Brief Genetics Report

Association of Amino Acid Variants in the Activating Transcription Factor 6 Gene (*ATF6*) on 1q21-q23 With Type 2 Diabetes in Pima Indians

Farook Thameem,[1,2] Vidya S. Farook,[1,3] Clifton Bogardus,[1] and Michal Prochazka[1]

Activating transcription factor 6 (ATF6) is important for protective cell response to accumulation of unfolded and misfolded proteins in endoplasmic reticulum, and disturbances of this process can contribute to β-cell apoptosis. We analyzed the structural gene located within a region on 1q21-q23 linked with type 2 diabetes in several populations for variants in the Pima Indians. Functionally important segments of *ATF6* were sequenced in 15 diabetic and 15 nondiabetic Pimas and representative single nucleotide polymorphisms (SNPs) tested for association with type 2 diabetes in 900–1,000 subjects. We identified 20 variants including three amino acid substitutions [Met(67)Val, Pro(145)Ala, and Ser(157)Pro]. Pro(145)Ala and Ser(157) Pro were in a complete linkage disequilibrium and showed a nominal association with type 2 diabetes ($P = 0.05$; odds ratio 2.3 [95% CI 1.0–5.2]) and with 30-min plasma insulin during oral glucose tolerance test in 287 nondiabetic individuals ($P = 0.045$). Although the associations with type 2 diabetes and plasma insulin levels are marginal and their functional consequences are yet unknown, all three amino acid substitutions are located in a functionally important part of ATF6. Because these variants are not unique to the Pimas, it will be feasible to investigate their association with type 2 diabetes in other populations to better evaluate their significance for a predisposition to the disease. *Diabetes* 55:839–842, 2006

Secretory and transmembrane proteins of eukaryotic cells are transported into the lumen of endoplasmic reticulum where they undergo posttranslational modification and folding before subsequent transport to other subcellular compartments (1). Properly folded proteins will eventually be cleared out, whereas excessive accumulation of unfolded and misfolded proteins within endoplasmic reticulum will perturb its function and lead to so-called endoplasmic reticulum stress, which may result in apoptosis of the involved cells (1). To survive and adapt under such conditions, cells have a self-protective mechanism called endoplasmic reticulum stress response involving upregulation of genes encoding endoplasmic reticulum chaperones to increase protein folding activity, translational attenuation to reduce the load of new proteins, degradation of proteins misfolded in the endoplasmic reticulum, and apoptosis when the functions of endoplasmic reticulum are extensively impaired.

Excessive, chronic endoplasmic reticulum stress has been implicated in many diseases including diabetes (1). Studies of pancreatic endoplasmic reticulum kinase–deficient mice (1) or of Akita mice with Cys96Tyr mutation in the insulin 2 gene (2) indicated that β-cells are excessively vulnerable to endoplasmic reticulum stress, which predisposes these strains to the development of diabetes. Similarly, the loss-of-function mutation in a human pancreatic endoplasmic reticulum kinase has been implicated in the etiology of a monogenic form of diabetes known as Wolcott-Rallison syndrome, a rare autosomal recessive disorder characterized by diabetes, chondrodysplasia, osteoporosis, and growth retardation (3).

A major mediator of transcriptional induction of endoplasmic reticulum chaperones by endoplasmic reticulum stress is the basic leucine zipper protein–activating transcription factor 6 (ATF6) (1). The *ATF6* gene is located on human chromosome 1q21-q23, a region linked to type 2 diabetes in Pima Indians of Arizona, Caucasians, and Chinese populations (4,5). The Pima Indians have the world's highest prevalence of type 2 diabetes, and prospective studies indicate that insulin resistance and insulin secretory dysfunction are major predictors of the disease in this population (6). To investigate the involvement of genetic factors in type 2 diabetes, we chose *ATF6* as a candidate susceptibility gene for type 2 diabetes in Pimas based on its localization on 1q21 and the apparent protective role of the protein in β-cells.

RESULTS AND DISCUSSION

ATF6 consists of 16 exons, spans ~193 kb (http://genome.ucsc.edu), and is located ~1.6 Mb centromeric from the microsatellite *D1S1677*, which marks the peak of type 2 diabetes linkage in the Pima Indians (4). We examined all exons plus adjacent splice sites, 2 kb of 5′ flanking region, and the entire 3′-untranslated region by sequencing in 15 diabetic and 15 nondiabetic subjects. We detected 20 diallelic variants including seven in the exons, six in the promoter region, four in the intronic regions, and

From the [1]Clinical Diabetes and Nutrition Section, Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona; the [2]Department of Medicine, Division of Nephrology, University of Texas Health Science Center at San Antonio, San Antonio, Texas; and the [3]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas.

Address correspondence to Michal Prochazka, MD, Diabetes Molecular Genetics Section, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 445 N. 5th St., Suite 210, Phoenix, AZ 85004. E-mail: mprochazka@phx.niddk.nih.gov.

Received for publication 4 August 2005 and accepted in revised form 23 November 2005.

ATF6, activating transcription factor 6; LOD, logarithm of odds; SNP, single nucleotide polymorphism.

© 2006 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/55/3/839/651470/zdb00306000839.pdf by guest on 13 April 2024

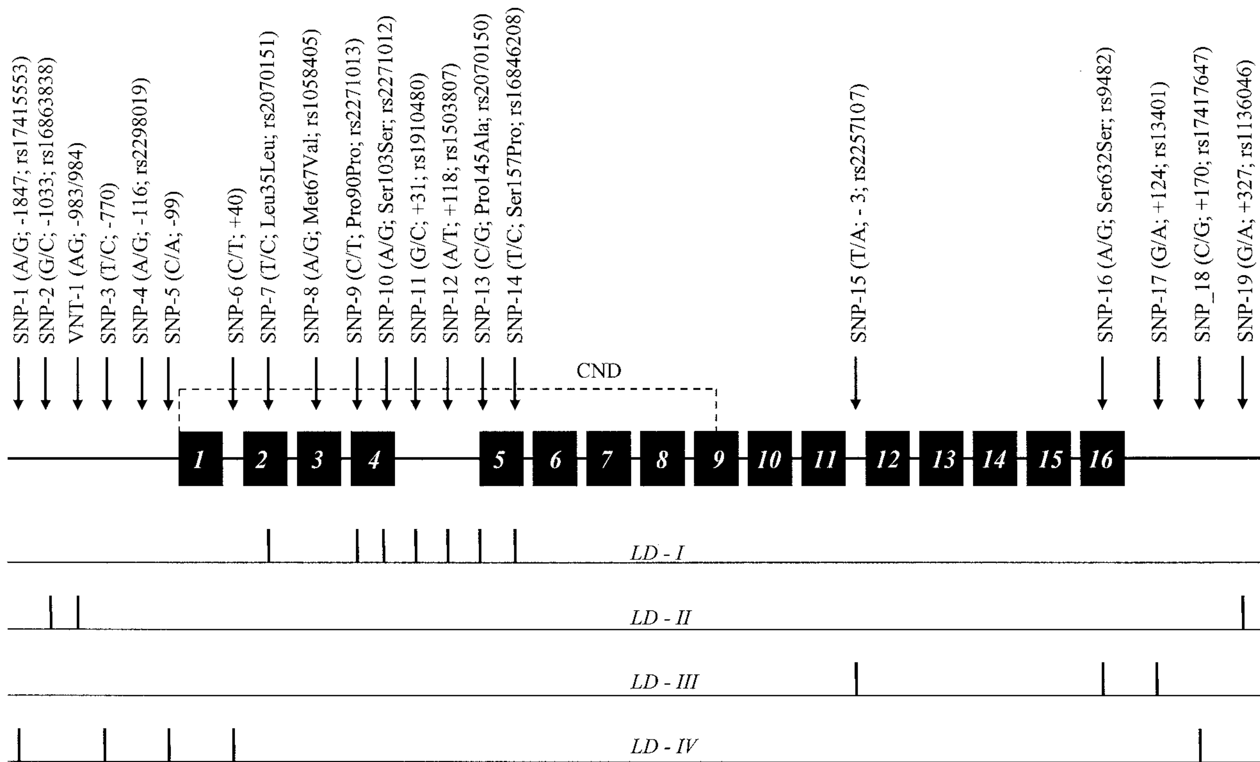

FIG. 1. Schematic diagram of human *ATF6* and DNA variants. Exons are represented by solid boxes showing their corresponding numbers, and introns, 5′ and 3′ untranslated region, and promoter are shown by a thin line. Variants are marked by vertical arrows, and the exact base positions, base changes, and rs # are given in parentheses (positive numbers show distance from the first base of the corresponding exon or intron, negative numbers indicate the distance from the following exon). Groups of markers in tentative linkage disequilibrium (I, II, III, and IV) are indicated below the gene structure. Exons 1–9 encode the cytoplasmic-localized nuclear domain (CND), which activates the transcription of endoplasmic reticulum chaperones.

three in the 3′ untranslated region (Fig. 1). One of the promoter variants is a 2-bp deletion/insertion (size variant-1 [VNT-1]), whereas the remaining are single nucleotide polymorphisms (SNPs). Three of the exonic SNPs predicted amino acid substitutions, including Met67Val (SNP-8 in exon 3), Pro145Ala (SNP-13 in exon 5), and Ser157Pro (SNP-14 in exon 5), respectively (Fig. 1). Based on the observed concordance of the initial genotypings obtained by sequencing in 30 individuals, we divided for practical reasons most of the variants into four tentative linkage disequilibrium groups. We defined as a tentative linkage disequilibrium group a set of SNPs/variants showing identical allele frequency and 100% genotypic concordance among 30 subjects. Based on our experience with SNP analysis in the Pima population, variants showing a 100% frequency and genotype concordance in 30 subjects used for the initial sequencing are typically found in a complete linkage disequilibrium ($\Delta^2 = 0.99–1.0$), when genotyped in 900–1,000 individuals used in our linkage and association studies.

The linkage disequilibrium groups we identified include SNP-7 and -9 to -14 (tentative linkage disequilibrium group I); SNP-2 and -19 and VNT-1 (group II); SNP-15 to -17 (group III); and SNP-1, -3, -5, -6, and -18 (group IV) (Fig. 1). SNP-4 and SNP-8 could not be assigned to any group and were analyzed individually. Several of the variants found in the Pimas (including Met67Val [rs1058405], Pro145Ala [rs2070150], and Ser157Pro [rs16846208]) have been already listed in the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/), consistent with their occurrence in other populations.

To examine the observed *ATF6* variants for association with type 2 diabetes and related metabolic traits, we selected one (presumably representative) marker from each tentative linkage disequilibrium group (SNP-13 from group I, SNP-2 from group II, and SNP-17 from group III), as well as SNP-4 and SNP-8, and examined them in ~900–1,000 Pimas in whom the type 2 diabetes linkage with 1q was originally observed. The minor allele from SNPs in group IV showed low frequency (0.03), and

TABLE 1
Association analysis of *ATF6* SNPs with type 2 diabetes in 960 Pima Indians

Representative markers	dbSNP no.	Alleles (frequency)	*P* value
SNP-13 (tentative linkage disequilibrium group I)	rs2070150 (Pro145Ala = CCG to GCG)	C/G(0.80/0.20)	0.05*
SNP-2 (tentative linkage disequilibrium group II)	rs16863838	C/G(0.82/0.18)	0.60
SNP-17 (tentative linkage disequilibrium group III)	rs13401	G/A(0.88/0.12)	0.70
SNP-4	rs2298019	G/A(0.84/0.16)	0.57
SNP-8	rs1058405 (Met67Val = ATG to GTG)	A/G(0.96/0.04)	0.62

*See Table 2 for details.

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/55/3/839/651470/zdb00306000839.pdf by guest on 13 April 2024

TABLE 2
Allele frequency and distribution of Pro(145)Ala (= SNP-13: CCG to GCG) genotypes in 960 Pima Indians

	Allele frequency* (C/G)	Genotype†			
		CC	CG	GG	Total
Type 2 diabetes	0.81/0.19	368	177	16	561
Non–type 2 diabetes	0.78/0.22	251	124	24	399

*Allele frequency difference between type 2 diabetes and non–type 2 diabetes are not statistically significant ($P > 0.2$). †Genotype frequency differences between type 2 diabetes, and non–type 2 diabetes are statistically significant ($P = 0.05$; dominant model adjusted for age, birth date, sex, and ethnicity).

therefore this group was not analyzed further due to the lack of statistical power. All analyzed variants were in Hardy-Weinberg equilibrium, and the data are summarized in Table 1. We found that SNP-13 [Pro(145)Ala, *C/G*] showed a nominal association with type 2 diabetes under a dominant model (the C allele being more frequent in type 2 diabetic subjects; $P = 0.05$; odds ratio =2.3 [95% CI 1.0–5.2]; Table 2) and was also associated with differences in plasma insulin at 30 min during an oral glucose tolerance test (geometric mean [95% CI] for CC = 1,284 [11,461–404]; for CG = 1,067 [929–1,225]; for GG = 1,470 [12,241–779] pmol/l; $P = 0.045$) in a subset of 287 nondiabetic Pimas. When this SNP was used for an adjustment of the linkage score (original logarithm of odds [LOD] = 2.6 [4]), a reduction of 0.2 was observed. The remaining analyzed SNPs showed no significant differences between diabetic and nondiabetic Pimas (Table 1 and not shown).

We also identified six variants in the promoter/5′-flanking region of *ATF6*. To determine whether any of them could be in a transcription factor binding site, we used the MatInspector program (TRANSFAC 4.0 database; http://www.gene-regulation.com/), which predicted several potential cis-acting binding elements for *CRE-BP1*, *CREB*, *STAT*, and *HNF-3β*. However, none of these variants were located in the predicted putative transcription factor binding sites, and neither SNP-4 nor VNT-1 (representing linkage disequilibrium group II) were found to be associated with type 2 diabetes.

ATF6 is constitutively expressed as a 90-kDa protein (670 aa) and cleaved to a 50-kDa protein (~400 aa) specifically in endoplasmic reticulum–stressed cells before the induction of its target genes (1). The resulting NH^2-terminal fragment of ~400 amino acids is translocated to the nucleus to activate endoplasmic reticulum stress-inducible genes required for cell survival. It has been shown that the first 373 aa (encoded by exons 1–9) are sufficient for ATF6-mediated transcriptional activity (7) and that the functionally responsible region resides within the NH^2-terminal 93 amino acids of ATF6 (7). Interestingly all three amino acid variants, including Pro(145)Ala and Ser(157)Pro showing association with type 2 diabetes in this study, are located within the functionally important NH^2-terminal part, also referred to as cytoplasmic-localized nuclear domain, which activates the transcription of endoplasmic reticulum chaperones (7).

Pro(145)Ala results in a change from a nucleophilic to a hydrophilic residue, and a Pro to Ala substitution can remove a helix-termination signal that is important for efficient protein folding (8). Furthermore, some Ser and Pro residues in ATF6 are targets for phosphorylation by p38 mitogen activated protein kinases (p38MAPK), which regulate gene transcription by phosphorylation of relevant transcription factors. It is noted that the NH^2-terminal fragment of ATF6 is phosphorylated by p38MAPK, which enhances the transactivation of its target genes (9). In this context, further studies are essential to address if Pro145Ala or Ser157Pro substitutions could affect the ATF6 folding and/or be a target for p38MAPK phosphorylation.

In summary, we report here an association of variants (including two amino acid substitutions) in *ATF6* with the risk for type 2 diabetes in the Pima Indians. Although the association by itself appears marginal and does not account for the linkage, two of the associated SNPs predict an amino acid change in a functionally important domain of ATF6. Combined with the importance of this protein in the protection of pancreatic β-cells, our finding of an additional association of the presumed diabetes risk alleles (145Ala and 157Pro) with lower plasma insulin levels measured during oral glucose tolerance test would fit the scenario of an impaired insulin release, which typically precedes the development of type 2 diabetes, and is one of the precursors of the disease in this population (6). One possible explanation for the lack of an obvious effect on the original linkage could be that the 1q linkage is the result of allelic effects and/or interactions at more than one susceptibility gene in this area, with the amino acid changes in ATF6 representing only a part of the overall effect of this locus contributing to the predisposition to type 2 diabetes. Although these two amino acids are not conserved between species for which the ATF6 sequence is available, both variants are deposited in the dbSNP database and thus are not unique to the Pimas. Therefore, it would be feasible to assess whether our observation of their association with type 2 diabetes can be replicated in other ethnic groups. It is intriguing that the substitutions are within an important part of the protein, and further studies will be necessary to assess their functional significance. It is also possible that the *ATF6* SNPs associated with type 2 diabetes could be markers for a diabetes susceptibility variant outside of this locus, and sequence analyses of adjacent genes are currently in progress to explore such possibility.

RESEARCH DESIGN AND METHODS

Pima Indians selected for the genomic sequencing of *ATF6* (15 who developed type 2 diabetes before the age of 25 years and 15 who remained nondiabetic until at least 45 years; there were no first-degree relatives among these subjects) were participants of ongoing longitudinal studies of type 2 diabetes in the Gila River Indian Community since 1965 (4). SNP genotyping was performed in ~900–1,000 subjects, and diabetes was diagnosed according to World Health Organization criteria (4). A subset of ~300 nondiabetic Pima Indians were also subjected to detailed clinical measurements related to insulin action and secretion, and substrate metabolism (6). This study was approved by the institutional review board of the National Institute of Diabetes and Digestive and Kidney Diseases and the Tribal Council of the Gila River Indian Community. All subjects gave written informed consent before participation.

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/55/3/839/651470/zdb00306000839.pdf by guest on 13 April 2024

Variant identification, genotyping, and data analysis. Oligonucleotide primers used for PCR and sequencing are available upon request. Sequencing was performed using ABI Prism Big Dye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA) and a capillary sequencer (Model 3730xl; Applied Biosystems). Genotyping was performed using the TaqMan assay (Applied Biosystems), which was carried out on a GeneAmp PCR system 9700 (Applied Biosystems), and fluorescent signals were detected on an ABI PRISM 7700 sequence detector (Applied Biosystems). To assure accuracy of the genotypings, coded blind replicate samples from 100 subjects were included in each genotyping assay. Only markers with an error rate between 0 and 2% (representing none to two discrepancies) are used for statistical analyses.

Differences in allelic frequencies between diabetic and nondiabetic subjects were analyzed with programs of the SAS Institute (Cary, NC). Associations were evaluated by logistic regression after adjustment for age, sex, birth date, and ethnicity (Pima heritage); generalized estimating equations were utilized to correct for familial relatedness between subjects. The power of these statistical analyses to detect association depends on the model used. To increase the power, and to take into consideration the possibility of various allelic effects and interactions, we used three different association models: dominant, recessive, and additive. The first model assumes that the more common allele is dominant over the less common allele; the second model assumes that the more common allele is recessive to the less common allele; the third model assumes an additive effect, where diabetes susceptibility due to the presence of one allele in a heterozygote carrier will proportionately increase in a homozygote.

To evaluate the degree to which a polymorphism associated with type 2 diabetes may account for the observed linkage on 1q, the LOD score was assessed after an adjustment for the association, including the genotype as an additional covariate factor in the linkage analysis. It is expected that the adjusted LOD score will become lower than the original value, depending on the extent of linkage disequilibrium between the causative variant and the SNP marker. Odds ratios were calculated for each model as another parameter to assess the risk contributed by the alleles/genotypes of the analyzed marker for developing type 2 diabetes.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Disease.

We thank the members of the Gila River Indian Community for their continued participation in the ongoing longitudinal studies of type 2 diabetes, Vicky Ossowski for technical support, and Sayuko Kobes and Dr. Robert Hanson for statistical analyses.

REFERENCES

1. Zhang K, Kaufman RJ: Signaling the unfolded protein response from the endoplasmic reticulum. *J Biol Chem* 279:25935–25938, 2004
2. Wang J, Takeuchi T, Tanaka S, Kubo SK, Kayo T, Lu D, Takata K, Koizumi A, Izumi T: A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the Mody mouse. *J Clin Invest* 103:27–37, 1999
3. Castelnau P, Le Merrer M, Diatloff-Zito C, Marquis E, Tete MJ, Robert JJ: Wolcott-Rallison syndrome: a case with endocrine and exocrine pancreatic deficiency and pancreatic hypotrophy. *Eur J Pediatr* 159:631–633, 2000
4. Hanson RL, Ehm MG, Pettitt DJ, Prochazka M, Thompson DB, Timberlake D, Foroud T, Kobes S, Baier L, Burns DK, Almasy L, Blangero J, Garvey WT, Bennett PH, Knowler WC: An autosomal genomic scan for loci linked to type 2 diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* 63:1130–1138, 1998
5. McCarthy MI: Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* 3:159–167, 2003
6. Lillioja S, Mott DM, Spraul M, Ferraro R, Foley JE, Ravussin E, Knowler WC, Bennett PH, Bogardus C: Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus: prospective studies of Pima Indians. *N Engl J Med* 329:1988–1992, 1993
7. Thuerauf DJ, Morrison LE, Hoover H, Glembotski CC: Coordination of ATF6-mediated transcription and ATF6 degradation by a domain that is shared with the viral transcription factor, VP16. *J Biol Chem* 277:20734–20739, 2002
8. Eyles SJ, Gierasch LM: Multiple roles of prolyl residues in structure and folding. *J Mol Biol* 301:737–747, 2000
9. Luo S, Lee AS: Requirement of the p38 mitogen-activated protein kinase signalling pathway for the induction of the 78 kDa glucose-regulated protein/immunoglobulin heavy-chain binding protein by azetidine stress: activating transcription factor 6 as a target for stress-induced phosphorylation. *Biochem J* 366:787–795, 2002

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/55/3/839/651470/zdb00306000839.pdf by guest on 13 April 2024