Linkage of Calpain 10 to Type 2 Diabetes
The Biological Rationale
Nancy J. Cox,1,2 M. Geoffrey Hayes,1 Cheryl A. Roe,1 Takafumi Tsuchiya,3 and Graeme I. Bell1–3

The follow-up studies to the original report of association of variation at calpain 10 (CAPN10) with type 2 diabetes in the Mexican-American population of Starr County, Texas, encompass a broad range of science. There are association studies on genetic variation at CAPN10 in different human populations over a range of phenotypes related to type 2 diabetes, physiological studies on the biological functions of calpain proteases, and evolutionary studies on CAPN10 and the NIDDM1 region. We review here the studies published to date on CAPN10, as well as the latest findings from positional cloning studies on a number of other complex disorders. Collectively, these studies provide perspective on the challenges of moving from the linkage mapping and positional cloning studies on which we have been focused to an understanding of the biology shaping the relationship of genotype to phenotype at loci influencing susceptibility to complex disorders like type 2 diabetes. Diabetes 53 (Suppl. 1):S19–S25, 2004

OVERVIEW OF STUDIES ON CAPN10 IN TYPE 2 DIABETES AND RELATED PHENOTYPES

Of the studies containing new primary data on CAPN10 or the biological effects of calpain proteases on phenotypes related to glucose homeostasis published since CAPN10 was proposed as a diabetes susceptibility locus, a number provide support for the association of polymorphisms and/or haplotypes with type 2 diabetes and/or related quantitative phenotypes consistent with the hypotheses put forward in the initial report (1). These include studies in a British population (2), studies in a Finnish population (3), studies in a population of South Indians (4), studies in a Utah population of European descent (5), studies in Pima Indians (6), and studies in African Americans (7).

Lynn et al. (2) found that non-diabetic British subjects with the high-risk haplotype combination (Fig. 1) had increased fasting and 2-h plasma glucose levels. This same variation was also associated with a significant decrease in the insulin secretory response adjusted for insulin resistance. Cassell et al. (4) reported that the high-risk haplotype combination was significantly associated with type 2 diabetes in a population of South Indians (odds ratio [OR] 6.52, 95% CI 1.32–35.3, for a series of unrelated case and control subjects; 5.78, 1.18–31.2, for a set of unrelated case subjects ascertained from families segregating for type 2 diabetes). Garant et al. (7) reported results of typing on one of the key polymorphisms (UCSNP-43) in African Americans, with homozygosity for the G allele being associated with an increased risk of type 2 diabetes (OR 1.38, 95% CI 1.04–1.83).

The original studies (1) reported that individuals with a combination of two different haplotypes (121/112; Fig. 1) were at higher risk for type 2 diabetes, but found no evidence that individuals homozygous for either the 112 or the 121 haplotype were at high risk for type 2 diabetes. In contrast, a number of studies in European populations, notably the studies of Orho-Melander et al. (3) in a Finnish population and of Malecki et al. (8) in a population from Poland, report associations between type 2 diabetes and the homozygous 121 haplotype. Orho-Melander et al. (3) observed increased frequencies of the G(1) allele at UCSNP-43 and the T(2) at UCSNP-63 in Finnish subjects with type 2 diabetes compared with control subjects. In addition, they reported a significantly increased risk for type 2 diabetes in individuals homozygous for the 121 haplotype (OR 1.93, 95% CI 1.07–3.47), and a similar risk for individuals with the high-risk haplotype combination (1.85, 0.90–3.81). The G allele at UCSNP-43 was also associated with higher fasting free fatty acid concentrations. In the Malecki et al. study (8), individuals homozygous for the 121 haplotype had a significantly increased risk of type 2 diabetes (1.93, 1.03–3.54). Individuals with the 121/112 haplotype combination found to be high risk in Mexican Americans were not at increased risk of type 2 diabetes in this Polish population (0.93, 0.39–2.23).

Elbein et al. (5) reported a significantly increased risk for individuals with the 111/221 haplotype combination at CAPN10 (1.48, 1.06–1.91) in their studies on a Utah
population of European descent. Additional studies on quantitative phenotypes in the unrelated members of the families studied indicated that homozygosity for the G allele at UCSNP-43 was associated with increased mean glucose. Variation at polymorphisms 19 and 63 showed association with a variety of the quantitative traits examined (polymorphism 19 with fasting insulin, 60-min glucose, insulin resistance by homeostasis model assessment, and $s_i$ determined by minimal-model analysis; polymorphism 63 with total insulin area under the curve, 90-min insulin, and 2-h insulin). Studies in Pima Indians on UCSNP-43 (6) are also consistent with a role for variation at $CAPN10$ in insulin sensitivity, with nondiabetic individuals homozygous for the G allele at UCSNP-43 found to have decreased rates of postabsorptive and insulin-stimulated glucose turnover that apparently result from decreased rates of glucose oxidation. Individuals homozygous for the G allele also showed significant alterations in nutrient partitioning, with preferential oxidation of fat when given a diet of mixed nutrient composition.

Two studies also reported no replication of the primary findings on the high-risk haplotypes or the individual variants comprising them, but report significant associations with amino acid polymorphisms in $CAPN10$ (9,10). Evans et al. (9) examined complete trios ascertained in the U.K. and found that none of the originally identified polymorphisms or haplotypes were overtransmitted from heterozygous parents to affected offspring. In contrast, the rarer C allele at UCSNP-44 was transmitted to affected offspring significantly more frequently than expected. This allele was noted to be in perfect linkage disequilibrium with an amino acid polymorphism, Thr504Ala. The frequency of the rarer overtransmitted allele at these sites was 0.16 in the data of Evans et al., but only 0.04 in the Mexican-American control subjects and 0.09 in Mexican-American case subjects originally reported (1). Horikawa et al. (10) also found no evidence for association of the variation at $CAPN10$ defining the high-risk haplotype combination with type 2 diabetes in a Japanese population. In studies conducted within subsets of patients matched for age, sex, BMI, and duration of disease, the high-risk haplotype combination was associated with higher levels of glucose-induced serum insulin at 60 min under a hyperglycemic clamp and with serum free fatty acid concentrations at the end point of a euglycemic clamp study. In addition, they observed a significant association with the rarer allele at an amino acid polymorphism, Phe200Thr, that was not observed in Mexican Americans.

Other studies provided no support for association of $CAPN10$ variation with diabetes. In a second study in the Finnish population, Fingerlin et al. (11) reported no association of variation at $CAPN10$ with type 2 diabetes, with ORs for the high-risk haplotype combination of 0.86 (95% CI 0.52–1.42) for FUSION1 (Finland-U.S. Investigation of NIDDM 1) case subjects and 0.72 (0.38–1.37) for FUSION2 case subjects. A number of associations of quantitative traits with individual polymorphisms, haplotypes, and/or haplotype combinations at $CAPN10$ were observed for these data (information available at http://www.sph.umich.edu/csg/CAPN10), but the number of significant associations was not necessarily any larger than might be expected given the number of comparisons. Hegele et al. (12) found no significant association of the single polymorphism they examined at $CAPN10$ (UCSNP-43) and type 2 diabetes in Oji-Cree, although the OR (1.30, 95% CI 0.93–1.81) was similar to that observed for studies in African Americans (7), when also examining only UCSNP-43. Similarly, Rasmussen et al. (13) reported no significant association of the high-risk haplotype combination in a Danish population (1.32, 0.60–2.89) and Tsai et al. (14) reported no significant association of the high-risk haplotype combination in a Samoan population (1.42, 0.68–2.98).

Thus, whereas many of the studies on variation at $CAPN10$ provide support for the general hypothesis that variation at $CAPN10$ can affect susceptibility to type 2 diabetes and related phenotypes, results of the studies on the high-risk haplotype combination and risk of type 2 diabetes have yielded variable results.

VARIABILITY IN ORs IN EUROPEAN SAMPLES

Closer examination of the numbers used in calculating ORs for populations of European descent (3,8,9,11,13), the most homogeneous of those studied on $CAPN10$ to date, reveals that there is much better agreement in the estimates of the proportion of diabetic individuals with the high-risk haplotype combination (0.056–0.073) than in the estimates of the proportion of control individuals with the high-risk haplotype combination (0.013–0.092). Although it seems clear that the variability in estimated ORs for populations of European descent (0.72–4.95) is due largely to the higher variability of the estimated proportion of control subjects with the high-risk haplotype combination, it is much less clear why this should be so. In general, case subjects from different studies all met the same criteria for diagnosis of type 2 diabetes but varied in how they were ascertained (e.g., as subjects in complete trios, as subjects in families with at least two sibs affected with type 2 diabetes, as subjects ascertained from clinic populations, etc.). In contrast, the control samples vary considerably both in the “diagnostic criteria” for control subjects and in their ascertainment. Some control samples might be better classified as random samples because the individuals were not examined with respect to diabetes status (9), whereas others were extensively phenotyped (13).

There is little correlation in the estimated proportion of control subjects with the high-risk haplotype combination with either the sample size or the geographic location of the sample. The samples of Orho-Melander et al. (3) from Finland and of Rasmussen et al. (13) from Denmark are relatively large (296 and 200, respectively) and have lower estimates of the proportion of control subjects with the high-risk haplotype combination (0.037 and 0.045, respectively). The Finnish samples of elderly and spouse control subjects that make up the total control sample reported in the FUSION studies (11) are also relatively large (207 and 174, respectively) and have higher estimates of the proportion of control subjects with the high-risk haplotype combination (0.092 and 0.075, respectively). These are all Northern European populations, and the estimates of the proportion of patients with type 2 diabetes having the high-risk haplotype combination are quite similar in these studies (0.061–0.072).
**VARIATION AT CAPN10 AND OTHER PHENOTYPES**

Several studies examined phenotypes less directly related to type 2 diabetes. Ehrmann et al. (15) reported association of the high-risk haplotype combination at CAPN10 with polycystic ovary syndrome (PCOS) and quantitative measures related to type 2 diabetes. Gonzalez et al. (16) also reported an association of CAPN10 variation with PCOS, and Escobar-Morreale et al. (17) reported an association of the variation with hirsutism. In contrast, Haddad et al. (18) reported no association of CAPN10 variation with PCOS. Shore et al. (19) reported an association of CAPN10 variation with microvascular function. Hoffstedt et al. reported association of CAPN10 variation with \( \beta_1 \)-adrenoceptor function in human fat cells (20) and with glucose metabolism in human fat cells (21). Daimon et al. (22) reported association of CAPN10 variation with serum cholesterol in Japanese. Hinney et al. (23) reported no association of CAPN10 variation with early-onset obesity.

**VARIABILITY IN EXPRESSION OF CAPN10 ASSOCIATED WITH GENETIC VARIATION AT CAPN10**

Yan et al. (24) examined allele-specific variability in expression of a number of genes implicated in human disease, including calpain 10, using cell lines from CEPH families. Their method requires use of variation within a coding sequence. They reported a 1.8-fold difference in expression of calpain 10 depending on the allele found at a synonymous polymorphism at CAPN10 (UCSNP-48). The UCSNP-48 site is in strong linkage disequilibrium with one of the sites (19) included in the polymorphisms defining the haplotype combination that is high risk for type 2 diabetes in the original studies on Mexican Americans. Ongoing studies are focused on identifying the particular variation(s) affecting this expression phenotype.

Baier et al. (6) reported that in vivo expression of calpain-10 mRNA in skeletal muscle is significantly lower (53% reduction in transcript levels) in individuals homozygous for the G allele at UCSNP-43. In addition, they observed a significant correlation \((r = 0.79, P = 0.003)\) between the level of CAPN10 expression in skeletal muscle and the rate of carbohydrate oxidation measured in a respiratory chamber over 24 h.

**PHYSIOLOGICAL STUDIES**

Sreenan et al. (25) examined the effects of calpain inhibitors, including calpain inhibitor II and E-64-d, on a variety of phenotypes important in glucose homeostasis. Exposure of mouse pancreatic islets to these inhibitors of calpain activity increased the insulin secretory response to glucose, due apparently to accelerated exocytosis of insulin granules. Exposure of muscle strips and adipocytes to calpain inhibitors led to a reduction in insulin-mediated glucose transport. Whereas these calpain inhibitors are not specific for the calpain 10 protease, but rather are likely to inhibit most if not all calpain proteases, the results of the studies are consistent with the possibility that calpain proteases play a role in glucose homeostasis.

A variety of studies on the effects of variation in CAPN10 in nondiabetic individuals also provide some insight into the physiological function of the calpain 10 protein. Several of the studies described in the above section provide evidence that the high-risk haplotype combination or one or more of the key individual polymorphisms at CAPN10 is associated with measures of insulin resistance (2,5,6). In addition, Stumvoll et al. (26) reported that nondiabetic German subjects homozygous for the G allele at UCSNP-43 show higher first-phase insulin secretion than those with other genotypes. Tschröter et al. (27) developed an approach for assessing the shape of the glucose curve during an oral glucose tolerance test and report that variation at UCSNP-44 shows association with the value of the composite measure they used to characterize the variability they observed in the shape of the glucose curve.

**SELECTION AND GENETIC VARIATION AT CAPN10**

The hypothesis that the same variation that now increases risk for type 2 diabetes may once have enjoyed a selective advantage (28,29) due to the relatively greater likelihood of individuals with such variation to survive the periodic famines affecting ancient human populations has long been an intriguing one. Moreover, the ability to use genomic features (the pattern and distribution of variation as well as the extent of linkage disequilibrium) to test for and localize a signature of natural selection offers a unique opportunity to identify regions of interest within a locus outside of the traditional disease association approaches. Fullerton et al. (30) reported preliminary studies on the population genetics and evolutionary history of variation at CAPN10. In studies of 589 individuals from 11 populations (African, European, Middle Eastern, Asian, Oceanian, North American, and South American), two of the three polymorphisms identifying the high-risk haplotype combination (polymorphisms 19 and 63) were found to have higher \( F_{ST} \) values (\( F_{ST} \) is a measure of allelic diversity) than had been observed for any of the 86 randomly chosen bi-allelic markers used as a baseline reference sample. These large \( F_{ST} \) values primarily reflect differences in haplotype frequencies between African and non-African populations and are broadly consistent with natural selection, having shaped, at least in part, the patterns of variation at CAPN10. Ongoing studies on the evolutionary history of variation at CAPN10 are focusing on regions within the gene showing patterns of variation atypical for regions evolving under neutrality.

It was also noted that the results of studies on the phenotypic effects of variation at CAPN10 (UCSNP-43) in the Pima Indians (6) were particularly congruent with hypotheses on metabolic responses to famine. Based on the Pima studies, individuals who are GG homozygotes at UCSNP-43 have reduced glucose oxidation but not glucose storage. This should preserve skeletal muscle protein and skeletal muscle glycogen that might fuel high-energy muscle contraction. A decreased rate of gluconeogenesis, indicated by the decreased sleeping metabolic rate, further reduces caloric requirements. And preferential oxidation of lipid (at the expense of protein and carbohydrate) enables an easier switch to ketone production to fuel the brain when carbohydrate supply is limited.

**THE GENETIC MODEL RELATING VARIATION AT CAPN10 AND SUSCEPTIBILITY TO TYPE 2 DIABETES**

For each of the two amino acid polymorphisms at CAPN10 associated with type 2 diabetes (Thr504Ala and Phe200Thr),
it is the rarer allele that increases risk in a dominant or additive fashion. The more common noncoding sequence variants at CAPN10 (polymorphisms 43, 19, and 63) are hypothesized to affect susceptibility to type 2 diabetes by altering expression of the calpain-10 mRNA and/or protein. These three noncoding sequence polymorphisms are not considered as separate and independent disease susceptibility alleles, but as part of a combination of variants needed to characterize the risk of disease. The highest risk for type 2 diabetes in Mexican Americans is observed in individuals heterozygous for two different haplotypes (Fig. 1), both of which are common in the Mexican-American population. Homozygotes for either haplotype show no increased risk in Mexican Americans, although studies in both Finland (3) and Poland (8) reported increased risk associated with homozygosity for the more common of these two haplotypes (121).

The variation at CAPN10 that best differentiates risk of type 2 diabetes also shows a relationship to age at diagnosis in Mexican Americans. Figure 2 summarizes the cumulative percent of Mexican Americans with type 2 diabetes by age at diagnosis. Data include all subjects from the original report (1) (■), subjects with the high-risk haplotype combination (121/112) (▲), and subjects with the lowest-risk haplotype combination (112/221) (●). Age at diagnosis has long been used in studies on diabetes to identify more homogeneous subgroups of patients and families, and it is assumed that earlier age at diagnosis is associated with higher genetic liability to type 2 diabetes. Thus, it is perhaps not surprising that the 221/112 haplotype combination associated with a significantly reduced risk of type 2 diabetes in both the original and replication samples of Mexican Americans is also associated with later age at diagnosis of type 2 diabetes.

FIG. 2. Cumulative percent of Mexican Americans with type 2 diabetes by age at diagnosis. Data include all subjects from the original report (1) (■), subjects with the high-risk haplotype combination (121/112) (▲), and subjects with the lowest-risk haplotype combination (112/221) (●).

MOLECULAR GENETIC MODELS FOR COMPLEX DISORDERS
The original report characterizing the nature and particular patterns of variation at CAPN10 most strongly associated with disease and the evidence for linkage was considered to depart radically from expectations derived from the identification of causal variation at simple Mendelian disorders. None of the key variation at CAPN10 encoded amino acid polymorphisms (although some amino acid polymorphisms may be susceptibility alleles), multiple polymorphisms were needed to characterize risk, and the overall model was complex. As it turns out, the findings at CAPN10 were simply the first of an ongoing set of challenging observations to arise in the context of positional cloning studies on complex disorders.

The NOD2 locus was identified as a susceptibility gene for Crohn’s disease in both positional cloning (31) and candidate gene studies (32,33). A number of mutations at NOD2 have been identified (including amino acid polymorphisms and truncating mutations) that are reproducibly associated with Crohn’s disease in multiple populations (33). However, it was clear from the
beginning that the mutations identified at NOD2 were probably not of sufficient frequency, even in aggregate, to account for the evidence for linkage in this region of chromosome 16, and subsequent studies have confirmed this (34). Sugimura et al. (35) reported an extended haplotype at NOD2 strongly associated with Crohn’s disease in an Ashkenazi Jewish populations. This haplotype is rare in other populations and does not contain any of the known mutations at NOD2 or variation of obvious functional consequence. Thus, there may be variation at NOD2 outside the known mutations that affects susceptibility to Crohn’s disease. Alternatively, the associated haplotype may be in linkage disequilibrium with variation at a gene other than NOD2, and it is entirely possible that there is both additional noncoding sequence variation within NOD2 and variation at one or more additional genes in the region that all affect susceptibility to Crohn’s disease.

A positional cloning study focused on a locus for asthma and bronchial hyper-responsiveness mapped to chromosome 20 recently implicated variation at the ADAM33 gene in these phenotypes (36). These studies identified primarily noncoding sequence variation as the variation affecting phenotype, and haplotypes comprised of two noncoding sequence polymorphisms predicted risk substantially better than individual sites. Subsequent reports on the progress in these studies confirm these preliminary findings and provide the additional information that the key haplotypes are actually comprised of polymorphisms in different haplotype blocks within the gene (37). Thus, it is unclear whether the observed association reflects a true haplotype effect or a type of allelic epistasis. It is unlikely, however, that the observed haplotype effects might be attributed to unstudied variation within ADAM33 that might be found preferentially on the associated haplotypes. All moderately common variation has been identified and tested. Whereas the key variation at ADAM33 is intrinsic, at least some of the implicated variation is located in regions with a higher than expected level of sequence similarity between the human and mouse genomes, increasing the likelihood that the surrounding sequence has functional consequence. There are as yet only limited reports on results of replication studies (38), but the preliminary reports have been positive.

A variety of loci have now been implicated in studies on schizophrenia, including neuregulin (NRG1 [39]) and dysbindin (DTNBP1 [40]). For these loci, the variation implicated in the risk is located in the noncoding sequence, and multiple polymorphisms and/or extended haplotypes are required to characterize the associations with disease. These studies included extensive survey of the variation in the surrounding region, allowing exclusion of the possibility that there is a coding sequence variation that is more strongly associated with disease than the extended haplotypes. For example, it was explicitly noted in the studies on NRG1 that all coding sequences and all conserved (with respect to the mouse genomic sequence) noncoding sequences had been screened without finding individual sites as strongly associated with schizophrenia as the extended haplotypes. However, in the larger physical regions delineated by the extended haplotypes for other genes (e.g., DTNBP1), ongoing studies continue to screen for and test variation in the region. Some of the variation implicated in susceptibility to schizophrenia has been replicated (41,42), and studies are ongoing.

CONCLUSIONS

The findings from these positional cloning studies in different common diseases challenge simplistic expectations of how genetic variation will affect complex phenotypes. Clearly, noncoding sequence variation will be found associated with common diseases with complex inheritance. Assaying the functional consequence of such variation will be difficult, although novel comparative genomic approaches can enhance our ability to identify functional elements in coding and noncoding sequence as well as improve our ability to assay functional consequence by highlighting control regions unlikely to be functional (43). The molecular models for common diseases are also likely to be complex within a single gene, involving multiple polymorphisms whose collective effects, whether considered at the level of the genotype or the haplotype, are difficult to predict given the marginal contributions of the individual sites. This aspect of model complexity is a sufficiently radical departure from the simple two-allele disease models that provided the foundation for the power studies commonly used to justify study designs and analytic strategies for linkage and association studies that we need to consider the consequences. In particular, it seems likely that consideration of simple two-allele genetic models leads us to underestimate the variance in the expected linkage signal for a given sample size and locus effect size relative to what would be obtained for genetic models allowing for the observed level of molecular complexity. Similarly, a consideration of these more complex molecular models should inform our decisions on study design and analytic strategies in association studies.

ACKNOWLEDGMENTS

This research was supported in part by U.S. Public Health Services Grants DK-55889, DK-20595, DK-47486, and DK-58026. G.I.B. is an investigator in the Howard Hughes Medical Institute.

REFERENCES

5. Elbein SC, Chu W, Ren Q, Hemphill C, Schay J, Cox NJ, Hanis CL, Hasstedt...


