Human Molecular Genetics, 2007, Vol. 16, No. 23 **2960–2971**
doi:10.1093/hmg/ddm256
Advance Access published on September 12, 2007

Common genetic variation in calpain-10 gene (*CAPN10*) and diabetes risk in a multi-ethnic cohort of American postmenopausal women

Yiqing Song[1], Nai-chieh You[2], Yi-Hsiang Hsu[3], James Sul[2], Lin Wang[4], Lesley Tinker[6], Charles B. Eaton[7] and Simin Liu[1,2,5,*]

[1]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02215, USA, [2]Department of Epidemiology, University of California at Los Angeles (UCLA), School of Public Health, Los Angeles, CA 90095, USA, [3]Molecular and Integrative Physiological Science Program, Department of Environmental Health, Harvard School of Public Health, Boston, MA 02115, USA, [4]Department of Human Genetics and [5]Department of Medicine, David Geffen School of Medicine UCLA, Los Angeles, CA 90095, USA, [6]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA and [7]Center for Primary Care and Prevention, Brown Medical School, Memorial Hospital of Rhode Island, Pawtucket, RI 02860, USA

Received July 30, 2007; Revised and Accepted September 3, 2007

Calpain-10 (*CAPN10*) protein may play a role in glucose metabolisms, pancreatic β-cell insulin secretion and thermogenesis. Emerging evidence has implicated a role of *CAPN10* genetic variants in the risk of type 2 diabetes mellitus (T2DM). Previous association studies, however, have focussed only on several variants initially reported and provided inconsistent results. We conducted a large nested case–control study to comprehensively investigate the associations between common variations in *CAPN10* gene and T2DM risk among postmenopausal women aged 50–79 years from the Women's Health Initiative Observational Study. After comprehensive screening in 244 randomly chosen control samples ($n=61$ for each of four ethnic groups), we selected a total of 12 tagging single nucleotide polymorphisms (tSNPs) spanning 91 kb in *CAPN10* and genotyped them in 1543 diabetes cases and 2132 matched controls (including 968 cases and 968 controls for whites, 366 and 732 for blacks, 152 and 303 for Hispanics and 98 and 195 for Asian/Pacific Islanders). There were no significant associations between any individual tSNP and T2DM, within either the full study sample or any specific ethnic group. Nor was there any evidence of association between common *CAPN10* haplotypes and diabetes risk (global tests for differences in risk were $P=0.31$ for overall common haplotypes, $P=0.44$ for haplotypes in block 1 and $P=0.37$ for haplotypes in block 2). In conclusion, we did not observe any significant associations of the common SNPs or haplotypes across the *CAPN10* gene with diabetes risk in our large and ethnically diverse cohort of postmenopausal women.

INTRODUCTION

Type 2 diabetes mellitus (T2DM) is a genetically heterogeneous and complex metabolic disease. Twin and family studies have shown that T2DM has a strong genetic component (1). Single-gene mutations causing rare monogenic forms of early-onset T2DM (<25 years), such as maturity-onset diabetes of the young, have been identified but they are estimated to account for only a small fraction of all diabetes cases (1–5%) (2). The more common, late-onset form of T2DM is thought to be modulated by multiple loci, each with modest genetic effects. However, specific genetic variants with low penetrance contributing to this polygenic disease remain largely undefined.

*To whom correspondence should be addressed at: Department of Epidemiology, UCLA School of Public Health, PO Box 951772, 650 Charles E. Young Drive South, Los Angeles, CA 90095, USA. Tel: +1 3102065862; Fax: +1 3102066039; Email: siminliu@ucla.edu

In a genome-wide linkage analysis of 330 Mexican-American affected sib pairs (3), calpain-10 gene (*CAPN10*) was first positionally cloned as a putative diabetes-predisposing gene within a T2DM susceptibility region localized to chromosome 2q37.3 (termed *NIDDM1*, non-insulin-dependent diabetes mellitus 1). *CAPN10* encodes a 672 amino acid intracellular protease, which is a member of the calpain family of calcium-dependent intracellular cysteine proteases (4). It has been reported that *CAPN10* is ubiquitously expressed, especially in heart, skeletal muscle, liver and pancreatic islet cells (5). CAPN10 protein may play a role in insulin-mediated glucose metabolisms (6,7), pancreatic β-cell insulin secretion (6,8–10) and thermogenesis (11). There is also some evidence that *CAPN10* genetic variants may affect either *CAPN10* gene expression or mRNA stability (3,6,12). The initial report documented the significant associations of one polymorphism [single nucleotide polymorphism (SNP)-43, rs3792267] and a 'high risk' haplotype combination '1-1-2/1-2-1' inferred from SNP-43 (rs3792267), -19 (3842570) and -63 (rs5030952) (1, major allele and 2, minor allele) with increased risk of T2DM in three independent populations (3). Since then, many replication association studies and hypothesis-driven experiments have generated inconsistent results. As indicated in our systematic assessment of the available data, the lack of association replication may reflect the fact that different populations with genetic and environmental diversity, case ascertainment schemes, control selection, sample sizes, analytic strategies and potential gene–gene and/or gene–environment interactions could produce either false-positive or false-negative results. Of note, almost all previous studies have focussed only on these three or four SNPs and their derived haplotypes highlighted in the initial report (13). As there is substantial heterogeneity of allelic and haplotypic frequencies among different populations with various ethnic or geographic backgrounds (14), these *CAPN10* variants and their derived haplotypes originally defined in a small sample of Mexican-Americans may not be sufficient to capture the genetic variability of *CAPN10* in multiple populations of various ethnic origins. The observed haplotype diversity could also decrease the power of individual studies and was thus likely to generate false-negative results. Most recently, a region of extensive linkage disequilibrium (LD) patterns was observed to be within the *CAPN10* locus (15), which makes it unlikely that the putative genetic variant responsible for the *CAPN10* genotype–T2DM association lies outside the *CAPN10* genomic region. However, few studies have systematically explored the underlying common genetic variation in *CAPN10* across diverse populations, especially when a body of information from the publicly accessible SNP databases, such as the HapMap database of the International HapMap Project (16), has provided the foundation to conduct a comprehensive association study of common genetic variation.

We therefore employed a comprehensive haplotype approach to assess the association between common variation in *CAPN10* and T2DM in a large case–control study nested in the Women's Health Initiative Observational Study (WHI-OS), an ethnically diverse cohort of postmenopausal women aged over 50 years including whites, blacks, Hispanics and Asian/Pacific Islanders. In particular, we also set out to replicate the associations of those specific SNPs and haplotypes in the *CAPN10* with T2DM highlighted in previous reports.

RESULTS

Estimation of MAFs and LD structures of 12 *CAPN10* gene tSNPs among controls

None of the SNPs showed statistically significant deviation from Hardy–Weinberg equilibrium (HWE) among controls (Table 1 and Fig. 1). With the exception of rs1138847 [5′-untranslated region (UTR)] and rs4676399 (3′-UTR), 10 of 12 tSNPs (one every ~3.1 kb) were within a 31.3 kb of *CAPN10* genomic region. The allelic frequencies of all tSNPs varied significantly by ethnic groups (all $P<0.0001$ except for rs4676399 at $P<0.003$). In particular, allele frequencies in blacks and Asian/Pacific Islanders were significantly different from those of other ethnic groups for most SNPs.

Figure 2 showed the LD structure and haplotype blocks based on these 12 tSNPs stratified by ethnic groups (for all controls). With the exception of rs1138847 and rs4676399, strong, pairwise LD was observed between all other 10 SNPs within *CAPN10*. As expected, the lengths and locations of defined haplotype blocks were slightly different among ethnic groups, partly reflecting the differences in allele frequencies between groups. Overall, two blocks with slightly different boundaries across four ethnic groups could be clearly defined. The LD pattern within block 2 was similar among whites, Hispanics and Asian/Pacific Islanders. For blacks, the sizes of both blocks were modestly reduced. There was no evidence for high LD of these two blocks with either of two SNPs in the 5′ and 3′ outside the *CAPN10* genomic region. This finding was consistent with a recent study reporting a high LD region within the *CAPN10* locus (17). Notably, the SNP44 and SNP43 loci are 12 bp apart and yet not in complete LD ($D'=0.94–1.00$; $r^2=0.007–0.058$). SNP-44 is in almost perfect LD with rs7607759 (Thr504Ala, a common non-synonymous SNP in exon 10) in all groups ($D'=0.988–1.00$; $r^2=0.934–1.00$).

Single-marker association analysis

As showed in Table 2, no significant associations between all tSNPs and diabetes risk were found in American blacks, Hispanics and Asians/Pacific Islanders. A marginally significant association was found for rs5030952 (SNP-63) in an additive genetic model among American whites ($P=0.035$). However, the false discovery rate (FDR) q-value was larger than 0.05 after adjusted for multiple testing. Tests of heterogeneity of the associations across ethnic groups were not statistically significant. We observed no evidence of significant associations between any of tSNPs and diabetes risk in all groups combined (Table 2).

Haplotype association analysis

We used 12 tSNPs to infer common haplotype within each block with a frequency of ≥5% in the combined data of all controls, regardless of ethnicity groups. We observed four common haplotypes in block 1 and five common haplotypes

Table 1. Characterization of 12 haplotype tagging SNPs (tSNPs) chosen in CAPN10 genomic region

dbSNP ID[a]	UC-SNP ID[b]	Location	Allele[c]	Minor allele frequency (%)					*P*-value for heterogeneity across ethnicity[d]
				Whites (n=968)	Blacks (n=732)	Hispanics (n=303)	Asian/Pacific Islanders (n=195)	Pooled (n=2198)	
rs1138847		5' Flanking	A/G	24.9	25.4	23.8	3.37	23.3	<0.0001
rs2953173		5' Flanking	T/A	29.7	21.6	22.6	13.3	24.7	<0.0001
rs2975760	SNP44	Intron 3	T/C	15.5	6.41	11.4	9.64	11.4	<0.0001
rs3792267	SNP-43	Intron 3	G/A	26.1	14.3	25.4	5.76	20.4	<0.0001
rs3842570	SNP19	Intron 6	INS/DEL	39.6	66.2	45.9	40.6	49.9	<0.0001
rs7607759	SNP110	Exon 10	A/G (Thr 504 Ala)	15.5	6.73	11.5	9.06	11.4	<0.0001
rs3749166	SNP48	Exon 11	A/G (Ala 620 Ala)	37.9	67.4	42.9	38.3	48.9	<0.0001
rs3749168		Intron 13	A/G	25.3	12.5	17.4	9.64	18.6	<0.0001
rs5030952	SNP63	Intron 13	C/T	7.70	43.4	20.5	24.9	23.1	<0.0001
rs4676348		Intron 13	A/G	37.9	59.4	42.8	40.4	46.2	<0.0001
rs7571442 (or rs1509)		3' Flanking	A/C	28.0	13.4	25.5	55.5	24.7	<0.0001
rs4676399		3' Flanking	C/G	11.3	15.4	12.7	8.84	12.7	0.0026

[a]From the NCBI dbSNP database.
[b]From the original report by Horikawa *et al*. (3).
[c]Major/minor allele.
[d]*P*-values were estimated by a chi-square test (df=3) for genotype distribution across the four ethnic groups.

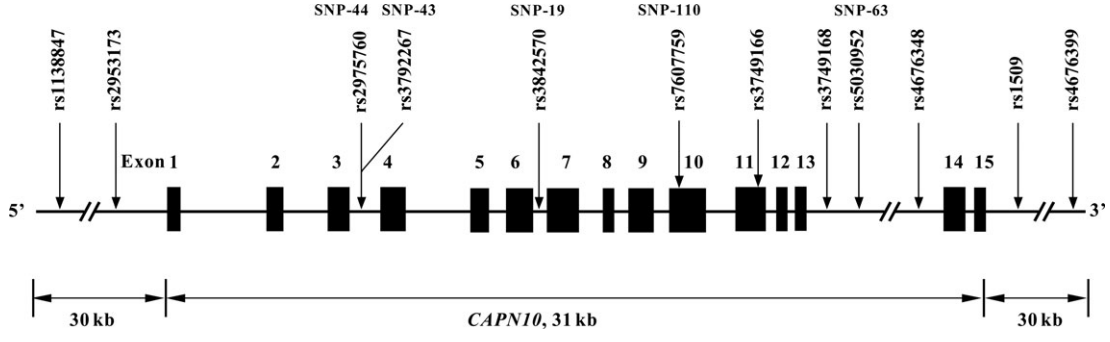

Figure 1. Schematic presentation of the position of 12 tSNPs spanning the *CAPN10* region. These markers within CAPN10 genomic region included one putative promoter SNP (rs2953173), one SNP in the 3′-UTR (rs1509), one missense SNP (rs7607759, Thr504Ala), one synonymous SNP (rs3749166, Ala620Ala) and six intronic SNPs. On the basis of physical positions, two SNPs (rs1138847 and rs4676399) are outside of the *CAPN10* genomic region without high LD with other SNPs; rs1138847 is located 24 kb upstream and rs4676399 is located 32 kb downstream of the *CAPN10* coding region.

in block 2 (Table 3). We first performed global tests for differences in the overall haplotype frequency between cases and controls and observed no significant haplotype effects in blocks 1 (P=0.44) and 2 (P=0.37). Within block 1 (rs2953173/rs2975760/rs3792267), we observed marginal haplotype effects; compared with non-carriers for 1-1-2, we observed increased risk for carriers of the 1-1-2 [odds ratio (OR)=1.82, 95% confidence interval (CI): 1.00–3.29; P=0.048] and lower risk of T2DM for carriers of the 2-1-1 (OR=0.67, 95% CI: 0.48–0.93; P=0.016). In block 2 (rs3749166/rs3749168/rs5030952/rs4676348), the only SNP63 (rs5030952) rare allele T-bearing haplotype 2-1-2-2 showed significant positive association with T2DM risk in whites (OR=1.50, 95% CI: 1.02–2.20; P=0.039) and all ethnic groups combined (OR=1.33, 95% CI: 1.05–1.69; P=0.0195). However, these associations were weak and did not reach nominal significance after we did an FDR correction for multiple testing.

We also reconstructed haplotypes on the basis of the haplotype block structure in each ethnic group (Table 4). The global likelihood ratio tests of haplotype associations in each block were not statistically significant. Significant ORs were only observed in whites. Besides the haplotype 2-1-1 in block 1 (same as that in Table 3), we also observed increased T2DM risk for all carriers of the SNP-63 T-bearing haplotype 2-1-2-2-1 in block 2 (OR=1.54, 95% CI: 1.02–2.26; P=0.030). Their statistical significance could not retain after taking into account multiple testing.

Replication of previously studied haplotype and diplotype associations

On the basis of SNP-43, -19 and -63, the haplotype 1-2-1 was most common in whites (36%), Hispanics (31%) and Asians/Pacific Islanders (54%), but only at a minor frequency in blacks (18%). The most common haplotype in blacks was 1-1-2 (41%). We found no evidence of any significant associations of haplotypes defined by SNP-43, -19 and -63 with T2DM (Table 5). When we inferred haplotype from SNP-44, -43, -19 and -63, the major haplotypes were 1-1-1-2 in blacks (40%) and 1-1-2-1 in other groups at 30–54%. There was evidence to suggest an association between the rare

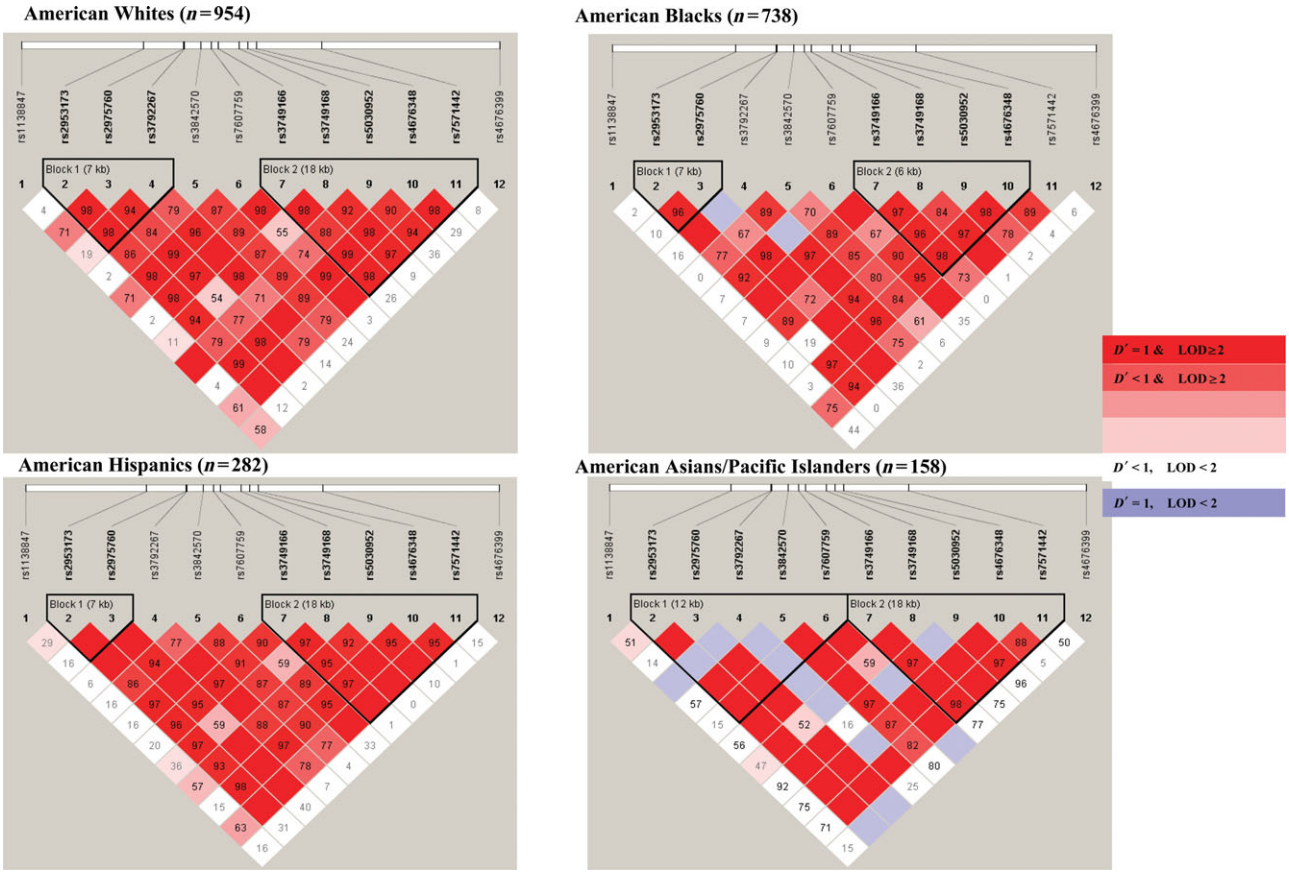

Figure 2. Haploview plot defining LD structures between the 12 tSNPs within or near *CAPN10* among all controls from four ethnic groups (American whites, blacks, Hispanics and Asian/Pacific Islanders). The relative physical position of each SNP is given in the upper diagram. Each diamond for each SNP combination indicates the pairwise LD between all tSNPs, with red indicating strong LD ($D' > 0.8$) and a logarithm of odds score of >2.0. LD strength between the chosen SNPs is determined by the 90% confidence limits of D' statistic.

Table 2. Single-SNP associations between each tSNP spanning the *CAPN10* genomic region and T2DM risk

SNP ID	Adjusted OR (95% CI)[a]					*P*-value for heterogeneity[b]
	White (968/968)	Black (366/732)	Hispanics (152/303)	Asian/Pacific Islanders (98/195)	All groups combined (1584/2198)	
rs1138847	1.02 (0.81–1.29)	0.95 (0.70–1.28)	0.99 (0.61–1.62)	0.17 (0.02–1.18)	0.99 (0.84–1.17)	0.24
rs2953173	0.89 (0.72–1.10)	1.00 (0.75–1.34)	1.09 (0.68–1.74)	0.60 (0.24–1.53)	0.94 (0.80–1.09)	0.75
rs2975760 (SNP44)	1.10 (0.84–1.43)	0.90 (0.51–1.57)	1.13 (0.58–2.19)	0.69 (0.25–1.94)	1.04 (0.85–1.29)	0.64
rs3792267 (SNP-43)	1.05 (0.84–1.30)	1.10 (0.77–1.58)	1.53 (0.94–2.51)	2.08 (0.60–7.21)	1.12 (0.95–1.32)	0.70
rs3842570 (SNP19)	0.94 (0.77–1.15)	1.11 (0.85–1.44)	0.99 (0.63–1.55)	1.15 (0.61–2.17)	0.99 (0.86–1.15)	0.57
rs7607759 (SNP110)	1.06 (0.80–1.39)	0.76 (0.42–1.36)	1.12 (0.57–2.22)	0.76 (0.27–2.11)	0.99 (0.80–1.23)	0.31
rs3749166 (SNP48)	1.02 (0.84–1.24)	0.91 (0.70–1.17)	1.09 (0.72–1.67)	0.91 (0.49–1.69)	0.99 (0.86–1.13)	0.80
rs3749168	0.87 (0.70–1.09)	1.17 (0.80–1.70)	1.10 (0.66–1.82)	0.37 (0.13–1.09)	0.92 (0.78–1.09)	0.29
rs5030952 (SNP63)	1.44 (1.03–2.02)[c]	1.01 (0.78–1.32)	1.14 (0.65–2.01)	1.09 (0.54–2.18)	1.16 (0.97–1.40)	0.21
rs4676348	1.00 (0.82–1.21)	0.98 (0.76–1.28)	1.11 (0.72–1.71)	0.74 (0.40–1.38)	1.00 (0.87–1.15)	0.77
rs7571442	1.01 (0.82–1.24)	1.08 (0.77–1.53)	0.89 (0.54–1.46)	0.89 (0.49–1.63)	0.99 (0.85–1.16)	0.48
rs4676399	1.03 (0.75–1.41)	1.06 (0.76–1.50)	0.93 (0.54–1.61)	0.50 (0.17–1.47)	1.03 (0.84–1.26)	0.57

[a]Odds ratio (OR) of each single SNP was calculated for additive genetic effect model; ORs were estimated using conditional logistic regression models adjusted for matching factors (e.g. age, clinical center, time of blood draw and ethnicity), BMI, cigarette smoking, alcohol intake, HRT, family history of diabetes and physical activity. The numbers of participants (cases/controls) were included in the parenthesis.
[b]*P*-value was estimated based on a log-likelihood ratio test for interaction between each genotype and ethnicity on T2DM risk.
[c]$P = 0.0352$ without adjusted for multiple testing; FDR q-value = 0.4224.

allele C in the SNP-44 locus alone or in combination with the 111 haplotype with risk of T2DM (18), whereas our study failed to replicate this finding with SNP-44. There was a suggestion that the haplotype 2-2-1 at SNP-43, -19 and -63 or with the common allele of SNP-44 (1-2-2-1) appeared to be associated with T2DM risk only in American Hispanics,

Table 3. Haplotype-based associations between *CAPN10* common haplotypes and T2DM risk

Haplotype[a]	OR (95% CI)[b]						*P*-values for global testing
					All groups combined		
	White (968/968)	Black (366/732)	Hispanics (152/303)	Asian/Pacific Islanders (98/195)	Carriers versus all others	Homozygous versus all non-carriers	
Block 1							
rs2953173(T/A)-rs2975760(T/C)-rs3792267(G/A)							
1-1-1	1.25 (0.93–1.69)	1.08 (0.63–1.83)	0.65 (0.32–1.31)	0.30 (0.03–3.38)	1.07 (0.85–1.36)	0.97 (0.64–1.47)	0.44
1-1-2	0.91 (0.67–1.19)	1.12 (0.77–1.63)	1.82 (1.00–3.29)	1.74 (0.51–5.94)	1.05 (0.86–1.28)	1.21 (0.66–2.19)	
2-1-1	0.67 (0.48–0.93)[c]	1.10 (0.75–1.61)	1.17 (0.59–2.33)	0.54 (0.07–4.50)	0.85 (0.68–1.07)	0.70 (0.34–1.44)	
2-2-1	1.01 (0.74–1.36)	0.90 (0.48–1.67)	1.14 (0.56–2.30)	0.74 (0.27–2.02)	1.00 (0.79–1.26)	1.37 (0.53–3.54)	
Block 2							
rs3749166(A/G)-rs3749168(A/G)-rs5030952(C/T)-rs4676348(A/G)							
1-1-1-1	0.90 (0.62–1.30)	1.51 (0.80–1.66)	0.98 (0.42–2.29)	0.99 (0.33–2.99)	0.98 (0.78–1.25)	1.15 (0.76–1.75)	0.37
2-1-2-2	1.50 (1.02–2.20)[d]	1.24 (0.82–1.85)	1.24 (0.64–2.41)	1.14 (0.49–2.62)	1.33 (1.05–1.69)[e]	1.50 (1.02–2.21)[f]	
2-2-1-2	0.83 (0.63–1.10)	1.27 (0.82–1.95)	1.14 (0.59–2.20)	0.49 (0.17–1.39)	0.92 (0.75–1.14)	1.05 (0.73–1.51)	
2-1-1-2	0.97 (0.63–1.49)	0.70 (0.38–1.27)	0.59 (0.15–1.59)	3.23 (0.58–18.2)	0.96 (0.70–1.32)	0.78 (0.08–7.21)	
2-1-1-1	—	0.69 (0.40–1.17)	0.55 (0.07–4.20)	—	0.62 (0.37–1.05)	0.72 (0.38–1.37)	

[a]Haplotypes observed with ≥0.05 frequency in at least one ethnic group in the multiethnic panel (1, major allele; 2, minor allele).
[b]ORs for each ethnic group was calculated for carriers versus all non-carriers; conditional logistic regression models were adjusted for matching factors (age, clinical center, time of blood draw and ethnicity) and other confounders including BMI, HRT, alcohol consumption, cigarette smoking, physical activity and family history of diabetes; the numbers of participants (cases/controls) were included in the parenthesis.
[c]$P=0.016$ without adjusted for multiple testing; FDR q-value=0.144.
[d]$P=0.039$ without adjusted for multiple testing; FDR q-value=0.176.
[e]$P=0.020$ without adjusted for multiple testing; FDR q-value=0.351.
[f]$P=0.040$ without adjusted for multiple testing; FDR q-value=0.360.

but their *P*-values were approximately marginal (0.042 for carriers of 2-2-1 and 1-2-2-1 versus non-carriers, respectively) (Table 5).

To replicate the associations of diplotypes with T2DM previously reported, we inferred the common diplotypes based on SNP-43, -19 and -63 with and without SNP-44 (Table 6). Of the diplotypes based on three SNPs, 10 genotypes with frequency >1% were observed in whites, blacks and Hispanics and 2-2-1/2-2-1 was not shown in Asians/Pacific Islanders. When SNP-44 was included, we observed 15 diplotypes with frequency >1% in whites, blacks and Hispanics. Similarly, 1-2-2-1/1-2-2-1 was not found in Asians/Pacific Islanders. Of 108 tests for the crude comparisons of diplotype frequencies between cases and controls, only seven tests showed notably different diplotype frequencies between cases and controls ($P<0.05$). These results could be false positives because they failed to remain statistically significant after adjusting for multiple comparisons. Overall, there were not statistically significant associations for these diplotypes and T2DM.

DISCUSSION

On the basis of 1543 T2DM cases and 2132 matched controls from a multi-ethnic cohort of American postmenopausal women, our study showed little evidence for the role of common genetic variants in *CAPN10* gene in T2DM risk. None of the SNPs originally featured as potential T2DM loci showed any significant association with T2DM. Moreover, our haplotype-based analyses did not reveal any significant findings. Such null results were consistent across four ethnic groups including whites, blacks, Hispanics and Asian/Pacific Islanders. The lack of replication in our large multiethnic case–control study strongly suggests that common genetic variants in *CAPN10*, including those highlighted in the original reports, may not confer a susceptibility to T2DM.

Given the large number of associations tested, any observed significant associations with any single marker or haplotype should be interpreted with caution. Although we did observe some significant associations, such findings may simply be explained by multiple testing (all corrected $P>0.05$) and are thus more likely to represent false-positive associations. First, the *P*-values for significant SNP and haplotype associations were above 5% after performing either an FDR or a Bonferroni correction for multiple testing, although the methods are quite conservative. Secondly, the significant results did not replicate any previously reported associations. Thirdly, the observed significant results were not robust to statistical model selection and adjustment for confounding. Fourth, the global tests did not support a significant association of the *CAPN10* haplotypes with T2DM.

Despite substantial heterogeneity in frequencies of the alleles and haplotypes in *CAPN10* gene among various ethnic groups, almost all previous studies did not determine the haplotype structure of *CAPN10* in their own study samples, but rather genotyping the SNPs (SNP-44, -43, -19, -63) originally identified in Mexican-Americans (3). As noted in our previous meta-analysis (13), the lack of association in previous association studies could be related to genetic heterogeneity because these four intronic *CAPN10* variants may not be sufficient to capture the vast majority of genetic variability of *CAPN10* in diverse populations. Emerging evidence has indicated that an optimal subset of most

Table 4. Associations between ethnic-specific haplotypes of *CAPN10* and T2DM risk

Haplotype[a]	Haplotype frequency		OR (95% CI) [b]		*P*-values for global testing
	Cases	Controls	All carriers versus all others	Homozygous versus all non-carriers	
Whites	*n*=968	*n*=968			
Block 1	*rs2953173(T/A)-rs2975760(T/C)-rs3792267(G/A)*				
1-1-1	0.43	0.44	1.25 (0.93–1.69)	1.18 (0.64–2.17)	0.20
1-1-2	0.28	0.26	0.90 (0.69–1.19)	0.87 (0.52–1.46)	
2-2-1	0.17	0.16	1.01 (0.74–1.36)	3.04 (0.90–10.3)	
2-1-1	0.13	0.14	0.67 (0.48–0.93)[c]	0.54 (0.31–0.97)[d]	
Block 2	*rs3749166(A/G)-rs3749168(A/G)-rs5030952(C/T)-rs4676348(A/G)-rs7571442(A/C)*				
1-1-1-1-1	0.33	0.34	0.90 (0.68–1.19)	1.12 (0.59–2.11)	0.27
1-1-1-1-2	0.27	0.27	1.00 (0.76–1.31)	0.96 (0.54–1.71)	
2-2-1-2-1	0.24	0.25	0.85 (0.64–1.12)	0.68 (0.32–1.42)	
2-1-2-2-1	0.09	0.07	1.54 (1.04–2.26)[e]	2.68 (0.46–15.8)	
2-1-1-2-1	0.06	0.06	0.96 (0.62–1.49)	1.57 (0.73–3.37)	
Blacks	*n*=366	*n*=732			
Block 1	*rs2953173(T/A)-rs2975760(T/C)*				
1-1	0.78	0.78	1.46 (0.66–3.25)	1.84 (0.75–4.55)	0.75
2-1	0.17	0.15	1.10 (0.75–1.61)	0.68 (0.23–1.99)	
2-2	0.05	0.06	0.97 (0.53–1.78)	—	
Block 2	*rs3749166(A/G)-rs3749168(A/G)-rs5030952(C/T)-rs4676348(A/G)*				
2-1-2-2	0.40	0.42	1.23 (0.82–1.84)	1.36 (0.71–2.63)	0.41
1-1-1-1	0.34	0.32	1.15 (0.80–1.65)	1.34 (0.59–3.04)	
2-2-1-2	0.12	0.11	1.28 (0.83–1.97)	1.37 (0.78–2.42)	
2-1-1-1	0.07	0.08	0.69 (0.40–1.17)	0.91 (0.28–2.96)	
2-1-1-2	0.06	0.05	0.70 (0.39–1.28)	0.14 (0.002–8.81)	
Hispanics	*n*=152	*n*=303			
Block 1	*rs2953173(T/A)-rs2975760(T/C)*				
1-1	0.77	0.77	0.87 (0.25–3.05)	0.87 (0.16–4.75)	0.71
2-2	0.13	0.11	1.14 (0.56–2.30)	0.25 (0.01–5.94)	
2-1	0.10	0.11	1.17 (0.59–2.33)	0.16 (0.008–2.95)	
Block 2	*rs3749166(A/G)-rs3749168(A/G)-rs5030952(C/T)-rs4676348(A/G)-rs7571442(A/C)*				
1-1-1-1-1	0.38	0.31	1.22 (0.65–2.29)	2.60 (0.18–37.5)	0.88
1-1-1-1-2	0.20	0.25	0.87 (0.43–1.76)	1.03 (0.34–3.10)	
2-1-2-2-1	0.18	0.19	1.10 (0.56–2.18)	1.75 (0.41–7.36)	
2-2-1-2-1	0.17	0.17	1.10 (0.57–2.12)	1.88 (0.28–12.7)	
2-1-1-2-1	0.04	0.05		—	
Asians/Pacific Islanders	*n*=98	*n*=195			
Block 1	*rs2953173(T/A)-rs2975760(T/C)-rs3792267(G/A)-rs3842570(insertion/deletion)-rs7607759(A/G)*				
1-1-1-1-1	0.57	0.53	0.76 (0.27–2.15)	—	0.68
1-1-1-2-1	0.28	0.28	1.19 (0.49–2.91)	2.00 (0.16–5.74)	
2-2-1-2-2	0.07	0.10	0.88 (0.28–2.75)	—	
1-1-2-1-1	0.07	0.06	1.90 (0.47–7.73)	—	
2-1-1-2-1	0.02	0.04	0.13 (0.003–5.31)	—	
Block 2	*rs3749166(A/G)-rs3749168(A/G)-rs5030952(C/T)-rs4676348(A/G)-rs7571442(A/C)*				
1-1-1-1-2	0.58	0.52	0.98 (0.37–2.65)	—	0.32
2-1-2-2-1	0.25	0.24	1.14 (0.49–2.67)	7.40 (0.36–150)	
2-2-1-2-1	0.06	0.10	0.50 (0.17–1.41)	—	
1-1-1-1-1	0.08	0.07	1.88 (0.53–6.62)	—	
2-1-1-2-1	0.03	0.04	3.65 (0.63-21.3)	—	

[a]Haplotypes with ≥0.05 frequency within each block were inferred in each of four ethnic groups (1, major allele; 2, minor allele).
[b]ORs for each haplotype was calculated using conditional logistic regression models with adjustment for matching factors (age, clinical center and time of blood draw) and other confounders including BMI, HRT, alcohol consumption, cigarette smoking, physical activity and family history of diabetes.
[c]*P*=0.0168 without adjusted for multiple testing; FDR *q*-value=0.588.
[d]*P*=0.030 without adjusted for multiple testing; FDR *q*-value=0.700.
[e]*P*=0.040 without adjusted for multiple testing; FDR *q*-value=0.700.

informative polymorphisms, defined as tSNPs based on the LD information, is more likely to capture most of common genetic variation. The haplotype-block information is believed to greatly increase power for detection of untyped common variants and improve localization of untyped variants associated with the trait of interest. Empirical data have also demonstrated that tSNPs selected from the HapMap population samples can effectively capture common variation and provide good power to detect an association under a disease model of modest risk in many other independent samples (19). On the basis of the HapMap database, our comprehensive association analysis of informative SNPs [minor allele frequency (MAF)≥5%] based on LD patterns in each ethnic group provides powerful null evidence against a main effect association between the overall risk of T2DM and variants in *CAPN10* that are common among four American ethnic

Table 5. Association of *CAPN10* haplotype previously reported and T2DM risk

Haplotypes[a]	Whites			Blacks			Hispanics			Asian/Pacific Islanders			All groups combined [OR (95% CI) [b]]
	Frequency		OR (95% CI)[b]	Frequency		OR (95% CI)[b]	Frequency		OR (95% CI)[b]	Frequency		OR (95% CI)[b]	
	Cases (968)	Controls (968)		Cases (366)	Controls (732)		Cases (152)	Controls (303)		Cases (98)	Controls (195)		
SNP-43-19-63													
1-1-1	0.32	0.31	1.00 (0.80–1.25)	0.24	0.25	0.91 (0.68–1.20)	0.24	0.24	0.87 (0.53–1.42)	0.09	0.16	0.62 (0.26–1.47)	0.95 (0.82–1.11)
1-1-2	0.09	0.07	1.36 (0.96–1.93)	0.39	0.41	1.06 (0.81–1.38)	0.18	0.19	1.34 (0.76–2.38)	0.25	0.24	1.14 (0.57–2.26)	1.18 (0.98–1.42)
1-2-1	0.31	0.36	0.82 (0.66–1.02)	0.19	0.18	1.07 (0.79–1.45)	0.27	0.31	0.67 (0.42–1.07)	0.59	0.54	1.02 (0.57–1.84)	0.87 (0.75–1.01)
2-2-1	0.26	0.24	1.10 (0.88–1.38)	0.15	0.13	1.14 (0.79–1.65)	0.30	0.22	1.71 (1.02–2.88)[c]	0.06	0.05	2.09 (0.60–7.29)	1.16 (0.98–1.38)
SNP-44-43-19-63													
1-1-1-1	0.16	0.16	0.83 (0.63–1.10)	0.20	0.19	0.96 (0.71–1.30)	0.12	0.13	0.78 (0.42–1.45)	0.02	0.07	0.48 (0.12–1.94)	0.89 (0.74–1.07)
1-1-1-2	0.09	0.07	1.36 (0.96–1.94)	0.39	0.40	1.07 (0.82–1.39)	0.17	0.19	1.30 (0.74–2.30)	0.25	0.24	1.13 (0.57–2.25)	1.19 (0.98–1.43)
1-1-2-1	0.30	0.35	0.85 (0.68–1.05)	0.19	0.17	1.05 (0.77–1.43)	0.26	0.30	0.64 (0.40–1.03)	0.58	0.54	1.05 (0.59–1.86)	0.88 (0.75–1.02)
1-2-2-1	0.26	0.24	1.11 (0.89–1.39)	0.15	0.13	1.14 (0.78–1.65)	0.30	0.22	1.73 (1.03–2.91)[d]	0.06	0.05	2.12 (0.60–7.49)	1.17 (0.98–1.39)
2-1-1-1	0.16	0.14	1.16 (0.88–1.53)	0.04	0.06	0.74 (0.40–1.38)	0.13	0.11	1.10 (0.55–2.19)	0.06	0.10	0.70 (0.25–2.01)	1.04 (0.84–1.30)

[a]Haplotypes were defined based on UC-SNP 43 (rs3792267, G/A), 19 (rs3842570, insertion/deletion), 63 (rs5030952, C/T) with or without SNP-44 (rs2975760, T/C); 1, major allele; 2, minor allele.
[b]OR for each diplotype was calculated for carriers versus all non-carriers; conditional logistic regression models were adjusted for matching factors (age, clinical center, time of blood draw and ethnicity) and other confounders including BMI, HRT, alcohol consumption, cigarette smoking, physical activity and family history of diabetes).
[c]$P=0.0424$ without adjusted for multiple testing; FDR $q=0.1908$.
[d]$P=0.0420$ without adjusted for multiple testing; FDR $q=0.1908$.

groups. In addition, our exploration of LD patterns and extent across *CAPN10* locus provided evidence to support the recent findings that an extensive LD region existed within the *CAPN10* gene (15) and made it unlikely that the putative genetic variants outside the *CAPN10* gene may explain the putative *CAPN10* genotype–T2DM association.

Our study focussed primarily on testing the common variant-common disease hypothesis. Indeed, the vast majority of common variations (85–95%) was observed to be shared between populations living in different continents (20). In the present study, we observed most of the common *CAPN10* haplotypes to be shared among whites, Hispanics and Asian/Pacific Islanders. In line with recent report from a population-based study, neither all the intronic variants previously reported in the initial study for haplotype definition nor functional variants (regulatory or coding polymorphisms) were significantly associated with T2DM in our four ethnic groups. Our strategies for common variant selection based on the available genetic information, although comprehensive and cost-effective, may be unable to identify any rare but potentially functional variants with relatively low frequencies ($<5\%$) in any specific ethnic group that were also not in high LD with any SNPs chosen in our study. Nevertheless, a rare polymorphism may not be a major contributor to T2DM with a high prevalence in the general population, given its low frequency and most likely modest genetic effects. Further large-scale studies with resequencing efforts are warranted to help identify any rare but causal variants underlying T2DM.

The lack of association for *CAPN10* variants and T2DM may also be most likely because of insufficient power to detect changes in phenotypes. It is clear that the size of effects of each *CAPN10* variant or haplotype was modest (1.15–1.25), requiring very large sample sizes to attain sufficient power for detection. We may not have enough power ($<80\%$) to detect significant signal, if estimated OR less than 1.25 for any single SNP or haplotype with a frequency of $<10\%$. Specifically, our study was well powered ($\geq 80\%$) to detect a relative risk of ≥ 1.15 for the combined samples and a relative risk of ≥ 1.25 for whites and blacks for a risk allele with frequencies of 10 and 70%. However, we did not have adequate statistical power ($<80\%$) if the relative risk is less than 1.25 in Hispanics and Asians/Pacific Islanders. To increase statistical power to detect modest effect sizes, meta-analysis has been increasingly applied in genetic association studies to synthesize the totality of evidence across available studies using either study-level data or individual data. However, we do not believe that sole reliance on the increased power in any meta-analysis could provide a true genetic effect size, because many inherent features contributing to between-study heterogeneity, such as the quality of study design and analysis, various ethnic backgrounds and publication bias, might influence the pooled estimates and complicate the interpretations. In this regard, it is essential to replicate the genetic association in carefully conducted, large-scale association studies.

As a complex metabolic disease, the pathogenesis of T2DM is multi-factorial involving both environmental and genetic factors. It seems plausible to speculate that the main effects of genetic factors may be modest and can be masked in the

Table 6. Association of *CAPN10* diplotypes previously reported and T2DM risk

Diplotypes[a]	Whites Frequency Cases (968)	Controls (968)	*P*-value[b]	Blacks Frequency Cases (366)	Controls (732)	*P*-value[b]	Hispanics Frequency Cases (152)	Controls (303)	*P*-value[b]	Asian/Pacific Islanders Frequency Cases (98)	Controls (195)	*P*-value[b]	All groups combined [OR (95% CI)[c]]
SNP-43-19-63													
111/111	0.10	0.08	0.10	0.07	0.07	0.81	0.05	0.07	0.41	0.01	0.04	0.22	0.96 (0.68–1.35)
111/112	0.06	0.04	0.06	0.18	0.18	0.99	0.09	0.05	0.20	0.05	0.05	0.94	1.36 (0.97–1.91)
111/121	0.21	0.25	0.04	0.09	0.10	0.63	0.16	0.17	0.88	0.13	0.19	0.23	0.83 (0.65–1.06)
112/112	0.01	0.01	0.42	0.15	0.16	0.51	0.02	0.05	0.17	0.09	0.05	0.31	1.12 (0.72–1.72)
112/121	0.06	0.05	0.63	0.16	0.17	0.59	0.10	0.11	0.76	0.27	0.27	0.91	1.08 (0.79–1.49)
121/121	0.09	0.11	0.16	0.04	0.03	0.52	0.05	0.10	0.07	0.30	0.27	0.61	0.82 (0.59–1.13)
111/221	0.14	0.14	0.84	0.07	0.06	0.39	0.11	0.10	0.72	0.03	0.01	0.45	0.97 (0.71–1.34)
112/221	0.05	0.04	0.53	0.13	0.12	0.75	0.14	0.12	0.58	—	0.05	—	1.11 (0.78–1.57)
121/221	0.17	0.19	0.33	0.07	0.04	0.01	0.17	0.12	0.19	0.11	0.05	0.06	1.04 (0.79–1.38)
221/221	0.09	0.06	0.02	0.01	0.02	0.27	0.09	0.05	0.20	—	—	—	1.51 (0.94–2.41)
All others	0.02	0.03	0.58	0.04	0.05	0.33	0.02	0.05	0.21	0.01	0.01	0.59	0.57 (0.31–1.05)
SNP-44-43-19-63													
1111/1111	0.03	0.02	0.45	0.06	0.05	0.39	0.01	0.03	0.37	—	0.01	—	0.93 (0.55–1.56)
1111/1112	0.04	0.03	0.18	0.14	0.14	0.96	0.04	0.04	0.87	0.03	0.02	0.71	1.38 (0.93–2.05)
1111/1121	0.10	0.12	0.26	0.06	0.07	0.49	0.09	0.07	0.69	0.05	0.08	0.34	0.75 (0.54–1.05)
1111/1221	0.06	0.08	0.12	0.06	0.04	0.17	0.04	0.04	0.85	—	0.01	—	0.81 (0.54–1.23)
1111/2111	0.04	0.04	0.86	0.01	0.02	0.22	0.03	0.03	0.85	0.01	0.01	0.97	0.70 (0.40–1.22)
1112/1112	0.01	0.01	0.57	0.15	0.16	0.51	0.02	0.05	0.17	0.09	0.05	0.31	1.08 (0.70–1.67)
1112/1121	0.06	0.06	0.94	0.17	0.17	0.81	0.09	0.12	0.39	0.27	0.27	0.99	1.11 (0.81–1.51)
1112/1221	0.05	0.04	0.60	0.13	0.12	0.65	0.14	0.11	0.51	—	0.05	—	1.14 (0.80–1.63)
1112/2111	0.03	0.02	0.19	0.05	0.04	0.86	0.05	0.01	0.03	0.03	0.04	0.51	1.22 (0.71–2.12)
1121/1121	0.09	0.10	0.26	0.04	0.03	0.76	0.05	0.10	0.07	0.30	0.27	0.59	0.82 (0.59–1.14)
1121/1221	0.17	0.18	0.48	0.07	0.04	0.03	0.17	0.12	0.12	0.11	0.05	0.09	1.06 (0.80–1.41)
1121/2111	0.10	0.12	0.21	0.02	0.01	0.83	0.06	0.08	0.53	0.08	0.10	0.60	0.94 (0.67–1.32)
1221/1221	0.09	0.06	0.02	0.01	0.02	0.27	0.09	0.05	0.20	—	—	—	1.51 (0.94–2.41)
1221/2111	0.08	0.07	0.38	0.01	0.02	0.41	0.07	0.07	0.88	0.03	0.01	0.20	1.09 (0.69–1.70)
2111/2111	0.03	0.02	0.04	—	0.01	0.16	0.01	0.01	0.74	—	1.81	0.23	1.83 (0.81–4.12)
All others	0.04	0.04	0.39	0.05	0.06	0.39	0.04	0.06	0.35	0.01	0.01	0.59	0.70 (0.42–1.15)

[a] Diplotypes were commonly defined based on UC-SNP 43 (rs3792267, G/A), 19 (rs3842570, insertion/deletion), 63 (rs5030952, C/T) with or without SNP-44 (rs2975760, T/C); 1, major allele; 2, minor allele. – indicates no observations.

[b] *P*-values determined by chi-square statistics; all FDR *q*-values >0.25.

[c] OR for each diplotype was calculated for carriers versus all non-carriers; conditional logistic regression models were adjusted for matching factors (age, clinical center, time of blood draw and ethnicity) and other confounders including BMI, HRT, alcohol consumption, cigarette smoking, physical activity and family history of diabetes).

presence of underlying gene–gene or gene–environment interaction. There was also evidence for a statistical interaction between the *NIDDM1* region harboring *CAPN10* of chromosome 2 and *CYP19* region of chromosome 15 (21); a recent study showed that gene expression of *CAPN3*, another gene on chromosome 15, was negatively associated with body fat and insulin resistance in 27 non-diabetic individuals (22). Without more data from independent association and function studies, it is, however, difficult to draw any conclusions about whether or not they are true disease-predisposing genes, and other candidate genes in the region may also contribute to disease predisposition. Although there is no a priori reason to suspect that these other genes are associated with T2DM, they cannot be definitely excluded as possible candidates until further characterization of this region is complete. To adequately explore potential gene–gene and gene–environmental interactions, further investigation in large-scale and well-designed association studies is clearly warranted.

In addition to T2DM, the variants in *CAPN10* have been associated with polycystic ovary syndrome (23), obesity (24,25) and the metabolic syndrome (26), which share some pathophysiological traits that also underlie T2DM (e.g. insulin resistance and/or compensatory hyperinsulinemia). Intriguingly, biological data, mostly from *in vitro* studies, point to a role of *CAPN10* in several key pathways for the pathophysiology of T2DM including insulin-mediated glucose metabolism (6,7), insulin production and release in pancreatic β cell (6,8–10) and thermogenesis (11). However, such results should be interpreted cautiously because different cell types, quantitative methods and experimental conditions as well as non-specific CAPN antagonists were used. Furthermore, *CAPN10* encodes at least eight different protein isoforms through the use of alternative translation or splicing processes. Each specific CAPN10 protein isoform may function as a tissue-specific metabolic modulator, leading to suggest that different tissues may have different protein profiles in terms of *CAPN10* protein type. To clarify the molecular consequences of genetic variation at *CAPN10* locus, additional study will be required to investigate the regulation mechanisms underlying *CAPN10* mRNA expression, splicing and degradation and the consequences on protein translation, stability, activity and tissue-specific expression patterns.

Finally, it is worth mentioning that population stratification may be a concern when all individuals in this study self-identified their ethnicity, and data from women with diverse ethnic backgrounds are included. However, the consistency of the null associations across ethnic groups suggests that population stratification is not a major concern in our study. Our prospective study design and matching strategy used in such a well-characterized study population could eliminate the bias because of different selection procedures between cases and controls.

In conclusion, our large multi-ethnic case–control study of postmenopausal women did not replicate previous reports implicating *CAPN10* as a T2DM susceptibility locus, although we cannot exclude the possibility of a modest genetic effect. Our results provide no evidence to support the notion that common genetic variants in *CAPN10* may contribute significantly to the pathogenesis of T2DM.

MATERIALS AND METHODS

Study population

The WHI-OS is a longitudinal study designed to examine the association between clinical, socio-economic, behavioral and dietary risk factors with subsequent incidence of health outcomes, including cardiovascular disease (CVD) and diabetes. Details of the scientific rationale, eligibility and other design aspects have been described elsewhere (27,28). Between September 1994 and December 1998, the WHI-OS enrolled a total of 93 676 postmenopausal women aged 50–79 years at 40 clinical centers throughout the USA. Women of all races and ethnicities were included with a priority to enroll racial/ethnic minority groups in their representation in the US population of women aged 50–79 years. At baseline, women completed screening and enrollment questionnaires and underwent a physical examination and fasting blood specimen collection. The study has been reviewed and approved by human subjects review committees at each participating institution, and signed informed consent was obtained from all women enrolled.

WHI-OS participants were followed by annual mailed self-administered questionnaires and completed an annual medical history. Incident cases of diabetes were identified on the basis of post-baseline self-report of first-time use of hypoglycemic medication (oral hypoglycemic agents or insulin) during a median follow-up period of 5.9 years (mean value=5.5 years). Of the 93 676 postmenopausal women enrolled into the WHI-OS cohort, approximately 82 069 have no prior history of CVD and/or diabetes at baseline. Following the principle of risk-set sampling (27), for each new case, controls were selected randomly from women who remained free of CVD and/or diabetes at the time the case was identified during follow-up. Controls were matched to the cases by age ($\pm$2.5 years), racial/ethnic group (white/Caucasian, black/African, Hispanic/Latino and Asian/Pacific Islander), clinical center (geographic location), time of blood drawn ($\pm$0.10 h) and length of follow-up. In our current nested case–control study, 968 cases in whites were randomly chosen and matched with one control each. Of 616 incident cases among ethnic minority women, 366 cases were black, 152 were Hispanics and 98 cases were Asian/Pacific Islanders. The 1:2 matching ratio was used for minorities to strengthen the power in these smaller sample sizes of cases. We estimated that we would have at least >80% statistical power to detect a relative risk $\geq$1.5 among whites or black/African Americans and a relative risk $\geq$2.0 among other ethnic groups. Our study did not include American-Indian or Alaskan-native women because of their limited numbers.

Characterizing LD and haplotype-tagging SNP selection

As described in detail previously (29), we implemented a two-stage approach to choose haplotype tagging SNP (tSNP) for genotyping in our large case–control samples. The first stage consists of comprehensive common SNP discovery followed by genotyping them in 244 samples randomly selected from the WHI-OS source population. The second stage is selecting the tSNPs based on LD patterns.

In the first stage, we initially surveyed common genetic variation from the National Center for Biotechnology Information

database SNP (NCBI dbSNP) supplemented by HapMap database. Our goal was to capture the common haplotype patterns across 91.8 kb of the *CAPN10* gene, covering its 30 kb 5′ upstream and 30 kb 3′ downstream regions. In total, an initial set of 35 SNPs was selected on the basis of the following criteria: (i) functionality priority: non-synonymous coding SNPs (cSNPs) and splicing-site SNPs (ssSNPs) were kept following the order: cSNPs>ssSNPs>synonymous SNPs>5′-UTR SNPs>3′-UTR SNPs>intronic SNPs and (ii) MAF≥5% in at least one ethnic group (30). The initial set of 35 SNPs (one every ∼2.6 kb) was genotyped in a multi-ethnic panel of 244 women randomly chosen from the entire case–control sample (n=61 for each group of whites, blacks, Hispanics and Asian/Pacific Islanders). This sample size guaranteed that any common haplotype with a frequency of >5% were sufficiently captured.

In the second stage, we identified tSNPs on the basis of the LD patterns of those 35 SNPs among 61 individuals from each ethnic population. Pairwise LD between SNPs was assessed using Lewontin's D' statistic and the squared correlation statistic r^2. The Haploview program was used to calculate the LD coefficient and to define haplotype blocks (31,32). We chose all common tSNPs with special focus on African and Caucasian-American samples. First, we selected tSNPs in the African-American sample using the r^2-based Tagger program (33). tSNPs in the African-American were chosen by finding the minimum set of SNPs with $r^2 \geq 0.80$ and MAF ≥5%. We then used backward trimming to reach a minimum set of tSNPs for other ethnic groups to ensure a sufficient and yet non-redundant parsimonious set of tSNPs. From the initial dense set of 35 SNPs, a total of 12 tSNPs (including one insertion/deletion polymorphism, SNP-19, rs3842570) were eventually selected and genotyped in all case–control samples (Table 1).

As described elsewhere (29,30), common tSNPs chosen using our two-stage strategy captured majority of genetic variations (29) and provided sufficient statistical power to detect an association under a disease model of modest risk (19).

SNP genotyping method

DNA was extracted from the buffy coat fraction of centrifuged blood using the QIAmp Blood Kit (Qiagen, Chatsworth, CA, USA). For the SNP set at first stage, the SNPs were genotyped using the high-throughput Illumina BeadArray™ platform for 264 DNA samples (San Diego, CA, USA). The large-scale tSNP genotyping in the entire case–control sample consisting of 1543 cases and 2132 matched controls was performed by the TaqMan allelic discrimination method using an ABI 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA) in the UCLA Molecular Epidemiology Laboratory (Dr Simin Liu). The primers and probes were custom-designed by the ABI Taqman system (PE Biosystems, Foster City, CA, USA). Following polymerase chain reaction (PCR) amplification, endpoint fluorescence was read with the Applied Biosystems Primer 7900HT instrument and genotypes were assigned using SDS2.2.2 Allelic Discrimination Software (Applied Biosystems) by two independent technicians blinded to sample identification numbers. A DNA fragment containing SNP-19 (rs3842573), which is a tandem-repeat insertion/deletion variant with either two or three copies of a 32 bp fragment, was amplified by a PCR using a forward primer of 5′-GTTTGGTTCTCTTCAGCG TGGAG-3′ and a reverse primer of 5′-CATGAACCCTG GCAGGGTCTAAG-3′. PCR products were separated by electrophoresis on a 3.5% agarose gel stained with ethidium bromide. Allele 1 (two repeats) comprised 154 bp and allele 2 (three repeats) comprised 185 bp. A total of 138 replicate samples were randomly selected and duplicated across all plates (5%). Concordance rate was >99% and the average genotyping drop-out rate was 1.5% (from 0.4 to 5.9%) for each of the 12 SNPs in the case–control study.

Statistical analyses

We first assessed each tSNP for the HWE test using the χ^2 test. We also tested for heterogeneity of genotype distributions across ethnic groups by the χ^2 test. In both single-SNP and halpotype-based analyses, we employed conditional logistic regression to calculate ORs and 95% CIs for each genetic variant with diabetes risk (using the PROC PHREG procedure in SAS v9.1, SAS Institute, Cary, NC, USA). We made adjustments for potential confounding variables including matching factors (age, clinical center, time of blood drawn and ethnicity), body mass index (BMI), cigarette smoking (never, past and current), alcohol intake (never, past and current), hormone replacement therapy (HRT) usage (never, past, current), family history of diabetes and physical activity per week at baseline (expressed as total METs; total expenditure of energy from recreational physical activity). For each tSNP, we tested for allelic association with diabetes risk under dominant, recessive and additive models. Likelihood ratio test was used to test the interaction effect between the genotypes and ethnicity on diabetes risk.

We estimated haplotype frequency in each ethnic group using the expectation–maximization (EM) algorithm applied to phase-unknown case–control genotype data (as implemented in SAS PROC HAPLOTYPE) (34). Haplotype frequency was validated via HAPLOTYPER version 2, a Bayesian haplotype inference algorithm (35). For each individual and each haplotype, h, the haplotype dosage estimate (i.e. an estimate of the number of copies of haplotype h) was computed using the individual's genotype data and haplotype frequency estimates obtained from the combined (cases and controls) data set (34,36). Only haplotypes with estimated frequencies ≥1% in the combined cases and controls were included for analyses. We first performed global likelihood ratio tests to examine whether the frequency distributions of the common haplotypes differed between cases and controls by comparing a model with additive effects on the log odds scale for each common haplotype (using the most common halpotype as the referent) with the intercept-only model. We then formally tested for ethnic differences in haplotype-associated risks by performing a likelihood ratio test following the inclusion of an interaction term between the risk haplotypes and ethnicity in the multi-variable model. To examine the association between the resulting haplotype/haplotype combinations and diabetes risk, the estimate of haplotype dosage was treated as a surrogate variable for the true haplotype. All the analyses were performed using SAS

statistical package (version 9.0 for window; SAS Institute), unless specified.

To account for the multiple comparisons made in this study, we calculated the *q*-value statistic to estimate the FDR by incorporating all *P*-values from multiple tests performed for SNPs, haplotypes and diplotypes in the Cox models (37). The *q*-values were obtained for each and all of the *P*-values, and the *q*-values for $P \leq 0.05$ were presented.

ACKNOWLEDGEMENTS

We would like to acknowledge all WHI centers and their principal investigators for their participation in this research. We are indebted to all dedicated and committed participants of the WHI-OS. We thank Alyssa Smith and Joseph Larson from the WHI Clinical Coordinating Center for their assistance. A short list of WHI investigators are shown in an on-line appendix file.

Conflict of Interest statement. None of the authors had any conflicts of interest.

FUNDING

This study was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) R01 grant DK062290 from the National Institutes of Health. The WHI program is funded by the National Heart, Lung and Blood Institute and by the US Department of Health and Human Services.

REFERENCES

1. Barroso, I. (2005) Genetics of type 2 diabetes. *Diabet. Med.*, **22**, 517–535.
2. Hansen, L. and Pedersen, O. (2005) Genetics of type 2 diabetes mellitus: status and perspectives. *Diabetes Obes. Metab.*, **7**, 122–135.
3. Horikawa, Y., Oda, N., Cox, N.J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T.H., Mashima, H., Schwarz, P.E. *et al.* (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.*, **26**, 163–175.
4. Goll, D.E., Thompson, V.F., Li, H., Wei, W. and Cong, J. (2003) The Calpain system. *Physiol. Rev.*, **83**, 731–801.
5. Ma, H., Fukiage, C., Kim, Y.H., Duncan, M.K., Reed, N.A., Shih, M., Azuma, M. and Shearer, T.R. (2001) Characterization and expression of calpain 10. A novel ubiquitous calpain with nuclear localization. *J. Biol. Chem.*, **276**, 28525–28531.
6. Baier, L.J., Permana, P.A., Yang, X., Pratley, R.E., Hanson, R.L., Shen, G.Q., Mott, D., Knowler, W.C., Cox, N.J., Horikawa, Y. *et al.* (2000) A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. *J. Clin. Invest.*, **106**, R69–R73.
7. Hoffstedt, J., Ryden, M., Lofgren, P., Orho-Melander, M., Groop, L. and Arner, P. (2002) Polymorphism in the calpain 10 gene influences glucose metabolism in human fat cells. *Diabetologia*, **45**, 276–282.
8. Sreenan, S.K., Zhou, Y.P., Otani, K., Hansen, P.A., Currie, K.P., Pan, C.Y., Lee, J.P., Ostrega, D.M., Pugh, W., Horikawa, Y. *et al.* (2001) Calpains play a role in insulin secretion and action. *Diabetes*, **50**, 2013–2020.
9. Marshall, C., Hitman, G. and Turner, M. (2003) Calpain-10 regulates insulin secretion. *Diabetes*, **52**, A373.
10. Marshall, C., Hitman, G.A., Partridge, C.J., Clark, A., Ma, H., Shearer, T.R. and Turner, M.D. (2005) Evidence that an isoform of calpain-10 is a regulator of exocytosis in pancreatic beta-cells. *Mol. Endocrinol.*, **19**, 213–224.
11. Hoffstedt, J., Naslund, E. and Arner, P. (2002) Calpain-10 gene polymorphism is associated with reduced beta(3)-adrenoceptor function in human fat cells. *J. Clin. Endocrinol. Metab.*, **87**, 3362–3367.
12. Tsuchiya, T., Yan, H., Vogelstein, B., Kinzler, K. and Bell, G.I. (2003) Identification of a haplotype associated with the decreased expression of calpain-10 in lymophoblastoid cell lines. *Diabetes*, **52**, A256.
13. Song, Y., Niu, T., Manson, J.E., Kwiatkowski, D.J. and Liu, S. (2004) Are variants in the CAPN10 gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *Am. J. Hum. Genet.*, **74**, 208–222.
14. Fullerton, S.M., Bartoszewicz, A., Ybazeta, G., Horikawa, Y., Bell, G.I., Kidd, K.K., Cox, N.J., Hudson, R.R. and Di Rienzo, A. (2002) Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *Am. J. Hum. Genet.*, **70**, 1096–1106.
15. Vander Molen, J., Frisse, L.M., Fullerton, S.M., Qian, Y., Del Bosque-Plata, L., Hudson, R.R. and Di Rienzo, A. (2005) Population genetics of CAPN10 and GPR35: implications for the evolution of type 2 diabetes variants. *Am. J. Hum. Genet.*, **76**, 548–560.
16. The International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
17. Tsuchiya, T., Schwarz, P.E., Bosque-Plata, L.D., Geoffrey Hayes, M., Dina, C., Froguel, P., Wayne Towers, G., Fischer, S., Temelkova-Kurktschiev, T., Rietzsch, H. *et al.* (2006) Association of the calpain-10 gene with type 2 diabetes in Europeans: results of pooled and meta-analyses. *Mol. Genet. Metab.*, **89**, 174–184.
18. Elbein, S.C., Chu, W., Ren, Q., Hemphill, C., Schay, J., Cox, N.J., Hanis, C.L. and Hasstedt, S.J. (2002) Role of calpain-10 gene variants in familial type 2 diabetes in Caucasians. *J. Clin. Endocrinol. Metab.*, **87**, 650–654.
19. de Bakker, P.I., Burtt, N.P., Graham, R.R., Guiducci, C., Yelensky, R., Drake, J.A., Bersaglieri, T., Penney, K.L., Butler, J., Young, S. *et al.* (2006) Transferability of tag SNPs in genetic association studies in multiple populations. *Nat. Genet.*, **38**, 1298–1303.
20. Hinds, D.A., Stuve, L.L., Nilsen, G.B., Halperin, E., Eskin, E., Ballinger, D.G., Frazer, K.A. and Cox, D.R. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science*, **307**, 1072–1079.
21. Cox, N.J., Frigge, M., Nicolae, D.L., Concannon, P., Hanis, C.L., Bell, G.I. and Kong, A. (1999) Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nat. Genet.*, **21**, 213–215.
22. Walder, K., McMillan, J., Lapsys, N., Kriketos, A., Trevaskis, J., Civitarese, A., Southon, A., Zimmet, P. and Collier, G. (2002) Calpain 3 gene expression in skeletal muscle is associated with body fat content and measures of insulin resistance. *Int. J. Obes. Relat. Metab. Disord.*, **26**, 442–449.
23. Vollmert, C., Hahn, S., Lamina, C., Huth, C., Kolz, M., Schoepfer-Wendels, A., Mann, K., Bongardt, F., Mueller, J.C., Kronenberg, F. *et al.* (2007) Calpain-10 variants and haplotypes are associated with polycystic ovary syndrome in Caucasians. *Am. J. Physiol. Endocrinol. Metab.*, **292**, 836–844.
24. Carlsson, E., Fredriksson, J., Groop, L. and Ridderstrale, M. (2004) Variation in the calpain-10 gene is associated with elevated triglyceride levels and reduced adipose tissue messenger ribonucleic acid expression in obese Swedish subjects. *J. Clin. Endocrinol. Metab.*, **89**, 3601–3605.
25. Shima, Y., Nakanishi, K., Odawara, M., Kobayashi, T. and Ohta, H. (2003) Association of the SNP-19 genotype 22 in the calpain-10 gene with elevated body mass index and hemoglobin A1c levels in Japanese. *Clin. Chim. Acta*, **336**, 89–96.
26. Hitman, G.A. and Sudagani, J. (2004) Searching for genes in diabetes and the metabolic syndrome. *Int. J. Clin. Pract.*, **143**, 3–8.
27. Anderson, G.L., Manson, J., Wallace, R., Lund, B., Hall, D., Davis, S., Shumaker, S., Wang, C.Y., Stein, E. and Prentice, R.L. (2003) Implementation of the Women's Health Initiative study design. *Ann. Epidemiol.*, **13**, S5–S17.
28. The Women's Health Initiative Study Group(1998) Design of the Women's Health Initiative clinical trial and observational study. *Control Clin. Trials*, **19**, 61–109.
29. Hao, K., Liu, S. and Niu, T. (2005) A sparse marker extension tree algorithm for selecting the best set of haplotype tagging single nucleotide polymorphisms. *Genet. Epidemiol.*, **29**, 336–352.

30. Wang, L., Liu, S., Niu, T. and Xu, X. (2005) SNPHunter: a bioinformatic software for single nucleotide polymorphism data acquisition and management. *BMC Bioinformatics*, **6**, 60.
31. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
32. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
33. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat. Genet.*, **37**, 1217–1223.
34. Kraft, P., Cox, D.G., Paynter, R.A., Hunter, D. and De Vivo, I. (2005) Accounting for haplotype uncertainty in matched association studies: a comparison of simple and flexible techniques. *Genet. Epidemiol.*, **28**, 261–272.
35. Niu, T., Qin, Z.S., Xu, X. and Liu, J.S. (2002) Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am. J. Hum. Genet.*, **70**, 157–169.
36. Stram, D.O., Leigh Pearce, C., Bretsky, P., Freedman, M., Hirschhorn, J.N., Altshuler, D., Kolonel, L.N., Henderson, B.E. and Thomas, D.C. (2003) Modeling and E–M estimation of haplotype-specific relative risks from genotype data for a case–control study of unrelated individuals. *Hum. Hered.*, **55**, 179–190.
37. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. and Golani, I. (2001) Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.*, **125**, 279–284.