Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Genetics and Metabolism 81 (2004) 16–21
Molecular Genetics and Metabolism
www.elsevier.com/locate/ymgme

The transcribed endosulfine α gene is located within a type 2 diabetes-linked region on 1q: sequence and expression analysis in Pima Indians☆

Farook Thameem,[a] Vidya S. Farook,[a] Xiaolin Yang,[b] Yong-Ho Lee,[a] Paskasari A. Permana,[c] Clifton Bogardus,[a] and Michal Prochazka[a,*]

[a] *Clinical Diabetes and Nutrition Section, Phoenix Epidemiology and Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix AZ, USA*
[b] *Lundberg Laboratory for Diabetes Research, Sahlgrenska University Hospital, Goteborg, Sweden*
[c] *Carl T. Hayden Veterans Affairs Medical Center, Phoenix AZ, USA*

Received 5 August 2003; received in revised form 18 September 2003; accepted 18 September 2003

Abstract

Endosulfine α (ENSA) is an endogenous ligand of the sulfonylurea receptor 1 (SUR1) that can stimulate insulin secretion by pancreatic β cells. Originally, an intronless gene coding for this protein was assigned to Chr. 14q, but more recent information available in public databases indicated the position of *ENSA* on 1q21. We show here that the 1q21 locus represents the expressed gene consisting of 6 exons, whereas the locus on Chr. 14q is apparently a pseudogene. The *ENSA* gene on 1q21 produces several alternatively spliced transcripts, and is located within a region linked with T2DM in diverse populations including the Pima Indians. We analyzed *ENSA* in this Native American population and identified seven variants, which fall into three linkage disequilibrium groups. Analysis of representative markers in over 1200 Pima Indians did not reveal any significant association with T2DM, or with differences in insulin action and insulin secretion in a subset of approximately 270 non-diabetic subjects. In addition, we did not detect any significant correlation of skeletal muscle *ENSA* transcript levels with differences in insulin action in 49 non-diabetic subjects. We conclude that sequence alterations in *ENSA* are an unlikely cause for the linkage of T2DM with 1q21–q23 in the Pima Indians.

Keywords: Endosulfine α gene; Chromosome 1q; Variants; Expression; Type 2 diabetes mellitus; Pima Indians

Introduction

Impaired insulin secretory response of pancreatic β-cells to postprandial elevation of plasma glucose is one of the main clinical characteristics preceding and accompanying the development and progression of type 2 diabetes mellitus (T2DM) [1]. Sulfonylureas (SUs) are the most common class of oral drugs used for treatment of T2DM, acting primarily by improving insulin release in response to glucose [2]. The insulin secretory effect of SUs is triggered by an inhibition of ATP-dependent potassium (K-ATP) channels in β cell membrane via binding of the drug to the sulfonylurea receptor-1 subunit of the channel (SUR1) [3]. SUs have been in clinical use for several decades, but an endogenous ligand for SUR1, endosulfine α (ENSA), was identified [4] and characterized relatively recently [5]. The reported 13 kDa ENSA consists of 121 amino acid residues, is phosphorylated by protein kinase A, and its broad expression parallels that of K-ATP channels [4,5]. Although the physiological role of ENSA is still unclear, its function as an insulin secretagogue was originally attributed to an inhibition of K-ATP channels in the β cells [5,6]. More recently a blocking effect of ENSA on voltage-gated calcium channels in these cells was also

☆ Sequences reported in this article were deposited with the GenBank Database under Accession Nos. AY326399–AY326403.

* Corresponding author. Fax: 1-602-200-5335.
E-mail address: mprochazka@phx.niddk.nih.gov (M. Prochazka).

1096-7192/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymgme.2003.09.013

reported, indicating a more complex mechanism by which this peptide may affect insulin secretion [6]. In addition to its effects in the pancreatic islets, the nearly ubiquitous expression of ENSA [4,5,7] is consistent with other, yet unknown biological roles of the protein.

The human *ENSA* was originally described as an intronless gene, which was assigned to 14q24.3–q31 by fluorescent in situ hybridization [8]. In addition, an *ENSA* pseudogene (*ENSAP*) with 84% identity to the reported cDNA is found on Chr. 20 (GenBank #NG_001035). However, based on more recent information deposited in the public domain (OMIM #603061), two overlapping BAC clones (RP11-54A4: GenBank #AL356356; and RP11-243G22: GenBank #AC053497) from Chr. 1q21 also contain *ENSA* sequences. Here we addressed this inconsistency and demonstrate that the transcribed *ENSA* is located on 1q21. Furthermore, we present the results of variant screening and expression analysis of this gene in the Pima Indians, a population with the highest reported prevalence of T2DM, in which we have previously detected a linkage of the disease with 1q21–q23 [9].

Subjects and methods

The participants are members of the Gila River Indian Community involved in prospective studies of T2DM, diagnosed by the WHO criteria [9]. Parameters of insulin action and insulin secretion were measured in a subset of unaffected subjects with normal glucose tolerance as detailed in [10,11]. Briefly, individuals were admitted to the clinical research center for 8–10 days, and on separate days underwent an oral glucose tolerance test (OGTT), two-step hyperinsulinemic–euglycemic clamp to determine whole body glucose disposal in response to physiologic (M-low) and maximal (M-high) insulin concentrations, and an intravenous GTT to measure acute insulin secretory response (AIR). Body composition was determined by dual X-ray absorptiometry (DEXA). Collection of skeletal muscle biopsies, RNA extraction, and cDNA synthesis were described previously [11]. The studies were approved by the Institutional Review Board of the National Institutes of Health and by the Gila River Tribal Council, and signed informed consent was obtained from all participants.

Genomic DNA samples from approximately 1200 Pima individuals were the same as in the original linkage study [9]. DNA samples from somatic cell hybrids containing either human Chr. 1 (NA13139), Chr. 14 (NA11535), or Chr. 20 (NA13140) were obtained from the NIGMS Human Genetic Mutant Cell Repository (Coriell Institute, Camden, NJ). Human brain and skeletal muscle cDNA samples were purchased from Clontech (Palo Alto, CA).

To distinguish between PCR products from the Chr. 1, Chr. 14, and Chr. 20 loci, we utilized the following primer combinations: AGAACCCTGCGGAGGAGA or CAG GAGAAAGAAGGTATTCTG + GTCCAATCCTAG CCTTTCCT (Chr.1-specific), CAGGAGAAAGAAGG TATTCTC + GTCCAATCCTAGCCTTTCCT (Chr.14-specific), and GAAACTAAACGAGAACACCGT + AT TTTCTGGGCAGATCCTGTGT (Chr.20-specific). Amplifications of genomic DNA were performed in the GeneAmp System 9600 thermocycler (Applied Biosystems, Foster City, CA) using the Expand PCR System kit (Roche Diagnostics, Indianapolis, IN) as described [12]. An assessment of expression was performed with the same sets of primers and KlenTaq DNA polymerase (Clontech) using commercially available cDNAs (Clontech) as described [13].

Variants were detected in selected individuals by PCR and direct sequencing with primers in Table 1, using ABI Prism BigDye Terminator Cycle Sequencing kit and ABI 377 or 3700 Sequencers (ABI, Foster City, CA). Marker genotypings were performed by PCR-RFLP using primers listed in Table 2 and the appropriate restriction endonucleases according to manufacturer's recommendations (New England Biolabs, Beverly, MA). The genotypes did not deviate statistically from Hardy–Weinberg equilibrium.

Quantification of *ENSA* transcript III was performed by TaqMan real-time PCR on ABI PRISM Sequence Detection System (Applied Biosystems, Foster City, CA) using the forward primer 5′-CCAAGTGCAGGACCAG ACAA, reverse primer 5′-TAAACGTCCAGTGGTTG AGGAA, and probe 5′-FAM-AGAGAAAGTCCTCGC TCTGTCACCAGCA-TAMRA-3′. Amplifications were started with 50 °C for 2 min and 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Measurements of human cyclophilin (TaqMan assay control from Applied Biosystems) were used to normalize the *ENSA* transcript levels.

Statistical analyses were performed with the procedures of the SAS Institute (Cary, NC) as described [10–13].

Results and discussion

As a first step in clarifying the position of the gene we performed a BLAST search [14] of public databases using the longest *ENSA* cDNA entry in the GenBank (AF067170; 2520 bp). We verified that BACs 54A4 and 243G22 on 1q21 contain *ENSA* sequences with features of functional exons which, when conceptually spliced, are identical to the cDNA. Moreover, we identified two additional BACs (RP11-340E1, AC069588; and RP11-406A9, AL163171), which contain *ENSA*-like sequence and map to 14q31.1, thus coinciding with the reported cytogenetic assignment of the gene [8]. In comparison to the genomic sequence from 1q21, the 14q locus represents a contiguous sequence of about 2.4 kb equivalent

Table 1
ENSA primers for variant screening

Designation	Region	Sequence (5′ → 3′)
PROM1F	Promoter	ACACCAACCTGGAAAGAGAG
PROM1R		GGGTTCAAGCGATTCTCCTG
PROM2F	Promoter	GACCAGCCTGACCAACATG
PROM2R		GTACCATCCACATACAGACAG
PROM3F	Promoter	GAGGACAGTGTAAGTATGTGG
PROM3R		AGCGATTCTCCTGCCTCAG
PROM4F	Promoter	AACCATCCTGGTCAACATGG
PROM4R		TGAATGCCTTATTTTCTCTCTG
PROM5F	Promoter	GTGGCAGATAAGAAGTGCTC
PROM5R		GGCGAAACCCTGTCTCTAC
PROM6F	Promoter	TGCCTCGCGGGTTAAAGTG
PROM6R		TCAAAGTGTTGGGATTACAGG
PROM7F	Promoter	GTAAGTAAAGTTGTGGAAGTTG
PROM7R		TCCTTAACGTCTTGCGATTG
1F	Exon 1	CCATTTTGACTGAGCAACCC
1R		AGCATGTCGTGTTGAAGCTC
1BF	Exon 1B	CACCAGTGCGGATTGGTC
1BR		CAATAGACCATCCTTAGCTAC
2F	Exon 2	CTCACAATAGTCATCTTGAGC
2R		AGAACCTCTACAGACTTCATTC
2BF	Exon 2B	CTTGAAGTGAGCACCAGCC
2BR		CATCACATTATCATCTGTGAC
3F	Exon 3	GGAAGACTTCTTGAAAGATCC
3R		GAGAAGATCTGACTAGGTGC
4F	Exon 4	ATAGGATTGAGACATGCCAG
4R		TGGGACACCAACAGGAAAG
I3AF	Intron 3	TTCTTCATCTCCGATCTTCTC
I3AR		AGCAGGCTGTTTCAGAGGC
I3BF	Intron 3	TCACATCACAAATTCTGATATG
I3BR		ACTCATTGTGACCAGTAGTC
I3CF	Intron 3	TAAGTCATTTAGCCTCTCTGG
I3CR		TGGCATCAGACTGCATCTG
I3DF	Intron 3	CCTTCCTCAACCACTGGAC
I3DR		TAAGGATTCTCTAAACCTGC
I3EF	Intron 3	CACTTTCCCAGCCTACCTC
I3ER		ATGCACATCAGCAAGTCTATC
I3FF	Intron 3	AGGTAGATTGAATGAGACATG
I3FR		CAGTTAAGACTTCAGAATCAG
I3GF	Intron 3	TAGATTGACACTTCCACACAC
I3GR		GTCCCACTATCCCATGATAC
I3HR	Intron 3	CAATCTACCTTTCCCAACTAC
3UT2	3′-UTR	GAACTCAGCCGCATCTCAC
3UT3		TGAGCTCTGTAGATAAATGCC
3UT4F	3′-UTR	CCAGATCCTGAGACGCTTC
3UT4R		AGGAGCTGTTTGCATCTGTC

Table 2
Representative markers in *ENSA* genotyped in ~1200 Pima Indians

Marker	Variant[a]	Primers (5′ → 3′)[b]	Restriction Endonuclease	Frequency
SNP-3	C/T	F: CAGACAAGAACCTGGTGAC R: TAAGGATTCTCTAAACCTGC	*Apa*I	0.72/0.28
SNP-4	A/G	F: CAGACAAGAACCTGGTGAC R: GTCCCACTATCCCATGATAC	*Sml*I	0.62/0.38
9-bp ins/del	TTCCCAAGA/-	F: GGGTTTTACACCGAGACATC R: AAGTACCCAGAGACAATACG	*Hpa*II	0.69/0.31

[a] SNP-3, SNP-4, and 9-bp ins/del represent LD cluster A, B, and C, respectively. The exact positions of the variants are annotated in AY326399 (9-bp ins/del), and AY326402 (SNP-3, SNP-4).
[b] F, forward; R, reverse orientation.

to *ENSA* exons 1, 2, and 3, plus the intron downstream of exon 3. However, this sequence shows 4% differences from the published *ENSA* cDNA, including a single base deletion leading to a shift in the reading frame after codon 98 predicting an addition of 40 novel amino acid residues at the C-terminal end.

To investigate the expression of each locus, we designed unique PCR primers that specifically amplified the corresponding transcript produced either from the locus on 1q21 or 14q31.1, or Chr. 20 that corresponds to *ENSAP*. The primer specificity was verified by PCR with genomic DNA from monochromosomal human somatic cell hybrids NA13139 (Chr. 1-specific), NA11535 (Chr. 14-specific), and NA13140 (Chr. 20-specific; data not shown). Using cDNA templates from brain and skeletal muscle reported to contain high levels of *ENSA* mRNA [4,5], we found that only primer combinations specific for the sequences on 1q amplified *ENSA* transcript in these tissues, whereas no products were obtained with the primer pair unique for the *ENSA*-like sequence on Chr. 14, and as expected for *ENSAP* (not shown). Also, a BLAST search detected multiple cDNA and EST entries identical with the corresponding 1q21 spliced exon sequences, but none corresponding to the Chr. 14 locus. Therefore, of the three *ENSA*-related loci in the human genome, only the gene on Chr. 1 is transcribed in the expected tissues and encodes the complete peptide, and we conclude that it represents the functional gene. Because of the structural organization of the locus at 14q, lack of detectable transcription, and significant changes in the coding sequence, we reasoned that it may represent another pseudogene generated by reverse transcription and integration of *ENSA* transcript III, described below. As this report was being written, an entry of a 674 bp sequence from Chr. 14 annotated as *ENSA* pseudogene 2 (*ENSAP2*) appeared in the GenBank (NG_002495). This sequence is identical with a corresponding segment of the sequence we detected in the BAC clones on 14q, thus verifying our interpretation.

During the BLAST searches for clones containing *ENSA* sequences, we noticed several human cDNAs and ESTs that varied by an insertion within the coding region and/or by their 5′ or 3′ ends, and all represented alternatively spliced transcripts from the *ENSA* gene on 1q (Fig. 1). Using this information and our experimental data obtained by PCR we conclude that *ENSA* can give rise to at least six different transcripts (I–VI; Fig. 1) and designate cDNA AF067170 (consisting of exons 1, 2, 3, and 4, and encoding the reported 121 amino acid peptide) as transcript I. As indicated in Fig. 1, transcript II encompasses an additional 48 bp exon (2B in Fig. 1) contributing an in-frame insertion of 16 amino acids after codon 97, transcript III consists of exons 1, 2, 3,

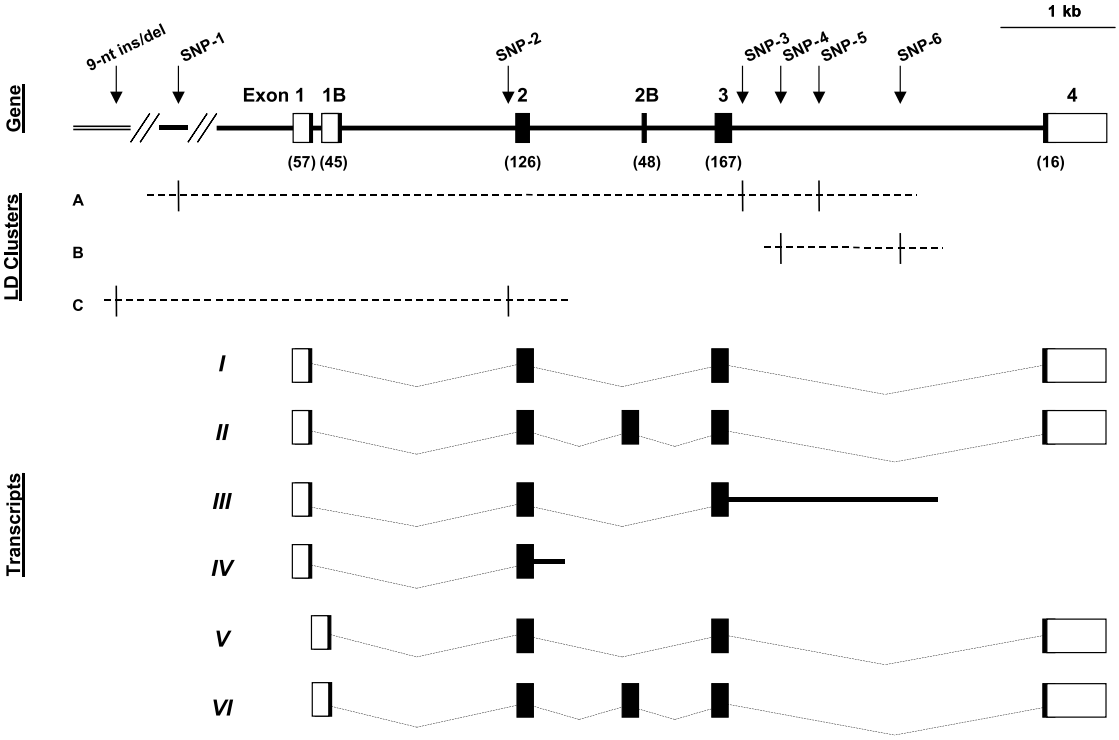

Fig. 1. Schematic diagram of *ENSA* genomic organization and transcripts. Exons are shown as boxes (black, coding; open, non-coding), and introns as a solid line. The sizes of the coding sequences (in bp) are indicated below the corresponding exons. LD clusters A, B, and C with the corresponding markers are indicated below the gene; transcripts are shown on the bottom. For exact positions of variants see GenBank Accession #AY326399–AY326403.

and retains intron 3 as 3′-UTR, and transcript IV is composed only of the first two exons and the beginning part of intron 2. Transcripts V and VI are similar to I and II, respectively, except of utilizing a different first exon which would replace the first 18 N-terminal amino acids of ENSA with 15 novel residues (1B in Fig. 1). Based on visual inspection of ethidium bromide-stained PCR products separated on agarose gels, transcripts I and III appear as the most abundant *ENSA* mRNA species in the examined tissues (not shown). It is presently unknown whether any of the alternatively spliced mRNAs are translated in vivo.

ENSA is located within a region on 1q21 previously linked with T2DM in the Pima Indians and other populations [9,15], and we therefore investigated this gene as a candidate for the disease in Pimas. We initially screened the gene for variants by PCR and sequencing of all six exons, adjacent intronic sequences, 2 kb of the promoter region, and the alternatively utilized 3′-UTRs in 10 diabetic and 10 non-diabetic Pima subjects. We identified six single nucleotide polymorphisms (SNPs) including SNP-1 in the promoter, SNP-2 in intron 1, and four (SNP-3 through SNP-6) in the last intron (Fig. 1), whereas no variants were found in the coding sequences. SNP-3 corresponds to rs1053732 already deposited in dbSNP. Upon examination of the genomic sequences further upstream from the first exon, we uncovered a diallelic length variant caused by an insertion/deletion of a 9-bp segment (TTCCCAAGA; 9-bp ins/del; GenBank AY326399) located about 5.9 kb upstream from the starting ATG codon in the first exon.

By comparison of the genotypes obtained initially by sequencing in 10 diabetic and 10 non-diabetic subjects, we were able to divide the markers into three separate linkage disequilibrium clusters (LD cluster A = SNP-1 + SNP-3 + SNP-5; cluster B = SNP-4 + SNP-6; cluster C = SNP-2 + 9-bp ins/del; Fig. 1A). For statistical analysis, we subsequently genotyped markers representative of each cluster (SNP-3, allelic frequency 0.72/0.28; SNP-4, 0.62/0.38; 9-bp ins/del, 0.69/0.31) in 149 affected (diabetic) and 150 control subjects. Based on these results, markers showing nominal associations (defined as $p < 0.05$) with T2DM (SNP-3: allelic frequency in affected subjects 0.69/0.31, in controls 0.77/0.23, $p < 0.04$; and 9-bp ins/del: frequency in affected subjects 0.74/0.26, in controls 0.64/0.36, $p = 0.04$) were subsequently analyzed in approximately 1200 Pimas used in the original genome scan that led to the identification of the linkage at 1q21–q23 [9]. In this extended group of subjects, neither marker was associated with T2DM ($p > 0.6$). Furthermore, in a subset of about 270 non-diabetic subjects who underwent detailed measurements of insulin action (M-low, M-high) and insulin secretion parameters (AIR, plasma insulin during OGTT), no associations of the markers with these clinical parameters, adjusted for age, sex, and % body fat, could be detected ($p > 0.1$).

ENSA has a ubiquitous expression with high levels of mRNA found in the brain and skeletal muscle [5]. Based on a previous study designed to compare mRNA profiles between skeletal muscle biopsies from non-diabetic, insulin-sensitive, and insulin-resistant Pimas by microarray hybridization [11], transcript III appears to represent the most abundant ENSA mRNA in this tissue (not shown). To assess potential inter-individual expression differences, we measured transcript III by real-time (RT) PCR in skeletal muscle biopsies from 49 non-diabetic Pimas including insulin-sensitive and insulin-resistant individuals. While the relative level of this transcript varied up to fourfold between subjects, we did not detect any statistically significant correlation with M-low, M-high, or plasma glucose and insulin during OGTT, adjusted for age, sex, and % body fat ($p > 0.3$). This indicates that variability of skeletal muscle *ENSA* mRNA between individuals is not related to differences in insulin action or secretion in non-diabetic Pimas.

In summary, we demonstrated that the expressed *ENSA* gene is located within a region on 1q21 linked with T2DM in several populations. After performing a detailed molecular analysis in diabetic and non-diabetic Pima Indians, we conclude that variants in this gene are unlikely to contribute to T2DM, insulin secretory dysfunction or insulin resistance in this population. However, verification of the position of the expressed gene, identification of variants, and detection of multiple transcripts provide valuable information for further evaluation of the function of the protein, and examination of this locus for alterations that may contribute to the development of T2DM or other diseases in humans.

Acknowledgments

We would like to express our appreciation to the members of the Gila River Indian Community for their participation and cooperation in the ongoing studies of T2DM, John Gerry and Shawnell Damon for technical assistance, Dr. Rob Hanson and Sayuko Kobes for statistical analyses, and Dr. Joy Bunt and Tom Brookshire for collection of skeletal muscle biopsies.

References

[1] K.S. Polonsky, Evolution of beta-cell dysfunction in impaired glucose tolerance and diabetes, Exp. Clin. Endocrinol. Diabetes 107 (1999) S124–127.
[2] S.E. Inzucchi, Oral antihyperglycemic therapy for type 2 diabetes: scientific review, JAMA 287 (2002) 360–372.
[3] F.M. Ashcroft, F.M. Gribble, Tissue-specific effects of sulfonylureas: lessons from studies of cloned K(ATP) channels, J. Diabetes Complications 14 (2000) 192–196.

[4] D. Bataille, L. Heron, A. Virsolvy, K. Peyrollier, A. LeCam, L. Gros, P. Blache, Alpha-endosulfine, a new entity in the control of insulin secretion, Cell. Mol. Life Sci. 56 (1999) 78–84.
[5] L. Heron, A. Virsolvy, K. Peyrollier, F.M. Gribble, A. Le Cam, F.M. Ashcroft, D. Bataille, Human alpha-endosulfine, a possible regulator of sulfonylurea-sensitive KATP channel: molecular cloning, expression and biological properties, Proc. Natl. Acad. Sci. USA 95 (1998) 8387–8391.
[6] A. Virsolvy, P. Smith, G. Bertrand, L. Gros, L. Heron, G. Salazar, R. Puech, D. Bataille, Block of $Ca^{(2+)}$-channels by alpha-endosulphine inhibits insulin release, Br. J. Pharmacol. 135 (2002) 1810–1818.
[7] L. Gros, B. Breant, B. Duchene, C. Leroy, G. Fauconnier, D. Bataille, A. Virsolvy, Localization of alpha-endosulphine in pancreatic somatostatin delta cells and expression during rat pancreas development, Diabetologia 45 (2002) 703–710.
[8] L. Heron, A. Virsolvy, F. Apiou, A. Le Cam, D. Bataille, Isolation, characterization, and chromosomal localization of the human *ENSA* gene that encodes alpha-endosulfine, a regulator of beta-cell K(ATP) channels, Diabetes 48 (1999) 1873–1876.
[9] R.L. Hanson, M.G. Ehm, D.J. Pettitt, M. Prochazka, D.B. Thompson, D. Timberlake, T. Foroud, S. Kobes, L. Baier, D.K. Burns, L. Almasy, J. Blangero, W.T. Garvey, P.H. Bennett, W.C. Knowler, An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians, Am. J. Hum. Genet. 63 (1998) 1130–1138.
[10] R.E. Pratley, D.B. Thompson, M. Prochazka, L. Baier, D. Mott, E. Ravussin, H. Sakul, M.G. Ehm, D.K. Burns, T. Foroud, W.T. Garvey, R.L. Hanson, W.C. Knowler, P.H. Bennett, C. Bogardus, An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians, J. Clin. Invest. 101 (1998) 1757–1764.
[11] X. Yang, R.E. Pratley, S. Tokraks, C. Bogardus, P.A. Permana, Microarray profiling of skeletal muscle tissues from equally obese, non-diabetic insulin-sensitive and insulin-resistant Pima Indians, Diabetologia 45 (2002) 1584–1593.
[12] M. Prochazka, V.S. Farook, V. Ossowski, J.K. Wolford, C. Bogardus, Variant screening of *PRKAB2*, a type 2 diabetes mellitus susceptibility candidate gene on 1q in Pima Indians, Mol. Cell. Probes 16 (2002) 421–427.
[13] F. Thameem, J.K. Wolford, J. Wang, M.S. German, C. Bogardus, M. Prochazka, Cloning, expression and genomic structure of human *LMX1A*, and variant screening in Pima Indians, Gene 290 (2002) 217–225.
[14] D.L. Wheeler, D.M. Church, S. Federhen, A.E. Lash, T.L. Madden, J.U. Pontius, G.D. Schuler, L.M. Schriml, E. Sequeira, T.A. Tatusova, L. Wagner, Database resources of the national center for biotechnology, Nucleic Acids Res. 31 (2003) 28–33.
[15] M.I. McCarthy, Growing evidence for diabetes susceptibility genes from genome scan data, Curr. Diab. Rep. 3 (2003) 159–167.