*Brief Genetics Report*

# Isolation and Characterization of the Human *PAX4* Gene

**Tsuyoshi Tao, Jon Wasson, Ernesto Bernal-Mizrachi, Philip S. Behn, Susan Chayen, Laura Duprat, Joanne Meyer, Benjamin Glaser, and M. Alan Permutt**

*PAX* genes are members of a family of developmental control genes that encode nuclear factors and play critical roles during fetal development (1). Results of recent gene targeting experiments revealed that *Pax4* and *Pax6* are involved in pancreatic islet development: *Pax4* mutant mice lacked β- and δ-cells (2), whereas *Pax6* mutant mice lacked α-cells (3). Because impaired insulin production by pancreatic islet β-cells is a major hallmark of type 2 diabetes (4), genes that regulate pancreatic islet development and insulin biosynthesis might contribute to the pathogenesis of this disorder. As the first step to test the hypothesis that *PAX4* might be one of these genes, we now report the isolation and characterization of the human *PAX4* gene.

The Basic Local Alignment Search Tool (BLAST) was used to search public databases with the partial sequence of the mouse *Pax4* cDNA (GenBank accession no. Y09584). The search revealed highly homologous sequences in human cosmid clone g1572c264 (GenBank accession no. AC000359). We found that this human homolog appeared to be present in five separate fragments within a 2.5-kb region of this cosmid clone. Based on the sequence of this human homolog, we designed a pair of oligonucleotide primers (*GSP1*, 5′-TGGAATGCGGCCCTGTGACAT-3′; *GSP2*, 5′-AGCTGCATTTCCCACTTGAGCT-3′). The partial cDNA of human *PAX4* was amplified with these primers by polymerase chain reaction (PCR) with human placenta Marathon-Ready cDNA (Clontech, Palo Alto, CA) used as template. Thermal cycling was accomplished with Advantage cDNA Polymerase Mix (Clontech) at an annealing temperature of 60°C. Reaction products were subcloned and identified as the partial human *PAX4* cDNA by sequencing. To isolate the 5′- and 3′-ends of the human *PAX4* cDNA, 5′ and 3′ rapid amplification cDNA ends (RACE) reactions were performed on human placental Marathon-Ready cDNA. Gene-specific primers corresponding to the sequence of identified partial human *PAX4* cDNA (*GSP3* for 5′-RACE, 5′-CCTTAAGGATCCGTGAGATGTCACA-3′; and *GSP4* for 3′-RACE, 5′-ATGGCGTCGGCAAGAGAAGCTCAAGT-3′) were designed. The reaction products were subcloned and sequenced.

These experiments yielded three identical clones for 5′-RACE and four clones for 3′-RACE, in which one clone was the longest; the other three clones were identical to each other and 516 base pair (bp) shorter than the longest. As a result, we isolated cDNA containing the complete coding region of the human *PAX4* gene (Fig. 1). At the 5′-end, there was an in-frame stop codon 190 bp upstream of the first methionine codon. At the 3′-end, a consensus AATAAA polyadenylation signal was located distantly upstream from the poly(A)$^+$ addition site of the longest clone, but 21 bases upstream of the 3′-end of the other three clones. The differences in size among the 3′-RACE clones occurred within the 3′-untranslated region (UTR); thus, the open reading frames (ORFs) were identical. This ORF encoded a predicted protein of 343 amino acids. Interestingly, the isolated human *PAX4* cDNA had relatively low homology to mouse *Pax4* (76.4% at amino acid level) compared with that between human and mouse for other members of the *PAX* gene family (>95% at amino acid level). These results originally suggested that our isolated cDNA was not *PAX4*, but perhaps a new *PAX* gene.

To examine this hypothesis, human genomic DNA was subjected to Southern blot analyses with fragments of the human placental *PAX4* cDNA and the mouse *Pax4* cDNA as probes under conditions of low stringency. Mouse *Pax4* cDNA was amplified with reverse transcription (RT)-PCR from total RNA of transfected mouse insulinoma cell line (βTC6-F7, supplied by S. Efrat, Albert Einstein Medical School) using a standard protocol. Southern blot analysis using a fragment of human *PAX4* cDNA (88.8% identical to mouse *Pax4* cDNA and <61.2% homologous to other human *PAX* genes) as a probe showed a single band that matched the restriction fragment of the cosmid g1572c264, and Southern blot analysis using the fragment of mouse *Pax4* cDNA (89.0% identical to the human *PAX4* cDNA) showed the same result on prolonged exposure (data not shown). These results indicated that the isolated cDNA represented the product of the real human *PAX4* gene as a single copy in the human genome.

The number and size of exons and the intron/exon boundaries (Table 1) were determined by comparison of the cDNA with the entire sequence of the cosmid g1572c264. The sequence of the nucleotide and deduced amino acid

From the Division of Metabolism, Endocrinology and Diabetes (T.T., J.W., E.B.-M., P.S.B., M.A.P.), Washington University Medical School, St. Louis, Missouri; Millennium Pharmaceuticals (L.D., J.M.), Cambridge, Massachusetts; and the Department of Endocrinology and Metabolism (S.C., B.G.), Hebrew University Hadassah Medical Center, Jerusalem, Israel.

Address correspondence and reprint requests to M. Alan Permutt, MD, Metabolism Division, Washington University School of Medicine, 660 S. Euclid Ave., Campus Box 8127, St. Louis, MO 63110. E-mail: apermutt@imgate.wustl.edu.

Received for publication 23 February 1998 and accepted in revised form 7 July 1998.

BLAST, Basic Local Alignment Search Tool; bp, base pair; HET, heterozygosity; NPL, nonparametric linkage; ORF, open reading frame; PCR, polymerase chain reaction; RACE, rapid amplification cDNA ends; RT, reverse transcription; UTR, untranslated region.

```
   1  CAAAGACTCACCCGTGAGCCAGCTCTCAAAGAAAGCAGCTTGCGTTGACAGCCTGGGGGC
  61  AGCAAGGATGCAGTCTCCCAGGAGAGGATGCACTCGGTGGTGGGAAGCCAGGCTGGAGGG
 121  GCCTGAGTGACCCTCTCCACAGGCGGGCAGGGCAGTGGGAGAGGTGGTGTGTGGATACCT

   1                                M  N  Q  L  G  G  L  F  V  N  G  R
 181  CTGTCTCACGCCCAGGGATCAGCAGCATGAACCAGCTTGGGGGGCTCTTTGTGAATGGCC

  13  P  L  P  L  D  T  R  Q  Q  I  V  R  L  A  V  S  G  M  R  P
 241  GGCCCCTGCCTCTGGATACCCGGCAGCAGATTGTGCGGCTAGCAGTCAGTGGAATGCGGC
                                                        GSP1

  33  C  D  I  S  R  I  L  K  V  S  N  G  C  V  S  K  I  L  G  R
 301  CCTGTGACATCTCACGGATCCTTAAGGTATCTAATGGCTGTGTGAGCAAGATCCTAGGGC
          GSP3                          RT1F

  53  Y  Y  R  T  G  V  L  E  P  K  G  I  G  G  S  K  P  R  L  A
 361  GTTACTACCGCACAGGTGTCTTGGAGCCAAAGGGCATTGGGGGAAGCAAGCCACGGCTGG

  73  T  P  P  V  V  A  R  I  A  Q  L  K  G  E  C  P  A  L  F  A
 421  CTACACCCCCTGTGGTGGCTCGAATTGCCCAGCTGAAGGGTGAGTGTCCAGCCCTCTTTG

  93  W  E  I  Q  R  Q  L  C  A  E  G  L  C  T  Q  D  K  T  P  S
 481  CCTGGGAAATCCAACGCCAGCTTTGTGCTGAAGGGCTTTGCACCCAGGACAAGACTCCCA
                                              RT1R

 113  V  S  S  I  N  R  V  L  R  A  L  Q  E  D  Q  G  L  P  C  T
 541  GTGTCTCCTCCATCAACCGAGTCCTGCGGGCATTACAGGAGGACCAGGGACTACCGTGCA

 133  R  L  R  S  P  A  V  L  A  P  A  V  L  T  P  H  S  G  S  E
 601  CACGGCTCAGGTCACCAGCTGTTTTGGCTCCAGCTGTCCTCACTCCCCATAGTGGCTCTG
                          RT2F

 153  T  P  R  G  T  H  P  G  T  G  H  R  N  R  T  I  F  S  P  S
 661  AGACTCCCCGGGGTACCCACCCAGGGACCGGCCACCGGAATCGGACTATCTTCTCCCCAA

 173  Q  A  E  A  L  E  K  E  F  Q  R  G  Q  Y  P  D  S  V  A  R
 721  GCCAAGCAGAGGCACTGGAGAAAGAGTTCCAGCGTGGGCAGTATCCTGATTCAGTGGCCC
          RT2R
                RT3F

 193  G  K  L  A  T  A  T  S  L  P  E  D  T  V  R  V  W  F  S  N
 781  GTGGAAAGCTGGCTACTGCCACCTCTCTGCCTGAGGACACGGTGAGGGTCTGGTTTTCCA

 213  R  R  A  K  W  R  R  Q  E  K  L  K  W  E  M  Q  L  P  G  A
 841  ACAGAAGAGCCAAATGGCGTCGGCAAGAGAAGCTCAAGTGGGAAATGCAGCTGCCAGGTG
                    RT3R                  GSP2
          GSP4

 233  S  Q  G  L  T  V  P  R  V  A  P  G  I  I  S  A  Q  Q  S  P
 901  CTTCCCAGGGGCTGACTGTACCAAGGGTTGCCCCAGGAATCATCTCTGCACAGCAGTCCC

 253  G  S  V  P  T  A  A  L  P  A  L  E  P  L  G  P  S  C  Y  Q
 961  CTGGCAGTGTGCCCACAGCAGCCCTGCCTGCCCTGGAACCACTGGGTCCCTCCTGCTATC

 273  L  C  W  A  T  A  P  E  R  C  L  S  D  T  P  P  K  A  C  L
1021  AGCTGTGCTGGGCAACAGCACCAGAAAGGTGTCTGAGTGACACCCCACCTAAAGCCTGTC

 293  K  P  C  W  G  H  L  P  P  Q  P  N  S  L  D  S  G  L  L  C
1081  TCAAGCCCTGCTGGGGCCACTTGCCCCCACAGCCGAATTCCCTGGACTCAGGACTGCTTT

 313  L  P  C  P  S  S  H  C  P  L  A  S  L  S  G  S  Q  A  L  L
1141  GCCTTCCTTGCCCTTCCTCCCACTGTCCCCTGGCCAGTCTTAGTGGCTCTCAGGCCCTGC

 333  W  P  G  C  P  L  L  Y  G  L  E  *
1201  TCTGGCCTGGCTGCCCACTACTGTATGGCTTGGAATGAGGCAGGAGTGGGAAGGAGATGG
1261  CATAGAGAAGATCTAATACCATCCTGCCCATTGTCCTTACCGTCCTGCCCATACAGACTG
1321  TGGCTCCTTCCTCCTTCCTGTGATTGCTCCCTCCTGTGTGGACGTTGCCTGGCCCTGCCT
1381  CGATGCCTCTCTGGCGCATCACCTGATTGGAGGGGCTGGTAAAGCAACACCCACCCACTT
1441  CTCACACTGGCCTTAAGAGGCCTCCACTCAGCAGTAATAAAAGCTGTTTTTATTAGCAGT
1501  AGTTCTGTTGTCCATCATGTTTTCCCTATGAGCACCCCTATGCCCACTCTAATATTCAAC
1561  AATTATAGACAATTTGCCCTATCATTTATTTACATCTATGTATCTACCATCTAATCTATG
1621  CATGTATGTAGGCAATACATGTATCTAAACAATGTATTTGTCAATGCATCAATTTACCTA
1681  CTCTATGTATGCATCTATATGTGTATTATGTATGTATGTGTGCATGCGTGCGCGCACACA
1741  CACACACACACATTGATATTATATCATGGCATTTTATTCCTAAATCTTCCAGCATGCATC
1801  CCCAAAAAACAAGAAACTTGTCTTACATAATCACAATAATATATCCACATCTAAGAAAAT
1861  TTACTGTAACTTCTTAATCTAAGAAAATTATGTATTTTTGTCATATGTATTTTGTCATAT
1921  GTATTTTGTATTTGCATATGTATTTTGTATTTGCATATGTATTTTTGTCATAGCAGCAAA
1981  CAGAGTGAAATGCCATTTTTCATATTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

**FIG. 1. Nucleotide and predicted amino acid sequences of the ORF of the human *PAX4* gene. One-letter code is used for amino acid sequences. The $NH_2$-terminal paired domain (aa 2-124) and the paired-type homeodomain (aa 163-222) are underlined. The termination codon (*), and the oligonucleotide primers for generating cDNA, Southern blot analysis, and RT-PCR analysis (*arrows*) are indicated.**

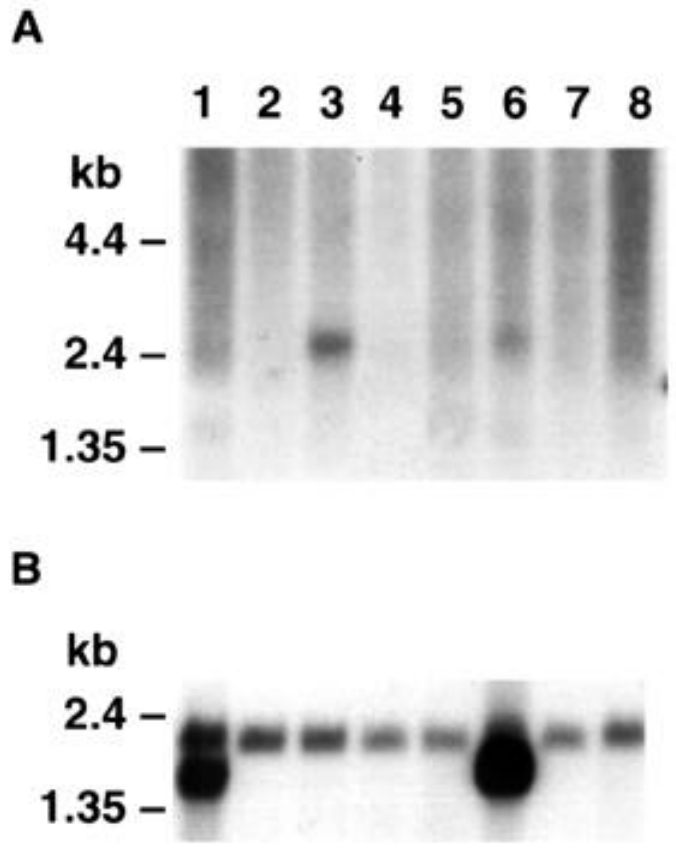


**FIG. 2. Northern blot analysis of 2 μg of poly(A)$^+$ mRNA for *PAX4* in human tissues: heart (*lane 1*), brain (*lane 2*), placenta (*lane 3*), lung (*lane 4*), liver (*lane 5*), skeletal muscle (*lane 6*), kidney (*lane 7*), and pancreas (*lane 8*). *A*: Detection of mRNAs for *PAX4*. *B*: Detection of mRNAs for β-actin. The relative positions of size markers are indicated.**

sequences of all exons were identical to the cDNA-derived sequences except for the alteration of C to T after the deletion of a $(CA)_4$ dinucleotide repeat in the 3′-UTR. Recently, a full-length cDNA of mouse *Pax4* was isolated, and the deduced sequence of human *PAX4* gene was proposed solely by comparison of the sequence of mouse *Pax4* cDNA and human cosmid g1572c264 (5). The deduced human homolog included an additional exon on the 5′-end not present in our *PAX4* cDNA. Furthermore, the proposed human exon 9 was completely different from the exon shown to exist in our *PAX4* cDNA. This additional exon might be an alternatively spliced isoform, yet it should be noted that the expression of this putative homolog was not confirmed in any human tissues, whereas our sequence is derived from cDNA obtained from amplified human mRNA.

The expression of mRNA for the human *PAX4* gene in normal adult human tissues was determined by Northern blot analysis (Fig. 2). The Multiple Tissue Northern Blot (Clontech) containing 2 μg of poly(A)$^+$ RNA from various human tissues was hybridized with $^{32}$P-labeled entire human *PAX4* cDNA. A single transcript of ~2.5 kb was detected, predominantly in placenta and skeletal muscle. With longer exposure, faint bands were observed in heart and pancreas. To confirm the expression of *PAX4* mRNA in pancreatic islets, RT-PCR was performed using human islet RNA and three sets of primers: *RT1F* and *RT1R*; *RT2F*, 5′-CTGTTTTGGCTCCAGCTGTC-3′ and *RT2R*, 5′-CTTTCTCCAGTGCCTCTGCT-3′; *RT3F*, 5′-GAGGCACTGGAGAAAGAGTT-3′ and *RT3R*, 5′-ACTTGAGCTTCTCTTGCCGA-3′. Human pancreatic islets were obtained from Drs. Paul Lacy and Thalachallour Mohanakumar (Washington University) as described (6).

TABLE 1
Splice junction, exon, and intron sequences for the human *PAX4* gene

| Exon | Size (bp) | Intron/exon junctions: Acceptor sequence | Intron/exon junctions: Donor sequence | Intron | Size (bp) |
|---|---|---|---|---|---|
| 1 | 326 | gggcaaggaaCAAAGACTCA | GATCCTTAAGgtaatgggcc | 1 | 305 |
| 2 | 216 | cctgactcagGTATCTAATG | GACTCCCAGTgtaagtaccc | 2 | 322 |
| 3 | 76 | acactcccagGTCTCCTCCA | AGGTCACCAGgtgggtcttg | 3 | 600 |
| 4 | 126 | tccttttcagCTGTTTTGGC | CTGGAGAAAGgtgctgggct | 4 | 224 |
| 5 | 83 | atctccgcagAGTTCCAGCG | CACGGTGAGGgtgagtgagc | 5 | 358 |
| 6 | 70 | cctcgctcagGTCTGGTTTT | CAGCTGCCAGgtgatttctc | 6 | 1,013 |
| 7 | 56 | cattttatagGTGCTTCCCA | CTCTGCACAGgtactcggga | 7 | 268 |
| 8 | 142 | ttttttttagCAGTCCCCTG | CCCTGCTGGGgtaagaattc | 8 | 328 |
| 9 | 913 | gtccccacagGCCACTTGCC | TTCATATTCTgctttccatt | | |

First-strand cDNA was synthesized from 5 µg of total human islet RNA using the SuperScript Preamplification System with oligo(dT) primer (GibcoBRL, Gaithersburg, MD). A portion of cDNA was then used as the template in subsequent PCR amplification with three sets of primers. Each amplification (30 cycles) was performed under standard conditions with 2.0 mmol/l $MgCl_2$ and the following annealing temperatures: 64°C for *RT1F* and *RT1R*, 57°C for *RT2F* and *RT2R*, and 55°C for *RT3F* and *RT3R*. The reaction products were analyzed by electrophoresis on a 1% low melting point agarose gel. The amplified DNA was not well visualized after the first round of amplification, so a portion of the first amplification sample was used for a second 30-cycle amplification under the same conditions. The DNA was recovered from the gel by Wizard PCR Preps DNA Purification System kit (Promega, Madison, WI), followed by subcloning and sequencing. Three fragments were successively amplified (data not shown). These results also indicated that within the amplified regions no alternatively spliced isoforms of *PAX4* mRNA could be detected between placenta and pancreatic islets.

We identified a $(GT)_n$ dinucleotide repeat (GenBank accession no. AF047018) 2.3 kb downstream of the end of the *PAX4* gene and a $(CA)_n$ dinucleotide repeat (GenBank accession no. AF047019) 16 kb downstream of the *PAX4* gene (Fig. 3). We observed 6 alleles for AF047018 (observed heterozygosity [HET] = 0.4995) and 11 alleles for AF047019 (HET = 0.8278) among 116 unrelated Ashkenazi Jews. Sequence analysis of AF047019 revealed that it was a composite of three polymorphic $(CA)_n$ repeats exhibiting extensive polymorphism (data not shown).

To further refine the location of this gene on the genetic and physical maps of chromosome 7q, AF047018 was used to screen the Stanford radiation hybrid G3 panel (Research Genetics, Huntsville, AL). The closest marker was shown to be Cda1dh10 (logarithm of odds [LOD] score = 1,000.0), which is located between D7S2876 and D7S635 (Fig. 3*A*). Further BLAST searching revealed that Cda1dh10 matched a sequence in the cosmid g1572c101 (GenBank Accession No. AC000357), and this cosmid overlapped the cosmid g1572c264. A contig of the cosmid clones suggested the physical relationship among the genetic markers (Fig. 3*B*).

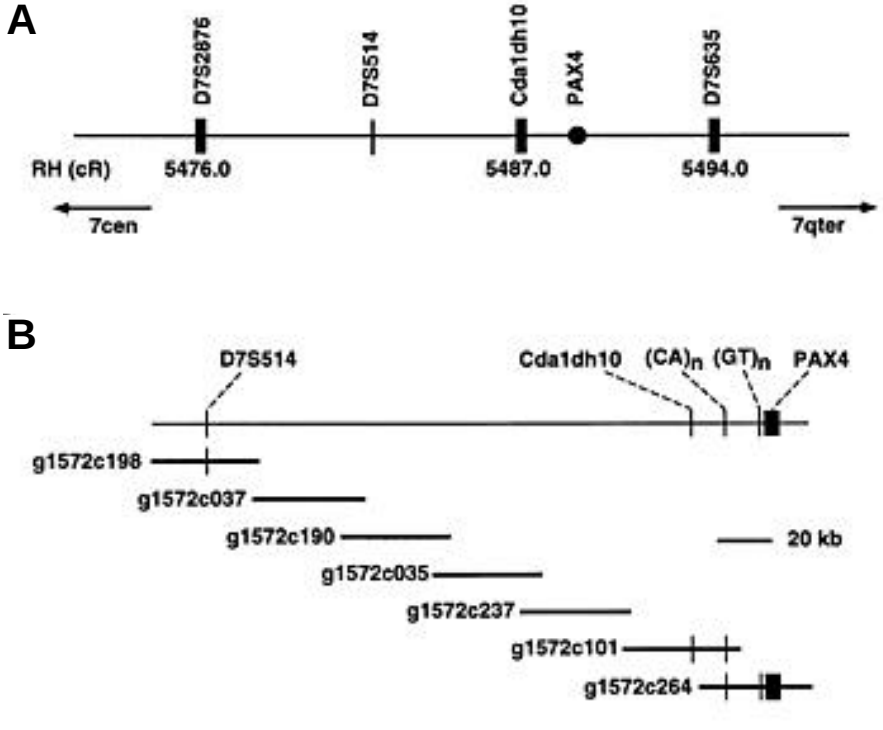


**FIG. 3. *A*: Radiation hybrid (RH) map of genetic markers flanking the human *PAX4* gene on chromosome 7q. Marker names and the *PAX4* gene are noted above the line. Numbers directly below the line represent genetic distances in centiRads (cR) from the top of the chromosome 7 linkage group on the Stanford radiation hybrid map. cen, centromere; qter, long arm terminal. *B*: Physical map of genetic markers flanking the *PAX4* gene as determined by defining a contig of human cosmid clones. Additional information on these cosmid clones is available at GenBank database. Accession numbers are AC000127 (g1572c198), AC000125 (g1572c037), AC000126 (g1572c190), AC000124 (g1572c035), AC000358 (g1572c237), AC000357 (g1572c101), and AC000359 (g1572c264).**

This result indicated that the gene maps to human chromosome 7q31.1-32 border. No crossovers were found between D7S514 and D7S635 in eight Centre d'Etudes du Polymorphisme Humain (CEPH) reference families, which indicates that extended haplotypes comprising the new markers and close genetic markers would be useful for further linkage studies of the *PAX4* gene with human inherited disease.

To assess the possible implication of the *PAX4* gene in diabetes, we performed affected sib-pair analysis by genotyping the new markers. Eighty-five Ashkenazi Jewish families with at least two siblings with overt type 2 diabetes were selected. Diagnosis was based on the World Health Organization's 1985 criteria. The number of affected individuals available was 205 (114 affected sib-pairs). The mean age at the time of examination of the diabetic probands was 60.4 years (range, 42–82), with mean age of diabetes diagnosis at 45.8 years (range, 30–65). We subjected the data to both single-locus and multipoint parametric analyses using GENEHUNTER (7). Allele frequencies were estimated from the data on 116 unrelated Ashkenazi Jews. We found no evidence for excess allele sharing at either region: the maximum nonparametric linkage (NPL) score was 0.509 ($P$ = 0.303) at AF047018, and the NPL score was 0.877 ($P$ = 0.188) at AF047019. Our results indicate that genetic variation in the human *PAX4* gene region is unlikely to be a major contributor to the pathogenesis of type 2 diabetes in Ashkenazi Jews. However, sib-pair analysis in a polygenic disease with only 114 sib-pairs has limited power to detect linkage, especially if multiple genes contribute to the total genetic susceptibility. The actual power to detect linkage is difficult to estimate without knowledge of the total genetic risk, the model of inheritance, or the number of genes contributing to the total risk. Yet power to detect linkage is >95% with a total genetic risk of 5- to 10-fold, with a single major gene contributing (8). These results do not exclude the possibility of a minor role for the *PAX4* gene in a polygenic model, however. It is also possible that *PAX4* contributes to type 2 diabetes in other racial or ethnic groups and that the markers defined here will be useful for these analyses.

## ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health Grants DK-16746 (M.A.P.), DK-49583 (M.A.P., S.C., L.D., J.M., B.G.), and DK-07120 (P.S.B.). T.T. was the recipient of a Mentor-Based Fellowship Award from the American Diabetes Association. Oligonucleotides and human islets were provided by the Protein Chemistry and Islet Cores of the Diabetes Research and Training Center (National Institutes of Health Grant 2P60-DK-20579), respectively.

The authors would also like to thank Jeannie Wokurka for preparation of the manuscript. We also wish to acknowledge Roche Pharmaceuticals for their support in collection of families.

## REFERENCES

1. Dahl E, Koseki H, Balling R: Pax genes and organogenesis. *Bioessays* 19:755–765, 1997
2. Sosa-Pineda B, Chowdhury K, Torres M, Oliver G, Gruss P: The Pax4 gene is essential for differentiation of insulin-producing β-cells in the mammalian pancreas. *Nature* 386:399–402, 1997
3. St-Onge L, Sosa-Pineda B, Chowdhury K, Mansouri A, Gruss P: Pax6 is required for differentiation of glucagon-producing α-cells in mouse pancreas. *Nature* 387:406–409, 1997
4. Porte DJ: β-Cell in type II diabetes mellitus. *Diabetes* 40:166–180, 1991
5. Matsushita T, Yamaoka T, Otsuka S, Moritani M, Matsumoto T, Itakura M: Molecular cloning of mouse paired-box-containing gene (Pax)-4 from an islet β cell line and deduced sequence of human Pax-4. *Biochem Biophys Res Commun* 242:176–180, 1998
6. Ricordi C, Lacy PE, Finke EH, Olack BJ, Scharp DW: Automated method for isolation of human pancreatic islets. *Diabetes* 37:413–420, 1988
7. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347–1363, 1996
8. Risch N: Linkage strategies for genetically complex traits: II. The power of affected relative pairs: *Am J Hum Genet* 46:229–241, 1990

Author Queries (please see Q in margin and underlined text)

Q1: In sentence beginning “This encoded a predicted protein...,” please add word(s) after “This” to clarify its meaning. Thanks.
Q2: **Figure 2 legend—OK to change “of mRNA (2 µg polyA$^+$)” to “2 µg of poly(A)$^+$ mRNA,” as given in text?**
Q3: Please provide first name for Dr. Mohanakumar. Thanks.
Q4: Were the 6 alleles for AF047018 (as well as the 11 alleles for AF047019) observed among Ashkenazi Jews? Current punctuation of sentence indicates that they were not.
Q5: Please provide city and state in which Research Genetics, Inc. is located. Thanks.
**Q6:** Please provide expansion of the abbreviation CEPH. Thanks.
Q7: **Is “centiRays (cR)”. Fig. 3 legend—Is “radiation hybrid” the correct expansion of “RH,” shown in part *A*? ><<AU: Fig. 3 legend—Is “centiRays (cR)” correct? I could not locate this unit in our standard reference sources. Fig. 3 legend—Are “centromere” and “long arm terminal” the correct expansions of “cen” and “qter”?**
Q8: Is GENEHUNTER the name of a software program?
Q9: Should “5–10” in “total genetic risk of 5–10” be followed by “%”?

Ref. 3—Is St-Onge correct last name of first author?