

# The Impact of Pro-Inflammatory Cytokines on Alternative Splicing Patterns in Human Islets

## *Supplemental Material*

### **RNA Isolation, Library Preparation, and Sequencing**

Total RNA was isolated using the SeraMir RNA isolation kit (SBI Biosciences). RNA quantification and quality were assessed by measuring the 260/280 ratio using a Nanophotometer (Implen, Munich, Germany). RNA integrity was evaluated using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA). The average ( $\pm$ S.D.) RIN value for islet samples utilized in this study was  $9.8 \pm 0.28$ . Next, 500 nanograms of RNA were used to prepare a single-indexed strand-specific cDNA library using a TruSeq Stranded mRNA Library Prep Kit (Illumina). The resulting libraries were assessed for quantity and size distribution using a Qubit and Agilent 2100 Bioanalyzer; 1.5 pM pooled libraries were sequenced with 2 $\times$ 75bp paired-end configuration on an Illumina NextSeq500 using a NextSeq 500/550 High Output Kit with an average of 52.6M reads. A Phred quality score (Q score) was used to measure the quality of sequencing; >90% of the sequencing reads reached Q30 (99.9% base call accuracy).

### **RNA-Sequencing Alignment and Differential Expression Analysis**

Sequencing data were first assessed for quality using FastQC (Babraham Bioinformatics, Cambridge, UK). Next, all sequenced libraries were mapped to the human genome (GENCODE GRCh37) using the STAR RNA-seq aligner (1). Read distributions across the genome were assessed using bamutils (from ngsutils) (2). Differential expression analysis between cytokine-treated and untreated human islets was performed using edgeR v3.22.3 from the Bioconductor package (3). Biological coefficients of variation between the samples were estimated using an empirical Bayes approach under the assumption that data followed a negative binomial distribution. We filtered out low expression transcripts based on the percentage of samples (less than 50%) and using a count per million (CPM) cutoff of 0.5. A total of 15,892 mRNAs remained after filtering and were used in differential expression analysis. Age, donor sex, and BMI were adjusted as co-variables in our statistical model. Statistical significance was defined as an FDR  $P$  value  $\leq 0.05$  for comparison between cytokine-treated and untreated human islets.

### **RNA Sequencing and Alternative Splicing Analysis**

To provide additional validation of our results, we analyzed one external dataset of human islets treated using the same cytokines (*GSE108413*)(4). We also analyzed alternative splicing patterns in islet from 4 donors with type 1 diabetes (*GSE121863*) and in islets from 12 non-diabetic control donors (5). These datasets were retrieved from NCBI GEO and analyzed using the same computational pipeline described in the Material and Methods. Overlapping events between *GSE121863* and our RNA-Seq dataset were defined as those 1) with more than half of each comparison group replicate having a sum of inclusion junction counts and skipping junction counts per sample  $\geq 10$ , 2) FDR  $\leq 0.05$ , and 3) exhibiting consistent alternative exon change trend between the two datasets.

### **NMD Prediction for Retained Intron Events.**

To predict non-sense mediated decay targets among mRNAs with significant differentially-spliced retained intron events, we first determined the primary transcripts (source transcripts) that harbored the retained intron (IR) events. The transcript was specified as containing both flanking exons of the IR event, and the transcript with the longest coding sequence (CDS) was identified as primary transcript for downstream analysis. Then we retrieved the genomic intron sequences of each transcript from the human genome reference repository (GENCODE GRCh37). We inserted the intron sequence into the CDS of the transcript at the junction of the intron, and the transcript sequence was translated using its source transcript open reading frame (ORF). For each retained intron event, the location of the translation termination site and the last exon-exon junction could then be determined. Transcripts were identified as NMD-sensitive if the stop codon fell more than 50 nt upstream of the final exon-exon junction, indicating a pre-mature stop codon (PTC).

### **Pairwise and Network Analysis of RNA Binding Protein**

To calculate the presence of motifs on the 7 sequence regions around differentially spliced skipped exon events, we used FIMO with a custom library of all TRANSFAC (and Jaspar) motifs at a  $P$  value threshold of  $10^{-4}$  (6). We explored previously known motif binding patterns of RBP's by performing hierarchical clustering. The distance between motifs was defined as distance = 1 - correlation. Spearman correlations were used to calculate motif co-occurrence correlation coefficients. A network diagram was constructed using Cytoscape (version 3.7.2) (7) to visualize an alternative splicing regulatory network with key known splicing factors by applying a yFiles Organic layout. Only significant skipped-exon events with an ExonImpact prediction score  $\geq 0.5$  were included as nodes in the network. For RBPs having two or more known motifs, the average of the motif counts was utilized for testing.

### **Quantitative PCR Validation of HLA-DMB Skipped-Exon Transcripts**

To demonstrate the presence of HLA-DMB transcripts with and without exon 5, a pair of primers was designed within exon 4 and exon 6 (Supplementary Table 2), based on the NCBI reference sequence NM\_002118.4 utilizing PrimerQuest software (IDT, Coralville, IA). Amplicons with and without exon 5 (184 and 148 bp, respectively) were PCR-amplified and visualized on a 2.5% agarose gel with SYBR safe DNA stain. In addition, SYBR green-based quantitative PCR (qPCR) was performed using primers that specifically amplified the exon-inclusive transcript or the skipped-exon transcript (Supplementary Table 2). Ct values derived from the skipped-exon transcript and the exon-inclusive transcript were used to calculate  $\Delta Ct$  ( $\Delta Ct = C_{t}^{\text{inclusive-exon}} - C_{t}^{\text{skipped-exon}}$ ), the ratio (R) between the two transcripts ( $R = 2^{-\Delta Ct}$ ), and the inclusion ratio ( $R/(R+1) \times 100\%$ ).

### **RNAscope In-situ Hybridization Assay.**

RNA FISH was performed to quantify expression of HLA-DMB total RNA and the HLA-DMB splice variant in human pancreatic sections from the Network of Pancreatic Organ Donors with Diabetes (nPOD). Formalin fixed paraffin embedded (FFPE) human pancreatic sections with a thickness of  $\sim 4\mu\text{m}$  were received from nPOD. RNA FISH for total RNA was performed using the RNAscope multiplex Fluorescent Reagent Kit v2 (Cat# 323100), and the BaseScope Detection

Reagent kit v2 (Cat# 323900) was used for detection of the mRNA splice variant as per the manufacturer's instructions. Probe sets for detection of HLA-DMB total RNA were obtained from Advance Cell Diagnosis Inc. For detection of the HLA-DMB splice variant, a custom designed probe set flanking exon 4 and exon 6 (973-1059 bp) was used. To confirm the validity of our staining in pancreatic tissues, we included a positive (PPIB) and a negative control probe set (DapB) in parallel with our mRNAs of interest. All stained slides were counterstained with DAPI and mounted using a coverslip with fluorescent mounting media (ProLong Gold Antifade Reagent, Life Technologies). Images were obtained using a Zeiss LSM800 confocal microscope with AiryScan (Carl Zeiss, Germany). Images were analyzed and processed using the NIH ImageJ (8) and an in-house single molecule image analysis algorithm as described below.

### **Quantitative Single Cell FISH Analysis.**

Expression of the HLA-DMB full length and splice variant RNAs was quantified using an in-house single molecule quantification algorithm, allowing for quantification at the single cell level with single molecule resolution (Supplementary Figure 1). Analysis began with segmentation of single nuclei, which was achieved via a binary Watershed algorithm in ImageJ (8). Manual annotation was utilized when the watershed algorithm proved inadequate for accurately deconvolving clumped nuclei. Based on the nuclei segmentation masks, the cytoplasm of single cells and the cell-cell boundaries were generated by dilating the nuclear mask with 150 pixels (5 $\mu$ m) in CellProfiler (9). Intersecting regions of dilated masks were removed along with masks touching the image boundary. Classification of cells were enabled according to partitions of the glucagon and insulin channels, and the mean intensity of both channels was calculated over each cellular mask and scattered in 2D intensity-space. For each sample, a pair of linear equations with manually defined parameters were used to partition the 2D intensity-space into four unique regions: Alpha (glucagon+), Beta (insulin+), Ambiguous (insulin+&glucagon+), and Double-Negative (insulin-&glucagon-). Individual RNA transcripts were registered in fluorescent images in a three-stage procedure: deconvolution, denoising, and detection. Deconvolution involved the subtraction of a local background estimate provided by a boxcar kernel, while denoising was achieved by convolution with a gaussian kernel. The Laplacian of Gaussian (LoG) blob-detector was used to identify the localization and intensity of each smFISH foci. For each slide, the foci intensity ( $I_{foci}$ ) was summarized in a histogram and the lowest peak of the intensity histogram was identified as the intensity value of a single mRNA ( $I_{mRNA}$ ). The absolute number of mRNA molecules clustered in a smFISH foci ( $N$ ) was determined by normalizing the smFISH foci intensity by the intensity value of a single mRNA (10-13),  $N = I_{foci}/I_{mRNA}$ . Finally, each RNA localization was assigned to the nucleus or the cytoplasm of a given cell by encoding masked regions with a binary label and assigning the appropriate binary label based on transcript coordinates. Data were graphed as the ratio of the expression level of the alternatively spliced (AS) mRNA to the mean value of full length mRNA number per cell, following an equation, AS/Full-length Ratio = (number of AS mRNAs per cell) / [Mean (number of Full-length mRNAs per cell)].

### **Western Blot and Immunofluorescence**

**Western blot.** Human islets were harvested and lysed with lysis buffer and protein concentration was determined using the Lowry method. Next, 20 ug of protein was denatured for 10 minutes

at 70°C. A total of 20 ug of protein was separated using 4-12% Bis-Tris gels (Invitrogen, Cat# NW04122BOX) and transferred to PVDF membrane using a 1X NuPAGE transfer buffer. The membranes were blocked for 30 minutes with Intercept Blocking Buffer (LI-COR, Cat# 927-70001) and incubated with primary antibodies for  $\beta$ -Tub (Cell Signaling, Cat # 2146S), HLA-DMB (Sigma, Cat# HPA012298) overnight at 4°C. Secondary antibodies for anti-rabbit (LI-COR, Cat# P/N: 926-68073) antibodies were used for quantification of protein expression.

**Immunofluorescence staining.** Human islets treated with or without pro-inflammatory cytokines were fixed in 4% PFA for 30 minutes and were then washed with 1X PBS twice for 5 minutes. Islets underwent blocking with 1X animal-free blocking buffer (Vector laboratories, cat# SP-5030) for 60 minutes, followed by an overnight incubation with primary antibodies for insulin (Dako, cat#IR002), HLA-DMB (Sigma, Cat # HPA012298) and Lamp1 (Developmental Studies Hybridoma Bank, cat# ID4B-s). The islets were washed with 1x PBS twice for 5 minutes and incubated with secondary antibodies for anti-guinea pig (Invitrogen, Cat# A11073), anti-mouse (Invitrogen, cat#A32727), and anti-rabbit (Invitrogen, Cat#31573) for 2 hours, followed by DAPI (Invitrogen, Cat#D1306) nuclear staining for 15 minutes at room temperature. Imaging was performed using an LSM 800 (Carl Zeiss, Germany) confocal microscope.

## Supplementary Tables

**Supplementary Table 1. Characteristics of the organ donors and human islet preparations used in RNA-Sequencing and smFISH.**

### *RNA-Sequencing*

ID	Gender	Age	BMI	Islet Purity (%)	Islet Viability (%)
1	Female	33	34.2	95	98
2	Female	43	29.5	80	96
3	Male	35	24.2	95	95
4	Male	30	35.9	90	95
5	Female	52	25.9	90	95
6	Male	23	24.8	95	95
7	Female	44	34.5	90	90
8	Male	28	24.6	91	95
9	Male	15	24.6	80	97
10	Male	28	30.7	95	95

### *smFISH*

nPOD Donor_ID	Donor Type	Gender	Age	BMI	T1D Duration (yrs)	Autoantibody
6250	No Diabetes	M	40	27.9	N/A	No
6048	No Diabetes	M	30	20.6	N/A	No
6179	No Diabetes	F	20	20.7	N/A	No
6447	No Diabetes	M	8	16.3	N/A	No
6328	T1D	M	39	24.0	20	GADA <sup>+</sup> mIAA <sup>+</sup>
6088	T1D	M	31	27.0	5	GADA <sup>+</sup> mIAA <sup>+</sup> IA2A <sup>+</sup> ZnT8A <sup>+</sup>
6459	T1D	F	22	36.7	10	mIAA <sup>+</sup> IA2A <sup>+</sup> ZnT8A <sup>+</sup>
6443	T1D	F	16	34.5	9	mIAA <sup>+</sup> IA2A <sup>+</sup> ZnT8A <sup>+</sup>
6197	AAb <sup>+</sup>	M	22	28.2	N/A	GADA <sup>+</sup> IA-2A <sup>+</sup>
6310	AAb <sup>+</sup>	F	28	23.9	N/A	GADA <sup>+</sup>
6450	AAb <sup>+</sup>	F	22	24.4	N/A	GADA <sup>+</sup> ZnT8A <sup>+</sup>

**Supplementary Table 2. Primer information for HLA-DMB qPCR**

Name	Note	Sequence (5'→3')
HLA-DMB-F	Agarose gel electrophoresis	ATGCAGACCCTGAAGGTTTC
HLA-DMB-R	Agarose gel electrophoresis	CACTTGGAGTGGAAGTTGTAGG
Inclusive-exon-F (Inc-F)	qPCR Quantification	GCC TCA TCA TCT TCT CTC TTG GTG TGA
Inclusive-exon-R (Inc-R)	qPCR Quantification	TCC TTC TGA ATA ATT GGA CCC AGG AAG
Exclusive-exon-F (Ex-F)	qPCR Quantification	ATG CAG ACC CTG AAG GTT TCT GTG T
Exclusive-exon-R(Ex-R)	qPCR Quantification	GAA ATG TGC CAT CCT AGA GTG G

**Supplementary Table 3. NMD prediction for Retained Intron events (Excel spreadsheet).**  
NMD prediction results for significant differentially-spliced retained intron events.**Supplementary Table 4. Characteristics of the organ donors and human islet preparations used in external GEO datasets*****GSE121863***

ID	Gender	Age	Health Status
1	Male	24	T1D
2	Female	10	non-diabetic
3	Male	16	non-diabetic
4	Male	20	T1D
5	Female	12	non-diabetic
6	Male	15	non-diabetic
7	Male	19	non-diabetic
8	Male	11	non-diabetic
9	Male	30	non-diabetic
10	Male	23	non-diabetic
11	Female	12	non-diabetic
12	Female	16	non-diabetic
13	Male	11	non-diabetic
14	Male	18	non-diabetic
15	Male	23	T1D
16	Female	12	T1D

***GSE108413***

ID	Gender	Age	BMI	Islet Viability (%)
1	Female	23	22.5	95.9

2	Male	31	27.8	Not Known
3	Male	77	24.5	94.9
4	Female	64	29.4	92.0
5	Female	58	21.3	94.6

**Supplementary Table 5. ExonImpact prediction and APPRIS database annotation (Excel spreadsheet with two tabs).** (A) ExonImpact prediction and alternative splice isoforms annotation (APPRIS) of splicing events with a prediction score  $\geq 0.5$ ; (B) annotation of protein structural features.

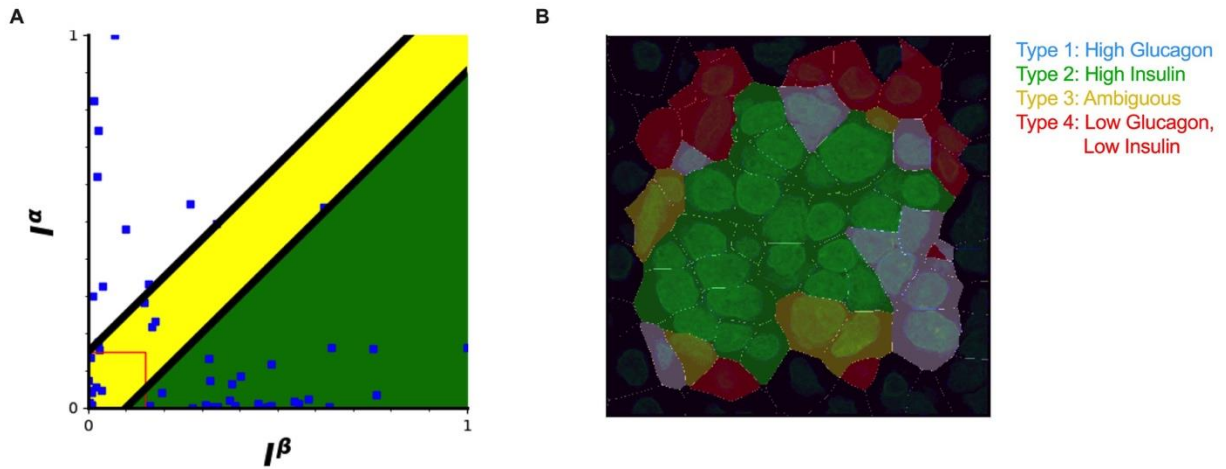
**Supplementary Table 6. FIMO prediction of SRSF2 targeted mRNAs (Excel spreadsheet).** Column “Region\_FIMOPrediction”: RNA-binding protein (RBP) motif enrichment analysis was performed on 7 sequence regions around differentially spliced skipped exon events. We searched for intronic splicing regulators within introns 300-bases upstream (region 3) and downstream of a skipped exon (region 5); 300-bases downstream of the upstream splice junction (region 2); and 300-bases upstream of the downstream splice junction (region 6). We searched for exonic splicing regulators within exons 150-bases upstream of the upstream splice junction (region 1); within the skipped exon itself (region 4); and 150-bases downstream of the downstream splice junction (region 7) (for details see Figure 1).

**Supplementary Table 7. GWAS and sQTL co-localization analysis (Excel spreadsheet).** Colocalization analysis of pancreas tissue specific sQTLs that overlap GWAS traits ( $P < 5 \times 10^{-8}$ ) and all variants in high linkage disequilibrium (EUR,  $R^2 > 0.8$ ).

**Supplementary Table 8. Significant alternatively spliced exons in MHC class II pathway genes (Excel spreadsheet).**

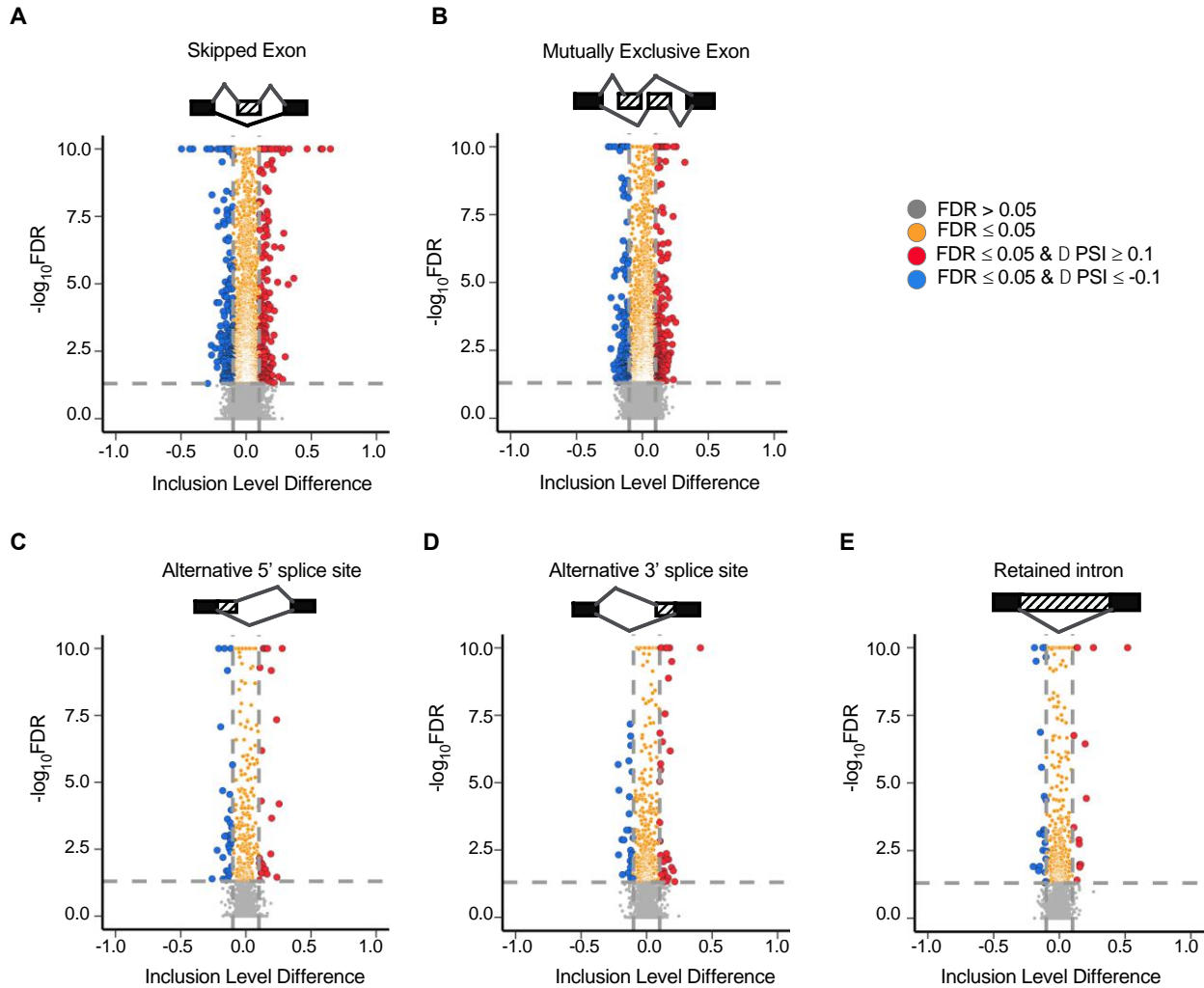
## Supplementary Figures

**Supplementary Figure 1. Classification of cells based on the expression of insulin and glucagon in human islets.** (A) Scatter plot of the glucagon and insulin intensities for each single cell. (B) Segmented cells were classified into four groups, which are coded by color. Light blue: high glucagon and low insulin; Green: low glucagon and high insulin; Yellow: high glucagon and high insulin; Red: low glucagon and low insulin. Cells shown in green (low glucagon and high insulin) were used for quantification.

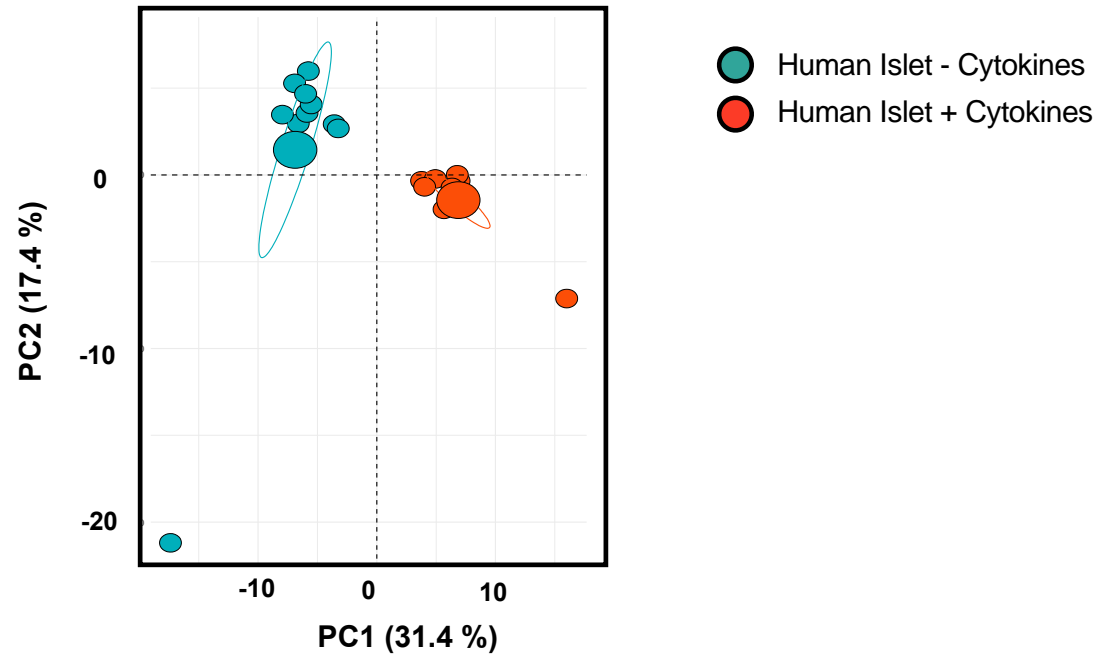




**Supplementary Figure 2. Volcano plot of differential spliced variations.** Volcano plot of all identified splicing events (requiring at least 10 counts per junction in more than half of each comparison group replicate) and dot plot of representative genes, illustrating splicing inclusion level differences between control and cytokine-treated human islets for five different alternative splicing categories. (A) Skipped Exon; (B) Mutually Exclusive Exon; (C) Alternative 5' splice site; (D) Alternative 3' splice site; (E) Retained introns.

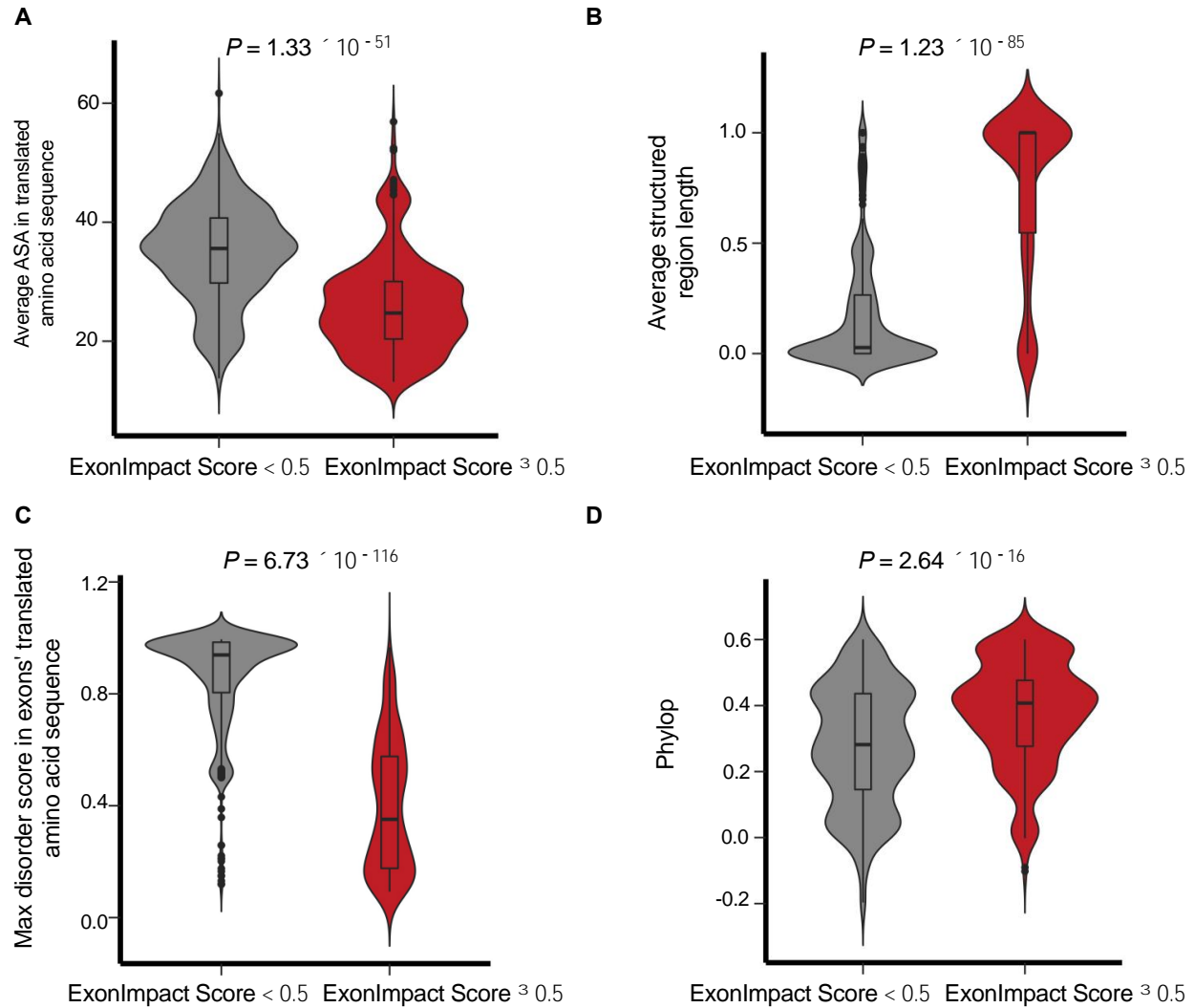


**Supplementary Figure 3. PCA plot of differential spliced variations.** Principle component analysis of differentially spliced variants.

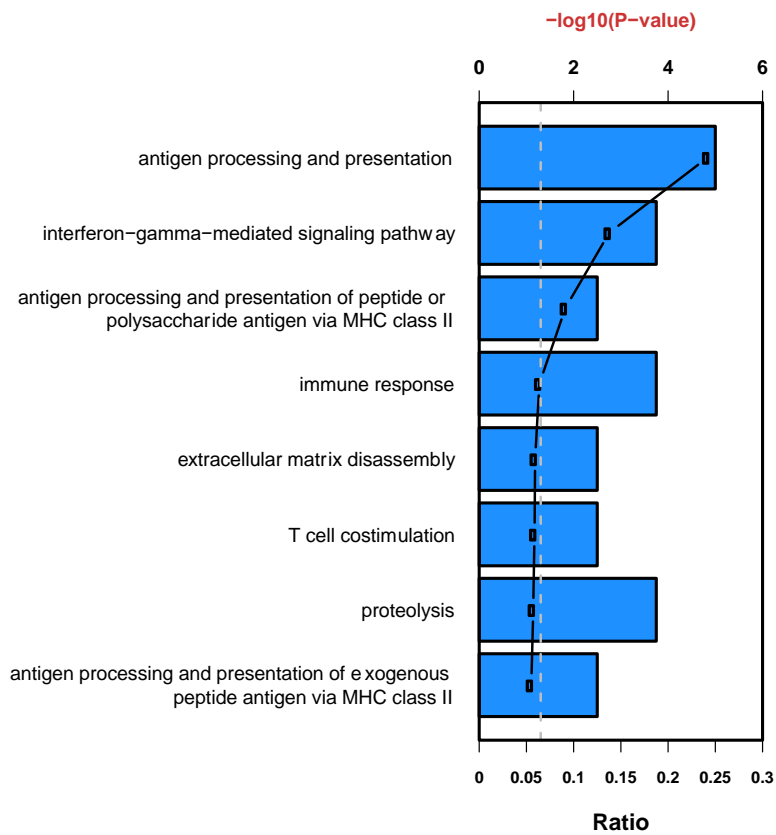


**Supplementary Figure 4.  $\beta$ -cell specifically expressed alternative splicing events in islets from type 1 diabetes donors (*GSE121863*).** (A) Volcano plot of all the splicing events detected with  $\text{FDR} \leq 0.05$  (requiring at least 10 counts per junction in more than half of each comparison group replicate). Splicing variants annotated with gene names requiring 1) at least 10 counts per junction in more than half of each comparison group replicate, and 2) with the same splicing change trend, 3)  $\text{FDR} \leq 0.05$  in both datasets (our data *GSE169221* and *GSE121863*), 4) a cut-off of 10% on  $\Delta\text{PSI}$  was applied to *GSE121863*, 5) ranked with top 15 differential changes in  $\Delta\text{PSI}$ ; (B) Correlation of splicing inclusion levels between the Russell MA, et al dataset (*GSE121863*) and our dataset (*GSE169221*) for 80 concordant splicing events. The x axis represents differential splicing inclusion levels in our data *GSE169221*. The y axis indicates differential splicing inclusion levels observed in Russell MA, et al *GSE121863*; (C) Gene ontology analysis of overlapped genes based on Biological Process terms. Overlapped genes were defined as events with more than half of each comparison group replicate having the sum of inclusion junction counts and skipping junction counts per sample greater than ten, reaching a cutoff of  $\text{FDR} \leq 0.05$ , with the same splicing change trend in both sets.

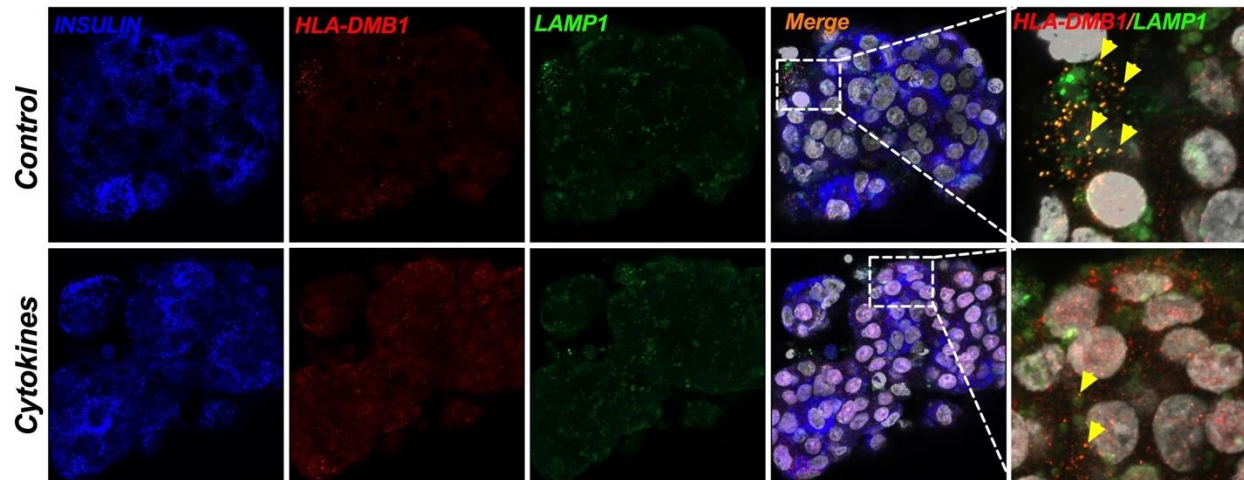
**Supplementary Figure 5. Comparison of the molecular features by ExonImpact scores.** Violin-box plots for molecular features of dichotomized groups of AS events (low=prediction score < 0.5) and (high=prediction score  $\geq$  0.5). Shown are the kernel density plots and median value with interquartile ranges for the (A) Average accessible surface area (ASA) in the translated amino acid sequence ( $P = 1.33 \times 10^{-51}$ ); (B) Average structured region length ( $P = 1.23 \times 10^{-85}$ ); (C) Max disorder score in translated amino acid sequence of an exon ( $P = 6.73 \times 10^{-116}$ ); (D) average phylop score of the exon ( $P = 2.64 \times 10^{-16}$ ).  $P$  values were calculated by Wilcoxon rank-sum tests.



**Supplementary Figure 6.** Gene ontology analysis of eGenes (genes with an sQTL) based on Biological Process terms.



**Supplementary Figure 7. Immunofluorescence staining of HLA-DMB protein in human islets.** Immunofluorescence staining of human islets treated with or without pro-inflammatory cytokines probed for insulin (blue), HLA-DMB (red), LAMP1 (green) and DAPI (white), showing decreased colocalization of HLA-DMB with LAMP1 in cytokine treated islets shown with yellow arrow heads.



## References

1. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15-21
2. Breese MR, Liu Y: NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics* 2013;29:494-496
3. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-140
4. Gonzalez-Duque S, Azoury ME, Colli ML, Afonso G, Turatsinze JV, Nigi L, Lalanne AI, Sebastiani G, Carre A, Pinto S, Culina S, Corcos N, Bugliani M, Marchetti P, Armanet M, Diedisheim M, Kyewski B, Steinmetz LM, Buus S, You S, Dubois-Laforgue D, Larger E, Beressi JP, Bruno G, Dotta F, Scharfmann R, Eizirik DL, Verdier Y, Vinh J, Mallone R: Conventional and Neo-antigenic Peptides Presented by beta Cells Are Targeted by Circulating Naive CD8+ T Cells in Type 1 Diabetic and Healthy Donors. *Cell Metab* 2018;28:946-960 e946
5. Russell MA, Redick SD, Blodgett DM, Richardson SJ, Leete P, Krogvold L, Dahl-Jorgensen K, Bottino R, Brissova M, Spaeth JM, Babon JAB, Haliyur R, Powers AC, Yang CX, Kent SC, Derr AG, Kucukural A, Garber MG, Morgan NG, Harlan DM: HLA Class II Antigen Processing and Presentation Pathway Components Demonstrated by Transcriptome and Protein Analyses of Islet beta-Cells From Donors With Type 1 Diabetes. *Diabetes* 2019;68:988-1001
6. Grant CE, Bailey TL, Noble WS: FIMO: scanning for occurrences of a given motif. *Bioinformatics* 2011;27:1017-1018
7. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-2504
8. Soille P, Vincent L: Determining watersheds in digital pictures via flooding simulations. *SPIE*, 1990
9. McQuin C, Goodman A, Chernyshev V, Kamensky L, Cimini BA, Karhohs KW, Doan M, Ding L, Rafelski SM, Thirstrup D, Wiegand W, Singh S, Becker T, Caicedo JC, Carpenter AE: CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol* 2018;16:e2005970
10. Ding F, Elowitz MB: Constitutive splicing and economies of scale in gene expression. *Nat Struct Mol Biol* 2019;26:424-432
11. Trcek T, Lionnet T, Shroff H, Lehmann R: mRNA quantification using single-molecule FISH in *Drosophila* embryos. *Nat Protoc* 2017;12:1326-1348
12. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36
13. Maynard KR, Tippani M, Takahashi Y, Phan BN, Hyde TM, Jaffe AE, Martinowich K: dotdotdot: an automated approach to quantify multiplex single molecule fluorescent in situ hybridization (smFISH) images in complex tissues. *Nucleic Acids Res* 2020;48:e66