THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 275, No. 47, Issue of November 24, pp. 36910–36919, 2000
Printed in U.S.A.

Autoregulation and Maturity Onset Diabetes of the Young Transcription Factors Control the Human PAX4 Promoter*

Received for publication, June 15, 2000, and in revised form, August 11, 2000
Published, JBC Papers in Press, August 30, 2000, DOI 10.1074/jbc.M005202200

Stuart B. Smith‡§, Hirotaka Watada‡§, David W. Scheel‡, Caroline Mrejen‡¶, and Michael S. German‡‖**

From the ‡Hormone Research Institute and ‖Department of Medicine University of California, San Francisco, California 941423-0534

During pancreatic development, the paired homeodomain transcription factor PAX4 is required for the differentiation of the insulin-producing beta cells and somatostatin-producing delta cells. To establish the position of PAX4 in the hierarchy of factors controlling islet cell development, we examined the control of the human *PAX4* gene promoter. In both cell lines and transgenic animals, a 4.9-kilobase pair region directly upstream of the human *PAX4* gene transcriptional start site acts as a potent pancreas-specific promoter. Deletion mapping experiments demonstrate that a 118-base pair region lying approximately 1.9 kilobase pairs upstream of the transcription start site is both necessary and sufficient to direct pancreas-specific expression. Serial deletions through this region reveal the presence of positive elements that bind several pancreatic transcription factors as follows: the POU homeodomain factor HNF1α, the orphan nuclear receptor HNF4α, the homeodomain factor PDX1, and a heterodimer composed of two basic helix-loop-helix factors. Interestingly, mutations in the genes encoding four of these factors cause a dominantly inherited form of human diabetes called Maturity Onset Diabetes of the Young. In addition, PAX4 itself has at least two high affinity binding sites within the promoter through which it exerts a strong negative autoregulatory effect. Together, these results suggest a model in which PAX4 expression is activated during pancreatic development by a combination of pancreas-specific factors but is then switched off once PAX4 protein reaches sufficient levels.

Organogenesis of the mammalian pancreas and its differentiation into discrete populations of endocrine and exocrine cells depends on the tightly regulated temporal and spatial expression of an array of transcription factors (for reviews see Refs. 1–3). Some of these factors, such as HNF3β (4, 5), PDX1 (6, 7), and HLXb9 (8, 9), are involved in the formation and outgrowth of the dorsal and ventral pancreatic buds. Within the forming pancreas, the basic helix-loop-helix transcription factor neurogenin3 functions as a pro-endocrine gene, initiating the program of endocrine differentiation in selected cells (10, 11). The complete differentiation and maturation of the endocrine cells requires additional factors, including neuroD1/BETA2 (12), NKX2.2 (13), ISL1 (14), and PAX6 (2, 15).

Once neurogenin3 expression sets a progenitor cell on a course of endocrine differentiation, additional factors are necessary to determine which of the four endocrine cell types it will become. One of these islet cell-type determination factors is the paired homeodomain transcription factor PAX4. PAX4 functions as a potent transcriptional repressor and is expressed only transiently in the fetal pancreas, peaking during the period of beta and delta cell differentiation (16). Mice with a targeted disruption of the *pax4* gene have a marked decrease in beta and delta cells with a commensurate increase in alpha cells (17), suggesting that PAX4 functions by repressing the alpha cell differentiation program and thereby deviates cells from a default alpha cell fate to an alternate beta or delta cell fate. In this model, once beta/delta cell type fate has been set, PAX4 expression is switched off.

To understand the mechanisms that control the expression of PAX4 and thereby impact endocrine cell type fate decisions, we mapped and characterized the regulatory regions of the human *PAX4* gene. We found that a small promoter element lying approximately 1.9 kb[1] upstream from the transcription start site of the *PAX4* gene is both necessary and sufficient to drive pancreas-specific expression. Remarkably, this sequence contains at least four functionally important binding sites for transcription factors, including four factors, HNF4α, HNF1α, pdx1, and neuroD1, that have been implicated in human beta cell dysfunction and early onset diabetes. Simultaneous exogenous expression of the above factors can stimulate the activity of the defined promoter element by 9-fold in a fibroblast cell line. In addition, the *PAX4* gene promoter contains several binding sites for PAX4 itself, suggesting a model in which PAX4 inhibits its own expression, providing transient, self-limited expression of the *PAX4* gene during pancreatic development.

MATERIALS AND METHODS

5′-Rapid Amplification of cDNA Ends (RACE)—The 5′ ends of human and murine PAX4 cDNA were identified by 5′-RACE, using a modification of the protocol from the 5′-RACE System 2.0 (Life Technologies, Inc.). 2.5 pmol of either human PAX4-specific primer (5′-ggacactcacccttcagctg-3′) or murine PAX4 primer (5′-tcacacagccattagata-3′) were annealed to 1 μg of human islet RNA or murine E15.5 fetal pancreas RNA. Reverse transcription was carried out using Superscript II reverse transcriptase (Life Technologies, Inc.) according to

* This work was supported in part by National Institutes of Health Grants DK55340 and DK21344. 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.
§ Recipient of fellowship from the Juvenile Diabetes Foundation.
¶ Recipient of an American Diabetes Association fellowship.
** To whom correspondence should be addressed: Hormone Research Institute, University of California, San Francisco, CA 94143-0534. Tel.: 415-476-9262; Fax: 415-731-3612; E-mail: german@cgl.ucsf.edu.

[1] The abbreviations used are: kb, kilobase pair; CMV, cytomegalovirus; EMSA, electrophoretic mobility shift assay; HSV-TK, herpes simplex virus thymidine kinase; RT-PCR, reverse transcriptase polymerase chain reaction. MODY, maturity onset diabetes of the young; RACE, rapid amplification of cDNA ends; bp, base pair; bHLH, basic helix-loop-helix protein.

the manufacturer's instructions. After first strand synthesis, the mRNA template was digested with RNase, and a homopolymeric tail added to the 3′ end using terminal deoxynucleotidyltransferase. 35 cycles of primary PCR were used to amplify fragments, followed by 35 cycles of nested PCR according to the manufacturer's instructions, using Universal Amplification primers (Life Technologies, Inc.) in conjunction with PAX4-specific primers: human primer 1 (5′-ggcaattcgagccaccacag-3′) and primer 2 (5′-gcggatccaatgccctttggctccaag-3′); murine primer 1 (5′-ggctccgtgaaatgtcacag-3′) and primer 2 (5′-ttgctagctgcacaatctgc-3′). PCR products were subcloned into BluescriptII (Stratagene) and sequenced.

Cell Culture and Transfection—NIH3T3, αTC1.6, and βTC3 cells were cultured as described previously (18) and transiently transfected using TRANSFAST lipid reagent (Promega) according to the manufacturer's instructions. 2 μg of DNA were transfected per million cells. Where expression vectors were cotransfected, 0.2 μg were transfected per million cells. Cells were harvested and extracts assayed for luciferase activity as described previously (19) using 2 μg of cellular protein extract.

The human *PAX4* gene was obtained from human cosmid clone g1572c264, and the sequences between −4958 bp and +46 relative to the transcription start site were subcloned upstream of the firefly luciferase gene in the pFOXLuc1 plasmid (20). Truncated forms of the PAX4 promoter and a full-length promoter with a deletion between −1963 and −1845 bp were amplified by PCR from the full-length promoter and confirmed by sequencing. The construction of the plasmid expressing luciferase under the control of the Rous sarcoma virus-long terminal repeat has been described previously (21).

Construction of pBAT12.mPAX4, the DNA plasmid expressing mouse PAX4 under the control of the human cytomegalovirus (CMV) immediate-early gene promoter, has been described previously (16). Mutations in the paired domain (S43P and R44P) and the homeodomain (N220P and R222P) in the mouse PAX4 cDNA were produced in the pBAT12.mPAX4 plasmid and confirmed by sequencing. Each mutation has previously been shown to disrupt DNA binding through the paired domain and homeodomain, respectively (22, 23).

Two complementary oligonucleotides encompassing the pancreas-specific domain of the human PAX4 promoter (bp −1960 to −1909, 5′-gatctggcaagacctttgagttaatgtataattgtgagcagatggcggggg-3)′ were ligated into *Bgl*II/*Bam*HI sites upstream of the herpes simplex virus thymidine kinase (HSV-TK −109 to + 51bp) promoter driving luciferase expression in the pFOXluc1 vector and confirmed by sequencing.

In Vitro Protein Expression—The construction of the *in vitro* expression vector containing the mouse PAX4 cDNA has been described previously (16). We used a carboxyl-terminally truncated PAX4 cDNA terminating at codon 274 because the synthesis of full-length PAX4 is less efficient but gives the same binding specificity (16); all results with truncated PAX4 were confirmed with full-length PAX4. The mouse PAX6 expression vector (pmPAX6) (24) was a gift from M. Busslinger, Research Institute of Molecular Pathology, Vienna, Austria. PAX4 and PAX6 proteins were produced by *in vitro* transcription/translation reaction, using the TNT-coupled reticulocyte lysate system (Promega, Madison, WI) according to the manufacturer's instructions.

Electromobility Shift Assays (EMSA)—Single-stranded oligonucleotide probes were 5′-end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase. Labeled oligonucleotides were column-purified and annealed to an excess of the complementary strand. EMSA buffers and electrophoresis conditions were as described previously (18) using 1.5 μg of poly(dI-dC)·poly(dI-dC) per lane. For PAX4 and PAX6 protein, 1 μl of the 50-μl *in vitro* translation reaction mix was added. Where antisera were used, 1 μl was added to the binding reaction 10 min prior to probe addition. HNF4α and HNF1α antisera were purchased from Santa Cruz Biotechnology; PDX1 antisera were raised against bacterially expressed PDX1. Blocking peptides were added at a concentration of 1 μg per 10-μl binding reaction 10 min prior to the addition of probe. Unlabeled competing oligonucleotides were added at a 50-fold excess.

The oligonucleotide probes used in EMSA, and their locations in relation to the transcriptional start site for the gene are shown (top strand) as follows: rat insulin I C2, bp −328 to −304, 5′-ctgggaaatgaggtggaaaatgctc-3′; rat insulin I E2, bp −242 to −234, 5′-gatcccaggccatctggcccg-3′; human PAX4 H4, bp −1960 to −1939, 5′-gtggcaagacctttgagttaa-3′; human PAX4 TAAT1/2, bp −1950 to −1926, 5′-tttgagttatgtataatgtgag-3′; human PAX4 E, bp −1928 to −1907, 5′-ttgtgagcagatggcgggggct-3′; human PAX4 TAAT 3/4, bp −1886 to −1860, 5′-ccagctaatgagggacaattagcccca-3′; human PAX4 P4.1; bp −4629 to −4583, 5′-gataagcagggtggaaaccggaatgatccatgtaataatggatta-3′; human PAX4 P4.2; bp −4164 to −4116, 5′-cccaattgtcaaaggtggaataatttgatcaacaaaataatgtattg-3′.

All of EMSA results shown are typical of those seen on at least three separate occasions.

RT-PCR—RT-PCR was performed as described previously (16) using cDNA from mPAC. αTC1.6, βTC3, and NIH3T3 cell lines, adult rat liver and purified islets, and dated mouse embryonic pancreas. The primers used to amplify all cDNAs were separated by an intron to distinguish contaminating genomic DNA from cDNA. Primer sequences for HNF4α amplification were primer 1 (5′-tgccctctcacctcagcaatg-3′) and primer 2 (5′-cccctcagcacacggttttg-3′). Products were analyzed by agarose gel electrophoresis and ethidium bromide staining. The identity of PCR products was confirmed by cloning into the TA vector (Invitrogen) and sequencing.

Transgenic Mice—Both the wild-type full-length PAX4 promoter and the promoter containing a deletion between −1963 to −1845 were cloned into *Kpn*I/*Xho*I sites upstream of the β-galactosidase gene in the plasmid pβgal-Enhancer (CLONTECH). The plasmid was linearized and microinjected (1.5 ng/μl) into murine pronuclei. Injected embryos were transferred to pseudopregnant females, and the fetal pancreata were harvested at E14.5. Tissues genotyping positive by PCR for the transgene were fixed for 30 min at 4 °C in 4% paraformaldehyde and then incubated overnight in phosphate-buffered saline supplemented with 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) (400 μg/ml) at 37 °C. Tissues were then fixed again in 4% paraformaldehyde for 30 min, paraffin-embedded, sectioned at 5 μm, and used for immunohistochemistry as described previously (2). Commercially available primary antibodies were used for staining of glucagon (dilution 1:10000) and insulin (1:8000) (Linco Research Immunoassay) according to the manufacturer's instructions. The rabbit antisera raised against mouse neurogenin3 has been described previously (25). Secondary antibodies were linked to alkaline phosphatase (insulin and glucagon) or peroxidase (neurogenin3) used at a dilution of 1:5000. Vector Red (Vector Laboratories) was used as a substrate for alkaline phosphatase staining.

RESULTS

Mapping the PAX4 Promoter—By using RNA purified from adult human islets and from embryonic day 15.5 (E15.5) fetal mouse pancreas, the pancreatic transcription start sites of the human and murine *PAX4* genes were determined by 5′-RACE. Although not identical, the 5′-RACE products identify tightly clustered start sites and the same first intron in both genes (Fig. 1*A*). Interestingly the pancreatic transcripts contain an additional two exons compared with the transcript previously characterized from human placenta (26).

To establish whether the 5′-flanking sequences can act as a cell-specific promoter, approximately 5 kb of flanking sequence was amplified by PCR from human cosmid clone g1572c264. The full-length promoter (−4958 bp) and four 5′-truncated promoters were placed in front of a luciferase reporter gene and transfected into the pancreatic beta cell line βTC3, the pancreatic alpha cell line αTC1.6, and non-islet NIH3T3 cells (Fig. 1*B*).

Removal of the region between −4958 and −2153 bp increases reporter activity in all cell types, and deletion of sequences between −2153 and −1614 bp markedly decreases promoter activity in the pancreatic cell lines. An additional (although smaller) decrease in pancreatic promoter activity results from the deletion of sequence from −1614 to −1078 bp. Addition of 5′ genomic sequences between +46 bp and the translation start site at +2543 bp gives no increase in promoter activity (data not shown).

Identification of a Pancreas-specific Promoter Element—Inspection of the promoter sequence between −2153 and −1614 bp reveals several potential binding sites for pancreatic transcription factors, all clustered between −1960 and −1861 bp (Fig. 2*A*). These include a potential HNF4α-binding site (27), four TAAT sequences that are commonly found in homeodomain protein-binding sites, and an E box, a recognition sequence for the binding of basic helix-loop-helix protein (bHLH).

To test the importance of these potential binding sites, a series of small deletions were made between −1960 and −1866 bp in the PAX4 promoter (Fig. 2*B*). In the βTC3 cells, removal

A

```
-201    tctgtccctg cttccgggtg gtaggtggac acaaagtggg ctctgaggat
-151    tccccatgga aggacacagt gccctgcctc agggctggtg tcagtcagcc
-101    catgcacacc tccctcctcc tctctcctct gctcatggcc tgtgacatca
-51     ttggctactc ccagaaggct gccctctgct cttgagcacg ggctctctat
+1      AGCTCTGGAC CCCCTGGCAG GACTGAAGCA GGAGCAGAGT GGAGGCTGTT
+51     ACCAAGACCA GATCATCACC ACCCCTGGAG CCTGCAGGGG TCTCAGAAGC
+101    TGAGACCTCC CACTGAAGCT CCCACCTCAT TTCCTCCACA CAGgtagctc
                              2kbp
+2002   agaattttct gattttagcc aacccactgg atttcaagca aaagacctca
+2052   cccaaataac cccccattat tgcagagCCA GCCCCCAAGC TGGAAAAGGA
+2102   GCTCCTAGAA GGAACTTCCC TTCCCAGCCT TCCCAGCTGG AGTCTCATCC
+2152   TTCTGAGGAG GGTCAGCCTG GAGCATGCAT CAGGACGgtg aggagcctgg
+2202   gggaagtggg aagacaggga ggacaacaga tcatctgcta ttggccaatg
+2252   ttctggggaa gctttttcca gtccctccct tcctgccctg tcctgcttcc
+2302   cttagatcag gagagttggc gcgtatgggc aaggaacaaa gactcacccg
+2352   tgagccagct ctcaaagaaa gcagcttgcg ttgacagcct gggggcagca
+2402   aggatgcagt ctcccaggag aggatgcact cggtggtggg aagccaggct
+2452   ggaggggcct gagtgaccct ctccacaggc gggcagggca gtgggagagg

+2502   tggtgtgtgg atacctctgt ctcacgccca gGGATCAGCA GCATGAACCA
+2552   GCTTGGGGGG CTCTTTGTGA ATGGCCGGCC CCTGCCTCTG GATACCCGGC
```

B

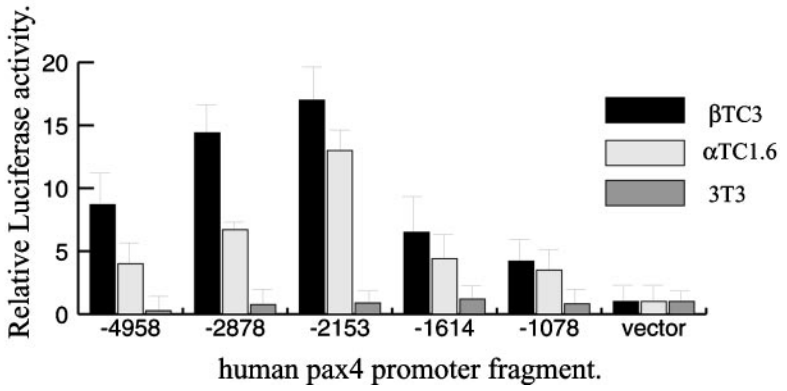

FIG. 1. **Mapping the PAX4 promoter.** *A,* the human *PAX4* gene sequence is shown with the transcriptional start sites for murine fetal pancreas (●), human adult pancreatic islets (■), and human placenta (−). The translational start site is indicated in *bold*, and the intron sequence is shown in *lowercase*, and the exon sequence is shown in *uppercase*. *B,* the activity of various lengths of the human PAX4 promoter designated by their 5′ ends and driving a luciferase reporter gene were transfected into the indicated cell types. Resulting luciferase activity is expressed relative to the activity of the promoterless expression vector. The mean ± S.E. is shown for transfections performed in duplicate on at least three separate occasions.

of each of these sites in series, the potential HNF4α site (deletion −1947), TAAT1 and -2 (deletion −1930), the E box (deletion −1909), and TAAT3 and -4 (deletion 1866), produces a stepwise reduction in promoter activity. By contrast, in αTC1.6 cells, only the last of these deletions (TAAT3 and -4, −1866) significantly reduces promoter activity.

We also deleted the sequences between −1963 and −1845 bp in the intact PAX4 promoter to test the importance of the pancreas-specific element in the context of the full-length promoter. In the βTC3 cells, this deletion completely disables the promoter (Fig. 3*A*). In addition, a single copy of the 5′ portion of the pancreas-specific element (−1960 to −1909 bp) can function as a pancreas-specific enhancer when linked to a heterologous promoter, the TK promoter from the herpes simplex virus (Fig. 3*B*). Although the enhancement given by the pancreas-specific element is small (2-fold), this represents a significant increase above the high basal activity of the TK promoter. With two copies of the (−1960 to −1909) region ligated 5′ of the TK promoter, reporter gene activity increases to 12-fold above the basal level in βTC3 but not 3T3 cells (Fig. 3*B*).

Activity of the Pancreas-specific Element in Vivo—Whereas the pancreas-specific element may be essential for the function of the PAX4 promoter in βTC3 and αTC1.6 cells, these tumor cells may not be representative of the cells in the developing pancreas where PAX4 is predominantly expressed. Therefore, we produced mice carrying a transgene with the −4958-bp

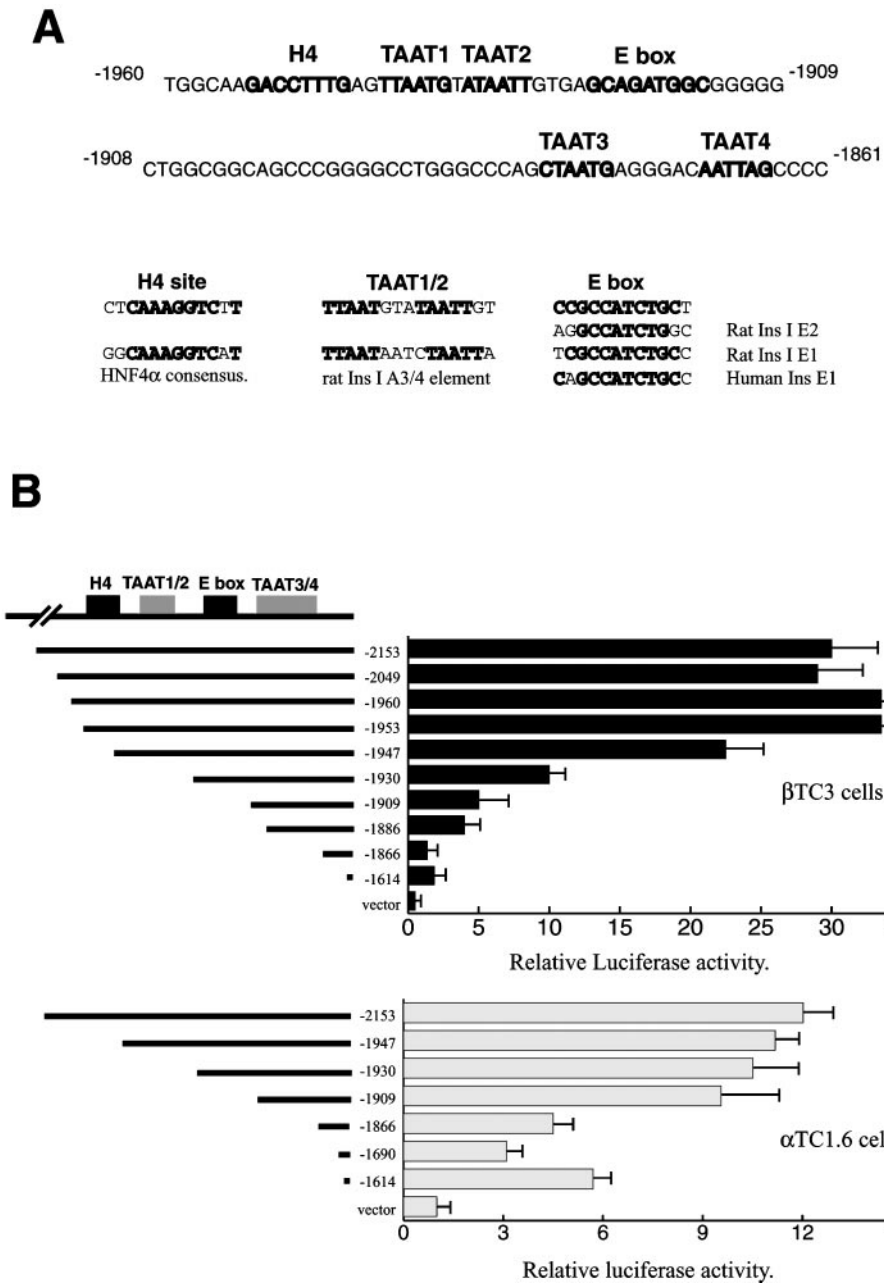

FIG. 2. **Identification of a pancreas-specific promoter element.** *A,* the sequence between −1960 and 1861 bp of the human PAX4 promoter is shown with the location of potential transcription factor binding sites. *B,* 5′ deletions were made between −2153 and −1614 bp in the human PAX4 promoter and tested in the indicated cell lines. Resulting luciferase activity is expressed relative to the activity of the promoterless expression vector. The mean ± S.E. is shown for transfections performed in duplicate on at least three separate occasions.

human PAX4 promoter driving the bacterial gene encoding β-galactosidase. Initial analysis of mice harvested at E12.5 reveals that β-galactosidase expression is restricted to the pancreas (Fig. 4).

Additional founder mice were harvested at embryonic day 14.5, 24 h after the peak of PAX4 mRNA expression in the fetal pancreas (16) for further analysis (Fig. 5). Out of 12 independent transgenic founders, strong β-galactosidase activity was detected in the pancreas of 8. In contrast, in a parallel experiment, no β-galactosidase activity was detectable in 9 founders using the −4958-bp promoter with the sequences between −1963 and −1845 bp (the pancreas-specific element) deleted.

In the transgenic mice carrying the wild-type promoter construct, we used immunohistochemistry to identify the cells expressing β-galactosidase. The β-galactosidase expressing cells are located predominantly in the central area of the pancreas where the main ducts lie and endocrine cell genesis normally occurs. Most of the insulin-expressing cells at this stage also express β-galactosidase, but there are also many non-insulin-expressing cells that express β-galactosidase. Occasional cells co-express glucagon and β-galactosidase; but these cells are uncommon; and most alpha cells do not express β-galactosidase. Interestingly, not all β-galactosidase expressing cells express pdx1, and a subset express neurogenin3, a basic helix-loop-helix transcription factor and a marker for islet cell precursors (11, 25). Together these results suggest that the PAX4 promoter is active in beta cell precursors and possibly occasionally in alpha cell precursors as well.

Multiple Factors Bind to the Pancreas-specific Element—To identify nuclear factors that bind to the pancreas-specific element in the PAX4 promoter, we synthesized a series of oligonucleotides spanning the transcriptionally active region and tested them for binding to nuclear proteins by electromobility shift assay (EMSA).

An oligonucleotide (H4) encompassing the putative HNF4α-binding site binds to a low mobility complex present in liver and βTC3 cell nuclear extracts. The mobility of this complex from both liver and βTC3 cells is retarded (supershifted) by the addition of antisera directed to the carboxyl terminus of mouse HNF4α, an effect that is prevented by the addition of excess

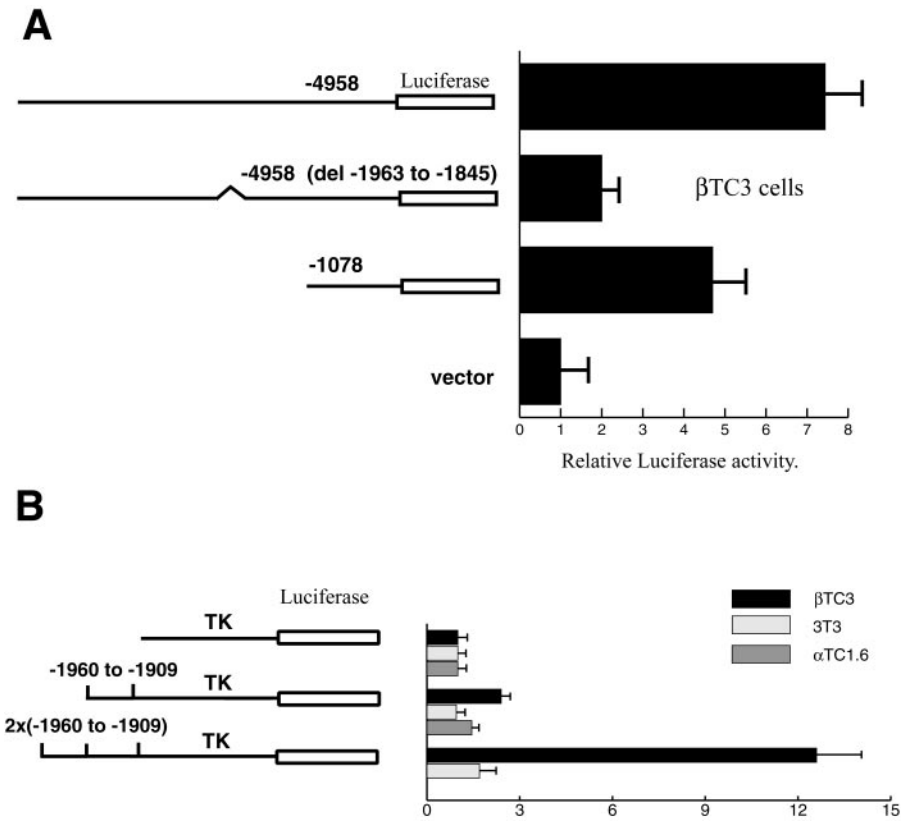

FIG. 3. **Elements between −1963 and −1845 are necessary and sufficient to activate transcription.** *A,* the full-length PAX4 promoter, a similar construct containing a 118-bp deletion (Δ-1963 to −1845 bp) and a truncated promoter fragment were transfected into βTC3 cells. Resulting luciferase activity is expressed relative to the activity of the promoterless expression vector. *B,* PAX4 promoter sequences from −1960 to −1909 were ligated upstream of the HSV-TK gene promoter driving the luciferase reporter gene. The resulting construct and the parent vector lacking the PAX4 promoter sequences were transfected into the indicated cell line. Resulting luciferase activity is expressed relative to the activity of the promoterless expression vector. The mean ± S.E. is shown for transfections performed in duplicate on at least three separate occasions.

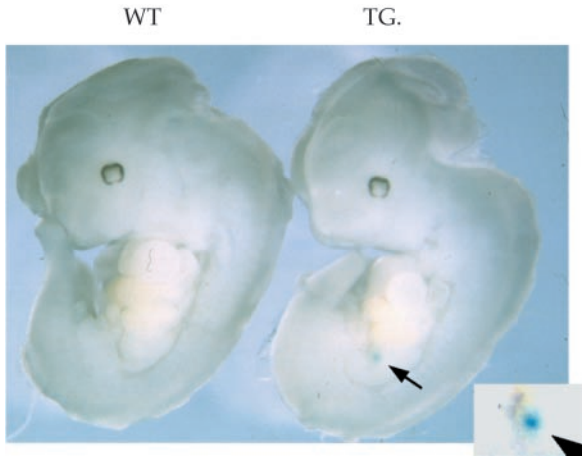

FIG. 4. **Activity of the PAX4 promoter in transgenic mice.** The image shows a mouse embryo carrying a transgene (*TG*) with the −4958 human PAX4 promoter ligated upstream of the reporter gene β-galactosidase. The *blue* coloring shown at E12.5 indicated by the *arrow* and shown in the *inset* demonstrates the restriction of transgene expression to the pancreatic bud. A wild-type (*WT*) control is shown for comparison.

HNF4α carboxyl-terminal peptide (Fig. 6*A*). A similar complex and supershifted complex is not seen in nuclear extracts from αTC1.6 cells (data not shown). In addition, *in vitro* produced HNF4α protein readily binds to the H4 oligonucleotide (data not shown).

These results demonstrate that HNF4α is present in the βTC3 cell line and binds to the H4 site. HNF4α expression has previously been demonstrated in the pancreas but with unspecified cellular distribution (28, 29). To assess the expression of HNF4α in the fetal pancreas when the *PAX4* gene is actively transcribed, we gauged HNF4α mRNA expression by RT-PCR and found that HNF4α mRNA peaks in the pancreas at E13.5 (Fig. 6*B*), in parallel with the peak of PAX4 mRNA (16). Unlike PAX4, however, HNF4α mRNA persists in the adult islet at detectable levels. Two HNF4α-derived PCR products of very similar size were detected in the products depicted in Fig. 6*B*; these were sequenced and represent two previously reported alternate splice forms of HNF4α mRNA (30). Interestingly, no HNF4α mRNA was detected in αTC1.6 cells consistent with the observation that removal of the H4 site had no effect on PAX4 promoter activity in these cells.

An oligonucleotide (oligonucleotide TAAT 1/2) containing the 5′ two TAAT sequences binds three major complexes (Fig. 7*A*). The lowest mobility complex present in pancreatic cell types is shifted by HNF1α antiserum (Fig. 7*A*). The highest mobility complex in βTC3 extracts is recognized by pdx1 antiserum (Fig. 7*B*). The αTC1.6 nuclear extract produces a complex with the same mobility but less intensity than the PDX1 band in βTC3 extracts, but this complex is not recognized by the pdx1 antiserum (data not shown).

The E box in the pancreas-specific element from the PAX4 promoter is identical to the conserved 8-base pair sequence of the E boxes present in the human and rat insulin promoters. As shown in Fig. 7*C*, identical slow migrating complexes present in both αTC1.6 and βTC3 extracts bind to both the human insulin promoter E1 element and the PAX4 promoter E element (oligonucleotide E). This complex has previously been shown to consist of a heterodimer containing ubiquitously expressed bHLH proteins such as E2A gene products (31–34) and neuroD1/BETA2 (35), a protein of limited tissue distribution.

Finally, an oligonucleotide (oligonucleotide TAAT3/4) containing the two 3′ TAAT sequences binds three complexes recognized by pdx1 antiserum in βTC3 extracts (Fig. 7*D*). Whereas one of the two more slowly migrating complexes most likely results from PDX1 binding to both TAAT sites, the other band probably is produced by interactions with an additional protein, such as PBX1 (36).

Activation of the −1960 to −1909 Promoter Region in a Non-pancreatic Cell Line—To test the ability of the identified transcription factors to activate the pancreas-specific element

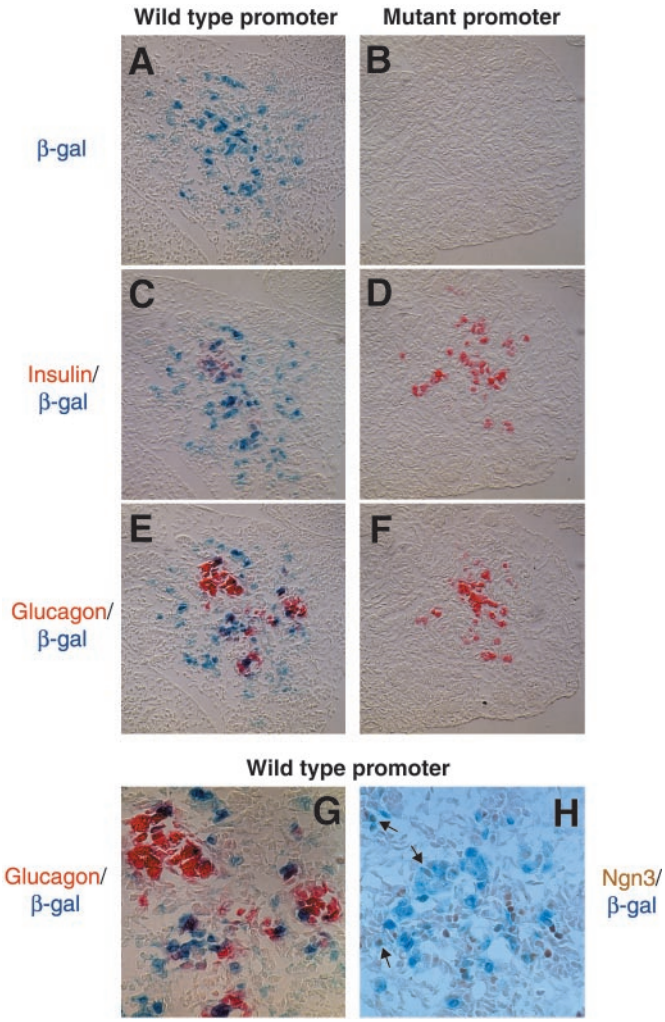

FIG. 5. **Transgenic expression within the pancreas.** Sections are shown from the fetal pancreas at E14.5 from transgenic mice expressing the bacterial gene β-galactosidase (*β-gal*) under the control of the wild-type (−4958 bp) human PAX4 promoter (*H* and serial sections *A, C, E,* and *G*) or a mutated promoter lacking the pancreas-specific element (sequences from −1963 to −1845 bp) (serial sections *B, D,* and *F*). β-Galactosidase expression is indicated by *blue* staining. *C* and *D*, immunohistochemical staining for insulin is in *red*. *E–G,* immunohistochemical staining for glucagon is in *red*. *H*, staining for neurogenin3 is in *brown*. Magnification, × 200 (*A–F*) and × 400 (*G* and *H*).

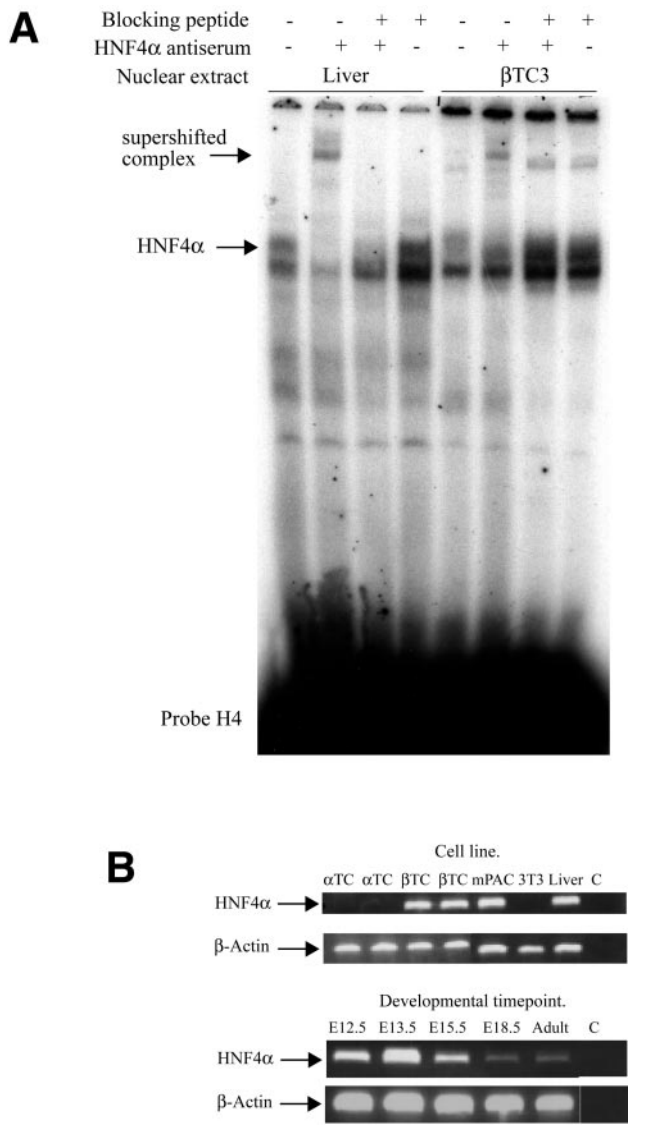

FIG. 6. **HNF4α binds to the pancreas-specific element.** *A*, EMSA using the H4 DNA probe is shown. Probe was incubated with 2 μg of nuclear extract from rat liver or βTC3 cells. 1 μl of antiserum and/or 5 μl of blocking peptide were preincubated with the samples as indicated prior to probe addition and electrophoresis. *B*, RT-PCR was performed using 250 ng of RNA derived from rat liver, cell lines, or mouse embryonic pancreas. HNF4α or β-actin cDNA were amplified for 30 cycles. PCR products were analyzed by agarose gel electrophoresis and ethidium bromide staining.

in a non-pancreatic cell line, we expressed each of the factors in 3T3 fibroblast cells along with a luciferase reporter construct driven by two copies of pancreas-specific element upstream of the minimal TK promoter. Although no single cDNA activates the pancreas-specific element (Fig. 8*A*), a combination of the transcription factors can activate the construct up to 9-fold (Fig. 8*B*). We also tested the ability of neurogenin3 to replace neuroD1 in this combination of factors because neurogenin3 is expressed abundantly in some progenitor cells that express PAX4 *in vivo* (Fig. 5). Neurogenin3 and neuroD1 are closely related, and when dimerized with Pan1/E47 both bind to the same E box consensus sequence(37).[2] Although the complex in βTC3 nuclear extract that binds to the E box in the pancreas-specific element does not include neurogenin3 (data not shown), neurogenin3 is expressed at much lower levels in βTC3 cells than in the embryonic pancreas (Fig. 5 and data not shown). Expressing neurogenin3 in place of neuroD1 in the transcription factor combination results in a further increase in the activity of the pancreas-specific element in 3T3 cells (Fig. 8*B*).

Autoregulation of the PAX4 Promoter—Inspection of the PAX4 promoter sequence reveals a number of sequence elements that match or closely resemble the consensus DNA binding sequence for the PAX4 protein (38, 39), including both paired domain and homeodomain binding motifs. To test whether the PAX4 promoter is autoregulated, the PAX4 promoter reporter plasmid was co-transfected with a plasmid vector expressing the PAX4 protein (Fig. 9*A*). PAX4 protein represses the activity of the full-length promoter, and this effect is most remarkable in αTC cells, consistent with previous evidence that PAX4 repression is most fully realized in α-cell lines (16). Removal of 5′ sequences from the promoter reduces the repression by PAX4. This repression depends on an intact homeodomain and to a lesser extent on the paired domain (Fig. 9*B*).

[2] H. Watada and M. German, unpublished data.

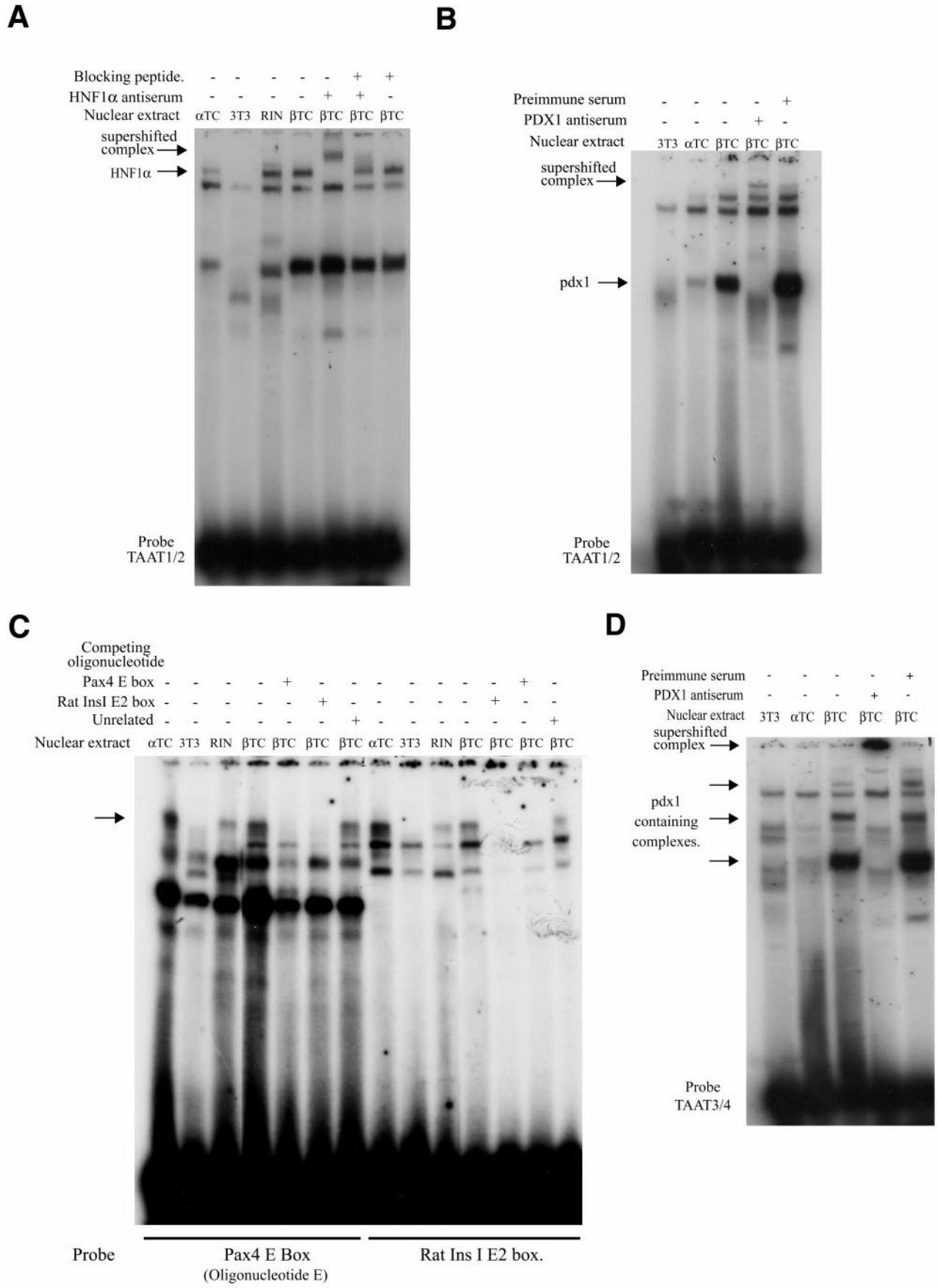

FIG. 7. **Multiple factors bind to the pancreas-specific element.** *A* and *B*, EMSA using the TAAT1/2 probe is shown. Probe was incubated with 2 μg of nuclear extract from the indicated cell lines. Samples were preincubated with 1 μl of HNF1α antiserum and/or 5 μl of blocking peptide (*A*) or with 1 μl of PDX1 antiserum or preimmune serum (*B*) as indicated prior to probe addition and electrophoresis. *C*, EMSA using the E probe and the E2 element of the rat insulin I promoter is shown. Probes were incubated with 2 μg of nuclear extract from the indicated cell line.

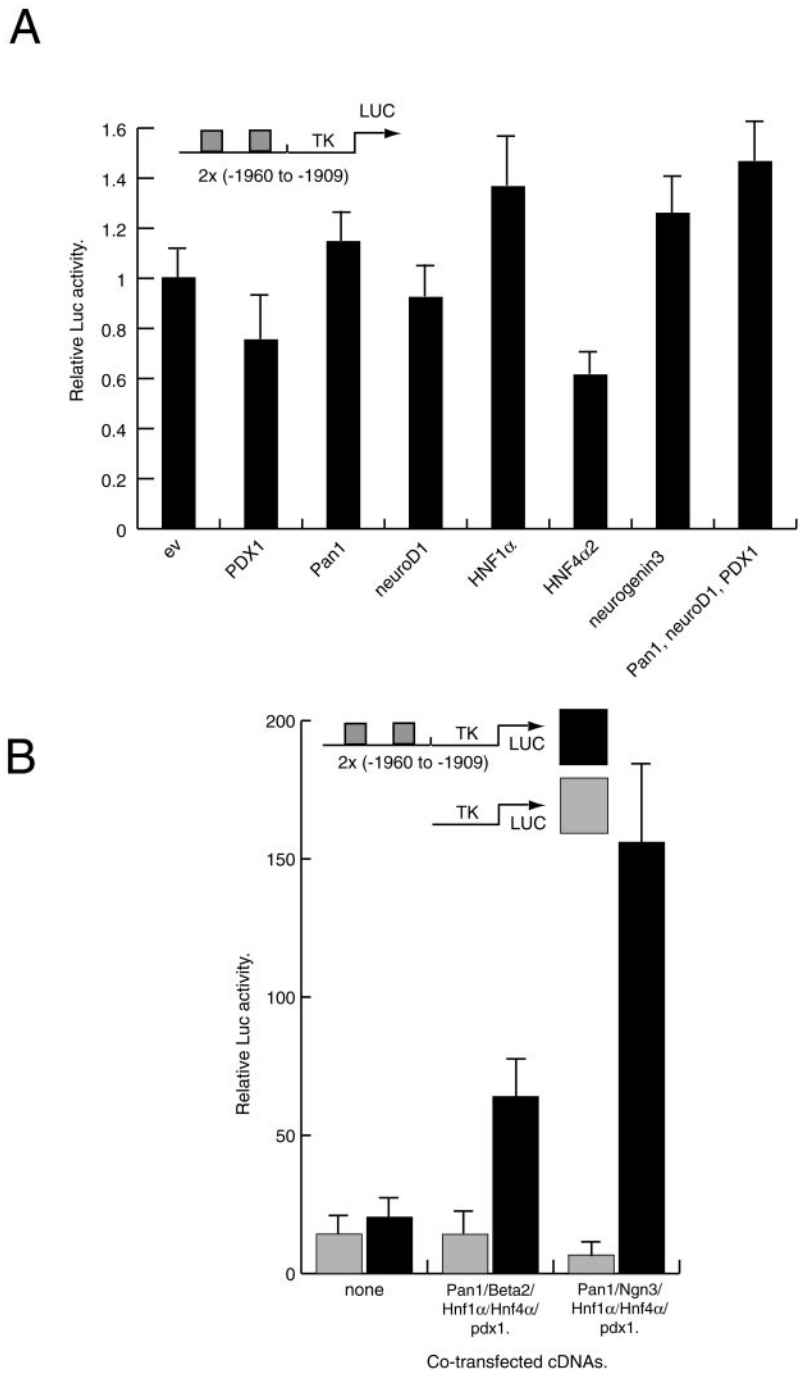

FIG. 8. **A combination of transcription factors activates the PAX4 promoter region −1960 to −1909 in the NIH3T3 fibroblast cell line.** *A*, NIH3T3 cells were transfected with a reporter construct containing 2 copies of the (−1960 to −1909)) PAX4 promoter region located 5′ to the HSV-TK promoter, upstream of the firefly luciferase (*LUC*) gene (2 μg per million cells). Additionally the cDNA for the indicated transcription factor (*ev* is an expression vector with no cDNA) was co-transfected under the control of the CMV promoter (0.1 μg per million cells). Relative luciferase activity is shown as a multiple of that from the promoterless reporter construct. *B*, transfections were performed as in *A*, except a combination of transcription factors were co-transfected together (HNF4α2, HNF1α, PDX1, and PAN1) with either NeuroD1 or neurogenin3 as indicated.

Two oligonucleotides (oligonucleotides P4.1 and P4.2) corresponding to PAX4 sites located at bases −4611 and −4148 of the PAX4 promoter bind PAX4 (Fig. 9*C*), with an affinity comparable to the C2 element of the insulin promoter. In contrast to the C2 element (16, 39, 40), the PAX4-binding sites of the PAX4 promoter show a significantly higher affinity for PAX4 than for PAX6.

DISCUSSION

In the present study, we have shown that the human and mouse *PAX4* genes have adjacent transcription start sites and use the same first two exons in the developing and mature pancreas. It should be noted, however, that a previous study using human placental mRNA identified a more 3′ transcription start site without the first two exons and introns described here (26). These differences most likely result from the use of alternate promoters in distinct tissues. A previous study of PAX4 mRNA in the rat pancreas identified alternate splice forms, but none of these differences involved the 5′ end (41).

The experiments using the 5′-flanking sequence from the *PAX4* gene confirm the presence of an active, pancreas-specific promoter driving expression from the mapped start sites. Although this promoter lacks a consensus TATAA element, there

Competition was performed with a 50-fold excess of the indicated unlabeled oligonucleotide. *D*, EMSA using the TAAT3/4 probe is shown. Probe was incubated with 2 μg of nuclear extract from the indicated cell line. Samples were preincubated with 1 μl of HNF1α antiserum and/or 5 μl of blocking peptide as indicated prior to probe addition and electrophoresis.

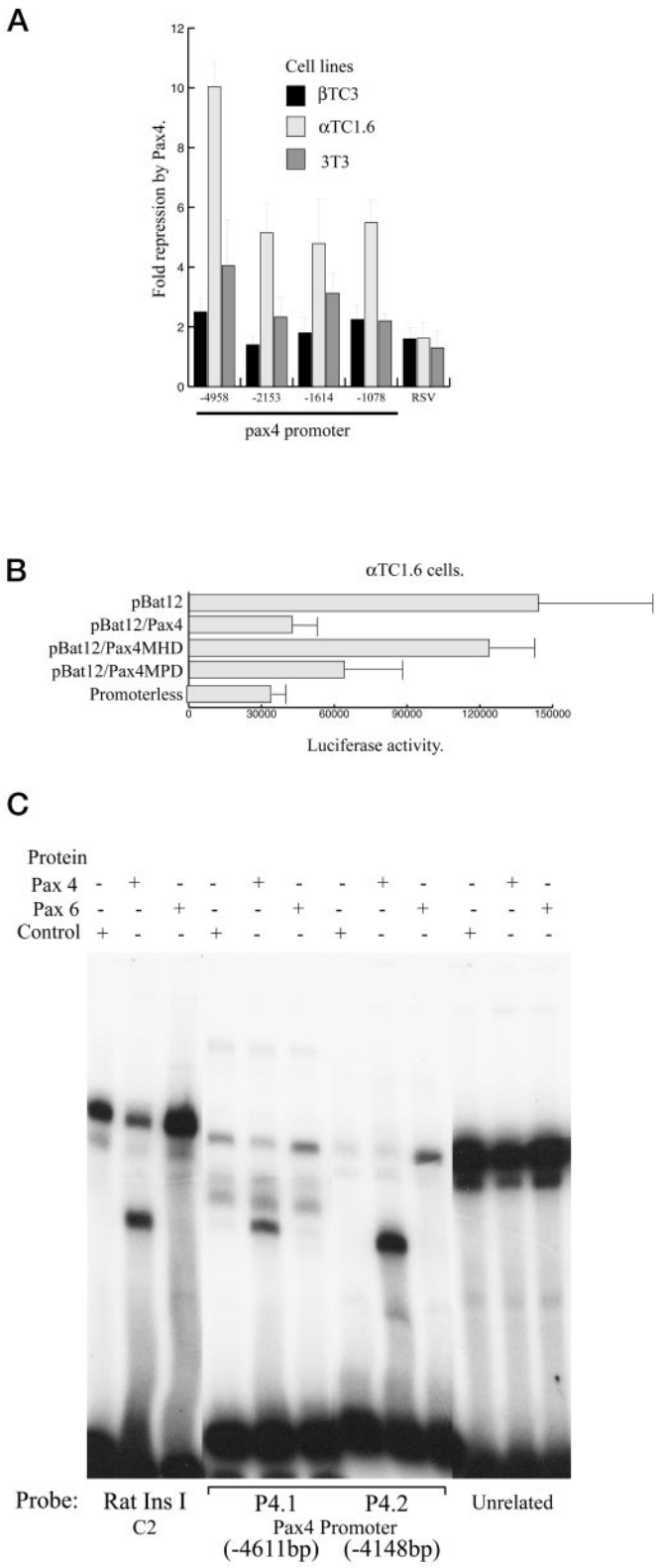

FIG. 9. **The PAX4 promoter is subject to autoregulation by a mechanism requiring DNA binding.** *A*, luciferase reporter vectors driven by the indicated PAX4 or Rous sarcoma virus promoters were transfected into the cell line shown. Cells were co-transfected with either the PAX4 expression vector pBAT12.mPAX4 which produces full-length PAX4 protein under the control of the CMV promoter or the parent vector pBAT12 without cDNA insert. Relative luciferase activities were calculated with the activity of cells transfected with the parent pBAT12 vector without cDNA insert set at 1. In order to highlight differences in repression, data are shown as fold repression, the inverse of relative luciferase activity. *B,* the long PAX4 promoter (−4958 bp) upstream of the luciferase reporter gene in pFOXLuc1 was transfected into αTC1.6 cells and co-transfected with pBAT12, pBAT12 expressing wild-type PAX4 cDNA, or pBAT12 expressing PAX4 cDNA carrying a mutation in the homeodomain (N220P and R222P) or paired domain (S43P and R44P). Resulting luciferase activity is expressed relative to the activity of the promoterless expression vector. The mean ± S.E. is shown for transfections performed in duplicate on at least three separate occasions. *C*, EMSA using the rat insulin I promoter C2 site and the human PAX4 P4.1 and P4.2 probes is shown. Probes were incubated with 1 μl of *in vitro* produced PAX4, PAX6, or luciferase (control).

are two CCAAT sequences adjacent to the transcription start sites, in keeping with observations of other TATAA-less promoters (42). A more distal, pancreas-specific element, however, contributes an essential component to the overall activity of the promoter.

Interestingly, the pancreas-specific element has remarkable similarities to the E2-A3/4 element (previously called the Far-FLAT or FF minienhancer) present in the rat insulin I promoter (18, 43). Like the E2-A3/4 element in the insulin promoter, the pancreas-specific element is required for full activity of the PAX4 promoter and is also sufficient to drive pancreas-specific transcription when linked to a heterologous promoter. But it is the similarity in DNA-binding sites that is most remarkable. The two elements contain identical sequences that can function as binding sites for bHLH proteins. In βTC3 cells, these sites bind a heterodimer of two bHLH proteins, a ubiquitous class A bHLH protein such as E47 (31–34, 44) and the neuroendocrine-specific class B bHLH protein neuroD1/BETA2 (35). In the developing pancreas, neurogenin3 may replace neuroD1/BETA2 in the heterodimer binding to this sequence.[3] Adjacent to the bHLH-binding sites in both elements lie nearly identical AT-rich sequences that function as binding sites for both the pou-homeodomain protein HNF1α (45) and the pancreatic/duodenal homeodomain protein PDX1 (46–48).

But if the same factors control *insulin* and *PAX4* gene expression, why do the two genes show marked differences in their temporal expression pattern? In the developing pancreas, PAX4 mRNA levels peak at E13.5, when insulin mRNA levels are just starting to rise, and PAX4 expression is shut off by the time insulin mRNA reaches its peak (16). This pattern is confirmed in individual cells in the transgenic mice with the PAX4 promoter driving β-gal, since β-gal is expressed in many cells that do not express insulin and is co-expressed in neurogenin3 expressing cells that represent progenitor cells in which insulin is not yet expressed (25). This pattern of PAX4 activation in non-hormone-expressing progenitor cells is consistent with a role for PAX4 in beta/delta-cell type determination.

The differences in insulin and *pax4* gene expression suggest that additional, distinct factors play essential roles in the expression of the two genes. HNF4α could be one such factor, since the pancreas-specific element from the PAX4 promoter contains an HNF4α-binding site that is not present in the rat insulin I promoter, and HNF4α mRNA shows an expression pattern that is similar to PAX4 mRNA. In addition, the deletion experiments demonstrate that sequences immediately downstream from the bHLH-binding site contribute to PAX4 promoter activity in both αTC and βTC cells. Although there are at least two PDX1-binding sites in this region, it seems unlikely that PDX1 alone can explain the activity, since the αTC cells do not express PDX1. Other factors, possibly including other homeodomain factors, must explain this activity in αTC cells and likely in progenitor cells *in vivo* as well. Finally, although the pancreas-specific element is critical for the overall activity of the PAX4 promoter, and it can function on its own,

[3] V. Schwitzgebel, H. Watada, and M. German, unpublished data.

it may require cooperative interactions with other sequences binding to distinct factors in order to realize its full activity.

One of the most interesting aspects of the *PAX4* gene is its capacity for auto-repression. If PAX4 functions solely to determine islet cell type fate during the differentiation of the endocrine pancreas, then persistent expression of PAX4 after this decision may be detrimental, possibly by repressing insulin gene expression (16, 40). As has been shown for the *Drosophila* homeobox gene *goosecoid*, auto-repression is one mechanism to prevent persistent expression (49). Such a model allows for the rapid activation and subsequent extinction of *PAX4* gene expression observed during pancreas development. Of course other mechanisms would be required to keep the gene off, and our data demonstrate that portions of the promoter can repress transcription in the absence of PAX4 protein.

But if PAX4 limits its own expression, why does it persist in the beta-cell tumor line βTC3? Tumor lines do not reflect completely normal beta cell biology, and in addition we have not consistently found PAX4 in other beta cell tumor lines (data not shown). The unique inability of βTC3 cells to extinguish *PAX4* gene expression could result from differences in expression or function of co-repressors that normally cooperate with the PAX4 protein in affecting transcriptional repression. Evidence that PAX4 functions poorly as a transcriptional repressor in βTC3 cells supports this explanation (16). Alternatively, persistent low level *PAX4* gene expression may allow for a balance of *PAX4* gene expression and feedback repression by PAX4 protein. This pattern is quite different from the rapid accumulation of PAX4 protein and subsequent repression of *PAX4* gene expression that normally occurs in the developing pancreas.

Finally, the presence of HNF4α, HNF1α, PDX1, and neuroD1 among the proteins binding to the active region of the PAX4 promoter is intriguing given the role of these proteins in human diabetes. Mutations in the genes encoding any of these four transcription factors can cause a form of autosomal dominant diabetes in humans called maturity onset diabetes of the young (50, 51). Although studies in mice carrying targeted mutations in the genes encoding HNF4α and HNF1α suggest that secondary defects in the expression of genes involved in glucose sensing may cause the insulin secretory defects observed in patients with MODY1 (52) and MODY3 (53), defects in beta cell development could also contribute to the eventual beta cell deficiencies seen in the human syndromes. By understanding the hierarchy of gene regulation events that control the development of the pancreas and the differentiation of the insulin-producing beta cells eventually may provide new insights into the inherited defects that cause insulin deficiency and diabetes.

Acknowledgments—We thank J. Johnson (Metabolex, Hayward, CA) for supplying human RNA and S. Iadonato (Washington University) for supplying the PAX4 genomic clone. We also thank J. Leung and Y. Zhang for technical assistance and members of the German laboratory for helpful discussions.

REFERENCES

1. Slack, J. M. (1995) *Development* **121,** 1569–1580
2. Sander, M., Neubuser, A., Kalamaras, J., Ee, H. C., Martin, G. R., and German, M. S. (1997) *Genes Dev.* **11,** 1662–1673
3. Edlund, H. (1998) *Diabetes* **47,** 1817–1823
4. Wu, K. L., Gannon, M., Peshavaria, M., Offield, M. F., Henderson, E., Ray, M., Marks, A., Gamer, L. W., Wright, C. V., and Stein, R. (1997) *Mol. Cell. Biol.* **17,** 6002–6013
5. Sharma, S., Jhala, U. S., Johnson, T., Ferreri, K., Leonard, J., and Montminy, M. (1997) *Mol. Cell. Biol.* **17,** 2598–2604
6. Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994) *Nature* **371,** 606–609
7. Offield, M. F., Jetton, T. L., Labosky, P. A., Ray, M., Stein, R. W., Magnuson, M. A., Hogan, B. L., and Wright, C. V. (1996) *Development* **122,** 983–995
8. Li, H., Arber, S., Jessell, T. M., and Edlund, H. (1999) *Nat. Genet.* **23,** 67–70
9. Harrison, K. A., Thaler, J., Pfaff, S. L., Gu, H., and Kehrl, J. H. (1999) *Nat. Genet.* **23,** 71–75
10. Gradwohl, G., Dierich, A., LeMeur, M., and Guillemot, F. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 1607–1611
11. Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D. J., Honjo, T., Hrabe de Angelis, M., Lendahl, U., and Edlund, H. (1999) *Nature* **400,** 877–881
12. Naya, F. J., Huang, H. P., Qiu, Y., Mutoh, H., DeMayo, F. J., Leiter, A. B., and Tsai, M. J. (1997) *Genes Dev.* **11,** 2323–2334
13. Sussel, L., Kalamaras, J., Hartigan-O'Connor, D. J., Meneses, J. J., Pedersen, R. A., Rubenstein, J. L., and German, M. S. (1998) *Development* **125,** 2213–2221
14. Ahlgren, U., Pfaff, S. L., Jessell, T. M., Edlund, T., and Edlund, H. (1997) *Nature* **385,** 257–260
15. St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A., and Gruss, P. (1997) *Nature* **387,** 406–409
16. Smith, S. B., Ee, H. C., Conners, J. R., and German, M. S. (1999) *Mol. Cell. Biol.* **19,** 8272–8280
17. Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G., and Gruss, P. (1997) *Nature* **386,** 399–402
18. German, M. S., Moss, L. G., Wang, J., and Rutter, W. J. (1992) *Mol. Cell. Biol.* **12,** 1777–1788
19. German, M. S., Wang, J., Chadwick, R. B., and Rutter, W. J. (1992) *Genes Dev.* **6,** 2165–2176
20. German, M., Wang, J., Fernald, A., Espinosa, R., LeBeau, M., and Bell, G. (1994) *Genomics* **24,** 403–404
21. deWet, J., Wood, K. V., DeLuca, M., Helsinki, D. R., and Subramani, S. (1987) *Mol. Cell. Biol.* **7,** 725–737
22. Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., Maas, R. L., and Pabo, C. O. (1999) *Genes Dev.* **13,** 1263–1275
23. Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell* **80,** 639–650
24. Czerny, T., and Busslinger, M. (1995) *Mol. Cell. Biol.* **15,** 2858–2871
25. Schwitzgebel, V. M., Scheel, D. W., Conners, J. R., Kalamaras, J. R., Lee, J. E., Anderson, D. J., Sussel, L., Johnson, J. D., and German, M. S. (2000) *Development* **127,** 3533–3542
26. Tao, T., Wasson, J., Bernal-Mizrachi, E., Behn, P. S., Chayen, S., Duprat, L., Meyer, J., Glaser, B., and Permutt, M. A. (1998) *Diabetes* **47,** 1650–1653
27. Sladek, F. M., Zhong, W. M., Lai, E., and Darnell, J. E., Jr. (1990) *Genes Dev.* **4,** 2353–2365
28. Miquerol, L., Lopez, S., Cartier, N., Tulliez, M., Raymondjean, M., and Kahn, A. (1994) *J. Biol. Chem.* **269,** 8944–8951
29. Duncan, S. A., Manova, K., Chen, W. S., Hoodless, P., Weinstein, D. C., Bachvarova, R. F., and Darnell, J. E., Jr. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 7598–7602
30. Sladek, F. M., Ruse, M. D., Jr., Nepomuceno, L., Huang, S. M., and Stallcup, M. R. (1999) *Mol. Cell. Biol.* **19,** 6509–6522
31. Aronheim, A., Ohlsson, H., Park, C. W., Edlund, T., and Walker, M. D. (1991) *Nucleic Acids Res.* **19,** 3893–3899
32. German, M. S., Blanar, M. A., Nelson, C., Moss, L. G., and Rutter, W. J. (1991) *Mol. Endocrinol.* **5,** 292–299
33. Cordle, S. R., Henderson, E., Masuoka, H., Weil, P. A., and Stein, R. (1991) *Mol. Cel. Biol.* **11,** 1734–1738
34. Sheih, S. Y., and Tsai, M.-J. (1991) *J. Biol. Chem.* **266,** 16708–16714
35. Naya, F. J., Stellrecht, C. M., and Tsai, M. J. (1995) *Genes Dev.* **9,** 1009–1019
36. Peers, B., Sharma, S., Johnson, T., Kamps, M., and Montminy, M. (1995) *Mol. Cell. Biol.* **15,** 7091–7097
37. Huang, H. P., Liu, M., El-Hodiri, H. M., Chu, K., Jamrich, M., and Tsai, M. J. (2000) *Mol. Cell. Biol.* **20,** 3292–3307
38. Smith, K. M., Olson, D. C., Hirose, R., and Hanahan, D. (1997) *Int. Immunol.* **9,** 1355–1365
39. Fujitani, Y., Kajimoto, Y., Yasuda, T., Matsuoka, T. A., Kaneto, H., Umayahara, Y., Fujita, N., Watada, H., Miyazaki, J. I., Yamasaki, Y., and Hori, M. (1999) *Mol. Cell. Biol.* **19,** 8281–8291
40. Campbell, S. C., Cragg, H., Elrick, L. J., Macfarlane, W. M., Shennan, K. I., and Docherty, K. (1999) *FEBS Lett.* **463,** 53–57
41. Tokuyama, Y., Yagui, K., Sakurai, K., Hashimoto, N., Saito, Y., and Kanatsuka, A. (1998) *Biochem. Cell Biol.* **248,** 153–156
42. Mantovani, R. (1998) *Nucleic Acids Res.* **26,** 1135–1143
43. Karlsson, O., Walker, M. D., Rutter, W. J., and Edlund, T. (1989) *Mol. Cell. Biol.* **9,** 823–827
44. Peyton, M., Moss, L. G., and Tsai, M.-J. (1994) *J. Biol. Chem.* **269,** 25936–25941
45. Emens, L. A., Landers, D. W., and Moss, L. G. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7300–7304
46. Miller, C. P., McGehee, R. E., Jr., and Habener, J. F. (1994) *EMBO J.* **13,** 1145–1156
47. Peers, B., Leonard, J., Sharma, S., Teitelman, G., and Montminy, M. R. (1994) *Mol. Endocrinol.* **8,** 1798–1806
48. Petersen, H. V., Serup, P., Leonard, J., Michelsen, B. K., and Madsen, O. D. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10465–10469
49. Danilov, V., Blum, M., Schweickert, A., Campione, M., and Steinbeisser, H. (1998) *J. Biol. Chem.* **273,** 627–635
50. Yamagata, K., Oda, N., Kaisaki, P., Menzel, S., Furuta, H., Vaxillaire, S., Southam, L., Cox, R., Lathrop, G., Boriraj, V., Chen, X., Cox, N., Oda, Y., Yano, H., Le Beau, M., Yamada, S., Nishigori, H., Takada, J., Fajans, S., Hattersley, A., Iwasaki, N., Hansen, T., Pedersen, O., Polonsky, K., Turner, R., Velho, G., Chevre, J., Froguel, P., and Bell, G. (1996) *Nature* **384,** 455–458
51. Yamagata, K., Furuta, H., Oda, N., Kaisaki, P. J., Menzel, S., Cox, N. J., Fajans, S. S., Signorini, S., Stoffel, M., and Bell, G. I. (1996) *Nature* **384,** 458–460
52. Dukes, I. D., Sreenan, S., Roe, M. W., Levisetti, M., Zhou, Y.-P., Ostrega, D., Bell, G. I., Pontoglio, M., Yaniv, M., Philipson, L., and Polonsky, K. S. (1998) *J. Biol. Chem.* **273,** 24457–24464
53. Stoffel, M., and Duncan, S. A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 13209–13214

GENES: STRUCTURE AND REGULATION:

Autoregulation and Maturity Onset Diabetes of the Young Transcription Factors Control the Human PAX4 Promoter

Stuart B. Smith, Hirotaka Watada, David W. Scheel, Caroline Mrejen and Michael S. German

J. Biol. Chem. 2000, 275:36910-36919.
doi: 10.1074/jbc.M005202200 originally published online August 30, 2000