HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2020 January 02.

Published in final edited form as:

Cell. 2019 August 08; 178(4): 1016–1028.e13. doi:10.1016/j.cell.2019.07.009.

T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes

Tomasz Kula[1,2], **Mohammad H. Dezfulian**[1,2], **Charlotte I. Wang**[1,2,3], **Nouran S. Abdelfattah**[1,2], **Zachary C. Hartman**[4], **Kai W. Wucherpfennig**[5], **Herbert Kim Lyerly**[6], **Stephen J. Elledge**[1,2,7,*]

[1]Division of Genetics, Department of Medicine, Howard Hughes Medical Institute, Brigham and Women's Hospital, Boston, MA 02115, USA

[2]Department of Genetics, Harvard University Medical School, Boston, MA, USA

[3]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA

[4]Departments of Surgery and Pathology, Duke University Medical Center, 571 Research Drive, Suite 433, Box 2606, Durham, NC 27710, USA

[5]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Department of Immunobiology, Harvard Medical School, Boston, MA 02115, USA

[6]Departments of Surgery, Immunology, and Pathology, Duke University Medical Center, 571 Research Drive, Suite 433, Box 2606, Durham, NC 27710, USA

[7]Lead Contact

SUMMARY

T cell recognition of specific antigens mediates protection from pathogens and controls neoplasias, but can also cause autoimmunity. Our knowledge of T cell antigens and their implications for human health is limited by the technical limitations of T cell profiling technologies. Here, we present T-Scan, a high-throughput platform for identification of antigens productively recognized by T cells. T-Scan uses lentiviral delivery of antigen libraries into cells for endogenous processing and presentation on major histocompatibility complex (MHC) molecules. Target cells functionally recognized by T cells are isolated using a reporter for granzyme B activity, and the antigens mediating recognition are identified by next-generation sequencing. We show T-Scan correctly

*Correspondence: selledge@genetics.med.harvard.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, T.K., M.H.D., and S.J.E.; Methodology, T.K., M.H.D., C.I.W., N.S.A., Z.C.H., H.K.L., and S.J.E.; Experimental Investigation, T.K., M.H.D., C.I.W., and N.S.A.; Resources, Z.C.H., H.K.L., and K.W.W.; Writing – Original Draft, T.K. and S.J.E.; Writing – Review & Editing, M.H.D., C.I.W., N.S.A., H.K.L., Z.C.H., and K.W.W.

DATA AND CODE AVAILABILITY

The published article includes all datasets generated in this study.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.07.009.

DECLARATION OF INTERESTS

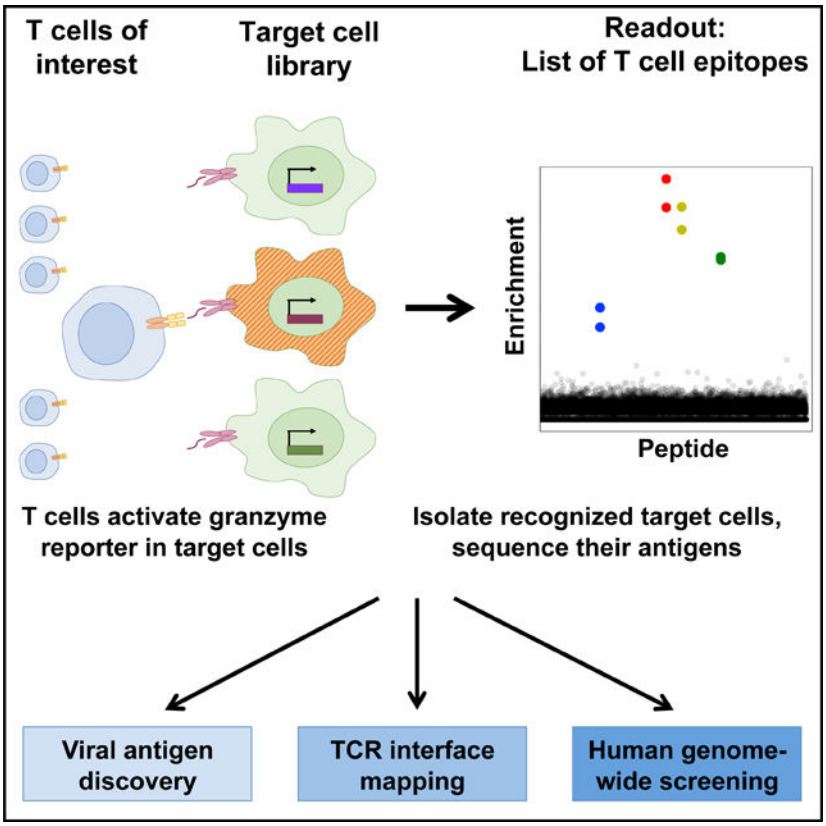
T.K., M.H.D. and S.J.E. are inventors of and have submitted a patent on the T-Scan technology. T.K. and S.J.E. are also founders of, and hold equity in, TScan Therapeutics, Inc., a company that has licensed this technology. The interests of S.J.E. were reviewed and are managed by BWH and Partners HealthCare in accordance with their conflict of interest policies.

identifies cognate antigens of T cell receptors (TCRs) from viral and human genome-wide libraries. We apply T-Scan to discover new viral antigens, perform high-resolution mapping of TCR specificity, and characterize the reactivity of a tumor-derived TCR. T-Scan is a powerful approach for studying T cell responses.

In Brief

T-Scan is a cell-based, pooled screening approach for high-throughput identification of antigens productively recognized by T cells.

Graphical Abstract



INTRODUCTION

The immune system is divided into innate and adaptive subsystems that together work to eliminate or inactivate pathogens and eliminate neoplasias. The adaptive immune system generates an immunological memory through memory B and T lymphocytes, potent effectors of this system. Understanding the specificity of this memory is central to understanding the ways in which pathogens are identified and eliminated, tumors are rejected, and pathogenic autoimmunity emerges.

B and T cells evolve antigen specificity through the generation of somatically rearranged B cell receptors (BCRs) and T cell receptors (TCRs). T lymphocytes fall broadly into two

categories, $CD4^+$ helper and $CD8^+$ cytotoxic T cells. Of these, cytotoxic T lymphocytes (CTLs) directly eliminate pathogens by recognizing and killing cells infected by intracellular pathogens. CTLs use TCRs to survey antigens presented on major histocompatibility complex (MHC) class I on the surface of cells. Upon TCR recognition of cognate antigen-MHC I complexes, CTLs secrete cytokines and cytolytic molecules, thereby killing the target cell. CTLs are required for the control of many infections, including HIV, cytomegalovirus (CMV), and malaria (Li et al., 2016; Riddell et al., 1992; Rowland-Jones et al., 1997; Russell et al., 2017; Sobao et al., 2002).

Aberrant antibodies and T cell responses to self-antigens can cause autoimmune diseases such as type 1 diabetes (Gravano and Hoyer, 2013). Additionally, CTL recognition of tumor cells serves as the foundation for promising immunotherapies such as adoptive T cell transfer and T cell immune checkpoint blockade (Yang, 2015). A major ongoing challenge is the characterization of the antigens driving T cell activity in these contexts. Understanding the targets of T cell responses is critical to enable the effective harnessing and modulation of CTLs across human disease.

We and others have recently developed high-throughput approaches to interrogate B cell specificities. These technologies rely on displaying large candidate antigen libraries using phage display (PhIP-Seq, VirScan) (Larman et al., 2011; Xu et al., 2015), ribosome display (PLATO) (Zhu et al., 2013), or protein microarray (Forsström et al., 2014). Such unbiased profiling of antibody specificities reveals biomarkers of disease and insights about humoral immunity (Xu et al., 2016; Zhu et al., 2013). However, equivalent tools for comprehensive profiling of T cell specificities have not kept pace, limiting our ability to understand the adaptive immune system on a systems-wide scale.

Identifying T cell specificities is challenging for several reasons. First, T cell antigens are presented as short peptides non-covalently bound to MHC molecules, complicating the prediction and synthetic generation of candidate antigens. Second, TCRs have relatively low affinity for their targets (Stone et al., 2009). Finally, TCR signaling is complex as antigen binding does not uniformly lead to functional TCR signaling (Sibener et al., 2018).

Classic approaches for understanding T cell specificity rely on readouts of T cell function, which include assays for cytotoxicity, cytokine release, and proliferation in the presence of candidate antigens (Sharma and Holt, 2014), augmented by peptide-MHC tetramers for antigen-specific populations (Altman et al., 1996) and others. However, these are primarily useful for per-determined sets of 10–100 s of antigens but are unsuitable for unbiased discovery of antigens at genome scale (Bentzen et al., 2016, 2018; Hondowicz et al., 2012; Newell et al., 2013; Zhang et al., 2018).

Several other approaches have been taken to map unknown T cell specificities. A recent approach uses display of peptides as single-chain fusions to MHC on the surface of target cells. T cell binding to cognate antigen results in trogocytosis (Li et al., 2019) or activation of a synthetic signaling molecule (Joglekar et al., 2019), enabling the isolation of recognized target cells. Another approach uses display of genetically encoded random peptides covalently attached to MHC molecules on the surface of yeast (Birnbaum et al., 2014) or

baculovirus (Crawford et al., 2004) and use soluble TCRs to pan the yeast library for peptide-MHC binding. This approach suffers from the formidable challenge of mapping the random peptides identified onto the endogenous antigens they resemble (Gee et al., 2018). Moreover, neither the yeast display nor single-chain approaches require endogenous processing of antigens nor functional activation of T cells, which leaves uncertainty as to whether the identified antigens are physiologically meaningful *in vivo*. Notably, even high affinity TCR binders identified by yeast display do not uniformly lead to T cell function (Adams et al., 2011; Sibener et al., 2018). Overall, additional genome-wide tools are required to gain a more complete picture of the antigens functionally driving T cell responses and to enable the level of systematic profiling that is already accessible for the humoral response.

Here, we present a TCR epitope scanning method we call T-Scan, a high-throughput, genome-wide platform for the systematic identification of antigens productively recognized by T cells. T-Scan leverages genome-wide libraries encoding pathogen or human proteomes that are processed and presented on endogenous MHC-I molecules. Antigen recognition by cognate TCRs is monitored through a fluorescent reporter of granzyme B (GzB) activity in the target cells. This unbiased strategy eliminates the need for predictive algorithms, and by relying on the physiological activity of T cell killing, is not solely dependent on measuring TCR binding affinity to peptide-MHC complexes. We validate the utility of T-Scan by showing that it can identify known peptide epitopes of both viral and autoreactive TCRs from proteome-wide libraries and go on to exploit the method to generate high-resolution maps of TCR specificity and identify the epitopes recognized by multiple orphan TCRs.

RESULTS

Development of Reporters that Allow Genetic Detection of T Cell Epitopes

We set out to establish a genome-wide screening platform to uncover the specificity of TCRs. We initially considered two categories of approaches to T cell antigen discovery: a target depletion approach and a target enrichment approach. The former would rely upon killing of cognate target cells, allowing for depletion of the relevant target from a pool. However, dropout approaches suffer from low dynamic range and are difficult to multiplex, and detectable depletion would require potent T cells with the ability to all but eliminate a target from a population. In contrast, target enrichment approaches have the advantage that they have a wider potential dynamic range. For example, if one were to obtain 20% killing, for a dropout screen one would need to detect a difference between cognate antigen abundance of 100% and 80%, whereas in an enrichment screen with background killing of 1%, 20% killing could provide a 20-fold enrichment of the antigen. Thus, we set out to employ a target enrichment strategy to develop a T cell epitope screening method.

The T-Scan platform that we developed employs a cell-based pooled screen to identify the cognate antigens of T cells (Figure 1A). In T-Scan, target cells express a library of lentivirally-delivered candidate antigens that are processed and presented endogenously on MHC molecules. This library of target cells is co-cultured with T cells of interest, which secrete cytotoxic granules into cells displaying their cognate antigen. A GzB reporter, described below, is used to detect the target cells that receive these cytotoxic granules and to

enable their isolation by fluorescence-activated cell sorting (FACS). Finally, we use PCR and next-generation sequencing (NGS) to identify the antigens that these cells are programmed to express, thereby permitting a sequencing-based readout of T cell specificity.

To isolate target cells functionally recognized by a T cell, we developed reporters of GzB activity. GzB is a serine protease that is contained in the cytotoxic granules and which, when released, cleaves a broad set of proteins to trigger apoptosis (Casciola-Rosen et al., 2007). We reasoned that GzB activity is an attractive readout of T cell recognition because (1) GzB provides robust enzymatic activity that can amplify the signal of a limited number of secreted molecules, which provides sensitivity, (2) GzB activity is completely specific to target cells that have received cytotoxic granules, allowing for the isolation of rare cells killed by T cells while avoiding contamination from cells undergoing apoptosis for other reasons, and (3) GzB activity is a functional readout of physiologically relevant T cell activation.

We explored several reporters of GzB activity. The first examined allows for direct amplification of the antigen cassette from cells that have received GzB (Figure S1A). Cre recombinase is expressed in target cells, but is held inactive by being tethered to the plasma membrane; GzB cleaves the membrane tether, releasing Cre into the nucleus where it can catalyze the inversion of a reporter cassette adjacent to the expressed antigen. This inversion event can be detected by PCR, enabling the selective amplification of antigen cassettes expressed in GzB-positive cells from bulk genomic DNA without the need for cell sorting of targeted cells. We demonstrated a 4-fold increase in cassette inversion following the delivery of GzB into target cells (Figures S1B and S1C). However, despite efforts to decrease background, the signal-to-noise was not as high as we desired.

The second reporter is a genetically encoded protein that fluoresces following proteolysis by GzB (Figure 1B). We generated IFP^{GZB}, an infrared fluorescent protein (IFP)-based GzB reporter, by introducing a GzB-specific cleavage sequence (Choi and Mitchison, 2013) into the scaffold reported by To et al. (2015). This reporter is activated in cells co-cultured with cytotoxic T cells in the presence of cognate antigen, but, crucially, not in the presence of non-cognate antigen (Figure 1C). In screens, target cells activating the IFP^{GZB} reporter following co-culture are isolated by FACS. We observed better signal-to-noise with the fluorogenic GzB reporter than the Cre reporter, and hence we selected this reporter for further development.

A key element of the success of any screening strategy is the reduction of noise and preservation of signal. To promote efficient antigen screening, we made two additional modifications to the target cells. First, to reduce noise through the potential for presentation of cross-reactive peptide epitopes on irrelevant zMHC proteins, we used CRISPR to mutate all 6 endogenous HLA-A, HLA-B, and HLA-C MHC-encoding genes and re-expressed the individual HLA allele of interest. This ensured that any response detected would be restricted to the correct MHC protein and reduced background reactivity to peptides presented on other MHC alleles.

GzB delivery into cells results in the activation of caspases and subsequent apoptosis. During the process of GzB-induced apoptosis, caspases cleave the inhibitor of caspase-activated DNase (ICAD) protein, releasing active CAD nuclease. CAD fragments genomic DNA, which could impair our NGS analysis. Thus, second, to preserve signal, we expressed a caspase-resistant version of ICAD protein, $ICAD^{CR}$ (Sakahira et al., 1998), in the target cells to prevent genomic DNA fragmentation during apoptosis, enabling more efficient recovery of antigen information from the genomic DNA of GzB-positive cells. We termed the MHC-I null target cells expressing IFP^{GZB} and $ICAD^{CR}$ epitope discovery cells (EDCs).

T-Scan Allows Discovery of CMV-Encoded T Cell Epitopes

As a model for testing the T-Scan platform, we chose CMV-specific T cell responses. T cell responses to CMV have been intensively investigated and many reagents exist that facilitate our efforts to validate the T-Scan platform, including known immunodominant epitopes, antigen-specific TCRs, and peptide-specific tetramers (Klenerman and Oxenius, 2016; Schub et al., 2009). For these reasons, we set out to use CMV genome-wide screens with known CMV-reactive T cells to validate the T-Scan platform.

We synthesized a library of 5,764 oligonucleotides encoding 2,882 56-amino acid (aa) fragments that collectively tiled across the entire CMV proteome with 28-aa overlap between adjacent fragments (Figure 2A). To generate a comprehensive list of CMV antigens for our screens, we supplemented the reference CMV proteome with open reading frames (ORFs) identified by ribosome footprinting (Stern-Ginossar et al., 2012). To provide a measure of reproducibility, each of the peptides was encoded by two distinct nucleic acid sequences, which act as internal duplicates. This library was cloned into a lentiviral vector and transduced into EDCs expressing HLA-A2. To generate T cells for the screen, we transduced donor CD8 T cells with a vector expressing the CMV pp65 protein-specific NLV2 TCR (Schub et al., 2009). First, we verified that these T cells activated the GzB reporter in EDCs that expressed either full-length pp65 or a 56-aa fragment of pp65 containing the NLV epitope (Figure 1D). Then, after co-culture of the T cells with the target cell library, we sorted for GzB-positive EDCs, identified the antigens that they expressed by Illumina sequencing, and compared the abundance of each antigen before and after the sort. Strikingly, the two most enriched peptides identified by the T-Scan screen were the only two antigens in the library that harbored the NLV epitope (Figure 2B). This result was reproducible across the duplicate nucleic acid encodings of each antigen and across independent replicates of the screen (Figures 2C and S2).

Characterization and Optimization of the Performance of the CMV Screen

We repeated the screen using a second transduced NLV-specific TCR, NLV3 (Schub et al., 2009), and observed robust enrichment of the 56-mers expressing the cognate target across different effector:target ratios (Figure 2D). By performing the screen under various conditions, we found that T-Scan is robust to varying concentrations of T cells, the number of antigens expressed by each target cell (MOI), and the dilution of the cognate T cells with non-transduced T cells (Figures 2D–2F). We also demonstrated that assay performance is not affected by the presence of additional MHC alleles in the EDCs, suggesting that the assay can be used to map T cell responses on multiple alleles simultaneously (Figure S3). To

verify that T-Scan is compatible with non-transduced primary cells, we generated NLV-reactive T cells by expanding CD8 T cells from an HLA-A2-positive, CMV-positive donor in the presence of the NLV peptide. We screened these cells against the CMV-wide library and confirmed that the NLV-containing antigens were again the top hits (Figures 2G and 2H). The fold-enrichment observed with the primary T cells (60-to 100-fold) was greater than that achieved with the NLV2 TCR T cells (10- to 20-fold), consistent with increased signal to noise of GzB reporter activation by these cells. Unexpectedly, two overlapping 56-mers from the IE1 gene of CMV were also reproducibly enriched by the NLV-expanded T cells. IE1 was also observed in subsequent screens with these T cells and will be further explored below.

T-Scan Allows Virome-wide Discovery of T Cell Specificity

To determine whether T-Scan can be used with far larger numbers of candidate antigens, we performed a virome-wide screen (Figure 3A). We introduced a library of 93,904 56-aa fragments that collectively tiled across the entire human virome (Xu et al., 2015) into HLA-A2 EDCs and performed T-Scan on this antigen pool using the NLV-expanded primary T cells. In spite of the greatly increased complexity of the target library, the two most enriched peptides in the screen were, once again, the only two peptides in the library that contained the known NLV epitope (Figure 3B). Intriguingly, we also observed reproducible enrichment of the same two 56-mers from the IE1 gene of CMV that scored in the CMV-wide screen (Figures 2G and 2H). Because there was no sequence similarity between the NLV epitope and the IE1 peptide, we hypothesized that a subset of the T cells in the donor CD8 repertoire were reactive to an epitope in IE1 contained within the enriching 56mers. Analysis of the 28-aa region common to both 56-mers revealed a high-affinity HLA-A2 binding epitope using the NetMHC4.0 algorithm (Figure 3C), and we found that the NLV-expanded T cells, but not the NLV3 TCR-transduced T cells, were reactive to this peptide (Figure 3D). Moreover, tetramer staining revealed that while ~25% of the T cells in the NLV-expanded population recognized the NLV epitope, a distinct subpopulation comprising only ~2% of the T cells recognized the IE1 epitope (Figure 3E). Altogether, these experiments demonstrated that T-Scan can identify both known and novel targets of subclonal populations of T cells from highly complex antigen libraries and highlighted the potential of multiplexing T-Scan to discover the specificities of many T cells at once.

T-Scan-Mediated Discovery of Immunodominant CMV T Cell Epitopes

As T cells are key mediators of protective immunity to diverse pathogens, understanding the landscape of T cell antigens is critical to inform effective vaccine design. Having demonstrated the capacity of T-Scan to accurately identify T cell targets and uncovered the potential for multiplexing, we sought to identify T cell epitopes using collections of T cells with unknown specificities. Thus, we applied T-Scan to perform an unbiased, genome-wide search for T cell responses to CMV using T cells isolated directly from peripheral blood. Therefore, we purified bulk CD8 memory T cells from the blood of a different CMV-positive, HLA-A2 positive donor, expanded these cells *in vitro*, and used them to screen the CMV-wide library of 2,882 peptides in EDCs expressing only the HLA-A2 allele (Figure 4A).

This complex library-on-library screen identified a set of peptides that displayed reproducible enrichment (Figure 4B). Strikingly, these peptides could be grouped into six sets of overlapping 56-mers, suggesting that common epitopes lay in each 28-aa overlap region. Indeed, high-affinity HLA-A2 binding epitopes were predicted in all six 28-aa overlap regions, at a threshold where only 10% of random 28-aa stretches are predicted to contain a high-affinity epitope (Figure 4C). Furthermore, two peptides (1336 and 1337) contained the known immunodominant NLV epitope while two others (405 and 406) contained the previously reported LIEDFDIYV epitope (Elkington et al., 2003). We found that just 0.2% of the memory T cells employed for the screen recognized the NLV epitope, but we did not detect T cells specific for the LIEDFDIYV epitope using tetramer staining (Figure 4D).

None of the other four epitopes have been previously reported (Immune Epitope Database; Vita et al., 2018). We generated tetramers of predicted high-affinity peptides contained in these epitopes. We used these tetramers to stain the total memory T cells used in the screen and detected antigen-specific populations against all four, ranging from 0.04%–0.23% of the total memory T cells (Figure 4D). The limited sensitivity of tetramer staining and the possibility of alternative peptides encoded by each fragment may explain the absence of antigen 405-specific T cells and suggest that the detected cell abundances are a lower bound estimate (Rius et al., 2018). Together, these experiments validate that T-Scan can operate in a T cell library-versus-peptide library setting and demonstrate its utility in unbiased discovery of novel antigens. Furthermore, had our reporter cell line expressed the full set of MHC-I alleles from the donor, it is likely that we would have identified many additional epitopes.

Comprehensive Mapping of the TCR-Peptide Interface

Mapping the TCR-peptide interface in high resolution has enabled the identification of off-target reactivities of therapeutic TCRs (Cameron et al., 2013) and revealed insights about the cross-reactivity and signaling of T cells (Bulek et al., 2012; Wooldridge et al., 2012). We reasoned that T-Scan might enable facile mapping of the interaction landscape of various TCRs and their cognate antigens by epitope saturation mutagenesis. To test this, we generated a comprehensive single mutant library of the NLV epitope, with each mutant epitope present in the context of both a 56-aa fragment and a 9-aa fragment immediately followed by a stop codon (Figure 5A). In the 56-aa version, we also mutagenized the two amino acids immediately upstream and downstream of the NLV epitope, residues that in general are not predicted to affect TCR recognition, as controls. We synthesized oligonucleotides encoding this set of 418 mutants, together with four copies of the unmutated NLV epitope encoded by distinct synonymous codons that serve as positive controls, and 1,266 unrelated peptides that serve as negative controls.

This mutant library was introduced into HLA-A2-expressing EDCs, and T-Scan was performed using the NLV-expanded primary T cells. Verifying the robustness of our platform, we found that all (76/76) of the constructs with mutations in the irrelevant residues flanking the NLV epitope enriched comparably to the wild-type (WT) peptides and more strongly than all 1,266 negative control peptides (Figure S4). To delineate the critical binding residues, we compared the enrichment of each mutant to the WT NLV peptides

(Figure 5B). As expected, most mutations abrogated T cell killing, but we observed significant differences across the epitope: for example, almost all mutations at position one were tolerated, while any substitution at position four and five prevented recognition. Thus, these patterns likely reflect the exact contact residues within the epitope.

This capability provided an opportunity to probe the extent of variation in the interface of different TCRs recognizing the same epitope. We repeated the mutagenesis screen with T cells expressing two different TCRs specific for the NLV epitope (Figures 5C and 5D; Schub et al., 2009). Each TCR had a distinct footprint of recognized mutants, which was highly reproducible between the 56-aa and 9-aa versions of the libraries (Figure S5). We observed several commonalities between the TCR footprints. All of the TCRs were intolerant of any mutations at position four but tolerant of a largely conserved set of hydrophobic amino acid substitutions at position two, consistent with the requirement for hydrophobic amino acids at position two for HLA-A2 binding (Falk et al., 1991). Conversely, we observed striking differences in the mutations permitted at position one, with the NLV2 TCR showing complete specificity for asparagine, the NLV-expanded T cells permitting all mutants except proline and aspartic acid, and the NLV3 TCR showing an intermediate phenotype.

To confirm the accuracy of these footprinting studies, we selected mutants differentially recognized by the NLV2 and NLV3 TCRs for validation. Preferential recognition was conferred by the expected TCR for each of the 9 peptides examined, including several examples of completely TCR-specific peptides (Figure 5E). Thus, T-Scan can map the critical residues for TCR recognition and identify TCR-specific mimotopes.

Finally, we compared the landscape of recognized mutants to the predicted MHC-I binding affinity for each mutant. Because our assay requires endogenous processing and loading onto MHC-I, we reasoned that all enriched mutants should bind MHC-I. We used the NetMHC4.0 algorithm (Andreatta and Nielsen, 2016) to calculate the predicted affinity of each variant for HLA-A2. As expected, the vast majority of recognized mutants were predicted to retain MHC binding (Figure 5F). Notably, the sole mutations at position one (asparagine replaced with aspartic acid or proline) that abrogated recognition by all NLV TCRs are also predicted to have significantly lower affinity for MHC; thus, the lack of enrichment here may be due to lack of presentation rather than reduced TCR binding. We examined 2 variants in detail, L2A and L2C, which were recognized by our NLV-specific TCRs but had predicted HLA-A2 binding affinities of 2141 nM and 4404 nM respectively, much lower than the affinities typically observed for effective peptide-MHC antigens. We synthesized peptides containing these mutants and validated that they were recognized by the NLV3 TCR, but only at concentrations of peptide 10- to 100-fold higher than the WT, consistent with the predicted lower peptide binding affinity (Figure 5G). Thus, these experiments demonstrate that T-Scan can detect functional T cell interactions with low-affinity peptide antigens.

Genome-wide Discovery of Tumor-Reactive TCR Specificity

A key application of T cell antigen discovery is to understand the specificity of self-reactive TCRs implicated in autoimmunity and anti-tumor immunity. This is challenging due to the enormous complexity of the human proteome and the fact that self-reactive TCRs tend to

have low affinity for their antigens (Falkenburg et al., 2011). To determine whether T-Scan can be used to identify the targets of self-reactive TCRs, we performed a genome-wide screen using a tumor-derived TCR specific for an HLA-A1-restricted epitope of MAGE-A3 (Karanikas et al., 2003). We introduced a library of 259,345 antigens that tile across the entire human proteome in 90-aa fragments with 45-aa overlap into EDCs expressing HLA-A1 (Figure 6A). We infected the EDCs at an MOI of 5 to increase the representation of each antigen in our library as we observed good performance of a CMV screen performed with MOI 5 (Figure 2E). We introduced the MAGE-A3 TCR into donor CD8 T cells and used these cells to perform T-Scan with the genome-wide library. We observed strong and reproducible enrichment of only 8 antigens in the library, which strikingly encoded four sets of overlapping peptides (Figure 6B). Two of the top-scoring peptides were the only 2 library antigens containing the known MAGE-A3 epitope (Figure 6B). Furthermore, the 2 corresponding 90-mers from the related MAGE-A6 gene— that contain the same 9-mer HLA-A1 epitope apart from a single leucine to valine substitution—also enriched reproducibly (Figure 6C). The remaining 4 enriching peptides were from PLD5 and FAT2. We identified predicted high-affinity HLA-A1 binding peptides in the overlap region of both pairs of enriching peptides (Figure 6C). Notably, these epitopes share only three or four identical amino acids with the MAGE-A3 epitope, although they do share other structural similarities such as hydrophobic residues at positions 5 and 9. Moreover, all four identified targets contain a conserved EXDP motif at the N terminus of the epitope.

To validate the TCR reactivity to these 4 proteins, and to verify that the epitopes we identified are functionally presented in the context of full-length proteins, we tested the reactivity of the MAGE-A3 TCR to each antigen in the context of an exogenously added 9-aa peptide, an endogenously expressed 90-aa fragment and the full-length ORF. Due to its size (13 kb), we were unable to test full-length FAT2, but we generated 2 synthetic constructs (188 aa and 157 aa long) that retained the cellular localization of the epitope through a fusion to the signal peptide and transmembrane domain of FAT2 or CD8, respectively. We observed reactivity to the peptides, 90-aa fragments, and ORFs for all 4 antigens but not controls (Figure 6D). Overall, these experiments demonstrated that T-Scan can be used to discover the targets of self-reactive TCRs at genome scale.

DISCUSSION

A significant gap in our ability to translate the language of the immune system has been the limited ability to decipher the identity of epitopes recognized by T lymphocytes. In this study we describe the development of T-Scan, a high-throughput screening platform for the discovery of cytotoxic T cell antigens. T-Scan differs fundamentally from traditional approaches to antigen discovery: rather than measuring responding T cells directly, T-Scan detects the focused action of cytotoxic T cells on a library of candidate antigens. This is enabled through the development of a GzB reporter that allows enrichment of rare cognate target cells from candidate antigen libraries with high sensitivity and specificity. The key advance enabled by this approach is the ability to interrogate highly complex sets of candidate antigens in a high-throughput screen.

The validity of the T-Scan platform was demonstrated in the context of multiple T cell sources and different candidate antigen libraries. On the T cell side, we used donor T cells programmed with 2 different TCRs of known specificity and primary T cells expanded against a known antigen. As candidate antigens, we screened a CMV genome-wide library of 2,882 peptides and a virome-wide library of 93,904 peptides. In each screen, T-Scan robustly and reproducibly identified the known cognate antigen for the T cells used.

We applied T-Scan to three applications. First, we performed unbiased discovery of CMV antigens recognized by bulk memory CD8 T cells from a CMV seropositive normal donor peripheral blood sample. Despite extensive prior efforts to profile anti-CMV responses, four out of the six antigens we identified in this individual had not been previously reported (Elkington et al., 2003). Notably, one of the novel antigens that we identified is from the UL150A protein. UL150A was only recently discovered by high-throughput transcriptome profiling and is encoded in the antisense direction within the better-studied UL150 ORF (Gatherer et al., 2011; Stern-Ginossar et al., 2012). This demonstrates the ability of T-Scan to reveal unexpected new antigens and highlights the utility of combining systematic antigen discovery with unbiased approaches to characterizing protein expression. Importantly, this includes proteomic characterization of presented MHC peptides in various contexts that can then be encoded genetically and assayed with T-Scan.

Second, we used T-Scan to comprehensively map the TCR-peptide interface for three sets of NLV-specific T cells. We showed that T-Scan can accurately identify subtle differences in TCR specificity. Finally, we screened a self-reactive TCR against a library of 259,345 fragments encoding the entire human proteome. We identified the known target of the MAGE-A3 TCR and validated reactivity to a related protein, MAGE-A6 whose epitope differed by a single amino acid. Moreover, we discovered and validated 2 additional proteins, PLD5 and FAT2, which were also recognized by the TCR but less efficiently than the cognate epitopes. These epitopes shared structural similarity with the MAGE-A3 target but only had identical amino acids at three or four positions, illustrating the challenge of predicting additional TCR specificities in the absence of an unbiased experimental approach.

T-Scan Platform Features

The utility of the T-Scan platform is further enhanced by its ability to be multiplexed, leading to higher throughput. While the precise extent of multiplexing is likely to be determined by the potency of each T cell clone, antigen library complexity, and the screen scale, our experiments suggest that the collective specificity of 10 s to 100 s of T cells can be identified simultaneously at genome scale, thus enabling complex collections of T cells such as TILs to be simultaneously profiled. Notably, this ability is made possible by the use of an enrichment screen; rare subclonal T cell populations are unlikely to detectably deplete target antigens from complex libraries but can provide sufficient killing to enrich antigens. If desired, the precise TCR specific for each discovered target in a pooled screen can be identified in a second step of deconvolution, for example by using tetramer staining to identify and characterize T cells with particular antigen specificities or simply expressing the candidate antigens in target cells and examining T cell activation.

T-Scan maintains critical elements of the TCR-peptide-MHC interaction: natural TCR recognition and signaling, endogenous antigen processing, transportation into the ER, and proper loading onto MHC for presentation. These must occur for peptides to enrich in the screen, thereby enhancing the biological relevance of identified antigens. With respect to MHC binding, we identified two predicted low-affinity peptides that were functionally recognized by CMV-specific TCRs. These would be missed by discovery approaches that rely on MHC binding prediction algorithms to pre-select candidates, highlighting the advantage of unbiased and comprehensive antigen profiling.

Employing endogenous protein processing has the added feature of dramatically increasing the interrogated search space of antigens. Each expressed protein fragment enables hundreds of candidate peptides derived from the fragment to be assayed; in contrast, a total of 326 individual peptides would be required to encode all of the 8- to 11-aa linear candidate epitopes that can be derived from a single 90-aa fragment used in our human library. Full length ORFs are also compatible with T-Scan (Figures 1D and 6D), and when comprehensive collections are available it will enable proteome-wide screens with even lower practical screen complexity that will also include post-translational modifications that may not be present on some of the synthetic protein fragments.

Endogenous processing of proteins also enables the discovery of antigens formed by post-translational peptide splicing. Recent work has uncovered the substantial contribution of post-translational peptide splicing in shaping the repertoire of peptides bound to MHC, with a bias observed toward *cis*-splicing of proximate regions of proteins (Faridi et al., 2018; Liepe et al., 2016). While such spliced peptides are not directly encoded in the primary protein sequence, they could be formed during the endogenous processing of our antigens.

Furthermore, the synthetic approach of employing oligonucleotides covering entire proteomes and the ability to encode any expressed RNA element also provides advantages over current cDNA library approaches: (1) all peptides are present at similar frequencies; (2) the half-lives of full-length proteins are likely to differ drastically, while peptides are generally unstable (Koren et al., 2018) which may enhance presentation and detection; and (3) large libraries for any organism can be readily synthesized using long oligos. This allows the inclusion of unusual features such as peptides resulting from alternative RNA splicing or included introns.

Like all methods, T-Scan has limitations. First, candidate antigens must be able to be genetically encoded and therefore some post-translational modifications will be missed. This also prevents the mapping of non-peptide antigens such as CD1-bound lipids (Le Nours et al., 2018). Second, TCRs recognizing antigens endogenously expressed by our EDCs cannot be screened. This limitation can be potentially overcome by generating alternative target cell lines from other lineages.

Applications of T-Scan

We envision three main areas for T-Scan application: infectious disease, autoimmunity, and cancer immunology. In infectious disease, T-Scan can systematically profile the T cell targets in patients who exhibit protective immunity (e.g., immunity from controlled malaria

infection or elite control of HIV). These antigens inform vaccine design. In autoimmunity, T-Scan can be used to uncover the scope and features of auto-reactive T cell antigens. In addition, identified auto-reactive TCRs can be screened for cross-reactivity to viral or bacterial antigens, directly addressing hypotheses regarding the etiology of various autoimmune diseases. Finally, T-Scan has multiple applications to cancer immunology. First is the profiling of neoantigen reactivities in patient TILs or peripheral blood by screening patient-specific candidate neoantigen libraries. This can enable the rapid identification of productive neoantigens to guide personalized vaccines. Second is the genome-wide identification of the targets of orphan TCRs that are reactive to patient tumors. Third is the screening of therapeutic TCRs for potential off-target reactivities that can lead to unexpected patient toxicity (Cameron et al., 2013).

Altogether, the T-Scan suite of tools enables the high-throughput discovery and analysis of the targets of cytotoxic T cells. This will further enable the community of immunologists to explore the functionality of T cells in mice and humans. The knowledge gained from these efforts will expand our understanding of the roles of T cells in human health and lead to new therapeutics with which to treat human disease.

STAR★METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and request for reagents may be directed to the corresponding author Stephen J. Elledge (selledge@genetics.med.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines—HEK293T (female) cells were obtained from ATCC and were cultured in DMEM with 10% (v/v) FBS (Hyclone), 100 units/mL penicillin, and 0.1 mg/mL streptomycin. T cells were cultured in RPMI (Life Technologies) with 10% (v/v) FBS (Hyclone), 100 units/mL penicillin, 0.1 mg/mL streptomycin, and 50U/ml IL-2 (Sigma).

T cell expansion—Apheresis collars were obtained from the Brigham and Women's Hospital Specimen Bank under protocol T0276. Primary blood mononuclear cells (PBMCs) were purified on a Ficoll gradient. Briefly, donor blood was diluted 1:1 with PBS and gently layered on Ficoll-Paque (Thermo) and centrifuged at 400 g for 30 min. with the brake off. The cells at the interface were extracted, washed four times with PBS, and irradiated with 60 Gy IR.

For expansion, 1E6 T cells were added to 20E6 irradiated PBMCs in 20 mL final volume of RPMI, 10% FBS, 100 units/mL penicillin, 0.1 mg/mL streptomycin, 50U/ml IL-2 (Sigma), and 0.1 ug/ml anti-CD3 antibody (OKT3, ebioscience).

METHOD DETAILS

Knockout of endogenous HLA-A, HLA-B, and HLA-C genes—Guide RNAs were designed against sequences conserved across the HLA-A, HLA-B and HLA-C locus using the multicrispr.net tool (Prykhozhij et al., 2015). The following guides were selected:

CRISPR-ALL-1: CGGCTACTACAACCAGAGCG

CRISPR-ALL-2: AGATCACACTGACCTGGCAG

CRISPR-ALL-3: AGGTCAGTGTGATCTCCGCA

The gRNA were cloned into the LentiCRISPR V2 vector using the BsmBI sites. 10E6 HEK293T cells were transfected with 3 ug of each plasmid guide construct using Mirus TransIT. After 7 days, MHC-negative cells were sorted using the Pan-MHC antibody (BioLegend 311410). Clones were grown up from single cells and the absence of MHC was verified by flow cytometry and western blot.

Generation of IFP^{GZB}—The pcDNA3.1 iCasper (pTLT427) plasmid was kindly provided by Xiaokun Shu. To generate the IFP^{GZB} reporter, we replaced the caspase cleavage site (aa409–415, GDEVDjG) with a GzB-specific cleavage sequence (VGPDjFGR) by PCR mutagenesis. The resulting IFP^{GZB} construct (IFPGZB T2A HO1) was amplified by PCR and cloned into the pENTR/D-TOPO vector (Invitrogen) according to manufacturer's protocol. The IFP^{GZB} construct was then cloned into the pHAGE CMV DEST Hygromycin and pHAGE EF1a DEST Hygromycin lentiviral destination vectors by Gateway cloning (Thermo Fisher).

Generation of ICAD^{CR}—The ICAD^{CR} protein sequence was generated by modifying the ICAD protein sequence (DFF35 isoform, UniProt entry O00273–2) to include the D117E and D224E mutations and an N-Terminal Flag tag. The protein sequence was reverse-translated, synthesized as a gBlock (IDT), and cloned into the pENTR/D-TOPO vector (Invitrogen). The ICAD^{CR} cassette was cloned into the pHAGE CMV DEST Blasticidin and pHAGE EF1a DEST Blasticidin destination vectors by Gateway cloning (Thermo Fisher).

Generation of MHC allele expression constructs—The protein sequences for HLA-A1 (Uniprot P30443) and HLA-A2 (Uniprot P01892) were reverse translated and synthesized as gBlocks (IDT) and cloned into the pENTR/D-TOPO vector (Invitrogen). The HLA-A1 construct was cloned into the pHAGE EF1a DEST Hygromycin lentiviral destination vector and the HLA-A2 construct was cloned into the pHAGE CMV DEST Hygromycin lentiviral destination vector by Gateway cloning. The HLA-C202, HLA-C702, HLA-C401, and HLA-B702 sequences were obtained from the IPD-IMGT/HLA database (Robinson et al., 2015) and synthesized as gBlocks (IDT). They were cloned into pDONR221. For expression, the HLA-C202 allele was cloned into the pHAGE_EF1a_DEST_hyg lentiviral vector, the HLA-C702 was cloned into the pHAGE_EF1a_DEST_flox_BFP lentiviral vector, the HLA-C401 allele was cloned into the pHAGE_EF1a_DEST_NAT vector, and HLA-B702 was cloned into pHAGE_EF1a_DEST_hyg lentiviral vector (all by Gateway cloning).

Lentiviral Production

Lentivirus was produced by transfecting HEK293T cells with the lentiviral transfer vector in addition to the plasmids encoding Tat, Rev, Gag-Pol and VSV-G. Transfection was performed using the TransIT-293 transfection reagent (Mirus) or PolyJet *In Vitro* DNA Transfection Reagent (SignaGen) according to manufacturer's protocol. Lentiviral

supernatants were collected 48 h post-transfection and filtered through a 0.45 um filter and added to cells in the presence of 4 ug/ml Polybrene.

TCR design and cloning—The NLV2 and NLV3 TCRs were reconstructed from Schub et al. (2009). Each TCR was encoded as a single construct containing TCRa P2A TCRb with the human TCR constant regions. The sequences encoding these TCRs were synthesized as gBlocks (IDT) and cloned into the pENTR/D-TOPO vector and then transferred by Gateway cloning into the pHAGE EF1a DEST ZsGreen vector.

The MAGE-A3 TCR was reconstructed from Karanikas et al. (2003). The TCR was encoded as a single construct containing TCRα P2A TCRβ with the mouse TCR constant regions. The sequence encoding this TCR was synthesized as a gBlock (IDT) and cloned into the pENTR/D-TOPO vector and then transferred by Gateway cloning into the pHAGE EF1a DEST ZsGreen vector.

T cell transduction—CD8 T cells were purified from donor blood using the RosetteSep CD8 purification kit (StemCell) according to manufacturer's protocol. 1E6 cells were seeded per well of a 24-well plate and stimulated with a 1:1 ratio of Dynabeads Human T-Activator CD3/CD28 (Life Technologies). At the time of activation, cells were simultaneously transduced with 100 ul of lentivirus encoding the TCR of interest. After three days, transduced T cells were sorted based on the presence of the GFP marker and purified for CD8 expression with an anti-CD8 APC antibody (Biolegend).

Generation of NLV-expanded T cells—PBMCs were seeded at 1E6/ml in RPMI with 10% Human AB serum and 1ug/ml NLV peptide (NLVPMVATV). The next day, 600 IU/ml IL-2 was added to the cells. The cells were frozen 11 days post-stimulation.

GzB reporter activation assays with peptides—EDCs were labeled with CellTrace violet (Thermo) according to manufacturer's protocol and seeded at 3E4 cells/well in 96-well plates. The next day, cells were pulsed with the peptides of interest at the desired concentrations (1 uM unless otherwise specified) for 1 h at 37C. The media was then aspirated and replaced by the T cells of interest in T cell media. Plates were spun down for 1 min at 300 g to contact the T cells and EDCs. The plates were incubated for 4–6 h at 37C, then the cells were harvested by pipetting up and down, and immediately analyzed by flow cytometry on the BD LSRII. Target cells were identified by the presence of the CellTrace violet dye (BV421 channel) and IFP^{GZB}-positive cells were identified by a shift in the APC-Cy7 channel. The following peptides were used (Genscript):

PP65 peptide (NLV): NLVPMVATV

IV9 peptide (negative control): ILKEPVHGV

MAGE-A3 peptide: EVDPIGHLY

MAGE-A6 peptide: EVDPIGHVY

PLD5 peptide: ETDPLTFNF

FAT2 peptide: ETDPVNHMV

GzB reporter activation assays with endogenously expressed antigens—The following peptides were reverse translated and synthesized as gBlocks (IDT) with 5’ and 3’ BP recombination sites.

PP65 56-mer #1:
GVMTRGRLKAESTVAPEEDTDEDSDNEIHNPAVFTWPPWQAGILARNLVPMV ATV Q

PP65 56-mer #2: HNPAVFTWPPWQAGILARNLVPMVATV
QGQNLKYQEFFWDANDIYRIFAELEGVWQ

MAGEA3 90-mer:
MGEPVTKAEMLGSVVGNWQYFFPVIFSKASSSLQLVFGIELM**EVDPIGHLY**IF ATCLGLSYDGLLGDNQIMPKAGLLIIVLAIIAREGDCAP

MAGEA6 90-mer:
MGEPVTKAEMLGSVVGNWQYFFPVIFSKASDSLQLVFGIELM**EVDPIGHVY**I FATCLGLSYDGLLGDNQIMPKTGFLIIILAIIAKEGDCAP

PLD5 90-mer:
MGYLPISSTSTKRTYWPDLDAKIREALVLRSVRVRLLLSFWK**ETDPLTFNF**ISS LKAICTEIANCSLKVKFFDLERENACATKEQKNHTFPR

FAT2 90-mer:
MGPPLSASVRLHIEWIPWPRPSSIPLAFDETYYSFTVM**ETDPVNHMV**GVISVE GRPGLFWFNISGGDKDMDFDIEKTTGSIVIARPLDTRRR

FAT2 ORF CD8TM:
MALPVTALLLPLALLLHAARPSQFRVSPPLSASVRLHIEWIPWPRPSSIPLAFDE TYYSFTVM**ETDPVNHMV**GVISVEGRPGLFWFNISGGDKDMDFDIEKTTGSIV IARPLDTRRRAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCNHRN

FAT2 ORF FAT2TM:
MTIALLGFAIFLLHCATCEKPLEGILSSSPPLSASVRLHIEWIPWPRPSSIPLAFD ETYYSFTVM**ETDPVNHMV**GVISVEGRPGLFWFNISGGDKDMDFDIEKTTGSI VIARPLDTRRREMEARGCSEGHCLVTPEIQRGDWGQQELLIITVAVAFIIISTVG LLFYCRRCKSHKPVAMEDPDLLARS

The full-length PP65 gene was synthesized as a gBlock (IDT) with 5’ and 3’ BP recombination sites. The constructs were cloned into pDONR221 using the Gateway BP Clonase (Thermo). The PLD5 full-length ORF was sourced from the ORFeome v8.1 collection. MAGEA3 and MAGEA6 full-length ORFs were sourced from the Ultimate ORF collection (Thermo). The antigen constructs were cloned into the pHAGE CMV NFlagHA DEST IRES Puro (pp65) or pHAGE EF1a DEST puro (MAGEA3, MAGEA6, PLD5, FAT2) lentiviral vector by Gateway cloning. Antigen constructs were packaged into lentivirus and introduced into HLA-A1 EDCs (MAGE-A3, MAGE-A6, PLD5, and FAT2) or HLA-A2 EDCs (pp65 56mers and ORF) at an MOI of < 0.5. After 48 h, cells expressing the antigens were selected in 1 ug/ml puromycin for 3 days.

EDCs expressing antigens were labeled with CellTrace violet (Thermo) according to manufacturer's protocol and seeded at 3E4 cells/well in 96-well plates. The next day, the media was aspirated and replaced by the T cells of interest in T cell media. Plates were spun down for 1 min. at 300 g to contact the T cells and EDCs. The plates were incubated for 4–6 h at 37° C, then the cells were harvested by pipetting up and down, and immediately analyzed by flow cytometry on the BD LSRII. Target cells were identified by the presence of the CellTrace violet dye (BV421 channel) and IFP^{GZB}-positive cells were identified by a shift in the APC-Cy7 channel.

Tetramer staining validation—The following peptides were synthesized (Genscript):

PP65 peptide (NLV): NLVPMVATV

IV9 peptide (negative control): ILKEPVHGV

2140: GVLDAVWRV

2058: ALWDVALLEV

1560: TLIVNLVEV

670: CLLESVYTA

405: LIEDFDIYV

Peptides were loaded at 100 ug/ml onto the QuickSwitch Quant HLA-A*02:01 Tetramers (PE or APC labeled) (MBL International) according to manufacturer's instructions. Tetramers were used for staining at 5 ug/ml final concentration. Where specified, cells were additionally stained with a Pacific Blue anti-CD8 antibody (Biolegend) and Alexa Fluor 488 anti-human TCR α/β Antibody (Biolegend).

Genomic DNA purification and NGS library preparation—Sorted EDCs were mixed with 2E5 carrier wild-type HEK293T cells and gDNA was purified using the GeneJet gDNA purification kit (Thermo) according to manufacturer's protocol. At least 100x representation of unsorted library cells were collected, prepped and prepared as an input sample for each screen. Antigen libraries were prepared for multiplexed Illumina sequencing using a modified version a previously published protocol (Xu et al., 2015). The libraries were amplified in three rounds of PCR using the hot start Q5 polymerase according to the manufacturer-suggested protocol (Thermo).

For the CMV library, Virome library, and tiling mutagenesis library screens, in the first PCR, 1 ug of gDNA per 100 ul reaction was amplified using primers flanking the antigen cassette and adding on adaptors (T Cell PCR1_F: 5' TCCAGTCAGGTGTGATGCTCGGGGATCCAGGAATTCAGTTTGTACAAAAAAGCAGGCTCA; T cell PCR1_R: 5' CGAGCTTATCGTCGTCATCCCCACTTTGTACAAGAAAGCTGGGTCA). PCR1 reactions from a single screen replicate were pooled and 1ul was used as input for a 50ul PCR2 reaction (T cell PCR2_F: 5' ACACTCTTTCCCTACACGACTCCAGTCAGGTGTGATGCTC; T cell PCR2_R: 5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCGAGCTTATCGTCGTCATCC).

One ul of PCR2 was used as input for a 50ul PCR3 reaction to add on the sequencing adaptors and sample-specific index (T cell PCR3_F: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACTCCAGT; T cell PCR3_R: CAAGCAGAAGACGGCATACGAGATxxxxxxxGTGACTGGAGTTCAGACGTGT where "**xxxxxxx**" denotes a unique 7-nt indexing sequence). Samples were pooled and gel extracted using the QiaQuick Gel Extraction columns (QIAGEN). Samples were sequenced on the Illumina MiSeq or Illumina NextSeq using the T7-Pep2.2_Illumina_SP (5' GGTGTGATGCTCGGGGATCCAGGAATTC).

For the human genome-wide screen, in the first PCR, 1 ug of gDNA per 100 ul reaction was amplified using primers flanking the antigen cassette (Pep_PCR1_F: 5' agctcgatggaggataccca; Pep_PCR1_R: 5' ccaaaagacggcaatatggtgg). PCR1 reactions from a single screen replicate were pooled and 1 ul was used as input for a 50 ul PCR2 reaction with a pool of 8 "stagger" forward primers (Pep_PCR2_F: 5' tccctacacgacgctcttccgatct[]GTTTGTACAAAAAAGCAGG, where the [] denotes between 0 and 7 bases (full sequences in Key Resources Table); Pep_PCR2_R: 5' gtgactggagttcagacgtgtgctcttccgatctGCCTTATTCCAAGCGGCTTC). One ul of PCR2 was used as input for a 50 ul PCR3 reaction to add on the sequencing adaptors and sample-specific index (Pep_PCR3_F: 5' aatgatacggcgaccaccgagatctacactcttTCCCTACACGACGCTCTTCCG; T cell PCR3_R: CAAGCAGAAGACGGCATACGAGATxxxxxxxGTGACTGGAGTTCAGACGTGT where "xxxxxxx" denotes a unique 7-nt indexing sequence). Samples were pooled and gel extracted using the QiaQuick Gel Extraction columns (QIAGEN). Samples were sequenced on the Illumina MiSeq or Illumina NextSeq using the standard Illumina Sequencing primer.

CMV library design and cloning—To generate a complete set of potential CMV antigens, we used the ribosomal footprinting data generated by Stern-Ginossar et al. (2012). We included all ORFs and uORFs detected that were at least 8aa long. For cases of multiple overlapping in-frame ORFs, we selected the longest version ORF to include in our library. This set of ORFs was supplemented with 20 ORFs from the Merlin strain of CMV that were not detected by ribosomal footprinting, for a total of 561 ORFs. We tiled across each ORF in 56 aa fragments with 28 aa overlap between adjacent tiles for a total of 2882 protein fragments. Fragment tiles were reverse translated using non-rare human codons. Each fragment was encoded by two nucleic acid sequences that differed by at least two bases in the 5' 50bp for a total of 5764 oligo sequences.

The oligo library was synthesized by Twist Biosciences with the following adaptors:

five prime adaptor: 5' TGAATTCTGAGCTCG

three prime adaptor: 5' TCGGGTGCTCGAGCT

The library was amplified with the following primers, which contain overhangs encoding the BP recombination sites:

CMV_BP F: 5' ggggacaagtttgtacaaaaaagcaggctcaGGAATTCTGAGCTCG

CMV_BP R: 5’ ggggaccactttgtacaagaaagctgggtcagctagttaAGCTCGAGCACCCGA

The amplified library was cloned into the pDONR221 vector using BP Clonase (Thermo) and then transferred into the pHAGE CMV NFlagHA DEST IRES puro lentirival vector using Gateway cloning. This vector provided a uniform start codon followed by N-terminal Flag and HA tag for the expression of each peptide. At least 100x library representation was maintained during all cloning steps.

The library was packaged into lentivirus and transduced at a representation of 1000 and MOI of 0.5 into EDCs stably expressing pHAGE CMV ICAD^{CR}, pHAGE CMV IFP^{GZB}, and pHAGE CMV HLA-A2. The cells were selected with 1 ug/ml puromycin for 3d beginning 48 h post-transduction. Two replicates of library were generated with independent infections and used for the NLV2 CMV screen, NLV-expanded T cell screen of CMV library, and CMV library versus library screen.

NLV2 CMV screen—Three replicates of 10E6 EDCs (1700x library representation each) expressing HLA-A2 and the CMV library were co-cultured with 50E6 NLV2 TCR T cells for 12 h, after which IFP^{GZB}-positive target cells were sorted (FacsAria II). For data visualization, the geometric mean of the fold-enrichment of each peptide across the three replicates was calculated and added to 0.25 to enable plotting in log format.

CMV optimization screens—For each condition, 10E6 EDCs (1700x library representation) expressing HLA-A2 and the CMV library were co-cultured with NLV3 TCR T cells for 12h, after which IFP^{GZB}-positive target cells were sorted (FacsAria II). Unless otherwise specified, 5E6 T cells were used for each condition and EDCs expressing antigens at an MOI of 1 were used. For the effector to target ratio experiment, EDCs were co-cultured with 20E6 (2:1), 5E6 (1:2), or 1.25E6 (1:8) NLV3 TCR T cells. For the MOI experiment, HLA-A2-expressing EDCs were transduced with the CMV library at an MOI of 1 or 5 and selected for 3 d with 1 ug/ml puromycin before use in the screen. For the T cell dilution experiment, 1.25E6 NLV3 TCR T cells were mixed with 3.75E6 MAGE-A3 TCR T cells and compared to a pure population of 5E6 NLV3 TCR T cells.

NLV-expanded T cell screen of CMV library—Three replicates of 10E6 EDCs (1700x library representation each) expressing HLA-A2 and the CMV library were co-cultured with 25E6 NLV-expanded T cells for 12 h, after which IFP^{GZB}-positive target cells were sorted (FacsAria II).

NLV-expanded T cell screen of virome-wide library—The virome-wide library was previously described (Xu et al., 2015). The library was cloned into the pHAGE CMV NFlagHA DEST IRES puro lentiviral vector by Gateway cloning. This vector provided a uniform start codon followed by N-terminal Flag and HA tag for the expression of each peptide. At least 100x library representation was maintained at each step of cloning. Two replicates of 100E6 EDCs (library representation ~1000x) expressing HLA-A2 were transduced with the virome library at an MOI of 1 and selected with 1 ug/ml puromycin for 3 d.

For the screen, 4 replicates (2 from each independent infection) of 80E6 EDCs (850x representation per replicate) were co-cultured with 50E6 NLV-expanded T cells for 8h, after which IFP^{GZB}-positive target cells were sorted (FacsAria II).

Tiling mutagenesis library design and cloning—The NLV epitope was mutagenized in the context of a 56-mer (PWQAGILARNLVPMVATVQGQNLKYQEFFWDANDIYRIFAE LEGVWQPAAQPKRRR) and a 9-mer (**NLVPMVATV**). Each amino acid in the NLV epitope (bold) was mutated to each of the other 19 amino acids. As a control, the two adjacent amino acids outside the NLV epitope in the 56 aa version (underlined) were also mutated to each of the other 19 amino acids. This set of 418 mutants was combined with 4 versions of the WT epitope and 1266 unrelated peptide sequences for a library of 1688 epitopes. Each peptide was reverse translated with non-rare human codons in two different nucleic acid sequences for a total of 3376 oligo sequences. The library was synthesized by Twist Biosciences.

The library was amplified with the following primers, which contain overhangs encoding the BP recombination sites:

CTL_MUT_BP F: 5’ ggggacaagtttgtacaaaaaagcaggctcaAGAATTCTCCGTGGC

CTL_MUT_BP R: 5’ ggggaccactttgtacaagaaagctgggtcagctagttaCACTCGAGAGCTCAC

The amplified library was cloned into the pDONR221 vector using BP Clonase (Thermo) and then transferred into the pHAGE CMV NFlagHA DEST IRES puro lentirival vector using Gateway cloning. This vector provided a uniform start codon followed by N-terminal Flag and HA tag for the expression of each peptide. At least 100x library representation was maintained during all cloning steps.

The library was packaged into lentivirus and transduced at a representation of 1000 and MOI of 0.5 into EDCs stably expressing pHAGE CMV $ICAD^{CR}$, pHAGE CMV IFP^{GZB}, and pHAGE CMV HLA-A2. The cells were selected with 1 ug/ml puromycin for 3 d beginning 48 h post-transduction and used for the tiling mutagenesis screens with the NLV2 TCR, NLV3 TCR, and NLV-expanded T cells.

Tiling mutagenesis screen—For the NLV-expanded cell screen, 3 replicates of 25E6 HLA-A2 EDCs expressing the tiling mutagenesis library (6500x library representation per replicate) were co-cultured with 25E6 NLV-expanded cells. For the NLV2 and NLV3 TCRs, 3 replicates of 12.5E6 HLA-A2 EDCs expressing the tiling mutagenesis library (3200x library representation per replicate) were co-cultured with 6.2E6 NLV2 or NLV3 TCR T cells. After 12 h, IFP^{GZB}-positive target cells were sorted (BD FacsAria II).

For analysis, the reads for the two barcodes of each peptide across three replicates of the screen were summed. To calculate the fold-change, the fractional abundance of each peptide was then divided by the fractional abundance of that peptide in the pre-sorted input library. The normalized fold-change that was depicted in the heatmaps was generated by dividing

the fold-change of each peptide by the average fold-change of the wild-type NLV peptides in the library.

Peptidome-wide screen—The 90 aa peptidome-wide library was described previously (Xu et al., 2016). The library was cloned into the pHAGE CMV NFlagHA DEST IRES puro lentiviral vector by Gateway cloning. This vector provided a uniform start codon followed by N-terminal Flag and HA tag for the expression of each peptide. At least 100x library representation was maintained at each step of cloning. 100E6 EF1a EDCs (EF1a ICAD^{CR}, EF1a IFP^{GZB}, EF1a HLA-A1) were transduced with the genome-wide library at an MOI of 5 (2000x library representation) and selected with 1 ug/ml puromycin for 3 d.

Eight replicates of 100E6 EDCs (2000x library representation per replicate) were co-cultured with 85E6 MAGE-A3 TCR T cells for 8–12 h and sorted for IFP^{GZB}-positive cells (BD FacsAria II).

To calculate the fold-change, the fractional abundance of each peptide was then divided by the fractional abundance of that peptide in the pre-sorted input library.

CMV library versus library screen—PBMCs from an HLA-A2 positive, CMV-positive donor were obtained (Astarte Biologicals). Memory CD8 T cells were purified using the Miltenyi CD8+ Memory T cell isolation kit according to manufacturer's instructions and expanded using the T cell expansion protocol described above.

Four replicates of 30E6 HLA-A2 EDCs expressing the CMV library (5000x library representation per replicate) were co-cultured with 25E6 T cells for 8h after which IFPGZB-positive target cells were sorted (BD FacsAria II).

For analysis, the relative fractional abundance of each peptide compared to the pre-sort input library was independently calculated for each replicate. For data visualization, we calculated a modified geometric mean of performance of each peptide in the four replicates. To penalize but not completely discount peptides that enriched in 3 out of 4 replicates, we added 0.1 to the relative fractional abundance of each peptide in each replicate prior to taking the geometric mean.

Cre inversion experiment—The Cre^{membrane} construct was designed by fusing together a signal peptide (MALPVTALLLPLALLLHAARPSQ), a flag tag (DYKDDDDK), the CD8 transmembrane domain, a GzB cleavage site (VGPDFGR), the Cre recombinase protein, another GzB cleavage site, and the ERT2 domain. This construct was synthesized as a gBlock (IDT) and cloned into the pENTR/D-TOPO (ThermoFisher) and then into the lentiviral pHAGE CMV hygro destination vector and introduced into K562 target cells by lentiviral transduction.

The single-inversion reporter of Cre activity consisted of GFP and E2-Crimson fluorescent proteins in a head-to-head orientation flanked by two sets of orthogonal Cre recombination sites designed to trap single inversion events based on the FLEX system (Schnütgen et al., 2003). Inversion of this reporter was detected by digital droplet PCR (BioRad) with the following primers and probe:

FLEX inversion FWD: GCCGCCCCCTTCACCATAG

FLEX inversion REV: TCCAGCTCGACCAGGATG

FLEX inversion probe: TGAGCGATCACGCATAAGGG

The double-inversion reporter of Cre activity consisted of inverted primer binding sites separated by a 99 bp spacer sequence flanked by sets of orthogonal suicide Cre recombination sites (loxRE/loxLE and lox66/lox71). Cre activity that inverted both sets of Cre recombination sites was detected by droplet PCR (BioRad) with the following primers and probe:

Double-inversion FWD: TCACCGTTCCTTGTGGTAATC

Double-inversion REV: CGTTGATGTGGATCGACTCTATAA

Double-inversion probe: ACCCTTATGCGTGATCGC

This reporter cassette was cloned into a pHAGE CMV lentiviral vector and introduced into reporter cells by lentiviral transduction (hygromycin selection 4d at 200ug/ml). Finally, the caspase-resistant D117E ICAD gene was introduced into the target cells by lentiviral transduction (blasticidin selection 5 d at 40 ug/ml). The target cells were mixed with a 2:1 excess of activated primary NK cells for 4 h. Genomic DNA was purified using the GeneJet purification kit and the inversion of the Cre reporter was quantified by qPCR using inversion-specific primers. The frequencies of inversion events were normalized to the abundance of the reporter cassette, based on the following primers and probes that constitutively amplified the blasticidin resistance cassette:

Bsd FWD: TGGCAACCTGACTTGTATCG

Bsd REV: GTCCATCACTGTCCTTCACTATC

Bsd probe: CGACAGGTGCTTCTCGATCTGCAT

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details of experiments can be found in the figure legends. Data analysis was performed in Python, Excel, and GraphPad Prism. All error bars in figures indicate standard deviation.

Sequence alignment and analysis—Illumina reads were aligned using Bowtie (Langmead et al., 2009). For the human genome-wide screen, adaptors were trimmed using Cutadapt (Martin, 2011) prior to alignment. The fractional abundance of each antigen in each screen replicate was divided by the fractional abundance in the pre-sort input library to calculate the fold-change of the antigen.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank G. Xu, C. Harvey, B. Larman, E. Shrock, R.Timms, and P. Bruno for helpful discussions. This research was supported by NIH (AI116833 to S.J.E.), Department of Defense Distinguished Investigator Award (W81XWH-18-1-0469 to S.J.E.), NCI (T32CA009216 to C.I.W), the Ludwig Center at Harvard, the Department of Defense (BC113107 and BC132150P1 to H.K.L. and BC170281 to Z.C.H.), and Susan G. Komen (CCR14299200 to Z.C.H.). T.K. was supported by the NSF Graduate Research Fellows Program and the Harvard Innovation Grant Program. S.J.E. is an Investigator with the Howard Hughes Medical Institute.

REFERENCES

Adams JJ, Narayanan S, Liu B, Birnbaum ME, Kruse AC, Bowerman NA, Chen W, Levin AM, Connolly JM, Zhu C, et al. (2011). T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. Immunity 35, 681–693. [PubMed: 22101157]

Altman JD, Moss PA, Goulder PJ, Barouch DH, McHeyzer-Williams MG, Bell JI, McMichael AJ, and Davis MM (1996). Phenotypic analysis of antigen-specific T lymphocytes. Science 274, 94–96. [PubMed: 8810254]

Andreatta M, and Nielsen M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics 32, 511–517. [PubMed: 26515819]

Bentzen AK, Marquard AM, Lyngaa R, Saini SK, Ramskov S, Donia M, Such L, Furness AJS, McGranahan N, Rosenthal R, et al. (2016). Large-scale detection of antigen-specific T cells using peptide-MHC-I multi-mers labeled with DNA barcodes. Nat. Biotechnol. 34, 1037–1045. [PubMed: 27571370]

Bentzen AK, Such L, Jensen KK, Marquard AM, Jessen LE, Miller NJ, Church CD, Lyngaa R, Koelle DM, Becker JC, et al. (2018). T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide-MHC complexes. Nat. Biotechnol Published online 11 19, 2018 10.1038/nbt.4303.

Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, Dobbins J, Ozkan E, Davis MM, Wucherpfennig KW, and Garcia KC (2014). Deconstructing the peptide-MHC specificity of T cell recognition. Cell 157, 1073–1087. [PubMed: 24855945]

Bulek AM, Cole DK, Skowera A, Dolton G, Gras S, Madura F, Fuller A, Miles JJ, Gostick E, Price DA, et al. (2012). Structural basis for the killing of human beta cells by CD8(+) T cells in type 1 diabetes. Nat. Immunol. 13, 283–289. [PubMed: 22245737]

Cameron BJ, Gerry AB, Dukes J, Harper JV, Kannan V, Bianchi FC, Grand F, Brewer JE, Gupta M, Plesa G, et al. (2013). 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.

Casciola-Rosen L, Garcia-Calvo M, Bull HG, Becker JW, Hines T, Thornberry NA, and Rosen A. (2007). Mouse and human granzyme B have distinct tetrapeptide specificities and abilities to recruit the bid pathway. J. Biol. Chem. 282, 4545–4552. [PubMed: 17179148]

Choi PJ, and Mitchison TJ (2013). Imaging burst kinetics and spatial coordination during serial killing by single natural killer cells. Proc. Natl. Acad. Sci. USA 110, 6488–6493. [PubMed: 23576740]

Crawford F, Huseby E, White J, Marrack P, and Kappler JW (2004). Mimotopes for alloreactive and conventional T cells in a peptide-MHC display library. PLoS Biol. 2, E90.

Elkington R, Walker S, Crough T, Menzies M, Tellam J, Bharadwaj M, and Khanna R. (2003). Ex vivo profiling of CD8+−T-cell responses to human cytomegalovirus reveals broad and multispecific reactivities in healthy virus carriers. J. Virol. 77, 5226–5240. [PubMed: 12692225]

Falk K, Rötzschke O, Stevanović S, Jung G, and Rammensee H-G (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351, 290–296. [PubMed: 1709722]

Falkenburg WJJ, Melenhorst JJ, van de Meent M, Kester MGD, Hombrink P, Heemskerk MHM, Hagedoorn RS, Gostick E, Price DA, Falkenburg JHF, et al. (2011). Allogeneic HLA-A*02-restricted WT1-specific T cells from mismatched donors are highly reactive but show off-target promiscuity. J. Immunol. 187, 2824–2833. [PubMed: 21821799]

Faridi P, Li C, Ramarathinam SH, Vivian JP, Illing PT, Mifsud NA, Ayala R, Song J, Gearing LJ, Hertzog PJ, et al. (2018). A subset of HLA-I peptides are not genomically templated: Evidence for cis-and trans-spliced peptide ligands. Sci. Immunol. 3, eaar3947.

Forsström B, Axnäs BB, Stengele K-P, Bühler J, Albert TJ, Richmond TA, Hu FJ, Nilsson P, Hudson EP, Rockberg J, and Uhlen M. (2014). Proteome-wide epitope mapping of antibodies using ultra-dense peptide arrays. Mol. Cell. Proteomics 13, 1585–1597. [PubMed: 24705123]

Gatherer D, Seirafian S, Cunningham C, Holton M, Dargan DJ, Baluchova K, Hector RD, Galbraith J, Herzyk P, Wilkinson GWG, and Davison AJ (2011). High-resolution human cytomegalovirus transcriptome. Proc. Natl. Acad. Sci. USA 108, 19755–19760. [PubMed: 22109557]

Gee MH, Han A, Lofgren SM, Beausang JF, Mendoza JL, Birnbaum ME, Bethune MT, Fischer S, Yang X, Gomez-Eerland R, et al. (2018). Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. Cell 172, 549–563. [PubMed: 29275860]

Gravano DM, and Hoyer KK (2013). Promotion and prevention of autoimmune disease by CD8+ T cells. J. Autoimmun. 45, 68–79. [PubMed: 23871638]

Hondowicz BD, Schwedhelm KV, Kas A, Tasch MA, Rawlings C, Ram-churren N, McIntosh M, D'Amico LA, Sanda S, Standifer NE, et al. (2012). Discovery of T cell antigens by high-throughput screening of synthetic minigene libraries. PLoS ONE 7, e29949.

Joglekar AV, Leonard MT, Jeppson JD, Swift M, Li G, Wong S, Peng S, Zaretsky JM, Heath JR, Ribas A, et al. (2019). T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. Nat. Methods 16, 191–198. [PubMed: 30700902]

Karanikas V, Lurquin C, Colau D, van Baren N, De Smet C, Lethé B, Connerotte T, Corbiére V, Demoitié M-A, Liénard D, et al. (2003). Monoclonal anti-MAGE-3 CTL responses in melanoma patients displaying tumor regression after vaccination with a recombinant canarypox virus. J. Immunol. 171, 4898–4904. [PubMed: 14568971]

Klenerman P, and Oxenius A. (2016). T cell responses to cytomegalovirus. Nat. Rev. Immunol. 16, 367–377. [PubMed: 27108521]

Koren I, Timms RT, Kula T, Xu Q, Li MZ, and Elledge SJ (2018). The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. Cell 173, 1622–1635. [PubMed: 29779948]

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. [PubMed: 19261174]

Larman HB, Zhao Z, Laserson U, Li MZ, Ciccia A, Gakidis MAM, Church GM, Kesari S, Leproust EM, Solimini NL, and Elledge SJ (2011). Autoantigen discovery with a synthetic human peptidome. Nat. Biotechnol. 29, 535–541. [PubMed: 21602805]

Le Nours J, Shahine A, and Gras S. (2018). Molecular features of lipid-based antigen presentation by group 1 CD1 molecules. Semin. Cell Dev. Biol. 84, 48–57. [PubMed: 29113870]

Li X, Huang J, Zhang M, Funakoshi R, Sheetij D, Spaccapelo R, Crisanti A, Nussenzweig V, Nussenzweig RS, and Tsuji M. (2016). Human CD8+ T cells mediate protective immunity induced by a human malaria vaccine in human immune system mice. Vaccine 34, 4501–4506. [PubMed: 27502569]

Li G, Bethune MT, Wong S, Joglekar AV, Leonard MT, Wang JK, Kim JT, Cheng D, Peng S, Zaretsky JM, et al. (2019). T cell antigen discovery via trogocytosis. Nat. Methods 16, 183–190. [PubMed: 30700903]

Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A, Kloetzel PM, Stumpf MPH, Heck AJR, and Mishto M. (2016). A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science 354, 354–358. [PubMed: 27846572]

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17, 10–12.

Newell EW, Sigal N, Nair N, Kidd BA, Greenberg HB, and Davis MM (2013). Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. Nat. Biotechnol. 31, 623–629. [PubMed: 23748502]

Prykhozhij SV, Rajan V, Gaston D, and Berman JN (2015). CRISPR multi-targeter: a web tool to find common and unique CRISPR single guide RNA targets in a set of similar sequences. PLoS ONE 10, e0119372.

Riddell SR, Watanabe KS, Goodrich JM, Li CR, Agha ME, and Greenberg PD (1992). Restoration of viral immunity in immunodeficient humans by the adoptive transfer of T cell clones. Science 257, 238–241. [PubMed: 1352912]

Rius C, Attaf M, Tungatt K, Bianchi V, Legut M, Bovay A, Donia M, Thor Straten P, Peakman M, Svane IM, et al. (2018). Peptide-MHC Class I Tetramers Can Fail To Detect Relevant Functional T Cell Clonotypes and Underestimate Antigen-Reactive T Cell Populations. J. Immunol. 200, 2263–2279. [PubMed: 29483360]

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, and Marsh SGE (2015). The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 43, D423–D431. [PubMed: 25414341]

Rowland-Jones S, Tan R, and McMichael A. (1997). Role of cellular immunity in protection against HIV infection. Adv. Immunol. 65, 277–346. [PubMed: 9238512]

Russell CD, Unger SA, Walton M, and Schwarze J. (2017). The Human Immune Response to Respiratory Syncytial Virus Infection. Clin. Microbiol. Rev. 30, 481–502. [PubMed: 28179378]

Sakahira H, Enari M, and Nagata S. (1998). Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis. Nature 391, 96–99. [PubMed: 9422513]

Sanjana NE, Shalem O, and Zhang F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784. [PubMed: 25075903]

Schnütgen F, Doerflinger N, Calléja C, Wendling O, Chambon P, and Ghyselinck NB (2003). A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. Nat. Biotechnol. 21, 562–565. [PubMed: 12665802]

Schub A, Schuster IG, Hammerschmidt W, and Moosmann A. (2009). CMV-specific TCR-transgenic T cells for immunotherapy. J. Immunol. 183, 6819–6830. [PubMed: 19864595]

Sharma G, and Holt RA (2014). T-cell epitope discovery technologies. Hum. Immunol. 75, 514–519. [PubMed: 24755351]

Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, McAffee D, Birnbaum ME, Yang X, Su LF, Yu W, et al. (2018). Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. Cell 174, 672–687. [PubMed: 30053426]

Sobao Y, Tomiyama H, Sugi K, Tokunaga M, Ueno T, Saito S, Fujiyama S, Morimoto M, Tanaka K, and Takiguchi M. (2002). The role of hepatitis B virus-specific memory CD8 T cells in the control of viral replication. J. Hepatol. 36, 105–115. [PubMed: 11804672]

Stern-Ginossar N, Weisburd B, Michalski A, Le VTK, Hein MY, Huang S-X, Ma M, Shen B, Qian S-B, Hengel H, et al. (2012). Decoding human cytomegalovirus. Science 338, 1088–1093. [PubMed: 23180859]

Stone JD, Chervin AS, and Kranz DM (2009). T-cell receptor binding affinities and kinetics: impact on T-cell activity and specificity. Immunology 126, 165–176. [PubMed: 19125887]

To T-L, Piggott BJ, Makhijani K, Yu D, Jan YN, and Shu X. (2015). Rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis in vivo. Proc. Natl. Acad. Sci. USA 112, 3338–3343. [PubMed: 25733847]

Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, and Peters B. (2018). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 47 (D1), D339–D343.

Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, Tan MP, Dolton G, Clement M, Llewellyn-Lacey S, Price DA, et al. (2012). A single autoimmune T cell receptor recognizes more than a million different peptides. J. Biol. Chem. 287, 1168–1177. [PubMed: 22102287]

Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, Ruxrungtham K, Sanchez J, Brander C, Chung RT, et al. (2015). Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. Science 348, aaa0698.

Xu GJ, Shah AA, Li MZ, Xu Q, Rosen A, Casciola-Rosen L, and Elledge SJ (2016). Systematic autoantigen analysis identifies a distinct subtype of scleroderma with coincident cancer. Proc. Natl. Acad. Sci. USA 113, E7526–E7534. [PubMed: 27821747]

Yang Y. (2015). Cancer immunotherapy: harnessing the immune system to battle cancer. J. Clin. Invest. 125, 3335–3337. [PubMed: 26325031]

Zhang S-Q, Ma K-Y, Schonnesen AA, Zhang M, He C, Sun E, Williams CM, Jia W, and Jiang N. (2018). High-throughput determination of the antigen specificities of T cell receptors in single cells. Nat. Biotechnol. Published online November 12, 2018 10.1038/nbt.4282.

Zhu J, Larman HB, Gao G, Somwar R, Zhang Z, Laserson U, Ciccia A, Pavlova N, Church G, Zhang W, et al. (2013). Protein interaction discovery using parallel analysis of translated ORFs (PLATO). Nat. Biotechnol. 31, 331–334. [PubMed: 23503679]

Highlights

- T-Scan is a high-throughput method to identify the functional targets of CD8 T cells
- We apply T-Scan to multiplex discovery of CMV antigens from bulk memory T cells
- Genome-wide T-Scan finds the cognate antigen and off-targets of a self-reactive TCR

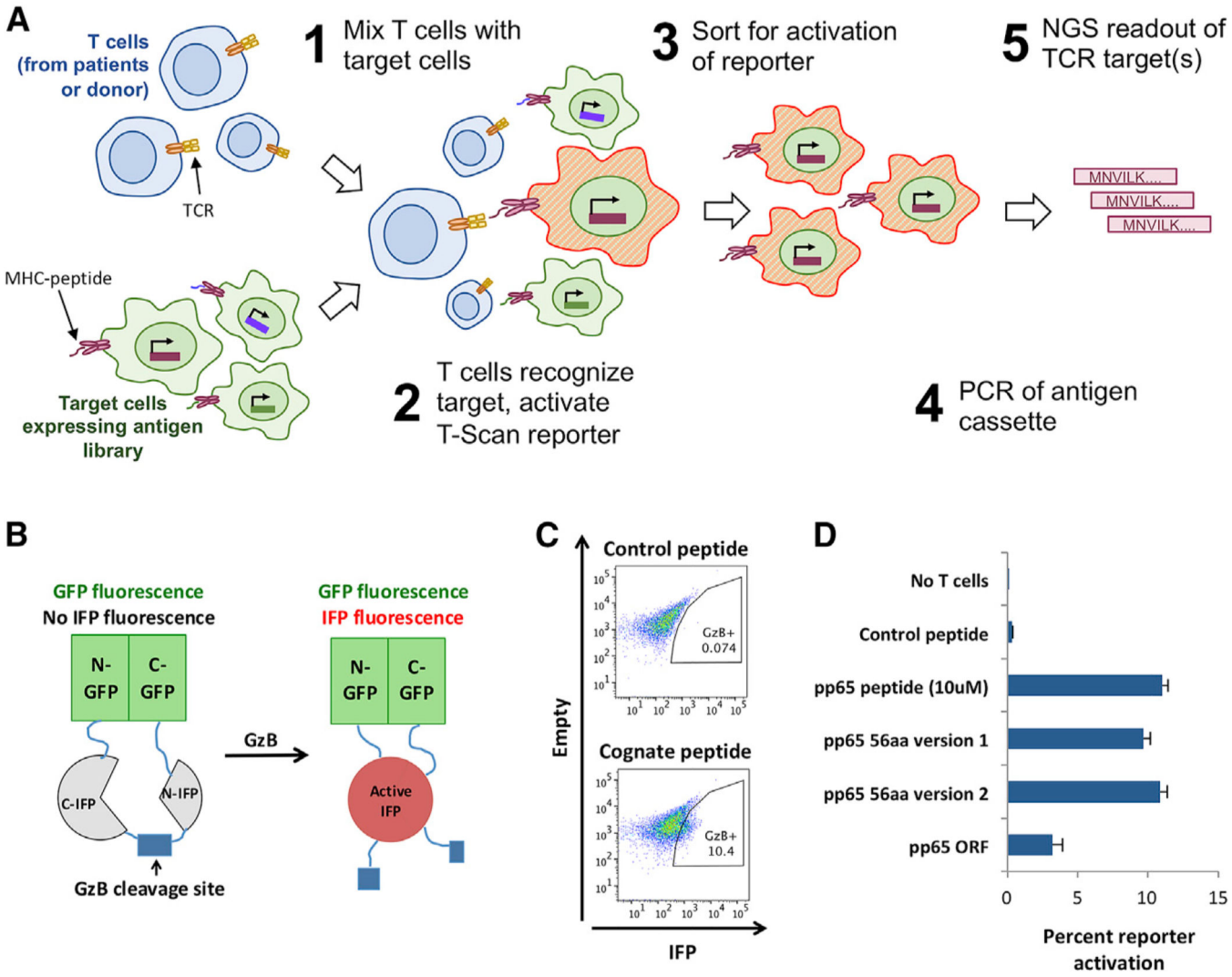


Figure 1. Design of the T-Scan Platform

(A) Schematic of T-Scan workflow. T cells of interest are co-cultured with target cells expressing a library of candidate antigens. Target cells recognized by the T cells are isolated by FACS. The antigens these cells had displayed are identified by PCR and next generation sequencing of the antigen cassette from the isolated cells.

(B) Schematic of the fluorogenic GzB reporter. GzB cleaves a constraining linker sequence between two domains of IFP to enable protein maturation and fluorescence.

(C) Representative flow cytometry plots of IFP fluorescent signals in target cells co-cultured with cytotoxic T cells in the absence (top) or presence (bottom) of cognate antigen.

(D) Efficiency of GzB reporter activation in target cells co-cultured with pp65-specific T cells. Target cells were pulsed with control peptide (HIV IV9), the cognate pp65 peptide, or express one of two 56-aa fragments or the full-length pp65 ORF that contain the cognate pp65 peptide. Error bars indicate the SD across three replicates. See also Figure S1.

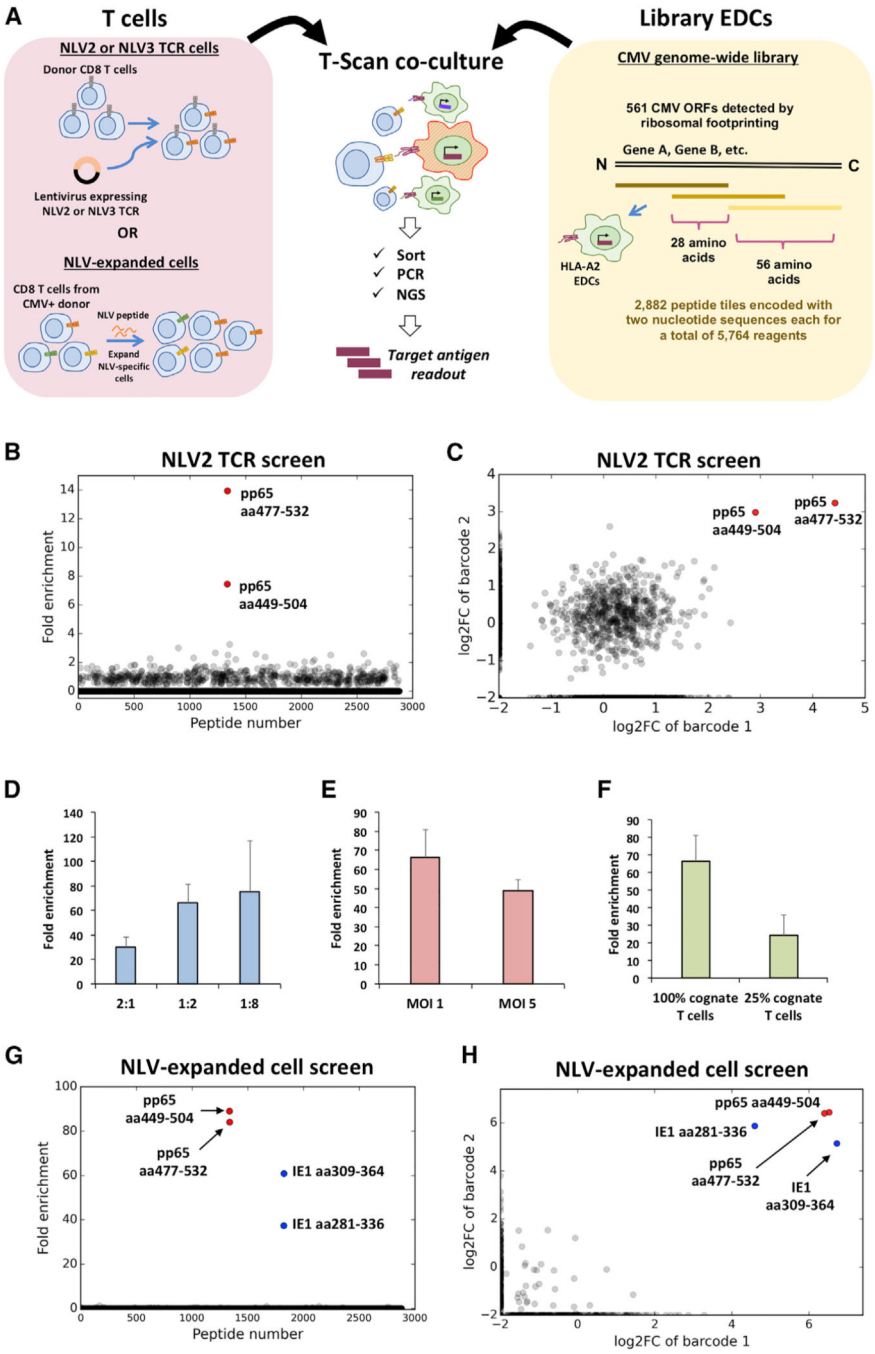


Figure 2. CMV Genome-wide T Cell Antigen Discovery

(A) Overview of CMV screens. Peptides tiling across the CMV genome in 56-aa steps with 28-aa overlap are expressed via lentiviral transduction of target cells expressing HLA-A2, IFP^{GZB}, and $ICAD^{CR}$. For NLV2 TCR T cells, donor CD8 T cells are transduced with lentivirus expressing the NLV2 TCR. For NLV-expanded T cells, CD8 T cells from a CMV-positive donor are expanded in the presence of the NLV peptide.

(B) T-Scan screen of NLV2 TCR cells against CMV genome-wide library. Each dot represents one peptide with the y axis plotting the geometric mean of the enrichment of the

peptide across six total replicates (three screen replicates with two internal barcode replicates each). Fold enrichment is defined as the ratio of the abundance of the peptide in the sorted population relative to the input library. Peptides highlighted in red contain the known cognate antigen of the NLV2 TCR. Data presented in Table S1.

(C) Reproducibility of internal replicates of NLV2 TCR T cell screen. Each dot represents one peptide, with the X and Y axes plotting the geometric mean of the fold-change of each barcode across three replicates.

(D–F) Performance of T-Scan assays in various conditions. Bars show the average fold-enrichment of the four NLV-containing peptides in screens performed with NLV3 TCR T cells with: (D) a 2:1 (4× T cells), 1:2 (standard), and 1:8 (0.25× T cells) ratio of NLV3 TCR T cells. (E) antigens introduced into target cells at an MOI of 1 or 5, and (F) a 1:2 ratio of NLV3 TCR T cells (standard), a 1:8 ratio of NLV3 TCR T cells (0.25× T cells), and a 1:8 ratio of NLV3 TCR T cells mixed with a 3-fold excess of non-specific T cells (0.25× T cells in 1:4 mix). Data for (D)–(F) are shown in Table S2. Error bars for (D)–(F) indicate SD across four target peptides.

(G) T-Scan screen of NLV-expanded cells against CMV genome-wide library, plotted as in (B). Peptides highlighted in red contain the NLV peptide and the blue peptides highlight additional enriching peptides. Data presented in Table S3.

(H) Reproducibility of internal replicates of NLV-expanded T cell screen, plotted as in (C). See also Figures S2 and S3.

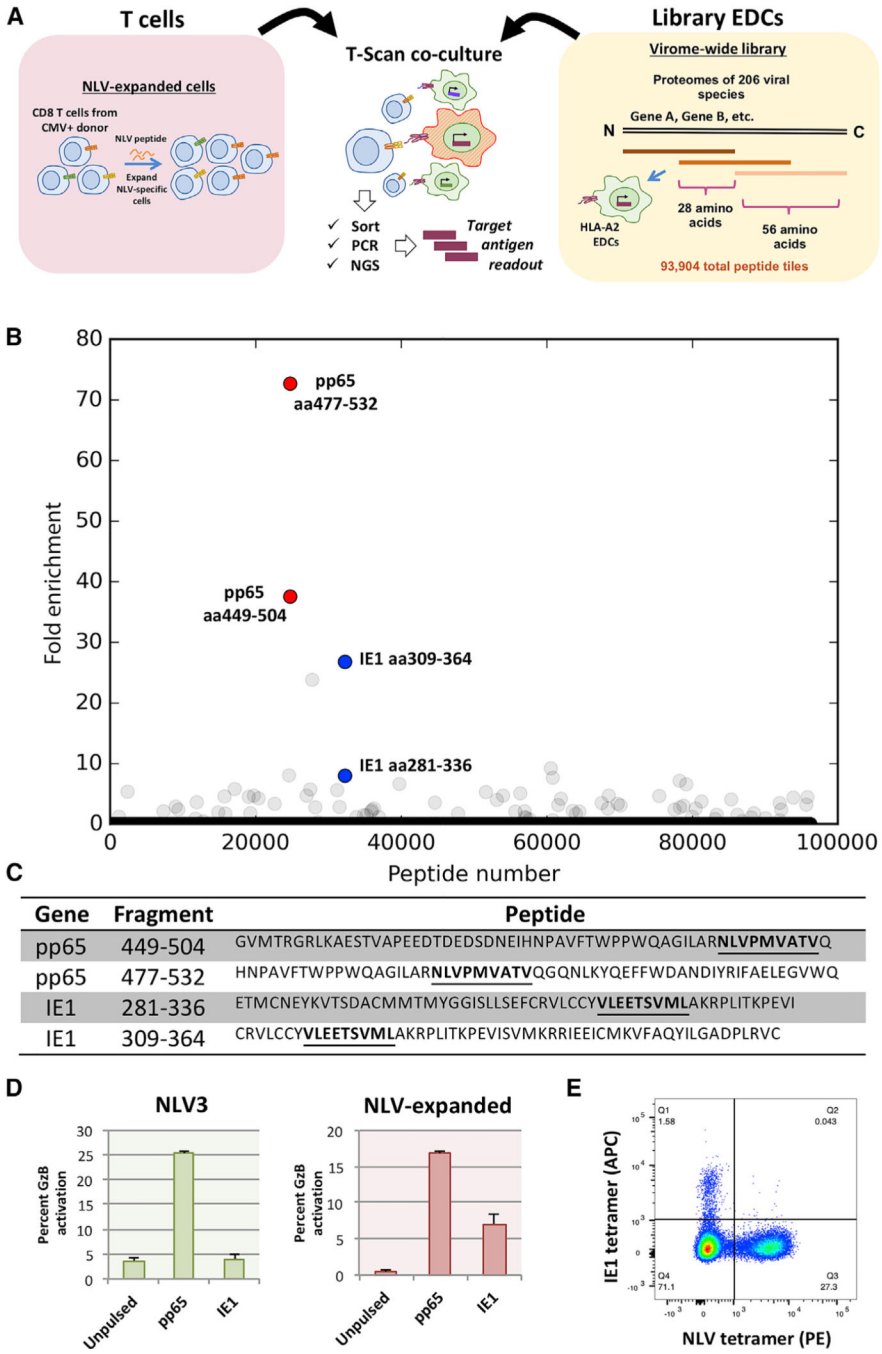


Figure 3. Virome-wide T Cell Antigen Discovery

(A) Schematic of virome-wide T-Scan screen. Peptides tiling across the proteomes of 206 viral species in 56-aa steps with 28-aa overlap are expressed via lentiviral transduction of target cells expressing HLA-A2, IFP^{GZB}, and ICAD^{CR}.

(B) T-Scan screen of NLV-expanded T cells against CMV genome-wide library. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold-change of each peptide across four replicates. Peptides highlighted in red contain the NLV peptide and the blue peptides highlight additional enriching peptides. Data presented in Table S4.

(C) Table of enriching peptides highlighted in (B) and Figure 2E. Predicted HLA-A2 binding peptides, including the known NLV epitope are bolded.
(D) GzB reporter activation in target cells pulsed with the pp65 peptide (NLVPMVATV) or IE1 peptide (VLEETSVML) in the presence of NLV3 TCR T cells (top panel) or NLV-expanded T cells (bottom panel). Error bars indicate SD across three replicates.
(E) Tetramer staining of the NLV-expanded T cells with the pp65 peptide and the IE1 peptide. Gated $CD8^+$ cells are shown.

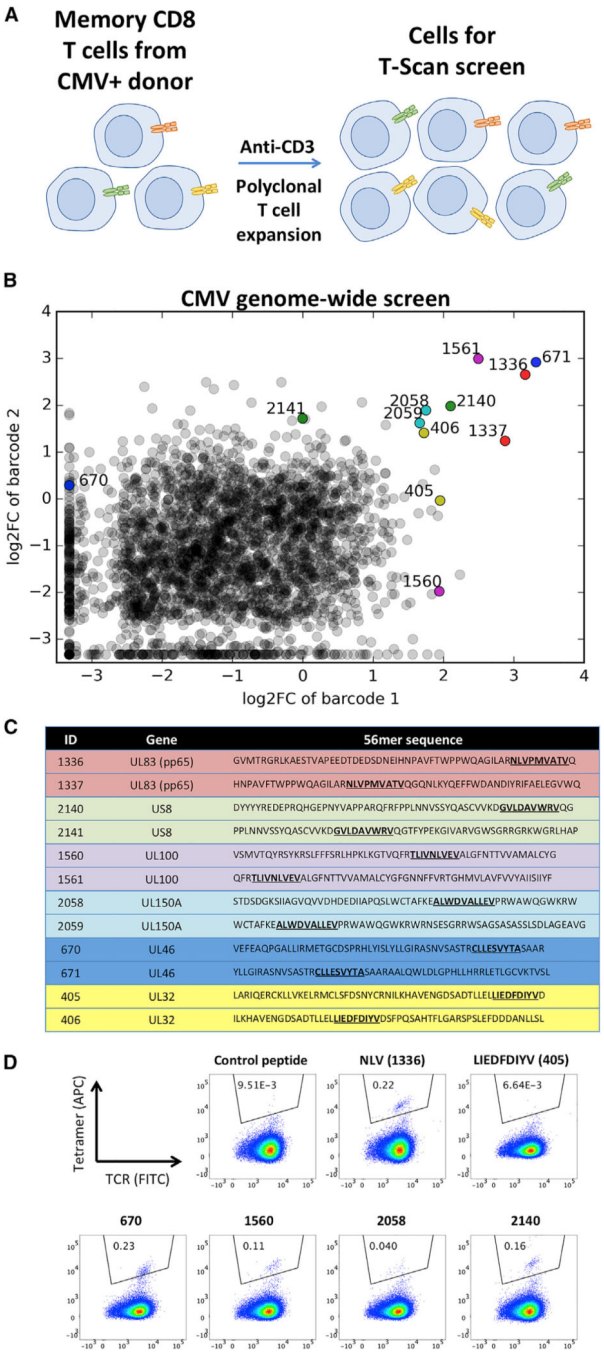


Figure 4. CMV Genome-wide Screen with Bulk Memory CD8 T Cells

(A) Generation of memory CD8 T cells. Bulk memory CD8 T cells are purified from donor blood and undergo polyclonal expansion in the presence of feeder cells and anti-CD3.

(B) T-Scan screen of bulk memory CD8 cells against CMV genome-wide library. Each dot represents one peptide, with the X and Y axes plotting the performance of two unique barcodes for the peptide. The X and Y values indicate the modified geometric mean (see STAR Methods) of the fold-change of each barcode across four replicates. Colored dots highlight pairs of overlapping peptides. Data presented in Table S5.

(C) List of enriching peptides highlighted in (B). Predicted high-affinity HLA-A2 binding epitopes in the peptide overlap regions are bolded and underlined.
(D) Tetramer staining of the memory CD8 T cells used in the screen with the predicted peptides bolded in (C).

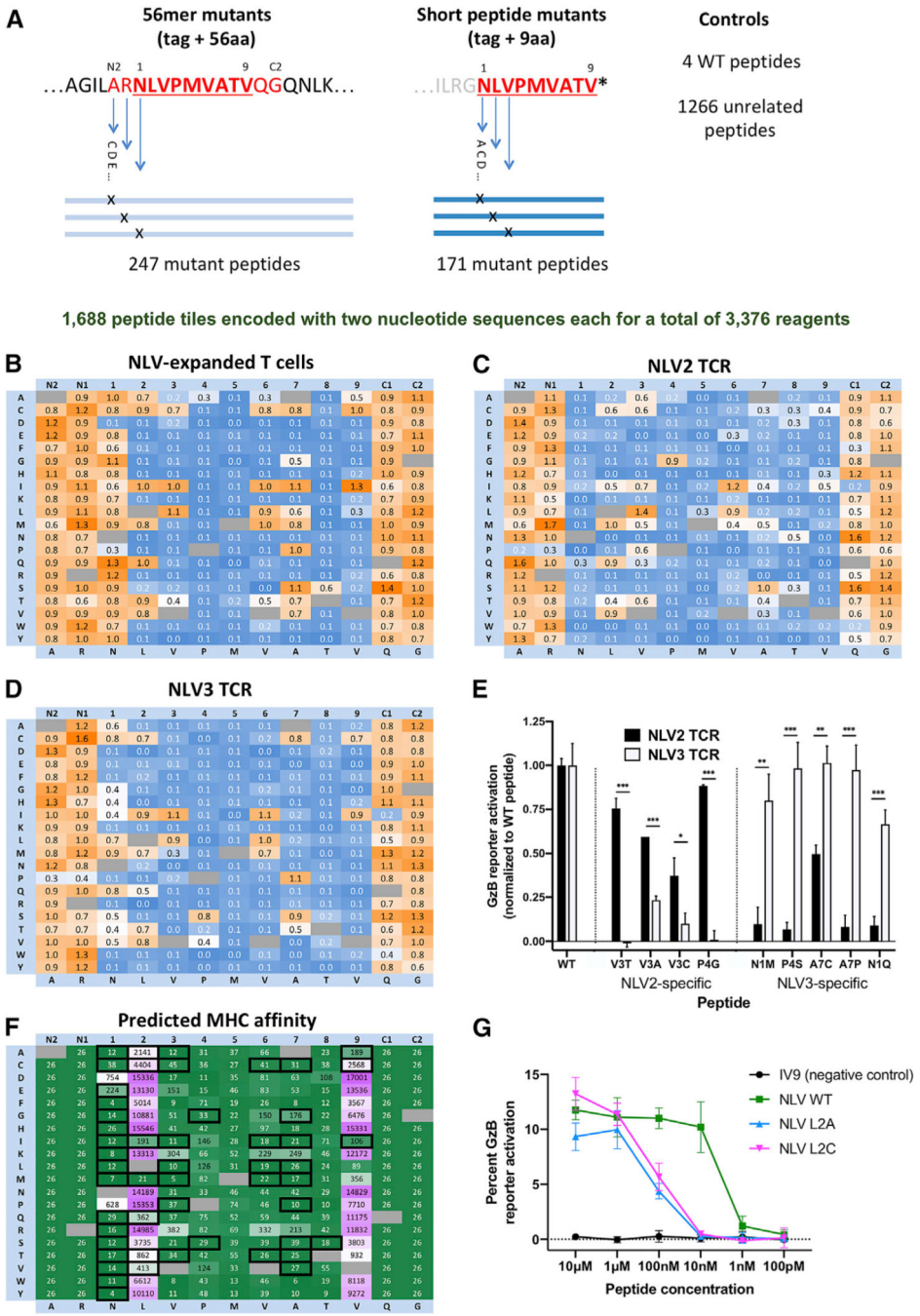


Figure 5. Comprehensive Mutagenesis Analysis of NLV-Specific T Cells

(A) Design of the NLV epitope comprehensive mutagenesis library. Every position in the NLV epitope (NLVPMVATV) is mutated to each of the 19 alternative amino acids. The mutant epitopes are expressed in the context of a 56-aa fragment with an N-terminal tag or as short peptides with an N-terminal tag. The two amino acids directly N-terminal and C-terminal to the epitope are also mutated in the 56-aa versions.

(B–D) Recognition of mutant peptides in the context of 56-aa fragments by NLV-specific T cells. Each box in the heatmap represents one mutant, where the amino acid along the x axis

is mutated to the amino acid indicated along the y axis. The value in the heatmap represents the enrichment of this mutant compared to the WT NLV epitope. Heatmaps are plotted for (B) NLV-expanded T cells, (C) T cells expressing the NLV2 TCR, and (D) T cells expressing the NLV3 TCR. Data for (B)–(D) are presented in Table S6.

(E) Relative activation of the GzB reporter in cells pulsed with identified NLV2 TCR-specific and NLV3 TCR-specific mutant peptides and co-cultured with NLV2 or NLV3 TCR T cells. Error bars indicate SD across three replicates. p values were determined by a two-tailed t test and are shown with asterisks, *<0.05, **<0.01, ***<0.001.

(F) Predicted MHC binding affinity of each mutant. The mutants are displayed as in (B)–(D) and the values plotted are nM affinities based on the NetMHC algorithm. Boxes are drawn around all mutants that enriched to at least 50% of WT levels in one or more of the T cell experiments from (B)–(D).

(G) GzB reporter activation following NLV3 T cell co-culture with target cells that were pulsed with dilutions of the WT NLV epitope or two predicted low-affinity mutants of the NLV epitope. Error bars indicate SD across three replicates.

See also Figures S4 and S5.

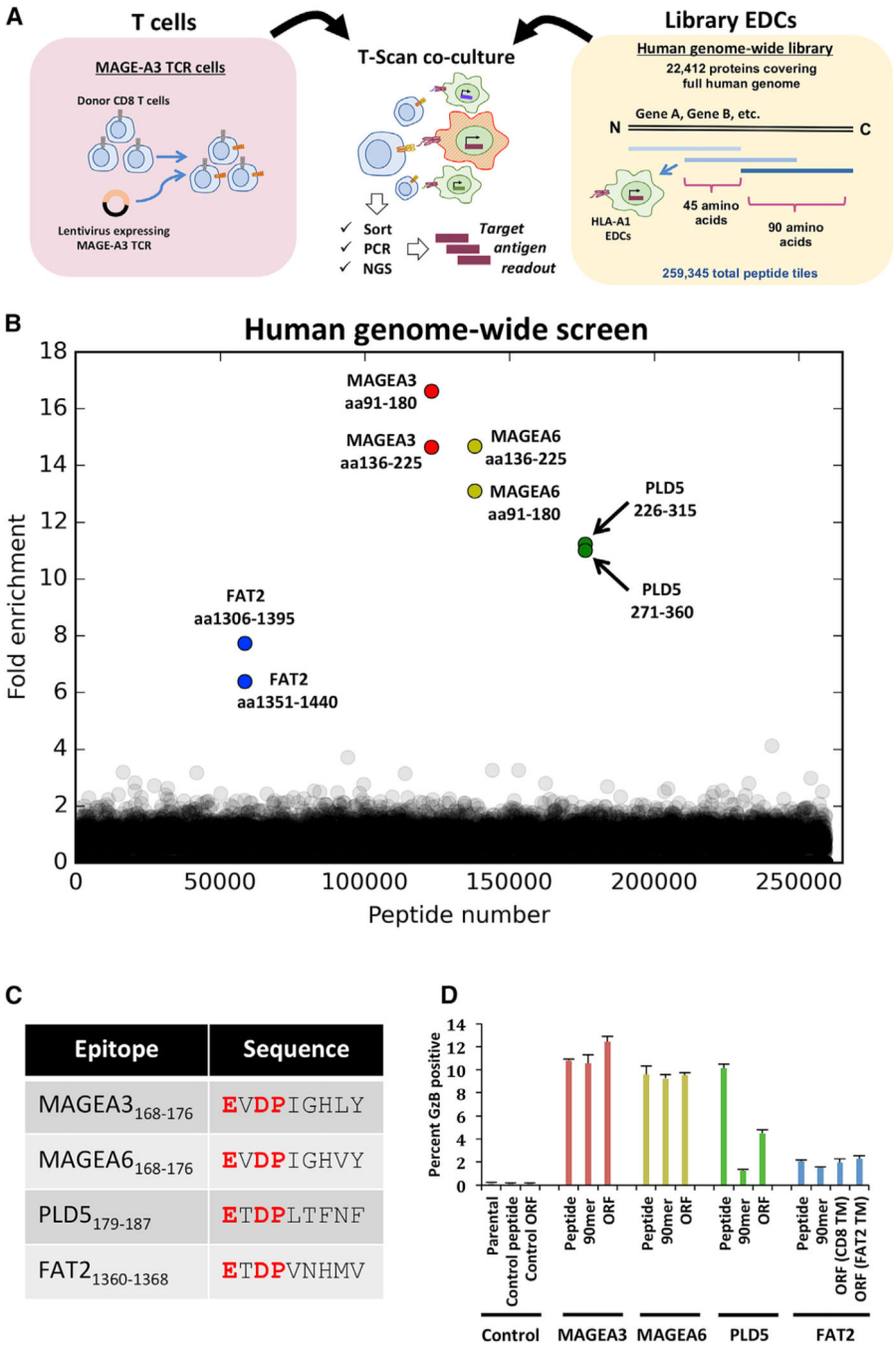


Figure 6. Human Genome-wide T Cell Antigen Discovery

(A) Schematic of human genome-wide T-Scan screen. Peptides tiling across the human proteome in 90-aa steps with 45-aa overlap are expressed via lentiviral transduction of target cells expressing HLA-A1, IFP^{GZB}, and ICAD^{CR}. MAGE-A3 TCR T cells are generated by lentiviral transduction of donor T cells with a construct encoding the MAGE-A3 TCR.

(B) T-Scan screen of MAGE-A3 TCR expressing T cells against human genome-wide library. Each dot represents one peptide, with the y axis plotting the geometric mean of the fold-change of each peptide across eight replicates. Data presented in Table S7.

(C) List of the predicted HLA-A1 binding epitopes from the antigens identified in (B) from MAGE-A3 (NM_005362.4), MAGE-A6 (NM_005363.3), PLD5 (NM_152666.1), and FAT2 (NM_001447.2).

(D) Activation of GzB reporter by MAGE-A3 TCR T cells in the presence of the candidate antigens. Untransduced EDCs (parental), EDCs pulsed with the HIV IV9 peptide (peptide), and EDCs transduced with the GFP ORF (ORF) were used as controls. Peptides were added at 1 μM. Fragments of 90 amino acids and full-length ORFs containing each antigen were stably expressed by lentiviral transduction. The FAT2 synthetic ORF fragments include the region surrounding the FAT2 epitope fused to the CD8 or FAT2 signal peptide and transmembrane domain. Error bars indicate SD across three replicates.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD3 antibody (OKT3)	Thermo Fisher Scientific	Cat: 14-0037-82 RRID: AB_467057
anti-human TCR α/β Antibody	Biolegend	Cat: 306711 RRID: AB_528966
Alexa Fluor 647 anti-human CD8 Antibody	Biolegend	Cat: 344725 RRID: AB_2563451
APC anti-human HLA-A,B,C Antibody	Biolegend	Cat: 311410 RRID: AB_314879
Biological Samples		
PBMC 50M (Donor #224)	Astarte Biologics	Cat: 1001-3575MY17
Chemicals, Peptides, and Recombinant Proteins		
Interleukin-2, human (hIL-2)	Sigma (Millipore)	Cat: 11147528001
Critical Commercial Assays		
GeneJET Genomic DNA Purification Kit	Thermo Fisher Scientific	Cat: K0722
QuickSwitchTM Quant HLA-A*02:01 Tetramer Kit-PE	MBL International Corporation	Cat: TB-7301-K1
QuickSwitchTM Quant HLA-A*02:01 Tetramer Kit-APC	MBL International Corporation	Cat: TB-7300-K2
Experimental Models: Cell Lines		
HEK293T cells	ATCC	ATCC Cat# CRL-3216 RRID:CVCL_0063
HEK293T HLA-A1 Epitope Discovery Cells (MHC KO, IFPGzB, ICADCR, HLA-A1)	This paper	N/A
HEK293T HLA-A2 Epitope Discovery Cells (MHC KO, IFPGzB, ICADCR, HLA-A2)	This paper	N/A
Oligonucleotides		
See Table S8	N/A	N/A
Recombinant DNA		
CMV sublibrary	This paper	N/A
Virame library	Xu et al., 2015	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Tiling mutagenesis library	This paper	N/A
Human peptidome library	Xu et al., 2016	N/A
pHAGE_EF1a_ICADCR_bsd	This paper	N/A
pHAGE_EF1a_IFPGzB_hyg	This paper	N/A
pHAGE_CMV_NFlagHA_DEST_IRES_puro	This paper	N/A
pHAGE_EF1a_HLA-A1	This paper	N/A
pHAGE_EF1a_HLA-A2	This paper	N/A
lentiCRISPR V2 puro vector	Sanjana et al., 2014	Addgene Plasmid #52961
pHAGE_EF1a_DEST_puro	This paper	N/A
pHAGE_EF1a_DEST_Zsg	This paper	N/A
Software and Algorithms		
Bowtie	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
Cutadapt	Martin, 2011	http://cutadapt.readthedocs.io/en/stable/