Prospective identification of cross-reactive human peptide-MHC ligands for T cell receptor based therapies

Ron S. Gejman[1,2,3], Martin G. Klatt[2], Aaron Chang[2,3], Heather F. Jones[2,3], Claire Y. Oh[2,3], Smita S. Chandran[4,5], Tatiana Korontsvit[3], Viktoriya Zakahleva[3], Tao Dao[3], Christopher A. Klebanoff[2,4,5], David A. Scheinberg[2,3,5]

1. Tri-Institutional MD-PhD Program (MSKCC, Rockefeller University, Weill Cornell Medical College), New York, NY 10065, USA.
2. Weill Cornell Medicine, New York, NY, USA.
3. Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA
4. Center for Cell Engineering and Department of Medicine, MSKCC, New York, NY, USA.
5. Parker Institute for Cancer Immunotherapy, MSKCC, New York, NY, USA.

T cell receptor (TCR)-based therapeutic cells and agents (e.g. adoptive T cell[1,2], TCR-engineered T cells[3], ImmTACs[4], TCR mimic antibodies[5], neoantigen vaccines[6,7]) have emerged as a new class of effective and selective cancer therapeutics against intracellular cancer-associated proteins. These agents rely on presentation of short peptides derived from cellular, viral or phagocytosed proteins on major histocompatibility complex (MHC). However, cross-reactivities of these agents to off-target cells and tissues are poorly understood, difficult to predict, and have resulted in serious, sometimes fatal, adverse events[8]. Here we have developed a mammalian, minigene-based method (termed "PresentER") that encodes MHC-I peptide ligands for functional immunological assays as well as for determining the reactivities and potential cross-reactivities of TCR-like therapeutic drugs against libraries of MHC-I ligands. This system is highly specific to, and entirely dependent on, the genetically encoded MHC peptide sequence, because it does not require proteasome cleavage, transporter associated with antigen processing (TAP) or processing, for immune presentation. For the expression of defined MHC-I ligands, this system is superior to expression of full-length cDNA. PresentER-expressing cells can be bound by TCR and TCR mimic (TCRm) antibodies, activate antigen-specific T cells, lead to antigen-specific cell death *in vitro* and tumor rejection in wild-type mice. Using PresentER in a pooled library screen, we find dozens of MHC-I ligands encoded in the human proteome that are cross-reactive with two TCR mimic antibodies and are not predictable by other methods. PresentER has broad and immediate utility for both basic and translational studies in immunology and oncology.

TCR based therapeutics are structurally similar to the TCR on CD8 T cells and thus share both their potential advantages and challenges. For instance, CD8 T cells can theoretically discern whether any 8-12 amino acid peptide is self, foreign or altered-self. Yet, the number of possible peptide sequences that can be encoded by the twenty proteinogenic amino acids is significantly larger than the number of circulating T cells in the human body. In order to account for this discrepancy, TCR can be highly cross-reactive: a single TCR may be capable of recognizing over 1 million distinct pMHC[9]. Thymic selection *in vivo* is critical to ensure that few circulating T cells are auto-reactive. However, TCR-based therapeutics that do not undergo negative selection for the human pMHC repertoire or are further engineered for high affinity binding can be auto-reactive. A prominent example is an affinity-enhanced engineered anti-MAGE-A3 T cell, which induced lethal cardiotoxicity in two patients during a phase I clinical trial. Extensive preclinical testing failed to uncover off-target reactivity of the anti-MAGE TCR, but afterwards it was discovered that an epitope derived from the Titin protein (a structural protein highly expressed by cardiomyocytes), was cross-reactive with the MAGE-A3 TCR[8]. Hence, a major challenge to the development of safe TCR based therapeutics is the prospective identification of off-tumor off-targets[10].

Identifying off-tumor off-targets is challenging for three reasons: (1) the scope and extent of the repertoire of MHC ligands in humans is unknown, despite multiple reports of isolation by mass spectrometry, (2) cross-reactive pMHC are not readily predictable from crystal structures or alanine scanning and (3) animal models of cross-reactivity are not possible due to the species-specific peptide processing[11] and structure of MHC. Methods to search the MHC-I ligandome for TCR targets have been developed with yeast[12,13], insect-baculovirus[14] and tetramer[15,16] technologies. These systems are

powerful and have been used to discover cross-reactive targets of some TCRs, in addition to elucidating fundamental biology of TCRs[17]. However, the cellular systems use a synthetic covalent linker to enforce peptide-MHC proximity, which allows presentation of peptides that would not ordinarily be presented and may distort the structure of the epitope. The tetramer-based screening systems rely on peptide synthesis, which is expensive and time consuming. Furthermore, tetramer, yeast and insect systems cannot be used for *in vitro* and *in vivo* immunology assays, such as T cell recognition and cytotoxicity. The PresentER system, by relying on the native MHC-I molecules of mammalian cells avoids these difficulties.

We designed a TAP independent antigen presentation minigene by encoding an endoplasmic reticulum (ER) signal sequence upstream of an MHC-I ligand, building on the methods of other groups[18]. We modified a Mouse mammary tumor virus ER signal sequence to include SfiI (type IIS) restriction sites flanking a removable cassette for efficient, directional cloning of DNA sequences **(Figure 1a** and **Sup. Fig. 1)**. PresentER was designed for pooled screening applications; therefore, the amino acid sequence corresponding to a peptide is encoded only once per minigene.

We have previously isolated and characterized two anti-cancer TCR mimic (TCRm) antibodies (ESK1[5] and Pr20[19]), which bind to the HLA-A*02:01 ligands RMFPNAPYL (WT1:126-134), and ALYVDSLFFL (PRAME:300-309), respectively. These antibodies selectively bound T2 cells expressing their respective PresentER epitopes but not irrelevant epitopes **(Fig. 1b-c and Suppl. Fig. 2a-b)**.

While ESK1 and Pr20 have nanomolar affinity to pMHC, TCR typically have micromolar affinity to their targets, making functional binding assays difficult to perform. T2 cells expressing PresentER-NLVPMVATV (Cytomegalovirus pp65:495-503; HLA-A*02:01 ligand) were specifically bound by a TCR multimer directed to this epitope (**Fig 1d**). In addition, an A6 TCR[20] tetramer showed specific binding to T2 cells expressing its target MHC-I ligand PresentER LLFGYPVYV (HTLV-1 Tax:11-19; HLA-A*02:01 ligand) (Sup. **Fig. 2c**).

PresentER also was useful in presenting antigens corresponding to different MHC-I alleles. For example, an antibody against mouse MHC-I H2-Kb/SIINFEKL bound the correct epitope in the TAP2 deficient mouse RMA-S cell line expressing PresentER SIINFEKL (Chicken Ovalbumin 257-264) or MSIIFFLPL (PEDF:271–279). (**Fig. 1d)**.

PresentER can be used in functional immunology assays. Genetically engineered T cells expressing the F5 TCR, specific to an HLA-A*02:01 peptide from MART-1 (27-35:AAGIGILTV)[3,21] released IFN-γ when exposed to peptide pulsed T2 cells or T2 cells expressing the PresentER-MART-1 minigene (**Fig. 1e**). F5 T cells specifically killed PresentER MART-1 expressing T2 cells in an *in vitro* co-culture assay (**Fig. 1f**). Finally, we challenged wild type C56B6/N mice with RMA/S cells expressing either foreign or wild type H2-Kb ligands to test whether PresentER minigenes could mediate antigen-specific tumor rejection *in vivo*. We found that tumors expressing foreign MHC-I ligands were rejected, in contrast to tumors expressing non-immunogenic self-peptide (**Fig. 1g-h**).

We hypothesized that the PresentER system would proteasomal cleavage and peptide transport in the generation of peptide-MHC. To test whether any peptide processing was occurring, we immunoprecipitated peptide-MHC complexes from T2 cells expressing PresentER-RMF or PresentER-ALY and identified bound peptides by mass spectrometry. RMFPNAPYL and ALYVDSLFFL were identified only in cells encoding those PresentER constructs. Truncated versions of the RMF and ALY peptides were not found, nor were peptides derived from the ER signal sequence identified (**Sup. Fig. 3a**). We scrambled the ER signal sequence to test whether peptides were associating with MHC via a signal-sequence independent mechanism and found no binding to cells expressing these constructs **(Sup. Fig. 3b)**. Importantly, by performing low MOI transductions, we show that the PresentER system is capable of driving sufficient antigen presentation from a single copy of the retroviral minigene, thereby enabling its use in a pooled screen **(Sup. Fig. 4)**.

Pr20 and ESK1 have known cross-reactivities to off-target MHC-I ligands. Based on alanine/residue scanning and structural[22] data, we determined that ESK1 binding to RMFPNAPYL depended primarily on the R1 and P4 residues and for Pr20, the C-terminus of the peptide was important[19]. Therefore we constructed a biased library of possible ESK1 cross-reactive targets by searching the human proteome *in silico* for 9-mer peptide sequences containing arginine in position 1 and proline in position 4. We performed a similar search for Pr20 cross-reactive targets using a 10-mer peptide motif based on prior biochemical data **(Fig. 2a)**. We located 1,157 and 24,539 potential cross-reactive peptides of ESK1 and Pr20, respectively, with NetMHCPan[23] predicted HLA-A*02:01 affinity of less than 500nM. We synthesized a pool of 12,472 oligonucleotides that together encoded all of the ESK1 cross reactive peptides and half of the Pr20 cross-reactive targets plus the one amino acid mutants of RMF and ALY (termed "CR-ESK1" and "CR-Pr20", respectively), as well as positive/negative controls **(Fig. 2b)**. The oligonucleotides were cloned into the PresentER vector and Illumina sequencing was performed to demonstrate that library representation was maintained during cloning. The flow-based screen assaying for cross-reactivity was performed as follows: T2 cells were transduced at low multiplicity of infection (MOI) and transductants selected with puromycin. Library representation was kept at >1,000x at all stages before the final sort. Transduced T2 cells were stained with ESK1 or Pr20, sorted and sequenced (**Fig. 2c**).

The abundance of each minigene in the unsorted samples was well correlated (r = 0.93), indicating that sample handling and the days between sorts did not affect the library representation (**Sup. Fig. 5a)**. Minigenes encoding previously known ESK1 ligands had higher enrichment scores than those encoding non-ligands (p=0.032), suggesting that the flow-based screen was able to separate ESK1 binders from non-binders (**Sup. Fig. 6a)**. Surprisingly, several of the most enriched peptides that emerged in the ESK1 screen were CR-Pr20 peptides (**Fig. 2d)**. Although these peptides are 10-mers, some bear sequence similarity to the target ligand of ESK1.

To validate screen hits, we synthesized a subset of the enriched peptides and assayed them for binding to ESK1 by pulsing T2 cells. Of the 27 peptides tested, 22 (81%) showed increased binding to ESK1, including several which had originally been selected for Pr20 cross-reactivity and did not contain a proline in position 4 (**Fig. 2e)**. These unusual targets could not have been predicted from either the crystal structure of ESK1 or the alanine scanning data.

Recently, large databases of HLA-A*02:01 peptide ligands isolated from tumors and normal tissue have become available[24-27]. Within these databases (including personal correspondence with Department of Immunology members at Tübingen) we found 48 nine-mer peptides with R1 and P4. Forty-five had been included in the library of CR-ESK1 peptides identified *in silico* from the human proteome. We synthesized 27 of these and found that 17 (63%) bound to ESK1 when pulsed on T2 cells. The median ESK1 enrichment in the flow-based screen of these 17 peptides was 1.8 whereas the median enrichment of the non-binders was 0.86 **(Sup. Fig. 6b**), indicating that these ESK1 ligands were enriched in the screen, even if they were not among the top hits.

Two WT1-negative[28] cell lines contained known MHC-I ligands corresponding to the ESK1 off-targets discovered in the PresentER screen (TPC-1: RLPPPFPGL, RVMPSSFFL, RLGPVPPGL, JY: KLYNPENIYL, RLVPFLVEL). RMFPNAPYL was not found among the MHC-I ligands immunoprecipitated from these lines. We tested ESK1 binding in each of these lines and found that JY cells bound ESK1 at high levels while TPC-1 was marginally positive for ESK1 binding (**Fig. 2f-g**). Thus, PresentER may be used to identify both theoretical and, in some cases, actually presented TCR mimic off-targets.

A screen of Pr20 cross-reactive ligands was performed in the same manner as described above. The abundance of each minigene in the unsorted samples was highly correlated (r=0.90-0.94) (**Sup. Fig. 5b)**. Positive control Pr20 binders were not enriched relative to the negative controls (p=0.71) (**Sup. Fig. 7a**). Twenty peptides more than 5-fold enriched for Pr20 binding with predicted ic50s of less than 100nM (**Sup. Fig. 7b**). An additional 47 peptides that matched the Pr20 ligand consensus motif were

located in the HLA-A*02:01 ligand databases and 13 had been included in the CR-Pr20 library. These peptides were not overall enriched for Pr20 binding (0.73-1.65 enrichment scores).

45 peptides were synthesized, including those enriched in the Pr20 off-target screen and several HLA-A*02:01 ligands identified by mass spectrometry that matched the Pr20 ligand consensus sequence. Of these, 28 (62%) were found to be Pr20 ligands when the peptides were pulsed onto T2 cells, including 3 known HLA-A*02:01 ligands. The 28 validated peptides were more enriched for Pr20 binding in the screen than the 17 non-validated peptides (**Sup. Fig. 7c**).

Examining only the CR-ESK1 subset of peptides, we noticed that the peptides most enriched for ESK1 binding were also predicted to have the highest affinity for HLA-A*02:01 (**Fig. 3a)**. The peptides that are ≥5-fold enriched for ESK1 binding have a median affinity to HLA-A*02:01 of 31nM, compared to 95nM for the library as a whole and 102nM for ≥5-fold depleted peptides (**Fig. 3b**). We found the same result in the Pr20 screen: the most enriched Pr20 ligands also had the highest affinity to MHC-I. (**Fig. 3c**). We cloned minigenes for four of the most enriched CR-ESK1 (RLFPLAWTV 31.8x; KLMGAISFFI 41.9x) and CR-Pr20 (WLLGDSSFFL 6.5x; LLIQEGPFFV 6.6x) peptides and tested them for binding to ESK1 and Pr20. Compared to RMFPNAPYL and ALYVDSLFFL, cells expressing these peptides bound ESK1 and Pr20 at much higher levels (**Fig. 3d**).

Previously known ligands of ESK1 and Pr20 were not all identified by flow-based screening of PresentER minigenes. These ligands were originally identified by pulsing T2 cells with saturating quantities of each peptide. We speculate that some of these peptides may not be well presented when expressed genetically, either because of inefficient loading onto MHC-I or negative selection during peptide editing, e.g. by TAPBPR[29]. The skew we observed in both ESK1 and Pr20-enriched minigenes towards high-affinity HLA-A*02:01 ligands suggests that genetic expression of peptides selects for presentation of ligands with the highest affinities for HLA-A*02:01. This is be an unexpected feature of PresentER, as affinity to MHC-I is the most important factor in determining if a peptide is presented on MHC-I (although high expression levels may overcome low affinity)[30]. This suggests that methods that incorporate a flexible linker to covalently retain a peptide in the vicinity of MHC may lead to artificial presentation of peptides that would never be presented endogenously. Additional study of the difference between genetic expression of MHC-I ligands and peptide pulsing should be pursued to help investigators decide which cross-reactive peptides are likely to be endogenous MHC-I ligands and thus pose a risk to patients in a clinical setting.

Preclinical evaluation methods for novel therapeutic agents directed towards peptide-MHC have been insufficient to prevent harmful off-tumor off-target toxicity. In PresentER, we have developed a mammalian screening approach to prospectively identify cross-reactive MHC-I ligands. While PresentER may not detect all naturally expressed cross-reactive epitopes and may detect epitopes that are never presented in an endogenous setting, the system can help to identify potential cells and tissues at risk for closer clinical surveillance.

In this report we have shown that PresentER can be used for biochemical evaluation of potentially therapeutic TCR based agents. PresentER can also be used as an immune presentation platform *in vitro* and *in vivo*; thus, this work can be expanded to recapitulate the MHC restricted antigenic diversity of human cancer. Libraries of MHC-I ligands could be used to ask how tumor neoantigen heterogeneity affects progression and treatment of tumors in immunocompetent animals, and address areas such as neoantigen immunogenicity and clonality, cancer vaccination and immunoediting.

Acknowledgements
We thank Eureka Therapeutics for their generous gift of the ESK1 and Pr20 antibodies; Brian Baker and Lance Hellman for the gift of A6 TCR plasmids and Yael David for assistance in expressing and refolding; Scott Lowe and Eusebio Manchado Robles for the gift of the MSCV backbone and assistance with pooled library cloning; Steven A. Rosenberg for the F5 TCR. We thank the Integrated Genomics

Operation at Memorial Sloan Kettering Cancer Center and the Flow Cytometry core facility for their assistance.

Funding

The project was supported by the Parker Institute for Cancer Immunotherapy and the Functional Genomics Initiative at Memorial Sloan Kettering Cancer Center. RSG is supported by NCI F30 CA200327 and NIGMS T32GM07739. DAS is supported by NCI PO1 CA 55349, CA23766. MM is supported by Deutsche Forschungsgemeinschaft Grant no. KL 3118/1-1.

References

1. O'Reilly, R. J., Dao, T., Koehne, G., Scheinberg, D. & Doubrovina, E. Adoptive transfer of unselected or leukemia-reactive T-cells in the treatment of relapse following allogeneic hematopoietic cell transplantation. *Semin. Immunol.* **22,** 162–172 (2010).
2. Rosenberg, S. A. & Restifo, N. P. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* **348,** 62–68 (2015).
3. Johnson, L. A., Heemskerk, B., of, D. P. T. J.2006. Gene Transfer of Tumor-Reactive TCR Confers Both High Avidity and Tumor Reactivity to Nonreactive Peripheral Blood Mononuclear Cells and Tumor-Infiltrating Lymphocytes. *Am Assoc Immnol* **177,** 6548–6559 (2006).
4. Oates, J., Hassan, N. J. & Jakobsen, B. K. ImmTACs for targeted cancer therapy: Why, what, how, and which. *Mol. Immunol.* **67,** 67–74 (2015).
5. Dao, T. *et al.* Targeting the intracellular WT1 oncogene product with a therapeutic human antibody. *Sci Transl Med* **5,** 176ra33 (2013).
6. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547,** 222–226 (2017).
7. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547,** 217–221 (2017).
8. Cameron, B. J. *et al.* Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells. *Sci Transl Med* **5,** 197ra103–197ra103 (2013).
9. Singh, N. K. *et al.* Emerging Concepts in TCR Specificity: Rationalizing and (Maybe) Predicting Outcomes. *J Immunol* **199,** 2203–2213 (2017).
10. Chang, A. Y. *et al.* Opportunities and challenges for TCR mimic antibodies in cancer therapy. *Expert Opin Biol Ther* **16,** 979–987 (2016).
11. Kotturi, M. F. *et al.* Of mice and humans: how good are HLA transgenic mice as a model of human immune responses? *Immunome Res* **5,** 3 (2009).
12. Birnbaum, M. E. *et al.* Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157,** 1073–1087 (2014).
13. Gee, M. H. *et al.* Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. *Cell* (2017). doi:10.1016/j.cell.2017.11.043
14. Crawford, F. *et al.* Use of baculovirus MHC/peptide display libraries to characterize T-cell receptor ligands. *Immunol. Rev.* **210,** 156–170 (2006).
15. Hadrup, S. R. *et al.* Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat Meth* **6,** 520–526 (2009).
16. Bentzen, A. K. *et al.* Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* **34,** 1037–1045 (2016).
17. Adams, J. J. *et al.* Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat Immunol* **17,** 87–94 (2016).
18. Bacik, I., Cox, J. H., Anderson, R., Yewdell, J. W. & Bennink, J. R. TAP (transporter associated with antigen processing)-independent presentation of endogenously synthesized peptides is enhanced by endoplasmic reticulum insertion sequences located at the amino- but not carboxyl-terminus of the peptide. *The Journal of Immunology* **152,** 381–387 (1994).
19. Chang, A. Y. *et al.* A therapeutic T cell receptor mimic antibody targets tumor-associated PRAME peptide/HLA-I antigens. *J Clin Invest* **127,** 2705–2718 (2017).
20. Garboczi, D. N. *et al.* Assembly, specific binding, and crystallization of a human TCR-alphabeta with an antigenic Tax peptide from human T lymphotropic virus type 1 and the class I MHC molecule HLA-A2. *The Journal of Immunology* **157,** 5403–5410 (1996).
21. Johnson, L. A. *et al.* Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. *Blood* **114,** 535–546 (2009).

282 22. Ataie, N. *et al.* Structure of a TCR-Mimic Antibody with Target Predicts Pharmacogenetics.
283 *J. Mol. Biol.* **428,** 194–205 (2016).
284 23. Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans.
285 *Immunogenetics* **61,** 1–13 (2009).
286 24. Schuster, H. *et al.* The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl*
287 *Acad Sci USA* **114,** E9942–E9951 (2017).
288 25. Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes
289 presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7,**
290 13404 (2016).
291 26. Bassani-Sternberg, M. *et al.* Deciphering HLA-I motifs across HLA peptidomes improves
292 neo-antigen predictions and identifies allostery regulating HLA specificity. *PLoS Comput*
293 *Biol* **13,** e1005725 (2017).
294 27. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass
295 spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of
296 protein abundance and turnover on antigen presentation. *Molecular & Cellular Proteomics*
297 **14,** 658–673 (2015).
298 28. Li, Z. *et al.* Identification of a WT1 protein-derived peptide, WT1, as a HLA-A 0206-
299 restricted, WT1-specific CTL epitope. *Microbiol. Immunol.* **52,** 551–558 (2008).
300 29. Hermann, C. *et al.* TAPBPR alters MHC class I peptide presentation by functioning as a
301 peptide exchange catalyst. *Elife* **4,** 26 (2015).
302 30. Abelin, J. G. *et al.* Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-
303 allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **46,** 315–326 (2017).
304

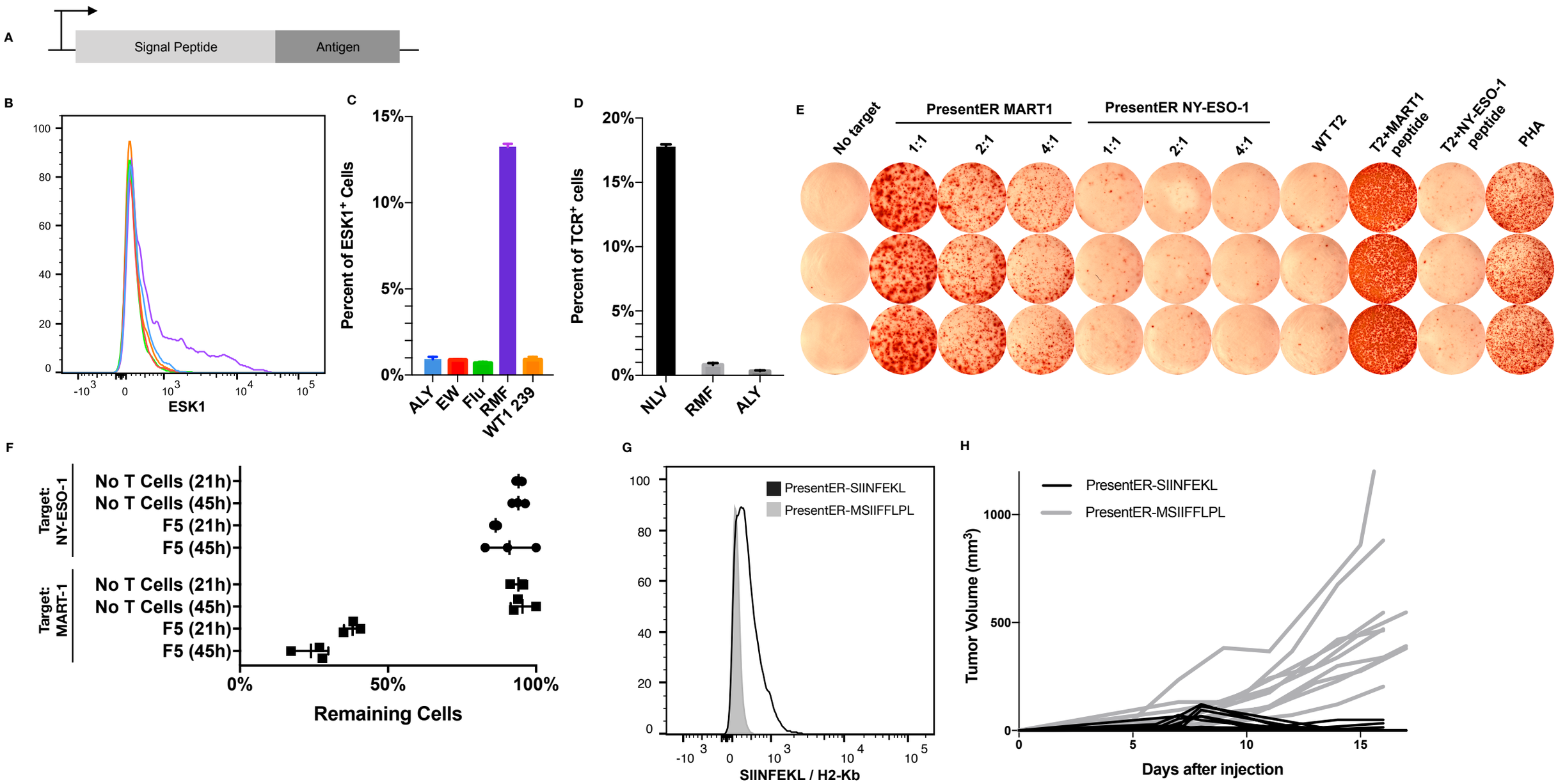


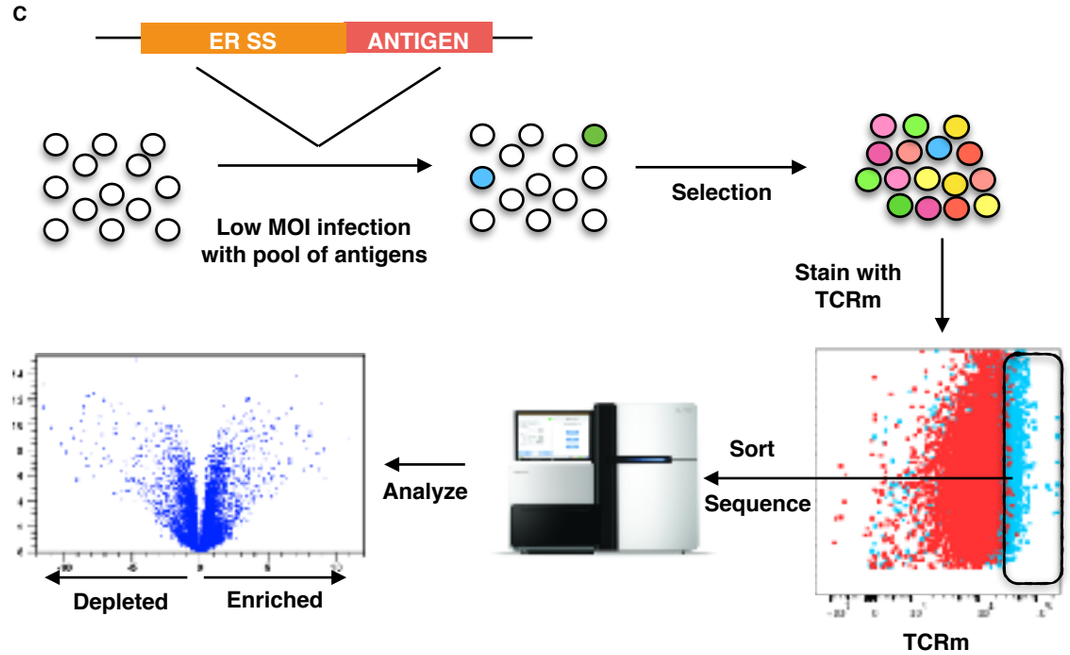
Figure 1: Design and characterization of PresentER. **(A)** The PresentER minigene encodes a signal sequence upstream of a peptide antigen and followed by a stop signal. **(B)** T2 cells expressing a PresentER minigene encoding the peptide RMFPNAPYL are bound by fluorescently labeled ESK1, a TCR mimic antibody to the complex of RMFPNAPYL/HLA-A*02:01. The color of each sample in the histogram correspond to the color of each sample in the barplot **(C)** of the same data. Antibody binding is quantified by gating all events with fluorescence intensity higher than the irrelevant-antigen controls. **(D)** A multimerized and fluorescently labeled TCR binds to T2 cells expressing its target peptide CMV pp65/NLVPMVATV in complex with HLA. **(E)** ELISPOT of genetically engineered T cells expressing the F5 TCR directed against the HLA-A*02:01 ligand MART-1 27-35 (AAGIGILTV) challenged with T2 cells pulsed with peptide, expressing the PresentER MART-1 minigene or irrelevant NY-ESO-1 157-165 (SLLMWITQC) peptide at Effector:Target ratios from 1:1 to 4:1. **(F)** The results of an *in vitro* co-culture killing assay where F5 expressing T cells were incubated with PresentER MART-1 or PresentER NY-ESO-1 minigene (GFP) expressing T2s mixed with PresentER ALYVDSLFFL (mCherry) expressing T2s. The fraction of GFP positive cells is reported at 21h and 45h, normalized to the percentage in the wells without T cells. **(G)** A fluorescently labeled antibody to SIINFEKL/H-2Kb (clone 25-D1.16) binds to RMA-S cells expressing PresentER-SIINFEKL, but not to PresentER-MSIIFFLPL. **(H)** C57BL/6N mice were injected subcutaneously with $5x10^6$ RMA-S cells expressing PresentER-SIINFEKL (immunogenic; black; n=14) or PresentER-MSIIFFLPL (not immunogenic; gray; n=10). Spider plots of tumor growth across several independent experiments are plotted.

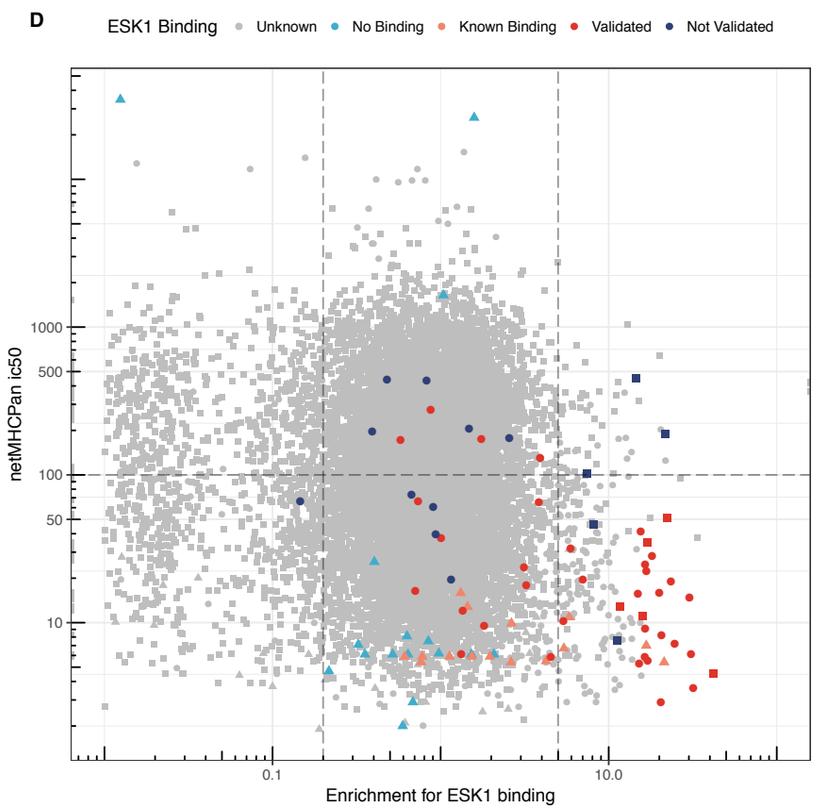
A

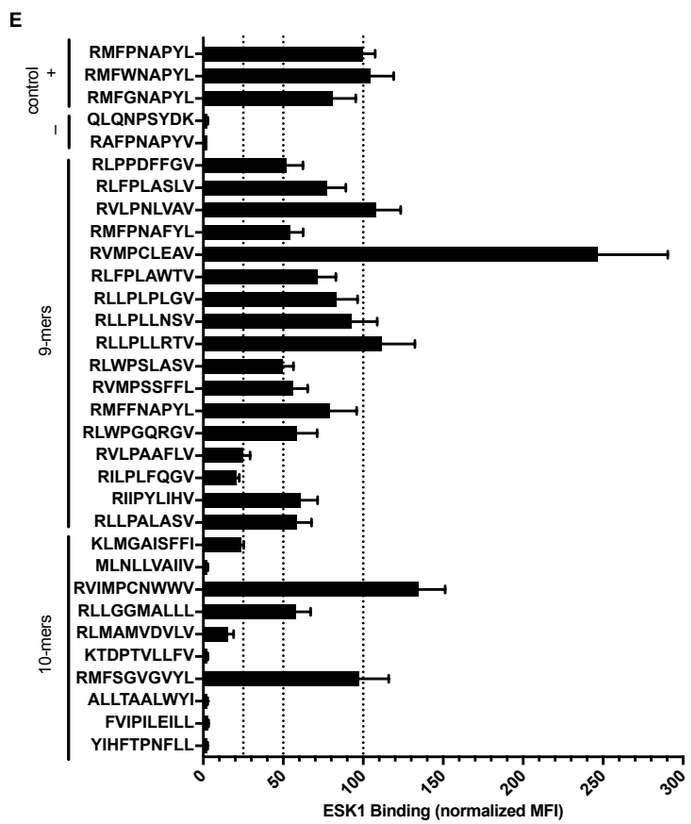
	1	2	3	4	5	6	7	8	9	10
ESK1 Target	R	M	F	P	N	A	P	Y	L	–
ESK1 Library	R	*	*	P	*	*	*	*	*	–
Pr20 Target	A	L	Y	V	D	S	L	F	F	L
Pr20 Library	*	*	*	DE	KH	K	K	FW	FW	L
					R	H	H	YV	YV	V
						R	R	LI	LI	I

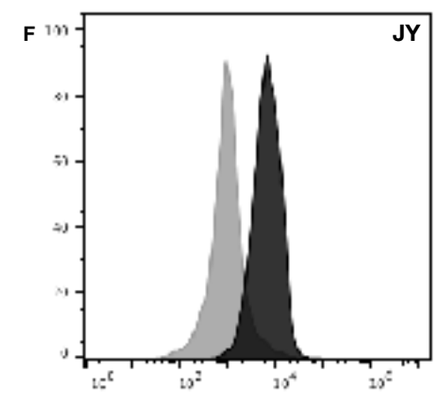
B

		# Constructs (%)
	Controls	13 (0.1%)
CR-ESK1	RMFPNAPYL: all 1 AA mismatch	180 (1.45%)
	ESK1 genomic off-targets	1,157 (9.3%)
CR-Pr20	ALYVDSLFFL: all 1 AA mismatch	190 (1.5%)
	Pr20 genomic off-targets	10,893 of 24,539 (87.6%)
	Total	**12,433**









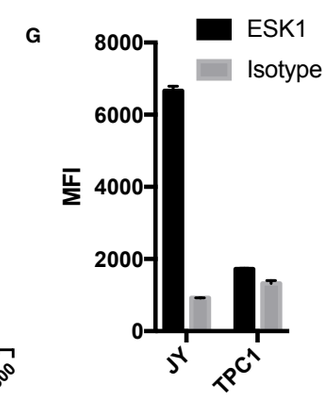


Figure 2: Design of PresentER library and ESK1 screen. **(A)** Design of sequence constraints on the target peptide library based on prior biochemical data. The human exome was mined for peptides matching the specified ESK1 and Pr20 consensus motifs. Asterisks indicate any amino acid is allowed. Red characters indicate prohibited amino acids at that position and black characters indicate allowed amino acids at that position. **(B)** The library that was constructed included 13 control peptides, 1,337 CR-ESK1 peptides and 11,083 CR-Pr20 peptides. **(C)** Schematic of the flow-based screen. T2s are transduced at low MOI (<0.3) with retrovirus encoding a pool of PresentER minigenes. Transduced cells are selected by puromycin and then cultured until sufficient cells are obtained. Cells are stained with the TCR mimic antibody and fluorescent activated cell sorting is used to sort binding and non-binding populations of cells. Genomic DNA is extracted from sorted cells and sequenced with Illumina sequencing. **(D)** Scatterplot of the ESK1 library screen. Each point is a unique peptide minigene with the x-axis indicating minigene enrichment for ESK1 binding (with 1 set as no enrichment) and y-axis indicating the peptide's predicted ic50 (in nM) to HLA-A*02:01. Lower ic50 indicates higher affinity. Marked control peptides and known ESK1 targets are plotted as triangles; CR-ESK1 as circles and CR-Pr20 as squares. **(E)** 27 peptides that were highly enriched for ESK1 binding and had high predicted affinity to HLA-A*02:01 (from Figure 2C) were synthesized at microgram scale, pulsed onto T2 cells and stained with a fluorescently labeled ESK1. Previously identified cross-reactive targets were included as positive controls. The median fluorescence intensity (MFI) of ESK1 binding is plotted, normalized to RMFPNAPYL, set at 100 units. **(F)** Representative ESK1 and isotype staining of the JY cell line. **(G)** Quantification of ESK1 and isotype staining of the JY and TPC1 cell lines.

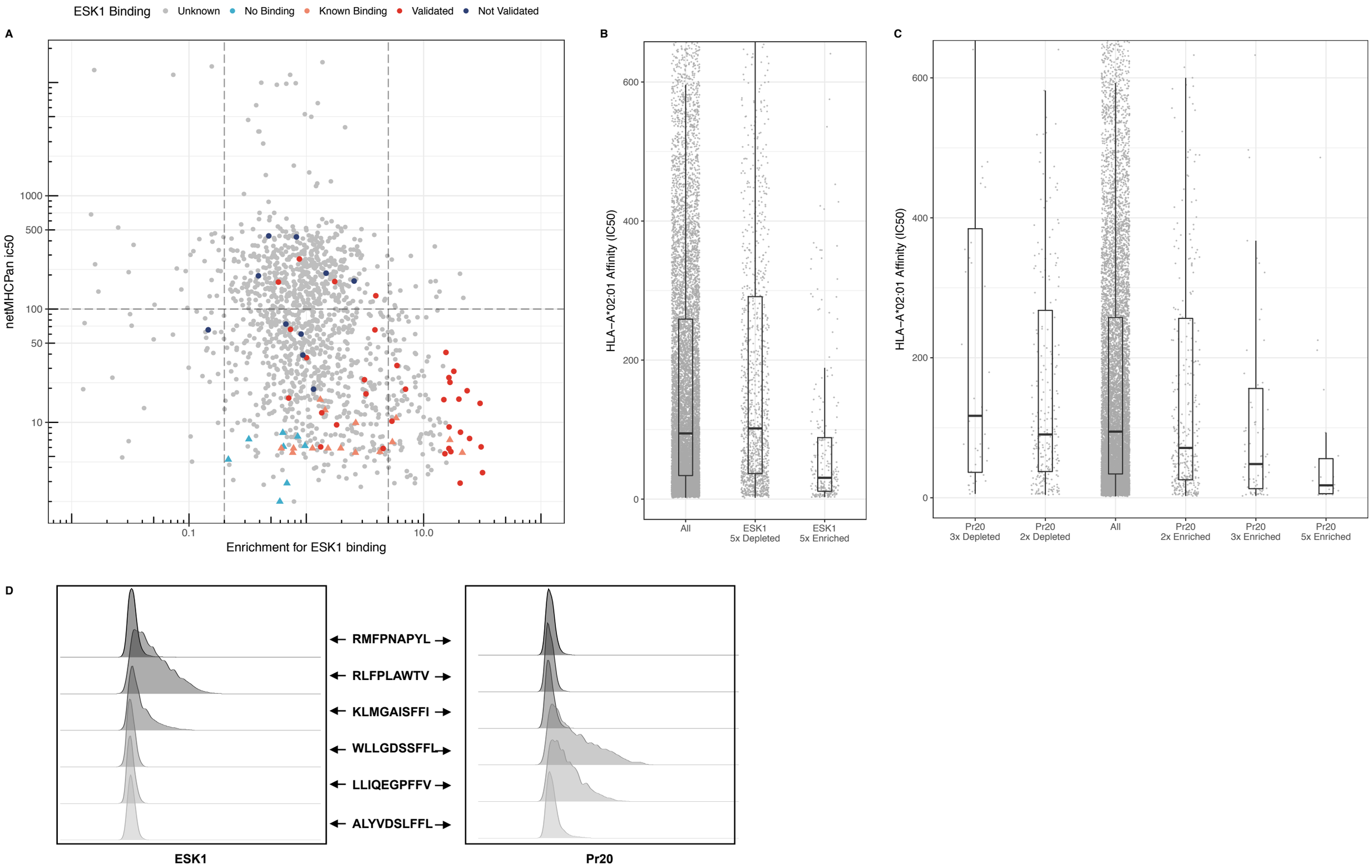


Figure 3: Peptides enriched in TCRm screening are high affinity MHC-I ligands **(A)** Scatterplot of the ESK1 library screen with only CR-ESK1 peptides (and controls) plotted. Each point is a unique peptide minigene with the x-axis indicating minigene enrichment for ESK1 binding (with 1 set as no enrichment) and y-axis indicating the peptide's predicted ic50 (in nM) to HLA-A*02:01. Lower ic50 indicates higher affinity. Marked control peptides and known ESK1 target peptides are plotted as triangles and CR-ESK1 peptides as circles. **(B)** The netMHCPan predicted HLA-A*02:01 affinity in ic50 (nM) of all screened peptides compared to peptides which were ≥5-fold depleted in the ESK screen and peptides that were ≥5-fold enriched for ESK1 binding. **(C)** The netMHCPan predicted HLA-A*02:01 affinity in ic50 (nM) of all screened peptides compared to peptides that were ≥5-fold, ≥3-fold or ≥2-fold enriched in the Pr20 screen and peptides that were ≤3 or ≤2-fold depleted for Pr20 binding. **(D)** ESK1 and Pr20 staining of 6 PresentER minigenes.