



ELSEVIER

Gene 241 (2000) 143–148

GENE

AN INTERNATIONAL JOURNAL ON
GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

Molecular characterization of the human *PEA15* gene on 1q21–q22 and association with type 2 diabetes mellitus in Pima Indians ☆

Johanna K. Wolford *, Clifton Bogardus, Victoria Ossowski, Michal Prochazka

Clinical Diabetes and Nutrition Section, Phoenix Epidemiology and Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 4212 N. 16th Street, Phoenix, AZ 85016, USA

Received 24 June 1999; received in revised form 1 October 1999; accepted 8 October 1999

Received by K. Gardiner

Abstract

The *PEA15* gene encoding a protein kinase C substrate is widely expressed, and its overexpression may contribute to impairment of glucose uptake. *PEA15* is located within a region on human 1q linked with type 2 diabetes in both Pima Indians and Caucasians. To assess the potential contribution of genetic alterations within this locus to disease susceptibility in the Pimas, we have investigated its genomic sequences. The *PEA15* locus is composed of four exons spanning approximately 10.2 kb of genomic DNA, flanked upstream by an potentially expressed *Alu* element, downstream by the *H326* gene, and is located within 250 kb of *KCNJ9*. We also sequenced over 2 kb of the promoter region and identified various motifs analogous to known transcription factor binding sites. By analysis of 22 Pimas, including 13 diabetic subjects, we detected four single nucleotide polymorphisms (SNPs) in the non-coding regions of *PEA15*, including three frequent variants that were in allelic disequilibrium, and one variant found only in a single Pima. The three SNPs were not associated with type 2 diabetes mellitus in 50 affected and 50 control Pimas ($p=0.12$ – 0.17), and we conclude that mutations in this gene probably do not contribute significantly to disease susceptibility in this Native American tribe. However, knowledge of the genomic structure of *PEA15* provides the basis for similar systematic examinations of this candidate locus in relation to type 2 diabetes and other metabolic disorders in other populations. Published by Elsevier Science B.V.

Keywords: Association; Candidate gene; Chromosome 1q; Linkage; Single nucleotide polymorphism; Type 2 diabetes mellitus

1. Introduction

Susceptibility to type 2 diabetes mellitus is heritable and has a complex genetic basis [reviewed in Elbein (1997)]. We have recently reported the results of a genome-wide linkage study of type 2 diabetes in the Pima Indians of Arizona, a population with the highest documented prevalence of this disease world-wide. We have found suggestive evidence for linkage of diabetes with a 30 cM interval on 1q21–q23, showing a maximum multipoint LOD score 2.48 at *DIS1677* (Hanson et al., 1998). This finding was consistent with a recent observa-

tion of linkage of type 2 diabetes with markers within the same region in Caucasian families living in Utah (Elbein et al., 1999). Considering the multiple genes that likely contribute to type 2 diabetes, evidence for linkage of the disease with 1q21–q23 obtained independently in Pimas and Caucasians strongly indicates the presence of a diabetes susceptibility locus in this area.

The 1q21–q23 region is gene-rich (<http://www.ncbi.nlm.nih.gov/LocusLink>). Our strategy to identify the susceptibility locus included investigation of potential candidates among known genes previously mapped to this interval. One gene, previously called *PED/PEA-15* [(phosphoprotein enriched in diabetes/phosphoprotein enriched in astrocytes; henceforth referred to by its official symbol *PEA15* (<http://www.ncbi.nlm.nih.gov/Omim/>))] that has been localized to 1q21–q22 by radiation hybrid mapping, appeared particularly relevant based on recent findings in diabetic subjects (Condorelli et al., 1998). The 15 kDa *PEA15* protein encoded by this gene is a substrate for protein

Abbreviations: *PEA15*, gene encoding PED/PEA-15; PED/PEA-15, phosphoprotein enriched in diabetes/ phosphoprotein enriched in astrocytes; SNP, single nucleotide polymorphism.

☆ Sequences reported in this article were deposited with the GenBank database under Accession Nos AF153272–AF153274.

* Corresponding author. Tel.: +1-602-200-5341; fax: +1-602-200-5335.

E-mail address: jwolford@exchange.nih.gov (J.K. Wolford)

kinase C, and, although originally identified in astrocytes (Arujo et al., 1993), it is now known to be widely expressed (Danziger et al., 1995; Condorelli et al., 1998; multiple ESTs of diverse tissue/cell origin listed in dbEST). Using differential display, Condorelli et al. (1998) discovered that *PEA15* mRNA and protein are significantly elevated in primary fibroblast cultures, as well as skeletal muscle and adipose tissue biopsies from people with type 2 diabetes mellitus. Although the biological function of the product is not yet known, overexpression of the protein in rodent and human cell lines inhibits glucose transport, suggesting a mechanism for the development of insulin resistance and diabetes (Condorelli et al. 1998). *PEA15* therefore represented an excellent candidate for further investigation in the Pima population due to its localization to 1q21–q22, potential effects on glucose transport, and overexpression in some subjects with type 2 diabetes mellitus.

The goal of this study was to determine the genomic structure of *PEA15* and identify potential allelic variants as a first step toward exploring this gene as a possible candidate for diabetes susceptibility. In this report, we resolved the exon–intron structure of the gene, sequenced ~2 kb of the promoter region, and found that *PEA15* is flanked upstream by an *Alu* element and downstream by the *H326* gene. Furthermore, we describe four single nucleotide polymorphisms, including three located within the first two introns plus one in the 5' flanking sequence, and report the results of their association analyses with type 2 diabetes mellitus in the Pimas.

2. Materials and methods

2.1. Subjects and genomic DNA

Subjects selected for genomic screening are participants of ongoing longitudinal studies of diabetes conducted among members of the Gila River Indian Community since 1965 (Lillioja et al., 1993). Diabetes was diagnosed by WHO criteria as described previously (Hanson et al., 1998). Association studies were performed by comparing allele and genotype frequencies at individual SNPs in 100 Pimas, representing 50 subjects diagnosed with type 2 diabetes mellitus before the age of 2 years and 50 unaffected subjects with normal glucose tolerance who were at least 45 years of age.

Pima genomic DNA samples were prepared as previously described (Prochazka et al., 1993). Genomic DNA samples from 54 unrelated Center d'Etude Polymorphisme Humain (CEPH) Caucasian individuals were obtained from the NIGMS Human Genetic Mutant Cell Repository (Coriell Institute, Camden, NJ).

2.2. YAC clones, PCR, sequencing, and primer extension

Using the Whitehead Institute database (<http://www-genome.wi.mit.edu/>), we identified several YACs containing STS WI-9524, which represents the *PEA15* 3' UTR. Of these YACs, clones 915G8 and 930F5 were used to determine the structure of *PEA15* by long distance PCR, Alu- and Vectorette-PCR as described previously (Wolford and Prochazka, 1998). All sequences were determined with an ABI Automated Sequencer Model 373A or 377 using the ABI Prism BigDye Terminator Cycle Sequencing kit (PE Applied Biosystems, Foster City, CA). The promoter region was analyzed through the Baylor College of Medicine search launcher (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html>) using programs designed for promoter and transcription binding site predictions (TESS, NNPP/Eukaryotic, and MatInspector TRANSFAC V2.2).

Adult human brain and skeletal muscle QUICK-Clone cDNAs were purchased from Clontech (Palo Alto, CA). The 5'RACE experiment was performed with human fetal brain Marathon-ready cDNA (Clontech), and PCR was performed using *PEA15* cDNA-specific antisense primer 5'-AAGAGGA-ACTAGATGAGCTAG-3' in combination with an adapter primer (AP1) provided with the cDNA. The products were subcloned into the TA-vector (Invitrogen, Carlsbad, CA) for sequencing.

Primer extension analysis was performed according to the manufacturer's instructions (GeneAmp ThermoStable rTth Reverse Transcriptase, PE Biosystems), with some modifications using 0.3, 1.0, and 5.0 µg of human fetal brain poly-A⁺ RNA (Clontech). RNA was heated at 95°C for 10 min, [γ -³²P] ATP (NEN LifeScience; Boston, MA) 5'-end labeled oligonucleotide PE2R (5'-GGTTCCTAAGCAGCGTCC) was added, and the reaction was heated at 95°C for an additional 5 min. Reverse transcriptase was added, and extension was carried out at 70°C for 20 min. Extension products were resolved on a 6% urea-polyacrylamide gel. Gels were dried and exposed to Biomax-MR film at -70°C with an intensifying screen.

2.3. Comparative sequencing and SNP genotyping

Information on primers used for PCR and comparative sequencing of *PEA15* is presented in Table 1. PEA-SNP1 was amplified from genomic DNA with primers 5'-CACATGAGTTCCTACCAGAC-3' and 5'-CTAAC-AGTAACCCTGGCAAC-3', PEA-SNP3 with primers 5'-GACACTAAGGCTTTCAGGC-3' and 5'-ATAGC-ATTTCTCTGCTGTTTC-3', and PEA-SNP4 with primers 5'-CTGCTGGTCTACATGTGC-3' and 5'-CCAAACAACAATACAGAGGAC-3'. Digestion with the appropriate restriction enzymes (PEA-SNP1

Table 1
PEA15 primers for genomic PCR and comparative sequencing^a

Designation	PCR primer (5'→3')	Gene segment
PEAPROM2	GAT CAG GTG TAA AGT TGT TGC	5' region
PEAPROM1R	CCT GAA CTG CGA AAC AAT TC	
PEAPROM3A	GTG GAT AAG TCT GAT CAG TC	Exon 1
PEA1R	TCC CAA ACT GCC GTT CAG C	
PEAEX2F	CCA TGT TCC AGG CAC ATG C	Exon 2
PEAEX2R	ATT TGC TGA ATG AAC TGA GCC	
PEAEX3F	AGC TTT GGT TCC AGG TCA C	Exons 3+4
PEA2R	AAG AGG AAC TAG ATG AGC TAG	
PEA2A	GGA GAG TGG AGG CAG ATG	3' UTR-a
PEA7R	GAC TAT GTA GCA ATT TGA TTC TGC	
PEA7	CAC TAC AGT ACA GTG AGT GA	3' UTR-b
H326-1	TCT GCC TCA GAT CTC CTA C	

^a The primer direction is indicated by F (forward) and R (reverse), respectively, according to the transcriptional orientation of *PEA15*. The 5' region and 3' UTR were amplified using the Expand Long Template PCR system, the GC-rich exon 1 was amplified in the presence of 7deazadGTP (Rowles et al., 1996), and the remaining exons were amplified by standard PCR (Wolford and Prochazka, 1998). Because of the size, 3' UTR was amplified in two overlapping segments (3' UTR-a and -b, respectively). For analysis of exons 1, 2, 3+4, respectively, the same primers were used for PCR as well as sequencing. Additional internal primers were designed to complete sequencing of the 5' region and 3' UTR.

with *Bfa*I, *PEA*-SNP3 with *Fnu*4HI, *PEA*-SNP4 with *Hae*III) was carried out as recommended by the manufacturer (New England Biolabs, Boston, MA; Gibco BRL, Gaithersburg, MD), and the fragments were resolved on 3–4% agarose gels. *PEA*-SNP2, a single base insertion/deletion, was analyzed by amplification with ³²P labeled primer 5'-ACTTTGAGAGACTAC-TGCTTG-3' combined with unlabeled primer 5'-ATGCCCTCCTTTGGATCTTG-3' under conditions described previously (Wolford et al., 1998).

Only SNPs having a frequency of at least 0.2 for the less common allele were used in association analyses comparing 50 affected and 50 unaffected control Pimas who were not first-degree relatives. Genotyping results were scored independently by two investigators, and 17 'blind' duplicate samples were included to guarantee data reproducibility; only data with no discrepancies between the duplicate typings were analyzed. Chi-square analysis (SAS Institute, Cary, NC) was used to assess associations of allelic variants at individual SNPs with diabetes.

3. Results and discussion

3.1. Genomic structure of *PEA15*

By searching the STS database with the published cDNA sequence (Estelles et al., 1996; GenBank No. X86809), two *PEA15*-specific STS markers were identified: WI-9524 and TIGR-A002626. Based on the STS content information of CEPH YACs available from the Whitehead Institute database, we identified several clones containing *PEA15*, including 915G8 and 930F5, which were subsequently used to determine the genomic organization of this locus.

PEA15 has previously been found to have a 393 bp open reading frame encoding a protein of 130 amino acids with a predicted molecular weight of 15 kDa (Arujo et al., 1993; Estelles et al., 1996). We found that *PEA15* consists of four exons spanning approximately 10.2 kb. The first exon and beginning of exon 2 contain untranslated sequences, whereas the coding sequence is represented in exons 2–4 (Fig. 1). All exon–intron boundaries (Table 2) are in agreement with the GT–AG consensus motif of splice donor and acceptor sites in eukaryotic genes (Shapiro and Senapathy, 1987). To examine the locus for additional, alternatively spliced exons or utilization of alternate splice sites, we used cDNA templates from skeletal muscle and brain for RT-PCR with selected primer pairs spanning the entire transcript. We detected single bands consistent with the size predicted from the published cDNA and conclude that the genomic structure shown in Fig. 1 represents a complete *PEA15* gene.

3.2. Characterization of *PEA15* promoter and flanking regions

To determine the 5' flanking area overlapping the promoter region, we isolated approximately 2.5 kb of sequence upstream from the first exon using *Alu*- and *Vectorette*-PCR with YAC 915G8. Two human *PEA15* transcripts (1.5 and 2.4 kb, respectively) were previously detected by Northern blotting, and their size difference was explained by the alternate usage of two polyadenylation sites (Estelles et al., 1996). As the size of the largest published cDNA (2385 bases) accounts almost completely for the 2.4 kb mRNA band, we anticipated that the transcription start site should be located close to the 5' end of the first (non-coding) exon.

Primer extension analysis using 5 µg of human fetal

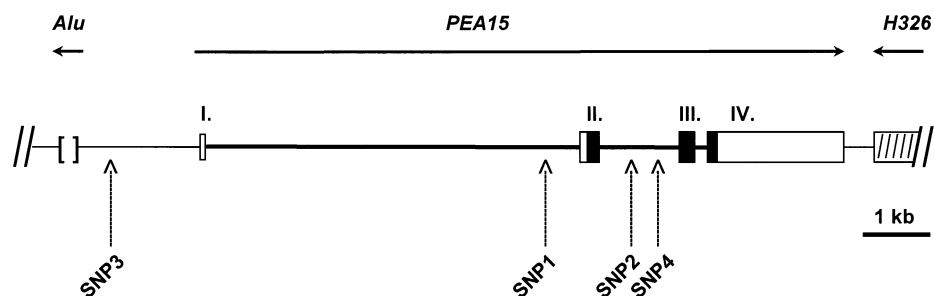


Fig. 1. Structure of the human *PEA15* gene and flanking regions. *PEA15* coding and non-coding sequences are shown as black and white boxes, respectively, and the 3' end of *H326* is shown as a striped box. *PEA15* introns are depicted as thick horizontal lines; 5' and 3' flanking sequences are indicated by thin horizontal lines. Horizontal arrows above the diagram correspond to the transcriptional orientation of the corresponding genes. The positions of SNP markers are shown by dotted vertical arrows at the bottom of the figure (see Table 2 for details on each SNP). The location of the tentatively expressed *Alu* element (AA491823) 5' of *PEA15* is represented by brackets.

brain poly-A⁺ RNA revealed two clustered bands. A comparison with a sequencing ladder produced with the same primer from a genomic DNA segment encompassing this area revealed that the larger band mapped to a C and the second band to a G residue located 89 and 85 bases, respectively, upstream from the end of the first exon. Analogous information was obtained with the 'NNPP/Eukaryotic' program of the Baylor College of Medicine search launcher (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html>), which predicted three transcription start sites clustered within a 25 bp interval spanning the same region. In addition, we performed a 5'RACE experiment using a human fetal brain cDNA template that yielded a single prominent band (as determined by agarose gel electrophoresis) of a size consistent with a relatively narrow region of the transcription start point (not shown). Sequencing of several 5'RACE clones revealed microheterogeneity of the 5' ends that were all clustered around the predicted sites. For orientation purposes, we therefore designated the C residue detected by the primer extension as +1 (Fig. 2).

Analysis of the 5' genomic sequence revealed that this gene lacks a TATA box and the first 600 bp flanking the transcription start region have a high GC content (64%). Such features are typical for housekeeping genes (Gardiner-Garden and Frommer, 1987) and are consistent with the nearly ubiquitous expression of *PEA15*. As shown in Fig. 2, this region contains matches with

the consensus-binding sites for numerous regulatory proteins, including AP2, CREB, MZF1, SP1, etc. Functional analyses will be necessary to define which, and how many, of these elements are necessary for the expression of *PEA15* in vivo. However, none of the human sequence motifs implied in mediating the regulatory effects of insulin or glucose was detected in this region.

Analysis of adjacent genomic sequences revealed that *PEA15* is flanked immediately downstream by *H326*, a homologue of the mouse *Pc326* gene encoding a protein characterized by similarities to the members of the β -transducin gene family (Bergsagel et al., 1992). Moreover, approximately 2 kb upstream from the first exon, we detected an *Alu* element that is 100% identical with the EST entry AA491823 isolated from a human liposarcoma cDNA library (Krizman et al., 1996). Although this genomic sequence has up to 89% homology with several *Alu* subclasses in the public databases, the insert of AA491823 (including 276 bases of the repeat plus 69 bases of the preceding unique sequence) perfectly matches the 5' genomic sequence flanking *PEA15*. Some human *Alu* sequences have the potential for being expressed (Szmulewicz et al., 1998), and, based on our data, we conclude that AA491823 may represent such a transcriptionally active repeat.

As indicated in Fig. 1, *H326* and AA491823 have an opposite transcriptional orientation relative to *PEA15*. Clustering of three transcription units within a few

Table 2
Characteristics of SNP markers detected in *PEA15*

	SNP1 ^a	SNP2	SNP3	SNP4
Location ^b	Intron 1 (–428)	Intron 2 (–504)	5' region (–1325)	Intron 2 (–285)
Alleles	G/A	AAA/AA	C/T	A/G
Pimas (frequency)	0.49/0.51	0.49/0.51	0.47/0.53	0.99/0.01
Caucasians (frequency)	0.44/0.56	0.48/0.52	0.43/0.59	Not determined

^a The numbering of SNPs reflects their chronological detection during comparative sequencing of the gene in the Pimas.

^b Numbers in parentheses indicate distances (in bp) from the beginning of the following exon (SNP1, SNP2, SNP4) or from the first transcription start site (SNP3).

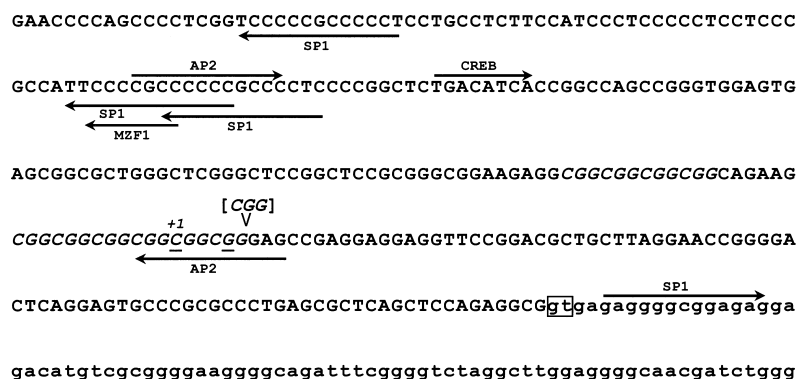


Fig. 2. Partial 5' flanking genomic sequence encompassing the first exon of *PEA15*. The two transcription start sites determined by primer extension are underlined. For orientation, the first site is numbered as +1. The beginning of intron 1 is shown in lower case, and the splice donor site is boxed. Both CGG clusters are italicized, and an additional triplet above the sequence marks the putative variant observed in YAC 915G8. Arrows indicate predicted binding sites for selected common transcription factors with their direction corresponding to the location of the motifs (the forward and reverse orientation corresponds to the sense and complementary sequence, respectively). AP2, activator protein 2; CREB, cyclic AMP-responsive element binding protein; MZF1, myeloid zinc finger protein 1; SP1, stimulating protein 1. The entire upstream sequence was deposited in GenBank under Accession No. AF153272.

kilobases apart is consistent with this chromosomal region being gene-rich. Furthermore, we found that *PEA15* and *APOA2* both reside on YAC clones 915G8 and 930F5, which have reported insert sizes of approximately 250 and 490 kb, respectively (<http://www.genome.wi.mit.edu>). It has been demonstrated that 1q21–q25 in human is highly syntenic with regions on mouse chromosomes 1 and 3 (Seldin et al., 1988). The physical proximity between *PEA15* and *H326* near *APOA2* found in this study is also similar for the homologous murine genes (Underhill et al., 1999).

3.3. Comparative sequencing and polymorphism analysis

All exons and flanking areas (including parts of introns, approximately 2 kb of the 5' region, and the entire 3'UTR encompassing the second polyadenylation site) were initially screened for variants by comparative sequencing in 13 diabetic and nine non-diabetic Pimas, one Caucasian, and YAC 915G8. No variants were detected within the coding sequences or splice junctions in the Pimas, but we identified four SNPs (designated PEA-SNP1 through PEA-SNP4) in the non-coding parts of the gene. PEA-SNP1, located 428 bp upstream of exon 2, is a G→A substitution that produces a *BfaI* restriction site. PEA-SNP2 is an insertion/deletion of a single A residue (AA vs. AAA) between bases 504 and 506 in intron 2. PEA-SNP3, located 1325 bases upstream from the first predicted transcription start site, consists of a C→T substitution that yields a *Fnu4HI* restriction site. PEA-SNP4 is an A→G exchange 285 bases upstream from the beginning of exon 3 that results in a gain of a *HaeIII* restriction site (Table 2).

With the exception of PEA-SNP4, which was detected in only one subject, the frequencies of the remaining variants were subsequently determined in 100 Pimas (50

affected and 50 control subjects) and 54 unrelated Caucasian subjects (Fig. 1). Chi-square analysis of these three markers found no associations with diabetes ($p = 0.12$ – 0.17). Because the alleles at PEA-SNP1, -SNP2, and -SNP3 were in linkage disequilibrium among all but two subjects ($p = 0.001$), no haplotype analysis was performed.

As presented in Fig. 2, the region encompassing the transcription start site contains two short tandem CGG repeats [(CGG)₄CAGAAG (CGG)₆]. Although we did not detect any differences of this motif among the Pimas or the Caucasian DNA NA10858, we found that the corresponding sequence in YAC 915G8 contains an additional triplet within the second cluster [(CGG)₄CAGAAG(CGG)₇], indicating that this segment may exhibit a length variation. As shown in Fig. 2, the region containing the CGG cluster encompasses the putative transcription start sites and could serve as a potential AP2-binding site. Moreover, expression of many genes is regulated by methylation of CpG motifs in their 5' flanking region (Dyner, 1989). Therefore, allelic differences in the number of CGG motifs within this segment may affect the expression of *PEA15*, e.g. by modifying the interaction of this region with transcription initiation factors and/or by contributing to differential methylation between the alleles.

Previously, Condorelli et al. (1998) noted four codon discrepancies predicted by single base differences between the open reading frames of the cDNA reported by their group and the sequence published by Arujo et al. (1993). We found that the corresponding genomic sequences that we obtained in all Pimas, one Caucasian, and YAC 915G8 match the coding sequences published by Condorelli et al. (1998), indicating that the reported discrepancies may be spurious.

4. Conclusion

In summary, our investigation of *PEA15* by comparative sequencing and association analysis of diallelic markers did not reveal any statistically significant frequency differences between diabetic and non-diabetic Pima Indians. Based on these observations and the absence of any mutation(s) affecting the expression or function of the protein, we conclude that an alteration within *PEA15* is unlikely to underlie the linkage of 1q21–q23 with diabetes in this Native American population. However, elucidation of the genomic structure of *PEA15* and detection of informative SNPs will be valuable in evaluating this locus for potential involvement in the pathogenesis of type 2 diabetes in other populations.

Acknowledgements

We thank the members of the Gila River Indian Community for their participation in the ongoing longitudinal studies. The authors would also like to acknowledge the assistance of Dr Jeffrey Bunkelmann, Bonnie Cooper, and Nancy Riebow.

References

- Arujo, H., Danziger, N., Cordier, J., Glowinski, J., Chneiweiss, H., 1993. Characterization of PEA-15, a major substrate for protein kinase C in astrocytes. *J. Biol. Chem.* 268, 5911–5920.
- Bergsagel, P.L., Timblin, C.R., Eckhardt, L., Laskov, R., Kuehl, W.M., 1992. Sequence and expression of a murine cDNA encoding PC326, a novel gene expressed in plasmocytomas but not normal plasma cells. *Oncogene* 7, 2059–2064.
- Condorelli, G., Vigliotta, G., Iavarone, C., Caruso, M., Tocchetti, C.G., Andreozzi, F., Cafieri, A., Tecce, M.F., Formisano, P., Beguinot, L., Beguinot, F., 1998. *PED/PEA-15* gene controls glucose transport and is overexpressed in type 2 diabetes mellitus. *EMBO J.* 17, 3858–3866.
- Danziger, N., Yokoyama, M., Jay, T., Cordier, J., Glowinski, J., Chneiweiss, H., 1995. Cellular expression, developmental regulation, and phylogenetic conservation of PEA-15, the astrocytic major phosphoprotein and protein kinase C substrate. *J. Neurochem.* 64, 1016–1025.
- Dynan, W.S., 1989. Understanding the molecular mechanism by which methylation influences gene expression. *Trends Genet.* 5, 35–36.
- Elbein, S.C., 1997. The genetics of human noninsulin-dependent (type 2) diabetes mellitus. *J. Nutr.* 127, 1891S–1896S.
- Elbein, S.C., Hoffman, M.D., Teng, K., Leppert, M.F., Hasstedt, S.J., 1999. A genome-wide search for type 2 diabetes susceptibility gene in Utah Caucasians. *Diabetes* 48, 1175–1182.
- Estelles, A., Yokoyama, M., Nothias, F., Vincent, J.D., Glowinski, J., Vernier, P., Chneiweiss, H., 1996. The major astrocytic phosphoprotein PEA-15 is encoded by two mRNAs conserved on their full length in mouse and human. *J. Biol. Chem.* 271, 14800–14806.
- Gardiner-Garden, M., Frommer, M., 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* 196, 261–282.
- Hanson, R.L., Ehm, M.G., Pettitt, D.J., Prochazka, M., Thompson, D.B., Timberlake, D., Foroud, T., Kobes, S., Baier, L., Burns, D.K., Almasy, L., Blanger, J., Garvey, W.T., Benneett, P.H., Knowler, W.C., 1998. 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.
- Krizman, D.B., Chuaqui, R.F., Meltzer, P.S., Trent, J.M., Duray, P.H., Linehan, W.M., Liotta, L.A., Emmert-Buck, M.R., 1996. Construction of a representative cDNA library from prostatic intraepithelial neoplasia. *Cancer Res.* 56, 5380–5383.
- Lillioja, S., Mott, D.M., Spraul, M., Ferraro, R., Foley, J.E., Ravussin, E., Knowler, W.C., Bennett, P.H., Bogardus, C., 1993. Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus: prospective studies in the Pima Indians. *N. Engl. J. Med.* 329, 1988–1992.
- Prochazka, M., Lillioja, S., Tait, J.F., Knowler, W.C., Mott, D.M., Spraul, M., Bennett, P.H., Bogardus, C., 1993. Linkage of chromosomal markers on 4q with a putative gene determining maximal insulin action in Pima Indians. *Diabetes* 42, 514–519.
- Rowles, J., Scherer, S.W., Xi, T., Majer, M., Nickle, D.C., Rommens, J.M., Popov, K.M., Harris, R.A., Riebow, N.L., Xia, J., Tsui, L.C., Bogardus, C., Prochazka, M., 1996. Cloning and characterization of PDK4 on 7q21.3 encoding a fourth pyruvate dehydrogenase kinase isoenzyme in human. *J. Biol. Chem.* 271, 22376–22382.
- Seldin, M.F., Morse, H.C., LeBoeuf, R.C., Steinberg, A.D., 1988. Establishment of a molecular genetic map of distal mouse chromosome 1: Further definition of a conserved linkage group syntenic with human chromosome 1q. *Genomics* 2, 48–56.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15, 7155–7174.
- Szmulewicz, M.N., Novick, G.E., Herrera, R.J., 1998. Effects of *Alu* insertions on gene function. *Electrophoresis* 19, 1260–1264.
- Underhill, D.A., Mullick, A., Groulx, N., Beatty, B.G., Gros, P., 1999. Physical delineation of a 700-kb region overlapping the *Looptail* mutation on mouse Chromosome 1. *Genomics* 55, 185–193.
- Wolford, J.K., Prochazka, M., 1998. Structure and expression of the human *MTG8/ETO* gene. *Gene* 212, 103–109.
- Wolford, J.K., Bogardus, C., Prochazka, M., 1998. Polymorphism in the 3' untranslated region of *MTG8* is associated with obesity in Pima Indians. *Biochem. Biophys. Res. Commun.* 246, 624–626.