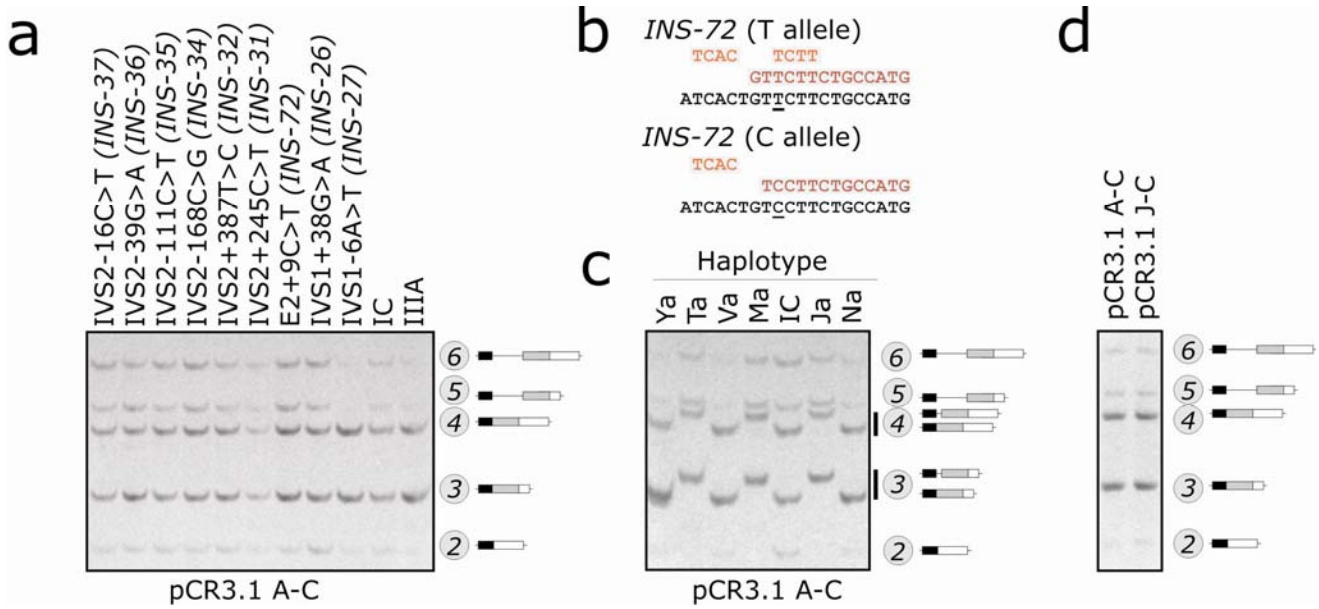
SUPPLEMENTAL DATA

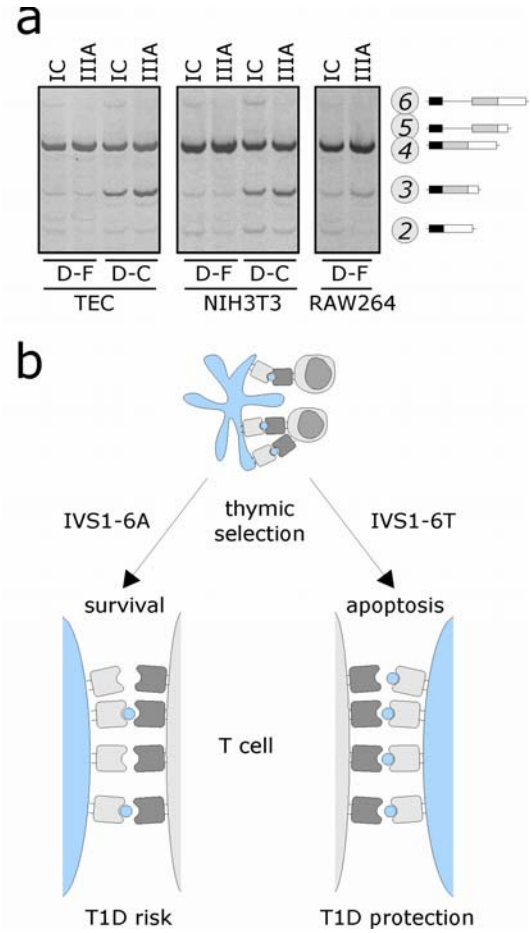
SUPPLEMENTAL FIGURES

SUPPLEMENTAL FIGURE 1 Dissection of haplotype-dependent *INS* pre-mRNA splicing



Legend: *(A)* Splicing pattern of reporter constructs mutated at naturally occurring *INS* variants. Mutations are denoted by the traditional nomenclature (http://www.hgvs.org/mutnomen/). The original designation (Stead et al. 2003) of the variant is followed in parentheses. RNA products are shown schematically to the right; template reporters for overlap-extension PCR are at the bottom. *(B)* Predicted auxiliary splicing elements at *INS-72*. Putative splicing regulatory motifs are shown above the sequence; variant residues at *INS-72* are underlined. *(C)* Splicing pattern of African-specific haplotypes with (Ta, Ma, Ja) or without (Ya, Va, Na) the 4-nt insertion allele at *INS-69*. The IC haplotype is shown as a control. Haplotype designation was as previously described (Stead et al. 2003). *(D)* Splicing of reporter constructs with (primers J-C) and without (primers A-C) upstream minisatellite (VNTR) sequences is identical. Amplifications shown for panels *B* and *C* was with primers PL3 (Kralovicova et al. 2006a) and E. In experiment shown in panel *D* we employed PCR primers D and PL4 (Kralovicova et al. 2006a) (Figure 1).

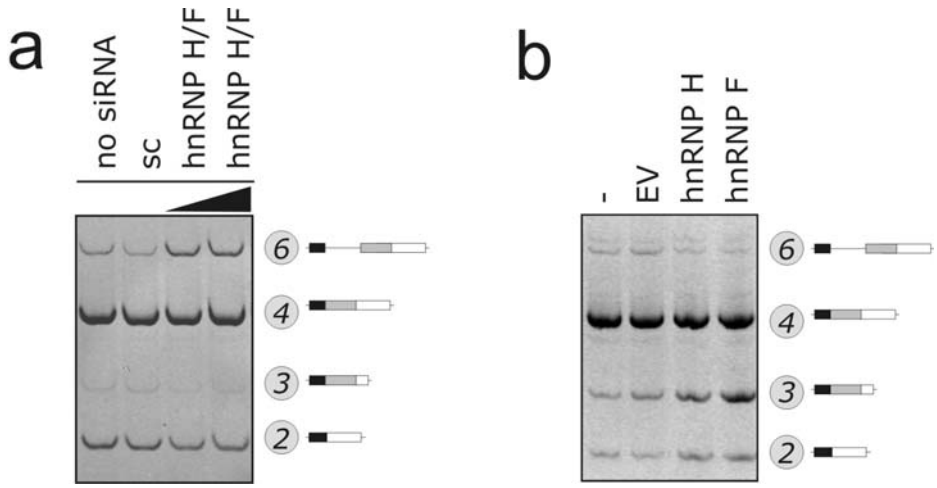
SUPPLEMENTAL FIGURE 2 Allele-specific *INS* splicing in mouse cell lines



Legend: *(A)* Increased IR of the *INS* class IC haplotype as compared to the class IIIA haplotype in the *Aire*-expressing TEC cell line derived from thymic epithelial cells and in the macrophage cell line RAW264; NIH3T3 cells are shown as controls. The *INS* splicing pattern was not significantly different in a TEC cell line lacking *Aire* (data not shown). Haplotypes are at the top of each panel and mRNA isoforms to the right. Cloning primers are shown at the bottom; their location is depicted in Figure 1A. The extended 3' segment of D-F reporters inhibits the cryptic 3' splice site in *INS* exon 3 (cr3'ss+126), which is most likely mediated by the 3'end processing machinery that can influence selection of 3'ss of the last intron (Rigo and Martinson 2008). *(B)* A proposed model for genetic susceptibility to T1D at *IDDM2*. Unlike *INS-69*, which is confined to Africans, the *INS-27*A allele (IVS1-6A) has a high frequency in whites (Stead et al. 2003), indicating that this variant is critical for the haplotype-dependent proinsulin expression in this population. Antigen presenting cells will have a lower number of the MHC/proinsulin peptide complexes in carriers of the low-expressing and T1D-predisposing A alleles than in carriers of the T alleles. Their lower number would be associated with the lower probability of inducing apoptosis of autoreactive T cells in the thymus during negative selection. Antigen-presenting cells/antigenic peptides are in blue. MHC molecules and T-cell receptors are highlighted in light and dark gray, respectively.

SUPPLEMENTAL FIGURE 3 Heterogenous nuclear RIBONUCLEOPROTEINS F/H in *INS* intron 1 splicing

Legend: *(A)* Increased intron 1 retention and decreased utilization of cr3'ss+126 in hnRNP F/H-depleted cells. siRNAs are at the top, RNA products to the right; sc, scrambled control. The final concentration of siRNAs in culture media was 80 (sc), 60 and 100 nM (hnRNP H/F). The D-F constructs used for this experiment contained the class IC haplotype. *(B)* Improved efficiency of intron 1 removal and cr3'ss+126 activation in cells overexpressing hnRNP H (1 μg of pcDNA-hnRNP-H) and hnRNP F (1 μg of pCI-hnRNP-F). EV, empty vector.



SUPPLEMENTAL FIGURE 4 Sequence alignment of the 5' untranslated region of the proinsulin gene in primates

Legend: Intronic sequences are in lower case, exonic sequences are in upper case. Lineage-specific changes observed in at least two species/animals are highlighted in gray. Naturally occurring human polymorphisms are in red. Alignment was carried out using the Clustal-Wallis algorithm (v. 2.0). Hs, *Homo sapiens* (Stead et al. 2003)(Genbank accession numbers AY138590 and L15440); Pt1-2, *Pan troglodytes* (Stead et al. 2003) (this study); Gg, *Gorilla gorilla* (Stead et al. 2003)(AY137500); Pp1-2, *Pongo pygmeus* and *abelii* (AY137503 and AC199962); Hg1, Hg2, *Hylobates gabriellae* (this study); Eryp, *Erythrocebus patas* (this study); Cae, *Chlorocebus aethiops* (X61092); Cc116, *Cercopithecus campbelli* (this study); Cd2A, *Cercopithecus diana* (this study); Ca303, *Cercopithecus ascanius* (this study); Mt114, *Macaca tonkeana* (this study); Mf, *Macaca fuscata* (this study); Mm1,Mm2, *Macaca mulatta* (Indian and Chinese, respectively; this study); Ms, *Macaca sylvanus* (this study); Ph100, *Papio hamydras* (this study); Msp119, Msp102, *Mandrillus sphinx* (this study); Cang, *Colobus angolensis* (this study); Ac1af, *Colobus guereza* (this study); Se117, Se2, *Semnopithecus entellus* (this study); Pn, *Pygathrix nemaeus* (this study); To, *Trachypithecus obscurus* (this study); Ta, *Trachypithecus auratus* (this study); Sv, *Semnopithecus vetulus* (this study); Pm, *Presbytis melalophos* (this study); Nl, *Nasalis larvatus* (this study); Cj, *Callithrix jacchus* (this study); So1, *Saguinus oedipus* (this study); Smid92; *Saguinus midas* (this study); Sfl15, *Saguinus fuscicollis* (this study); Calgo6, *Callimico goeldii* (this study); Aab203, *Aotus azarae* (this study); At, *Aotus trivirgatus* (J02989); Mur, *Microcebus murinus* (Ensembl scaffold sequence, www.ensembl.org). Y, C or T; W, A or T; N, C or G or T or A. Genbank accession numbers of new sequences: GU901169-GU901198.

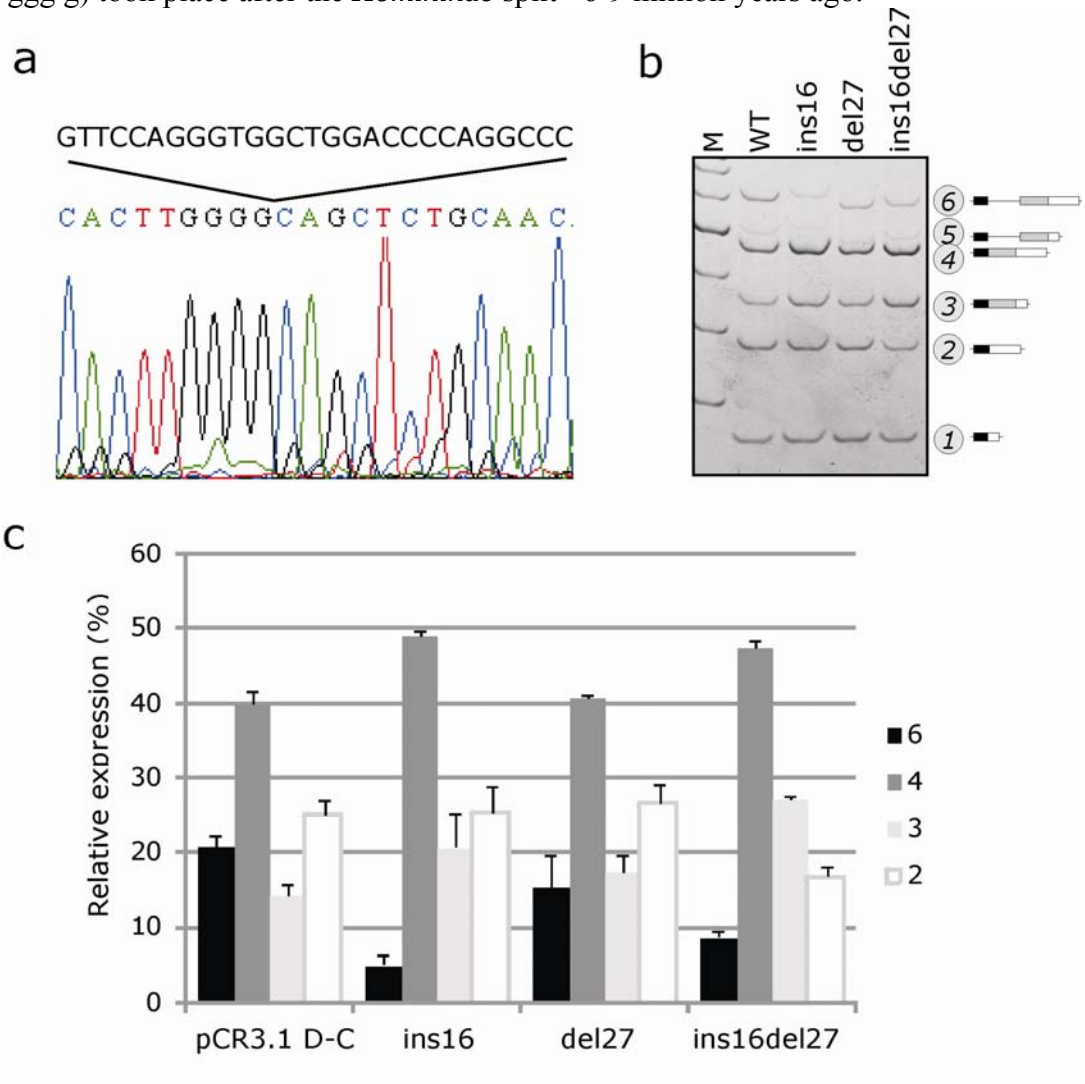
```
                                                                    G0
Hs              CAGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGgtcttgcgttccaagg 60
Pt1             CAGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGgtct----gttccaagg 56
Pt2             CAGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGgtct----gttccaagg 56
Gg              CAGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGgtct----gttccaagg 56
Pp1             CAGCCCTCCGGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGgtct----gttccaagg 56
Pp2             CAGCCCTCCGGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGgtct----gttccaagg 56
Hg1             CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Hg2             CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Eryp            CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Cae             CAACCCTC--GGAC-G-CTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 52
Cc116           CAACCCTCCGGGAC-GGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 55
Cd2A            CAACCNTCCGGGAC-GGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 55
Ca303           CACCCCTCCAGGAC-GGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 55
Mt114           CAACCCTCCGGGACAGGCTGCNTCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Mf              CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Mm1             CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Mm2             CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Ms              CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Ph100           CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Msp119          CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Msp102          CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Cang            CAACCCTCCGGGACAGGCNGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Ac1af           CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Se117           CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Se2             CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Pn              CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
To              CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Ta              CAACCCTCCAGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Sv              CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Pm              CAACCCTCCAGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Nl              CAACCCTCCGGGACAGGCTGCATCAGAAGAGGCCAGCAAGCAGgtct----gttccaagg 56
Cj              CAGCCCTCCGGGACCAGCTGCATCGCAGGAGACCAGCCAGCAGgtgt----gttccaagg 56
So1             CAGCCCTCCGGGACCAGCCGCATCACAGGAGACCAGCCAGCAGgtct----gttccaagg 56
Smid92          CAGCCCTCCGGGACCAGCCGCATCACAGGAGACCAGCCAGCAGgtct----gttccaagg 56
Sfl15           CAGCCCTCCGGGACCAGCCGCATCACAGGAGACCAGCCAGCAGgtct----gttgcaagg 56
Calgo6          CAGCCCTCCGGGACCAGCTGCATCACA---GACCAGCCAGCAGgtct----gttccaagg 53
Aab203          CAGCCCTCCGGGACCAGCTGCATCACAGGAGGCCAGCGAGCAGgtct----gttccaagg 56
At              CTCACGCCCGGGACCAGCTGCATCACAGGAGGCCAGCGAGCAGgtct----gttccaagg 56
Mur             CAGCCCTGC-GGACCAGCTGTTTCCCCGGCCGTCAGCGAGCAGgtct----gtgccaggg 55
```

```
                          G1                                 G2
Hs            gcctttgcgtcaggtggg--ctcaggr----------------ttccagggtggctggac 102
Pt1           gcctttgcgtcaggtggg--ctcaggg----------------ttccagggtggctggac 98
Pt2           gcctttgcgtcaggtggg--ctcaggg----------------ttccagggtggctggac 98
Gg            gcctttgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Pp1           gcctttgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Pp2           gcctttgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Hg1           gccttcacgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Hg2           gccttcacgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Eryp          gccttcacgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Cae           gccttcgcgtcaggtggg--ctcagggctgccc-acttgggggttccagggtggctggac 109
Cc116         gccttcacgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 113
Cd2A          gccttcgcgtcaggtggg--ctcrgggcygccccacttgggggtcccagggtggctggac 113
Ca303         gccttcgcgtcaggtggg--ctcagggctgccccacttgggggtcccagggtggctggac 113
Mt114         gccttcacgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Mf            gccttcgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Mm1           gccttcgcgttaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Mm2           gccttcgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Ms            gccttcgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Ph100         gccttcgngtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Msp119        gccttcgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Msp102        gccttcgcgtcaggtggg--ctcagggctgccccacttgggggttccagggtggctggac 114
Cang          gccttcgcgtcaggtggg--ntcagggctgcccgacttgggggttccagggtggctggac 114
Ac1af         gccttcgcgtcaggtggg--ctcagggctgcccgacttgggggttccagggtggctggac 114
Se117         gccttcgcgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
Se2           gccttcgcgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
Pn            gccttcgtgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
To            gccttcgcgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
Ta            gccttcgcgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
Sv            gccttcgcgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
Pm            gccttcgcgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
Nl            gccttcgcgtcaggtggg--ctcagggctgccccacttgggg----------------- 96
Cj            gcctttgagccagcctggg-cccagggctgccccacttgggggttgcagagcagttggac 115
So1           gcctttgagccagcctgga-cccagggctgccccacttgggggtcccagagcagttggac 115
Smid92        gcctttgagccagcctgga-cccagggctgccccacttgggggtcccagagcagttggac 115
Sfl15         gcctttgagccagcctggg-cccagggctgccccacttgggggtcccagagcagttggac 115
Calgo6        gcctttgagccagcctggg-cccagggctgccccacttgggggttccagagcagttggac 112
Aab203        gccttcgagccagtctggg-cccagggctgccccactcgggggttccagagcagttggac 115
At            gccttcgagccagtctgggccccagggctgccccactcgggg-ttccagagcagttggac 115
Mur           gcctccggtccgggcggt-----gggacc------------------------cgggac 85

                                 G3    uORF1     G4         uORF2  G5
Hs            cccaggccc-cagctctgcagcagggagg-acgtgg-ctgggctcgtgaagcatgtgggg 159
Pt1           cccaggccc-cagctctgcagcagggagg-acgtgg-ctgggctcttgaagcatgtgggg 155
Pt2           cccaggccc-cagctctgcagcagggagg-acgtgg-ctgggctcttgaagcatgtgggg 155
Gg            cccaggccc-cagctctgcagcagggagg-acgtgg-ctgggctcttgaagcatgtgggg 171
Pp1           cccaggctc-cagctctgcagctgggagg-acgtgg-ctgggctcttgaagcatttgggg 171
Pp2           cccaggctc-cagctctgcagctgggagg-acgtgg-ctgggctcttgaagcatttgggg 171
Hg1           cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Hg2           cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Eryp          cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Cae           cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgagg 166
Cc116         cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgagg 170
Cd2A          cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgagg 170
Ca303         cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgagg 170
Mt114         cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Mf            cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Mm1           cccaggccc-cagctctgcaacagggagg-acgtgg-ctgggctcttgaagcgtttgggg 171
Mm2           cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Ms            cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtctgggg 171
Ph100         cccaggccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Msp119        cccaggccc-cagctctgcaacaggggag-acatgg-ctgggctcttaaagcgtttgggg 171
Msp102        cccaggccc-cagctctgcaacaggggag-acatgg-ctgggctcttaaagcgtttgggg 171
Cang          cccagactc-cagctctgcaacagggagg-acatgg-ntgggctcttgaagcgtttgggg 171
Ac1af         cccagaccc-cagctctgcaacagggagg-acatgg-ctgggctcttgaagcgtttgggg 171
Se117         ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcgtttgggg 144
Se2           ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcgtttgggg 144
Pn            ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcgtttgggg 144
To            ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcgtttgggg 144
Ta            ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcgtttgggg 144
Sv            ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcgtttgggg 144
Pm            ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcgtttgggg 144
Nl            ----------cagctctgcaacagggaga-acatgg-ctgggctcttgaagcatttgggg 144
Cj            cccaggccc-cagctctgcagtgggaggg--catgg-ctgggctc-tgaagcattt--gg 168
So1           cccaggccc-cagctctgcagtgggaggg--catgg-ctgggctc-tgaagcattt--gg 168
Smid92        cccaggccc-cagctctgcagtgggaggg--catgg-ctgggctc-tgaagcattt--gg 168
Sfl15         cccaggccc-cagctctgcagtgggaggg--catgg-ctgggctc-tgaagcattt--gg 168
Calgo6        cccaggccc-cagctctgcagtgggaggg--catgg-ctgggctc-tgaagcattt--gg 165
Aab203        cc--------caggcctgcagcgggaggg--tgtgg-ctgggctc-tgaagcattt--gg 161
At            cc--------caggtct-cagcgggaggg--tgtgg-ctgggctc-tgaagcattt--gg 160
Mur           cccagctctgcatggtggtggtggggagggacgtgggctcctctcgtggggcatttgggg 145
```

```
                      G6          G7
Hs          gtgagcccaggggccccaaggcagggcacctggccttcagcctgcctcagccctgcctgt 219
Pt1         gtgagcccaggggccccaaggcagggcacctggccttcagccggcctcagccctgcctgt 215
Pt2         gtgagcccaggggccccaaggcagggcanctg-ccttcagccggcctcagccctgcctgt 214
Gg          gtgagcccaggggccccaaggcagggcaactggccttcagccggcctcagccctgcctgt 231
Pp1         gtgagcccaggggccccagggcagggcacctggccttcagccgacctcagctctgcctgt 231
Pp2         gtgagcccaggggccccagggcagggcacctggccttcagccgacctcagctctgcctgt 231
Hg1         gtgagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Hg2         gtgagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Eryp        gtgagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Cae         gtgaacccaggggccc-agggcagg-cacctggcct-cagctggcctcaggc-tgcctgt 222
Cc116       gtgaacccaggggccccagggcagggcacctggccttcagctggcctcaggcctgcctgt 230
Cd2A        gtgaacccaggggccccagggcagggcacctggccttcagctggcctcaggcctgcctgt 230
Ca303       gtgaaccctggggccccagggcagggcacctggccttcagctggcctcaggcctgcctgt 230
Mt114       gtgagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Mf          gtaagcccaggggccccagggcagggcanctggccttcagccggcctcaggcctgcctgt 231
Mm1         gtaagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Mm2         gtaagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Ms          gtgagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Ph100       gtgagcccaggggccccagggcagggcacctggccttcagccggcctcaggcctgcctgt 231
Msp119      gtgagcccaggggccccagggcagggcacctggcctccagccggcctcaggcctgcctgt 231
Msp102      gtgagcccaggggccccagggcagggcacctggcctccagccggcctcaggcctgcctgt 231
Cang        gtgagcccagggaccccaggccagggcacctggccttcagccggcctcaggcctgcctgt 231
Ac1af       gtgagcccagggaccccaggccagggcacctggccttcagccggcctcaggcctgcctgt 231
Se117       gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
Se2         gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
Pn          gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
To          gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
Ta          gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
Sv          gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
Pm          gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
Nl          gtgagcccagggaccccagggcagggcacctggccttcagccagcctcaggcctgcctgt 204
Cj          gtgagcccaggggccccagggcagggcacctgaccttcagccggcctcagccctgcctgt 228
So1         gtgagcccaggggccccagggcagggcacctggccttcagccggcctcagccctgcctgt 228
Smid92      gtgagcccaggggccccagggcagggcacctggccttcagccgacctcagccctgcctgt 228
Sfl15       gtgagcccaggggccccagggcagggcacctggccttcagccggcctcagccctgcctgt 228
Calgo6      gtgagcccaggggccccagggcagggcacctggccttcagccggcctcagccccgcctgt 225
Aab203      gtgagcccaggggctccagggcagggcacctggccttcagccggcctcagccctgcctgt 221
At          gtgagcccaggggctc-agggcagggcacctg-ccttcagc-ggcctcagcc-tgcctgt 216
Mur         gagcaagcgggggtcccggggcagggcgcccgcccac--gctggcctcagccccgctcct 203


Hs          cwcccagATCACTGTYCTT---CTGCCATG 246
Pt1         ctcccagATCACTGTCCTT---CTGCCATG 242
Pt2         ctcccagATCACTGTCCTT---CTGCCATG 241
Gg          ctcccagATCACTGTCCTT---CTGCCATG 258
Pp1         ctcccagATCACTGTCCTT---CTGCCATG 258
Pp2         ctcccagATCACTGTCCTT---CTGCCATG 258
Hg1         ctcccagGTCACTGTCCTT---CCGCCATG 258
Hg2         ctcccagGTCACTGTCCTT---CCGCCATG 258
Eryp        ctcccagGTCACTGTCCTT---CCGCCATG 258
Cae         ctctcagGTCACTGTCCTT---CCGCCATG 249
Cc116       ctcccagGTCACTGTCCTT---CCGCCATG 257
Cd2A        ctcccagGTCACTGTCCTT---CCGCCATG 257
Ca303       ctcccagGTCACTGTCCTT---CCGCCATG 257
Mt114       ctcccagGTCACTGTCCTT---CCGCCATG 258
Mf          ctcccagGTCACTGTCCTT---CCGCCATG 258
Mm1         ctcccagGTCACTGTCCTT---CCGCCATG 258
Mm2         ctcccagGTCACTGTC-TT---CCGCCATG 257
Ms          ctcccagGTCACTGTCCTT---CTGCCATG 258
Ph100       ctcccagGTCACTGTCC------TGCCATG 255
Msp119      ctcccagGTCACTGTCCTT---CCGCCATG 258
Msp102      ctcccagGTCACTGTCCTT---CCGCCATG 258
Cang        ctcccagGTCACTGTCCTT---CCGCCATG 258
Ac1af       ctcccagGTCACTGTCCTT---CCGCCATG 258
Se117       ctcccagGTCACTGTCCTT---CCGCCATG 231
Se2         ctcccagGTCACTGTCCTT---CCGCCATG 231
Pn          ctcccagGTCACTGTCCTT---CCGCCATG 231
To          ctcccagGTCACTGTCCTT---CCGCCATG 231
Ta          ctcccagGTCACTGTCCTT---CTGCCATG 231
Sv          ctcccagGTCACTGTCCTT---CCGCCATG 231
Pm          ctcccagGTCACTGTCCTT---CTGCCATG 231
Nl          ctcccagGTCACTGTCCTT---CCGCCATG 231
Cj          ctcccagGCCGCTGTGCTT---CCACCATG 255
So1         ctcccagGTCGCTGTCCTT---CCACCATG 255
Smid92      ctcccagGTCGCTGTCCTT---CCACCATG 255
Sfl15       ctcccagGTCGCTGTCCTT---CCACCATG 255
Calgo6      ctcccagGTCGCTGTCCTT---CCACCATG 252
Aab203      ctcccagGTCGCTGTCCTT---CCACCATG 248
At          ctcccagGTCTCTGTCCTT---CCACCATG 243
Mur         ctcccagGTCTTTGCCCCCGCCCCGCCATG 233
```

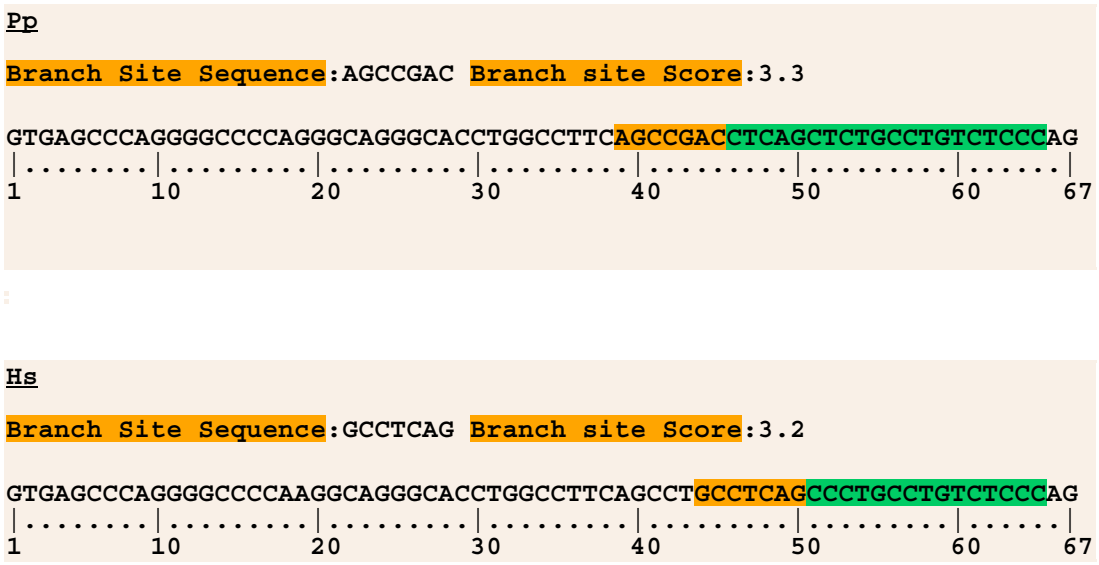
SUPPLEMENTAL FIGURE 5 Ancestral deletions of *INS* intron 1 in primates and their effect on splicing

Legend: *(A)* A colobine-specific, 27-nucleotide deletion in *INS* intron 1. The deleted sequence is shown at the top. *(B)* Splicing pattern of human reporter constructs with two large deletions that took place during primate evolution. 293T cells were transfected with pCR3.1 containing *INS* segments defined by primers D and C, as shown in Figure 1A. Deletions are at the top and RNA products to the right; M, 100-nt size marker. Total RNA was isolated 48 hours post-transfection, reverse transcribed using oligo-d$(T)_{15}$ and complementary DNA products were visualized using PCR with vector-specific primers PL3 and PL4 and/or PL3 and E (Kralovicova et al. 2004) (Figure 1A). *(C)* Relative expression of *INS* isoforms in the wild-type reporter and plasmids containing *Hominini*- and *Colobinae*-specific deletions and their combination. The alignment of intron 1 sequences in primates is shown in Supplemental Figure 4. The 16-nt deletion (ctg ccc cac ttg ggg g) took place after the *Homininae* split ~6-9 million years ago.



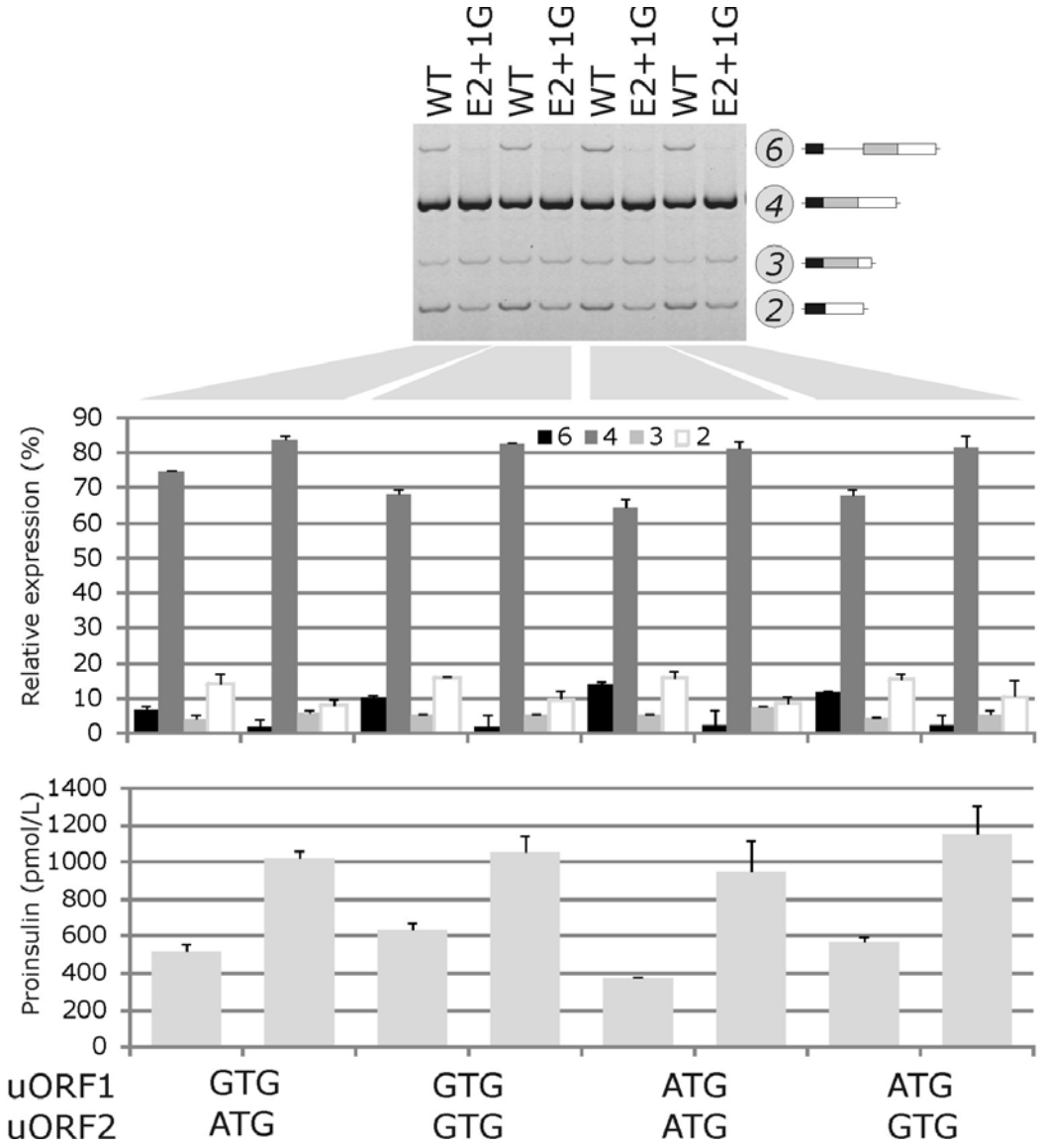
SUPPLEMENTAL FIGURE 6 Predicted branch point sequence and polypyrimidine tract in *INS* intron 1 in primates

Legend: Pp, orangutan; Hs, human. Predicted branch point sequence (orange) and polypyrimidine tracts (green) were determined using an algorithm derived from a comparative analysis of these signals that were conserved between human and mouse (Kol et al. 2005). Predicted branch point sequences in the remaining primates (Supplemental Figure 4) were identical to the human sequence.



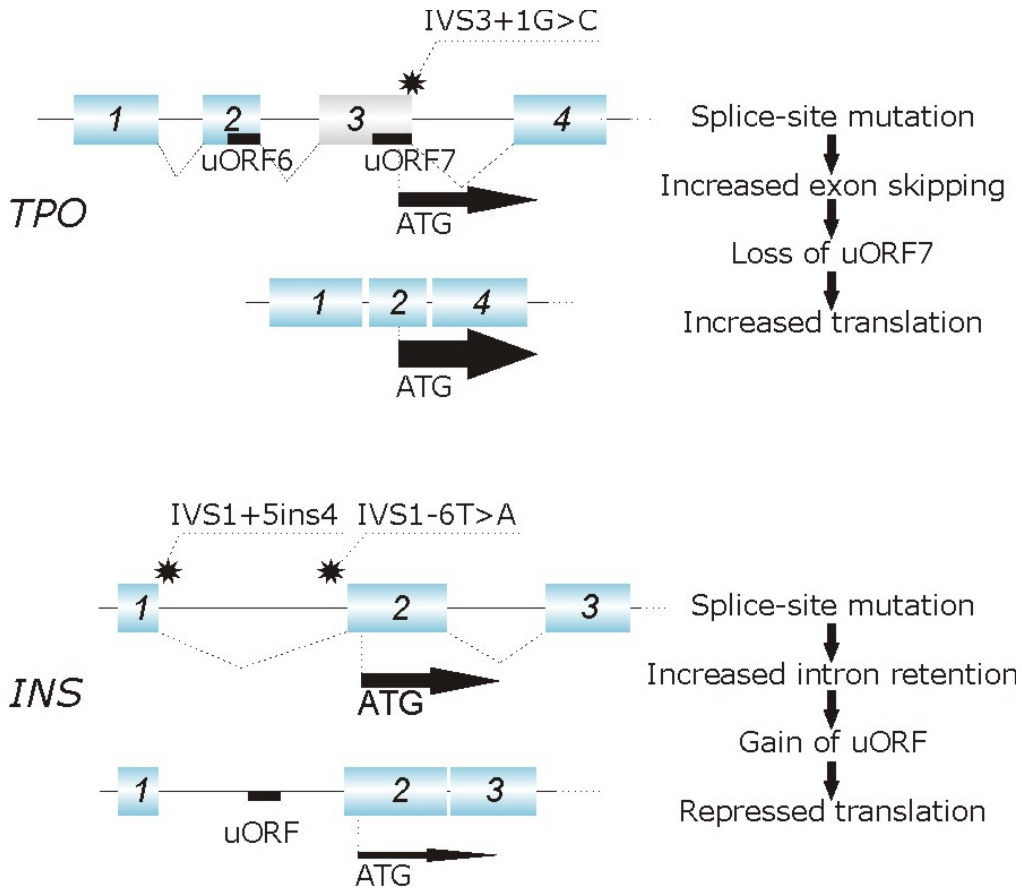
SUPPLEMENTAL FIGURE 7 Coupled translation and splicing regulation of *INS* expression by upstream open reading frames

Legend: *Upper panel*, the splicing pattern of wild-type (WT; E2+1A) and mutated (E2+1G) reporter constructs 48 hours post-trasfection into 293T cells. RNA products are schematically shown to the right; *middle panel,* quantification of mRNA isoforms. *Lower panel,* proinsulin concentration in cultures. Mutations introduced in two upstream open reading frames are shown at the bottom and their location and natural variability in primates is shown in Figure 3. E2+1A (WT) is specific for Great Apes, whereas E2+1G was found in all lower primates.



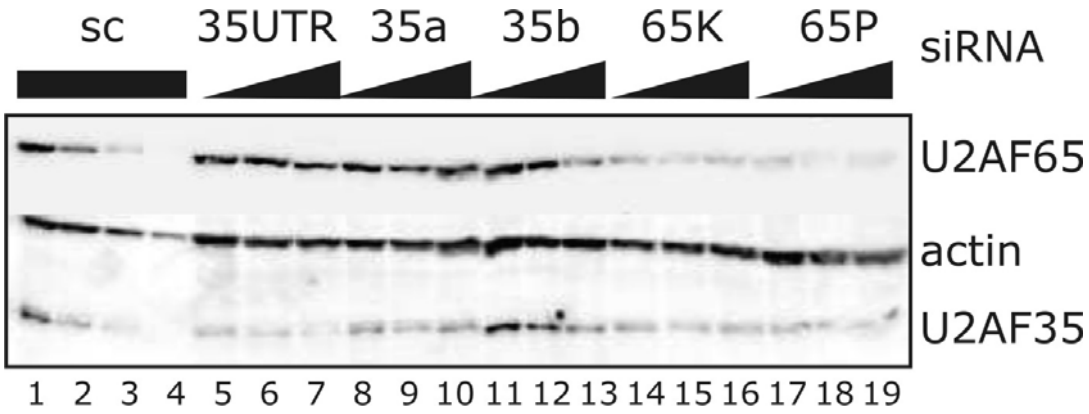
SUPPLEMENTAL FIGURE 8 Extending translational pathophysiology: gain and loss of upstream open reading frames (uORFs) in disease predisposition

Legend: Splicing mutations in the gene for thrombopoetin (*TPO*) induce exon skipping and a loss of the seventh uORF, leading to derepression of translation, excess of TPO and thrombocytosis (Wiestner et al. 1998) *(upper panel)*. In contrast, variants that influence splicing of *INS* intron 1 result in a gain of uORF that curtails translation of intron 1-containing transcripts *(lower panel)*. These transcripts are exported from the nucleus to the cytoplasm (Kralovicova et al. 2006a; Wang et al. 1997). Exons are denoted by numbered boxes, introns as lines, uORFs by black rectangles, disease-causing or -predisposing mutations by stars, and the increase *(upper panel)* or decrease *(lower panel)* of translated products by arrows. The degree of translation efficiency corresponds to the vertical size of the indicated arrows.



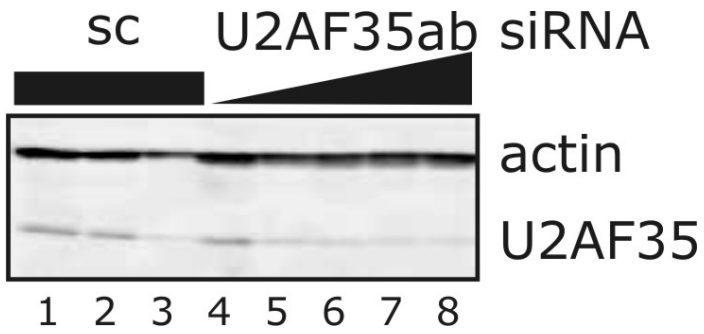
SUPPLEMENTAL FIGURE 9 RNAi-mediated depletion of U2AF subunits

Legend: Western blot analysis of HeLa cells transfected with siRNAs individually targeting U2AF subunits and their isoforms. Final concentrations of each siRNA in culture media were 10, 33 and 100 nM; a scrambled control (GC content 68%) was added at a final concentration of 48 nM. Antibodies are shown to the right. Lanes 2-4 contain two-, four- and eight-fold dilutions, respectively, of the total cell lysate shown in lane 1.



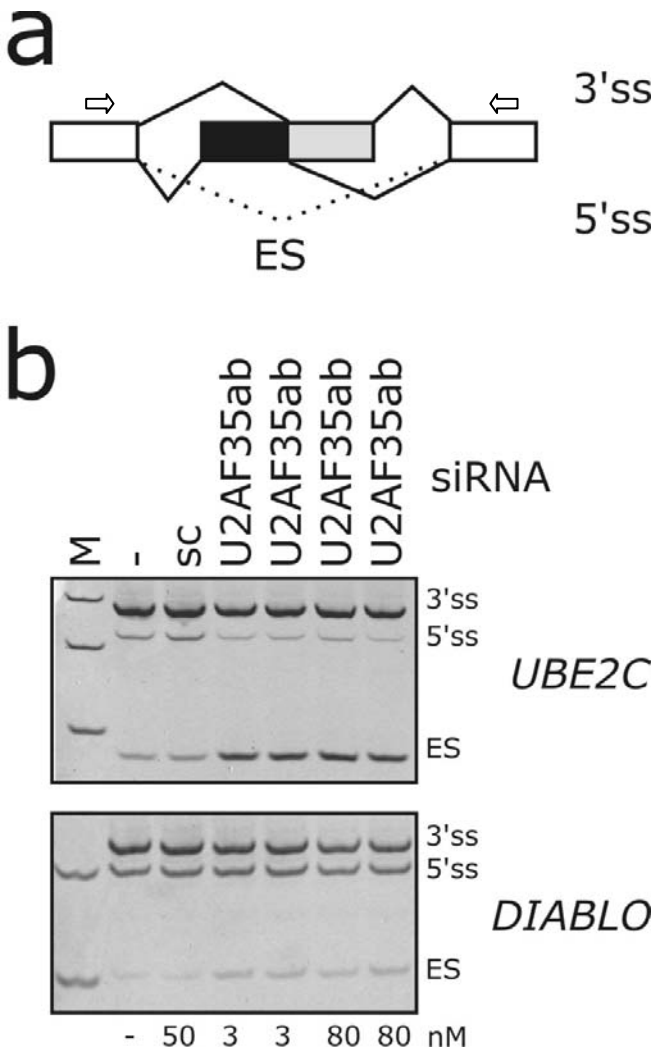
SUPPLEMENTAL FIGURE 10 U2AF35 is effectively downregulated at low concentrations of the U2AF35ab duplex

Legend: Western blot analysis was performed with antibodies shown to the right. Final concentration of the U2AF35ab siRNA in HeLa cell cultures was 0.1, 0.33, 1.0, 3.3 and 10 nM. sc, scrambled siRNA control (47% GC).



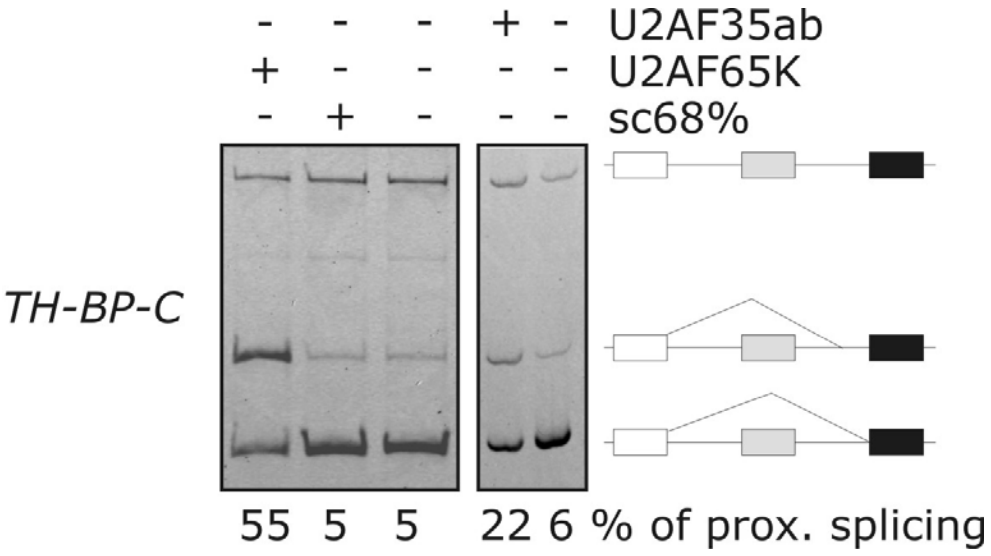
SUPPLEMENTAL FIGURE 11 Selection of dual-specificity splice sites by U2AF35

Legend: *(A)* Schematics of the splicing pattern and location of primers (arrows) used to detect isoforms that result from the use of the 5'ss or 3'ss in *UBE2C/DIABLO* transcripts. ES, exon skipping. *(B)* RT-PCR analysis of total RNA extracted from HeLa cells transfected with U2AF35ab siRNA and a scrambled control (sc). Final concentrations of siRNAs are shown at the bottom. M, 100-nt size marker.

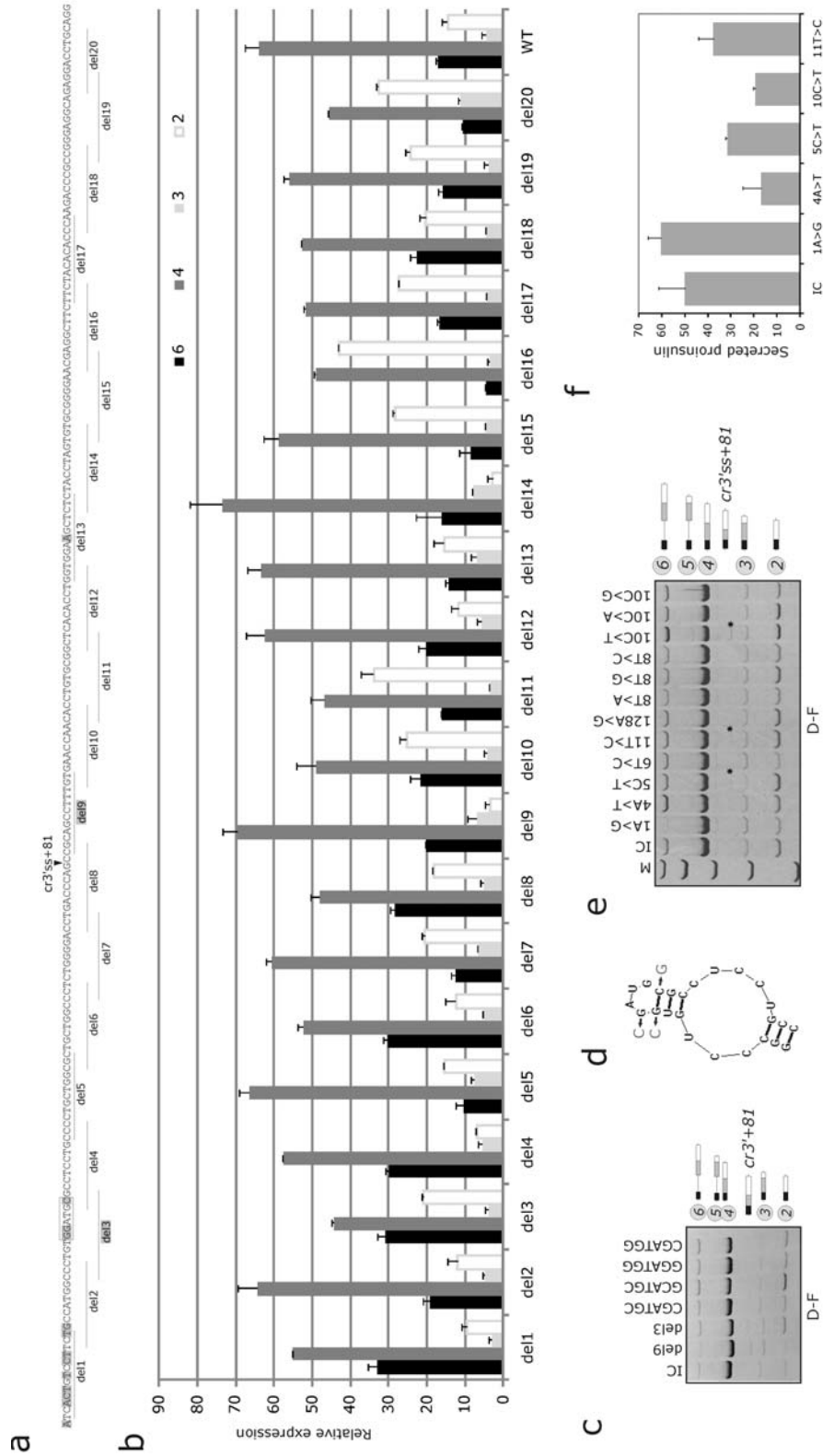


SUPPLEMENTAL FIGURE 12 Skipping of short exons in U2AF-depleted cells

Legend: An example of a reporter construct with a short central exon (96 nt). The reporter construct was described previously (Kralovicova et al. 2006b). siRNAs are shown at the top, RNA products to the right, the name of the reporter to the left and utilization of the intron-proximal 3'ss at the bottom.



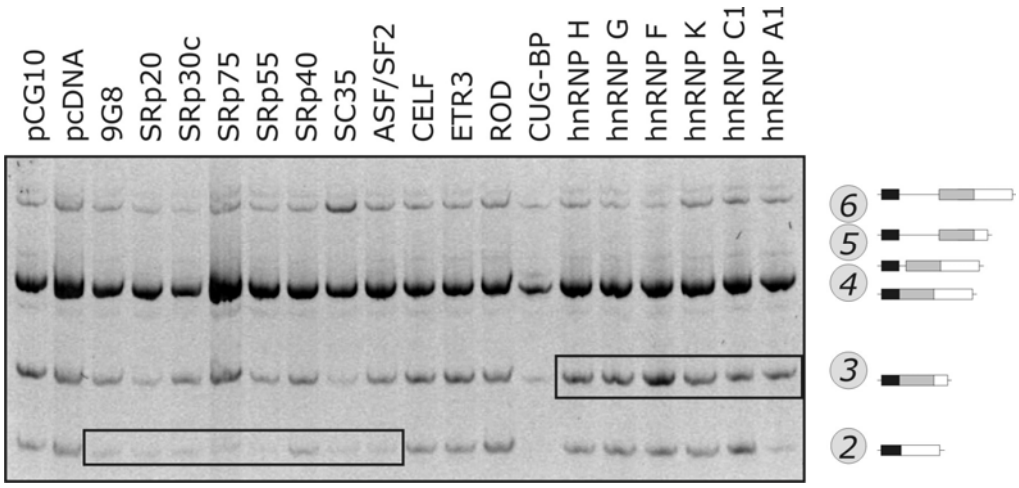
SUPPLEMENTAL FIGURE 13 Identification of auxiliary splicing sequences in exon 2



Legend: *(A) INS* exon 2 deletion constructs (D-F) on the class IC haplotype background. Deletions are shown as horizontal lines (denoted del1-del20). Mutated residues and two deletions that activated cr3'ss+81 are highlighted in gray. *(B)* Relative expression (%) of four *INS* mRNA isoforms in each deletion reporter. Error bars indicate s.d. determined from duplicate wells. The utilization of the remaining isoforms was negligible, except for activation of cr3'ss+81 in del3 and del9. *(C)* Identification of a splicing enhancer in exon 2. Mutations in a putative RESCUE-ESE (Fairbrother et al. 2004) (boxed in panel *(A)* Are shown at the top. RNA products are shown to the right and the *INS* reporter at the bottom. *(D)* A putative structure of the newly identified splicing enhancer in exon 2. Mutations are shown by arrows. *(E)* Activation of cr3'ss+81 by point mutations at exon positions 5, 10 and 11. Mutated positions are highlighted in gray in panel *(A)*. M, 100-nt size marker. *(F)* Secreted proinsulin is shown on the pmol/L scale. Error barrs denote standard deviations of a single transfection experiment in duplicate.

SUPPLEMENTAL FIGURE 14 Coexpression of SR proteins/hnRNPs and *INS* in 293T cells

Legend: Individual SR proteins/hnRNPs are shown at the top, RNA products to the right. Reduction of exon skipping upon individual overexpression of SR proteins is highlighted. An increase of cr3'ss+126 utilization observed for a subset of hnRNPs is also boxed.



SUPPLEMENTAL TABLES

SUPPLEMENTAL TABLE 1 Phylogeneny of G runs in *INS* intron 1 in primates

G run	Great Apes	Old World Monkeys	New World Monkeys	*Strepsirrhini*
G0	x	x	x	x
G1	x	x	x[1]	
G2	x	x[2]		
G3	x	x	x[1,3]	x
G4	x	x	x	
G5	x[4]	x[5]	x[6]	x[4]
G6	x[7]	x[8]	x[7]	x[7]
G7	x	x	x	x

Legend: G runs are numbered as in Supplemental Figure 4. 'x' denotes a conserved G run. [1], G run was shifted by 1 nucleotide in a subset of New World Monkeys; [2], G run was deleted in all examined genera of Asian colobines; [3], extra G_3 in the vicinity; [4], G_5; [5], G_5, except for *Cercopithecini;* [6] G_4; [7], G_4; [8] G_4, except for *Colobinae.*

SUPPLEMENTAL TABLE 2 Predicted strength of *INS* splice sites

Intron	Splice site[1]	Consensus sequence[2]	Maximum entropy score (allele)[3]
1	5'ss	CAG/GTCTKT	6.84 (*INS-69*del), 2.36 (*INS-69*ins)
1	Cr5'ss+30	CAG/GTGGRC	8.07 (*INS-26*G), 3.09 (*INS-26*A)
1	3'ss	CAGCCCTGCCTGTCWCCCAG/ATC	7.57 (*INS-27*T), 4.75 (*INS-27*A)
1	Cr3'ss+81	CCTCTGGGGACCTGACCCAG/CCG	-1.12
2	5'ss	AGG/GTGAGC	7.75
2	3'ss	TGCGCGGCACGTCCTGGCAG/TGG	4.36
2	Cr3'ss+35	GGGCGGGGGCCCTGGTGCAG/GCA	1.54
2	Cr3'ss+126	GCATCTGCTCCCTCTACCAG/CTG	8.24

Legend: [1]Cryptic (Cr) splice sites are denoted by their exon/intron position (in nucleotides) relative to their authentic counterparts; [2], IUB codes: K is G or T, R is A or G, W is A or T. GT and AG dinucleotides are underlined. Exon/intron boundaries are denoted by a slash. [3], Maximum entropy scores were computed as described (Buratti et al. 2007; Vorechovsky 2006; Yeo and Burge 2004). Location of the three variants is shown in Figure 1A. Their designation is according to the previously published nomenclature (Stead et al. 2003); del/ins, 4-nucleotide deletion/insertion allele at *INS-69*. The composite score of intron 1 splice sites is 11.59/14.41 (haplotype I/III, respectively), which is comparable to the composite score of intron 2 splice sites (12.01).

SUPPLEMENTAL TABLE 3 Oligonucleotide primers amplifying competing 3'ss in endogenous transcripts

Gene	Exon	ID#[1]	Primers (5'-3')	Size of PCR products	Distance between competing 3'ss
SERBP1	5	Far2	CTCACAACTGGGGAACTGTC TGTCTGCCACTGGATGATGT	102/147	45
SRRM1	6	Far3	AAAAAAGGGAGCGGTCTCGTA AAGTAGCCTCTTGTACTGAA	138/228	90
SLC15A4	3	Far12	CTGTTCCCAGAAGCGAAGTG GAAAACAGGGACAATCTTGA	114/172	58
CIT	30	Far13	TGAAAATGGAAGGCACTATT GAAGGGCTTCCTCTAGCTCT	165/210	45
LRRC28	6	Far15	GGATCTTAGTGACAATGCCT AAGTGGCAAAAATGCAAGAC	226/321	95
SEPT4	11	Far19	GGACCTGAAGGATGTGACAC CCTCATCTTTCTCTCGGATA	144/197	53
STAT3	23	Far20	CGCTGCCCCATACCTGAAGA TGAGGGTTCAGCACCTTCAC	89/139	50
ATP13A1	14	Far24	GGAGGTGACCCCAGTGTCCA TACGAGGCAAGCACGGACAT	182/252	70
FAM134A	4	Far25	CCTTCTTCCTACTCAGCGTC GCTGGATTCTGCCTCTTGTA	156/244	88
MAP4K4	11	Far27	GAAGAGGAGCAGCGGCACTT CTCGGTCAGCAGGCTCGTAG	187/280	93
C20orf30	3	Far28	TTATGATGCCGTCCCGTACC CCTTTGCTGATGTAGCCTGA	135/223	88
SLC35C2	3	Far29	TTTGTGGAAGGCGGTGTTGA CCCAGCTCAGCACCACACGG	141/212	71
SFRS15	19	Far30	CCTCACACTCCACCAATAAG GTGAAACAGGCTGGGTTACA	151/217	66
FMR1-201	15	E15	GACGCGGTCCTGGATATACTT CCTCCACGCCGCCGTCCGT	85/124/160	39/36
FMR1-201	16	E16	GTAATCCAAGAGAGGCTAAAG TTGCTGACCATCCACGCTGTC	134/185	51
ECM1[2]	4	Far1	CCCCACCCCTATCCCGAAGC GGGAGTTGGGCAGGTAGCAG	148/235	87
SPOPL[2]	4	Far26	AGGAAATGGGTGAAGTGTTA TTGTTTCTTCCCTTTTAGCA	136/207	71
PRP3[2]	12	Far4	TTTGATGCGAGTATTAGGAA TTTTAATTTTCTTGACCTTT	155/213	58
EIF3S7[2]	15	-	CTGATGGCACCTTCAGCTCT CCTCGTAGTCTCGGTGGAAG	127/295	168
PTK9[2]	12	-	GAGGAGCAGCCACTTCCTG CTCCACTGTCCCAGCTCTTC	128/247	119
CUEDC1	17	-	GCCTGCGAGAAGGACAGTAA TCGGTTCTCAAGCACCATCT	141/213	72

Legend: [1], Identity number corresponding to a previously described set of tandem 3'ss separated by >39 nt (Akerman and Mandel-Gutfreund 2007). [2], PCR products did not show the predicted size or failed to amplify.

SUPPLEMENTAL TABLE 4 Selection of competing 3'ss upon depletion of U2AF subunits

Gene	ID 3'ss[4]	ID 3'ss in U2AF35(-)[5]	3'ss promoted in U2AF35(-) cells	ID 3'ss in U2AF65(-)	3'ss promoted in U2AF65(-) cells	Intron- proximal 3'ss (ME score)	Intron-distal 3'ss (ME score)	Splice site more dependent on U2AF
MAP4K4[1]	50	39	Intron-proximal	38	Intron-proximal	AAG/G (7.8)	CAG/C (9.6)	Intron-distal (stronger)
SLC35C2[1]	12	51	Intron-distal	63	Intron-distal	CAG/A (5.7)	CAG/G (8.9)	Intron-proximal (weaker)
SRRM1[1]	10	1	Intron-proximal	1	Intron-proximal	TAG/A (4.5)	CAG/A (7.4)	Intron-distal (stronger)
FMR1-201[1]	17	8	Intron-proximal	9	Intron-proximal	TAG/G (7.4)	TAG/C (-0.7)	Intron-distal (weaker)
						CAG/A (4.0)		
CUEDC1[1]	82	90	Intron-distal	88	Intron-distal	CAG/G (8.3)	CAG/A (0.8)	Intron-proximal (weaker)
UBE2C[1,2]	78	90	Intron-distal	N.D.	N.D.	AAG/G (3.0)	CAG/G (6.7)	Intron-proximal (weaker)
DIABLO[1,2]	65	60	Intron-proximal	N.D.	N.D.	TAG/G (10.4)	CAG/G (8.8)	Intron-distal (weaker)
INS intron 1(cr3'ss+81)[3]	0	19	Intron-distal	0	N.I.	CAG/A (4.8)	CAG/C (-1.1)	Intron-proximal (stronger)
INS intron 2(cr3'ss+126)[3]	15 (IC D-F)[6]	0 (IC D-F)[6]	Intron-proximal	2 (IC D-F)[6]	Intron-proximal	CAG/T (4.4)	CAG/C (8.2)	Intron-distal (stronger)
	55 (IC D-C)[6]	3 (IC D-C)[6]		9 (IC D-C)[6]				
LIPC (ESS13, SS63)[3]	54	40	Intron-proximal	38	Intron-proximal	TAG/A (6.2)	CAG/T (-2.3)	Intron-distal (weaker)
TH-BP-C[3]	95	78	Intron-proximal	45	Intron-proximal	CAG/G (5.2)	CAG/C (5.1)	Intron-distal (weaker)

Legend: [1], endogenous transcripts; [2], dual-specificity splice sites (Zhang et al. 2007); [3], exogenous transcripts; *LIPC* and *TH* minigenes were described previously (Kralovicova et al. 2006b; Kralovicova and Vorechovsky 2007); [4], % of splicing to intron-distal 3'ss in untreated cells; [5], % of splicing to intron-distal 3'ss in U2AF35-deficient (-) cells; [6], haplotype and primers used for minigene cloning are in parentheses. N.I., not informative; N.D., not determined; ME (maximum entropy) scores were computed as described previously (Buratti et al. 2007; Vorechovsky 2006; Yeo and Burge 2004).

SUPPLEMENTAL TABLE 5 A list of synthetic small interfering RNAs

Target	siRNA	Source
U2AF35a	CCAUUGCCCUCUUGAACAU	(Pacheco et al. 2006)
U2AF35b	CCAUCUUGAUUCAAAACAU	(Pacheco et al. 2006)
hU2AF35ab	GGCUGUGAUUGACUUGAAU	(Pacheco et al. 2006)
U2AF35-UTR	AGUGUUGUAGUUGAUUGAC	This study
U2AF26-UTR	AGCCCCCUUCACUCUCCUG	This study
U2AF26-83	GGACAAGGUUAACUGCUCU	This study
U2AF65K	GCAAGUACGGGCUUGUCAA	(Hastings et al. 2007)
U2AF65P	GCACGGUGGACUGAUUCGU	(Pacheco et al. 2006)
PUF60	GCAGAUGAACUCGGUGAUG	(Hastings et al. 2007)
nPTB-N1	GAGAGGAUCUGACGAACUA	(Spellman et al. 2007)
nPTB-N3	UAAGAAACCUGGAUCCAAA	(Spellman et al. 2007)
PTB	CUUCCAUCAUUCCAGAGAA	(Wollerton et al. 2004)
9G8-Gao	AGAUCAAGAUCCAGGUCUA	(Gao et al. 2007)
9G8-UTR	GGGAUUUGUGAUGUCUGUA	This study
9G8-UTR2	GUGUAACCUAGGAAAGAUA	This study
SRp75-1	AGACCAAGCUGAAGAGAAG	This study
SRp75-2	AGCAGUCAUUCUAAGAGUA	This study
SRp55-1	GCAGAUCUAAGGAUGAGUA	This study
SRp55-2	AAGAUGAGGCUCUAAGGAA	This study
SRp40-1	UGCUCCACCUGUAAGAACA	This study
SRp40-2	ACGUGGUUCUUCAAGUAGA	This study
SRp30c	AGAGGAUGCUAUUUAUGGA	(Paradis et al. 2007)
SRp30b/SC35	AAUCCAGGUCGCGAUCGAA	(Gabut et al. 2005),
SRp30b/SC35	CAAGCAACUGGCUAUUGAA	This study
SRp30a/ASF	ACGAUUGCCGCAUCUACGU	(Karni et al. 2007)
SRp20-1	GGAAAUAGAAGACAGUUUG	(Bedard et al. 2007)
SRp20-2	GGUCCCUUUCUAGAGAUAG	(Bedard et al. 2007)
RSRC1-ORF1	GAACGCGUAGUCGGUCUCG	This study
RSRC1-ORF2	UGGGCAGUCCUGUGGCCUA	This study
Scrambled control (47% CG)	AGGUAGUGUAAUCGCCUUG	This study
Scrambled control (68% CG)	UGCGCUAGGCCUCGGUUGC	This study
hnRNP H	UCAGAAGAUGAAGUCAAAU	(Paul et al. 2006)
hnRNP-F	GCGACCGAGAACGACAUUU	(Garneau et al. 2005)
Tra2α-UTR	CGGAAUGGUUGCAAUUAAG	This study
Tra2β-UTR	GCAUGAAGACUUUCUGAAA	This study
ZRSR1/2	CAACAACUAGAAGAAGAGA	This study
ZRSR2-UTR	GGCUAUGAACGCAGAUGGG	This study
CAPERα-UTR	GGAUCUACUGUCAUUUGUA	This study
CAPERβ-UTR	UCAGUGGCACAGUAUACUG	This study
hSlu7-ORF	GUGAGUUAUGCUGGAGAUA	This study
hSlu7-UTR	GACCAUCCAAGAUAGAUGC	This study

References to Supplemental material

Akerman M, Mandel-Gutfreund Y (2007) Does distance matter? Variations in alternative 3' splicing regulation. Nucleic Acids Res 35:5487-5498

Bedard KM, Daijogo S, Semler BL (2007) A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation. EMBO J 26:459-467

Buratti E, Chivers MC, Kralovicova J, Romano M, Baralle M, Krainer AR, Vorechovsky I (2007) Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic Acids Res 35:4250-4263

Fairbrother WG, Yeo GW, Yeh R, Goldstein P, Mawson M, Sharp PA, Burge CB (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res 32:W187-190

Gabut M, Mine M, Marsac C, Brivet M, Tazi J, Soret J (2005) The SR protein SC35 is responsible for aberrant splicing of the E1alpha pyruvate dehydrogenase mRNA in a case of mental retardation with lactic acidosis. Mol Cell Biol 25:3286-3294

Gao L, Wang J, Wang Y, Andreadis A (2007) SR protein 9G8 modulates splicing of tau exon 10 via its proximal downstream intron, a clustering region for frontotemporal dementia mutations. Mol Cell Neurosci 34:48-58

Garneau D, Revil T, Fisette JF, Chabot B (2005) Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. J Biol Chem 280:22641-22650

Hastings ML, Allemand E, Duelli DM, Myers MP, Krainer AR (2007) Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF. PLoS ONE 2:e538

Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol 14:185-193

Kol G, Lev-Maor G, Ast G (2005) Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. Hum Mol Genet 14:1559-1568

Kralovicova J, Gaunt TR, Rodriguez S, Wood PJ, Day INM, Vorechovsky I (2006a) Variants in the human insulin gene that affect pre-mRNA splicing: is -23*Hph*I a functional single nucleotide polymorphism at *IDDM2*? Diabetes 55:260-264

Kralovicova J, Haixin L, Vorechovsky I (2006b) Phenotypic consequences of branchpoint substitutions. Hum Mutat 27:803-813

Kralovicova J, Houngninou-Molango S, Kramer A, Vorechovsky I (2004) Branch sites haplotypes that control alternative splicing. Hum Mol Genet 13:3189-3202

Kralovicova J, Vorechovsky I (2007) Global control of aberrant splice site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition. Nucleic Acids Res 35:6399-6413

Pacheco TR, Coelho MB, Desterro JM, Mollet I, Carmo-Fonseca M (2006) In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. Mol Cell Biol 26:8183-8190

Paradis C, Cloutier P, Shkreta L, Toutant J, Klarskov K, Chabot B (2007) hnRNP I/PTB can antagonize the splicing repressor activity of SRp30c. RNA 13:1287-1300

Paul S, Dansithong W, Kim D, Rossi J, Webster NJ, Comai L, Reddy S (2006) Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. EMBO J 25:4271-4283

Rigo F, Martinson HG (2008) Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. Mol Cell Biol 28:849-862

Spellman R, Llorian M, Smith CW (2007) Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. Mol Cell 27:420-434

Stead JD, Hurles ME, Jeffreys AJ (2003) Global haplotype diversity in the human insulin gene region. Genome Res 13:2101-2111

Vorechovsky I (2006) Aberrant 3' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic Acids Res 34:4630-4641

Wang J, Shen L, Najafi H, Kolberg J, Matschinsky FM, Urdea M, German M (1997) Regulation of insulin preRNA splicing by glucose. Proc Natl Acad Sci USA 94:4360-4365

Wiestner A, Schlemper RJ, van der Maas AP, Skoda RC (1998) An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. Nat Genet 18:49-52

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell 13:91-100

Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11:377-394

Zhang C, Hastings ML, Krainer AR, Zhang MQ (2007) Dual-specificity splice sites function alternatively as 5' and 3' splice sites. Proc Natl Acad Sci USA 104:15028-15033