

Single Nucleotide Polymorphisms in CAPN10 Gene of Chinese People and Its Correlation With Type 2 Diabetes Mellitus in Han People of Northern China¹

R3 A

SUN HONG-XIA^{*#}, ZHANG KUI-XING[#], DU WEI-NAN[^], SHI JIN-XIU[#], JIANG ZHENG-WEN[#], SUN HAO[△], ZUO JIN^{*}, HUANG WEI^{#2}, CHEN ZHU[#], SHEN YAN^{§ 2}, YAO ZHI-JIAN[§], QIANG BO-QIN[§], AND FANG FU-DE^{*2}

^{}National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China; [#] Chinese National Human Genome Center at Shanghai, Shanghai 201203, China; [△] Institute of Medical Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Kunming 650107, China; and [§] Chinese National Human Genome Center at Beijing, Beijing 100176, China*

Objective To investigate the distribution of single nucleotide polymorphisms (SNPs) in CAPN10 gene in Chinese population and their relation with type 2 diabetes mellitus in Han people of Northern China. **Methods** CAPN10 gene was sequenced to detect SNPs in different nationalities of China. Five SNPs were chosen to perform case-control study and haplotype analysis in 156 normal Han people of Northern China and 173 type 2 diabetes. One SNP was also analyzed with transmission-disequilibrium test (TDT) and sib transmission-disequilibrium test (STDT) in 68 type 2 diabetes pedigrees (377 people). **Results** A total of 40 SNPs were identified in length of 8 936bp, with an average of 1 in every 223bp. The SNPs in CAPN10 gene did not distribute evenly and the SNPs in Chinese were different from those reported in Mexican American. There was no significantly statistical difference in the allele frequency of the 5 SNPs between case and control, and the haplotype frequencies in the two groups were not significantly different. No positive results was found in TDT and STDT analysis. **Conclusions** The SNP distribution of CAPN10 gene differs in different nationalities. The studied SNPs in CAPN10 gene may not be the major susceptibility ones of type 2 diabetes mellitus in Han people of Northern China.

Key words: CAPN10; SNP; Different nationalities in China; Type 2 diabetes; Association study

¹This work was supported by the National Natural Science Foundation (Grant No. 30170441, 39896200), 863 High Tech Project (Grant No. 2001AA221161), National 973 Project (Grant No. G1998051016), Beijing Natural Science Fund (Grant No. 7002026) and the National High Education Science Fund (Grant No. 20010023024).

²Corresponding author: E-mail: Fang Fu-De (fangfd@public3. bra. net. cn), Huang Wei (huangwei@chgc. sh. cn), Shen Yan (shengy@cdm. imicams. ac. cn).

Biography of the first author: Sun Hong-Xia, female, Ph. D. candidate in National Lab of Medical Molecular Biology, Institute of Basic Medical Sciences, CAMS & PUMC.

0895-3988/2002
CN11-2816
Copyright © 2002 by CAPM

INTRODUCTION

Single nucleotide polymorphism (SNP) is DNA sequence polymorphism caused by single nucleotide diversity in genome level, which is the most common kind of inheritable diversity in human. In recent years, with the study process of human genome and genetics, SNP has begun to show greater importance in investigating the genome diversity of individual, group and disease studies.

Recently, Horikawa *et al.*^[1] made a new breakthrough in mapping type 2 diabetes susceptibility genes. Based on the former gene mapping work^[2], they had successfully found a susceptibility gene of type 2 diabetes, CAPN10, in Mexican American. In this gene, three SNPs, named UCSNP-43, UCSNP-19 and UCSNP-63 respectively, were reported to be associated with the disease and their different haplotypes made a very different risk to the population. The high risk haplotype also showed association with the high incidence of type 2 diabetes in Finnish and German. As SNPs differ in population and nationalities, it is very important to identify SNPs in CAPN10 in Chinese, investigate their distribution and their relation with type 2 diabetes in mapping and cloning susceptibility genes of type 2 diabetes mellitus in China.

MATERIALS AND METHODS

Sample Collection and Genomic DNA Extraction

27 unrelated individual samples from different nationalities in China were collected, including 10 Han, 2 Chuang, 2 Tibetan, 2 Bulang, 1 Korean, 2 Dai, 2 Mongolian, 2 Miao, 2 Yi and 2 Uygur. DNA was extracted from peripheral blood by the standard phenol/chloroform method, its concentration was measured and adjusted to 1.5 ng/ μ L.

173 sporadic type 2 diabetes affected from Beijing area versus 152 both sex and age-matched control were enrolled to take a case-control study. TDT and STDT were performed for 377 members in 68 type 2 diabetes pedigrees collected from Northern China, including 192 affected, 20 IGT and 165 unaffected. DNA was extracted as above and adjusted to 20 ng/ μ L.

SNP Identification in CAPN10

According to the CAPN10 gene sequence reported^[1], PCR primers were designed to amplify regions up to about 2kb upstream from transcription-initiation sites and all the coding regions by the Primer 3 computer program. Multiple nested PCR was used to amplify the extracted DNA.

After electrophoresis and resin purification of the PCR product, bi-directional sequencing was performed. Products were analyzed with ABI 377 autosequencer. The polyphred computer program^[3] was used to detect the presence of heterozygous SNPs by fluorescence-based sequencing of PCR products.

Selection and Primer Design of SNPs in Case-control Study

5 SNPs were selected to perform case-control study, including 1 in regulatory region (-1410), 1 in exon which could change the amino acid code (+9803), the other 3 were corresponded with the reported UCSNP-43 (+4852), UCSNP-19 (+7920) and UCSNP-

63(+16378). According to the reference^[1], 3 primers were designed for each locus, 2 for PCR and 1 for SBE.

SNPs Genotype and Data Analysis

According to the literature^[4], the procedures were as follows: amplifying the SNP-containing loci with flanking primers and digesting the PCR product with Exonuclease I and calf intestine alkaline phosphatase to rule out the redundant primers and unincorporated deoxynucleotide. SBE reaction was performed by using a primer adjacent to the SNP to incorporate a fluorescence labeled dideoxynucleotide. The SNP alleles were indicated according to the SBE product position and the fluorescence after electrophoresis on ABI377 sequencer. Thus, 5 SNPs in CAPN10 gene were genotyped in case and control groups and UCSNP-43 was also done in pedigrees.

Hardy-Weinberg equilibrium (HWE) was used to test the respectability of the samples. The difference of allele frequency of each SNP between control and case groups was analyzed with SPSS10.0 software. Logistic regression was used to analyze the association between disease and SNP genotypes. PM1.1 was used to analyze the haplotype frequencies of UCSNP-43 (+4852), UCSNP-19 (+7920) and UCSNP-63 (+16378) in the two groups^[5] and TDT-STDT program 1.1 was used to perform TDT and STDT for UCSNP-43.

RESULTS

SNPs Identified in CAPN10 Gene Region in Chinese

8 936bp of CAPN10 gene was sequenced, including 2 000bp in regulation region, 2 451bp in coding region, 177bp in 5' flanking region and 352bp in 3' flanking region. A total of 40 SNPs were identified, with an average of 1 in each 223bp. The SNPs distribution was unequal in the gene. 7 in regulation region, with an average of 1 in each 285bp; 6 in coding region, including 5 synonymous and 1 nonsynonymous, with an average of 1 in each 405bp; 26 in intron, 1 in each 120bp; and there was 1 SNP in the 3' flanking region. Among these SNPs, 28 were base transition, accounting for 70 per cent of all the identified; and 11 were transversion, accounting for 27.5 per cent. The ratio of the two was about 2.6 to 1. Another SNP was formed by the different repeats of a 32bp fragment (Table 1).

The Character of CAPN10 SNPs in Chinese

The SNPs identified in Chinese were much different from those reported by Horikawa *et al.* in Mexican American. Among the 40 SNPs, 16 were found in Chinese but not reported in Mexican American, which accounted for 40% of the total. In addition, 16 SNPs were reported in Mexican American, accounting for 40% of the total, the others were common in the two populations (Table 2).

The Case-control Study of 5 SNPs in CAPN10 Gene Region

SNPs used in these comparisons were in HWE. There were no obvious difference in SNP allele frequency between the two groups, logistic regression analysis showed no association between SNP genotypes and type 2 diabetes (Table 3).

TABLE 1

SNPs Position, Type and Allele Frequency in CAPN10 Gene Region in Chinese

No.	Position	SNP Name	Location	Allele 1	Allele 2	A1%	AA 1	AA 2
1	-1410		Promotor	G	A	16.7		
2	-1307		Promotor	A	G	3.7		
3	-1013		Promotor	C	T	14.8		
4	-690		Promotor	C	A	9.3		
5	-638		Promotor	A	G	9.3		
6	-489		Promotor	T	C	5.6		
7	-471		Promotor	G	C	9.3		
8	+2606		intron	A	G	19.2		
9	+2647		intron	C	G	16.0		
10	+4213		intron	A	T	4.2		
11	+4238		intron	G	A	4.2		
12	+4258		intron	T	A	4.2		
13	+4262		intron	T	C	4.2		
14	+4294		intron	T	A	20.8		
15	+4428		intron	G	A	4.2		
16	+4471		intron	G	A	10.4		
17	+4795	UCSNP45	intron	C	A	3.7		
18	+4841	UCSNP44	intron	C	T	9.6		
19	+4852	UCSNP43	intron	A	G	13.0		
20	+5091		coding	T	C	1.9	Thr	Thr
21	+5155		coding	A	C	1.9	Pro	Thr
22	+5157	UCSNP79	coding	G	A	14.8	Pro/Thr	Pro/Thr
23	+5253	UCSNP78	intron	G	A	1.9		
24	+5262	UCSNP77	intron	A	G	11.1		
25	+7189		intron	A	T	9.3		
26	+7236		intron	T	C	7.7		
27	+7522	UCSNP60	intron	T	C	1.9		
28	+7611		intron	C	T	5.6		
29	+7769		coding	T	C	3.8	Val	Val
30	+7840		intron	A	G	16.0		
31	+7865		intron	T	A	5.6		
32	+7920	UCSNP19	intron	2	3	37.0		
33	+9730	UCSNP109	intron	T	C	22.9		
34	+9803	UCSNP110	coding	G	A	13.5	Thr	Ala
35	+11098	UCSNP48	coding	G	A	20.0	Ala	Ala
36	+11246		intron	C	T	4.5		
37	+11738		intron	C	T	14.8		
38	+11906		3'UTR	A	G	16.7		
39	+16378	UCSNP63	intron	T	C	50.8		
40	+16398		intron	G	C			

Note. SNP position was marked +1 in the transcription-initiation site, those upstream from this site were marked - and downstream marked +. SNP Name indicated its reported name. A1% was allele1 frequency. AA1 was amino acid 1 and AA2 was amino acid 2. +7920 2/3 was 2 and 3 times repeat of a 32bp fragment respectively.

TABLE 2

The Similarities and Differences of SNPs in CAPN10 Gene Region in Chinese and Mexican American

Special SNPs in Chinese		Special SNPs in Mexican American		Common SNPs in the 2 Populations		
-1307G/A	-690A/C	-1193G/A	+2619G/A	-1410A/G	-1013T/C	+2606G/A
-638G/A	-489C/T	+3828C/G	+3829C/G	+2647G/C	+4428A/G	+4471A/G
-471C/G	+4213T/A	+4577T/C	+4780C/G	+4852G/A	+5155C/A	+5157A/G
+4238A/G	+4258A/T	+4794G/C	+4841C/G	+5253A/G	+5262G/A	+7189T/A
+4262C/T	+4294A/T	+5162G/A	+7022C/T	+7236C/T	+7522C/T	+7611T/C
+4795A/C	+4841T/C	+8143C/T	+9878G/T	+7769C/T	+7920 T/3	+9730C/T
+5091C/T	+7840G/A	+11076G/A	+11512G/A	+9803A/G	+11098A/G	+11246T/C
+7865A/T	+11906G/A	+11751G/A	+11877G/C	+11738T/C	+16378C/T	+16398C/G

TABLE 3

The Case-control Study of 5 SNP Loci in CAPN10

CAPN10-1410								
Group	Allele Types				Genotype Frequency			
	A	G	Total	P value	AA	AG	GG	P value
Control	279	29	308	0.958	82.47 (127)	16.23 (25)	1.30 (2)	
Case	313	33	346		81.50 (141)	17.92 (31)	0.58 (1)	0.748
Total	592	62	654					
CAPN10+4852								
Group	Allele Types				Genotype Frequency			
	G	A	Total	P value	GG	AG	AA	P value
Control	274	38	312	0.897	76.28 (119)	23.08 (36)	1.30 (1)	
Case	305	41	346		78.03 (135)	20.23 (35)	1.74 (3)	0.583
Total	579	79	658					
CAPN10+7920								
Group	Allele Types				Genotype Frequency			
	2	3	Total	P value	22	23	33	P value
Control	103	209	312	0.488	12.18 (19)	41.67 (65)	46.15 (72)	
Case	108	246	354		7.91 (14)	45.20 (80)	46.89 (83)	0.769
Total	211	455	666					
CAPN10+9803								
Group	Allele Types				Genotype Frequency			
	A	G	Total	P value	AA	AG	GG	P value
Control	288	24	312	0.931	85.26 (133)	14.10 (22)	0.64 (1)	
Case	320	26	346		84.97 (147)	15.03 (26)	0 (0)	0.905
Total	608	50	658					
CAPN10+16378								
Group	Allele Types				Genotype Frequency			
	C	T	Total	P value	CC	CT	TT	P value
Control	240	72	312	0.904	59.62 (93)	35.26 (55)	5.12 (8)	
Case	232	68	300		61.33 (92)	32.00 (48)	6.67 (10)	0.747
Total	472	140	612					

Haplotype Frequencies of UCSNP-43 (+4852), UCSNP-19 (+7920) and UCSNP-63+16378) in Case and Control Groups

There were 8 haplotypes formed by the 3 SNPs, and no significant difference in their frequencies between the two groups was found during PM1. 1 software analysis(Table 4).

TABLE 4

Haplotype Frequencies UCSNP-43, -19 and -63					
	Allele			Case Group	Control Group
	43	19	63		
	1	1	1	0.21	0.22
	1	1	2	0.06	0.07
	1	2	1	0.47	0.45
	1	2	2	0.14	0.14
	2	1	1	0.03	0.03
	2	1	2	0.01	0.01
	2	2	1	0.06	0.06
	2	2	2	0.02	0.02

Note. UCSNP-43, allele 1, G, allele 2, A; UCSNP-19, allele 1, 2 repeats of 32 bp sequence, allele 2, 3 repeats; UCSNP-63, allele 1, C, allele 2, T.

TDT and STDT Analysis of UCSNP-43

Of the 68 type 2 diabetes pedigrees, those in which parents genotypes were known and at least one was heterozygous were analyzed with traditional TDT to estimate whether the allele frequency transmitted from parents to offspring departed from 50% or not. Those pedigrees with at least one affected and one unaffected whose genotypes were not the same were analyzed with STDT to compare the genotype distribution of marked locus between affected and unaffected offspring and to test whether the marker allele frequency was significantly different between different state offspring or not. The results showed that the allele frequency transmitted from heterozygous parents to affected offspring did not significantly depart from 50% ($P > 0.05$) and there was no statistical difference in the allele distribution between affected and unaffected offspring, suggesting no linkage between marker and disease loci (Table 5).

TABLE 5

TDT and STDT Analysis of UCSNP-43											
Allele	TDT			Y	S-TDT			W	Combined Scores		
	b	c	Chi-Sq		Mean(A)	Var(V)	z'		Mean(A)	Var(V)	z'
1	39	32	0.690	5	5.250	0.188	-0.577	44	40.750	17.938	0.649
2	32	39	0.690	1	0.750	0.188	-0.577	33	36.250	17.938	0.649

DISCUSSION

Now, there are two most used strategies in the study of complex genetic diseases, such as diabetes mellitus, hypertension, and so on. One is mapping and cloning method, which

maps the disease susceptibility genes with whole-genome scan and then clones the genes. This method is much affected by the collected disease pedigrees. The other strategy is candidate genes method, which directly studies the relationship of the variation of disease candidate genes with disease phenotypes. This method is based on case-control association study of pedigrees or population. It is much direct and fast, and is widely used in complex genetic disease study.

On October 2000, Horikawa *et al.* reported the first NIDDM susceptibility gene cloned in Mexican American, CAPN10, which is mapped to the chromosome of 2q37.3 and consists of 15 exons. CAPN10 is ubiquitously expressed and its different transcripts can be expressed in different tissues. The CAPN10 protein is a member of the cysteine protease family, which suggests a new biochemical pathway involved in the regulation of blood glucose levels.

According to the CAPN10 sequence reported by Horikawa *et al.*, its SNPs were measured in different nationalities of China for the first time by using resequencing method. A total of 40 SNPs were identified. The SNPs in CAPN10 are characterized by: (1) The distribution of SNPs is unequal, 1 in each 120bp at the dense position and 1 in each 405bp at the sparse position, with an average of 1 in 223bp. There are more SNPs in non-transcription region than in the transcription region, because the SNPs in non-transcription region do not influence the protein transcription, and the natural selection pressure they suffered is more little^[6,7]. (2) In the SNPs the authors identified, transition is much more than transversion, the ratio of the two is about 2.6:1, which is consistent with that formerly reported and may be related to the structure of the bases, because cytosine can spontaneously deaminate to yield a thymidine residue^[8]. (3) There is an obvious difference of the SNPs detected in Chinese with those reported in Mexican American, which suggests the ethnic difference of the SNP distribution of CAPN10 and supports the authors' related study.

In the 5 SNPs used in case-control association study, none was found to be related with type 2 diabetes and there was no statistical difference in the haplotype frequencies of UCSNP-43 (+4852), UCSNP-19 (+7920) and UCSNP-63 (+16378) between the two groups. TDT and STDT analysis showed no linkage between the reported positive UCSNP-43 and type 2 diabetes in Han people of Northern China. These weren't accordant with those of Horikawa *et al.* and may be interpreted as ethnic difference. From above, it is concluded that these SNPs in CAPN10 gene may not be the major susceptibility loci of type 2 diabetes mellitus in Han people of Northern China, which is consistent with the authors' former work of gene mapping^[9].

REFERENCES

1. Horikawa, Y., Oda, N., Cox, N. J., Li, X., Orho-Melander, M., Hara, M., Hinokio, Y., Lindner, T. H., Mashima, H., Schwarz, P. E. H., Bosque-Plata, L. D., Horikawa, Y., Oda, Y., Yoshiuchi, I., Colilla, S., Polonsky, K. S., Wei, S., Concannon, P., Iwasaki, N., Schulze, J., Baier, L., Bogardus, C., Groop, L., Boerwinkle, E., Hanis, C. L., and Bell, G. I. (2000). Genetic variation in the gene coding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genet.* **26**, 163-175.
2. Hanis, C. L., Boerwinkle, E., Chakraborty, R., Ellsworth, D. L., Concannon, P., Stirling, B., Morrison, V. A., Wapelhorst, B., Spielman, R. S., Gogolin-Ewens, K. J., Shephard, J. M., Williams, S. R., Risch, N., Hinds, D., Iwasaki, N., Ogata, M., Omori, Y., Petzold, C., Poetsch, H., Schröder, H., Schulze, J., Cox, N. J., Menzel, S., Boriraj, V. V., Chen, X., Lim, L. R., Lindner, T., Mereu, L. E., Wang, Y.-Q., Xiang, K., Yamagata, K., Yang, Y., and Bell, G. I. (1996). A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nature Genet.* **13**, 161-166.
3. Nickerson, D. A., Tobe, V. O., and Taylor, S. (1997). Phred, Phrap, Consed. Polyphred. Reference: Polyphred: Automating the detection and genotyping of Single Nucleotide Substitution using fluorescence-based resequencing. *Nucleic Acid*

- Research*. **25**, 2745-2751.
4. Lindblad-Toh, K., Winchester, E., Daly, M. J., Wang, D. G., Hirschhorn, J. N., Lavoie, J. -P., Ardlic, K., Reich, D. E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J. -B., Gingeras, T., Warrington, J., Patil, N., Hudson, T. J., and Lander, E. S. (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genet.* **24**, 381-386.
 5. Zhao, J. H., Curtis, D., and Sham, P. C. (2000). Model-free analysis and permutation tests for allelic association. *Hum Hered.* **50**, 133-139.
 6. Li, W. and Sadle, L. A. (1991). Low nucleotide diversity in man. *Genetics*. **129**, 513-523.
 7. Nickerson, D. A., Taylor, S. L., Weiss, K. M., Clark, A. G., Hutchinson, R. G., Stengård, J., Salomaa, V., Vartiainen, E., Boerwinkle, E., and Sing, C. F. (1998). DNA sequence diversity in a 9.7kb region of the human lipoprotein lipase gene. *Nature Genet.* **19**, 233-240.
 8. Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kibum, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., and Lander, E. S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*. **280**, 1077-1082.
 9. Zhao, J. Y., Wang, H., Xiong, M. M., Huang, W., Zuo, J., Chen, Z., Qiang, B. Q., Sun, Q., Li, Y. X., Liu, Q. Y., Du, W. N., Chen, J. L., Ding, W., Yuan, W. T., Zhao, Y., Xu, H. Y., Jin, L., and Fang, F. D. (2000). The localization of type 2 diabetes susceptibility gene loci in Northern Chinese Han families. *Chin Sci Bull.* **45**, 1792-1795.

(Received June 23, 2001 Accepted November 17, 2001)