ONLINE APPENDIX – Azoury ME, Tarayrah M et al.

Peptides derived from insulin granule proteins are targeted by CD8$^+$ T cells across MHC Class I restrictions in humans and NOD mice.

Study phase	Group	Case ID	Age (yrs)	Gender (M/F)	T1D duration (wks)	GADA	IA-2A	ZnT8A	Therapy
Screening	Healthy (n=7)	H002N	44	M	NA	ND	ND	ND	NA
		H004N	44	M	NA	ND	ND	ND	NA
		H156C	25	F	NA	ND	ND	ND	NA
		H423C	25	F	NA	ND	ND	ND	NA
		H458C	33	F	NA	ND	ND	ND	NA
		H459C	25	M	NA	ND	ND	ND	NA
		H478C	32	F	NA	ND	ND	ND	NA
			32 (24-47)	42.9% M 57.1% F					
Validation	T1D (n=9)	D286T	38	F	0.3	+	+	+	insulin
		D309V	30	F	1.4	+	+	ND	insulin
		D187D	40	F	34.9	+	-	ND	insulin
		D406D	26	F	0.4	+	+	+	insulin
		D396D	25	F	3.4	+	-	+	insulin
		D204D	22	M	2.0	+	-	ND	insulin
		D275T	30	F	6.7	+	ND	ND	insulin
		D242D	23	M	0.1	+	-	ND	insulin
		D404V	27	M	0.3	+	-	ND	insulin
			26 (22-40)	33.4% M 66.6% F	1.4 (0.3-32.0)				
	Healthy (n=9)	H004N	44	M	NA	ND	ND	ND	NA
		H005N	38	F	NA	ND	ND	ND	NA
		H112C	42	F	NA	ND	ND	ND	NA
		H355C	32	M	NA	ND	ND	ND	NA
		H423C	25	M	NA	ND	ND	ND	NA
		H449C	20	F	NA	ND	ND	ND	NA
		H458C	33	M	NA	ND	ND	ND	NA
		H459C	25	F	NA	ND	ND	ND	NA
		H478C	32	F	NA	ND	ND	ND	NA
			32 (25-44)	44.5% M 55.5% F					

Table S1. Characteristics of HLA-A3$^+$ study participants. The healthy donors analyzed in Fig. 2 (screening phase) and the patients with type 1 diabetes (T1D) and age/sex-matched healthy donors analyzed in Fig. 3 (validation phase) are listed. The distribution of age, gender and T1D duration is shown at the bottom of each list (median and range for numerical variables). GADA, IA-2A, ZnT8A are anti-GAD, -IA-2 and -ZnT8 auto-antibodies, respectively. ND, not determined; NA, not available or not applicable.

Protein/mRNA	Peptide	Sequence	Type	MMr staining pattern	T-cell+ donors	Frequency in CD8+	Donors with ≥5 MMr+ cells	% antigen experienced	Validation
ABCC8	787-795	IIFESPFNK	Conventional	Spread	6/6	8.0×10^{-6}	5/6	14.3	No
****C15orf48-003**	**50-58**	**KVSAPAYCK**	**mRNA splice**	**Clustered**	**6/6**	**1.7×10^{-5}**	**5/6**	**22.7**	**Yes**
C15orf48-003	97-106	ATGRSQMVRK	mRNA splice	Spread	5/5	4.0×10^{-6}	0/5	NA	No
***C15orf48-003**	**103-111**	**MVRKGMNLK**	**mRNA splice**	**Clustered**	**5/5**	**8.3×10^{-6}**	**4/5**	**15.3**	**Yes**
CADPS1;CADPS2	780-788	SLLERVLMK	Conventional	Spread	5/6	4.0×10^{-6}	2/6	NA	No
CHGB	603-613	RVAQLDQLLHY	Conventional	NA	3/6	8.8×10^{-7}	0/6	NA	No
****CTNND1-026**	**852-860**	**QVSYPSMQK**	**mRNA splice**	**Clustered**	**5/5**	**3.4×10^{-5}**	**5/5**	**10.3**	**Yes**
G6PC2-001	317-326	HLFYVLSFCK	mRNA splice	ND	1/6	0	1/6	NA	No
G6PC2-001	343-351	HMLMKQSGK	mRNA splice	Spread	3/5	3.1×10^{-6}	1/5	NA	No
GNAS-036	22-31	KVKKVPLAEK	mRNA splice	Spread	3/5	1.4×10^{-6}	1/5	NA	No
GNAS-036	33-42	RQMRKEALEK	mRNA splice	Clustered	4/5	1.7×10^{-6}	1/5	NA	No
****GNAS-036**	**74-83**	**LLLGAGESGK**	**mRNA splice**	**Clustered**	**6/6**	**2.2×10^{-5}**	**5/6**	**8.7**	**Yes**
GNAS-050	16-24	RSSTRHSGK	mRNA splice	ND	1/5	0	0/5	NA	No
****GNAS-050**	**159-168**	**RLCHLSLMEK**	**mRNA splice**	**Clustered**	**5/5**	**6.2×10^{-6}**	**4/5**	**10.0**	**Yes**
****GNAS-050**	**477-485**	**RLLRGLLAR**	**mRNA splice**	**Clustered**	**5/6**	**1.2×10^{-5}**	**5/6**	**5.6**	**Yes**
GNAS-050	496-504	AAAACTTMK	mRNA splice	Spread	4/6	1.3×10^{-6}	0/6	NA	No
***GNAS-050**	**542-551**	**ATYSATFSCK**	**mRNA splice**	**Clustered**	**4/5**	**2.2×10^{-6}**	**2/5**	**NA**	**Yes**
INS/PTPRN	15-21/893-896	ALWGPDP/KVNK	Peptide splice	Spread	6/6	1.2×10^{-5}	1/6	NA	No
KCNK16	13-21	RVLPLLLAY	Conventional	Spread	6/6	4.4×10^{-6}	2/6	NA	No
KIF1A	153-161	RVRDLLNPK	Conventional	Spread	4/6	2.9×10^{-6}	1/6	NA	No
KIF1A	219-228	AVFNIIFTQK	Conventional	Spread	5/6	4.6×10^{-6}	1/6	NA	No
****KIF1A**	**860-868**	**LLYPVPLVH**	**Conventional**	**Clustered**	**5/6**	**9.9×10^{-6}**	**3/6**	**85.2**	**Yes**
***LARP4-006**	**222-230**	**KAINTFFAK**	**mRNA splice**	**Clustered**	**4/6**	**1.0×10^{-6}**	**2/6**	**NA**	**Yes**
MAP3K13-012	33-41	AMGNHPSPK	mRNA splice	Spread	5/6	1.9×10^{-6}	1/6	NA	No
***MNX1-001**	**217-225**	**STAGMILPK**	**mRNA splice**	**Clustered**	**6/6**	**5.9×10^{-6}**	**2/6**	**NA**	**Yes**
ONECUT1-003	5-14	KSFIKHAKGK	mRNA splice	Spread	5/6	1.4×10^{-6}	1/6	NA	No
***ONECUT1-003**	**47-56**	**KISVHPKNCK**	**mRNA splice**	**Clustered**	**4/6**	**1.7×10^{-6}**	**2/6**	**NA**	**Yes**
PCSK1	706-714	KLNKPSQLK	Conventional	NA	4/6	2.6×10^{-6}	0/6	NA	No
PDIA2-007	86-94	AVHGFPTLK	mRNA splice	Spread	6/6	1.2×10^{-5}	5/6	30.0	No
***PFN2-008**	**59-67**	**RVLVFVMGK**	**mRNA splice**	**Clustered**	**4/6**	**2.2×10^{-6}**	**3/6**	**14.3**	**Yes**
****PNMA2**	**50-58**	**RLLGKIFRK**	**Conventional**	**Clustered**	**6/6**	**1.4×10^{-5}**	**6/6**	**42.5**	**Yes**
PRKACB-008	2-10	GLLKEFLAK	mRNA splice	Spread	4/6	2.0×10^{-6}	1/6	NA	No
PTPRN	156-164	RLPQPPVGK	Conventional	Spread	4/6	2.4×10^{-5}	2/6	NA	No
PTPRN	965-975	AVAEEVNAILK	Conventional	Spread	3/6	2.6×10^{-6}	2/6	NA	No
****PTPRN/PTPRN**	**576-580/708-711**	**SVLLT/RLAK**	**Peptide splice**	**Clustered**	**6/6**	**5.3×10^{-6}**	**5/6**	**33.3**	**Yes**
REXO2-020	2-11	SVANALWIVK	mRNA splice	Spread	6/6	3.9×10^{-6}	3/6	33.3	No
****RTN1**	**120-129**	**STYFTGILQK**	**Conventional**	**Clustered**	**6/6**	**7.4×10^{-6}**	**5/6**	**11.1**	**Yes**
****SCG3**	**166-174**	**AVFDKIVSK**	**Conventional**	**Clustered**	**6/6**	**3.2×10^{-6}**	**2/6**	**NA**	**Yes**
****SCG5-009**	**193-201**	**RLKPSLVGK**	**mRNA splice**	**Clustered**	**6/6**	**7.1×10^{-6}**	**4/6**	**30**	**Yes**
SLC30A8	6-16	RTYLVNDKAAK	Conventional	Clustered	5/6	1.4×10^{-6}	0/6	NA	No
****UCN3**	**46-56**	**GQWEDASLLSK**	**Conventional**	**Clustered**	**5/6**	**4.9×10^{-6}**	**3/6**	**14.3**	**Yes**
***ZNF268-015**	**86-95**	**ILKAGKSKAK**	**mRNA splice**	**Clustered**	**4/6**	**1.8×10^{-6}**	**1/6**	**NA**	**Yes**

Table S2. Summary of islet peptides tested for recognition by circulating naïve CD8$^+$ T cells in healthy subjects. The peptides presented in Fig. 2 are alphabetically listed and classified according to type (conventional, mRNA spliced, spliced peptide). The other columns detail the MMr staining pattern (clustered, spread, not detected [ND], or not available [NA] when too few MMr$^+$ cells were counted), the number of T-cell$^+$ donors ($\geq 5/10^7$ MMr$^+$ cells counted), the median MMr$^+$ frequency, the number of donors with ≥5 MMr$^+$ cells, the median percent antigen-experienced cells within the MMr$^+$ fraction (for peptides with ≥3 donors displaying ≥5 MMr$^+$ cells), and the final validation outcome. The 19 validated islet peptides displaying the expected frequency out of total CD8$^+$ T cells ($\geq 1 \times 10^{-6}$; with ≥5 MMr$^+$ cells counted for at least one donor) and a clustered MMr staining pattern are marked with one asterisk (or two asterisks for the 12 peptides further analyzed) and highlighted in bold.

Protein	Peptide	Sequence	Predicted Kd affinity SYFPEITHI	Predicted Kd affinity NetMHC
Urocortin-3 (Ucn3, UniProt Q924A4)	**5-13**	**TYFLLPLLL**	**23**	**1405**
	32-40	**VFSCLNTAL**	**21**	**1177**
Proconvertase-2 (Pcsk2, UniProt P21661)	109-118	GYRDINEIDI	20	10594
	341-350	**LYDESCSSTL**	**21**	**428**
	501-510	**RYLEHVQAVI**	**24**	**1660**
Secretogranin-5 (Scg5, UniProt P12961)	4-13	RLVSAMLSGL	*14*	19542
	8-17	AMLSGLLFWL	*13*	26446
	26-35	AYSPRTPDRV	18	7158
	193-201	**DNVVAKKSV**	***18***	**9660**
Insulin gene enhancer protein-1 1 (Isl1, UniProt P61372)	27-36	QYILRVSPDL	21	2487
	65-74	TYCKRDYIRL	22	10474
	286-294	SYQPPWKVL	30	172
Tumor antigen P198 (Tum)	*148-156*	*KYQAVTTTL*	*30*	*6*
Insulin (Ins1/Ins2)	*B15-23*	*LYLVCGERG*	*17*	*9145*
Islet-specific glucose-6-phosphatase catalytic subunit-related protein (Igrp)	*206–214*	*VYLKTNVFL*	*24*	*195*

Table S3. Islet peptides tested for recognition by islet-infiltrating CD8$^+$ T cells of NOD mice. The 9-mer and 10-mer peptides from the 4 islet antigens Ucn3, Pcsk2, Scg5 and Isl1 were selected according to their *in silico* predicted H2-K^d binding affinity (top hits using the SYFPEITHI and NetMHC 4.0 algorithms; www.syfpeithi.de and www.cbs.dtu.dk/services/NetMHC) and, for secreted proteins (Ucn3, Pcsk2 and Scg5), according to a peptide position close to cleavage sites involved in the intermediate processing of their precursors. Peptides yielding positive IFN-γ responses (in terms of both percent IFN-γ$^+$ cells and percent positive mice) are shown in bold. Control peptides are shown in italics.

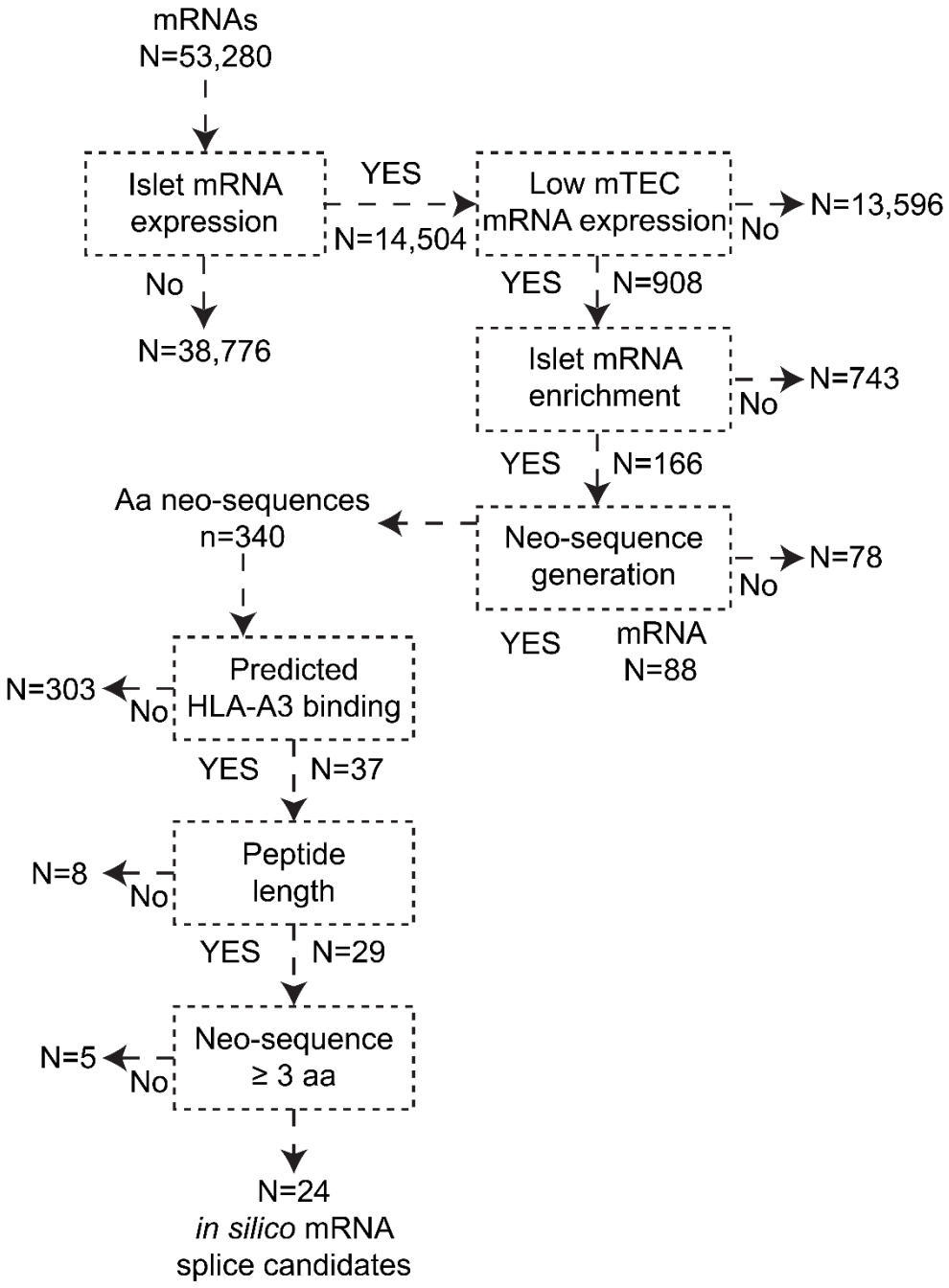

Figure S1. The bioinformatics pipeline used to identify candidate HLA-A3-restricted epitopes by *in silico* mining of mRNA splice variants. This analysis strategy was previously described for the identification of HLA-A2-restricted islet epitopes (see *Research Design and Methods*). Briefly, 53,280 mRNAs were filtered based on islet mRNA expression (n=14,504), low mTEC mRNA expression (n=908), islet enrichment compared with control tissues (n=166), and generation of a predicted aa neo-sequence (n=88). The 340 predicted aa neo-sequences obtained from these 88 mRNA variants were scanned for the presence of predicted HLA-A3 binders (n=37) with a 9- to 10-aa length (n=29) and a neo-sequence of ≥3 aa (n=24).

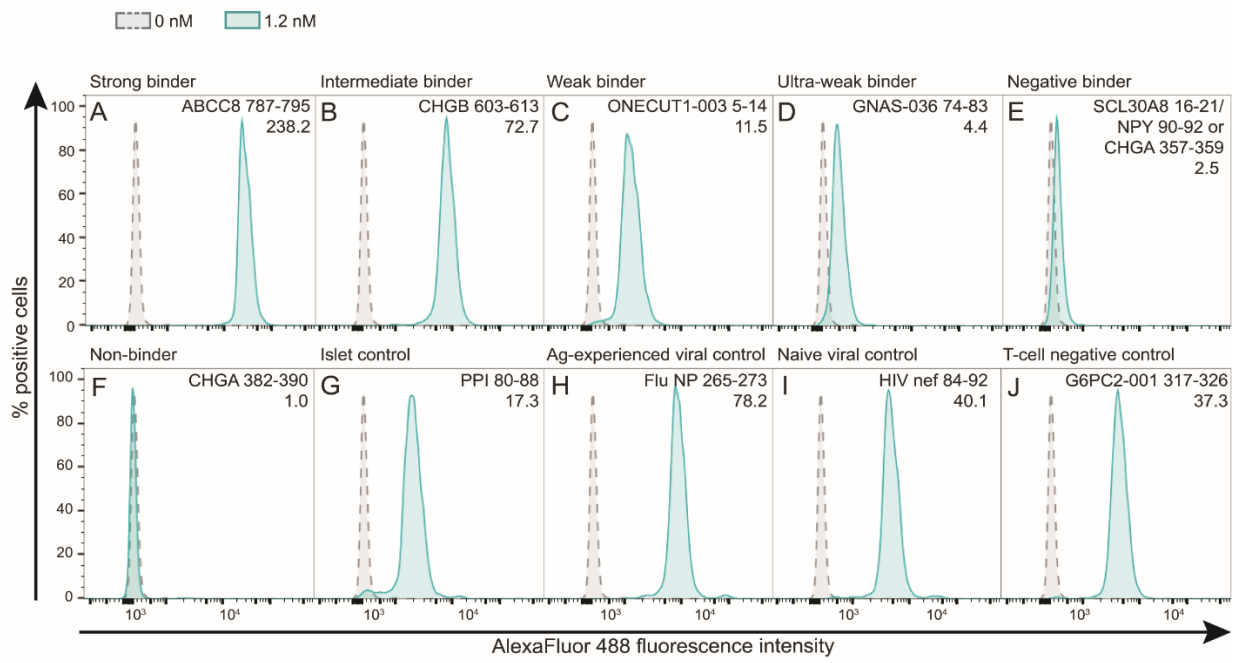

Figure S2. Representative staining from HLA-A3 binding assays. Bionylated recombinant monomeric HLA-A3 molecules were folded with the indicated peptides and β2-microglobulin and captured on streptavidin-coated beads, followed by an anti-β2-microglobulin antibody revealed with an AF488-labeled secondary antibody. A bead-associated fluorescence is only detected if the test peptide supports the folding of the HLA-A3 complex. Peptides were tested at 1.2 nM and exemples are shown for strong binders (ABCC8$_{787-795}$; **A**), intermediate binders (ChgB$_{603-613}$; **B**), weak binders (ONECUT1-003$_{5-14}$; **C**), an ultra-weak binder (GNAS-036$_{74-83}$; **D**) and a non-binder (SCL30A8$_{16-21}$/NPY$_{90-92}$; **E**), along with the non-binding (ChgA$_{382-390}$; **F**) and binding control peptides: PPI$_{80-88}$ for islet antigens, **G**; Flu NP$_{265-273}$ for a recall viral antigen, **H**; HIV nef$_{84-92}$ for a naïve viral antigen, **I**; and G6PC2-001$_{317-326}$ for the T-cell negative control (i.e. binding peptide but no T-cell recognition, **J**). The empty HLA-A3 complex control (no peptide added) is shown as a dashed grey profile in each histogram. Each peptide was tested in two separate experiments. The indicated values are arbitrary units of median fluorescence intensity normalized to the ChgA$_{382-390}$ non-binding control, as those reported in Table 1.

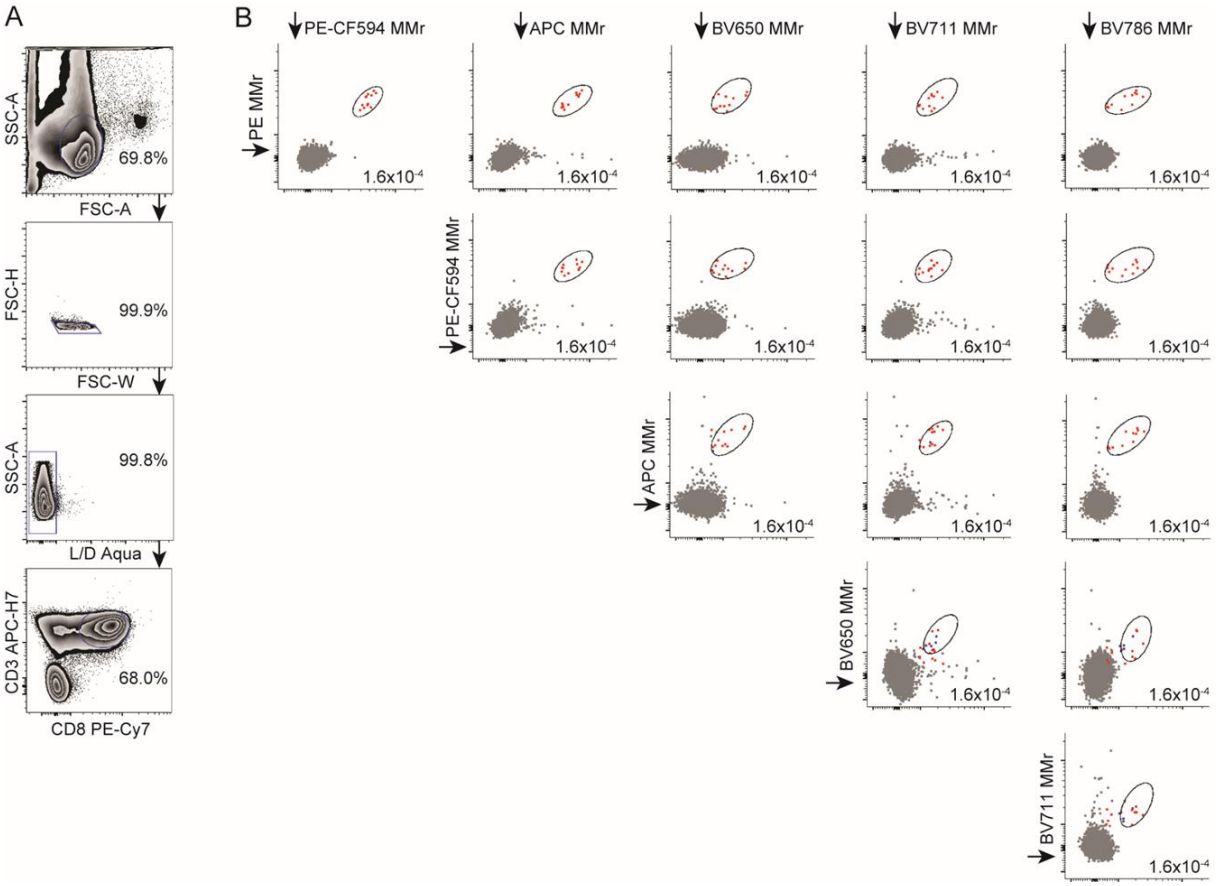

Figure S3. *Ex-vivo* detection of Flu $NP_{265-273}$-reactive $CD8^+$ T cells using a 6-color double-stained multimer (MMr) panel. PBMCs from HLA-$A3^+$ healthy donor H004N were stained with HLA-A3 MMrs loaded with the Flu $NP_{265-273}$ peptide and labeled with 6 fluorochromes (PE, PE-CF594, APC, BV650, BV711, BV786) organized in 15 unique pairs. **(A)** Gates were sequentially set on lymphocytes, single cells, viable cells (Live/Dead $Aqua^-$) and $CD3^+CD8^+$ cells. **(B)** No combinatorial MMr analysis was applied. Double-stained $PE^+PE\text{-}CF594^+$ T cells were gated (upper left dot plot) and overlaid (in red) on the other 14 double-stained populations. Numbers in each panel indicate the frequency of gated MMr^+CD8^+ T-cell out of total $CD8^+$ T cells. Although identical T-cell frequencies were counted for all fluorochrome pairs, the overlap with $PE^+PE\text{-}CF594^+$ events was incomplete for the dimmer BV650/BV711, BV650/BV786 and BV711/BV786 pairs (bottom left dot plots). Total $CD8^+$ events are overlaid in gray in each dot plot.