Melanocortin 4 Receptor Gene Variation Is Associated With Severe Obesity in Pima Indians

Lijun Ma, P. Antonio Tataraanni, Clifton Bogardus, and Leslie J. Baier

Heterozygous coding mutations in the melanocortin 4 receptor gene (MC4R) are the most common genetic causes of severe human obesity identified to date. To determine whether MC4R has a role in causing severe obesity in Pima Indians, we sequenced the coding region of this gene in 426 full-heritage, non–first-degree related, adult Pima Indians (300 severely obese and 126 nondiabetic nonobese control subjects). Three coding variations were detected as heterozygotes only in severely obese subjects. One variation, detected in three obese subjects, was a novel single-base insertion (A) at nucleotide 100, and it predicted a frameshift and premature STOP at codon 37. The second variant, detected in 10 obese subjects, predicted a previously identified arginine-to-glutamine substitution at codon 165, and the third variant, detected in one obese subject, predicted a novel glycine-to-serine substitution at codon 231. Three polymorphisms were also identified in the 5′ untranslated region, but these variants were detected in both obese and lean subjects and had similar allele frequencies. We conclude that variations in MC4R may account for a small portion of obesity in this Native American population. Diabetes 53:2696–2699, 2004

The genetic basis of human obesity is largely unknown. A few genes, such as the melanocortin 4 receptor (MC4R) located on chromosome 18q22, have been identified as the cause of rare forms of monogenic obesity in both humans and rodents (1,2). Heterozygous coding mutations in the MC4R gene are implicated in 1–6% of early onset or severe adult obesity, and they appear to exert a major gene effect in affected individuals (i.e., have a high penetrance and explain a high percentage of the phenotypic variance) (3,4). To identify obesity susceptibility loci in Pima Indians, a population with high prevalence of severe obesity (5), we screened MC4R as a candidate gene. The coding region of the single exon was sequenced in 426 full-heritage, non–first-degree related, adult Pima Indians (300 severely obese, as defined by maximum BMI >45 kg/m2, and 126 nonobese, as defined by maximum BMI <30 kg/m2 at age >35 years and nondiabetic) (Table 1). Three coding variations were identified. A novel single-base insertion (A) at nucleotide 100, which predicts a frameshift resulting in a premature STOP (TGA) at codon 37, was observed as a heterozygote in 3 of the 300 severely obese Pima subjects, but it was not observed in any of the 126 nonobese subjects. In addition, a previously described G-to-A substitution at nucleotide 494, which predicts an arginine (R)–to–glutamine (Q) substitution at codon 165 (R165Q), was observed as a heterozygote in 10 of the 300 obese Pima subjects but in none of the control subjects. Finally, a novel G-to-A substitution at nucleotide 691, predicting a glycine (G)–to–serine (S) substitution at codon 231 (G231S), was identified as a heterozygote in 1 of the 300 severely obese subjects but in none of the control subjects (Table 2). All of these coding variations occurred in different subjects of this sample.

To search for variants that could affect gene expression, ~2 kb of the upstream untranslated (putative promoter) region of MC4R was additionally sequenced in 84 of the extremely obese subjects (BMI 50.5–73.8 kg/m2). The DNA sequences from these obese Pima Indian subjects were compared with each other and also with the published human MC4R sequence. Three polymorphisms were detected. The untranslated polymorphisms were all C-to-T substitutions, at positions −1,538, −1,005, and −896 relative to the transcriptional start site, where the minor allele (T) had frequencies of 0.02, 0.05, and 0.49, respectively. These three upstream polymorphisms were further genotyped in the 300 severely obese and 126 nonobese subjects, and their allelic frequencies were similar between the obese and nonobese groups (data not shown). Genotypes of these three upstream single nucleotide polymorphisms (SNPs) were in Hardy-Weinberg equilibrium when analyzed in the obese group, the nonobese group, and the combined group. In our linkage disequilibrium analysis, we noted that many of the pairwise estimates of D2 are high, which is common for low-frequency polymorphisms because these are often recently introduced into the population and are typically carried on a single haplotype. However, D2 values are in general fairly low, indicating that there is little redundancy for the information provided by each SNP (data not shown).

All of the coding mutations were detected only in
severely obese subjects. As shown in Table 2, the nucleotide A insertion, resulting in the premature STOP codon, was detected as a heterozygous mutation in three severely obese subjects and was not observed in any of the control subjects, but its frequency (1%) was too rare to be statistically associated with obesity in this sample. Similarly, G231S was too rare (0.3%) to be statistically associated with severe obesity in this case-control sample. In contrast, the coding R165Q was detected in heterozygous form in 10 severely obese subjects (6 female and 4 male subjects), was monomorphic in all of the control subjects, and was statistically associated with severe obesity (P = 0.04, χ² test) in the case-control sample. Among the 300 severely obese subjects, the 10 individuals heterozygous for R165Q had a similar maximum BMI as compared with individuals who did not carry the Q allele (Table 3). However, the R165Q subjects tended to be younger when they reached their maximum BMI as compared with the other obese subjects (30 ± 7 vs. 37 ± 11 years, P = 0.04) (Table 3).

Among the 3 subjects who were heterozygous for the novel insertion at nucleotide 100, 1 subject had extended pedigree information and DNA to allow genotyping of 16 additional family members (Fig. 1A). Seven family members in this pedigree were confirmed to be heterozygous for the nucleotide insertion. These seven subjects had a maximum BMI that ranged from 33 to 54 kg/m², and some subjects reached their maximum BMI at a very young age (e.g., subject II₆ and III₁₅ had a BMI of 43 and 49 kg/m² at age 19 and 7, respectively) (Fig. 1A). Based on previous functional studies of MC4R mutations, an insertion resulting in a premature STOP at codon 37 should produce a truncated receptor with no activity (3). Critical ligand-binding and transmembrane domains lie COOH-terminal to this region (3,6–9). Complete loss of function of one MC4R allele is predicted to result in obesity because the gene appears to function in a codominant manner (2,3). However, we are unable to clearly demonstrate a link between severe obesity and the inheritance of this single nucleotide insertion in our Pima Indian pedigree because

<table>
<thead>
<tr>
<th>TABLE 1 Characteristics of subjects in case/control sample for severe obesity</th>
<th>Severe obesity</th>
<th>Control</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n (F/M)</td>
<td>300 (224/76)</td>
<td>126 (62/64)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (years)*</td>
<td>36.7 ± 11.0</td>
<td>40.3 ± 4.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Maximum BMI (kg/m²)</td>
<td>50.2 ± 4.9</td>
<td>26.5 ± 2.6</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Data are means ± SD unless otherwise indicated. *Age at maximum BMI.

there are at least two siblings in generation IV (Fig. 1A) who are severely obese (BMI of 50 and 45 kg/m²) but do not carry the nucleotide insertion. Therefore, we propose that there are additional genetic and/or environmental factors that have major effects on determining obesity in this pedigree.

The single subject with G231S had four more family members for whom clinical information and DNA were available. Genotyping of these four additional family members identified two members to be G231S heterozygotes, with maximum BMIs of 44 and 47 kg/m² at age 44 and 47 years, respectively (II₇ and II₈) (Fig. 1B). The lean sibling (I₂) (Fig. 1B) of the proband (I₁) did not carry the mutation, and the only available subject of the third generation, whose maximum BMI was 44 kg/m² at age 22 years, did not carry G231S. Based on our current data, we are not able to determine whether the novel heterozygous amino acid substitution G231S is the cause of morbid obesity in this family.

The R165Q substitution has been previously reported as a rare mutation causing severe obesity (3,9), and this variant is statistically associated with severe obesity in Pima Indians. This heterozygous mutation was observed in 3% of severely obese Pima Indians (maximum BMI >45 kg/m²), which is the highest frequency reported for any single functional MC4R variant. Subjects with the Q allele were also younger when they reached maximum BMI as compared with equally obese subjects homozygous for the R allele (Table 3). The finding that heterozygous R165Q Pima subjects had an earlier onset to their obesity is supported by other studies in humans (3,9–11). Amino acid 165 is located in a transmembrane domain, and functional studies have shown that the R165Q mutation dramatically reduces the binding activity of MC4R by 70% (3,12). In addition, another recent study demonstrated that the R165Q mutant receptor is poorly expressed on the cell surface (13). Therefore, R165Q should decrease MC4R signaling.

Haploinsufficiency at the MC4R locus, which is thought to be sufficient to cause obesity, typically results from translated nonfunctional MC4R protein rather than aberrant gene transcription. We specifically sought mutations that could reduce MC4R gene transcription in Pima Indians by sequencing a 2-kb region that included the minimal promoter region (14). Although three upstream variants were identified, none were associated with obesity, which is consistent with a recent report of MC4R promoter variants in 266 children and 165 adults (14).

We have previously published a report that failed to

| TABLE 2 Frequency of heterozygous coding mutations in 426 case/control subjects |
|---|---|---|---|
| Mutation | Severe obese | Control | P value (χ² test) |
| −/A insertion at nucleotide 100 | 3/300 | 0/126 | NS |
| R/Q at codon 165 | 10/300 | 0/126 | 0.04 |
| G/S at codon 231 | 1/300 | 0/126 | NS |

| TABLE 3 Characteristics of severely obese subjects based on R165Q genotype |
|---|---|---|---|
| n (F/M) | Arg/Arg (R165R) | Arg/Gln (R165Q) | P value |
| Maximum BMI (kg/m²) | 50.1 ± 4.8 | 51.9 ± 7.2 | NS |
| Age at maximum BMI (years) | 36.9 ± 11.1 | 29.7 ± 7.4 | 0.04 |

Data are the means ± SD unless otherwise indicated. *P value is adjusted for age and sex.
identify variation in the *MC4R* coding sequence in Pima Indians (15). Our inability to detect allelic variation in this previous study was likely caused by the smaller number of samples that were sequenced (10 lean and 10 obese subjects), which would allow detection of variant alleles with a frequency ≥2.5% in the general Pima population, or ≥5.0% if a variant occurred only in obese subjects. In the current study, 426 subjects were sequenced, which should allow detection of coding variants with an allele frequency of ≥0.12%.

Identification of two rare functional variants in *MC4R* suggests that this gene has an important role in contributing to severe obesity in some Pima Indian families, but these two variants do not account for the high prevalence of obesity in this Native American tribe. However, identifying individuals and pedigrees whose BMI is influenced...
by rare mutations may lessen the complexity of population-based studies to search for genes that cause common, multigenic obesity.

**RESEARCH DESIGN AND METHODS**

All of the subjects are part of our ongoing longitudinal study of the etiology of obesity and type 2 diabetes among the Gila River Indian Community in central Arizona (16). Sequencing of the coding region was performed in DNA from 300 severely obese subjects, as defined by a maximum BMI >45 kg/m², and 126 nonobese and nonobese control subjects, as defined by BMI <30 at age >35 years. Additional sequencing of 2 kb of the 5′/H11032 IV, for genotypic analysis (Fig. 1 only three generations data available for subjects) and 126 nonobese subjects. Diabetic status was determined by the genotyping the remaining 216 obese subjects (for a total of 300 obese subjects) and 126 nonobese subjects. Diabetic status was determined by the criteria of the World Health Organization (17). All subjects in the case-control set were from different nuclear families.

The pedigree analyzed for the insertion A at nucleotide 100 had clinical data available for five generations and DNA for 17 subjects, from generation II to IV, for genotypic analysis (Fig. 1A), whereas pedigree analyzed for G231S had only three generations’ clinical data and 5 subjects’ DNA available (Fig. 1B).

**Sequence variant identification and genotyping.** Genomic DNA for sequencing and genotyping was obtained from peripheral lymphocytes. Sequencing was performed using Big Dye terminator (Applied Biosystems) on an automated DNA capillary sequencer (model 3700; Applied Biosystems). The DNA region encompassing the insertion A at nucleotide 100 was sequenced using primers MCAR-insA100 forward 5′-TGAGACGACTCCCTGACC-3′ and reverse 5′-CCCCACGGCTTACGTCA-3′. The DNA region encompassing R165Q was sequenced using primers MCAR-R165Q forward 5′-TACGGATGCCACAGATTTCA-3′ and reverse 5′-CCCGAGACACAAAGAGCC-3′. The DNA region encompassing G231S was sequenced using primers MCAR-G231S forward 5′-GGGTTGGGATCATCATAAGT-3′ and reverse 5′-CCAGTACCCTACGCGAAGA-3′. The DNA region encompassing the MCAR mutations were genotyped using primers MCAR-896, -1,005, and -1,538 positions upstream of transscriptional start site was sequenced using primers MCAR-896 forward 5′-CTCTTCCAGGGATCTGTC-3′ and reverse 5′-TGAGACGACTCCCTGACC-3′, primers MCAR-1005 forward 5′-CCACCTCCTTCCAAACCTGATTA-3′ and reverse 5′-TAAGACCCGACGGATGTC-3′, and primers MCAR-L1538 forward 5′-GGGTTGGGATCATCATAAGT-3′ and reverse 5′-CTTGAACTTGGAAGAGAC-3′. Genotyping of MCAR-896, -1,005, and -1,538 was performed using TaqMan allelic discrimination assay (Applied Biosystems) for case-control association study. The TaqMan genotyping reaction was amplified on a GeneAmp PCR system 9700 (95 °C for 10 min, followed by 40 cycles of 95 °C for 30 s and 60 °C for 1 min 30 s), and fluorescence was detected on an ABI Prism 7700 sequence detector (Applied BioSystems). Sequence information for all oligonucleotide primers and probes is available upon request.

**Statistical analysis.** Statistical analyses were performed using the statistical analysis system of the SAS Institute (Cary, NC). Age and BMI are expressed as means ± SD. Differences between means were compared by unpaired Student’s t test or general linear model (GLM) when adjusted for age and sex, and differences of frequencies were compared by χ² test. All reported P values were two sided. P values <0.05 were considered to be of statistical significance.

**ACKNOWLEDGMENTS**

The authors are grateful to Dr. Michal Prochazka for his critical reading of this manuscript and Dr. Robert L. Hanson for his analysis of linkage disequilibrium.

**REFERENCES**