REVIEW

Calpain-10 (NIDDM1) as a Susceptibility Gene for Common Type 2 Diabetes

YUKIO HORIKAWA

Department of Diabetes and Endocrinology, Gifu University School of Medicine, Gifu; the Laboratory of Medical Genomics, Biosignal Genome Resource Center, Institute for Molecular and Cellular Regulation, Gunma University, Gunma; Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST), Kawaguchi, Japan

Key words: Calpain 10, SNP, Haplotype Type 2 diabetes mellitus, NIDDM

It is assumed that susceptibility genes associated with lifestyle-related diseases including type 2 or non-insulin-dependent diabetes mellitus (NIDDM) were positively selected for energy conservation but act adversely in modern conditions. These “thrifty genes” can be identified on the “common disease common variant” hypothesis, but detailed analysis of the genetic polymorphisms in many ethnic groups is required to clarify the molecular evolution of these susceptibility alleles. At present, single nucleotide polymorphisms (SNPs) represent the most useful data for genetic analyses of non-Mendelian polygenic lifestyle-related diseases, the common diseases including diabetes, hypertension, and obesity. Although many association studies have been conducted using single SNPs, they are problematical for several reasons, including ethnic differences within the study population, unknown environmental factors, misdiagnosed disease, and genetic mistyping. Haplotype analysis whereby several tag SNPs can be monitored simultaneously improves and complements the search. NIDDM1 (calpain-10) is the first susceptibility gene for type 2 diabetes to be identified by this method.

SNP and haplotype

When a single SNP is used in analysis, there are only two possible variants, but if multiple SNPs are used, there are $2^n$ possible variants in haplotypes, somewhat less in the case of linkage disequilibrium (LD). In addition, determination of the degree of LD by LD coefficients (e.g., $D$, $D^2$, $r^2$) permits estimation of the physical map distance from a susceptibility gene, narrowing a disease locus to 10–100 kb. Thus, LD mapping of haplotypes is potentially more fruitful in screening susceptibility genes for common diseases. When marker SNPs and a putative disease susceptibility allele are located in the same LD block, haplotype structure analysis is relevant for any variant regardless of its frequency. On the other hand, if the putative disease allele is not located in an LD block with the SNP markers, single SNP analysis may yet be more useful than analysis of the haplotype structure [1].

A meta-analysis assessed the collected data on single SNPs in a comprehensive manner to screen candidate polymorphisms for polygenic diseases, and found only 16% of the candidate polymorphisms to have a significant genetic association that could be replicated without heterogeneity or bias. However, the majority of these candidate polymorphisms were false-positive results, attributed later to insufficient sample size or first versus subsequent discrepancies [2]. LD analyses using single SNPs are clearly liable to type I errors, while simultaneous assay of several SNPs contained in haplotype reduces the confounding effects of ethnic
differences within the study population, environmental factors, misdiagnosed disease, and genetic mistyping, greatly facilitating molecular-level genetic analysis of naturally selected thrifty genes.

**Identification of type 2 diabetes susceptibility genes**

Nearly 50% of Mexican Americans aged 35 years or older have diabetes mellitus or family history in first-degree relatives. A genome-wide linkage study with affected sib-pairs for type 2 diabetes genes in a Mexican American population localized a type 2 diabetes susceptibility gene (*NIDDM1*) to a 12-cM 1-locus support interval in the distal long arm of chromosome 2 near D2S140 [3]. A physical map of the *NIDDM1* region was generated to identify the SNPs within the region by resequencing the expressed sequence tags (EST). SNPs were surveyed in eight patients of families with evidence of linkage at *NIDDM1* and in two patients of families without such evidence. SNPs with minor allele frequency of more than 10% or showing a unique pattern were genotyped in a patient group of 110 Mexican Americans with type 2 diabetes and a control group of 112 randomly sampled subjects to compare allele and haplotype frequency distributions between the groups as described below. Haplotype structure was determined by the expectation maximization (EM) algorithm [4]. We then classified the patients into three subgroups for comparisons of association with the *NIDDM1* allele. The first patient subgroup included all 110 patients, the second subgroup included the 37 patients from families showing evidence of linkage at *NIDDM1*, and the third subgroup included the 20 patients from families showing evidence of linkage at both *NIDDM1* and *CYP19*, a marker located on the second peak of chromosome 15 [5] (Fig. 1).

Of these, no putative single SNP was significantly associated with increased risk of diabetes. However, when three SNPs were combined in a computed LD block, haplotypes composed of the three intron SNPs in the region of calpain-10, namely SNP-43, -19, and -63, showed a significant association with type 2 diabetes. Specifically, the 112/121-haplotype combination (1 = major allele, 2 = minor allele) was associated with increased risk of diabetes in our Mexican Ameri-

---

**Fig. 1.** The guide map of identification of *NIDDM1* SNPs are numbered in the order in which they were identified. Genes and ESTs were found on screening the GenBank database with the indicated STS. The locations of the SNPs typed in 110 patients and the 112 random samples are shown by arrows.
can group (O.R., 3.02; 95% C.I., 1.37 to 6.64). This combination also was associated significantly with diabetes in Finnish and German populations [5]. The identification of NIDDM1 demonstrates the following: 1) several susceptibility alleles can be found in one gene; 2) some combinations of SNP increase risk of development of diabetes while other combinations decrease risk; 3) risk that cannot be associated with a single SNP can be identified by a haplotype including the SNP; and 4) SNPs in introns affect regulation of the transcription level of the gene.

**Genetic analysis of NIDDM1 in various populations**

An analysis of NIDDM1 including SNP-44, which is located near SNP-43 and affects transcripational activity, in British and Irish whites found a significant association of SNP-44 with type 2 diabetes. SNP-44 showed significant LD transmission singly and was in complete LD with the missense mutation T504A. Thus, proteins with mutations or altered transcription expression levels contribute to the development of type 2 diabetes in this population [6]. A recent meta-analysis with additional genotyping in 4213 individuals (2056 type 2 diabetes patients and 2157 healthy individuals) found the O.R. of the development of diabetes to be 1.17 (95% C.I., 1.02 to 1.34) for SNP-44 with 80% statistical power [7]. Another large-scale meta-analysis involving 2288 type 2 diabetes patients and 3041 healthy individuals found O.R. of 1.19 (95% C.I., 1.07 to 1.33) for SNP-43 alone [8]. Racial differences were reported in the association of three SNPs (SNP-43, -19, and -63) with the incidence of diabetes [9–16]. An analysis in Japanese found no significant association between diabetes and these haplotype combinations [17]. In addition, a case-control association analysis in nearly 1000 patients and controls found that the minor allele of SNP-19 and the 121 haplotype was associated with reduced risk in Japanese diabetes patients aged 50 years or over [18]. Unfortunately, most of the original studies included in the these meta-analyses were based on genotyping only SNP-43, SNP-19, and SNP-63, without consideration of LD blocks specific to each ethnic group. Thus, the frequencies of the haplotypes in each population remain to be carefully assessed. For example, the pattern of LD in CAPN10 was evaluated by calculation of \( r^2 \) both in Japanese and in Mexican Americans from the genotype data using 17 SNPs for all possible pairs with 96 control subjects (Fig. 2A & B). The distribution of LD was similar in the two populations, and at least four major SNP subgroups with minor differences were present. Since more SNPs are found in tight linkage with each other in Japanese than in Mexican Americans (Fig. 2B), Japanese may have higher LD in this locus.

To clarify the ancestral role of these high-risk haplotypes in the development of diabetes, a genotyping survey was undertaken in human individuals from various populations and other primates. The presence of positive natural selection at the calpain-10 locus that cannot be explained by genetic drift has previously been established [19]. Comparison of human individuals from various populations with primates revealed that the 111-haplotype was likely to carry the ancestral allele in all populations, while the 112-haplotype was selectively favored in African populations, and the 121- and 221-haplotypes were selected in populations outside Africa in the process of racial migration toward Europe, Asia, and America. In addition, the survey showed that the four major haplotypes (111-, 112-, 121-, and 221-) occur in Native Americans, an ancestral population of Mexican Americans, and that recent admixture between populations was not a factor [19]. Haplotypes with minor allele SNP-44 have almost no other polymorphisms, suggesting recent and rapid positive selection of these haplotypes. According to the common disease common variant hypothesis, major alleles can be disease susceptibility alleles, represented by PPAR\(\gamma\) (Pro12Ala) for type 2 diabetes [20]; minor alleles as susceptibility alleles are represented by ApoE (\(e4<e3\)) for coronary artery diseases [21] or Alzheimer’s disease [22] and the calpain-10 gene (SNP-44 C<T) for type 2 diabetes [6, 7]. For the latter two genes, the major allele may have originally been beneficial by reducing the risk of common metabolic syndromes but become detrimental relatively recently.

Summary statistics in various populations for polymorphisms on the 33465 base pairs in the NIDDM1 region including calpain-10 and GPR35 show a higher frequency of mutation at the calpain-10 locus than at the GPR35 locus. In addition, sliding window analysis reveal a high frequency of polymorphisms in intron 13 of calpain-10 that cannot be fully characterized by the neutral hypothesis. Simulation analysis indicates that this locus cannot be explained with a simple two-allele
selection model but requires a multiallele selection model with sequential turnover of polymorphisms. Thus, a complicated population structure other than natural selection may be involved in this feature, but that alone cannot explain the rapid decline in LD. For loci such as intron 13 of calpain-10 that exhibit a rapid decline in LD, a high recombination rate is generally reported. Although no studies have reported a correlation between the recombination rate and the frequency of polymorphisms, exceptional recombination or mutation in a neutral state is possible [23]. Recently, a close analysis of intron 13 showed that a selection model consisting of five clusters of haplotypes can explain the genetic findings in Mexican Americans, two of the five clusters appearing 2 to 3 million years ago.

Fig. 2. A. Exon-intron organization of CAPN10. The physical distance between SNP-118 and SNP-32 is approximately 40 kb. The locations of the SNPs are shown. B. Pairwise linkage disequilibrium in CAPN10 evaluated by \( r^2 \). All SNP numbers are denoted in Fig. 2A. Pairwise LD was determined using 136 and 105 marker pairs in Japanese (left panel) and Mexican Americans (right panel), respectively. SNP-134 and -135 were not identified in the previous study with Mexican Americans [5]. The color gradations from red (perfect LD, i.e., \( r^2 = 1 \)) to blue (no LD, i.e., \( r^2 = 0 \)) reflect the degree of the observed LD. Subgroups of SNPs found in tight linkage (\( r^2 > 0.5 \)) also are shown. Since there was no suitable polymorphic site upstream of CAPN10 [5], we used SNP-118 and -66 in the study, despite their low allele frequencies.

Fig. 3. Domain structures of the human calpain family. Typical calpains are composed of four domains (I–IV), but in the case of atypical calpains, certain domains have been deleted or replaced. The small subunit of calpain is composed of two domains.
in the glacial ages. Comparisons of the partial sequence at intron 13 among 10 species show that four USF1 and one HNF1 binding sites in a segment share common sequence motifs. In addition, an unknown gene is expressed around the segment that might influence calpain-10 expression [24].

Calpains

Calpains are a family of cytoplasmic cysteine proteases activated by Ca\(^{2+}\). At least 15 calpains have been identified as eight typical calpains, six atypical calpains, and one small subunit of calpain [25, 26]. The domain structures of the calpain family are shown in Fig. 3. Known substrates for calpains include cytoskeletal proteins, actin-binding proteins, calmodulin-binding proteins, hormone receptors, cell membrane hormone receptors, glucose-metabolizing enzymes, enzymes regulating signal transmission, and transcription factors. Calpains are known to play a physiological role in reconstruction of cytoskeleton, apoptosis, and reconstruction (proliferation, differentiation, and transformation) of tissue cells. Mutations in genes encoding calpains cause various disorders including diabetes related to calpain-10, gastric cancer related to calpain-9, muscular dystrophy related to calpain-3, neurodegenerative diseases (e.g., Alzheimer’s disease), cerebral infarction, spinal injury, myocardial infarction, hepatic ischemia, and renal impairment. Animal tests show death in the fetal period in calpain-4 (small subunit) knockout mice due to impaired development of the cardiovascular system. Calpain-1 knockout mice develop normally because the lost function of calpain-1 is compensated by calpain-2, but the mice often show platelet aggregation disorder.

Calpain-10

The human calpain-10 gene is located in chromosome band 2q37.3 and consists of 15 exons. There are at least eight isoforms (calpain-10a to calpain-10h) of the gene. The longest isoform, calpain-10a, consists of 672 amino acids. Calpain-10 is an atypical calpain that lacks domain IV and instead has a tandem linking domain, domain III. Calpain-10a is expressed most strongly in the heart, but is present in various tissues including those playing an important role in glucose metabolism, including liver, muscle, pancreatic islets, and adipocytes [5, 25]. Although calpain-10c and 10g can be detected in many tissues, calpain-10b, 10d, 10e, and 10f are much less abundant [5]. Because calpain-10 lacks calcium-binding sites in domains II and III, it is not known whether the protein is activated by calcium. Calpain-10 may react with calcium in a separate mechanism, as it can be found in sarcomere, the calcium storage in muscle fibre, and its expression is increased or its distribution is altered following calcium stimulation in the epithelium of crystalline [27].

Effects of calpain-10

Effects of calpain-10 on β cells

1) Calpain and apoptosis

Involvement of calpain-10 in ryanodine-induced apoptosis was reported based on finding that apoptosis was enhanced by ryanodine or palmitin acid in pancreatic β cell-specific calpain-10 transgenic mice by the rat insulin promoter but was not enhanced in calpain-10 knockout mice [28].

2) Calpain and insulin secretion

The relationship between calpain and insulin secretion was assessed with calpain inhibitors. Assessed with a nonspecific calpain inhibitor, glucose-responsive insulin secretion was enhanced in short-term culture with the inhibitor added, while insulin secretion was inhibited by reduced mitochondrial glucose metabolism in 48-hour culture with the inhibitor added [29, 30].

Calpain-1, a typical calpain, is reported to break ICA512, a tyrosine phosphatase-like protein located in insulin granules, dependent on the intracellular Ca\(^{2+}\) concentrations. A calpain inhibitor blocked the breakage, thereby impairing insulin secretion [31]. In a recently reported assay system in which calpain-10 was overexpressed stably in INS1 cells, glucose-responsive insulin secretion was enhanced. The breakage was reduced by the addition of E64, a calpain inhibitor, to the cells. When Ca\(^{2+}\) was removed from the supernatant, these reactions were not induced. These results indicate that the intracellular Ca\(^{2+}\) concentration increased by glucose stimulation can activate calpain-10 and break SNAP25 (a SNARE protein), thereby inducing fusion of insulin granules to the membrane of the β cells [32]. However, considering the
absence of key 54-kD proteins among the reported calpain-10 isoforms, further analysis is required.

Effects of calpain-10 on muscles

In clinical research assessing the relationship between SNP-43G/G and the level of calpain-10 mRNA expression in Pima Indians, individuals with the SNP-43G/G allele were found to have a low level of expression. A significant association was also found between oxidative utilization of glucose and the level of calpain-10 expression [33]. In an animal study using mice expressing calpastatin (a calpain inhibitor) in muscles, muscular glucose uptake and general glucose tolerance remained unchanged despite increased expression of GLUT4 protein and muscular hypertrophy mainly due to loss of insulin action resulting from reduced AKT kinase activity [34]. However, while calpastatin inhibits the activity of calpain-1 and calpain-2, action on calpain-10 is unlikely since calpain 10 does not bind with the calpain small subunit (calpain-4), the target of calpastatin. In an in-vitro study with human myoblasts to assess the relationship between calpain-10 and muscle differentiation, 60 kD calpain-10 protein levels were found to increase as differentiation progressed [35]. As differentiation of the myoblasts progressed, calpain-1 levels also increased and calpastatin expression decreased. In L8 cells over-expressing calpastatin, muscle differentiation was inhibited [36–38]. These findings demonstrate the involvement of calpain-10 as well as the calpain-calpastatin system in the process of myoblast differentiation.

Effects of calpain-10 on adipocytes

Some groups have investigated the function of adipocytes in association with calpains, but no relationship between SNP43 and obesity was noted. In an in-vitro study using human adipocytes, adipogenesis in adipocytes was increased in a group with the SNP-43G/G allele irrespective of the level of GLUT4 expression. Lipolysis function from β3 adrenalin receptors was reduced in a group with the SNP19-deficient allele to one-thirtieth that in the normal group [39, 40]. In a clinical study, the level of calpain-10 mRNA expression in adipocytes decreased dependent on the blood triglyceride level in obese patients with the SNP43G/G allele [41]. In calpain-10 antisense-expressing stable cell line 3T3L1 adipocytes, actin reconstruction was inhibited by insulin with an unchanged level of GLUT4 expression, resulting in reduced glucose uptake and inhibited transfer of GLUT4 to the membrane by insulin [42]. As in other reports on the relationship between calpains and adipocytes, the level of calpain-1 expression increased while calpastatin expression decreased as adipocyte differentiation progressed, and when the action of calpain was inhibited by forced expression of calpastatin or the addition of the calpain inhibitor, C/EBPα expression and adipocyte differentiation were inhibited [43].

Clinical assessment of calpain-10

An analysis in nondiabetic British subjects revealed that genetic variation in the CAPN10 gene influences blood glucose levels and that this is, at least in part, due to the effects of calpain-10 on early insulin secretory response [44]. An analysis in Finns showed that individuals with the 1121/1121-haplotype combination for SNP-44, -43, -19, or -63 have approximately two times higher risk of development of diabetes, and that SNP-43 is associated with high fasting insulin levels, high HOMA-R levels, and high fasting fatty acid levels [45]. An analysis in Japanese found no significant association between diabetes and these haplotype combinations, but did find an association with insulin resistance and high fatty acid levels under euglycemic hyperinsulinemic clamp in individuals with the 112/121 haplogenotype [17] (Table 1).

Relationship between calpain-10 and diabetes-related diseases

For polycystic ovary syndrome, no phenotypic differences were noted among non-diabetic European Americans with a single polymorphism or haplotype of SNP-43, -19, or -63. In non-diabetic African American probands, no single polymorphism of SNP-43, -19, or -63 was associated with any phenotype, but individuals with the 112/121-haplotype combination showed a significantly greater area under the insulin-time curve on oral glucose tolerance test. This result was evident after data adjustment for body mass index. In African Americans and European Americans, the 112/121-haplotype combination was associated with approximately two times higher risk of polycystic ova-
In Spanish, an association between UC SNP44 and the incidence of polycystic ovary syndrome was reported [47].

Protease inhibitors and diabetes

Protease inhibitors are effective in the treatment of HIV infection. However, long-term use appears to induce pathology of hyperlipidemia or diabetes, including peripheral fatty atrophy and central fatty hypertrophy [48]. The mechanism is thought to be inhibited adipocyte differentiation. Calpain inhibitors, a type of protease inhibitor, are considered in the treatment of several disorders including spinal injury, liver transplantation, and myocardial infarction. Since calpain inhibitors inhibit adipocyte differentiation, attention must be paid to the possibility of diabetes and hyperlipidemia when they are used.

Conclusions

Ethnic comparison of polymorphisms can clarify the association of genetic structure with disease and play an important role in supporting case-control results. The approach is especially useful in screening putative susceptibility genes for common diseases that may have undergone recent natural selection, including those for type 2 diabetes. Sequence determination of a naturally selected segment can both identify a susceptibility gene and the mechanism of its regulation. Because of the recent, remarkable progress in sequencing techniques, identification of all of the polymorphisms on the entire human genome, SNPs in particular, may soon be possible. However, the enormous cost of SNP typing remains a limiting factor for their use in investigations of ethnic variants in common diseases. Association studies based on haplotype analysis can identify susceptibility genes, calculate developmental risks, and predict drug responsiveness. Translational research on function using interactome, proteome, and model mice can be used to apply the results clinically as individualized therapy.

Acknowledgement

We thank S. Oike, R. Kawakami, Y. Yaginuma, I. Uda, J. Tsutsumi, Y. Ibe, and T. Takahashi for assistance. The study was supported by Grant-in-Aid for Scientific Research and for Scientific Research on Priority Areas (C) “Medical Genome Science” from the Japanese Ministry of Science, Education, Sports,

Table 1. Association studies of CAPN10 in various populations

<table>
<thead>
<tr>
<th>Population</th>
<th>Genotype</th>
<th>Odds Ratio</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexican American</td>
<td>112/121</td>
<td>3.02 (1.37–6.64)</td>
<td>Glucose utilization ↓</td>
</tr>
<tr>
<td>Pima Indian</td>
<td>SNP-43 G/G</td>
<td>111/111</td>
<td>2.04 (1.22–3.39)</td>
</tr>
<tr>
<td>British/Irish Whites</td>
<td>2111/2111</td>
<td>2111/1111</td>
<td>2.36 (1.19–4.66)</td>
</tr>
<tr>
<td></td>
<td>112/121</td>
<td>112/121</td>
<td>1.42 (0.68–2.98)</td>
</tr>
<tr>
<td>Samoans</td>
<td>111/221</td>
<td>1.48 (1.06–1.91)</td>
<td>Insulin AUC ↑</td>
</tr>
<tr>
<td></td>
<td>SNP-19-63</td>
<td>112/121</td>
<td>2.18 (1.06–4.45)</td>
</tr>
<tr>
<td>African-American</td>
<td>SNP-43 G/G</td>
<td>112/121</td>
<td>2.57 (1.22–5.44)</td>
</tr>
<tr>
<td>Spanish</td>
<td>SNP-44 CC, TC</td>
<td>121/121</td>
<td>1.93 (1.03–3.54)</td>
</tr>
<tr>
<td>Finnish</td>
<td>1121/1121</td>
<td>1.93 (1.07–3.47)</td>
<td>Fasting Insulin HOMA-RFFA ↑</td>
</tr>
<tr>
<td>Japanese</td>
<td>112/121</td>
<td>121/121</td>
<td>Protective against T2DM ↑</td>
</tr>
</tbody>
</table>

The results of studies with various populations are shown in this slide. The major (1) and minor (2) alleles are denoted as in Mexican Americans.

HORIKAWA

Culture and Technology; a Health and Labor Science Research Grant for Research on Human Genome and Tissue Engineering from the Japanese Ministry of Health, Labor and Welfare; and the Naito Foundation.

References


