actual dataset and repeated this procedure 1,000 times. In the time-to-event analyses where FINRISK-92 males and females were pooled together, permutation analysis suggested a critical p -value of 0.005 for a 5% level of significance for both CVD risk and mortality risk. All other critical p -values obtained from permutation analysis can be seen in Table 4.

Table 2. Baseline Characteristics of Cardiovascular Cases and Subcohort

Characteristic	FR92 Males		FR92 Females		FR97 Males		FR97 Females		FR92+FR97 Males		FR92+FR97 Females	
	CVD Cases ^a	Sub-cohort ^b	CVD Cases ^a	Sub-cohort ^b	CVD Cases ^a	Sub-cohort ^b	CVD Cases ^a	Sub-cohort ^b	CVD Cases ^a	Sub-cohort ^b	CVD Cases ^a	Sub-cohort ^b
Number of subjects	170	220	86	104	204	231	68	71	374	451	154	175
Smoking ^c	67 (39)	54 (25)	14 (16)	10 (10)	55 (27)	54 (23)	14 (21)	8 (11)	122 (33)	108 (24)	28 (18)	18 (10)
Diabetes ^c	16 (9)	10 (5)	13 (15)	4 (4)	28 (14)	13 (6)	12 (18)	5 (7)	44 (12)	23 (5)	25 (16)	9 (5)
Hypertension ^c	120 (71)	120 (55)	68 (79)	66 (63)	153 (75)	151 (65)	46 (68)	53 (75)	273 (73)	271 (60)	114 (74)	119 (68)
Age at baseline (years) ^d	54.2 (7.7)	52.8 (9.1)	55.0 (8.0)	55.9 (6.7)	61.9 (8.7)	59.9 (10.2)	62.8 (7.4)	63.1 (7.5)	58.4 (9.1)	56.4 (10.3)	58.5 (8.6)	58.8 (7.9)
Age at first CVD event (years) ^d	59.5 (7.4)	NA	60.6 (8.5)	NA	65.5 (8.9)	NA	66.3 (7.5)	NA	62.8 (8.8)	NA	63.1 (8.6)	NA
Total cholesterol (mmol/l) ^{d,e}	6.28 (1.11)	5.87 (0.99)	6.06 (1.30)	5.96 (0.98)	5.79 (1.03)	5.66 (0.97)	6.05 (0.93)	6.00 (1.11)	6.02 (1.09)	5.76 (0.99)	6.06 (1.15)	5.98 (1.03)
HDL cholesterol (mmol/l) ^{d,e}	1.17 (0.32)	1.27 (0.35)	1.35 (0.38)	1.58 (0.35)	1.17 (0.28)	1.26 (0.32)	1.41 (0.37)	1.53 (0.37)	1.17 (0.30)	1.27 (0.34)	1.37 (0.38)	1.56 (0.36)
Non-HDL cholesterol ^{d,e}	5.12 (1.13)	4.59 (1.02)	4.72 (1.36)	4.37 (1.02)	4.62 (1.00)	4.39 (0.97)	4.64 (0.95)	4.47 (1.05)	4.85 (1.09)	4.49 (1.00)	4.68 (1.19)	4.41 (1.03)
Triglycerides ^{d,e}	2.12 (1.17)	1.78 (0.99)	1.85 (1.14)	1.32 (0.76)	1.89 (1.09)	1.69 (1.05)	1.78 (0.92)	1.55 (0.86)	2.00 (1.14)	1.74 (1.02)	1.82 (1.05)	1.41 (0.81)
Waist-to-hip ratio ^d	0.97 (0.06)	0.95 (0.07)	0.84 (0.07)	0.80 (0.06)	0.97 (0.06)	0.94 (0.06)	0.84 (0.07)	0.83 (0.07)	0.97 (0.06)	0.95 (0.06)	0.84 (0.07)	0.81 (0.06)
BMI (kg/m ²) ^d	28.6 (3.9)	27.4 (3.9)	29.3 (5.2)	26.9 (4.7)	28.6 (3.5)	27.5 (3.9)	29.3 (4.5)	27.9 (4.4)	28.6 (3.6)	27.5 (3.9)	29.3 (4.9)	27.3 (4.6)

^aIncident CVD cases (includes incident CHD cases and incident ischemic stroke cases).

^bSubcohort members without CVD at baseline or during the follow-up.

^cData are given as number (%).

^dData are given as mean (SD).

^ePersons with lipid-lowering medication at baseline ($n = 44$) were excluded from calculations.

FR, FINRISK; NA, not applicable.

DOI: 10.1371/journal.pgen.0020069.t002

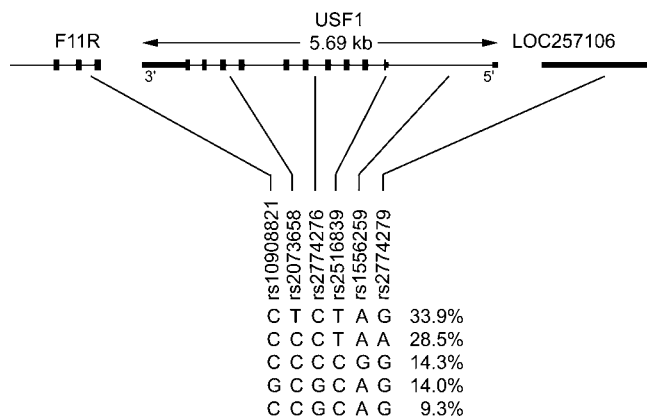


Figure 2. The Haplotype Structure in the *USF1* Gene Locus

The six genotyped SNPs from the *USF1* gene locus resulted in five common haplotypes. The haplotype frequencies were calculated for the two FINRISK subcohorts combined.

DOI: 10.1371/journal.pgen.0020069.g002

In our age-stratified study samples, the average age of the onset of CVD was about the same for males and females (Table 2). Since females generally have their first CVD event at an older age than men, we tested whether the observed sex difference was primarily due to a possible difference in relative age at event for males and females in our study sample. The time-to-event analysis combined for the two cohorts was repeated for males who had their first CVD event early (before age 55; $n = 70$), but no associations with CVD or mortality were observed (unpublished data).

The time-to-event results of individual htSNPs suggested that the *rs2073658* T-allele and *rs2774279* G-allele increase the risk of CVD and mortality among females. From the *USF1* haplotypes it was obvious that the *rs2073658* T-allele was present only in haplotype CTCTAG (population frequency 34%) (Figure 2), thus specifying it as a risk-increasing haplotype. The *rs2774279* G-allele was present in all other haplotypes but the CCCTAA haplotype (population frequency 29%), thus implying a protective role for this haplotype. We wanted to investigate more closely how the carriership of these haplotypes relates to the risk of CVD and mortality. Time-to-event analysis was performed to estimate the risk of CVD for females carrying the risk haplotype CTCTAG without the protective haplotype CCCTAA. The females carrying the risk haplotype CTCTAG, but without the protective haplotype CCCTAA, had the highest CVD risk when compared to any other group of haplotype carriers, this risk being 2.8-fold (HR 2.77; 95% CI 1.50–5.13; $p = 0.001$). This result remained significant after correcting for multiple testing: In permutation analysis, the critical p -value for 5% significance level was 0.003.

Finally, we tested whether the sequence variants of *USF1* have an effect on lipids, body mass index (BMI), or waist-to-hip circumference ratio, the traits characteristic of FCHL and other conditions with earlier-reported associations with *USF1* and known contributors to the CVD risk. The obtained results support previous findings of the best association observed for lipid values in males, and are summarized in Tables 5 and S1.

When males and females were analyzed together, significant associations among incident CVD cases were seen with

higher levels of cholesterol and non-high-density lipoprotein (HDL) cholesterol with the C-allele of *rs2073658* ($p = 0.02$ and $p = 0.02$ for cholesterol and non-HDL cholesterol, respectively), C-allele of *rs2516839* ($p < 0.0001$ and $p = 0.0001$), and G-allele of *rs1556259* ($p = 0.01$ and $p = 0.03$). Among the subcohort free of CVD, the G-allele of *rs2774276* was associated with higher waist-to-hip ratios ($p = 0.009$). For all these pooled analyses the critical p -value for 5% significance level was 0.006.

Discussion

We previously established the association of the *USF1* gene with FCHL in a set of rare Finnish families with multiple dyslipidemic individuals [2], known to be predisposed to a significantly increased risk for CVD. Here we show that allelic variants of the *USF1* gene have an impact on CVD risk also at the population level, a finding rarely established for any complex disease gene so far.

The biological importance of the *USF1* gene has been implied in several studies reporting an association between *USF1* and disorders of lipid and glucose metabolism, mostly in study samples ascertained for these traits. Following the first association report for *USF1* with FCHL in Finnish FCHL families [2], the association was replicated in Mexican FCHL families [3]. Further, in Utah pedigrees ascertained for early death due to coronary heart disease (CHD), early stroke, or early-onset of hypertension, *USF1* was found to be associated with FCHL-related lipid traits, especially in males [8]. *USF1* has also been implicated in the etiology of metabolic syndrome and type II diabetes [4], common conditions predisposing to premature CVD. Additional studies show variants of *USF1* as being associated with features of glucose and lipid homeostasis in healthy young men [9] and increased adipocyte lipolysis in healthy obese women [7]. Although multiple studies have provided evidence of the role of *USF1* in CVD-associated dyslipidemias, the direct contribution of the *USF1* gene to CVD and mortality at the population level has not been assessed.

Here we prospectively followed up two distinct cohorts representative of the general Finnish population of the study areas. Using Cox proportional hazards regression analysis measuring time-to-event, we found a female-specific association of two *USF1* htSNPs, defining a “risk allele” of *USF1*, with both CVD and all-cause mortality. We also tentatively identified a “protective” allele for this gene.

The *USF1* risk allele identified here for the cohort females differs from the risk allele segregating in FCHL families. In 60 Finnish FCHL families the risk was associated with the common allele of *rs2073658* [2], whereas in the present study the SNP haplotype CTCTAG, carrying the minor allele of *rs2073658*, was associated with a risk effect. However, our original FCHL study did not address the female risk allele in any detail, since the most significant association was observed in males with high triglycerides. We performed a simple analysis of the original FCHL data, choosing families (32 out of 60) in which at least 50% of those affected were female. Among these affected FCHL females the frequency of the minor allele of *rs2073658* (haplotype CTCTAG seen in FINRISK females) was higher than among the unaffected females of those families (45% versus 38%, respectively), thus supporting our findings. In addition, the frequency of the

Table 3. Minor Allele Frequencies of *USF1* htSNPs for CVD Cases and the Subcohort

Marker	FR92 Males		FR92 Females		FR97 Males		FR97 Females		FR92+97 Males		FR92+97 Females	
	CVD Cases ^a (n = 170)	Subcohort ^b (n = 220)	CVD Cases ^a (n = 86)	Subcohort ^b (n = 104)	CVD Cases ^a (n = 204)	Subcohort ^b (n = 231)	CVD Cases ^a (n = 68)	Subcohort ^b (n = 71)	CVD Cases ^a (n = 374)	Subcohort ^b (n = 451)	CVD Cases ^a (n = 154)	Subcohort ^b (n = 175)
<i>rs10908821</i> (minor allele G) ^c	0.14 (45)	0.16 (68)	0.16 (26)	0.15 (30)	0.13 (52)	0.16 (71)	0.11 (14)	0.11 (15)	0.13 (97)	0.16 (139)	0.13 (40)	0.13 (45)
<i>rs2073658</i> (minor allele T) ^c	0.34 (111)	0.32 (138)	0.38 (61)	0.30 (62)	0.37 (147)	0.37 (166)	0.38 (52)	0.31 (43)	0.36 (258)	0.35 (304)	0.38 (113)	0.31 (105)
<i>rs2774276</i> (minor allele G) ^c	0.24 (78)	0.25 (109)	0.26 (42)	0.22 (46)	0.20 (79)	0.26 (117)	0.20 (27)	0.24 (34)	0.22 (157)	0.26 (226)	0.23 (69)	0.23 (80)
<i>rs2516839</i> (minor allele C) ^c	0.39 (125)	0.41 (176)	0.43 (70)	0.38 (78)	0.35 (140)	0.37 (166)	0.32 (42)	0.36 (51)	0.37 (265)	0.39 (342)	0.38 (112)	0.37 (129)
<i>rs1556259</i> (minor allele G) ^c	0.15 (48)	0.16 (67)	0.17 (28)	0.16 (33)	0.16 (64)	0.11 (50)	0.13 (16)	0.12 (17)	0.16 (112)	0.13 (117)	0.15 (44)	0.14 (50)
<i>rs2774279</i> (minor allele A) ^c	0.27 (86)	0.27 (116)	0.20 (32)	0.32 (65)	0.27 (105)	0.26 (118)	0.30 (39)	0.32 (46)	0.27 (191)	0.26 (234)	0.24 (71)	0.32 (111)

^aIncident CVD cases.
^bSubcohort members without CVD at baseline or during the follow-up.
^cData are given as frequency (count).
 FR, FINRISK.
 DOI: 10.1371/journal.pgen.0020069.t003

“protective” FINRISK haplotype CCCTAA was 23% in affected females of those families and 30% in unaffected females. Thus the obtained results of the “risk allele” of *USF1* in these two studies are comparable, although we recognize that the ascertainment strategy for the FCHL families results in a biased selection of population alleles and complicates the estimation of any sex-specific effect in these rare families. Many previous studies have found the influence of *USF1* to be more prominent among males [2,8], or have studied only males [9], whereas here we saw the most significant evidence of association among females. The lack of association among males at the population level in our cohorts could be explained by environmental covariates. CVD events are more common in males, most probably due to clustering of numerous lifestyle risk factors, including smoking, obesity, and high blood pressure, this increasing the environmental “noise” and complicating genetic analyses in males compared to females.

The association of *USF1* with CVD in females of two distinct population cohorts supports the real involvement of *USF1* in CVD risk. Moreover, the trends of the hazard ratios for the risk-associated alleles of *USF1* were the same in both FINRISK cohorts, and the same risk alleles of *USF1* were associated both with CVD events and all-cause mortality among females, further supporting their significance. Finally, correcting for multiple testing with permutations providing a more stringent significance threshold for the original *p*-values still retained a significant association between *USF1* variation and the risk of CVD and mortality among females.

In the association analyses of SNPs with lipid parameters, we observed *USF1* htSNPs to strongly associate with plasma cholesterol levels and non-HDL cholesterol especially among the male CVD cases. These findings confirm prior reports that *USF1* variants contribute to differences in lipid profiles [2–4,7–9]. However, no associations with traditional risk factors were obtained that would explain the increased risk of CVD and mortality among female carriers of *USF1* risk alleles. The question remains whether the impact of *USF1* on lipid parameters among females is as powerful as among males, or if the gene variants have alternative ways of increasing the CVD risk—for example, through inflammation-related pathways.

As noted previously, the strong LD of the *USF1* locus extends the critical FCHL region to 46 kb in Finnish FCHL families [2]. This DNA region may harbor additional variants that represent true functional variants that affect CVD risk. Sequencing of the complete region in sufficiently large study samples would shed light on the risk-increasing variants of *USF1*, and would enable the detailed functional studies needed to establish their role. So far, functional data exist only for the most significantly associated SNP of the original FCHL study, *rs2073658* [1].

In conclusion, here we have shown that the risk effect of a gene identified in a rare set of FCHL families who are at exceptionally high risk for cardiovascular events can be confirmed in a population-based prospective follow-up study. We observed one “risk” and one “protective” allelic haplotype of *USF1* that significantly contribute to the risk for CVD and all-cause mortality among females. To evaluate the significance of these results obtained in the Finnish cohorts for CVD risk more globally, other populations need to be analyzed. Still, this study demonstrates, to our knowl-

Table 4. Risk of a Cardiovascular Event and All-Cause Mortality for the USF1 SNPs^a

Marker	Allele ^b	FR92 Males		FR97 Males		FR92+FR97 Males		FR92 Females		FR97 Females		FR92+FR97 Females		
		Hazard ratio (95% CI)	p-Value	Hazard ratio (95% CI)	p-Value	Hazard ratio (95% CI)	p-Value	Hazard ratio (95% CI)	p-Value ^c	Hazard ratio (95% CI)	p-Value ^d	Hazard ratio (95% CI)	p-Value ^e	
rs10908821	G	CVD	0.91 (0.54–1.52)	0.71	0.73 (0.43–1.22)	0.23	0.77 (0.26–2.29)	0.64	0.91 (0.35–2.34)	0.84	1.25 (0.39–4.03)	0.70	1.02 (0.54–1.94)	0.95
		Mortality	1.04 (0.66–1.64)	0.86	1.00 (0.65–1.54)	0.99	1.07 (0.54–2.11)	0.85	1.05 (0.50–2.24)	0.90	1.47 (0.49–4.45)	0.49	1.21 (0.69–2.14)	0.51
rs2073658	C	CVD	0.63 (0.33–1.21)	0.16	1.04 (0.52–2.11)	0.90	0.83 (0.16–4.15)	0.82	0.36 (0.09–1.52)	0.16	0.52 (0.17–1.60)	0.25	0.45 (0.20–0.99)	0.05
		Mortality	0.73 (0.40–1.35)	0.32	0.62 (0.35–1.10)	0.10	0.71 (0.17–2.94)	0.64	0.50 (0.14–1.70)	0.26	1.17 (0.28–4.86)	0.83	0.73 (0.33–1.60)	0.43
T	CVD		1.02 (0.65–1.61)	0.92	1.07 (0.68–1.70)	0.77	1.01 (0.38–2.64)	0.99	1.68 (0.74–3.80)	0.22	3.24 (1.21–8.69)	0.02	2.02 (1.16–3.53)	0.01
		Mortality	0.99 (0.66–1.49)	0.95	0.72 (0.46–1.13)	0.15	0.84 (0.24–2.94)	0.79	1.87 (0.94–3.71)	0.07	4.73 (1.49–15.06)	0.009	2.52 (1.46–4.35)	0.0009
rs2774276	C	CVD	1.14 (0.46–2.80)	0.78	0.97 (0.39–2.41)	0.95	1.13 (0.10–13.00)	0.92	3.87 (0.19–80.56)	0.38	1.25 (0.27–5.75)	0.78	2.44 (0.56–10.66)	0.23
		Mortality	2.37 (0.95–5.89)	0.06	0.85 (0.36–2.00)	0.70	1.36 (0.10–17.93)	0.81	1.56 (0.18–13.56)	0.69	1.07 (0.19–5.97)	0.94	1.29 (0.40–4.18)	0.67
G	CVD		0.95 (0.60–1.49)	0.81	0.64 (0.40–1.01)	0.05	0.74 (0.32–1.71)	0.48	1.44 (0.64–3.21)	0.38	1.12 (0.44–2.83)	0.81	1.17 (0.68–2.00)	0.57
		Mortality	0.98 (0.65–1.49)	0.93	0.93 (0.61–1.41)	0.72	1.02 (0.58–1.80)	0.93	1.25 (0.64–2.45)	0.51	1.07 (0.43–2.65)	0.88	1.17 (0.71–1.93)	0.53
rs2576839	C	CVD	0.90 (0.56–1.44)	0.65	0.76 (0.49–1.18)	0.22	0.81 (0.44–1.50)	0.51	1.20 (0.51–2.85)	0.68	1.76 (0.70–4.44)	0.23	1.36 (0.77–2.42)	0.29
		Mortality	0.91 (0.60–1.38)	0.65	0.85 (0.56–1.28)	0.44	0.97 (0.52–1.79)	0.92	1.09 (0.54–2.17)	0.82	2.00 (0.74–5.43)	0.17	1.34 (0.80–2.23)	0.27
T	CVD		1.08 (0.60–1.94)	0.81	0.92 (0.50–1.70)	0.79	1.03 (0.34–3.12)	0.96	1.08 (0.37–3.20)	0.89	1.67 (0.48–5.78)	0.42	1.35 (0.62–2.91)	0.45
		Mortality	1.83 (1.00–3.36)	0.05	0.69 (0.39–1.22)	0.21	1.06 (0.35–3.24)	0.92	1.56 (0.61–4.01)	0.35	1.29 (0.34–4.91)	0.71	1.49 (0.75–2.95)	0.25
rs1556259	G	CVD	1.07 (0.66–1.74)	0.79	1.18 (0.72–1.94)	0.52	1.18 (0.45–3.12)	0.74	0.99 (0.38–2.54)	0.98	1.86 (0.54–6.37)	0.33	1.36 (0.72–2.57)	0.35
		Mortality	0.87 (0.54–1.39)	0.55	0.96 (0.61–1.51)	0.87	0.98 (0.46–2.07)	0.95	0.79 (0.34–1.84)	0.58	2.50 (0.74–8.46)	0.14	1.05 (0.57–1.92)	0.89
rs2774279	A	CVD	0.87 (0.56–1.37)	0.55	1.16 (0.72–1.86)	0.55	1.03 (0.39–2.76)	0.95	0.33 (0.13–0.85)	0.02	0.67 (0.28–1.63)	0.38	0.53 (0.31–0.91)	0.02
		Mortality	1.12 (0.74–1.70)	0.60	0.98 (0.65–1.47)	0.91	0.97 (0.45–2.05)	0.93	0.57 (0.30–1.11)	0.10	0.76 (0.30–1.90)	0.56	0.62 (0.38–1.01)	0.06
G	CVD		0.99 (0.44–2.21)	0.97	0.86 (0.39–1.91)	0.72	0.89 (0.07–10.98)	0.93	1.41 (0.28–7.26)	0.68	15.38 (2.49–94.87)	0.003	4.01 (1.30–12.39)	0.02
		Mortality	0.62 (0.30–1.26)	0.19	0.57 (0.26–1.21)	0.14	0.65 (0.04–9.66)	0.75	1.36 (0.45–4.11)	0.58	75.26 (4.97–1,139.47)	0.002	4.43 (1.58–12.40)	0.005

^aAnalyses adjusted for cohort, age, sex, hypertension, smoking, diabetes, total cholesterol-to-HDL ratio, and BMI, stratified by eastern and western Finland.

^bPersons carrying the allele are pooled together and analyzed against noncarriers.

^cCritical p-value for 5% level of significance is 0.0006 when analyzing CVD.

^dCritical p-value for 5% level of significance is 0.001 when analyzing CVD and 0.0006 when analyzing mortality.

^eCritical p-value for 5% level of significance is 0.003 when analyzing CVD and 0.004 when analyzing mortality.

FR, FINRISK.

DOI: 10.1371/journal.pgen.0020069.t004



Table 5. Association between *USF1* SNP Alleles and Lipids, BMI, and Waist-To-Hip Ratio in FINRISK CVD Case Males and Females

Marker	Allele ^a	Sex	Cholesterol (mmol/l)	p-Value ^{b,c}	HDL (mmol/l)	p-Value ^b	Non-HDL Cholesterol	p-Value ^{b,d}	Triglycerides (mmol/l)	p-Value ^{b,e}	BMI (kg/m ²)	p-Value ^{b,f}	Waist-To-Hip Ratio	p-Value ^b
<i>rs10908821</i>	G	Males	6.05 (0.97)/ 6.01 (1.13)	0.99	1.16 (0.28)/ 1.17 (0.31)	0.90	4.88 (1.04)/ 4.84 (1.11)	0.99	2.06 (1.18)/ 2.01 (1.14)	0.87	28.9 (3.6)/ 28.7 (3.7)	0.68	0.97 (0.06)/ 0.97 (0.06)	0.36
		Females	6.18 (1.04)/ 5.98 (1.17)	0.33	1.36 (0.35)/ 1.38 (0.39)	0.69	4.84 (1.04)/ 4.60 (1.23)	0.25	1.75 (1.01)/ 1.85 (1.07)	0.69	28.5 (4.1)/ 29.5 (5.0)	0.40	0.82 (0.07)/ 0.84 (0.07)	0.21
<i>rs2073658</i>	C	Males	6.08 (1.09)/ 5.70 (1.00)	0.02	1.18 (0.30)/ 1.16 (0.34)	0.60	4.90 (1.08)/ 4.54 (0.99)	0.03	2.01 (1.12)/ 2.00 (1.25)	0.60	28.6 (3.6)/ 28.4 (4.0)	0.85	0.97 (0.06)/ 0.97 (0.06)	0.53
		Females	6.08 (1.19)/ 5.83 (0.84)	0.44	1.37 (0.37)/ 1.44 (0.46)	0.92	4.71 (1.24)/ 4.39 (0.75)	0.35	1.89 (1.04)/ 1.37 (1.04)	0.03	29.6 (4.9)/ 27.3 (4.2)	0.06	0.84 (0.07)/ 0.83 (0.07)	0.74
T	Males		5.92 (1.10)/ 6.20 (1.03)	0.03	1.18 (0.31)/ 1.17 (0.29)	0.83	4.74 (1.09)/ 5.03 (1.04)	0.02	1.91 (1.05)/ 2.15 (1.25)	0.04	28.6 (3.6)/ 28.6 (3.7)	0.98	0.97 (0.06)/ 0.97 (0.06)	0.61
		Females	6.03 (1.16)/ 6.05 (1.13)	0.91	1.39 (0.42)/ 1.36 (0.31)	0.93	4.64 (1.21)/ 4.70 (1.16)	0.77	1.81 (1.05)/ 1.83 (1.07)	0.89	28.8 (4.6)/ 29.9 (5.0)	0.13	0.84 (0.07)/ 0.83 (0.08)	0.39
<i>rs2774276</i>	C	Males	6.00 (1.06)/ 6.68 (1.16)	0.01	1.18 (0.30)/ 1.09 (0.22)	0.20	4.82 (1.06)/ 5.58 (1.17)	0.004 ^g	2.00 (1.13)/ 2.24 (1.33)	0.19	28.7 (3.6)/ 29.1 (3.9)	0.75	0.97 (0.06)/ 0.98 (0.06)	0.44
		Females	6.07 (1.16)/ 5.49 (0.81)	0.34	1.39 (0.39)/ 1.23 (0.26)	0.13	4.68 (1.19)/ 4.26 (0.98)	0.67	1.82 (1.06)/ 1.76 (0.92)	0.81	29.3 (4.7)/ 28.4 (5.3)	0.98	0.84 (0.07)/ 0.80 (0.10)	0.21
G ^h	Males		6.18 (1.07)/ 5.96 (1.08)	0.13	1.17 (0.30)/ 1.17 (0.30)	0.88	5.01 (1.13)/ 4.78 (1.05)	0.12	2.13 (1.30)/ 1.95 (1.04)	0.34	28.8 (3.8)/ 28.6 (3.6)	0.66	0.97 (0.06)/ 0.97 (0.06)	0.82
		Females	6.11 (1.16)/ 5.98 (1.14)	0.53	1.37 (0.33)/ 1.39 (0.41)	0.93	4.75 (1.15)/ 4.60 (1.21)	0.44	1.80 (0.97)/ 1.83 (1.11)	0.91	29.4 (4.8)/ 29.1 (4.7)	0.65	0.83 (0.07)/ 0.84 (0.07)	0.18
<i>rs2516839</i>	C ^h	Males	6.18 (1.05)/ 5.77 (1.08)	0.001 ^g	1.17 (0.29)/ 1.17 (0.31)	0.71	5.01 (1.08)/ 4.61 (1.05)	0.001 ^g	2.12 (1.23)/ 1.87 (1.00)	0.09	28.7 (3.6)/ 28.6 (3.7)	0.94	0.97 (0.06)/ 0.97 (0.06)	0.84
		Females	6.21 (1.17)/ 5.77 (1.06)	0.02	1.39 (0.36)/ 1.35 (0.40)	0.42	4.83 (1.19)/ 4.42 (1.16)	0.04	1.87 (1.04)/ 1.75 (1.09)	0.40	30.0 (4.6)/ 28.1 (4.9)	0.006 ^g	0.84 (0.08)/ 0.84 (0.07)	0.94
T	Males		5.95 (1.07)/ 6.33 (1.13)	0.04	1.17 (0.30)/ 1.17 (0.29)	0.86	4.78 (1.06)/ 5.15 (1.15)	0.04	2.02 (1.15)/ 2.01 (1.14)	0.86	28.7 (3.7)/ 28.5 (3.5)	0.75	0.97 (0.06)/ 0.97 (0.06)	0.68
		Females	6.06 (1.17)/ 5.88 (1.05)	0.52	1.38 (0.39)/ 1.35 (0.32)	0.88	4.68 (1.23)/ 4.53 (1.01)	0.60	1.81 (1.09)/ 1.88 (0.87)	0.50	29.2 (4.8)/ 29.3 (4.9)	0.77	0.84 (0.07)/ 0.83 (0.08)	0.29
<i>rs1556259</i>	G	Males	6.22 (1.05)/ 5.93 (1.09)	0.03	1.20 (0.30)/ 1.16 (0.30)	0.22	5.02 (1.02)/ 4.77 (1.10)	0.06	2.01 (1.12)/ 2.00 (1.14)	0.97	28.4 (3.6)/ 28.7 (3.7)	0.42	0.97 (0.06)/ 0.97 (0.06)	0.49
		Females	6.23 (1.22)/ 5.92 (1.11)	0.22	1.41 (0.39)/ 1.37 (0.38)	0.41	4.82 (1.23)/ 4.55 (1.16)	0.36	1.95 (1.08)/ 1.74 (1.05)	0.31	30.4 (4.5)/ 28.8 (4.9)	0.07	0.85 (0.08)/ 0.83 (0.07)	0.29
<i>rs2774279</i>	A	Males	6.05 (1.07)/ 6.05 (1.10)	0.52	1.16 (0.30)/ 1.18 (0.30)	0.54	4.80 (1.05)/ 4.87 (1.11)	0.62	2.03 (1.14)/ 1.99 (1.15)	0.47	28.7 (3.6)/ 28.7 (3.7)	0.90	0.97 (0.06)/ 0.97 (0.06)	0.90
		Females	5.97 (1.22)/ 6.08 (1.10)	0.45	1.34 (0.37)/ 1.39 (0.39)	0.54	4.63 (1.33)/ 4.69 (1.09)	0.64	1.87 (1.14)/ 1.79 (1.01)	0.89	29.2 (5.2)/ 29.3 (4.6)	0.77	0.84 (0.08)/ 0.83 (0.07)	0.59
G ^h	Males		6.03 (1.10)/ 5.78 (0.85)	0.36	1.17 (0.30)/ 1.17 (0.28)	0.94	4.86 (1.10)/ 4.61 (0.82)	0.35	2.01 (1.15)/ 1.95 (1.06)	0.97	28.6 (3.7)/ 29.1 (3.7)	0.48	0.97 (0.06)/ 0.98 (0.05)	0.49
		Females	6.03 (1.16)/ 6.13 (1.10)	0.88	1.38 (0.38)/ 1.31 (0.25)	0.99	4.65 (1.19)/ 4.83 (1.20)	0.83	1.81 (1.06)/ 2.00 (1.19)	0.75	29.1 (4.7)/ 31.4 (6.1)	0.30	0.84 (0.07)/ 0.84 (0.04)	0.69

Data on cholesterol, HDL, non-HDL, triglycerides, BMI, and waist-to-hip ratio are presented as mean (SD) for carriers of the allele/mean (SD) for noncarriers of the allele.

^aPersons carrying the allele are pooled together and analyzed against noncarriers.

^bp-Value for analysis of covariance. Persons with lipid-lowering medication (21 males and five females) were excluded from analysis of lipid variables.

^cCritical p-value for 5% level of significance is 0.007 when analyzing males and 0.005 when analyzing females.

^dCritical p-value for 5% level of significance is 0.006 when analyzing males or females.

^eCritical p-value for 5% level of significance is 0.006 when analyzing males and 0.005 when analyzing females.

^fCritical p-value for 5% level of significance is 0.007 when analyzing males.

^gp-Value less than the critical value corrected for multiple testing.

^hThis allele showed suggestive association when analyzing the subcohort (Table S1). DOI: 10.1371/journal.pgen.0020069.t005



edge for the first time, that *USF1* gene variation may have a long-term predictive effect on CVD risk in females, and thus adds another piece of proof to the accumulating evidence that the gene plays an important role in the progression of CVD.

Materials and Methods

Study populations. Our population cohorts were collected via the FINRISK [5] studies. FINRISK surveys are carried out every 5 y and are designed to assess the prevalence and risk factors of cardiovascular diseases in Finland. The FINRISK-92 study contacted a random sample of 8,000 persons aged 25–64 y from four geographical areas of Finland (Helsinki region, southwestern Finland, North Karelia, and Kuopio region) as part of the WHO MONICA Project (Multinational Monitoring of Trends and Determinants in Cardiovascular Disease, an international study conducted under the auspices of the World Health Organization) [10]. The persons were selected through random sampling of the population, stratified by sex, area, and 10-y age group (25–34, 35–44, 45–54, and 55–64 y). The protocol was as established by the WHO MONICA [10] Project.

The FINRISK-97 study contacted 10,000 persons from the same regions, selected using the same sampling procedure and using the same protocols as the FINRISK-92 study, but including also the province of Oulu. In addition, 250 females and 500 males aged 65–74 y were included from two of the areas. At the beginning of the follow-up period, the overall response rates covering participants with all required information for the genetic study were 71% for males and 79% for females in the FINRISK-92 study, and 68% for males and 74% for females in the FINRISK-97 study, resulting in a total of 14,140 participants out of the 19,500 in the original sample (Figure 1). No overlap exists between the FINRISK-92 and the FINRISK-97 cohorts.

The FINRISK-92 and FINRISK-97 studies were approved by the Ethical Committee of the National Public Health Institute, the participants gave an informed consent, and the principles of the Helsinki Declaration were followed.

Both FINRISK-92 and –97 cohorts are participating in the ongoing MORGAM Project (<http://www.ktl.fi/morgam>) [11], a multinational collaborative study.

The follow-up period for participants of FINRISK-92 lasted from spring 1992 until the end of December 2001, a total of 10 y. In FINRISK-97 the participants were followed up for 7 y, from 1997 until the end of December 2003. More detailed information about the FINRISK cohorts and the follow-up procedures can be found in the cohort descriptions of the MORGAM Project (<http://www.ktl.fi/publications/morgam/cohorts>).

Baseline examination and follow-up of FINRISK studies. At the baseline, all participants were asked to fill in a self-administered questionnaire and were physically examined. The physical examination included blood pressure, waist, hip, weight, and height measurements. The self-administered questionnaire asked respondents for information on the most established environmental cardiovascular risk factors, e.g., demographic variables, history of CVD, history of diabetes, medication, smoking, and additional lifestyle and dietary habits. Blood was collected in a “semifasting” state, i.e., the participants were instructed to fast for 4 h and to avoid fatty meals earlier during the day. Serum total cholesterol, HDL cholesterol, and triglyceride levels were measured. Whole blood was stored at –20 °C for DNA extraction.

Case-cohort design. We used a case-cohort design [12], in which DNA was genotyped only for stratified random samples—subcohorts—of the FINRISK-92 and –97 cohorts, and for participants experiencing death, coronary event, or stroke during the follow-up periods (Figure 1). In addition, DNA of participants experiencing venous thromboembolic event or reporting history of CVD at the baseline examination was genotyped (<http://www.ktl.fi/publications/morgam/cohorts>); however, venous thromboembolism and baseline CVD were not analyzed as endpoints in our genetic study. Further, participants with prevalent CVD were excluded from our genetic analyses, except when analyzing all-cause mortality, for which all subcohort members and all deaths were included regardless of their CVD history. Information about coronary and stroke events during the follow-up periods was obtained from several sources: specific myocardial infarction and stroke registers (FINAMI [13] and FINSTROKE [14]) complemented with record linkage of the study data with the National Causes of Death Register and the National Hospital Discharge Register. These country-wide, computerized registers cover every death of Finnish citizens and every hospital-

ization in Finland, and thus the coverage of follow-up for symptomatic CVD events was almost 100%. In the National Causes of Death Register International Classification of Diseases (ICD)-9 codes 410–414 and 798, and ICD-10 codes I21–I25, I46, R96, R98 and R99 were taken as coronary deaths. ICD-9 codes 433 (excluding 4330X, 4331X, 4339X) of the Finnish modification of ICD-9; see <http://www.ktl.fi/publications/morgam/cohorts>) and 434 (excluding 4349X), and ICD-10 code I63, were taken as ischemic stroke deaths. In the National Hospital Discharge Register ICD-9 codes 410–411, and ICD-10 codes I21–I22 and I20.0, were taken as nonfatal coronary events; and ICD-9 codes 433 (excluding 4330X, 4331X, 4339X) and 434 (excluding 4349X), and ICD-10 code I63, were taken as nonfatal ischemic stroke events. ICD-9 was used in Finland until 31 December 1995, and ICD-10 after that. The combination of FINAMI register and the National Hospital Discharge Register was also used to identify revascularization procedures (CABG and PTCA) during the follow-up period. The validity of cardiovascular diagnoses in the Finnish Causes of Death Register and the Hospital Discharge Register has been documented in several publications [15–17].

Unique Finnish social security numbers of the participants were compared between the two cohorts, and participants already enrolled in the FINRISK-92 cohort were excluded from the FINRISK-97 cohort.

In the present study, incident CVD and all-cause mortality were analyzed as endpoints. Incident CVD cases included the persons without prevalent CVD at baseline who suffered a fatal or nonfatal ischemic stroke event, or fatal or nonfatal coronary event during the follow-up period (Figure 1 and Table 1). The number of main study endpoints was used as reference to determine the size of the random subcohorts to be selected from the original population cohorts and age was controlled through sampling weights (<http://www.ktl.fi/publications/morgam/manual/contents.htm>) (Figure 1). The selection of the subcohort members was independent of the selection of cases, thus of the 374 male and 154 female incident CVD cases, 72 were also part of the subcohorts. Altogether 610 persons (180 females and 430 males) died during the follow-ups of the two cohorts. Table 2 shows the baseline characteristics of the incident CVD cases and subcohort members free of CVD.

DNA extraction. DNA was extracted from EDTA-treated whole-blood samples using a standard phenol-chloroform method modified from Vandenplas et al. [18]. DNA of 23 FINRISK-92 samples was extracted by a salt-precipitation method.

Markers. To capture the allelic diversity of the 5.69 kb *USF1* gene in the FINRISK-92 and –97 cohorts, we genotyped six htSNPs (Figure 2). *Rs2073658* in intron 7 of the *USF1* gene was the most significantly associated SNP in our previous FCHL study [2]. Additional htSNPs were selected using the SeattleSNPs Variation Discovery Resource [6]. All common (average minor allele frequency > 4%) LD select bins for European descent were covered with our SNP selection.

Rs2516839 and *rs1556259* were included in our family-based study [2], but showed no association with FCHL. *Rs10908821* is located in an intron of the F11 receptor gene (*F11R*), 2.23 kb downstream from *rs2073658*. *Rs2774276* resides in intron 5 of *USF1*, 0.95 kb upstream of *rs2073658*. SNP *rs2774279* results in a synonymous amino acid substitution (874 R/R) in an exon of a predicted gene LOC257106, and is located 2.90 kb upstream from *rs1556259*.

Genotyping. Most of the genotyping was performed using the MassARRAY System (Sequenom, San Diego, California, United States), with a protocol described elsewhere [19]. SNPs *rs2073658*, *rs2516839*, and *rs2774279* were genotyped in the FINRISK-92 study sample using allele-specific primer extension on microarrays with a protocol described elsewhere [19,20]. For MassARRAY System genotyping, 81 low-yield DNA samples of the FINRISK-92 study sample and 19 low-yield DNA samples of the FINRISK-97 study sample were first amplified in replicates using GenomePhi DNA amplification kit (GE Healthcare UK, Buckinghamshire, England) as specified in the manufacturer’s instructions and genotyped as described [19].

The staff in the genotyping laboratory were blinded to sex and disease status of participants. To control for sample mix-ups, all samples were genotyped for sex, and the data were compared to the sex of participants as recorded in the database. Plate-specific duplicates and water samples were used to control for plate-handling errors. Of the samples genotyped, 5% were blinded duplicates. The *USF1* marker genotypes of the blinded duplicate pairs were compared after genotyping and were all found to be consistent. The success rate of genotyping was 98% for SNP *rs2774276* and 97% for all other SNPs.

Haplotype estimations. LD and haplotype block analysis and haplotype population frequency estimations were performed with Haploview version 3.2 [21] using the two FINRISK subcohorts

combined. Subcohort members with over 50% missing genotypes were excluded from the analysis, thus 782 individuals were used in the haplotype estimations.

Data analyses. Allele and genotype frequencies were determined from the data, and deviation from the Hardy-Weinberg equilibrium was tested using Pearson's chi-square. Pearson's chi-square statistic was used to compare allele frequencies between incident CVD cases and subcohorts free of CVD.

A weighted Cox proportional hazards model, modified to account for the case-cohort sampling design, was used for risk analyses with the variance correction based on the published literature using SAS PHREG procedure [12,22,23]. Age represented the time measure in the models. Covariates known to associate with CVD (age, sex, hypertension, smoking, diabetes, total cholesterol to HDL ratio, and BMI) were included in multivariate models as potential confounders. When analyzing CVD, individuals with prevalent CVD at the baseline were excluded from analyses. Analyses were stratified for eastern and western Finland, and the cohort was used as a covariate. Carriership of the minor and major allele was analyzed for each SNP, except for the two SNPs with lowest minor allele frequencies (*rs10908821* and *rs1556259*), for which only the carriership of the minor allele was used in the analysis due to the low number of individuals homozygous for the minor allele.

To determine the relation between the *USF1* htSNPs and several baseline variables, we computed age-, sex-, geographical area-, and cohort-adjusted mean values for SNP allele groups and tested the difference with analyses of covariance implemented in the SAS "PROC GLM" procedure. Before analyses, the variables HDL and triglyceride were log-transformed to correct them to be normally distributed. Mean values of the variables are displayed as crude data on Tables 5 and S1. Individuals deviating ± 4 standard deviations (SD) from the mean were excluded as outliers from the analyses. Further, the analyses were conducted using subcohort members free of CVD at baseline and incident CVD cases.

To evaluate the significance of our findings, we permuted the genotype while retaining the phenotype data within sex/cohort groups, and repeated the same analyses that were performed with the actual dataset, recording the minimum *p*-value observed. We repeated this procedure 1,000 times and took the fifth percentile of these minimum *p*-values as the new multiple-testing corrected threshold for the *p*-values obtained with the original data.

Statistical analyses were carried out using the SAS statistical software (versions 8.2 and 9.1 for Windows) (SAS OnlineDoc, Version 8, 1999; SAS Institute, Cary, North Carolina, United States).

Supporting Information

Table S1. Association between *USF1* SNP Alleles and Lipids, BMI, and Waist-To-Hip Ratio in FINRISK Subcohort Males and Females
Found at DOI: 10.1371/journal.pgen.0020069.st001 (27 KB XLS).

References

- Naukkarinen J, Gentile M, Soro-Paavonen A, Saarela J, Koistinen HA, et al. (2005) *USF1* and dyslipidemias: Converging evidence for a functional intronic variant. *Hum Mol Genet* 14: 2595–2605.
- Pajukanta P, Lilja HE, Sinshheimer JS, Cantor RM, Lusi AJ, et al. (2004) Familial combined hyperlipidemia is associated with upstream transcription factor 1 (*USF1*). *Nat Genet* 36: 371–376.
- Huertas-Vazquez A, Aguilar-Salinas C, Lusi AJ, Cantor RM, Canizales-Quintero S, et al. (2005) Familial combined hyperlipidemia in Mexicans: Association with upstream transcription factor 1 and linkage on chromosome 16q24.1. *Arterioscler Thromb Vasc Biol* 25: 1985–1991.
- Ng MC, Miyake K, So WY, Poon EW, Lam VK, et al. (2005) The linkage and association of the gene encoding upstream stimulatory factor 1 with type 2 diabetes and metabolic syndrome in the Chinese population. *Diabetologia* 48: 2018–2024.
- Vartiainen E, Jousilahti P, Alftan G, Sundvall J, Pietinen P, et al. (2000) Cardiovascular risk factor changes in Finland, 1972–1997. *Int J Epidemiol* 29: 49–56.
- SeattleSNPs. NHLBI Program for Genomic Applications, SeattleSNPs, Seattle, WA. Available: <http://pga.gs.washington.edu>. Accessed 15 October 2005.
- Hoffstedt J, Ryden M, Wahrenberg H, van Harmelen V, Arner P (2005) Upstream transcription factor-1 gene polymorphism is associated with increased adipocyte lipolysis. *J Clin Endocrinol Metab* 90: 5356–5360.
- Coon H, Xin Y, Hopkins PN, Cawthon RM, Hasstedt SJ, et al. (2005) Upstream stimulatory factor 1 associated with familial combined hyperlipidemia, LDL cholesterol, and triglycerides. *Hum Genet* 117: 444–451.
- Putt W, Palmen J, Nicaud V, Tregouet DA, Tahri-Daizadeh N, et al. (2004)

Accession Numbers

The Entrez Gene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>) accession numbers for *USF1* and *FT1R* are 7391 and 50848, respectively. The online Mendelian Inheritance in Man (OMIM) <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=omim> accession numbers for FCHL, metabolic syndrome, and type II diabetes are 144250, 605552, and 125853, respectively.

Acknowledgments

We gratefully acknowledge the Finnish individuals who volunteered to participate in this study. We thank MSc. Minttu Jussila, Mrs. Siv Knaappila, Ms. Minna Levander, Mrs. Anne Nyberg, and Ms. Arja Tapio for their excellent technical assistance. Mr. Juri Ahokas, Mr. Zygimantas Cepaitis, and Mr. Tero Hiekkalinna are thanked for their assistance in the data management. Dr. Juha Karvanen and Mr. Olli Saarela are thanked for their assistance in the data analysis.

Author contributions. K. Komulainen participated in the genotyping of the samples and data analyses. M. Alanne was in charge of the sample management for the study and participated in the data analyses. K. Auro participated in the genotyping of the samples and data analyses. R. Kilpikari and K. Salminen participated in the data analyses. P. Pajukanta provided SNP information and participated in the SNP selection for the study. J. Saarela participated in the optimization of the microarray genotyping and analyzing the results of the genotyping. P. Ellonen participated in the optimization of the microarray genotyping and analyzing the results of the genotyping. S. Kulathinal participated in the study design, the assessment of the quality of the phenotypic data, and data analyses. K. Kuulasmaa participated in the study design and the assessment of the quality of the phenotypic data and data analyses. K. Silander participated in the genotyping of the samples and the whole genome amplification of DNA. V. Salomaa participated in the design of the study, recruitment of the FINRISK patients, assignment of clinical diagnoses, and data analyses. M. Perola participated in the DNA extraction, study design, and data analyses. L. Peltonen is the principal investigator, with the responsibility of overall study design and data analyses. All authors participated in writing the paper.

Funding. This study has been supported by the Center of Excellence in Disease Genetics of the Finnish Academy, Biocentrum Helsinki; awards from the Finnish Academy (grant 53646), Sigrid Juselius Foundation, Finnish Heart Foundation, and Jenny and Antti Wihuri Foundation; NIH grants HL-28481 and HL-70150; and the American Heart Association grant 0430180N. GenomEUtwin-project is supported by the European Commission under the program "Quality of Life and Management of the Living Resources" of 5th Framework Programme (QLG2-CT-2002-01254).

Competing interests. The authors have declared that no competing interests exist. ■

- Variation in *USF1* shows haplotype effects, gene:gene and gene:environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. *Hum Mol Genet* 13: 1587–1597.
- Tunstall-Pedoe Heditor (2003) MONICA Monograph and Multimedia Sourcebook. Geneva: World Health Organization.
 - Evans A, Salomaa V, Kulathinal S, Asplund K, Cambien F, et al. (2005) MORGAM (an international pooling of cardiovascular cohorts). *Int J Epidemiol* 34: 21–27.
 - Prentice R (1986) A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 73: 1–11.
 - Salomaa V, Ketonen M, Koukkunen H, Immonen-Raiha P, Jerkkola T, et al. (2003) Trends in coronary events in Finland during 1983–1997. The FINAMI study. *Eur Heart J* 24: 311–319.
 - Sivenius J, Tuomilehto J, Immonen-Raiha P, Kaarisalo M, Sarti C, et al. (2004) Continuous 15-year decrease in incidence and mortality of stroke in Finland: The FINSTROKE study. *Stroke* 35: 420–425.
 - Rapola JM, Virtamo J, Korhonen P, Haapakoski J, Hartman AM, et al. (1997) Validity of diagnoses of major coronary events in national registers of hospital diagnoses and deaths in Finland. *Eur J Epidemiol* 13: 133–138.
 - Leppala JM, Virtamo J, Heinonen OP (1999) Validation of stroke diagnosis in the National Hospital Discharge Register and the Register of Causes of Death in Finland. *Eur J Epidemiol* 15: 155–160.
 - Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Raiha P, et al. (2005) The validity of the Finnish Hospital Discharge Register and Causes of Death Register data on coronary heart disease. *Eur J Cardiovasc Prev Rehabil* 12: 132–137.
 - Vandenplas S, Wiid I, Grobler-Rabie A, Brebner K, Ricketts M, et al. (1984) Blot hybridisation analysis of genomic DNA. *J Med Genet* 21: 164–172.

19. Silander K, Komulainen K, Ellonen P, Jussila M, Alanne M, et al. (2005) Evaluating whole genome amplification via multiply-primed rolling circle amplification for SNP genotyping of samples with low DNA yield. *Twin Res Hum Genet* 8: 368–375.
20. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, et al. (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* 10: 1031–1042.
21. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
22. Lin D, Wei L (1989) The robust inference for the Cox Proportional Hazards Model. *J Am Stat Assoc* 84: 1074–1079.
23. Barlow WE (1994) Robust variance estimation for the case-cohort design. *Biometrics* 50: 1064–1072.