

Brief Genetics Report

Variants in the Human Insulin Gene That Affect Pre-mRNA Splicing

Is $-23HphI$ a Functional Single Nucleotide Polymorphism at *IDDM2*?

Jana Královičová, Tom R. Gaunt, Santiago Rodriguez, Peter J. Wood, Ian N.M. Day, and Igor Vořechovský

Predisposition to type 1 diabetes and juvenile obesity is influenced by the susceptibility locus *IDDM2* that includes the insulin gene (*INS*). Although the risk conferred by *IDDM2* has been attributed to a minisatellite upstream of *INS*, intragenic variants have not been ruled out. We examined whether *INS* polymorphisms affect pre-mRNA splicing and proinsulin secretion using minigene reporter assays. We show that *IVS1-6A/T* ($-23HphI+/-$) is a key *INS* variant that influences alternative splicing of intron 1 through differential recognition of its 3' splice site. The A allele resulted in an increased production of mature transcripts with a long 5' leader in several cell lines, and the extended mRNAs generated more proinsulin in culture supernatants than natural transcripts. The longer mRNAs were significantly overrepresented among β -cell-expressed sequenced tags containing the A allele as compared with those with T alleles. In addition, we show that a rare insertion/deletion polymorphism *IVS1+5insTTGC* (*IVS-69*), which is exclusively present in Africans, activated a downstream cryptic 5' splice site, extending the 5' leader by 30 bp. These results indicate that $-23HphI$ and *IVS-69* are the most important *INS* variants affecting pre-mRNA splicing and suggest that $-23HphI+/-$ is a common functional single nucleotide polymorphism at *IDDM2*. *Diabetes* 55:260–264, 2006

Type 1 diabetes results from the autoimmune destruction of the insulin-producing pancreatic β -cells. In addition to a major susceptibility locus in the HLA region, termed *IDDM1*, genetic predisposition to this disease is conferred by a locus on chromosome 11, designated *IDDM2* (1). The genetic risk at *IDDM2* has been attributed to the *INS* minisatellite

(2–6), which is composed of a variable number of tandem repeat sequences. However, reanalysis of allelic association data in type 1 diabetes did not rule out intragenic variants, including a single nucleotide polymorphism (SNP) $-23HphI$ (7), which is located in position -6 relative to the 3' splice site of intron 1 (*IVS1-6A/T*). This SNP has been used as a surrogate marker for *INS* genotyping in a large number of studies to infer minisatellite haplotypes (class I/III) in disease susceptibility (2,3,6 and refs. therein). *IVS1-6A/T* is located in the polypyrimidine tract (PPT), a splicing signal of central importance for vertebrate 3' splice site recognition, and in a position exhibiting a great depletion of purine residues in the PPT (8). Since uridine is the preferred PPT nucleotide in pre-mRNA (9), the A allele, which reduces the Shapiro-Senapathy matrix score and other splicing prediction scores for the acceptor splice site of intron 1 (online appendix Table 1, available at <http://diabetes.diabetesjournals.org>), is likely to weaken efficient splicing of this intron. As intron 1 separates noncoding and coding exons, differential utilization of its 3' splice site by the uridine- and adenosine-containing pre-mRNAs would be predicted to result in distinct representation of mature transcripts with short and long 5' untranslated regions (UTRs) with potential effects on translation.

To test whether the *IVS1-6A*→T base change promotes PPT recognition and splicing of intron 1, we transiently transfected the wild-type and mutated *INS* minigenes (Fig. 1A) into several cell lines and examined their splicing pattern. Nucleotide sequencing of the resulting mRNAs identified six alternatively spliced isoforms designated 1–6 (Fig. 1B). Isoforms 2, 4, and 6 have correctly spliced the 3' splice site of intron 2, whereas isoforms 1, 3, and 5 lack the 5' part of exon 3 due to a cryptic 3' splice site activation (Fig. 1A and B), confirming an earlier observation (10). As predicted, minigenes with the *IVS1-6A*→T mutation produced a significantly lower proportion of isoforms lacking exon 2 and isoforms retaining intron 1, which was confirmed by selective amplification of naturally spliced transcripts with primer B upstream of the cryptic 3' splice site (Fig. 1A, C, and D). In an attempt to remove the cryptic 3' splice site activation in exon 3, we truncated both the wild-type and mutated minigenes with primer B and transfected shorter reporter constructs into 293T cells. We found that skipping of exon 2 and retention of intron 1 was

From the Human Genetics Division, University of Southampton School of Medicine, Southampton, U.K.

Address correspondence and reprint requests to Igor Vořechovský, University of Southampton School of Medicine, Human Genetics Division, Duthie Building, MP808, Tremona Road, Southampton SO16 6YD, U.K. E-mail: i.vorechovsky@soton.ac.uk.

Received for publication 17 June 2005 and accepted in revised form 13 October 2005.

Additional information for this article can be found in an online appendix at <http://diabetes.diabetesjournals.org>.

EST, expressed sequence tag; PPT, polypyrimidine tract; SNP, single nucleotide polymorphism; SR, serine/arginine rich; UTR, untranslated region.

© 2006 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.

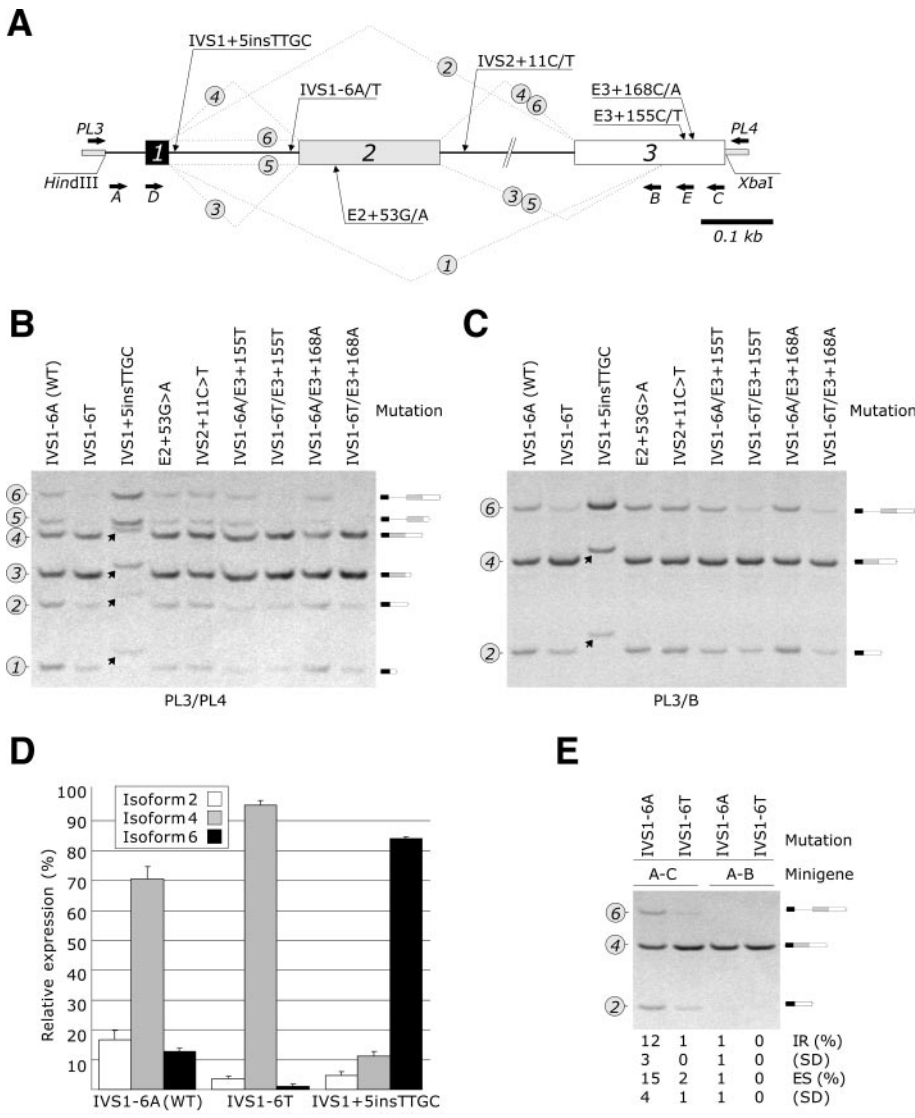


FIG. 1. Insulin variants that affect pre-mRNA splicing. **A:** Schematic representation of the *INS* reporter construct, tested DNA variants, and alternatively spliced mRNA isoforms. Exons 1–3 are shown as black, gray, and white boxes, respectively. Thick lines denote introns, dotted lines represent mRNA isoforms identified by sequencing (numbered in circles). Primers are shown as black arrows, and polylinker sequences are shown as gray rectangles. The designation of DNA variants involves the distance (in nt) from the splice sites: E2+53G/A (previously *INS-70*) (12), IVS1+5insTTGC (*INS-69+/-*), IVS1-6A/T (*INS-27* or *-23HphI+/-*), IVS2+11C/T (*INS-28*), E3+155C/T (*INS-38*), and E3+168C/A (*INS-39*). Scale units are kilobases (kb). Intron 2 (not to scale) truncation is indicated by a slash. **B** and **C:** RNA products detected after transfection of wild-type and mutated *INS* minigenes into 293T cells. Cells were transiently transfected as described (16). RT-PCR was carried out with primers PL3 and PL4 (**B**) or primers PL3 and B (**C**) using total RNA extracted as described (16) and reverse transcribed with oligo-d(T)₁₅. Mutations introduced in *INS* constructs are shown at the top, and PCR primers are shown at the bottom. RNA products are schematically shown on the right side of each panel. Arrows indicate RNA products spliced to the cryptic 5' splice site of intron 1. Nucleotide sequence of the cryptic donor site is shown in Fig. 2A. **D:** Relative expression of correctly spliced mRNA isoforms in 293T cells transfected, with reporter pre-mRNAs indicated at the bottom. Error bars denote SDs obtained from four transfection experiments in duplicates. **E:** Comparison of the splicing pattern generated by truncated (primers A and B) and full-length (primers A and C) minigene constructs (A). Means and SDs of four transfection experiments in duplicate are shown as the percentage of isoforms 6 (intron retention, IR) or 2 (exon skipping, ES) relative to the sum of RNA products spliced to the authentic 3' splice site of intron 2.

reduced for both pre-mRNAs, suggesting that the deleted segment contains splicing regulatory sequences (Fig. 1E).

To test the influence of other intragenic polymorphisms (11,12) on splicing, we examined constructs mutated in all common SNPs and in less prevalent *INS* variants that are located close to the 3' or 5' splice site (Fig. 1A). Interestingly, a 4-nt insertion *IVS-69+* (IVS1+5insTTGC) in the vicinity of the authentic 5' splice site of intron 1 activated a downstream cryptic 5' splice site, extending the 3' end of noncoding exon 1 (Figs. 1B–D and 2A). The insertion allele, which is present in ~25% of Africans but absent in Caucasians (12), introduces a G-to-T mutation in position +5 of the 9-nt 5' splice site consensus, thus reducing the strength of the authentic 5' splice site and increasing relative expression of intron 1–containing transcripts (Fig. 1B–D). Identical 5' splice site alterations in other genes have been previously shown to cause splicing defects that resulted in genetic disease (13), suggesting that this variant may affect proinsulin expression. The remaining SNPs had no major influence on splicing of reporter pre-mRNAs (Fig. 1B and C), although we could not exclude minor effects.

To examine the splicing pattern of *INS* in the natural context, we transfected our reporter constructs into a cell line derived from β -cells (Fig. 2B). We observed a similar

decrease of intron 1–retaining transcripts for the IVS1-6T allele in insulin-producing rat INS-1E cells, as well as an activation of the cryptic 5' splice site of intron 1 for pre-mRNAs carrying the 4-nt insertion. Similar splicing patterns were also observed for pancreatic ductal adenocarcinoma PANC1 cells, HeLa cells, and H1299 cells derived from small-cell lung carcinoma.

The more efficient splicing of T-containing intron 1 observed in both β - and non- β -cells (Figs. 1B–D and 2B) would be predicted to result in a biased representation of *INS* isoforms in gene sequencing databases. We therefore determined distribution of transcripts that lack intron 2 in expressed sequence tag (EST) libraries created from pancreatic β -cells. Of 19 *INS* EST libraries, two contained transcripts with complete retention of intron 1 (Table 1). Analysis of aligned cDNA sequences in both libraries indicated that the HR85 library contained only class I haplotypes (IVS1-6A alleles), whereas the human insulinoma library had only class III haplotypes (IVS1-6T alleles), suggesting that they were prepared from homozygous donors. Comparison of correctly spliced transcripts and transcripts retaining intron 1 showed that the latter mRNAs were significantly overrepresented in the HR85 library as compared with a human insulinoma library (Table 1) ($\chi^2 = 62.2$, 1 df, $P = 3.0 \times 10^{-15}$).

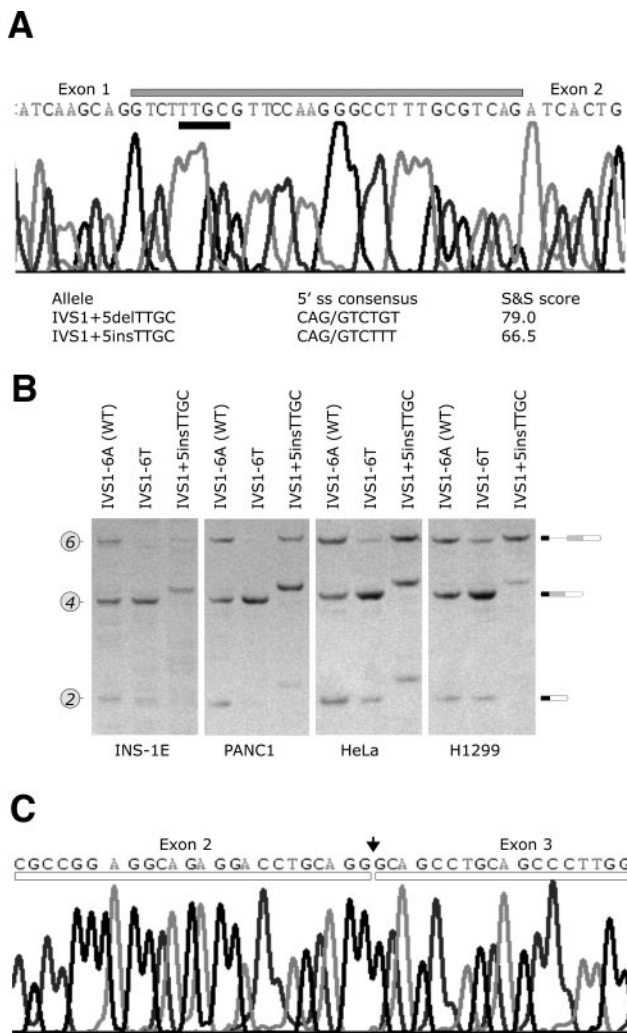


FIG. 2. Activation of the cryptic 3' and 5' splice sites in *INS* pre-mRNAs. **A:** Nucleotide sequence of RNA products spliced to the cryptic 5' splice site of intron 1. The products are indicated by arrows in Fig. 1B and C. The insertion (black bar) allele at the IVS1+5insTTGC polymorphism alters the 5' splice site selection and extends the 5' UTR to include the 5' end of intron 1 (gray bar). The Shapiro and Senapathy (S&S) scores were computed using a splice site analysis tool available at <http://ast.bioinfo.tau.ac.il/>. The S&S score of the cryptic 5' splice site consensus in intron 1 (CAG/GTGGGC) is 78.2. **B:** Exogenous expression of *INS* isoforms in INS-1E cell non-β-cell lines. Mutations are shown at the top, and cell lines are indicated at the bottom. RNA products are shown on the right side and isoform numbers on the left side. **C:** Nucleotide sequence of RNA products spliced to the second cryptic 3' splice site (arrow) in exon 3 located 36 nt downstream of intron 2 and 89 nt upstream of a major cryptic 3' splice site that defines isoforms 1, 3, and 5 (Fig. 1A).

TABLE 1
In silico analysis of *INS* transcripts derived from EST libraries

EST library*	Total number of available ESTs	<i>HphI</i> allele/minisatellite†	Isoform 4 (%)	Isoform 6 (%)
HR85 islet	192	A/class I	47	15 (24)‡
Human insulinoma	3,026	T/class III	825	25 (3)§

*Genbank accession numbers for each database entry and corresponding sequences are available in online appendix Table 2. The proportion of informative ESTs was similar in both libraries (~30%). †*INS* haplotypes and isoform structure were determined using Python scripts (www.python.org) available in online appendix Table 3. ‡Of 15 ESTs, 10 were informative for the IVS1+5insTTGC polymorphism (*INS*-69+/-), and all of them had the deletion allele. §Of 25 ESTs, 10 were informative for the IVS1+5insTTGC polymorphism, and all of them had the insertion allele, suggesting that the insulinoma donor was of African origin.

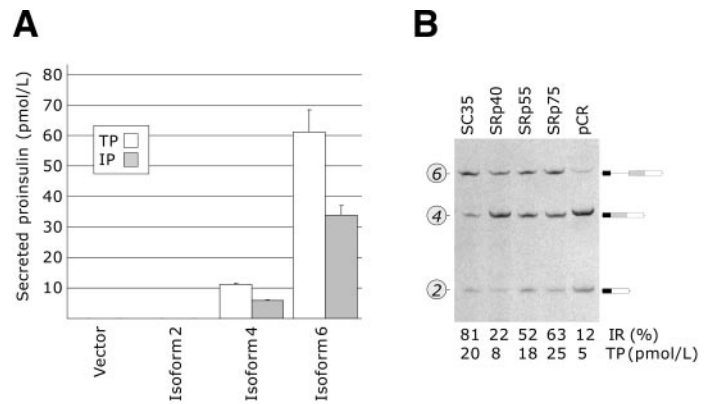


FIG. 3. Functional significance and regulation of *INS* isoforms 4 and 6. **A:** Total (TP) and intact (IP) proinsulin secretion by 293T cells transfected with the indicated *INS* cDNA constructs using solid-phase sandwich fluoroimmunoassays and monoclonal antibodies as described (25). Error bars indicate SDs of duplicate transfection experiments. **B:** Increased representation of intron 1-retaining transcripts in cells expressing a subset of SR proteins. Cotransfections with SR-expressing plasmids were as described (16).

A search for transcripts spliced to the cryptic 3' splice site in exon 3 revealed one such transcript in the HR85 library (0.52%) and nine clones in the insulinoma library (0.29%), indicating that these transcripts are rare in β-cells in vivo. In addition, the insulinoma library contained several transcripts (3/3,026 [0.10%]) spliced to another cryptic 3' splice site of intron 2 (online appendix Table 2). We found such transcripts by sequencing a rare RT-PCR product amplified from a cDNA sample prepared from pancreatic mRNA (Fig. 2C), suggesting that this splice site is occasionally used in normal β-cells in vivo. RNA products spliced to the cryptic 5' splice site in intron 1 (Fig. 2A) were present only in the insulinoma library (974/3,026 [32%]). Of 974 transcripts, 688 were informative at *IVS*-69, and all of them had the TTGC insertion, consistent with cryptic 5' splice site utilization of our mutated reporters (Fig. 1B and C) and the presence of this insertion in the donor's germline. Five of these transcripts (0.72%) were spliced to cryptic 3' splice site E3 + 37 (Fig. 2C), further expanding the complexity of *INS* mRNAs.

Unlike isoform 2, isoforms 4 and 6 encode a functional peptide, but the two differ by the length of the 5' UTR, which often contains sequences that influence translation (14). We therefore cloned cDNAs representing each isoform and measured levels of secreted proinsulin in culture supernatants after their transient transfection into 293T cells. Significantly, equimolar amounts of plasmids containing isoform 6 repeatedly generated higher proinsulin levels in culture supernatants than those with isoform 4 (Fig. 3A). Transfection experiments with shorter reporters

prepared with primers D and E to create exact replicas of previously used minigenes (15) gave a similar difference (data not shown).

Serine/arginine-rich (SR) proteins are a family of RNA-binding proteins that regulate splicing, mRNA export, translation, and other cellular processes. To investigate how SR proteins affect the splicing pattern of *INS* reporter pre-mRNAs, we cotransfected our reporter constructs with plasmids expressing SC35, SRp40, SRp55, and SRp75. Reporter pre-mRNAs coexpressed with SR proteins in 293T cells increased the proportion of transcripts retaining intron 1 (Fig. 3B), and this increase correlated with proinsulin levels measured in culture supernatants of cotransfected cells ($r = 0.90$, $P < 0.05$), further supporting a causal link between the enhanced isoform 6 expression and higher proinsulin secretion.

Our results identified *INS* variants that affect pre-mRNA splicing and illustrated how transcript complexity of a small gene could be significantly influenced by intragenic DNA polymorphism. They also showed marked differences in the proinsulin production generated by these splice variants and suggested that proinsulin production is controlled by a subset of SR proteins. Uridine substitutions in PPT position -6 have been shown to increase exon skipping and retention of weakly spliced introns (16), most likely by diminishing interaction between the PPT and the 65-kDa subunit of the U2 small nuclear RNA auxiliary factor or competing polypyrimidine binding proteins. *INS* intron 1 is weakly spliced in both the rat (17) and the mouse gene, where transcripts retaining this intron have been reported with higher translational efficiency (18). Absence of introns reduced the cytoplasmic level of rat insulin mRNA and protein expression by approximately sixfold (19), which is comparable with our observation (Fig. 3A) and the mouse data (18). Conversely, extension of the rat 5' UTR by inserting unstructured oligonucleotides resulted in repeat-dependent enhancement of translation (14). Chicken embryonic proinsulin is also regulated through an extended 5' UTR and upstream open reading frames (20). It will therefore be interesting to examine translation efficiency of mutated constructs to identify underlying *cis*-acting elements in the first introns. Assuming a direct action of coexpressed SR proteins, the observed reduction of canonical transcripts in cotransfected cells (Fig. 3B) could be explained by the promotion of proximal splicing and/or reduction of distal splicing (16 and refs. therein) or altered mRNA export of intron 1-retaining transcripts.

The existence of functional *INS* variants that are more frequent or unique in Africans raises the hypothesis that they may have been subject to selection following a shift of out-of-Africa ancestral population to primitive agriculture with consequent increased carbohydrate intake. This could be the case for both the IVS1-6T allele, which has a frequency of ~82% in Africans and ~19% in non-Africans, and the insertion allele at IVS-69, with a frequency of ~25 and 0%, respectively (12). Utilization of the cryptic 5' splice site in intron 1 was found to be increased in human insulinomas, but the ethnic origin of donors was not provided and primers amplifying the extended mRNA were designed to amplify the deletion not the insertion allele (21). Both our EST data and transfection experiments (Figs. 1B-D and 2B) indicate that utilization of the cryptic 5' splice site is allele dependent. Future studies should therefore confirm that use of this splice site in insulinomas is indeed increased both in IVS-69- and

IVS-69+ individuals and that β -cells of the IVS-69+ donors produce higher amounts of transcripts spliced to the cryptic 5' splice site of intron 1.

Marked differences in the proinsulin secretion observed for isoforms 4 and 6 in humans (Fig. 3A) and mice (18) well illustrate the power of alternative pre-mRNA splicing to regulate gene expression. One can speculate that even minute differences in the ratios of these isoforms in pancreatic β -cells may account for a small but significant increase in fasting insulin levels and more rapid weight gain in late childhood and adolescence observed in carriers of the -6A allele (6). Since previous studies correlating *INS* genotypes and steady-state mRNA expression used primers in exon 3/3' UTR (3,4) or bridging intron 1 (5), which could not measure the relative ratios of *INS* isoforms, future work should attempt to determine the expression of alternatively spliced mRNAs in β -cells of AA, AT, and TT carriers.

Finally, the allele-dependent alternative splicing described here may affect "promiscuous" intrathymic selection processes and tolerance to the central antigenic determinant of diabetic autoimmunity. Such a hypothesis is supported by the recent identification of oligoclonally expanded T-cells from diabetic subjects with DR4, a susceptibility molecule for type 1 diabetes. These cells have been shown to recognize the DR4-restricted insulin epitope A1-15 (22), which is encoded by the 5' part of exon 3, corresponding exactly to a segment absent in transcripts spliced to the downstream cryptic 3' splice site (Fig. 1A). Smaller transcripts that may represent mRNAs lacking this segment were seen in developing thymi (Fig. 3 in ref. 23), although this finding was not discussed by the authors. Pre-mRNAs spliced to this acceptor site observed in our reporter system (Fig. 1B) are most likely to be preferentially degraded *in vivo* by "nonstop" mRNA surveillance mechanisms as these transcripts lack termination codons. The existence of noncanonical transcripts in transfected cells (Figs. 1B and 2A and C) and in EST libraries (online appendix Table 2) would be consistent with this concept, which appears to be further supported by the observed genetic linkage and allelic association of *INS* with type 1 diabetes (1) but not type 2 diabetes or other metabolic traits (24), except in childhood obesity (6). Subsequent studies should therefore examine each of these possibilities in greater detail, since alternative splicing and associated pathways might become suitable targets for therapeutic approaches in the future.

RESEARCH DESIGN AND METHODS

Splicing reporter constructs. The wild-type *INS* reporters (Fig. 1A) were cloned into the *HindIII/XbaI* restriction sites of the mammalian expression vector pCR3.1 (Invitrogen) using primers A (5' ATC TAA GCT TGG GAG ATG GGC TCT GAG ACT A) and C (GTC ATC TAG ATG GTT CAA GGG CTT TAT TCC A). Minigenes lacking the 3' part of exon 3 to avoid the cryptic 3' splice site were subcloned by amplifying the full-length clones with primer B (ATA ATC TAG ACA CAA TGC CAC GCT TCT GC). In addition, we cloned splicing reporters using primers D (ACC AAG CTT AGC CCT CCA GGA CAG GCT) and E (ACC TCT AGA GGC TGC GTC TAG TTT CAG TA) to obtain minigenes with shorter UTRs as described (15). Mutated minigene constructs were prepared by overlap-extension PCR using *Pfu* (Promega) or *Pwo* (Roche) polymerases. Mutagenic oligonucleotide primers are available on request. The wild-type and mutated reporters were fully sequenced as described (16) to confirm intended mutations and exclude clones with undesired alterations. Plasmids expressing SR proteins were obtained as described (16).

Cell lines and transfections. The human embryonic kidney 293T, HeLa, PANC1, and H1299 cells were grown under standard conditions in RPMI-1640 supplemented with 10% (vol/vol) FCS (Gibco) as described (16). INS-1E cells were grown in RPMI-1640 supplemented with 10% FCS, 10 mmol/l Hepes, 50

$\mu\text{mol/l}$ 2-mercaptoethanol, and 1 mmol/l sodium pyruvate. Transient transfections were performed in 12-well plates or Falcon T25 using FuGENE 6 (Roche) or GeneCarrier-1 (Epoch-Biolabs). The plating density was 10^5 cells $17\text{--}24$ h before transfection. The medium was changed 2 h before adding a DNA mixture prepared by combining 2–6 μl transfection reagent and 50 μl serum-free medium, followed by the addition of 500 ng purified plasmid DNA (Wizard Plus SV Minipreps; Promega). The DNA mixture was incubated for 20 min at room temperature before transfection. Cells were harvested 48 h post-transfection.

Detection of spliced products. Total RNA was extracted as described (16), treated with DNase I (Ambion), and reverse transcribed using oligo(dT)₁₅ primers and Moloney murine virus reverse transcriptase (Promega) according to the manufacturer's recommendations. Three microliters of the first-strand cDNA reaction were amplified with vector-specific primers PL3 (GGG AGA CCC AAG CTG GCTA) and PL4 (AGTC GAG GCT GAT CAG CGG) and primers PL3/B (Fig. 1) to validate the ratios of RNA products in independent PCRs. The number of PCR cycles was 29 or lower to maintain an approximately linear relationship between the RNA input and signal. DNA fragments were extracted from the gel using the BIO101 GeneClean Kit (Q-BIOgene) and sequenced to confirm the identity of each fragment. RNA products were measured with FluorImager 595 using ImageQuant and Phoretix software (Nonlinear Dynamics, Sunnyvale, CA) as described (16).

Quantification of secreted proinsulin. Total and intact proinsulin levels were measured in culture supernatants (48 h post-transfection) by dissociation-enhanced lanthanide fluoroimmunoassay using monoclonal antibodies 3B1 (capture), CPT3F11 (total), and A6 (intact) as described (25).

Analysis of *INS* EST libraries. *INS* cDNA sequences (online appendix Table 2) were obtained from EST libraries available at from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/dbEST/>). The statistical significance of their distribution was estimated by a standard contingency-table χ^2 test. The Python scripts to analyze *INS* isoforms from EST libraries and their haplotypes are available in online appendix Table 3. Unique cDNA clones were aligned with a contig assembly program (CAP) available at <http://bio.ifom-firc.it/ASSEMBLY/assemble.html>, and haplotypes/isoforms were re-examined manually.

ACKNOWLEDGMENTS

This work was supported by the University of Southampton, the Wessex Medical Trust, and the Medical Research Council of the U.K. T.R.G. is a British Heart Foundation Intermediate Fellow (FS/05/065/19497). This work is subject to U.K. patent application 0510555.6.

We thank Professor C. Wollheim, Dr. P. Maechler (University of Geneva), and Dr. J. Blaydes (University of Southampton) for providing cell lines; Dr. J. Caceres (University of Edinburgh) and Dr. G. Sreaton (University of Oxford) for providing plasmids expressing SR proteins; and Ms. C. Glenn and D. Smith for technical help with dissociation-enhanced lanthanide fluoroimmunoassay.

REFERENCES

- Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM, et al.: A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371:130–136, 1994
- Kennedy GC, German MS, Rutter WJ: The minisatellite in the diabetes susceptibility locus *IDDM2* regulates insulin transcription. *Nat Genet* 9:293–298, 1995
- 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
- 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
- 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
- Le Stunff C, Fallin D, Schork NJ, Bougeres P: The insulin gene VNTR is associated with fasting insulin levels and development of juvenile obesity. *Nat Genet* 26:444–446, 2000 [erratum in 28:97, 2001]
- 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-Tirgoviste 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
- Yeo G, Hoon S, Venkatesh B, Burge CB: Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc Natl Acad Sci U S A* 101:15000–15005, 2004
- Roscigno RF, Weiner M, Garcia-Blanco MA: A mutational analysis of the polypyrimidine tract of introns: effects of sequence differences in pyrimidine tracts on splicing. *J Biol Chem* 268:11222–11229, 1993
- Laub O, Rutter WJ: Expression of the human insulin gene and cDNA in a heterologous mammalian system. *J Biol Chem* 258:6043–6050, 1983
- Stead JD, Jeffreys AJ: Allele diversity and germline mutation at the insulin minisatellite. *Hum Mol Genet* 9:713–723, 2000
- Stead JD, Hurler ME, Jeffreys AJ: Global haplotype diversity in the human insulin gene region. *Genome Res* 13:2101–2111, 2003
- Bardelli T, Donati MA, Gasperini S, Ciani F, Belli F, Blau N, Morrone A, Zammarchi E: Two novel genetic lesions and a common BH4-responsive mutation of the PAH gene in Italian patients with hyperphenylalaninemia. *Mol Genet Metab* 77:260–266, 2002
- Kozak M: Leader length and secondary structure modulate mRNA function under conditions of stress. *Mol Cell Biol* 8:2737–2744, 1988
- Shalev A, Blair PJ, Hoffmann SC, Hirschberg B, Peculis BA, Harlan DM: A proinsulin gene splice variant with increased translation efficiency is expressed in human pancreatic islets. *Endocrinology* 143:2541–2547, 2002
- Královičová J, Hounginou-Molango S, Krämer A, Vofechovský I: Branch sites haplotypes that control alternative splicing. *Hum Mol Genet* 13:3189–3202, 2004
- Wang J, Shen L, Najafi H, Kolberg J, Matschinsky FM, Urdea M, German M: Regulation of insulin preRNA splicing by glucose. *Proc Natl Acad Sci U S A* 94:4360–4365, 1997
- Minn AH, Lan H, Rabaglia ME, Harlan DM, Peculis BA, Attie AD, Shalev A: Increased insulin translation from an insulin splice-variant overexpressed in diabetes, obesity, and insulin resistance. *Mol Endocrinol* 19:794–803, 2005
- Lu S, Cullen BR: Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* 9:618–630, 2003
- Hernández-Sánchez C, Mansilla A, de la Rosa EJ, Pollerberg GE, Martínez-Salas E, de Pablo F: Upstream AUGs in embryonic proinsulin mRNA control its low translation level. *EMBO J* 22:5582–5592, 2003
- 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
- Kent SC, Chen Y, Bregoli L, Clemmings SM, Kenyon NS, Ricordi C, Hering BJ, Hafler DA: Expanded T cells from pancreatic lymph nodes of type 1 diabetic subjects recognize an insulin epitope. *Nature* 435:224–228, 2005
- Sospedra M, Ferrer-Francesch X, Dominguez O, Juan M, Foz-Sala M, Pujol-Borrell R: Transcription of a broad range of self-antigens in human thymus suggests a role for central mechanisms in tolerance toward peripheral antigens. *J Immunol* 161:5918–5929, 1998
- Hansen SK, Gjesing AP, Rasmussen SK, Glumer C, Urhammer SA, Andersen G, Rose CS, Drivsholm T, Torekov SK, Jensen DP, Ekstrom CT, Borch-Johnsen K, Jorgensen T, McCarthy MI, Hansen T, Pedersen O: Large-scale studies of the *HphI* insulin gene variable-number-of-tandem-repeats polymorphism in relation to type 2 diabetes mellitus and insulin release. *Diabetologia* 47:1079–1087, 2004
- Sobey WJ, Beer SF, Carrington CA, Clark PM, Frank BH, Gray IP, Luzio SD, Owens DR, Schneider AE, Siddle K, et al.: Sensitive and specific two-site immunoradiometric assays for human insulin, proinsulin, 65–66 split and 32–33 split proinsulins. *Biochem J* 260:535–541, 1989