# *Brief Report*

# Evaluation of Polymorphic Splicing in the Mechanism of the Association of the Insulin Gene With Diabetes

**Luc Marchand and Constantin Polychronakos**

**The association of type 1 diabetes with the insulin gene (*IDDM2* locus) has been mapped to a short haplotype encompassing two single nucleotide polymorphisms (SNPs) in perfect linkage disequilibrium ($r^2 = 1$) with each other and with the two allele classes at the variable number of tandem repeats (VNTR) polymorphism upstream of the transcription site. Although it is believed that the genetic effect is mediated through transcriptional effects of the VNTR, an alternative mechanism has been recently proposed: In transfected cells, the common A allele at one of the SNPs (−23A→T, in relation to the translation-initiation codon) weakens the splicing of intron 1, resulting in a minor (~15% of total RNA) transcript with a longer 5′ untranslated region and sixfold enhanced translational efficiency. The purpose of our study was to confirm these findings in RNA from normal human pancreas and thymus. We report that pancreas does contain the alternative transcript in an allele-dependent manner but at a very low proportion (<5% of total INS mRNA). We believe that this level would have a minor, if any, biological effect involved in the mechanism of the *IDDM2* locus. *Diabetes* 56: 709–713, 2007**

A haplotype within a 2-kb linkage disequilibrium (LD) block encompassing the human insulin gene (*INS*) is strongly associated with type 1 diabetes (1–3) and, possibly, also with plasma insulin levels (4), juvenile obesity (5), and type 2 diabetes (6). The association maps to a haplotype defined by a repeat polymorphism upstream of *INS* that consists of variable number of tandem repeats (VNTR). The short class I alleles predispose to type 1 diabetes, whereas the long class III alleles have a dominantly protective effect (2,3). However, because of tight LD, the effect cannot be genetically dissected from two single nucleotide polymorphisms (SNPs), rs689 (−23A→T, from translation start) and rs3842753 (+1140C→A), both of which have an $r^2$ of ~1 with the VNTR classes (I vs. III), and either of which may account for part or all of the genetic effect (3). Further dissection of the locus requires functional studies.

In the absence of coding polymorphisms, the association must be due to allelic effects on transcription, splicing, and/or RNA stability. In pancreas, steady-state mRNA derived from class III chromosomes is indeed ~20% lower than from class I (2,7). A much more dramatic effect in the opposite direction (class III is two- to threefold higher than class I) is seen in the thymus, where insulin is expressed at low levels (8,9) for the purpose of self-tolerance (10,11). Thus, a transcriptional effect by the VNTR seems likely but is not definitively proven. An alternative mechanism was recently proposed (12).

Exon 1 of *INS* is untranslated and the entire coding sequence is contained in exons 2 and 3 (Fig. 1). SNP −23A→T (often referred to as −23*Hph*I and renamed IVS-6A/T in 12) maps to near the 3′ end of the 179-bp intron 1. Allele T maintains while A disrupts a polypirimidine tract that promotes splicing at the 3′ end of intron 1. On this basis, Královičová et al. (12) predicted that allele A will promote alternative splicing in which intron 1 is retained and becomes part of the 5′ untranslated region (UTR). A minor transcript corresponding to this was indeed found by transfecting human mini-gene constructs in four nonpancreatic cell lines and, in a single experiment, in the rat INS-1E β-cell line; it was more abundant in constructs of the A than of the T allele (12). As the same report also found that the longer transcript is translated with sixfold higher efficiency, it could—even if it represents only ~15% of RNA—double peptide synthesis from chromosomes carrying the A allele. This effect could be involved in the biological mechanism the *IDDM2* locus.

However, splicing is tissue specific, and isoforms found in unrelated cells, or even tumors or cell lines derived from the tissue of interest, may be irrelevant. As an example, insulinomas preferentially use a cryptic splicing site that results in an isoform retaining 22 nucleotides of intron 1 (Fig. 1), only traces of which are found in normal β-cells (13). Interestingly, this cryptic splicing site is activated in the presence of a 4-bp insertion (12) and is frequent in Africans (~25%) but absent in the European-descent populations (14) in which the *IDDM2* effect was observed.

The aims of this study were to confirm the allelic effect in RNA from human pancreas and thymus and to assess its quantitative importance.

We first examined the available human expressed sequence tag (EST) database (Online Table 1, available at http://dx.doi.org/10.2337/db06-0402). Besides the two libraries analyzed in ref. 12, there were four more containing at least five clones informative for splicing and having at least one A allele at −23A→T. None of the four showed intron 1 retention (0 of a total of 119 informative clones;

From the Endocrine Genetics Laboratory, McGill University Health Center Research Institute, Montréal, Québec, Canada.

Address correspondence and reprint requests to Constantin Polychronakos, MD, McGill University Health Center, Montreal Children's Hospital, 2300 Tupper St., Ste. C244, Montreal, Quebec, Canada H3H 1P3. E-mail: constantin.polychronakos@mcgill.ca.

Received for publication 26 March 2006 and accepted in revised form 18 December 2006.

Additional information for this article can be found in an online appendix at http://dx.doi.org/10.2337/db06-0402.

LD, linkage disequilibrium; SNP, single nucleotide polymorphism; UTR, untranslated region; VNTR, variable number of tandem repeats.

DOI: 10.2337/db06-0402

© 2007 by the American Diabetes Association.

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

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/56/3/709/663605/zdb00307000709.pdf by guest on 13 July 2024

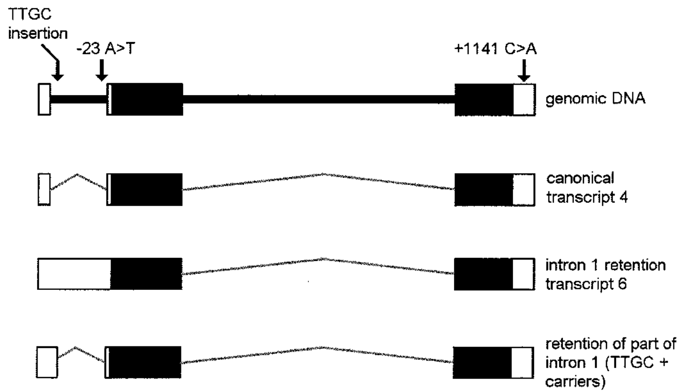


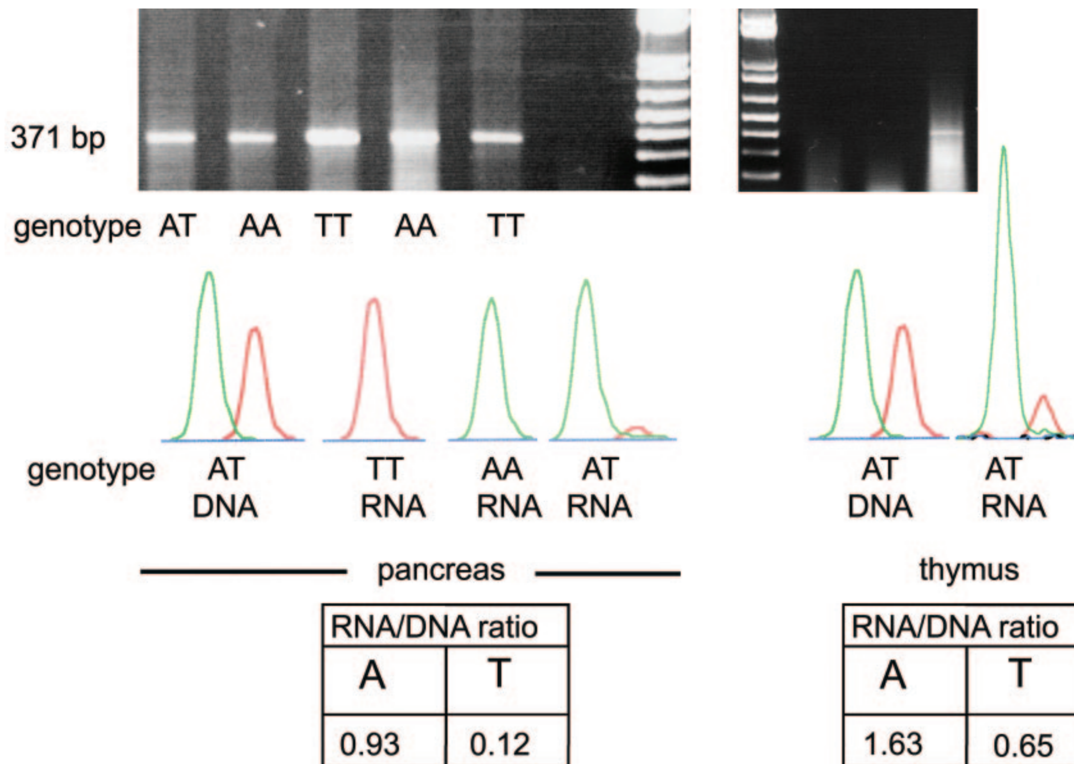

**FIG. 1. Schematic diagram of the insulin gene, the canonical transcript (T4), and the transcripts retaining all (T6) or part of intron 1. Coding region is solid black; UTRs are white. The position of the PCR primers is indicated. Numbering of the transcripts follows the scheme in ref. 12. Additional transcripts found in that in vitro study, in which coding sequences from exons 2 and 3 are spliced out, are not shown because we found no convincing evidence of their existence in pancreas or thymus.**

95% CI 0–4.5%). From the two libraries analyzed by Královičová et al. (12), the islet HR85 library appears to be the only exception where sequences retaining intron 1 are in significant abundance, representing 15 of 47 informative clones (24%). However, the inserts used to construct this library were selected for larger sizes (1.5 kb), thus skewing isoform representation (http://image.llnl.gov/image/html/humlib_info.shtml#HR85%20islet).

Transcripts splicing out part of coding sequences, observed in vitro (transcripts 1, 2, 3, and 5 of ref. 12) were rare—9 of 2,009 ESTs—underscoring the pitfall in extrapolating from in vitro studies.

In our examination of RNA from pancreas and thymus, we first confirmed the allele specificity of intron 1 retention by use of RT-PCR primers that specifically amplified mRNA retaining intron 1 (transcript 6 in Fig. 1, abbreviated T6 to maintain the numbering of ref. 12). All samples showed amplification product (Fig. 2, *top panel*), and genotype had no apparent effect on band intensity. However, because PCR was carried beyond the logarithmic phase, quantitative comparison between samples is not meaningful; alleles can still be compared within heterozygous samples (competitive RT-PCR). Thus, in AT heterozygotes, relative allele quantification by single nucleotide primer extension clearly showed that most of the T6 transcripts carry the A allele. After correction for labeling and fluorochrome efficiency using heterozygous DNA as standard for 1:1 stoichiometry, the A allele was five- to sixfold more abundant than the T allele (mean $\pm$ SD = 5.7 $\pm$ 1.7–fold in four samples). Therefore, we confirm the observation (12) that intron 1 retention is driven mostly by A alleles.

Given the allele dependence of splicing, we then asked whether our previous finding of allelic imbalance in the thymus might be due to the presence of different, allele-dependent isoforms in the 3′ UTR RT-PCR used to demonstrate it. To answer this question, we amplified transcripts containing all three exons, using primers S1 and AS1 from a thymus sample heterozygous for $-$23A$\rightarrow$T (Fig. 3). The band corresponding to the 451-bp canonical transcript T4 was gel-purified and examined for relative expression of the two alleles. There was no band corre-




| RNA/DNA ratio (pancreas) | |
|---|---|
| A | T |
| 0.93 | 0.12 |

| RNA/DNA ratio (thymus) | |
|---|---|
| A | T |
| 1.63 | 0.65 |

**FIG. 2. *Top panel*: RT-PCR amplification of a 371-bp fragment from intron 1 to exon 3 specific for transcripts retaining intron 1 (primers S2 and AS2) from five samples of pancreas (*left*) and one of thymus (*right*). *Bottom panel*: Determination of allele ratios at $-$23A$\rightarrow$T by single nucleotide primer extension. DNA (equal allele representation) is used to correct for labeling and fluorochrome efficiency. Representative examples are shown of heterozygous DNA; pancreatic RNA of AA, AT, and TT genotypes; and DNA and RNA from the one AT heterozygous thymus. Green = A, red = T. The AT heterozygous pancreas result is representative of a total of five different heterozygous pancreata, all showing a remarkably similar allele ratio.**

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/56/3/709/663605/zdb00307000709.pdf by guest on 13 July 2024

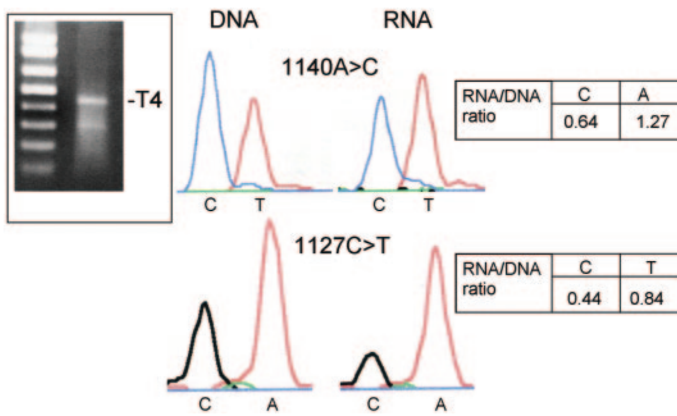



FIG. 3. RT-PCR product from a heterozygous thymus (primers S1 and AS1) was purified from the 451 band (canonical transcript T4) and subjected to competitive allele quantification as in Fig. 2 at two SNPs in the 3′ UTR. Relative allele strength in DNA and RNA at each of the two SNPs is shown. Because of LD (see text), C at 1140A→C marks the class I and A the class III. At 1127C→T, C marks class I and T marks class III.

sponding to T6 (Fig. 3). Because −23A→T is spliced out in these transcripts, we used two 3′ UTR SNPs as proxies. In +1140A→C, allele C is in perfect ($r^2$ = 1) LD with VNTR class I (A at −23) and A with VNTR class III (T at −23) (3). In +1127C→T (rs3842752), allele T invariably predicts a class III VNTR (T at −23) (3,8) and, in individuals heterozygous at both −23A→T and +1,127 C→T, C predicts class I-A with virtual certainty (99.93% of cases, see METHODS in ref. 8). By both SNPs, transcripts from the class III chromosome are approximately twice as abundant as those from the class I chromosome. Therefore, the allelic imbalance is present in gel-purified T4 transcripts and cannot be due to differential presence of allele-dependent splicing isoforms. A VNTR transcriptional effect remains the most compelling explanation. To confirm the absence of the T6 transcript in thymus, we amplified an additional five RNA samples using primers S1 and AS1. To maximize visualization of faint bands, RT-PCR products were subjected to Southern blot (Fig. 4). T4, the canonical transcript, was seen in all samples, but no band corresponding

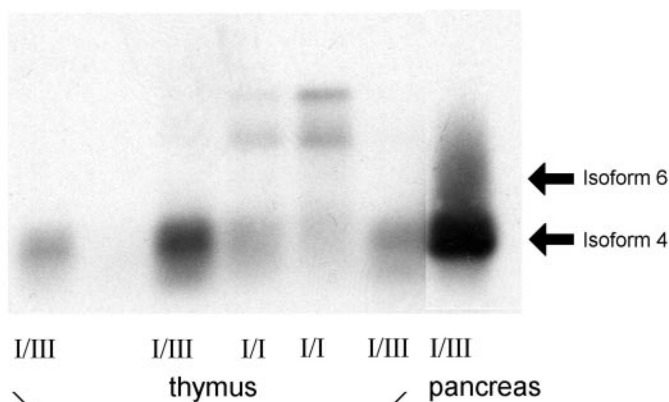



FIG. 4. To confirm the absence of isoform T6 in thymus RNA, as seen in Fig. 3, we amplified by RT-PCR an additional five samples with primers S1 and AS1. To optimize detection of the much lower level of insulin message in thymus, RT-PCR bands were blotted and probed with a dig-labeled probe corresponding to the sequence of isoform T4. The high–molecular weight bands seen in two of the samples do not correspond to any possible splicing isoform and represent some PCR artifact. This is representative of two separate experiments with five thymus samples each.

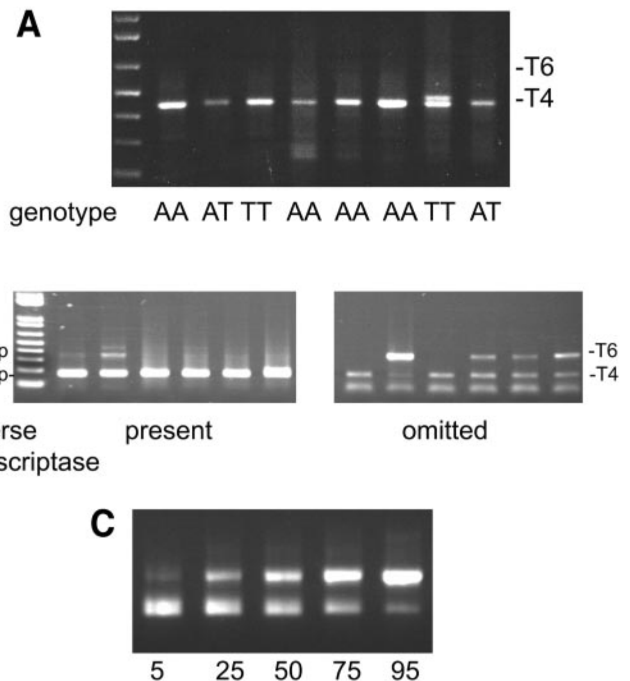



FIG. 5. *A*: Splicing-isoform spectrum of pancreatic insulin. RT-PCR amplification using primers S1 and AS1 encompasses all expected isoforms shown in Fig. 1. With the exception of the sixth sample, the canonical T4 transcript (451-bp band) is the only one seen. The sixth sample is from an individual carrying the TTGC insertion found to activate a cryptic site that adds 22 bp from the 5′ end of intron 1 (confirmed by sequencing). *B*: In an additional attempt to detect transcripts retaining intron 1, we amplified with a different set of primers from exon 1 to exon 2 (primers S1 and AS3). Only one sample shows a clear 630 band consistent with intron 1 retention (T6 isoform). The same sample also amplified a strong 630-bp band even when reverse transcriptase was omitted from the RT reaction, indicating that the 630 band was due to contamination with genomic DNA. *C*: Because longer sequences may amplify less efficiently in competitive RT-PCR, we defined our sensitivity to detect small proportions of T6 in competition against T4. A T4 DNA template was spiked with the indicated percentage of T6. The lowest concentration (5%) can be easily detected after PCR amplification with primers S1 and AS3.

to the intron-containing T6 isoform was present. Therefore, intron 1 retention in thymus, if it occurs at all, is quantitatively insignificant.

Finally, to address the question of whether the enhanced translational efficiency of T6 may be contributing to the *IDDM2* effect, we defined its relative representation within total pancreatic mRNA, which should determine the proportion of insulin peptide synthesis attributable to this isoform. Fig. 5 shows the results of competitive RT-PCR, which we used as the most direct way of defining the relative quantitative contribution of different isoforms amplified with the same primers in the same reaction. Full-length *INS* cDNA was amplified using primers S1 and AS1 (Fig. 1) from eight pancreas RNA samples (Fig. 5*A*). Transcript identity as inferred from size was confirmed in all cases by extracting and sequencing PCR bands. The canonical transcript (T4) is the only clearly visible band. No band corresponding to the size of T6 can be seen. The extra band seen in the seventh sample is due to retention of 22 bp at the 5′ end of intron 1, resulting from activation of the cryptic splice site (12) in an individual carrying the TTGC insertion (12) (confirmed by sequencing). Thus, we confirm the effect of +TTGC on splicing as previously reported (12). This insertion, common in Africans (∼25%), is never seen in Europeans, in whom the *IDDM2* effect was

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/56/3/709/663605/zdb00307000709.pdf by guest on 13 July 2024

observed. Association of this SNP with type 1 diabetes is worth studying in Africans, but it cannot explain the *IDDM2* effect observed in Europeans.

In a further attempt to co-amplify T6 with T4, RNA from six human pancreas samples was RT-PCR amplified with primers S1 and AS3, going from exon 1 to exon 2 (Fig. 5*B*). Because the sequence of RT-PCR products retaining intron 1 is identical with genomic DNA, we are also showing PCR products from the same RNA samples, processed identically except for the omission of reverse transcriptase. A distinct band corresponding to T6 is seen only in the second sample, which is clearly contaminated with genomic DNA. Of the remaining five samples, two AA homozygotes show no T6 bands, while an AT heterozygote has a very faint band. Thus, the relative contribution of T6 to the total amount of insulin mRNA is very low, even if multiplied by a sixfold greater translational efficiency.

However, because T6 is longer than T4 and may be amplified with lower efficiency, we had to evaluate the sensitivity of our technique to detect T6 at low proportions in competitive PCR. Figure 5*C* shows PCR amplification of templates generated by spiking T4 DNA with T6 at the proportions indicated. We easily detected T6 at 5% of total transcripts. This allows the conclusion that T6, when it is present at all, is well below 5% of total transcript, much lower than might be suggested by splicing in cultured cells (12).

Therefore, we confirm the allelic effect on splicing described in ref. 12. We also show that the in vitro system used in that report seriously overestimated the abundance of the allele-dependent alternative transcript and, hence, its biological significance. If in a few individuals T6 represents as much as 3–4% of *INS* mRNA (the most generous estimate from our data), the enhanced translation from T6 would result in a ~15–18% higher peptide level in AA than in AT individuals, the two genotypes mainly responsible for the *IDDM2* effect (TT is rare). The presence of detectable T6 sequence in only one of several samples raises the intriguing possibility of genetic heterogeneity in *INS* alternative splicing. If present, it must be controlled in *trans*, as other *INS* SNPs do not influence splicing in vitro (12). Thus, intriguing as this possibility might be, it still would not explain the *IDDM2* effect, even assuming that <20% higher expression in the occasional individual is of biologic significance.

It would also be interesting to speculate on the immune consequences, in view of its genetic behavior, of increased insulin synthesis associated with the A allele. Enhancing expression of the β-cell autoantigen (insulin) might promote the type 1 diabetes autoimmune process, as is compatible with the predisposing effect of the haplotype. At the level of the thymus, however, where allelic differences may be more important, higher levels would result in better insulin-specific self-tolerance. This is difficult to reconcile with the increased diabetes risk conferred by the A allele haplotype. We believe that our findings place this allelic alternative splicing in quantitative perspective and suggest minimal effect, if any at all.

## RESEARCH DESIGN AND METHODS

**Analysis of INS libraries.** INS EST libraries were accessed at the National Center for Biotechnology Information (NCBI) Web site (http://www.ncbi.nlm.nih.gov/UniGene/lbrowse2.cgi?TAXID=9606).

**Tissue samples and RNA preparation.** Pancreas and thymus samples were obtained from human fetal tissue from pregnancy terminations for reasons other than maternal or fetal disease, as approved by the local institutional review board.

**Primers and PCR conditions.** One microgram (pancreas) or 5 μg (thymus) RNA was reverse transcribed using random primers and Superscript II (Invitrogen) for pancreas or the Quantitect kit (Qiagen, Hilden, Germany) for thymus.

The primers were as follows:

For all isoforms: S1 5′ATCAGAAGAGGCCATCAAGC3′; AS1 5′TTCCATCTCTCTCGGTGCAG3′.

For isoform 6 specific: S2 5′GAAGCATGTGGGGGTGAG3′; AS2 5′CACAATGCCACGCTTCTG3′.

For the shorter PCR product of T4 and T6, S1 with this antisense primer: AS3 5′CCCCGCACACTAGGTAGAGA3′.

PCR cycling conditions were 35 cycles at 94°C for 20 s, 54°C for 20 s, and 72°C for 30 s.

**Allelic imbalance measurements.** Alleles were detected and quantified in RT-PCR products by single nucleotide primer extension, using allelic ddNTPs labeled with different fluorochromes. Specific bands from the PCR products were gel purified (Gel Extraction Kit; Qiagen, Hilden, Germany). Two microliters of extracted material was mixed with 2.5 pmol of probe and 5 μl of Snapshot mix (Applied Biosystems, Foster City, CA) in 10 μl total volume and cycled as follows: 25 cycles at 94°C for 10 s followed by 60°C for 35 s. After digestion with 1 unit of shrimp alkaline phosphatase (ABI) to remove unincorporated primers, reaction products were denatured in formamide at 95°C for 5 min and electrophoresed in an ABI 310 sequencing apparatus. Results were normalized taking the ratios in genomic DNA as 1:1 in order to obtain allelic ratios values. The probe for −23T→A was 5′GCCTCAGCCCGCCTGTC3′ and for +1140C→A 5′AGGAGGCGGCGGGTG3′.

## ACKNOWLEDGMENTS

This work was supported by a grant to C.P. from Genome Québec and Genome Canada. We thank Dr. Cindy G. Goodyer for providing the tissue samples.

## REFERENCES

1. Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 33:176–183, 1984
2. Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Kawaguchi Y, Dronsfield MJ, Pociot F, et al.: Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet* 9:284–292, 1995
3. Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, Concannon P, Gharani N, McCarthy MI, Olavesen MG, McCormack R, Guja C, Ionescu-Tîrgoviste C, Undlien DE, Rønningen KS, Gillespie KM, Tuomilehto-Wolf E, Tuomilehto J, Bennett ST, Clayton DG, Cordell HJ, Todd JA: Remapping the insulin gene/*IDDM2* locus in type 1 diabetes. *Diabetes* 53:1884–1889, 2004
4. Dos Santos C, Fallin D, Le Stunff C, LeFur S, Bougnères P: *INS* VNTR is a QTL for the insulin response to oral glucose in obese children. *Physiol Genomics* 16:309–313, 2004
5. Le Stunff C, Fallin D, Bougnères P: Paternal transmission of the very common class I INS VNTR alleles predisposes to childhood obesity. *Nat Genet* 29:96–99, 2001
6. Huxtable SJ, Saker PJ, Haddad L, Walker M, Frayling TM, Levy JC, Hitman GA, O'Rahilly S, Hattersley AT, McCarthy MI: Analysis of parent-offspring trios provides evidence for linkage and association between the insulin gene and type 2 diabetes mediated exclusively through paternally transmitted class III variable number tandem repeat alleles. *Diabetes* 49:126–130, 2000
7. Vafiadis P, Bennett ST, Colle E, Grabs R, Goodyer CG, Polychronakos C: Imprinted and genotype-specific expression of genes at the *IDDM2* locus in pancreas and leucocytes. *J Autoimmun* 9:397–403, 1996
8. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by *INS* VNTR alleles at the *IDDM2* locus. *Nat Genet* 15:289–292, 1997
9. Pugliese A, Zeller M, Fernandez A Jr, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD: The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet* 15:293–297, 1997
10. Chentoufi AA, Polychronakos C: Insulin expression levels in the thymus modulate insulin-specific autoreactive T-cell tolerance: the mechanism by which the IDDM2 locus may predispose to diabetes. *Diabetes* 51:1383–1390, 2002
11. Kyewski B, Derbinski J: Self-representation in the thymus: an extended view. *Nat Rev Immunol* 4:688–98, 2004
12. Královičová J, Gaunt TR, Rodriguez S, Wood PJ, Day INM, Vořechovský I:

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/56/3/709/663605/zdb00307000709.pdf by guest on 13 July 2024

Variants in the human insulin gene that affect pre-mRNA splicing: is –23HphI a functional single nucleotide polymorphism at *IDDM2*? *Diabetes* 55:260–264, 2006
13. Minn AH, Kayton M, Lorang D, Hoffmann SC, Harlan DM, Libutti SK, Shalev A: Insulinomas and expression of an insulin splice variant. *Lancet* 363:363–367, 2004
14. Stead JD, Hurles ME, Jeffreys AJ: Global haplotype diversity in the human insulin gene region. *Genome Res* 13:2101–2111, 2003

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/56/3/709/663605/zdb00307000709.pdf by guest on 13 July 2024