SUPPLEMENTARY DATA

Supplementary Table S1. Site directed mutation of signal peptide of preproinsulin

Designation	Sequence
Wild type PPI SP sequence	MALWMRLLPLLALLALWGPDPAAA
9L mutant	MALWMRLLLLLLLLLWGPDPAAA
12L mutant	MALWMRLLLLLLLLLWGPDPAAA
15L mutant	MALWMRLLLLLLLLLWGPDPAAA
9L/12L mutant	MALWMRLLLLLLLLLWGPDPAAA
9L/15L mutant	MALWMRLLLLLLLLLWGPDPAAA
12L/15L mutant	MALWMRLLLLLLLLLWGPDPAAA
9L/12L/15L mutant	MALWMRLLLLLLLLLWGPDPAAA

SUPPLEMENTARY DATA

Supplementary Table S2. Details of subjects studied for pancreatic expression of antigen processing enzymes

Case ID	Group	Source	Donor Status	Age (years)	Duration
485 88	No diabetes control	UK Pancreas Biobank	Autopsy	2	
315 89	No diabetes control	UK Pancreas Biobank	Autopsy	9	
65 71	No diabetes control	UK Pancreas Biobank	Autopsy	40	
8579	No diabetes control	UK Pancreas Biobank	Autopsy	7	
540 91	No diabetes control	UK Pancreas Biobank	Autopsy	11	
PAN8	No diabetes control	UK Pancreas Biobank	Autopsy	19	
E560	Type 1 diabetes	UK Type 1 diabetes Biobank	Organ Donor	42	1.5y
Sc115	Type 1 diabetes	UK Type 1 diabetes Biobank	Autopsy	1	0 'Recent'
E124B	Type 1 diabetes	UK Type 1 diabetes Biobank	Autopsy	17	0 'Recent'
E375	Type 1 diabetes	UK Type 1 diabetes Biobank	Autopsy	11	unknown
11746	Type 1 diabetes	UK Type 1 diabetes Biobank	Autopsy	6	1 week
11713	Type 1 diabetes	UK Type 1 diabetes Biobank	Autopsy	3	3 months

SUPPLEMENTARY DATA

Supplementary Table S3. Details of the staining protocol for each antibody employed in the current study. Where multiple antigens were stained on the same section, antibodies were applied sequentially (staining order: anti-TAP1 (overnight at 4°C) followed by anti-HLA-ABC (1h at room temperature) or anti-glucagon (1h at room temperature) then anti-insulin plus DAPI (1h at room temperature)).

Primary Antibody	Manufacturer and clone	Antigen Retrieval	Antibody Dilution	Incubation time with primary antibody	Secondary Detection System
HLA-ABC	Abcam C#ab70328 Mouse monoclonal EMR8-5	10mM citrate pH6.0	1/1000	1h at RT	Immunofluoroscence staining using anti-mouse IgG (H+L) Alexa Fluor™-conjugated secondary antibodies (1/400 for 1h)
Insulin	DakoC#A0564 Guinea-pig polyclonal	10mM citrate pH6.0	1/700	1h at RT	Immunofluoroscence staining using anti-guinea-pig IgG (H+L) Alexa Fluor™ -conjugated secondary antibodies (1/400 for 1h)
Glucagon	AbcamC#ab92517 Rabbitmonoclonal EP3070	10mM citrate pH6.0	1/4000	1h at RT	Immunofluoroscence staining using anti-rabbit IgG (H+L) Alexa Fluor™-conjugated secondary antibodies (1/400 for 1h)
TAP1	Protein Tech C#11114-1-AP Rabbit polyclonal	10mM citrate pH6.0	1/200	o/n 4^{o}C	Immunofluoroscence staining using anti-rabbit IgG (H+L) Alexa Fluor™ -conjugated secondary antibodies (1/400 for 1h)
ERAP	R & D Systems C#AF2334 Goat polyclonal	10mM Tris, 1mM EDTA pH9.0	1/200	o/n 4^{o}C	Dako REAL™ Envision™ Detection System with anti-goat IgG HRP-conjugated secondary antibodies (1/800 for 1h).

SUPPLEMENTARY DATA

Supplementary Table S4. Oligonucleotides used for CRISPR-mediated gene knockout.

	Forward	**Reverse**
Exon 2	GAAGCGGGCGGCATCCCGGCGTTTT	GCCGGGATGCCGCCCGCTTCCGGTG
Exon 3	GCAGGAGGTTGATGTACTCCGTTTT	GGAGTACATCAACCTCCTGCCGGTG
Exon 10	GCAGCCTACATCTTCGGCCTGTTTT	AGGCCGAAGATGTAGGCTGCCGGTG
Exon 11	CAGGACAGGAAAACCGATGCGTTTT	GCATCGGTTTTCCTGTCCTGCGGTG

SUPPLEMENTARY DATA

Supplementary Table S5. Eluted epitopes are not selectively derived from SP containing source protein nor their signal peptide region.

	HLA-A*02:01	HLA-A*11:01	HLA-A*24:02	HLA-B*18:01	HLA-B*38:01	HLA-B*39:06	All HLA	Human proteome
Source proteins	**458**	**667**	**190**	**510**	**378**	**259**	**2462**	**20226**
Number of source proteins containing a signal peptide	**26**	**29**	**17**	**25**	**35**	**21**	**153**	**3560**
Frequency of signal peptide containing source proteins	**5.68%**	**4.35%**	**8.95%**	**4.90%**	**9.26%**	**8.11%**	**6.21%**	**17.6%**
Number of proteins with signal peptide-derived epitope	**12**	**2**	**4**	**5**	**4**	**13**	**40**	
Frequency of identified SP epitopes from SP-Protein	**46.15%**	**6.90%**	**23.53%**	**20.00%**	**11.43%**	**61.90%**	**26.14%**	

Source Proteins are those represented by at least one epitope with Mascot Score >40 within our HLA class I elution data (1). Uniprot (2) (access via http://www.uniprot.org/) retrieve/ID mapping function was used to identify signal peptide containing source proteins. SP, signal peptide.

SUPPLEMENTARY DATA

Supplementary Table S6. Putative PPI epitopes identified by in silico binding prediction algorithms.

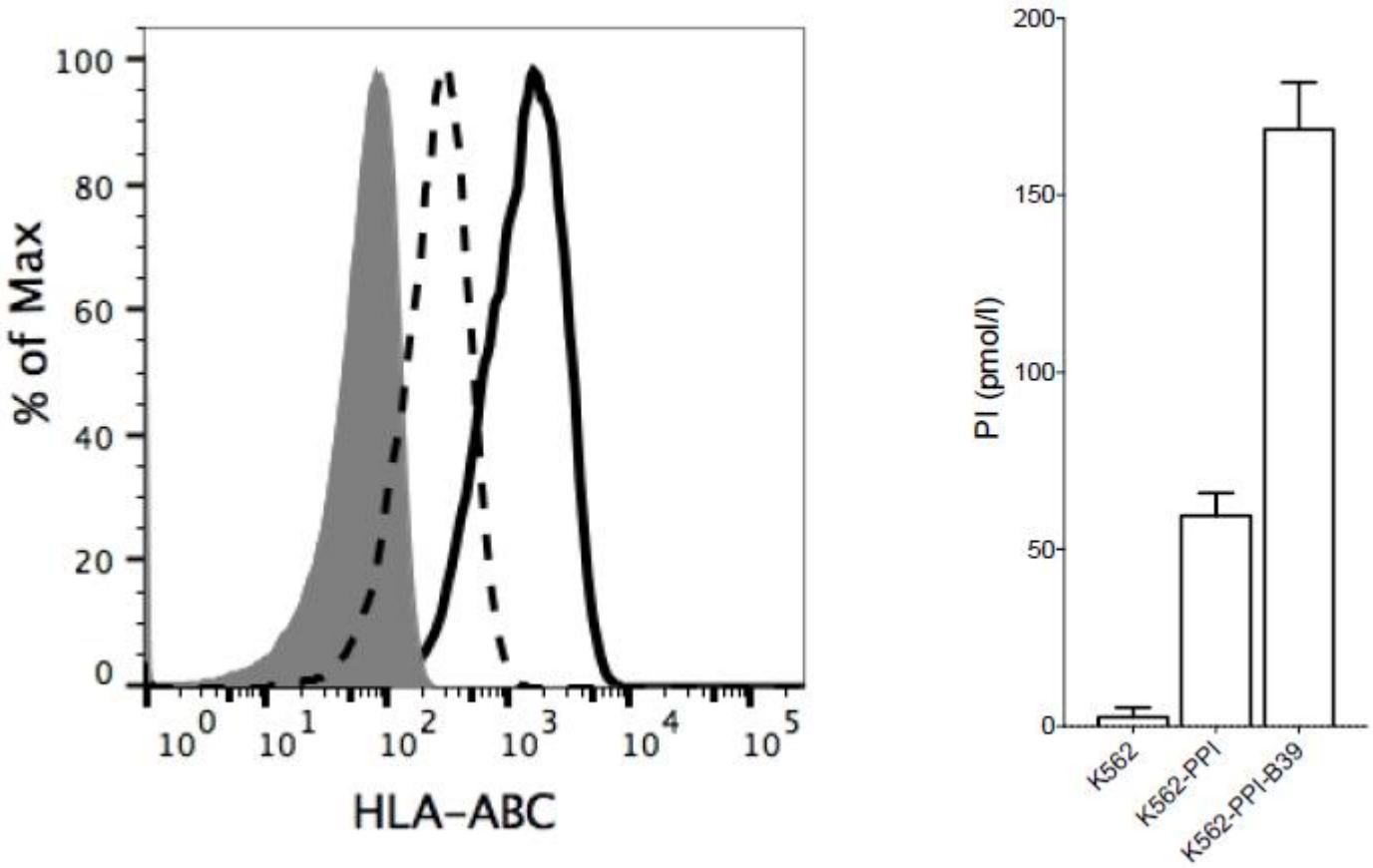
Binding prediction were performed by SYFPEITHI (3) (access via : www.syfpeithi.de)) to identify putative PPI epitopes to compare eluted epitopes to those identified in silico. Epitopes derived from the PPI signal peptide are highlighted in bold and those identified by elution in green.

	A2	9-mer	Score	A11	9-mer	Score	A24	9-mer	Score	B18	8-mer	Score	B38	9-mer	Score
1	6	**RLLPLLALL**	31	38	ALYLVCGER	18	6	**RLLPLLALL**	15	34	HLVEALYL	15	33	SHLVEALYL	21
2	2	**ALWMRLLPL**	28	81	ALEGSLQKR	18	17	**WGPDPAAAF**	13	36	VEALYLVC	14	5	**MRLLPLLAL**	18
3	34	HLVEALYLV	27	35	LVEALYLVC	16	39	LYLVCGERG	13	56	REAEDLQV	14	6	**RLLPLLALL**	14
4	60	DLQVGQVEL	25	75	GSLQPLALE	16	97	TSICSLYQL	13	61	LQVGQVEL	13	3	**LWMRLLPLL**	13
5	15	**ALWGPDPAA**	22	76	SLQPLALEG	16	3	**LWMRLLPLL**	12	32	GSHLVEAL	12	2	**ALWMRLLPL**	12
6	8	**LPLLALLAL**	20	32	GSHLVEALY	15	5	**MRLLPLLAL**	12	42	VCGERGFF	12	8	**LPLLALLAL**	12
7	3	**LWMRLLPLL**	19	80	LALEGSLQK	15	8	**LPLLALLAL**	12	58	AEDLQVGQ	12	31	CGSHLVEAL	12
8	5	**MRLLPLLAL**	19	62	QVGQVELGG	14	22	**AAAFVNQHL**	12	3	**LWMRLLPL**	11	60	DLQVGQVEL	12
9	81	ALEGSLQKR	19	84	GSLQKRGIV	14	31	CGSHLVEAL	12	6	**RLLPLLAL**	11	72	PGAGSLQPL	12
10	7	**LLPLLALLA**	18	85	SLQKRGIVE	14	33	SHLVEALYL	12	9	**PLLALLAL**	11	74	AGSLQPLAL	12
11	76	SLQPLALEG	18	96	CTSICSLYQ	14	49	FYTPKTRRE	12	33	SHLVEALY	11	97	TSICSLYQL	12
12	22	**AAAFVNQHL**	17	2	**ALWMRLLPL**	13	78	QPLALEGSL	12	41	LVCGERGF	11	28	QHLCGSHLV	11
13	69	GGGPGAGSL	17	4	**WMRLLPLLA**	13	2	**ALWMRLLPL**	11	44	GERGFFYT	11	69	GGGPGAGSL	11
14	97	TSICSLYQL	17	7	**LLPLLALLA**	13	53	KTRREAEDL	11	66	VELGGGPG	11	78	QPLALEGSL	11
15	9	**PLLALLALW**	16	47	GFFYTPKTR	13	69	GGGPGAGSL	11	75	GSLQPLAL	11	94	QCCTSICSL	11
16	10	**LLALLALWG**	16	10	**LLALLALWG**	12	74	AGSLQPLAL	11	82	LEGSLQKR	11	22	**AAAFVNQHL**	10
17	31	CGSHLVEAL	16	29	HLCGSHLVE	12	94	QCCTSICSL	11	92	VEQCCTSI	11	91	IVEQCCTSI	10
18	58	AEDLQVGQV	16	60	DLQVGQVEL	12	27	NQHLCGSHL	10	96	CTSICSLY	11	17	**WGPDPAAAF**	9
19	94	QCCTSICSL	16	98	SICSLYQLE	12	40	YLVCGERGF	10	1	**MALWMRLL**	10	27	NQHLCGSHL	9
20	12	**ALLALWGPD**	15	12	**ALLALWGPD**	11	41	LVCGERGFF	10	18	**GPDPAAAF**	10	41	LVCGERGFF	9
21	13	**LLALWGPDP**	15	41	LVCGERGFF	11	60	DLQVGQVEL	10	43	**CGERGFFY**	10	53	KTRREAEDL	9
22	18	**GPDPAAAFV**	15	45	ERGFFYTPK	11	72	PGAGSLQPL	10	54	TRREAEDL	10	83	EGSLQKRGI	9
23	28	QHLCGSHLV	15	50	YTPKTRREA	11	83	EGSLQKRGI	10	101	SLYQLENY	10	40	YLVCGERGF	8
24	33	SHLVEALYL	15	53	KTRREAEDL	11	91	IVEQCCTSI	10	4	**WMRLLPLL**	9	55	RREAEDLQV	8
25	38	ALYLVCGER	15	97	TSICSLYQL	11	102	LYQLENYCN	10	7	**LLPLLALL**	9	54	TRREAEDLQ	7

	A2	10-mer	Score	A11	10-mer	Score	A24	10-mer	Score	B18	9-mer	Score	B38	10-mer	Score
1	2	**ALWMRLLPLL**	28	79	PLALEGSLQK	21	39	LYLVCGERGF	22	100	CSLYQLENY	15	5	**MRLLPLLALL**	19
2	7	**LLPLLALLAL**	28	75	GSLQPLALEG	20	5	**MRLLPLLALL**	14	60	DLQVGQVEL	14	59	EDLQVGQVEL	14
3	4	**WMRLLPLLAL**	24	84	GSLQKRGIVE	19	26	VNQHLCGSHL	13	31	CGSHLVEAL	13	7	**LLPLLALLAL**	13
4	29	HLCGSHLVEA	24	32	GSHLVEALYL	18	1	**MALWMRLLPL**	12	44	GERGFFYTP	13	73	GAGSLQPLAL	13
5	15	**ALWGPDPAAA**	22	6	**RLLPLLALLA**	15	2	**ALWMRLLPLL**	12	56	REAEDLQVG	13	1	**MALWMRLLPL**	12
6	1	**MALWMRLLPL**	21	35	LVEALYLVCG	15	7	**LLPLLALLAL**	12	8	**LPLLALLAL**	12	2	**ALWMRLLPLL**	12
7	90	GIVEQCCTSI	21	85	SLQKRGIVEQ	15	30	LCGSHLVEAL	12	36	VEALYLVCG	12	4	**WMRLLPLLAL**	12
8	85	SLQKRGIVEQ	20	98	SICSLYQLEN	15	59	EDLQVGQVEL	12	58	AEDLQVGQV	12	32	GSHLVEALYL	12
9	6	**RLLPLLALLA**	19	44	**GERGFFYTPK**	14	77	LQPLALEGSL	12	74	AGSLQPLAL	12	33	SHLVEALYLV	12
10	10	**LLALLALWGP**	19	9	**PLLALLALWG**	13	90	GIVEQCCTSI	12	78	QPLALEGSL	12	68	LGGGPGAGSL	12
11	33	SHLVEALYLV	19	97	TSICSLYQLE	13	16	**LWGPDPAAAF**	11	82	LEGSLQKRG	12	93	EQCCTSICSL	12
12	76	SLQPLALEGS	19	29	HLCGSHLVEA	12	21	**PAAAFVNQHL**	11	3	**LWMRLLPLL**	11	26	VNQHLCGSHL	11
13	96	CTSICSLYQL	19	47	GFFYTPKTRR	12	49	FYTPKTRREA	11	5	**MRLLPLLAL**	11	28	QHLCGSHLVE	11
14	5	**MRLLPLLALL**	18	53	KTRREAEDLQ	12	52	PKTRREAEDL	11	6	**RLLPLLALL**	11	30	LCGSHLVEAL	11
15	13	**LLALWGPDPA**	18	2	**ALWMRLLPLL**	11	73	GAGSLQPLAL	11	32	GSHLVEALY	11	52	PKTRREAEDL	11
16	30	LCGSHLVEAL	17	12	**ALLALWGPDP**	11	93	EQCCTSICSL	11	42	VCGERGFFY	11	71	GPGAGSLQPL	11
17	73	GAGSLQPLAL	17	15	**ALWGPDPAAA**	11	40	YLVCGERGFF	10	95	CCTSICSLY	11	96	CTSICSLYQL	11
18	68	LGGGPGAGSL	16	34	HLVEALYLVC	11	68	LGGGPGAGSL	10	9	**PLLALLALW**	10	16	**LWGPDPAAAF**	10
19	80	LALEGSLQKR	16	41	LVCGERGFFY	11	71	GPGAGSLQPL	10	27	NQHLCGSHL	10	21	**PAAAFVNQHL**	10
20	12	**ALLALWGPDP**	15	100	CSLYQLENYC	11	82	LEGSLQKRGI	10	40	YLVCGERGF	10	77	LQPLALEGSL	10
21	57	EAEDLQVGQV	15	4	**WMRLLPLLAL**	10	96	CTSICSLYQL	10	41	LVCGERGFF	10	40	YLVCGERGFF	9
22	71	GPGAGSLQPL	15	7	**LLPLLALLAL**	10	4	**WMRLLPLLAL**	9	66	VELGGGPGA	10	54	TRREAEDLQV	9
23	14	**LALWGPDPAA**	14	25	FVNQHLCGSH	10	32	GSHLVEALYL	9	2	**ALWMRLLPL**	9	39	LYLVCGERGF	8
24	77	LQPLALEGSL	14	37	EALYLVCGER	10	24	**AFVNQHLCGS**	6	17	**WGPDPAAAF**	9	45	ERGFFYTPKT	8
25	101	SLYQLENYCN	14	46	RGFFYTPKTR	10	42	VCGERGFFYT	6	53	KTRREAEDL	9	55	RREAEDLQVG	8

SUPPLEMENTARY DATA

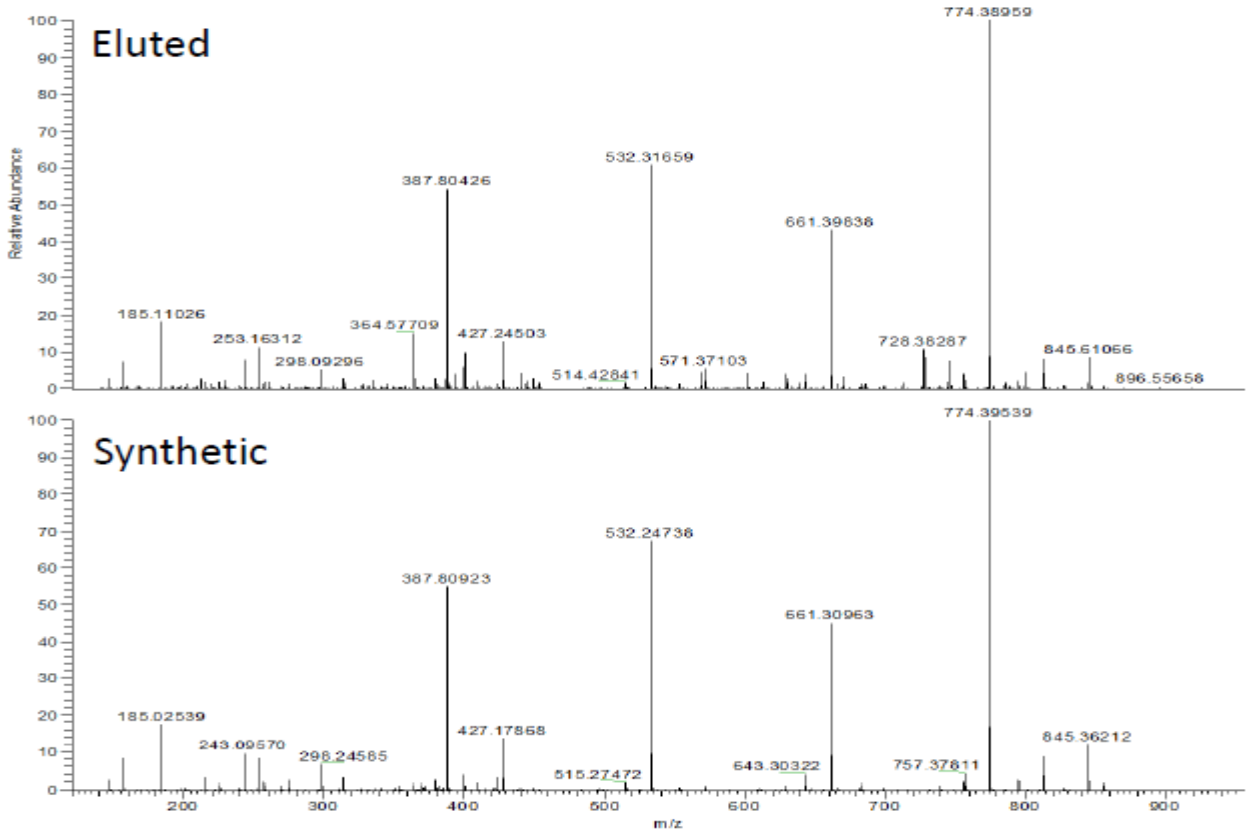
Supplementary Figure S1. Representative Expression of HLA class I and secretion of proinsulin. Expression of HLA class I and secretion of proinsulin by the generated K562-PPI (dashed line) and K562-PPI-HLA (solid line) cell lines (isotype control grey shaded). Single cell cloning was performed to select for the K562-PPI-HLA cell line with best combination of HLA class I expression (left) and Proinsulin expression (right). Data is representative, with B*3906 shown, and similar data obtained for A*11:01, B*1801 and B*38:01.



SUPPLEMENTARY DATA

Supplementary Figure S2. PPI epitope discovery Tandem mass spectrometry analysis of collision-induced dissociation revealing the tandem mass spectrum of a PPI peptide (A) HLA-A*1101 Epitope PPI_{80-88} (LALEGSLQQK), (B) HLA-B*3801 Epitope PPI_{5-14} (MRLLPLLALL) and (C) HLA-B*3801 Epitope PPI_{33-41} (SHLVEALYL) . The correct identity of the peptide was proven by tandem mass spectrometry of the synthetic compound. The table lists the amino acid sequence of the peptide with the expected b- and y-fragment ions (fragment ions extending from the amino- and carboxyl terminus respectively). Observed fragment ions are underlined.

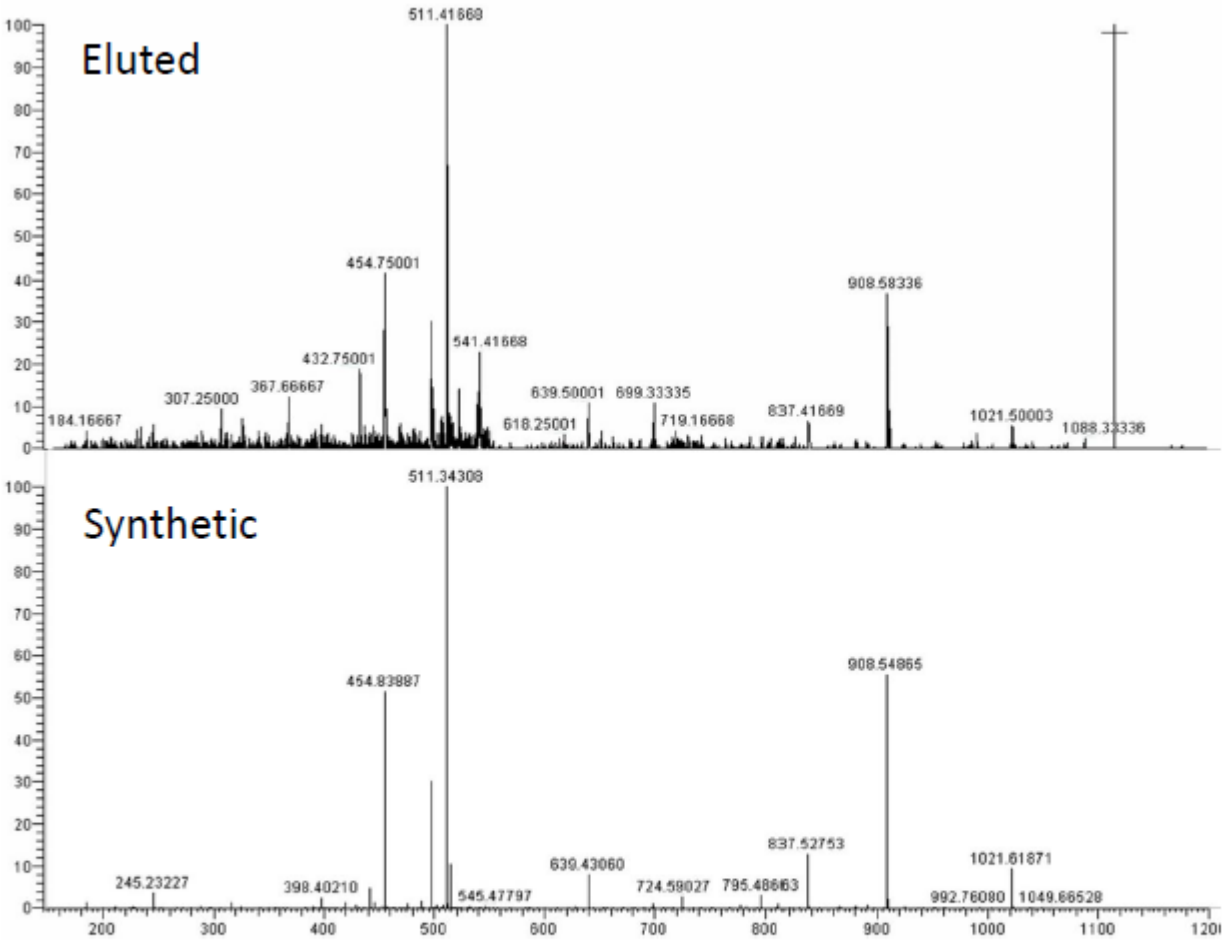
Supplementary Figure 2 A) HLA-A*1101 Epitope PPI_{80-88}



b_n ions	114.1	185.1	298.2	427.3	484.3	571.3	684.4	812.5	958.6
Peptide	L	A	L	E	G	S	L	Q	K
y_n ions	958.6	845.5	774.4	661.4	523.3	475.3	388.3	275.2	147.1

SUPPLEMENTARY DATA

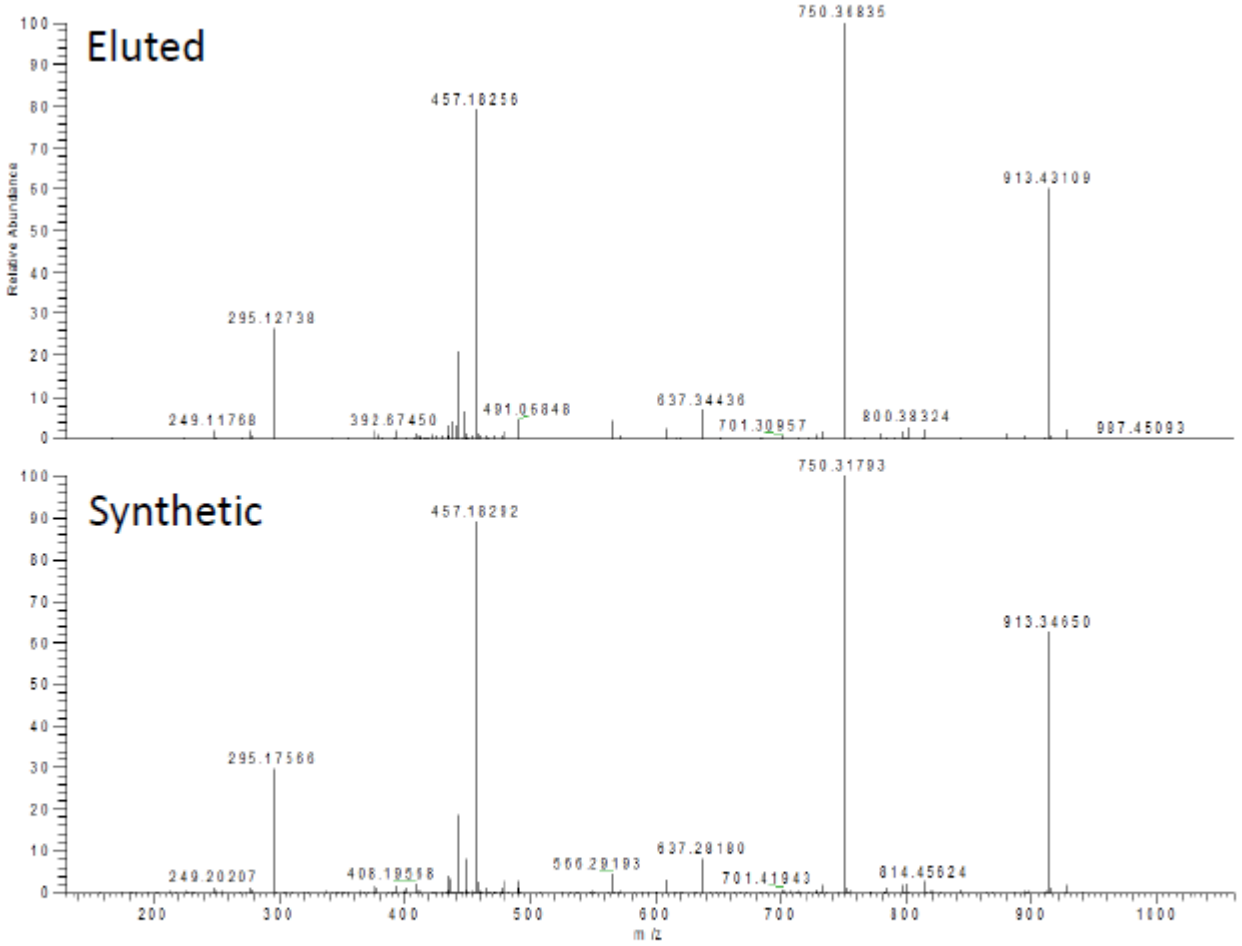
Supplementary Figure 2 B) HLA-B*3801 Epitope PPI_{5-14}



b_n ions	132	288.1	401.2	514.3	611.4	724.5	837.5	908.6	1021.7	1134.7
Peptide	M	R	L	L	P	L	L	A	L	L
y_n ions	1177.8	1021.7	865.6	752.5	639.4	542.4	429.3	316.2	245.2	132.1

SUPPLEMENTARY DATA

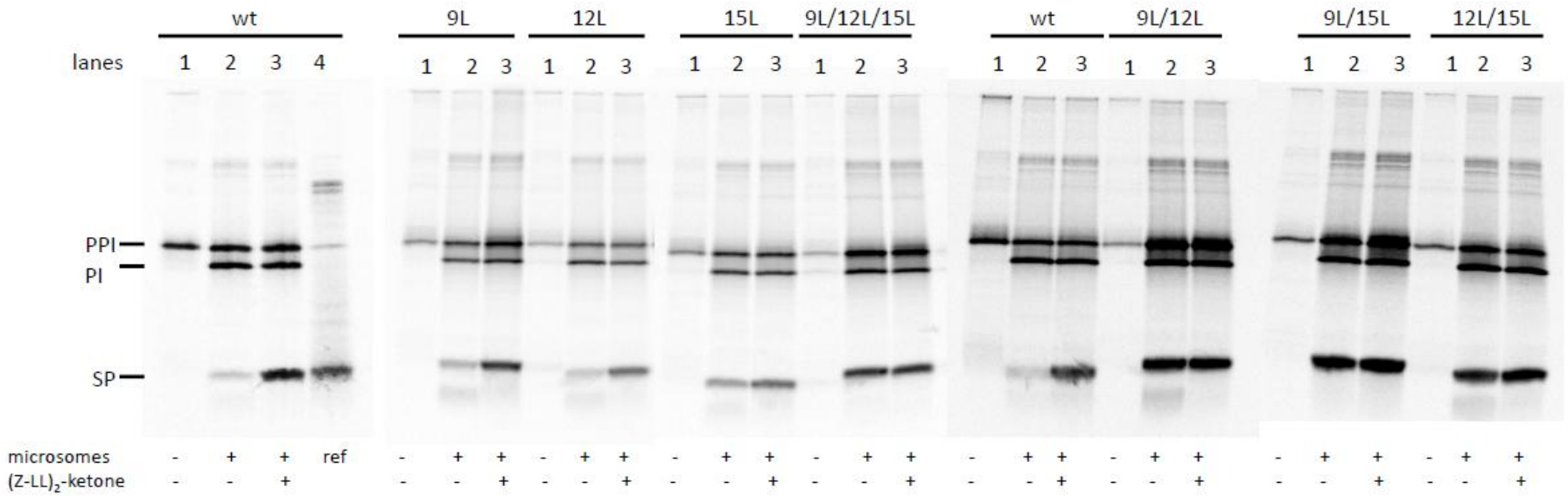
Supplementary Figure 2 C) HLA-B*3801 Epitope PPI_{33-41}



b_n ions	88	225.1	338.2	437.3	566.3	637.3	750.4	913.5	1044.6
Peptide	S	H	L	V	E	A	L	Y	L
y_n ions	1044.6	957.5	820.5	707.4	608.3	479.3	408.2	295.2	132.1

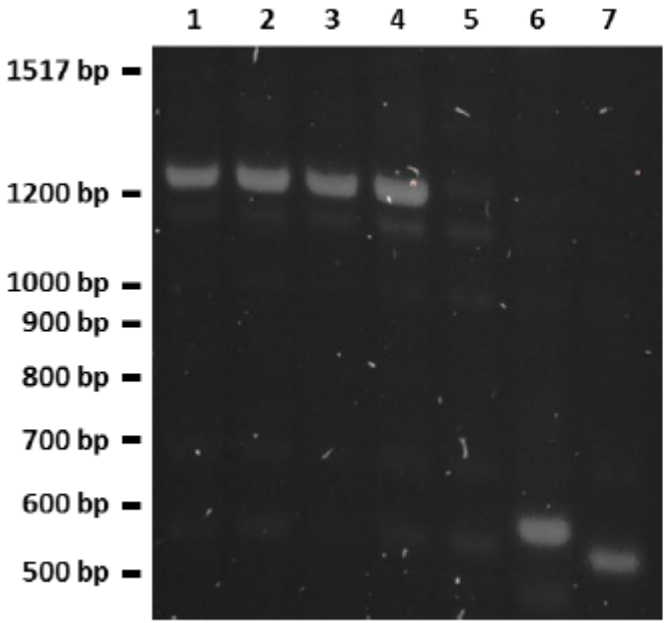
SUPPLEMENTARY DATA

Supplementary Figure S3. Processing of the PPI signal peptide in microsomes. *In vitro* translation of wt PPI mRNA or mutant PPI mRNA (P9L, A12L, A15L) in the absence (lanes 1) or in the presence of ER-derived microsomes (lanes 2 and 3) and SPP inhibitor $(Z\text{-}LL)_2$-ketone (lanes 3). Microsomes were isolated and analyzed by SDS-PAGE, and radiolabeled proteins visualized by phosphorimaging. Lane 4, *in vitro*-translated reference. Images are representative images of n=2, apart from wt where n=5. Equal translocation efficiency and PPI precursor availability for processing was controlled by comparing the amount of proinsulin between conditions.



SUPPLEMENTARY DATA

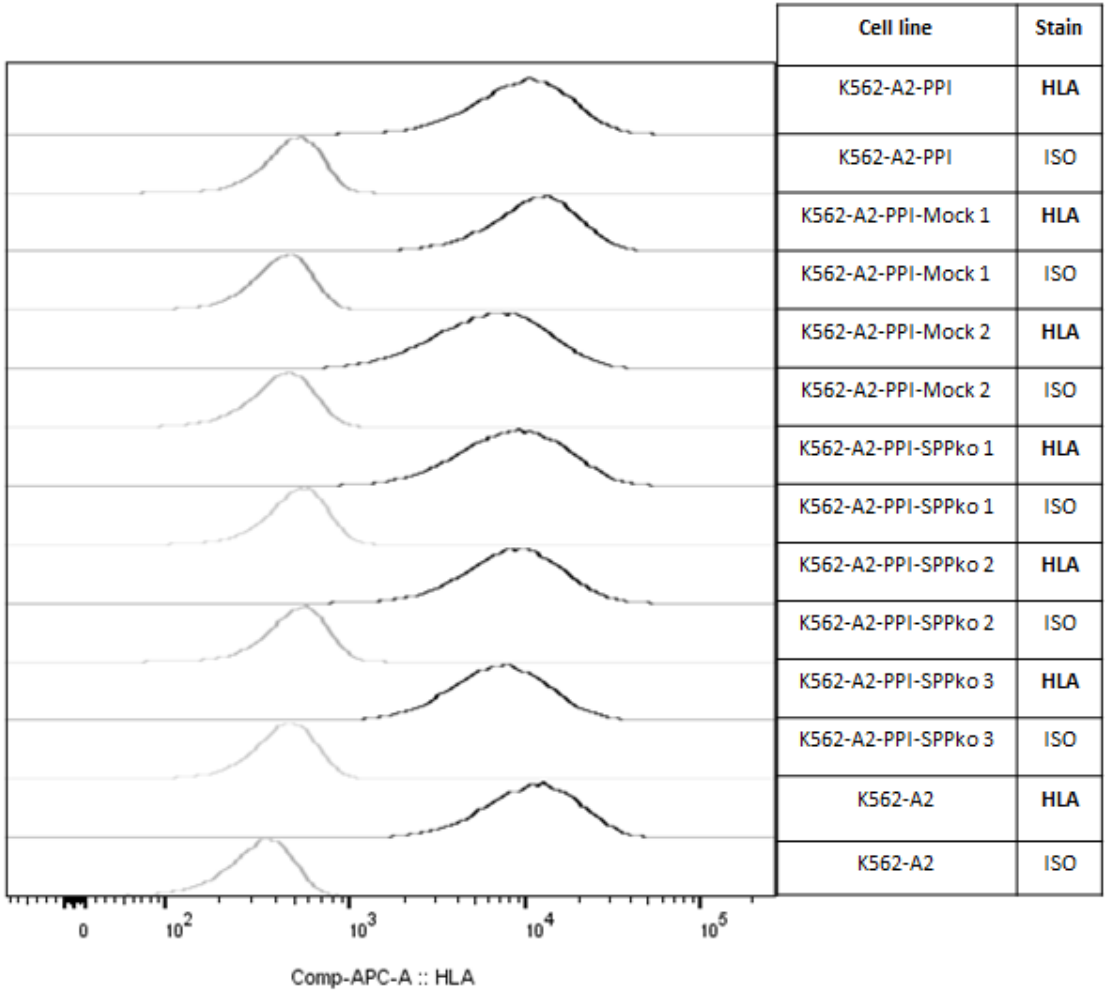
Supplementary Figure S4. Validation of SPP knockout (truncation) by PCR. Within the K562-A2-PPI cells line the SPP gene was truncated using CRISPR-Cas9 directed double targeting of each exon at the start and end of the gene. Simultaneous targeting leads to truncation of the gene in effect leading to functional gene knockout. Length of the gene was assessed using PCR with SPP specific primers on cDNA generated from isolated mRNA from the cell lines. WT (lane 1 and 2) and mock transfected (lane 3 and 4) cell lines harbour full length SPP, whereas reduction (lane 6 and 7) in size or abrogation (lane 5) was observed for cell lines with effective CRISPR-Cas9 targeting of SPP.



Lane	Cell line
1	K562-A2
2	K562-A2-PPI
3	K562-A2-PPI-Mock 1
4	K562-A2-PPI-Mock 2
5	K562-A2-PPI-SPPko 1
6	K562-A2-PPI-SPPko 2
7	K562-A2-PPI-SPPko 3

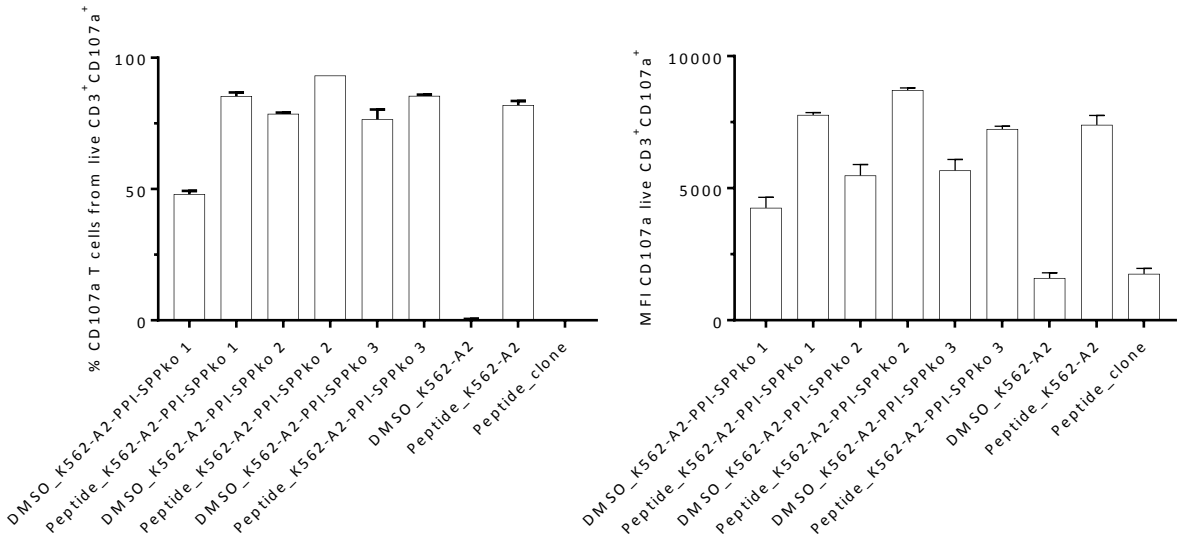
SUPPLEMENTARY DATA

Supplementary Figure S5. HLA class I surface expression on Mock and SPPko cell lines. Cell lines were stained with anti-pan HLA class I antibody (HLA, black line) and corresponding isotype control (ISO, grey line) antibody. Levels of surface HLA expression are similar for each experimental condition (WT, mock and SPP knockout).



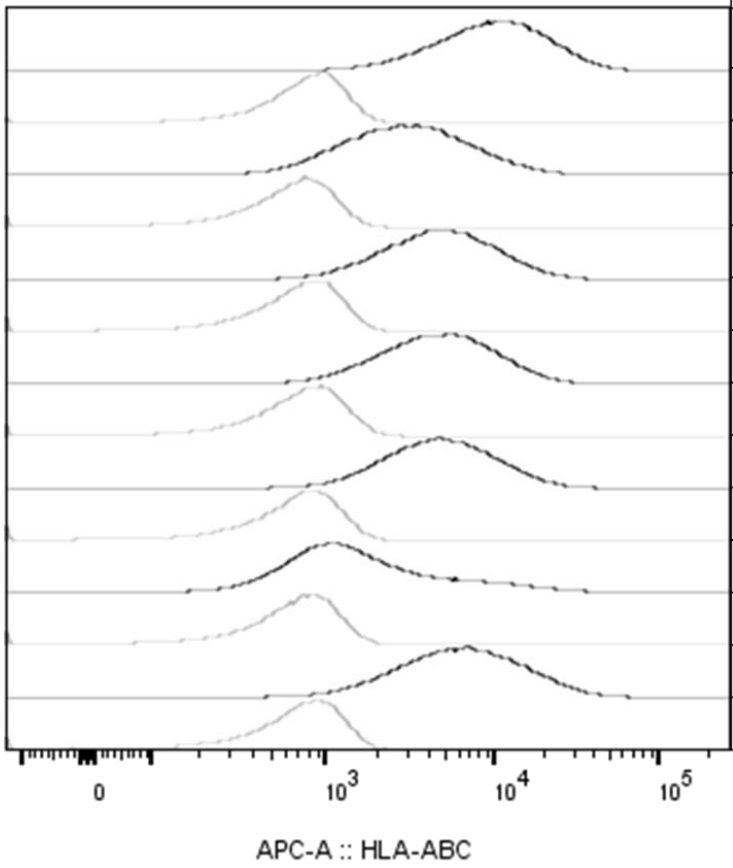
SUPPLEMENTARY DATA

Supplementary Figure S6. Pulsing SPPko cells with cognate peptide rescues their phenotype. K562-A2 and K562-A2-PPI-SPPko cell lines were pulsed for 1 hour with 10uM/ml peptide or peptide diluent (DMSO) prior to co-culture with T cell clone specific for PPI_{15-24}-HLA-A0201. T cell activation was assessed by CD107a expression. For each SPPko cell line, peptide pulsing leads to increased T cell activation as evidenced by both increases in percentage expression of CD107a (left) and median fluorescence intensity of CD107a (right) with comparable levels in the pulsed WT K562-A2 cell line. Three independent experiments.



SUPPLEMENTARY DATA

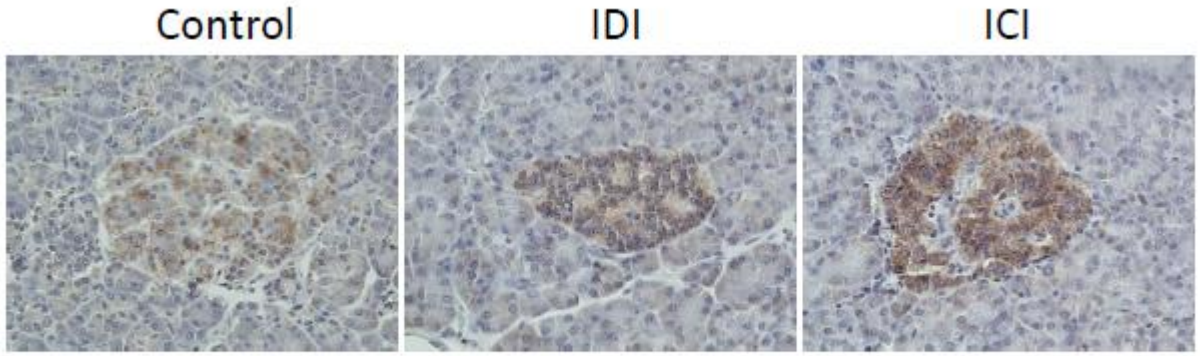
Supplementary Figure S7. HLA class I expression in siRNA experiments. Surface HLA-ABC (W6/32 clone) expression on b2M, ERAP1, ERAP2, TAP1, TAP2, scramble knockdown treated and untreated K562-A24-PPI respectively (HLA, black line) and isotype control (ISO, grey line). Median Fluorescence Intensity (MFI) of HLA staining is shown.



siRNA	Stain	MFI
Scramble	**HLA**	**9774**
Scramble	ISO	797
TAP1	**HLA**	**2909**
TAP1	ISO	655
TAP2	**HLA**	**4683**
TAP2	ISO	715
ERAP1	**HLA**	**4559**
ERAP1	ISO	740
ERAP2	**HLA**	**4694**
ERAP2	ISO	702
B2m	**HLA**	**1348**
B2m	ISO	691
Untreated	**HLA**	**8231**
Untreated	ISO	738

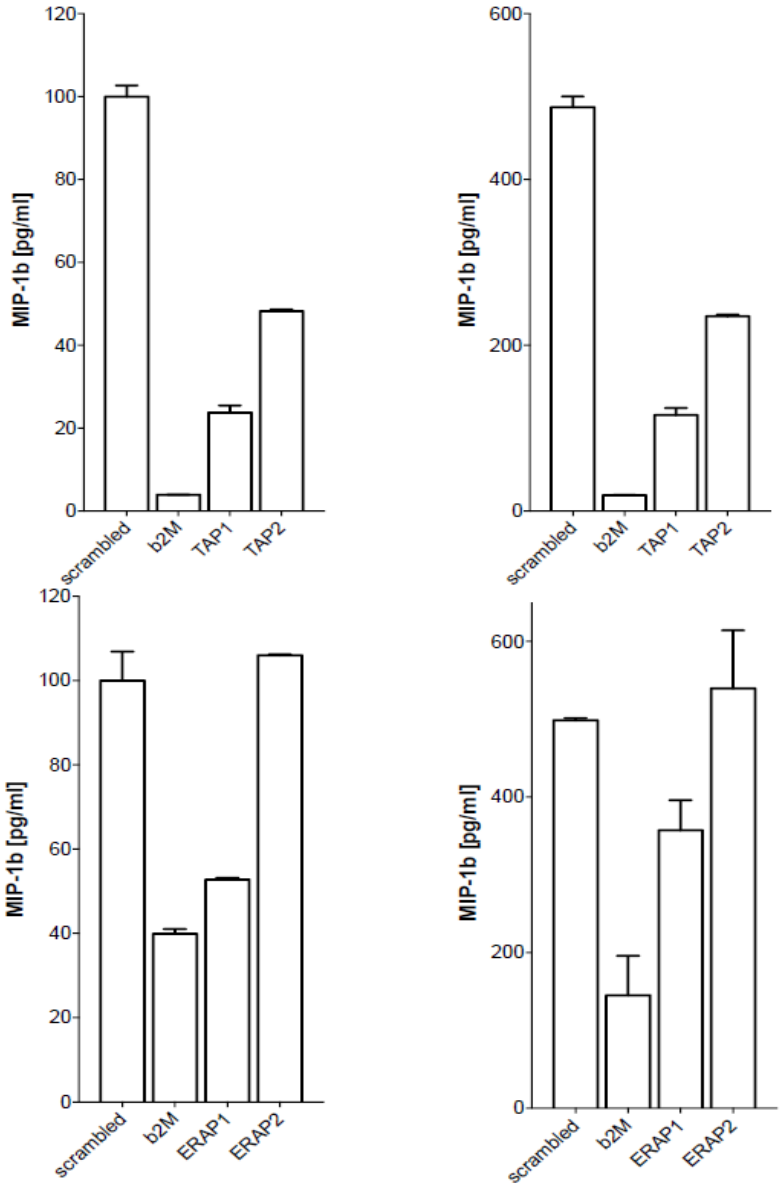
SUPPLEMENTARY DATA

Supplementary Figure S8. Representative immunohistochemistry staining of ERAP. Representative immunohistochemistry staining of ERAP on pancreas samples from patients with type 1 diabetes (n=3) and a representative control sample. ERAP is expressed in all islet cells irrespective of sample source and pancreas samples from type 1 diabetes patients show similar expression in insulin-deficient islets (IDI) and insulin-containing islets (ICI). These data suggest that ERAP expression is not significantly altered in the islets of patients with type 1 diabetes.



Supplementary Figure S9. MIP-1β in siRNA experiments. MIP-1β production of HLA-A2402-restricted PPI_{3-11} specific CD8 T cell clone 4C6 upon co-culture with *β2M*, *TAP1*, *TAP2*, genes (top panel) and *ERAP1*, *ERAP2* gene (bottom panel) knockdown in K562-A24-PPI cells. Two independent experiments.

SUPPLEMENTARY DATA



SUPPLEMENTARY DATA

References

1. Eichmann M, de Ru A, van Veelen PA, Peakman M, Kronenberg-Versteeg D: Identification and characterisation of peptide binding motifs of six autoimmune disease-associated human leukocyte antigen-class I molecules including HLA-B*39:06. Tissue antigens 2014;84:378-388
2. The UniProt C: UniProt: the universal protein knowledgebase. Nucleic Acids Res 2017;45:D158-D169
3. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S: SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 1999;50:213-219