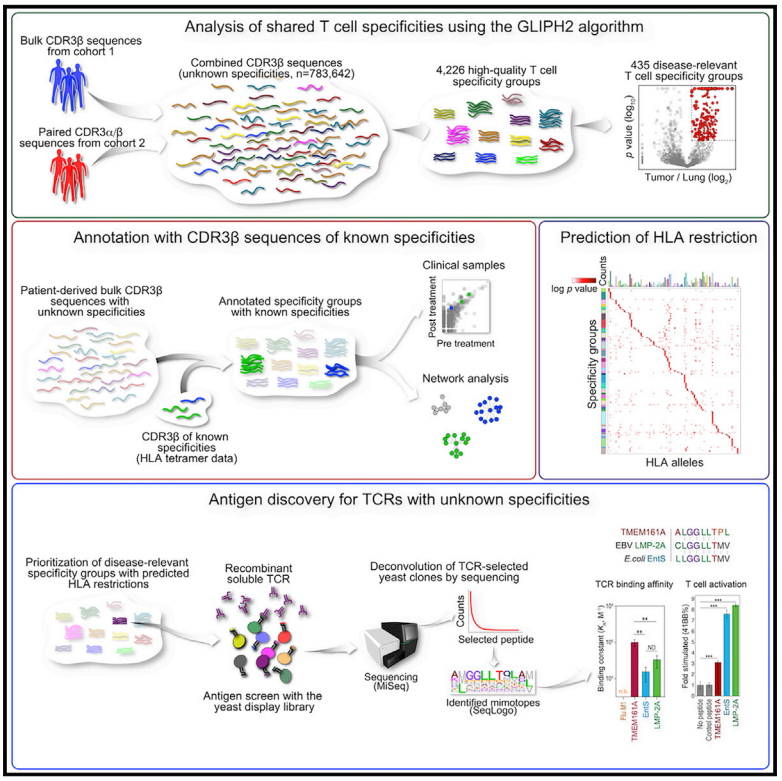
Resource

Immunity

Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery

Graphical abstract



Highlights

- The algorithm GLIPH2 enables analysis of shared TCR specificity and HLA prediction
- Tumor-infiltrating T cells cross-react to EBV antigens and shared tumor antigens
- EBV-specific T cells expanded in patients responding to immune checkpoint blockade
- Cross-reactive CD8 T cells express GZMK

Authors

Shin-Heng Chiou, Diane Tseng, Alexandre Reuben, ..., K. Christopher Garcia, Crystal L. Mackall, Mark M. Davis

Correspondence

mmdavis@stanford.edu

In brief

Chiou, Tseng, et al. analyze TCRβ chain sequences from 178 non-small cell lung cancer patients and identify shared specificity groups, which in turn enable antigen identification. One such antigenic epitope—a peptide from an epithelial protein—is cross-reactive to epitopes from Epstein-Barr virus and *E. coli*, suggesting that cross-reactivity may underlie the presence of pathogen-specific T cells in tumor infiltrates.

Chiou et al., 2021, Immunity *54*, 586–602
March 9, 2021
https://doi.org/10.1016/j.immuni.2021.02.014

CellPress

CellPress

Immunity

Resource

Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery

Shin-Heng Chiou,[1,20,22] Diane Tseng,[2,20,23] Alexandre Reuben,[3] Vamsee Mallajosyula,[1] Irene S. Molina,[1,20] Stephanie Conley,[4] Julie Wilhelmy,[5] Alana M. McSween,[1] Xinbo Yang,[6] Daisuke Nishimiya,[6] Rahul Sinha,[4] Barzin Y. Nabet,[7] Chunlin Wang,[1] Joseph B. Shrager,[8,9] Mark F. Berry,[8] Leah Backhus,[8,9] Natalie S. Lui,[8,9] Heather A. Wakelee,[2,9] Joel W. Neal,[2,9] Sukhmani K. Padda,[2] Gerald J. Berry,[10] Alberto Delaidelli,[11] Poul H. Sorensen,[11] Elena Sotillo,[12] Patrick Tran,[12] Jalen A. Benson,[8] Rebecca Richards,[12,13] Louai Labanieh,[12,14] Dorota D. Klysz,[12] David M. Louis,[1] Steven A. Feldman,[12] Maximilian Diehn,[4,7,9] Irving L. Weissman,[4] Jianjun Zhang,[3,16] Ignacio I. Wistuba,[17] P. Andrew Futreal,[16] John V. Heymach,[3] K. Christopher Garcia,[6,18] Crystal L. Mackall,[12,13,15,21] and Mark M. Davis[1,18,19,21,24,*]

[1]Institute for Immunity, Transplantation and Infection, Stanford University, Stanford, CA 94305, USA
[2]Department of Medicine, Division of Oncology, Stanford University, Stanford, CA 94305, USA
[3]Department of Thoracic Head and Neck Medical Oncology, Division of Cancer Medicine, MD Anderson Cancer Center, Houston, TX 77030, USA
[4]Institute for Stem Cell Biology and Regenerative Medicine Institute, Stanford University, Stanford, CA 94305, USA
[5]Stanford Genome Technology Center, Stanford University, Stanford, CA 94305, USA
[6]Department of Molecular and Cellular Physiology and Structural Biology, Stanford University, Stanford, CA 94305, USA
[7]Department of Radiation Oncology, Stanford University, Stanford, CA 94305, USA
[8]Department of Cardiothoracic Surgery – Thoracic Surgery, Stanford University, Stanford, CA 94305, USA
[9]Stanford Cancer Institute, Stanford, CA 94305, USA
[10]Department of Pathology, Stanford University, Stanford, CA 94305, USA
[11]Department of Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, BC V5Z 1L3, Canada
[12]Center for Cancer Cell Therapy, Stanford Cancer Institute, Stanford University, Stanford, CA 94305, USA
[13]Department of Pediatrics, Stanford University, Stanford, CA 94305, USA
[14]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[15]Department of Medicine, Stanford University, Stanford, CA 94305, USA
[16]Department of Genomic Medicine, Division of Cancer Medicine, MD Anderson Cancer Center, Houston, TX 77030, USA
[17]Department of Translational Molecular Pathology, Division of Pathology and Laboratory Medicine, MD Anderson Cancer Center, Houston, TX 77030, USA
[18]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA
[19]Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305, USA
[20]These authors contributed equally
[21]These authors contributed equally
[22]Present address: Rutgers Cancer Institute of New Jersey, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA
[23]Present address: Department of Medicine, Division of Medical Oncology, University of Washington, Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[24]Lead contact
*Correspondence: mmdavis@stanford.edu
https://doi.org/10.1016/j.immuni.2021.02.014

SUMMARY

To identify disease-relevant T cell receptors (TCRs) with shared antigen specificity, we analyzed 778,938 TCRβ chain sequences from 178 non-small cell lung cancer patients using the GLIPH2 (grouping of lymphocyte interactions with paratope hotspots 2) algorithm. We identified over 66,000 shared specificity groups, of which 435 were clonally expanded and enriched in tumors compared to adjacent lung. The antigenic epitopes of one such tumor-enriched specificity group were identified using a yeast peptide-HLA A*02:01 display library. These included a peptide from the epithelial protein TMEM161A, which is overexpressed in tumors and cross-reactive epitopes from Epstein-Barr virus and *E. coli*. Our findings suggest that this cross-reactivity may underlie the presence of virus-specific T cells in tumor infiltrates and that pathogen cross-reactivity may be a feature of multiple cancers. The approach and analytical pipelines generated in this work, as well as the specificity groups defined here, present a resource for understanding the T cell response in cancer.

INTRODUCTION

Despite the widespread use of immunotherapies for treating cancer, our understanding of T cell specificities in this disease is very limited (Sharma and Allison, 2020). Antigen specificity is the key determinant of T cell function, but challenges posed by T cell receptor (TCR) diversity and human leukocyte antigens (HLAs) allele polymorphism have been major obstacles to understanding the full scope of antigens recognized by tumor-infiltrating T cells (Arstila et al., 1999; Robins et al., 2010). Tumor-infiltrating T cells that recognize mutated proteins (i.e., neoantigens), non-mutated tumor-associated antigens (TAAs), and viral antigens have been described (Coulie et al., 1994; 1995; Kawakami et al., 1994; Koziel et al., 1995; Murray et al., 1992; Rehermann et al., 1995; Savage et al., 2008; van der Bruggen et al., 1991; Wölfel et al., 1995). In tumors with no known viral etiology, prior reports have identified virus-specific T cells infiltrating tumors, including those that recognize influenza (flu), Epstein-Barr virus (EBV), or cytomegalovirus (CMV) (Andersen et al., 2012; Rosato et al., 2019; Scheper et al., 2019; Simoni et al., 2018). In these tumors, virus-specific tumor-infiltrating T cells are presumed to not recognize tumor antigens and are often referred to as "bystander cells" (Scheper et al., 2019; Simoni et al., 2018).

With respect to the search for TAAs, next-generation sequencing has enabled rapid sequencing of large numbers of TCR variable regions in tumor-infiltrating T cells, but challenges remain in making use of the data generated. This is in part due to hundreds or thousands of distinct TCR sequences that can recognize the same peptide-major histocompatibility complex (MHC) ligand (Song et al., 2017). To reduce this immense sequence diversity to a much smaller number of specificities, we developed an algorithm, GLIPH (grouping of lymphocyte interactions by paratope hotspots; Glanville et al., 2017), and an improved version (GLIPH2; Huang et al., 2020), that parses large numbers of TCR sequences into shared specificity groups that are highly likely to recognize the same peptide-MHC ligands. These shared specificity groups are established based on identical amino acid sequence motifs or strong homologies within the complementarity-determining region 3 (CDR3) of the TCRβ chain.

Here, we used GLIPH2 to identify over 66,000 high-quality, shared specificity groups from 778,938 CDR3β sequences found in 178 non-small cell lung cancer (NSCLC) patients with surgically resectable tumors (Reuben et al., 2020). Four hundred thirty five shared specificity groups were clonally expanded in the tumor compared to the adjacent lung tissue. Among those, CDR3β sequences containing a "S%DGMNTE" sequence motif were prioritized for antigen discovery using HLA-A*02 yeast display library, where "%" denotes the amino acid that varied (Gee et al., 2018). T cells with the "S%DGMNTE CDR3β" motif responded to the non-mutated tumor antigen TMEM161A, as well as antigens from EBV and *E. coli*, demonstrating T cell cross-reactivity to TAAs and common pathogens. Furthermore, we uncovered a second example of cross-reactivity between an endogenous antigen and an EBV epitope and two other cases where EBV-specific CDR3β sequences were clonally expanded in patients who had clinically significant responses to anti-PD-1 treatment. This suggests that pathogen cross-reactivity may be an important feature in the interaction between neoplasia and T cell immunity. Overall, the approach presented here enables the comprehensive analyses of shared T cell specificities in human cancer and the identification of specific antigens using a yeast display library, with broader application to other cancer types.

RESULTS

Defining shared specificity groups for tumor-infiltrating T cells in human lung cancer

As described previously, GLIPH2 identifies CDR3β sequences that are highly likely to have shared peptide-MHC specificities based on local motifs and/or global homology (Glanville et al., 2017; Huang et al., 2020). To identify T cells recognizing shared tumor antigens in lung cancer, we applied GLIPH2 to a recently published MD Anderson Cancer Center (MDACC) dataset of 778,938 distinct CDR3β sequences from NSCLC tumors and from adjacent lungs. This clinical cohort represents 178 patients with surgically resectable disease and with available HLA data (Table S1) (Reuben et al., 2020). We first defined shared specificity groups with a set of specific filtering criteria and identified 66,094 shared specificity groups (Figure 1A; Table S2). To focus on the most disease-relevant TCRs, we further identified 4,226 specificity groups with evidence of clonal expansion, and of these, 435 were enriched in tumor compared to adjacent lung (Figures 1A and S1A; Table S3). Thus, the CDR3β members of these 435 tumor-enriched specificity groups are inferred to recognize yet undiscovered TAAs.

Next, we reasoned that T cells recognizing shared tumor antigens would undergo clonal expansion in NSCLC patients but not in individuals without cancer. We observed a significantly higher percentage of the expanded CDR3β clones in the MDACC NSCLC cohort (Figure 1B) belonging to the 435 tumor-enriched specificity groups compared to the remainder of less expanded TCRs. We made a similar observation in a validation cohort of 1,173,806 CDR3β sequences from 202 tumor samples representing 68 NSCLC patients (TRACERx; Joshi et al., 2019; Figure 1B). In contrast, adjacent lungs of cancer patients (not involved by tumor) (Figure S1B), lungs from healthy donors, or lungs from chronic obstructive pulmonary disease (COPD) patients (without cancer diagnoses) (Reuben et al., 2020) had fewer CDR3β clones that belonged to tumor-enriched specificity groups. (Figure 1B). Together, these data demonstrate that GLIPH2 successfully parsed a large dataset of CDR3β sequences into a few hundred tumor-enriched specificity groups with disease relevance to NSCLC.

Viral specificity group inferences from HLA tetramer datasets

To validate the shared specificity groups established by GLIPH2, we included CDR3β sequences from publicly available HLA tetramer databases in combination with the MDACC CDR3β sequences for a joint GLIPH2 analysis (Glanville et al., 2017; Shugay et al., 2018; Song et al., 2017). The publicly available tetramer CDR3β sequences primarily cover viral specificities and were experimentally shown to bind epitopes in the context of their respective HLAs. This allowed us to annotate some specificity groups with CDR3β sequences linked to unique epitopes in the context of their HLA and therefore infer the shared specificity

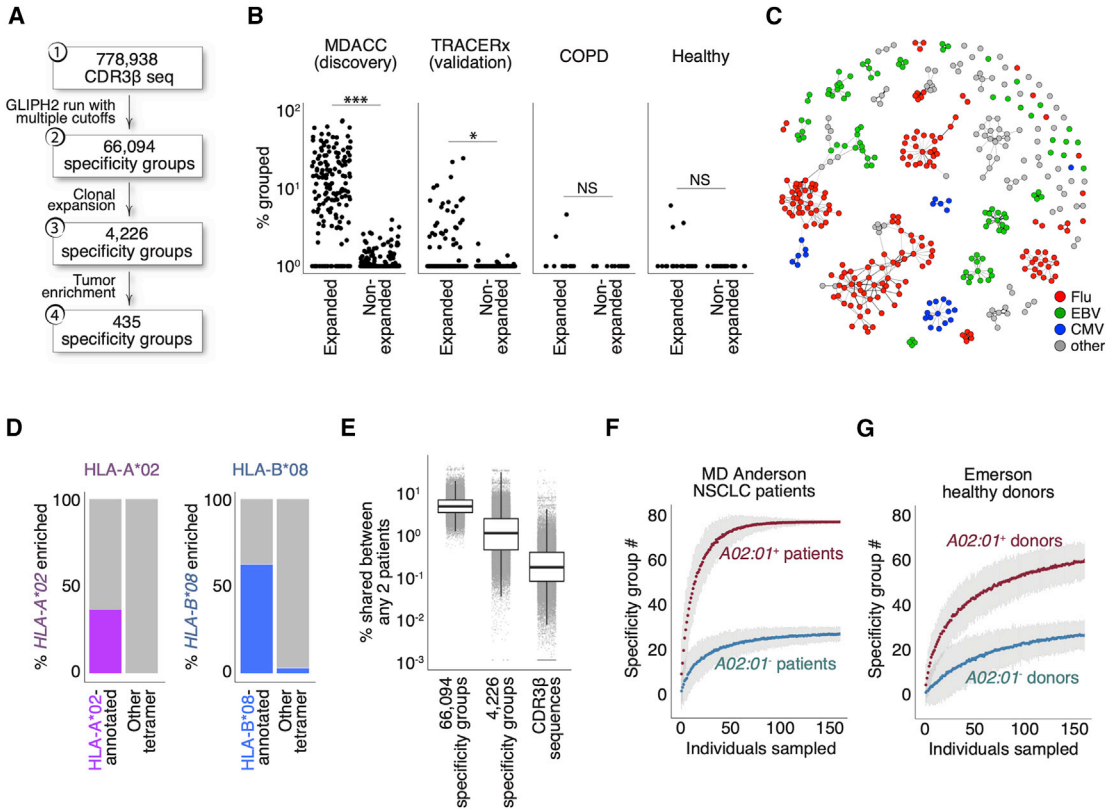


Figure 1. Establishing specificity groups with CDR3β sequences from lung cancer patients

(A) Analysis of shared T cell specificities with the GLIPH2 algorithm. Step 1: 778,938 CDR3β sequences from the MDACC cohort as input for GLIPH2 analysis. Step 2: establish 66,094 specificity groups with multiple criteria (Figure S1A). Step 3: establish 4,226 clonally expanded specificity groups. Step 4: establish 435 clonally expanded, tumor-enriched specificity groups.

(B) Clinical relevance of tumor-enriched specificity groups in lung cancer. The most clonally expanded CDR3β sequences from tumors belonged to the 435 tumor-enriched specificity groups, whereas those from lung tissues of healthy donors and COPD patients did not. The trend was validated with tumors from a second NSCLC cohort (the TRACERx consortium, n = 202, validation). ***$p < 0.001$; *$p < 0.05$ by paired t test. NS, not significantly different.

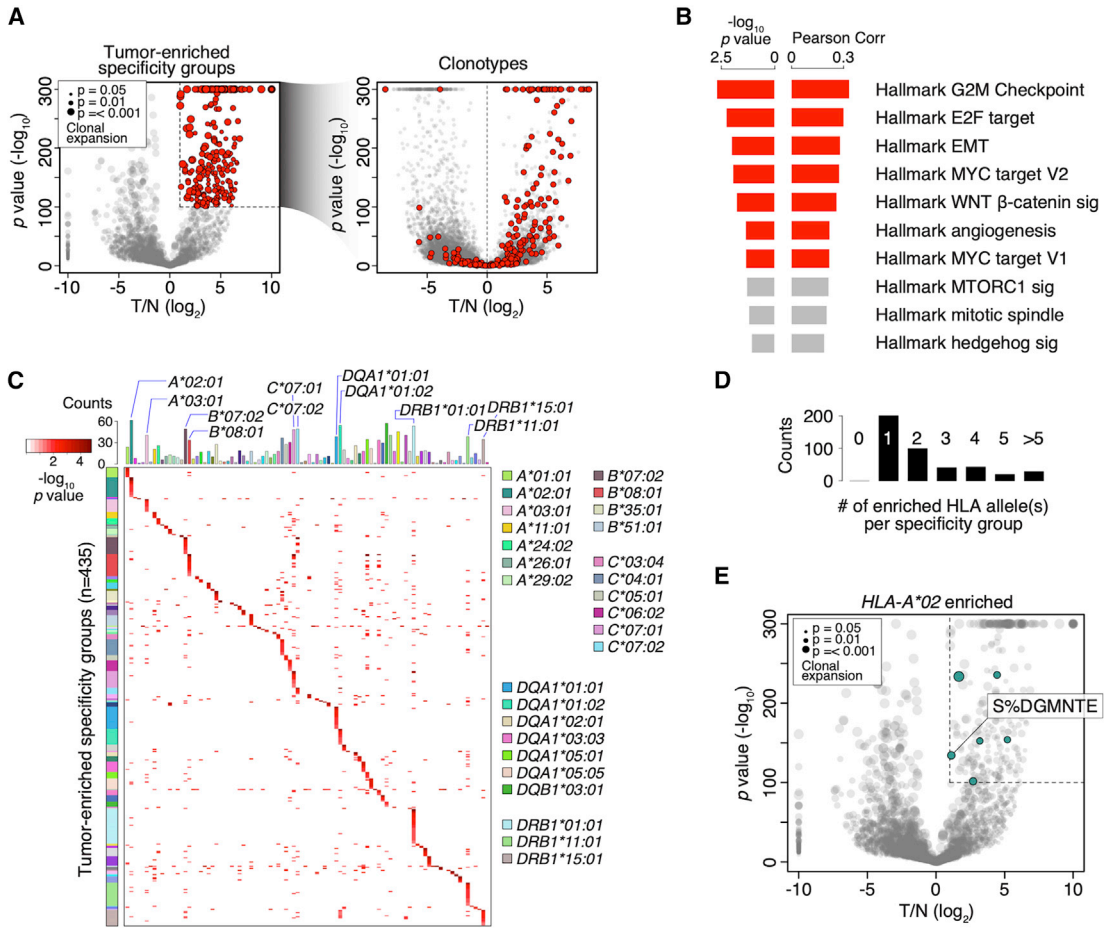
(C) Network analysis of 396 specificity groups annotated with CDR3β sequences from HLA tetramers with flu (red), EBV (green), and CMV (blue) antigens. Each dot is a specificity group, edges indicate the presence of identical CDR3β sequence(s) shared across two specificity groups.

(D) Percentage (%) of HLA-A*02 or HLA-B*08 tetramer-annotated specificity groups with significantly enriched the *A*02* (purple, left plot) or *B*08* (blue, right plot) supertype alleles, respectively. Specificity groups annotated with tetramers of other HLA alleles (other tetramer) were included for comparisons.

(E) Percentage of shared specificity between any two given MDACC NSCLC patients (% shared between any 2 patients, total n = 178) based on CDR3β membership in total specificity groups regardless of clonal expansion (n = 66,094), membership in clonally expanded specificity groups (n = 4,226), or comparison of identical CDR3β sequences. Boxes represent medians with the first (25^{th}) and third (75^{th}) quartiles.

(F and G) Bootstrapping of specificity group numbers (y axis, specificity group #) with varying sampling sizes (individuals sampled) for either HLA-A*02$^+$ or HLA-A*02$^-$ NSCLC patients (F) or healthy donors (G, Emerson study). Data represent means with 3× standard errors from repeated sampling.

of the remaining CDR3β members. The joint analysis annotated 394 of the 66,094 shared specificity groups (Figures 1A and 1C). Of these specificity groups, 71 were clonally expanded and annotated with 10 distinct tetramers (Figure S1C). We found that CDR3β sequences with inferred specificities to flu-, EBV-, or CMV-derived antigens collectively did not show biases in the tumor compared to the adjacent lung (data not shown). Furthermore, the estimated frequencies of these viral-specific CDR3β clones were well above the naive level (one in every 10^5–10^6) and on par with the previously reported ranges measured by HLA tetramer staining (data not shown) (Andersen et al., 2012; Rosato et al., 2019; Simoni et al., 2018). Thirteen of the 27 expanded flu M1-annotated specificity groups carry either the "RS" or "GxY" motifs known to be critical for the engagement with the flu-$M1_{58–66}$ peptide/HLA-A*02 (Figure S1D) (Song et al., 2017). Network analysis organized these tetramer-annotated specificity groups with identical CDR3β sequence members into communities (Figures 1C and S1C). Specificity groups belonging to a given community were consistently annotated with identical HLA tetramers (Figures 1C, S1C, and S1D), indicating that some antigen specificity groups, albeit sharing distinct sequence motifs, are exhibiting the same specificity and HLA restriction. Among the 394 shared specificity groups annotated with tetramers, 588 out of 634 identical CDR3β sequence members (93%) connected specificity groups annotated with the same tetramer (Figures S1E and S1F). Among the 71 clonally expanded specificity groups annotated with tetramers, 92 out of 92 identical CDR3β sequence members (100%) connected groups annotated with the same tetramer (Figures S1C and S1G). This result indicates that while CDR3β sequences



F

TCRβ	Vβ	Patient counts	Counts of *HLA-A*02*+ cases/total	Average clonal freq (%) Adjacent lung%	Tumor%	Blood%
CAGSqDGMNTEAFF	*TRBV28-1*	1	1/1	ND	0.051394	ND
CASSrDGMNTEAFF	*TRBV10-2*	2	1/2	0.027940	0.146732	0.001875125
CAWSpDGMNTEAFF	*TRBV30-1*	1	1/1	ND	0.004012	ND
CASSvDGMNTEAFF	*TRBV10-2*	3	1/3	0.02816867	ND	0.002013375
CASSlDGMNTEAFF	*TRBV5-1*	17	7/17	0.05602664	0.2180523	0.000631
* **CASSgDGMNTEAFF**	***TRBV10-2***	**10**	**9/10**	**0.07031889**	**0.2189447**	**0.009374636**
CASSsDGMNTEAFF	*TRBV10-2*	4	3/4	0.205308	0.5404552	0.01325683
CASSdDGMNTEAFF	*TRBV10-2*	8	5/8	0.0142875	0.063278	0.0088616
CASSpDGMNTEAFF	*TRBV10-2*	1	1/1	0.038671	0.082521	0.002561429
CASSwDGMNTEAFF	*TRBV11-3*	1	1/1	ND	0.066804	0.004829
CASSaDGMNTEAFF	*TRBV10-2*	8	6/8	0.08398783	0.1378564	0.01398162
CAISeDGMNTEAFF	*TRBV10-3*	1	1/1	ND	0.05393	ND
CASSeDGMNTEAFF	*TRBV10-2*	18	13/18	0.2954922	0.2520289	0.02317956
CASSqDGMNTEAFF	*TRBV5-1*	1	0/1	0.002324	ND	0.00301675
	(*TRBV10-2*, 85/113) $p = 0.001$		(36/59) $p = 0.035$	Clonal expansion $p = 0.005$		

Figure 2. The TCR members of the tumor-enriched specificity group with the motif "S%DGMNTE" are inferred to recognize tumor antigen in the context of HLA-A*02

(A) Left: volcano plot showing the comparison of the 4,226 clonally expanded specificity groups between tumor (T) and the adjacent lung (N) by Poisson test. The y axis represents the negative $\log_{10}$ converted p values of the Poisson test, and the x axis represents the $\log_2$ converted fold difference between tumor and adjacent lung (T/N). Dot size represents levels of clonal expansion. Tumor-enriched specificity groups (n = 435) are highlighted in red. Right: volcano plot of T/N comparison for CDR3β clonotypes. CDR3β clones of the 435 tumor-enriched specificity groups (left) are highlighted in red.

(B) Pearson correlations and the corresponding p values between the signature scores for the hallmark GSEA gene sets (n = 50) and the percentages of CDR3β clones belonging to the 435 tumor-enriched specificity groups. Significant comparisons are highlighted in red ($p < 0.05$).

(C) Heatmap showing the $-\log_{10}$ p values of top-enriched HLA allele(s) of the 435 tumor-enriched specificity groups. Top, number of MDACC patients carrying each indicated HLA alleles.

(legend continued on next page)

are not the sole determinant of specificity, GLIPH2 analysis of CDR3β sequences leads to correct specificity inferences in the vast majority of cases.

HLA allele enrichment within TCR specificity groups makes robust inferences of HLA restriction

We next examined whether HLA allele enrichment within a specificity group accurately reflected the HLA context annotated by the tetramer. We quantified the enrichment of HLA supertypes across all clonally expanded specificity groups annotated with tetramer CDR3β sequences (Harjanto et al., 2014; Sidney et al., 2008). We focused on the *HLA-A*02* and *HLA-B*08* supertypes since these tetramer-defined HLA contexts were the most abundant in the MDACC dataset (Figure S1C). We reasoned that if a given specificity group was annotated by an HLA/peptide tetramer, there should be a higher probability of observing enrichment of HLA allele(s) belonging to the same supertype by GLIPH2. Indeed, 36.7% of all HLA-A*02 tetramer-annotated specificity groups were enriched with *HLA-A*02* supertype alleles, whereas none of the groups annotated with non-A*02 tetramers were enriched (Figure 1D). While 62.5% of HLA-B*08 tetramer-annotated specificity groups were enriched with *HLA-B*08* supertype alleles, only 3.13% of the non-B*08 tetramer-annotated groups were enriched (Figure 1D). Therefore, the enrichment of a given HLA allele within a specificity group accurately reflected the HLA context of the cognate antigen. Previous work has also validated the inferred HLA restricting element by expressing TCR heterodimers in reporter T cells and identifying their peptide-MHC specificities (Glanville et al., 2017).

Inferred T cell specificities enable robust comparisons of T cell repertoires across patients

One of the major advantages of establishing TCR specificity groups with GLIPH2 is that it greatly facilitates TCR repertoire analysis across individuals. In the MDACC lung cancer dataset, an average ~0.4% of the repertoire was shared between any two patients (Figure 1E). The likelihood of measuring such shared specificities increased to 1.9% when considering the 4,226 shared specificity groups (enriched in clonally expanded TCR sequences) and to 5.3% when considering all 66,094 shared specificity groups (Figure 1E). This demonstrated that GLIPH2 captured shared specificities in the T cell repertoire to an extent that was not possible by only comparing CDR3β sequences across individuals.

Next, we reasoned that if a finite number of shared TCR specificities exist in a particular disease context, the number of specificity groups should reach saturation given enough patients. By bootstrapping from patients who carry at least one copy of the most prevalent *HLA-A*02:01* allele, we found that the number of *HLA-A*02:01*-enriched specificity groups reaches saturation at ~70 patients (Figure 1F). Repertoires from at least nine patients were needed to establish half of all the specificity groups (n = 77) (Figure 1F). In contrast, concurrent bootstrapping from *A*02:01*-negative patients accounted for far fewer *A*02:01*-enriched specificity groups (Figure 1F). In addition, bootstrapping from an independent, healthy cohort with comparable CDR3β sequencing depth did not reach saturation over similar sampling sizes, consistent with a higher prevalence of TCRs belonging to these specificity groups in NSCLC patients carrying the *A*02:01* allele (Figure 1G). Of note, the number of patients needed to establish half of specificity groups was dependent on the level of clonal expansion, the numbers of specificity groups, and the sequencing depth (Figures S1H–S1J). Thus, a complete set of TCR specificity groups could be established with finite patient numbers. Furthermore, these results showed that T cell specificity inference is strengthened by HLA allele enrichment.

Experimental validation of GLIPH2-inferred specificities

Given that experimental validation of T cell specificities requires TCRα/β pairs, we therefore performed single-cell TCR sequencing (scTCR-seq) from 15 early-stage NSCLC patients treated at Stanford (Table S4). Tumor-infiltrating T cells were prepared from surgically resected specimens and index sorted by fluorescence-activated cell sorting (FACS) before sequencing (Figure S2A). scTCR-seq yielded 4,704 paired CDR3α and CDR3β sequences. We combined these CDR3β sequences with the MDACC NSCLC sequences for a joint GLIPH2 analysis. We chose to validate four T cell clones belonging to three flu M1-annotated specificity groups (SV%SNQP, SIRS%YE, and S%RSTDT) and one EBV BMLF1-annotated specificity group (RTG%GNT). We used Jurkat 76 cells, deficient for both TCRα and TCRβ, to express the four TCR candidates and co-cultured them with HLA-A*02$^+$ T2 cells pulsed with their respective peptides (Figures S2B and S2C). Three of them responded to their predicted antigens in the context of HLA-A*02, showing the robustness of GLIPH2 for inferring T cell specificities (Figures S2B and S2C). Similar analyses of specificity group members in *M. tuberculosis* studies found that ~80%–90% of the TCRs recognized the predicted peptide-MHC ligands (Glanville et al., 2017).

Characterization of tumor-enriched specificity groups

To identify disease-relevant specificity groups, we focused on the 435 tumor-enriched specificity groups that revealed a strong clonal bias in the tumor compared to the adjacent lung (Figures 2A and S1A). Using the transcriptome data available from 84 patients (total n = 178), we found that the percentage of T cells belonging to these tumor-enriched specificity groups correlated with gene set enrichment analysis (GSEA) hallmark signatures of cancer progression, including *MYC* and the cell cycle programs (Figures 2B). In contrast, using the specificity groups expanded in the adjacent lungs (n = 114), we failed to observe any significant correlation with the GSEA hallmark gene sets (data not

(D) Number of top-enriched HLA allele(s) found in each of the 435 tumor-enriched specificity groups.
(E) Volcano plot for the 4,226 NSCLC specificity groups as in (A, left). The tumor-enriched specificity groups significantly enriched with *HLA-A*02* alleles ($p < 0.05$ by Fisher's exact test) are colored in green. The specificity group "S%DGMNTE" is highlighted.
(F) The distinct CDR3β sequence members of the "S%DGMNTE" specificity group. For each CDR3β sequence, the *Vβ* gene usage (*Vβ*), number of patients with each sequence (patient counts), number of HLA-A*02$^+$ patients (counts of HLA-A*02$^+$ cases/total), and the average clonal frequencies (% by patient) found in the adjacent lung, tumor, and peripheral blood are shown. ND, not detected. Bottom: p values for the enrichment of *Vβ* gene usage, *HLA-A*02* alleles, and the level of clonal expansion are shown.

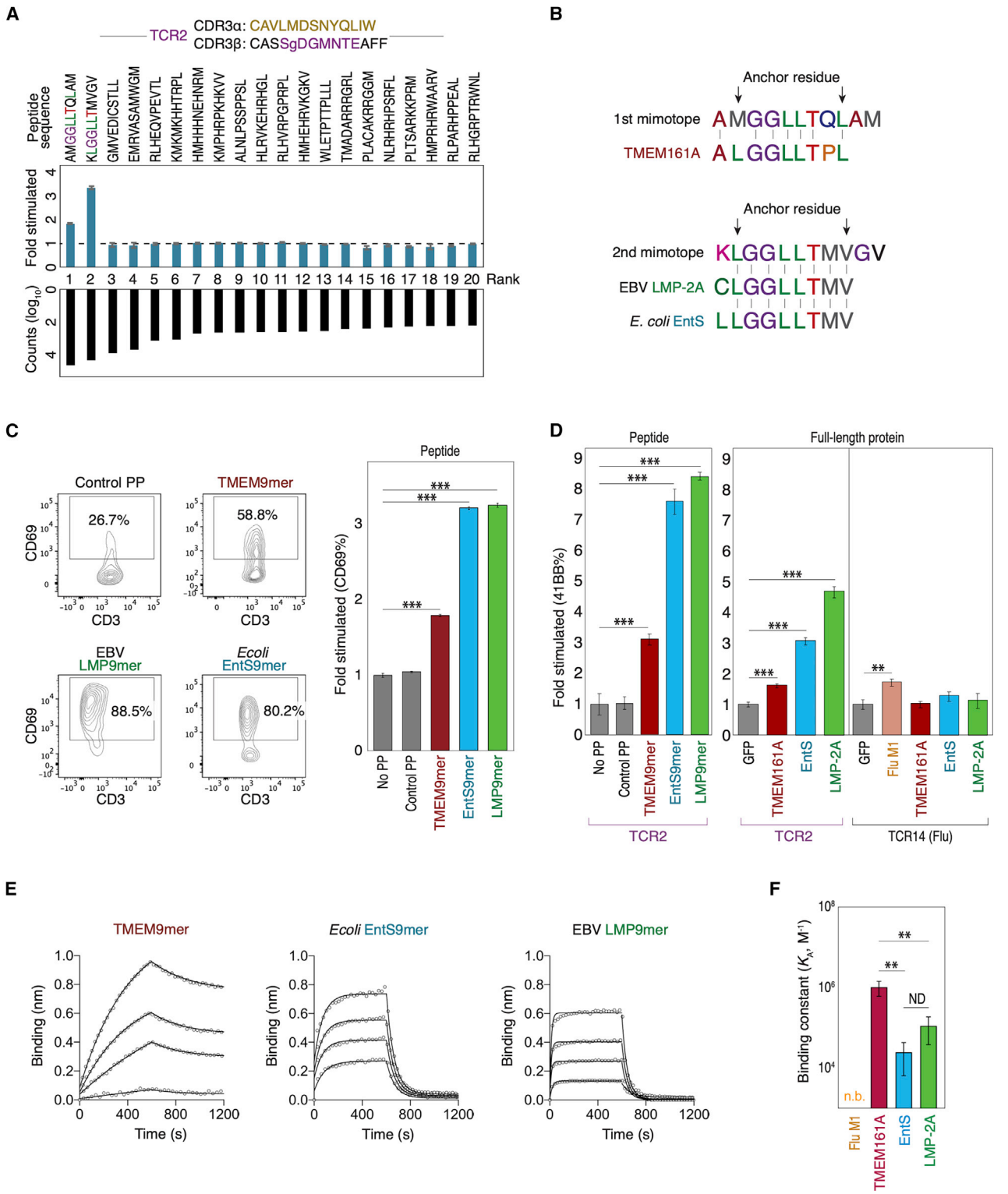


Figure 3. Identification of tumor and pathogen-derived antigens recognized by a tumor-enriched TCR in human lung cancer

(A) Top: top-20 mimotopes from the 4th round of selection on an 11-mer yeast library are used to stimulate Jurkat-TCR2 cells. CD69 fold change is shown compared to unstimulated control. Bottom: ranked raw counts (log_{10}) of the enriched mimotopes from the selection.

(B) Alignment of the top-two mimotopes with peptides from the human *TMEM161A* locus, EBV *LMP-2A*, and *E. coli EntS*. All peptides were 9-mers and predicted to bind HLA-A*02 with high affinities.

(legend continued on next page)

shown). Thus, this result showed a correlation between the 435 tumor-enriched specificity groups and an aggressive, highly proliferative cancer phenotype. Next, we systematically examined the enrichments of all HLA alleles in the MDACC cohort for the 435 tumor-enriched specificity groups and found only one predominant allele in most cases (n = 202/435; Figures 2C and 2D). Of note, we found that in cases when motifs were enriched with multiple predominant alleles, e.g., those co-enriched with both *HLA-B*07:02* and *HLA-C*07:02* (Figure 2C), strong linkage disequilibrium in the associated HLA alleles could be observed.

Identification of a shared specificity group cross-reactive to tumor and pathogen-derived antigens in human lung cancer

Of the 435 tumor-enriched specificity groups, we prioritized those that fulfilled the criteria of (1) having a paired TCRα/β clonotype from the Stanford cohort and (2) significantly enriched with *HLA-A*02* alleles by Fisher's exact test. This led us to focus on the specificity group with the "S%DGMNTE" CDR3β motif (Figures 2E and 2F). Hence, the candidate TCRα/β clonotype (referred to as TCR2) bearing the CDR3α sequence CAVLMDSNYQLIW and CDR3β sequence CASSGDGMNTEAFF was chosen for antigen identification (Figure 3A).

To identify the cognate epitopes of the candidate clone TCR2, we screened a yeast library displaying peptides of four different lengths (8–11 amino acids) in the context of wild-type HLA-A*02:01 (Gee et al., 2018). Four rounds of selection with a multimer of TCR2 led to the enrichment of peptide sequences (mimotopes) in the 11-mer library (Table S5). We performed an *in vitro* stimulation assay with the top-20 enriched mimotopes and showed that the top-two sequences "AMGGLLTQLAM" and "KLGGLLTMVGV" stimulated Jurkat cells expressing TCR2 (Jurkat-TCR2) (Figures 3A and S3A). A protein database search (UniParc) (UniProt Consortium, 2019) led to the identification of multiple endogenous 9-mers that shared close sequence similarities with the top-two mimotopes and were predicted to bind HLA-A*02:01 with anchors separated by six instead of eight amino acids (Figures 3B and S3B). Indeed, 9-mer variants of the top mimotope stimulated Jurkat-TCR2 cells to comparable levels as the 11-mer counterpart (Figure S3C). This result indicated that the identified HLA-A*02 antigens were *de facto* 9-mers.

We functionally validated all candidate endogenous peptide 9-mers resembling the top-two mimetopes (11-mer) (Figure S3B). We found that 9-mers from the mammalian protein TMEM161A (TMEM9-mer, ALGGLLTPL), the latent membrane protein 2a (LMP9-mer, CLGGLLTMV) from EBV, and the enterobactin exporter (EntS9-mer, LLGGLLTMV) from *E. coli* could all stimulate the Jurkat-TCR2 cells (Figures 3C, S3D, and S3E). These results demonstrated that TCR2 was cross-reactive to antigens from humans and pathogens. The accurate GLIPH2 inference of HLA restriction facilitated antigen discovery with the HLA-A*02:01 yeast library.

To show that the full-length proteins TMEM161A, LMP2, and EntS could be processed, presented on HLA-A*02:01, and activate specific T cells, we overexpressed these proteins in HLA-A*02$^+$ 293T cells and measured the responses of co-cultured primary T cells expressing TCR2. Similar to the pulsed peptides, 293T cells expressing full-length TMEM161A, LMP2, and EntS all stimulated the co-cultured TCR2-T cells, with TMEM161A appearing to be the weakest stimulator (Figure 3D). We further performed a biolayer interferometry to quantify the binding affinity of each cross-reactive epitopes to TCR2 and showed that the weakest stimulator TMEM9-mer revealed the most stable binding to TCR2 (Figures 3E and 3F). Thus, this result suggested a partial uncoupling of binding affinity and signaling strength, similar to the previous report (Sibener et al., 2018). In summary, we identified a tumor-enriched TCR specificity group with cross-reactivity to both a TAA and pathogen-derived antigens.

TMEM161A is overexpressed on human lung cancer

We found significantly higher levels of TMEM161A protein expression in human lung cancer compared to adjacent lung tissue (Figures 4A, 4B, and S4A). We also noted some heterogeneity in TMEM161A expression on some tumor sections (Figure S4B). We also examined *TMEM161A* gene expression in the Cancer Genome Atlas (TCGA) NSCLC dataset. Consistent with protein expression, we found higher levels of *TMEM161A* transcript in tumors compared to the adjacent lung. The level of *TMEM161A* expression was higher in squamous cell carcinomas (SCCs) of the lung compared to adenocarcinomas (Figure 4C). Whole-exome sequencing of specimens from the Stanford cohort did not identify any mutation within the coding region of the *TMEM161A* locus, supporting its role as a non-mutated TAA (Table S6). Similarly, less than 1% of deleterious mutations in the *TMEM161A* locus were found in the pan-lung cancer TCGA dataset (n = 6/1053; Figure S4C). In addition, *TMEM161A* expression in lung cancer associated with GSEA signatures related to cell proliferation programs and the proto-oncogene *MYC* targets, consistent with the general trend revealed by the 435 tumor-enriched specificity groups (Figures 2B, 4D, and 4E). In contrast, *TMEM161A* expression appeared to inversely correlate with gene sets related to inflammatory responses (Figures 4D and 4E). These data showed that TMEM161A is a TAA

(C) Left: representative FACS plots showing the stimulation of the Jurkat-TCR2 cells with 9-mers from the human *TMEM161A* locus (TMEM9-mer), *LMP-2A* of EBV (LMP9-mer), and *EntS* from *E. coli* (EntS9-mer); right: results of Jurkat-TCR2 cell stimulation in triplicate. Control PP, control peptide (GILGFVFTL); No PP, no peptide.

(D) Stimulation of primary T cells ectopically expressing TCR2 TCRα/β chains with either 9-mers (left) or full-length proteins (right). Stimulation of primary T cells expressing TCR14 by 293T-A*02 cells expressing full-length FluM1 protein was shown as control. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ by t test. Control PP, control peptide (GILGFVFTL).

(E) The binding of TCR2 to the indicated A*02/9-mers was determined by biolayer interferometry. An overlay of binding traces over a concentration series of the indicated A*02/9-mers from one representative experiment is shown. The data points are represented as open circles and the fits from a simple 1:1 Langmuir interaction model are indicated by solid lines. Each binding experiment was repeated three times.

(F) The equilibrium association constants (KA) of the surface plasma resonance as in (E). The flu M1 peptide showed no detectable binding (n.b.) to TCR2. Significance was determined by t test after one-way ANOVA. The reported p values were corrected for multiple comparisons. **$p < 0.01$. ND, not different. All error bars represent standard deviation of the mean.

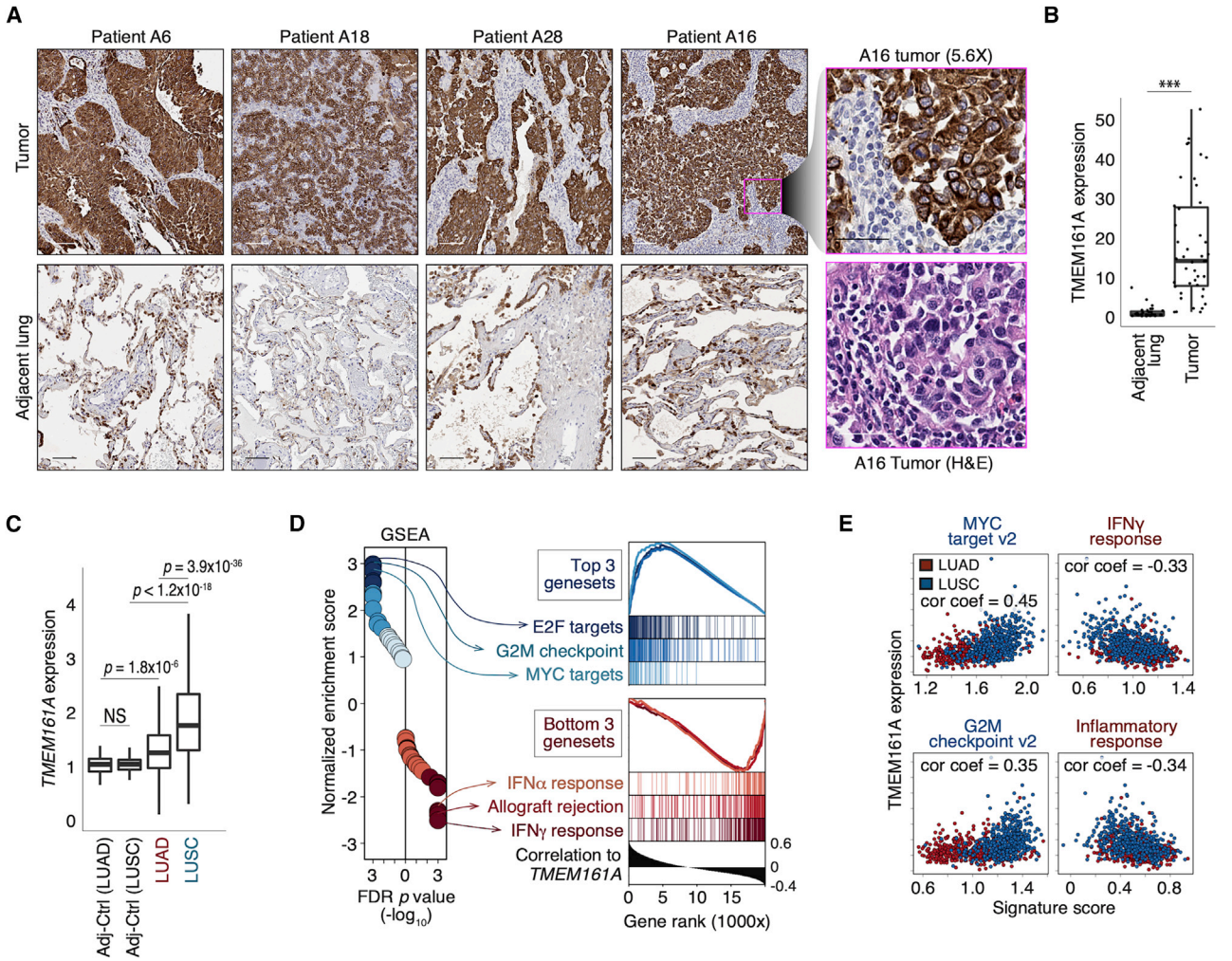


Figure 4. TMEM161A protein is highly expressed in human lung cancer

(A) Representative images of TMEM161A immunohistochemistry on tumor (top) and the adjacent lung (bottom) sections from four patients. Scale bar, 100 μm. Rightmost panels: zoomed in images of patient A16 tumor with TMEM161A immunohistochemistry (top) and H&E staining on a serial section (bottom). Scale bar, 40 μm.

(B) Quantification of TMEM161A immunohistochemistry on sections from the Stanford NSCLC cohort (n = 11). Boxplots show medians with the first (25th) and third (75th) quartiles with individual data points. ***p < 0.001.

(C) *TMEM161A* expression quantified by bulk RNA-seq of the indicated samples from TCGA (n = 958) is shown in boxplots. Adj-Ctrl, the adjacent lung control. *TMEM161A* expression normalized against Adj-Ctrl is shown. p values were calculated with the Wilcoxon Rank Sum test. ND, not significantly different. Boxplots represent medians with the first (25th) and third (75th) quartiles.

(D) Gene set enrichment analysis of the ranked gene list based on Pearson correlation with *TMEM161A* abundance in the pan-lung cancer TCGA dataset (n = 958). Left: hallmark gene sets with highest (blue) and lowest (red) normalized enrichment scores are indicated, and their enrichment curves are shown (right).

(E) Single-sample GSEA signature scores (Sig score) of two most and two least enriched hallmark signatures are plotted against *TMEM161A* expression. Pearson correlation coefficients are shown in plots (cor coef).

overexpressed in human NSCLC and associated with gene expression signatures such as MYC and cell cycle.

T cells recognizing TMEM161A antigen have the "S%DGMNTE" sequence motif

We further interrogated the TCR sequence identity of TMEM161A-specific CD8$^+$ T cells *in vivo* and examined their clinical relevance. TMEM161A-specific T cells could be detected in 31/78 (40%) of HLA-A*02$^+$ patients in the MDACC NSCLC cohort. We used TMEM9-mer/HLA-A*02 tetramers to sort T cells from the tumor of patient A6, where the TCR2 clone was first identified. scTCR-seq of TMEM9-mer/A02 tetramer$^+$ T cells from tumor and the adjacent lung confirmed that they carried the "S%DGMNTE" motif, consistent with their recognition of TMEM161A *in vivo* (Figures S4D and S4E). We next examined how tumor characteristics impact the recruitment of T cells with the "S%DGMNTE" motif among patients who were HLA-A02$^+$. We observed that T cells with the "S%DGMNTE" motif were observed more frequently in SCCs compared to adenocarcinomas, similar to the expression pattern of TMEM161A (Figures 4C and S4F). We also noted that the percentage of T cells with the "S%DGMNTE" motif in tumors with a mutation count

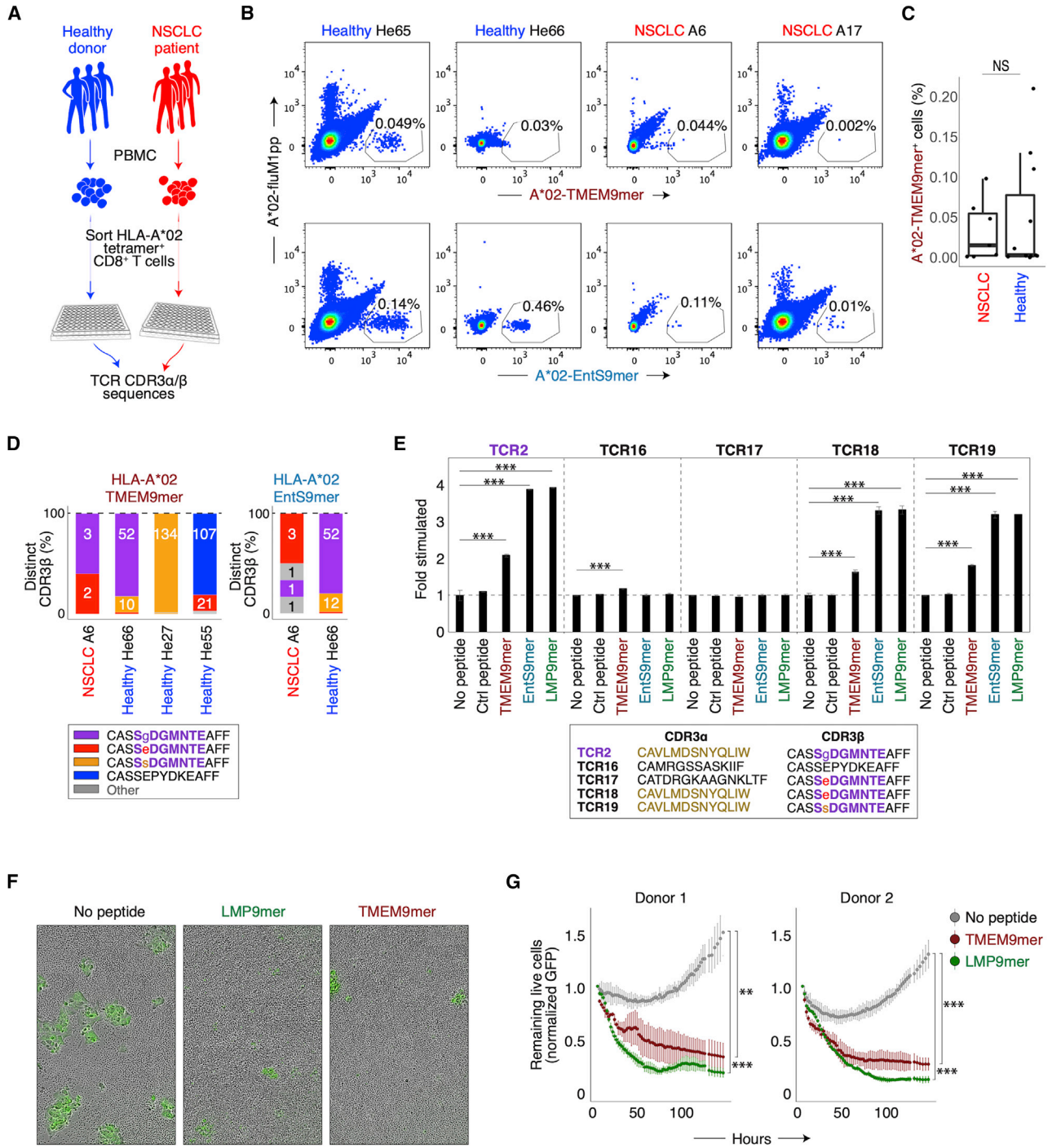


Figure 5. Isolation and characterization of cross-reactive TMEM161A-specific T cells from peripheral blood of healthy donors and lung cancer patients

(A) Schematic showing the procedure used to capture antigen-specific T cell clones from HLA-A*02$^+$ healthy donors and NSCLC patients. Cells were sorted by FACS directly into 96-well plates for scRNA-seq and scTCR-seq.

(B) Representative FACS plots of T cells sorted with indicated tetramers from the PBMC of HLA-A*02$^+$ healthy donors (He65 and He66) or HLA-A*02$^+$ NSCLC patients (A6 and A17).

(C) Percentage of tetramer$^+$ T cells from healthy donors (n = 11) and NSCLC patients (n = 7). Boxes represent medians with the first (25th) and third (75th) quartiles. NS, not significantly different.

(D) Percentage of distinct CDR3β sequences in tetramer-sorted T cells from healthy donors and NSCLC patient. Numbers in plots represent the cell counts.

(E) Indicated TCR clonotypes identified with tetramers were expressed in Jurkat cells and co-cultured with T2 cells pulsed with indicated 9-mers. y axis (fold stimulated) shows activation by CD69 fold change compared to unstimulated control. ***p < 0.001. Ctrl peptide, control peptide (GILGFVFTL).

(legend continued on next page)

of less than 500 was higher than in tumors with mutation count of greater than 500 (total n = 34), although this observation may be impacted by the association between total infiltrating T cell numbers and mutation burden (Figure S5A). Finally, although the presence of detected T cells with the "S%DGMNTE" motif in tumors alone did not predict patient outcome, we observed that T cells with the "S%DGMNTE" CDR3β motif were among 146 shared specificity groups enriched in patients without recurrence (Figures S5B–S5D).

CD8$^+$ T cells with the "S%DGMNTE" motif were also detected in healthy donors

To characterize the cross-reactive TMEM161A-specific and pathogen-specific clonotypes, we used TMEM9-mer/HLA-A*02 tetramers or EntS9-mer/HLA-A*02 tetramers to sort CD8$^+$ T cells from the peripheral blood of HLA-A*02$^+$ healthy donors and NSCLC patients by FACS (Figure 5A). We saw no difference in the frequency of HLA-A*02/TMEM9-mer$^+$ CD8 T cells in healthy donors and lung cancer patients (Figures 5B and 5C), suggesting that these T cells were likely maintained due to cross-reactivity to pathogen-derived antigens. Consistent with this, the frequencies of these specific T cells, as quantified by tetramers or GLIPH2, were approximately one in every 10^3–10^5 T cells (tetramer-measured: 0.0032%–0.0980%; GLIPH2-inferred: 0%–0.2643%), higher than naive level for human CD8$^+$ T cells (Yu et al., 2015).

Regardless of which tetramer was used to sort peripheral blood T cells, the CDR3β sequences of the sorted cells consistently carried the "S%DGMNTE" motif (Figure 5D). In fact, we found a variety of CDR3β sequences sharing the "S%DGMNTE" motif where % could be a glycine, glutamate, or serine, confirming the diversity seen in the GLIPH2 analysis using the MDACC data (Figures 2F and 5D). Furthermore, single-cell RNA sequencing (scRNA-seq) data suggested that HLA-A*02/TMEM9-mer$^+$ cells mostly manifested effector T cell states, indicating that they had previously encountered their cognate antigens, even in healthy individuals (Figures S5E and S5F).

To functionally validate CDR3α/β sequences from the tetramer-sorted clones, we generated stable Jurkat cells expressing the TCRα/β chains identified with the tetramers. We then quantified their reactivities to both TMEM9-mer and pathogen-derived 9-mers in the context of HLA-A*02:01. We found that the Jurkat cell clones with the "S%DGMNTE" CDR3β motif could respond to all cross-reactive peptides only when paired with the permissive TCR2α chain (CDR3α: CAVLMDSNYQLIW; Figure 5E). For example, we identified a CDR3α/β pair that did not carry the "S%DGMNTE" motif and recognized TMEM9-mer but not the microbial antigens (TCR16; Figure 5E). Finally, we quantified the cell-mediated cytotoxicity induced by the cross-reactive epitopes by co-culturing an HLA-A*02$^+$ lung cancer cell line H1395 with primary T cells expressing TCR2. Compared to the no peptide control, both LMP9-mer and TMEM9-mer induced more than 50% of target cell lysis (Figures 5F and 5G). Cancer cells pulsed with TMEM9-mer were weaker targets for cell-mediated cytotoxicity compared to those with LMP9-mer, consistent with the results of the T cell activation studies (Figures 3C and 3D). In summary, CD8$^+$ T cells with the "S%DGMNTE" motif cross-reacted with the TMEM161A tumor antigen and the pathogen-derived antigens EntS and LMP2 when paired with the permissive α chain. Recognition of these cross-reactive antigens on HLA-A*02 led to target cell lysis by CD8$^+$ T cells with the "S%DGMNTE" motif.

Phenotypic characterization of TMEM161A-specific CD8$^+$ T cells in lung cancer

We sequenced the full single-cell transcriptomes of 2,950 sorted, tumor-infiltrating T cells from 10 NSCLC patients using the SMART-seq method and acquired their paired CDR3α/β repertoires (Figure S2A) (Han et al., 2014; Stubbington et al., 2016). We identified 14 major cell states of which 13 could be mapped to those reported in a separate cohort (Figures 6A and S6A–S6C; Table S7) (Guo et al., 2018). Clusters c5, c6, c12 (CD8$^+$ T cells with effector phenotypes), c7, and c10 (CD8$^+$ T cells with resident memory phenotype) were among the most expanded (Figures 6B and 6C). To uncover the cell states of clones specific for shared antigens, we examined the scRNA-seq profiles of the TCR specificity groups members. We found that 2.9% of the T cells (n = 86/2950) belonged to the clonally expanded specificity groups (top, Figures 6D). Twelve of these T cells were members of the 435 tumor-enriched specificity groups, whereas 13 of these T cells were inferred to be specific to viral epitopes (Figures 1C and 6D). Interestingly, T cells belonging to the tumor-enriched specificity groups were biased toward the effector phenotype (c5) and differentially expressed EOMES, KLRG1, GZMK, and other genes expressed in activated natural killer cells (Figures 6D–6F; Table S7). Consistently, HLA-A*02/TMEM9-mer tetramer-sorted CD8$^+$ T cells from tumor also preferentially exhibited the effector T cell phenotype c5 (Figures 6D–6F). Pseudotime trajectories and activation/exhaustion signature scores indicate that these T cells adopt distinct cell states (Figures 6G and 6H). In comparison, T cells inferred to be virus-specific exhibited cell states that included effector (c5, c6, c12) and tissue resident-memory phenotypes (c7). In conclusion, TMEM161A-specific CD8$^+$ T cells showed a range of effector T cell states in NSCLC, consistent with recognition of their cognate antigen *in situ*.

Expansion of EBV-specific T cell clones in patients responding to immune checkpoint blockade

To see if pathogen-specific T cells might impact clinical responses to anti-PD1 checkpoint immunotherapy, we analyzed the TCR repertoire of two NSCLC patients who experienced a clinical response to treatment (Figure S7A). We sequenced paired CDR3α/β repertoires on both pre- and post-treatment blood samples and identified 102 CDR3β clonotypes that expanded in post-treatment samples (Figure 7A). Of these expanded clones, 41 belonged to 99 specificity groups identified in tumor-infiltrating T cell CDR3β repertoires (total n = 66,094; Figure S1A). We used tetramer-defined T cell CDR3β sequences

(F and G) Cell-mediated cytotoxicity of H1395 lung cancer cells. Primary T cells ectopically expressing TCR2α/β chains were co-cultured with the A*02$^+$ H1395 cancer cells and pulsed with either no peptide, TMEM9-mer, or LMP9-mer. Representative images (F) and results using cells from two different donors (G) are shown. **p < 0.01; ***p < 0.001 by t test. Error bars represent standard deviation of the mean.

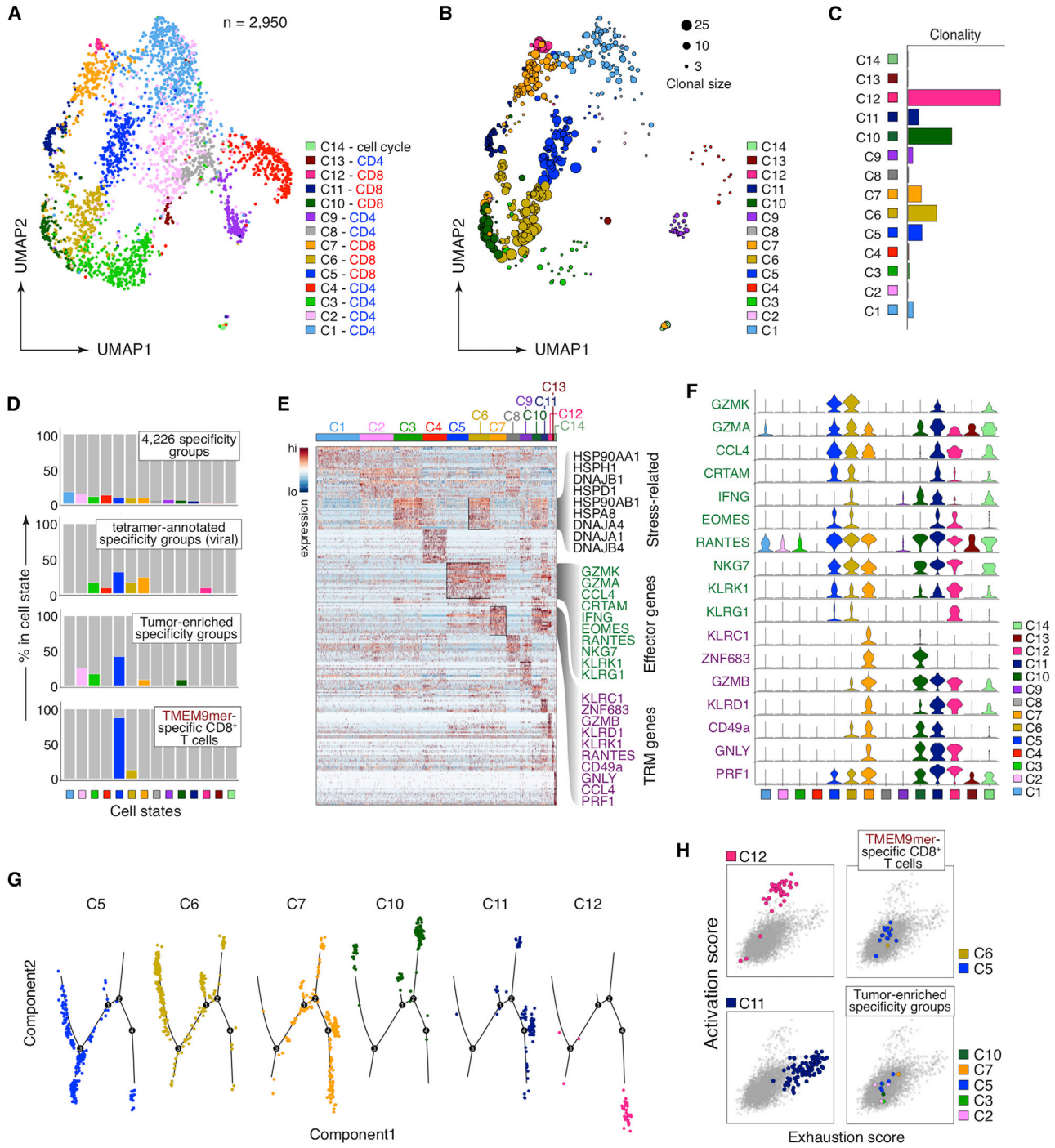


Figure 6. Phenotypic characterization of the TMEM161A-specific CD8⁺ T cells

(A and B) Dimension reduction by Uniform Manifold Approximation and Projection (UMAP) of the scRNA-seq data from 2,950 sorted tumor-infiltrating T cells from 10 NSCLC patients (Stanford cohort). The identified cell clusters (n = 14) are labeled with distinct colors (A) and shown with varying dot sizes representing the level of clonal expansion (B).

(C) Clonality of the 2,950 sorted T cells as in (B) quantified as 1 - Pielou's evenness.

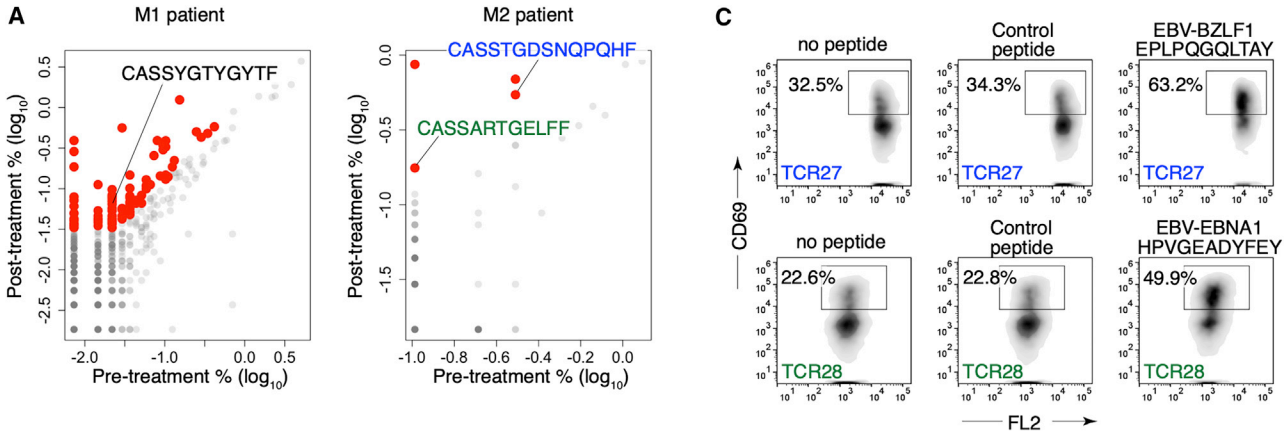
(D) Breakdown of cell states for T cell clones of the 4,226 specificity groups defined in Figure S1A (top), viral-related specificity groups (second from top), the 435 tumor-enriched specificity groups (third from top), and TMEM9-mer/A*02 tetramer-sorted CD8 T cells from tumor (bottom, patient A6).

(E) Heatmap showing differentially expressed genes for each cell cluster defined in (A). Select differential genes for cluster c5, c6, and c7 are highlighted.

(F) Stacked violin plot showing the expression of highlighted differential genes in (E) in all cell clusters.

(G) Pseudotime trajectory of CD8⁺ single cells by Monocle (v2.10.1).

(H) Exhaustion score versus activation score for CD8⁺ T cells sorted by the HLA-A*02/TMEM9-mer tetramer (top right) and those that belong to tumor-enriched specificity groups (bottom right), colored by the cluster identity. Exhausted CD8⁺ T cells (c11) and activated CTL (c12) are shown for comparison.



B

Motif sequence	Pathogen	Antigen	Tetramer HLA	CDR3b sequence	Patient ID	HLA-A*02 p value	HLA-B*35 p value
STGD%NQP	EBV	BZLF1	HLA-B*35	CASSTGDSNQPQHF	M2	-	$6.96x10^{-5}$
%TGDSNQP	EBV	BZLF1	HLA-B*35	CASSTGDSNQPQHF	M2	-	0.0004
ST%DSNQP	EBV	BZLF1	HLA-B*35	CASSTGDSNQPQHF	M2	-	0.0013
STG%SNQP	EBV	BZLF1	HLA-B*35	CASSTGDSNQPQHF	M2	-	0.0027
S%GDSNQP	EBV	BZLF1	HLA-B*35	CASSTGDSNQPQHF	M2	-	0.0141
SARTG%	EBV	EBNA1	HLA-B*35	CASSARTGELFF	M2	-	$3.45x10^{-6}$
S%RTGE	EBV	EBNA1	HLA-B*35	CASSARTGELFF	M2	-	0.0078
SAR%GE	EBV	EBNA1	HLA-B*35	CASSARTGELFF	M2	-	0.0009
SA%TGE	EBV	EBNA1	HLA-B*35	CASSARTGELFF	M2	-	0.0082
SART%E	EBV	EBNA1	HLA-B*35	CASSARTGELFF	M2	-	$9.21x10^{-5}$
S%GTYG	Flu	M1	HLA-A*2	CASSYGTYGYTF	M1	0.0433	-

TCR27α CAVRYPPGGYQKVTF TCR27β CASSTGDSNQPQHF
TCR28α CAVQDLETSGSRLTF TCR28β CASSARTGELFF

Figure 7. Virus-specific CD8 T cell clones expanded in patients responding to anti-PD1 treatment

(A) Comparisons of pre- and post-treatment CDR3β clonal frequencies (in $\log_{10}$ percent) in the peripheral blood of patient M1 (left) and M2 (right). CDR3β clones inferred to recognize viral antigens are highlighted.

(B) Specificity groups containing expanded CDR3β clones post-treatment (column 5, CDR3β sequence) from patients M1 or M2 (column 6, Patient ID) that are annotated with viral tetramer CDR3β sequences (column 2–4, antigen and HLA alleles of the tetramers). Enrichment of the *A*02:01* or *B*35:01* allele is shown (last two columns, p values from the hypergeometric tests are shown). CDR3α/β sequences of the two EBV-related expanded clones from patient M2 are shown at the bottom.

(C) TCR27- (CDR3β: CASSTGDSNQPQHF, top panels) and TCR28- (CDR3β: CASSARTGELFF, bottom panels) Jurkat cell lines were created and tested for their reactivities to the predicted EBV antigens in the context of B*35 as shown in (B). TCR27- and TCR28-Jurkat cells were co-cultured with T2-B*35 cells pulsed with indicated peptides (above each plot). Level of activation was quantified with CD69 expression. Control peptide: LPFDFTPGY.

to annotate these specificity groups and found 11 (total n = 99) containing 3 expanded CDR3β clones inferred to recognize EBV and flu antigens (Figure 7B). To validate the specificity inferences, we created two Jurkat cell clones expressing the TCRα/β chains inferred to recognize the EBV antigens and a T2 cell line expressing wild-type B*35 (Figures 7B and S7B). Indeed, upon co-culture with the T2-B*35 cells, both Jurkat-TCR27 and -TCR28 cells responded to the predicted EBV peptides (Figure 7C). Of note, these EBV-specific specificity groups were not only expanded post-treatment but also showed a bias in tumor compared to the adjacent lung, suggesting the potential cross-reactivities to unknown TAAs (Figure S7C). Furthermore, we found that the EBV-specific clone TCR15 (CDR3β: CSARTGVGNTIYF) identified from patient A11 (Figure S2C) was inferred to have the same antigen specificity as two previously reported clones detected in patients receiving immune checkpoint blockade at the time of clinical response (CDR3β: CSARVGVGNTIYF and CSARSGVGNTIYF) (Anagnostou et al., 2019). Our analysis suggested that these clones belonged to the "R%GVGNT" specificity group predicted to recognize EBV-BMLF1 (GLCTLVAML) in the HLA-A*02 context. We further tested three similar epitopes from the human ORFeome that were predicted to bind HLA-A*02 and found the endogenous "LLGTLVAML" from the human *CLDN2* locus also stimulated the Jurkat-TCR15 clone (Figure S7D), indicating that TCR15 was indeed cross-reactive to both EBV and a TAA. In summary, these results indicated that pathogen-specific T cells in patients might play a role in the anti-tumor immune responses upon treatment with immune checkpoint inhibitors.

DISCUSSION

While recent work on T cell specificities in cancer has focused on neoantigens that are typically unique to individuals, prior work

also describes shared tumor antigens that are inappropriately expressed or overexpressed in tumors. Here, we developed an approach to systematically survey the TCR repertoire of a substantial number of NSCLC patients to uncover shared T cell specificities. Using the GLIPH2 algorithm, we first distilled this raw TCR sequence data into a much smaller and more useful collection of shared specificity groups with inferred HLA restrictions. We then prioritized disease-relevant TCR candidates for antigen discovery. The enormous diversity of the yeast library greatly facilitated antigen identification and the discovery of cross-reactive antigens. Unlike other MHC/peptide libraries built in mammalian cells, the yeast libraries incorporate close to 10^9 randomly permutated peptide sequences (Gee et al., 2018; Joglekar et al., 2019; Kula et al., 2019; Li et al., 2019). While previously the uncertainty of HLA restriction limited the success of antigen identification using the yeast library (Gee et al., 2018), we overcame this limitation by using GLIPH2 to infer the correct HLA context of the candidate TCR.

Using this approach in lung cancer, we discovered examples of TCRs cross-reactive to both tumor and microbial antigens. Thus, this seems to be a likely explanation for the reports of pathogen-specific T cells infiltrating tumors (Andersen et al., 2012; Rosato et al., 2019; Scheper et al., 2019; Simoni et al., 2018). We previously proposed that maintaining a broad T cell repertoire to defend against pathogens may rely heavily on TCR cross-reactivity (Su et al., 2013). T cells specific to self-antigens have been detected in the peripheral blood of healthy individuals, pruned but not clonally deleted in the thymus, potentially to avoid immunologic "blind spots" to pathogens (Sewell, 2012; Yu et al., 2015). Because cancer cells overexpress self-antigens, T cell specificity for self-antigens may partly explain why previous studies observed low reactivities of tumor-infiltrating T cells to autologous tumor (Scheper et al., 2019). In this study, we observed that TMEM161A-specific T cells were relatively weak responders to the self-antigen TMEM161A compared to antigens from EBV and *E. coli*. Despite this weak reactivity, the data presented here show that the binding affinity of TCR2 to the TMEM9-mer/A*02:01 ligand is higher than LMP2 and EntS. This suggests that in tumors, the uncoupling of TCR binding from T cell activation may be yet another mechanism by which the natural course of specific responses against TAAs are dampened during tumor progression. This provides a possible explanation for why these T cells are localized to tumors where TMEM161A is overexpressed but where EBV and *E. coli* are likely absent. In this regard, previous reports show that EBV is rarely detected in lung cancer (Kheir et al., 2019) and *E. coli* is rarely detected in the lung sputum (Cameron et al., 2017; Dickson et al., 2016).

Previously, common pathogen-specific T cells found in tumors have been presumed to be "bystanders" and not specific for TAAs. Our data showed that T cell specificities for TAAs and pathogen-derived antigens were not mutually exclusive. Furthermore, these pathogen-specific T cells in tumors exhibited an effector phenotype rather than an exhausted or stressed state and lacked CD39 expression (Simoni et al., 2018). In this study, we described examples of cross-reactive T cells with weaker reactivities to shared, non-mutated tumor antigens compared to the cross-reactive microbial antigens. Despite this weaker reactivity, our data suggested that cross-reactive T cells might play a role in controlling cancer progression in the setting of anti-PD1 checkpoint blockade. Although it is still unclear what roles these cross-reactive T cells play in the anti-tumor immune response unleashed by immune checkpoint blockade, it is tempting to speculate that exposure to cross-reactive microbial antigens might overcome tolerance for non-mutated tumor or self-antigens (Ohashi et al., 1991; Röcken et al., 1992). The idea that pathogens could be the basis of immunotherapy was suggested originally from the work of William Coley who, in the late nineteenth century, pioneered a mixed bacterial vaccine termed Coley's toxin for the treatment of cancer patients with some success (McCarthy, 2006). Recently the gut microbiome has been shown to be a key determinant of immunotherapy responses in cancer (Gopalakrishnan et al., 2018; Matson et al., 2018; Routy et al., 2018; Sivan et al., 2015; Vétizou et al., 2015). In pancreatic cancer, a unique microbiome composition has been observed in patients with longest survival after surgery (Riquelme et al., 2019). Cross-reactive T cells recognizing both tumor antigens and microbial antigens have also been shown to control tumor growth in mouse models (Bessell et al., 2020; Fluckiger et al., 2020). In addition, EBV and flu have recently been shown to induce anti-tumor immunity against shared TAAs (Choi et al., 2020; Newman et al., 2020). Additional studies are needed to understand whether or not there is a causal relationship between the microbe/TAA-cross-reactive T cells and clinical benefit from immune checkpoint blockade.

In summary, we present a resource for comprehensively characterizing TCRs from a large NSCLC patient cohort using methodologies that could be applied to any tumor type. Thus, we reduced almost 800,000 TCR sequences from 178 patients to over 66,000 specificities shared by three or more individuals. Of the 66,000 specificity groups we identified, we then subsetted these into 435 specificities that were enriched in the tumors versus adjacent lung. This number may represent a much smaller number of antigens since a given peptide-MHC ligand can elicit five or more different specificity groups (Glanville et al., 2017). We found an intriguing cross-reactivity between non-mutated tumor antigens and pathogens, which could explain recent puzzling results describing nominally virus-specific T cells infiltrating tumors (Simoni et al., 2018), implying, as does other data presented here, that this cross-reactivity may be a common phenomenon. This raises the prospect that memory T cells to these pathogenic epitopes could trigger a cross-reactive response against cancer. Perhaps during the early phase of neoplasia, pre-cancerous cells that happen to overexpress self-antigens (mutated or not) that are cross-reactive to similar antigens from EBV interact with these T cells to create a chronic, low-grade inflammatory tumor microenvironment. In support of this, we observe that cross-reactive T cells express high levels of granzyme K, which has been reported in the context of inflammatory diseases and aging (Corridoni et al., 2020; Mogilenko et al., 2020). Since inflammation is known to promote neoplasia, this could then facilitate the process by which some cells become malignant.

Limitations of the study

The GLIPH2 algorithm infers T cell specificities based on TCRβ sequences only. Thus, it captures only a portion of all input sequences. Although the current study focuses on T cells from lung cancer, many of the shared specificity groups generated,

including the "S%DGMNTE" motif, are anticipated to overlap with other cancer types and can serve as a template for analysis. However, further studies are needed in order to establish a complete shared specificity landscape for other cancer types. Finally, we identified only a few examples of cross-reactive specificities and thus we cannot rule out the possibility that at least some pathogen-specific T cells in lung cancer infiltrates were not cross-reactive to TAA and therefore true "bystanders."

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Tissue processing
  - FACS analyses
  - Establishment of T cell specificity groups
  - Classification of TCRs and specificity groups
  - Annotation of specificity groups
  - Validation of HLA restriction inference
  - *HLA-A*02:01* specificity group bootstrapping
  - GSEA analysis of the TCGA data
  - FACS sorting of antigen-specific CD8 T cells
  - Single-cell RNA-seq (scRNA-Seq)
  - Single-cell TCR sequencing (scTCR-seq)
  - Data analyses of scRNA-Seq results
  - GLIPH2 analysis on TRACERx data
  - Soluble biotinylated TCRα/β synthesis
  - Antigen discovery with the yeast library
  - Lentiviral TCR transduction
  - Retroviral TCR transduction
  - Cell culture
  - *In vitro* stimulation of the Jurkat T cells
  - Expression of TCRα/β on primary T cells
  - Binding affinity measurements using BLI
  - *In vitro* cytotoxicity assay
  - Immunohistochemistry of TMEM161A
  - Whole-exome sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.immuni.2021.02.014.

ACKNOWLEDGMENTS

This work was supported by Parker Institute for Cancer Immunotherapy, Virginia and D.K. Ludwig Fund for Cancer Research, the Howard Hughes Medical Institute, NIH U54 CA232568-01, NIH 5P30CA124435 (C.L.M.), NIH 5RO1-AI03867, NIH U19 AI057229, and a St. Baldrick's-Stand Up to Cancer Dream Team Translational Research grant (SU2C-AACR-DT-27-17). Stand Up to Cancer is a division of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C. C.L.M. and M.M.D. are members of the Parker Institute for Cancer Immunotherapy, which provided partial funding for this project. D.T. was funded by a Lung Cancer Research Foundation scientific grant, Ellie Guardino Cancer Foundation Award from the Stanford Cancer Institute, a Stanford School of Medicine Honorary Dean's Fellowship, and a Conquer Cancer Foundation of ASCO Young Investigator Award. Any opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect those of the American Society of Clinical Oncology or the Conquer Cancer Foundation. This work was also funded in part by the National Cancer Institute of the National Institutes of Health Research Project grant (R01CA234629-01), the AACR-Johnson & Johnson Lung Cancer Innovation Science grant (18-90-52-ZHAN), the Cancer Prevention and Research Institute of Texas Multi-Investigator Research Award grant (RP160668), and the University of Texas Lung Specialized Programs of Research Excellence grant (grant number P50CA70907). Flow cytometry and cell-sorting services were provided by (1) the Stanford Shared FACS Facility supported by the NIH S10 Shared Instrument Grants (S10RR025518-01 and S10RR027431-01) and (2) the Flow Cytometry & Cell Sorting Core Facility, a shared resource of Rutgers - Robert Wood Johnson Medical School and the Rutgers Cancer Institute of New Jersey (P30CA072720-5921), and also supported by NIH Shared Instrumentation grant (1 S10 RR025468-01).

AUTHOR CONTRIBUTIONS

S.-H.C. and D.T. contributed equally. C.L.M. and M.M.D. are jointly supervising authors. M.M.D. and C.L.M. are both corresponding authors. S.-H.C., D.T., C.L.M., and M.M.D. conceived the project and wrote the manuscript. D.T., S.-H.C., A.R., V.M., I.S.M., P.T., and X.Y. performed experiments. S.-H.C. performed the bioinformatics. S.C. and R.S. helped with robotic-assisted SMART-seq2 data collection and analyses. X.Y. and D.N. assisted with yeast screens. C.W. wrote the GLIPH2 script. S.-H.C., D.T., K.C.G., E.S., I.L.W., C.L.M., and M.M.D. interpreted data. J.B.S., M.F.B., L.B., N.S.L., H.A.W., J.W.N., S.K.P., and J.A.B. contributed clinical samples. G.J.B. reviewed pathology. A.D. and P.H.S. assisted with immunohistochemistry data. S.-H.C., J.W., A.M.M., and D.M.L. assisted with scTCR-seq. R.R., L.L., D.D.K, and S.A.F. assisted with experimental design. B.Y.N. and M.D. contributed clinical samples and performed and interpreted exome-seq data. J.Z., A.R., I.I.W., J.V.H., and P.A.F. generated the TCR sequences, clinical data, and the exome-seq data of the MDACC NSCLC cohort. All authors reviewed and approved the final version of manuscript.

DECLARATION OF INTERESTS

C.L.M. is a founder of, holds equity in, and receives consulting fees from Lyell Immunopharma and receives consulting fees from NeoImmuneTech, Nektar, Apricity, and Roche. J.W.N. reports research support from Genentech/Roche, Merck, Novartis, Boehringer Ingelheim, Exelixis, Takeda Pharmaceuticals, Nektar Therapeutics, Adaptimmune, and GSK and has served in a consulting or advisory role for AstraZeneca, Genentech/Roche, Exelixis Inc., Jounce Therapeutics, Takeda Pharmaceuticals, Eli Lilly and Company, Calithera Biosciences, Amgen, Regeneron Pharmaceuticals, Natera, and Iovance Biotherapeutics. H.A.W. has received research support from Celgene, Clovis Oncology, Genentech/Roche, Arrys Therapeutics, Novartis, Merck, BMS, Exelixis, Lilly, Pfizer, and has participated on the advisory boards of Helsinn, Mirati, Cellworks, Genentech/Roche, Merck, and ITMIG. N.S.L. has received research funding from Intuitive Foundation and Auspex Diagnostics. E.S. is a consultant for Lyell Immunopharma. L.L. is a consultant for Lyell Immunopharma. S.A.F. is consulting for Lonza PerMed and Samsara BioCapital. M.D. reports research funding from Varian Medical Systems and Illumina; ownership interest in CiberMed and Foresight Diagnostics; patent filings related to cancer biomarkers; paid consultancy from Roche, AstraZeneca, BioNTech, Genentech, Novartis, and Gritstone Oncology; and travel/honoraria from Reflexion. K.C.G. is founder of 3T therapeutics. I.I.W. has received honoraria from Genentech/Roche, Bayer, Bristol-Myers Squibb, AstraZeneca/Medimmune, Pfizer, HTG Molecular, Asuragen, Merck, GlaxoSmithKline, Guardant Health, Oncocyte, and MSD. I.I.W. is also supported by Genentech, Oncoplex, HTG Molecular, DepArray, Merck, Bristol-Myers Squibb, Medimmune, Adaptive, Adaptimmune, EMD Serono, Pfizer, Takeda, Amgen, Karus, Johnson &

Johnson, Bayer, Iovance, 4D, Novartis, and Akoya. J.Z. reports grants from Merck and Johnson & Johnson, as well as adversary/consulting/Hornoraria fees from Bristol Myers Squibb, AstraZeneca, GenePlus, Innovent, OrigMed, and Roche outside the submitted work. This study was supported in part by a Cancer Prevention Research Institute of Texas Multi-Investigator Research Award (grant number RP160668) and the University of Texas Lung Specialized Programs of Research Excellence grant (grant number P50CA70907). S.-H.C., D.T., C.L.M., and M.M.D have a patent related to this work.

Received: July 17, 2020
Revised: December 8, 2020
Accepted: February 11, 2021
Published: March 9, 2021

REFERENCES

Altman, J.D., and Davis, M.M. (2003). MHC-peptide tetramers to visualize antigen-specific T cells. Curr Protoc Immunol. *Chapter 17*, Unit 17.13.

Anagnostou, V., Forde, P.M., White, J.R., Niknafs, N., Hruban, C., Naidoo, J., Marrone, K., Sivakumar, I.K.A., Bruhm, D.C., Rosner, S., et al. (2019). Dynamics of Tumor and Immune Responses during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. Cancer Res. *79*, 1214–1225.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andersen, R.S., Thrue, C.A., Junker, N., Lyngaa, R., Donia, M., Ellebæk, E., Svane, I.M., Schumacher, T.N., Thor Straten, P., and Hadrup, S.R. (2012). Dissection of T-cell antigen specificity in human melanoma. Cancer Res. *72*, 1642–1650.

Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. (1999). A direct estimate of the human alphabeta T cell receptor diversity. Science *286*, 958–961.

Bessell, C.A., Isser, A., Havel, J.J., Lee, S., Bell, D.R., Hickey, J.W., Chaisawangwong, W., Glick Bieler, J., Srivastava, R., Kuo, F., et al. (2020). Commensal bacteria stimulate antitumor responses via T cell cross-reactivity. JCI Insight *5*, e135597.

Cameron, S.J.S., Lewis, K.E., Huws, S.A., Hegarty, M.J., Lewis, P.D., Pachebat, J.A., and Mur, L.A.J. (2017). A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. PLoS ONE *12*, e0177062.

Choi, I.-K., Wang, Z., Ke, Q., Hong, M., Paul, D.W., Jr., Fernandes, S.M., Hu, Z., Stevens, J., Guleria, I., Kim, H.J., et al. (2020). Mechanism of EBV inducing anti-tumour immunity and its therapeutic use. Nature *590*, 157–162.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Corridoni, D., Antanaviciute, A., Gupta, T., Fawkner-Corbett, D., Aulicino, A., Jagielowicz, M., Parikh, K., Repapi, E., Taylor, S., Ishikawa, D., et al. (2020). Single-cell atlas of colonic $CD8^+$ T cells in ulcerative colitis. Nat. Med. *26*, 1480–1490.

Coulie, P.G., Brichard, V., Van Pel, A., Wölfel, T., Schneider, J., Traversari, C., Mattei, S., De Plaen, E., Lurquin, C., Szikora, J.P., et al. (1994). A new gene coding for a differentiation antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. J. Exp. Med. *180*, 35–42.

Coulie, P.G., Lehmann, F., Lethé, B., Herman, J., Lurquin, C., Andrawiss, M., and Boon, T. (1995). A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T lymphocytes on a human melanoma. Proc. Natl. Acad. Sci. USA *92*, 7976–7980.

Dickson, R.P., Erb-Downward, J.R., Martinez, F.J., and Huffnagle, G.B. (2016). The Microbiome and the Respiratory Tract. Annu. Rev. Physiol. *78*, 481–504.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Emerson, R.O., DeWitt, W.S., Vignali, M., Gravley, J., Hu, J.K., Osborne, E.J., Desmarais, C., Klinger, M., Carlson, C.S., Hansen, J.A., et al. (2017). Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat. Genet. *49*, 659–665.

Fluckiger, A., Daillère, R., Sassi, M., Sixt, B.S., Liu, P., Loos, F., Richard, C., Rabu, C., Alou, M.T., Goubet, A.G., et al. (2020). Cross-reactivity between tumor MHC class I-restricted antigens and an enterococcal bacteriophage. Science *369*, 936–942.

Gee, M.H., Han, A., Lofgren, S.M., Beausang, J.F., Mendoza, J.L., Birnbaum, M.E., Bethune, M.T., 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.e16.

Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., et al. (2017). Identifying specificity groups in the T cell receptor repertoire. Nature *547*, 94–98.

Gopalakrishnan, V., Spencer, C.N., Nezi, L., Reuben, A., Andrews, M.C., Karpinets, T.V., Prieto, P.A., Vicente, D., Hoffman, K., Wei, S.C., et al. (2018). Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science *359*, 97–103.

Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., Kang, B., Liu, Z., Jin, L., Xing, R., et al. (2018). Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat. Med. *24*, 978–985.

Han, A., Glanville, J., Hansmann, L., and Davis, M.M. (2014). Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat. Biotechnol. *32*, 684–692.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7.

Harjanto, S., Ng, L.F., and Tong, J.C. (2014). Clustering HLA class I superfamilies using structural interaction patterns. PLoS ONE *9*, e86655.

Hellmann, M.D., Nabet, B.Y., Rizvi, H., Chaudhuri, A.A., Wells, D.K., Dunphy, M.P.S., Chabon, J.J., Liu, C.L., Hui, A.B., Arbour, K.C., et al. (2020). Circulating tumor DNA analysis to assess risk of progression after long-term response to PD-(L)1 blockade in NSCLC. Clin. Cancer Res. *26*, 2849–2858.

Huang, H., Wang, C., Rubelt, F., Scriba, T.J., and Davis, M.M. (2020). Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. Nat. Biotechnol. *38*, 1194–1202.

Joglekar, A.V., Leonard, M.T., Jeppson, J.D., Swift, M., Li, G., Wong, S., Peng, S., Zaretsky, J.M., Heath, J.R., Ribas, A., et al. (2019). T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. Nat. Methods *16*, 191–198.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Res. *36*, W5-9.

Joshi, K., de Massy, M.R., Ismail, M., Reading, J.L., Uddin, I., Woolston, A., Hatipoglu, E., Oakes, T., Rosenthal, R., Peacock, T., et al.; TRACERx consortium (2019). Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. Nat. Med. *25*, 1549–1559.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J. Immunol. *199*, 3360–3368.

Kawakami, Y., Eliyahu, S., Delgado, C.H., Robbins, P.F., Rivoltini, L., Topalian, S.L., Miki, T., and Rosenberg, S.A. (1994). Cloning of the gene coding for a shared human melanoma antigen recognized by autologous T cells infiltrating into tumor. Proc. Natl. Acad. Sci. USA *91*, 3515–3519.

Kheir, F., Zhao, M., Strong, M.J., Yu, Y., Nanbo, A., Flemington, E.K., Morris, G.F., Reiss, K., Li, L., and Lin, Z. (2019). Detection of Epstein-Barr Virus Infection in Non-Small Cell Lung Cancer. Cancers (Basel) *11*, 759.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576.

Koziel, M.J., Dudley, D., Afdhal, N., Grakoui, A., Rice, C.M., Choo, Q.L., Houghton, M., and Walker, B.D. (1995). HLA class I-restricted cytotoxic T

lymphocytes specific for hepatitis C virus. Identification of multiple epitopes and characterization of patterns of cytokine release. J. Clin. Invest. *96*, 2311–2321.

Kula, T., Dezfulian, M.H., Wang, C.I., Abdelfattah, N.S., Hartman, Z.C., Wucherpfennig, K.W., Lyerly, H.K., and Elledge, S.J. (2019). T-Scan: a genome-wide method for the systematic discovery of T cell epitopes. Cell *178*, 1016–1028.e13.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, G., Bethune, M.T., Wong, S., Joglekar, A.V., Leonard, M.T., Wang, J.K., Kim, J.T., Cheng, D., Peng, S., Zaretsky, J.M., et al. (2019). T cell antigen discovery via trogocytosis. Nat. Methods *16*, 183–190.

Matson, V., Fessler, J., Bao, R., Chongsuwat, T., Zha, Y., Alegre, M.L., Luke, J.J., and Gajewski, T.F. (2018). The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science *359*, 104–108.

McCarthy, E.F. (2006). The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. Iowa Orthop. J. *26*, 154–158.

Mogilenko, D.A., Shpynov, O., Andhey, P.S., Arthur, L., Swain, A., Esaulova, E., Brioschi, S., Shchukina, I., Kerndl, M., Bambouskova, M., et al. (2020). Comprehensive Profiling of an Aging Immune System Reveals Clonal GZMK(+) CD8 (T Cells as Conserved Hallmark of Inflammaging. Immunity).

Murray, R.J., Kurilla, M.G., Brooks, J.M., Thomas, W.A., Rowe, M., Kieff, E., and Rickinson, A.B. (1992). Identification of target antigens for the human cytotoxic T cell response to Epstein-Barr virus (EBV): implications for the immune control of EBV-positive malignancies. J. Exp. Med. *176*, 157–168.

Newman, A.M., Bratman, S.V., To, J., Wynne, J.F., Eclov, N.C., Modlin, L.A., Liu, C.L., Neal, J.W., Wakelee, H.A., Merritt, R.E., et al. (2014). An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat. Med. *20*, 548–554.

Newman, J.H., Chesson, C.B., Herzog, N.L., Bommareddy, P.K., Aspromonte, S.M., Pepe, R., Estupinian, R., Aboelatta, M.M., Buddhadev, S., Tarabichi, S., et al. (2020). Intratumoral injection of the seasonal flu shot converts immunologically cold tumors to hot and serves as an immunotherapy for cancer. Proc. Natl. Acad. Sci. USA *117*, 1119–1128.

Oakes, T., Heather, J.M., Best, K., Byng-Maddick, R., Husovsky, C., Ismail, M., Joshi, K., Maxwell, G., Noursadeghi, M., Riddell, N., et al. (2017). Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile. Front. Immunol. *8*, 1267.

Ohashi, P.S., Oehen, S., Buerki, K., Pircher, H., Ohashi, C.T., Odermatt, B., Malissen, B., Zinkernagel, R.M., and Hengartner, H. (1991). Ablation of "tolerance" and induction of diabetes by virus infection in viral antigen transgenic mice. Cell *65*, 305–317.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Rehermann, B., Fowler, P., Sidney, J., Person, J., Redeker, A., Brown, M., Moss, B., Sette, A., and Chisari, F.V. (1995). The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. J. Exp. Med. *181*, 1047–1058.

Reuben, A., Zhang, J., Chiou, S.H., Gittelman, R.M., Li, J., Lee, W.C., Fujimoto, J., Behrens, C., Liu, X., Wang, F., et al. (2020). Comprehensive T cell repertoire characterization of non-small cell lung cancer. Nat. Commun. *11*, 603.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. *48* (W1), W449–W454.

Riquelme, E., Zhang, Y., Zhang, L., Montiel, M., Zoltan, M., Dong, W., Quesada, P., Sahin, I., Chandra, V., San Lucas, A., et al. (2019). Tumor microbiome diversity and composition influence pancreatic cancer outcomes. Cell *178*, 795–806.e12.

Robins, H.S., Srivastava, S.K., Campregher, P.V., Turtle, C.J., Andriesen, J., Riddell, S.R., Carlson, C.S., and Warren, E.H. (2010). Overlap and effective size of the human CD8+ T cell receptor repertoire. Sci. Transl. Med. *2*, 47ra64.

Röcken, M., Urban, J.F., and Shevach, E.M. (1992). Infection breaks T-cell tolerance. Nature *359*, 79–82.

Rosato, P.C., Wijeyesinghe, S., Stolley, J.M., Nelson, C.E., Davis, R.L., Manlove, L.S., Pennell, C.A., Blazar, B.R., Chen, C.C., Geller, M.A., et al. (2019). Virus-specific memory T cells populate tumors and can be repurposed for tumor immunotherapy. Nat. Commun. *10*, 567.

Routy, B., Le Chatelier, E., Derosa, L., Duong, C.P.M., Alou, M.T., Daillère, R., Fluckiger, A., Messaoudene, M., Rauber, C., Roberti, M.P., et al. (2018). Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science *359*, 91–97.

Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics *28*, 1811–1817.

Savage, P.A., Vosseller, K., Kang, C., Larimore, K., Riedel, E., Wojnoonski, K., Jungbluth, A.A., and Allison, J.P. (2008). Recognition of a ubiquitous self antigen by prostate cancer-infiltrating CD8+ T lymphocytes. Science *319*, 215–220.

Scheper, W., Kelderman, S., Fanchi, L.F., Linnemann, C., Bendle, G., de Rooij, M.A.J., Hirt, C., Mezzadra, R., Slagter, M., Dijkstra, K., et al. (2019). Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. Nat. Med. *25*, 89–94.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Sewell, A.K. (2012). Why must T cells be cross-reactive? Nat. Rev. Immunol. *12*, 669–677.

Sharma, P., and Allison, J.P. (2020). Dissecting the mechanisms of immune checkpoint therapy. Nat. Rev. Immunol. *20*, 75–76.

Shugay, M., Bagaev, D.V., Zvyagin, I.V., Vroomans, R.M., Crawford, J.C., Dolton, G., Komech, E.A., Sycheva, A.L., Koneva, A.E., Egorov, E.S., et al. (2018). VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res. *46* (D1), D419–D427.

Sibener, L.V., Fernandes, R.A., Kolawole, E.M., Carbone, C.B., Liu, F., McAffee, D., Birnbaum, M.E., Yang, X., Su, L.F., Yu, W., et al. (2018). Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. Cell *174*, 672–687.e27.

Sidney, J., Peters, B., Frahm, N., Brander, C., and Sette, A. (2008). HLA class I supertypes: a revised and updated classification. BMC Immunol. *9*, 1.

Simoni, Y., Becht, E., Fehlings, M., Loh, C.Y., Koo, S.L., Teng, K.W.W., Yeong, J.P.S., Nahar, R., Zhang, T., Kared, H., et al. (2018). Bystander $CD8^+$ T cells are abundant and phenotypically distinct in human tumour infiltrates. Nature *557*, 575–579.

Sivan, A., Corrales, L., Hubert, N., Williams, J.B., Aquino-Michaels, K., Earley, Z.M., Benyamin, F.W., Lei, Y.M., Jabri, B., Alegre, M.L., et al. (2015). Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science *350*, 1084–1089.

Song, I., Gil, A., Mishra, R., Ghersi, D., Selin, L.K., and Stern, L.J. (2017). Broad TCR repertoire and diverse structural solutions for recognition of an immunodominant $CD8^+$ T cell epitope. Nat. Struct. Mol. Biol. *24*, 395–406.

Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat. Methods *13*, 329–332.

Su, L.F., Kidd, B.A., Han, A., Kotzin, J.J., and Davis, M.M. (2013). Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. Immunity *38*, 373–383.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and

Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. *47* (D1), D506–D515.

van der Bruggen, P., Traversari, C., Chomez, P., Lurquin, C., De Plaen, E., Van den Eynde, B., Knuth, A., and Boon, T. (1991). A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. Science *254*, 1643–1647.

Vétizou, M., Pitt, J.M., Daillère, R., Lepage, P., Waldschmitt, N., Flament, C., Rusakiewicz, S., Routy, B., Roberti, M.P., Duong, C.P., et al. (2015). Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science *350*, 1079–1084.

Witwicka, H., Hwang, S.Y., Reyes-Gutierrez, P., Jia, H., Odgren, P.E., Donahue, L.R., Birnbaum, M.J., and Odgren, P.R. (2015). Studies of OC-STAMP in Osteoclast Fusion: A New Knockout Mouse Model, Rescue of Cell Fusion, and Transmembrane Topology. PLoS ONE *10*, e0128275.

Wölfel, T., Hauer, M., Schneider, J., Serrano, M., Wölfel, C., Klehmann-Hieb, E., De Plaen, E., Hankeln, T., Meyer zum Büschenfelde, K.H., and Beach, D. (1995). A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. Science *269*, 1281–1284.

Yu, W., Jiang, N., Ebert, P.J., Kidd, B.A., Müller, S., Lund, P.J., Juang, J., Adachi, K., Tse, T., Birnbaum, M.E., et al. (2015). Clonal Deletion Prunes but Does Not Eliminate Self-Specific αβ CD8(+) T Lymphocytes. Immunity *42*, 929–941.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CD4 antibody	Biolegend	Clone: OKT4
Anti-CD4 antibody	Biolegend	Clone: RPA-T4
Anti-CD8 antibody	Biolegend	Clone: SK1
Anti-CD8 antibody	Biolegend	Clone: HIT8a
Anti-CD3 antibody	Biolegend	Clone: OKT3
Anti-CD3 antibody	Biolegend	Clone: UCHT1
Anti-CD45 antibody	Biolegend	Clone: H130
Anti-CD25 antibody	Biolegend	Clone: BC96
Anti-PD-1 antibody	Biolegend	Clone: EH12.2H7
Anti-CD137 antibody	Biolegend	Clone: 4B4-1
Anti-HLA-DR antibody	Biolegend	Clone: L243
Anti-HLA-BC antibody	Thermo Fisher Scientific	Clone: B1.23.2
Anti-TCR$\gamma\delta$ antibody	Biolegend	Clone: B1
Anti-TCRα/β	Biolegend	Clone: IP26
Anti-CD19 antibody	Biolegend	Clone: H1B19
Anti-CD14 antibody	Biolegend	Clone: M5E2
Anti-CD38 antibody	Biolegend	Clone: HIT2
Anti-CD69 antibody	Biolegend	Clone: FN50
Anti-APC microbeads	Miltenyi Biotech	Cat#: 130-090-855
Anti-CD3/CD28 microbeads	Thermo Fisher Scientific	Cat#: 11141D
Anti-TMEM161A antibody	abcam	Clone: EPR14369
Chemicals, Peptides, and Recombinant Proteins		
Collagenase III	Worthington Biochemical	Cat#: LS004182
DNase I	Worthington Biochemical	Cat#: LS002007
Zombie Aqua	Biolegend	Cat#: 423102
Human TruStain FcX	Biolegend	Cat#: 422302
Live/dead near-IR dye	Thermo Fisher Scientific	Cat#: L34975
AMPure XP beads	Beckman Coulter	Cat#: A63881
Recombinant hIL-2	Peprotech	Cat#: 200-02
Recombinant RNase Inhibitor	Takara Bio	Cat#: 2313A
ERCC RNA Spike-In Mix	Ambion/Life Technologies	Cat#: 4456740
LNA-TSO	Exiqon	Cat#: 500100
Ni-NTA resin	QIAGEN	Cat#: 30210
BirA Biotin-protein ligase	Avidity	Cat#: BirA500
Streptavidin MicroBeads	Miltenyi Biotech	Cat#: 130-048-101
LS Columns	Miltenyi Biotech	Cat#: 130-042-401
Polyethylenimine (PEI)	Millipore-Sigma	Cat#: 408727
Opti-MEM	Thermo Fisher Scientific	Cat#: 31985062
RetroNectin® Recombinant Human Fibronectin Fragment	Takara Bio	Cat#: T100A
FuGENE® 6	Promega	Cat#: E2691
Amicon® Ultra-15 Centrifugal Filter Unit (30 kDa filter)	Millipore-Sigma	Cat#: UFC903024

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Streptavidin	Thermo Fisher Scientific	Cat#: 434302
Custom synthetic peptides	Alan Scientific	N/A
Flex-T HLA-A*02:02 Monomer UVX	Biolegend	Cat#: 280003
Betaine	Millipore-Sigma	Cat#: W422312
Critical Commercial Assays		
SMARTScribe Reverse Transcriptase kit	Takara Bio	Cat#: 639538
KAPA Library Quantification kit	Roche	Cat#: KK2602
AATI Fragment Analyzer	Agilent	Cat#: DNF-474-1000
Nextera XT DNA Library Preparation Kit	Illumina	Cat#: FC-131-1096
MiSeq Reagent Kit v2 (300-cycles; yeast screen)	Illumina	Cat#: MS-102-2002
MiSeq Reagent Kit v2 (500-cycles; scTCR-seq)	Illumina	Cat#: MS-102-2003
Chromium Single-Cell V(D)J kit (for TCR)	10x Genomics	Cat#: 1000005; 1000009; 120262; 1000084; 1000080; 1000014; 1000020
Immunoseq assay (Deep)	Adaptive Biotechnologies	Cat#: hsTCRB
RosetteSep human T cell enrichment cocktail	Stem Cell Technologies	Cat#: 15061
KAPA HyperPrep Kit	Roche	Cat#: KK8502
SeqCap EZ MedExome Enrichment Kit	Roche	Cat#: 07676581001
Gibson Assembly Cloning Kit	NEB	Cat#: E5510S
Zymoprep II kit	Zymo Research	Cat#: D2004
In-Fusion Cloning	Takara Bio	Cat#: 638947
Deposited Data		
scRNA-seq data (tumor-infiltrating T cells from 10 Stanford lung cancer patients)	This paper	Database: GSE151537 (SuperSeries #GSE151538)
scRNA-seq data (tetramer-sorted T cells from peripheral blood)	This paper	Database: GSE151531 (SuperSeries #GSE151538)
Human reference genome NCBI build 37, GRCh38	Genome Reference Consortium	https://genome.ucsc.edu/cgi-bin/hgTables
Bulk CDR3β sequences (n = 178 HLA-typed NSCLC patients, MDACC)	(Reuben et al., 2020); IMMUNOSEQ ANALYZER (Adaptive Biotechnologies)	https://clients.adaptivebiotech.com
HLA tetramer-derived CDR3β sequences	(Shugay et al., 2018)	https://vdjdb.cdr3.net
Reference CDR3β sequences for GLIPH2	(Huang et al., 2020)	http://50.255.35.37:8080/tools
Bulk RNA-seq data from tumors (pan-lung cancer)	The Cancer Genome Atlas	https://portal.gdc.cancer.gov
GSEA hallmark gene sets	Broad Institute	http://www.gsea-msigdb.org/gsea/downloads.jsp
The UniProt Archive (UniParc)	The UniProt consortium	https://www.uniprot.org/downloads
Emerson CDR3β dataset	(Emerson et al., 2017)	https://clients.adaptivebiotech.com
TRACERx CDR3β sequences	(Joshi et al., 2019)	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA544699
Experimental Models: Cell Lines		
Jurkat 76 cells	S.-A. Xue, University of College London	N/A
T2 cells	ATCC	Cat#: CRL-1992; RRID: CVCL_2211
HLA-A*02^{+} 293T cells	S. Feldman, Stanford University	N/A
293T HEK cells	ATCC	Cat#: CRL-11268
H1395 cells	ATCC	Cat#: CRL-5868; RRID: CVCL_1467
Sf9 cells	ATCC	Cat#: CRL-1711
Hi5 cells	Thermo Fisher Scientific	Cat#: BTI-TN-5B1-4

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
Custom gBlocks dsDNA fragments	IDT	N/A
pAcGP67a vector	BD Biosciences	N/A
Bestbac 2.0	Expression systems	Cat#: 91-002
Yeast A*02 display library constructs	K.C. Garcia, Stanford University	N/A
Soluble TCR baculoviral constructs	K.C. Garcia, Stanford University	N/A
EF1a-MCS-GFP-PGK-puro lentiviral vector	(Witwicka et al., 2015)	Addgene#: 73582
MSGV1 retroviral vector	S. Rosenberg, NIH	N/A
gag-pol plasmid (Δ8.9)	M.M. Winslow, Stanford University	N/A
pMD.G plasmid (VSV-G)	M.M. Winslow, Stanford University	N/A
Lenti-TMEM161A plasmid	GeneCopoeia	Cat#: EX-A1961-Lv241
Software and Algorithms		
GLIPH2 algorithm	(Huang et al., 2020)	http://50.255.35.37:8080
R version 4.0.2	CRAN	https://www.r-project.org/; RRID: SCR_001905
star/2.7.1a	(Dobin et al., 2013)	https://github.com/alexdobin/STAR; RRID: SCR_015899
samtools/1.4	(Li et al., 2009)	http://www.htslib.org; RRID: SCR_002105
python/2.7.3	Python Software Foundation	https://www.python.org; RRID: SCR_008394
htseq-count (HTSeq 0.5.4p5)	(Anders et al., 2015)	https://htseq.readthedocs.io/; RRID: SCR_011867
GSEA v2.2.2	Broad Institute; (Subramanian et al., 2005)	RRID: SCR_003199
GSVA/1.34.0	(Hänzelmann et al., 2013)	https://www.bioconductor.org/packages/release/bioc/html/GSVA.html
FIJI/2.0.0-rc-69/1.52p	(Schindelin et al., 2012)	https://imagej.net/Fiji; RRID: SCR_002285
TraCeR algorithm	(Stubbington et al., 2016)	https://github.com/Teichlab/tracer
HighV-QUEST	international ImMunoGeneTics information system (IMGT)	http://www.imgt.org/; RRID: SCR_018196
FlowJo software	FlowJo, LLC	https://www.flowjo.com; RRID: SCR_008520
varscan2/2.4.3	(Koboldt et al., 2012)	http://varscan.sourceforge.net; RRID: SCR_006849
gatk-3.7/MuTect2	Broad Institute; (Cibulskis et al., 2013)	RRID: SCR_000559
Strelka/2.9.10	Illumina; (Saunders et al., 2012)	RRID: SCR_005109
Seurat/3.1.4	Bioconductor	RRID: SCR_016341
Monocle/2.10.1	Bioconductor	RRID: SCR_018685
netMHCpan/4.0	(Jurtz et al., 2017; Reynisson et al., 2020)	RRID: SCR_018182
BLASTP	National Center for Biotechnology Information (NCBI); (Johnson et al., 2008)	RRID: SCR_001010

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Mark Davis (mmdavis@stanford.edu).

Materials availability

Further information and material requests should be addressed to Mark Davis (mmdavis@stanford.edu).

Data and code availability

The scRNA-Seq data from tumor-infiltrating T cells (n = 2950, GEO: GSE151537) and HLA tetramer-sorted peripheral blood T cells (n = 623, GEO: GSE151531) were deposited in the GEO database (SuperSeries accession number: GSE151538). The algorithm

GLIPH2, reference CDR3β sequences, and tutorial are available from the following link: http://50.255.35.37:8080 (Huang et al., 2020).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Protocols for collection of human tissue and blood were approved by the Stanford Institutional Review Board (IRB 15166). Inclusion criteria included adult patients (age > = 18 years), known or suspected diagnosis of NSCLC, primary tumor > 2 cm, and consent for research. Patients receiving neoadjuvant therapy or patients with underlying lung infection, inflammatory, or fibrotic disease were excluded. Overall, 21 patients with surgically-resectable NSCLC treated at Stanford were included in this study. A table of patient characteristics is provided (Table S4). DNA was extracted from peripheral blood PBMC (QIAGEN) for HLA tying. In addition, we analyzed samples from 2 patients with advanced/metastatic disease treated with anti-PD1 antibody on IRB 21319. Patients experienced clinical benefit at 6 months after initiation of treatment.

METHOD DETAILS

Tissue processing

Tissue was processed within 2 h from surgery. Tissue was divided and one section for cell suspensions and another section for histology. Cell suspensions were generated by mincing of tissue followed by digestion with collagenase III (200 IU/mL) and DNase I (100 U/mL) (Worthington Biochemical) for 40 min in RPMI and passing through a 70-um filter. Sections for histology were fixed in 4% paraformaldehyde and transferred to 70% ethanol solution the following day.

FACS analyses

T cells were isolated from tumor single cell suspensions by antibody staining followed by cell sorting on a 5-laser FACSAria Fusion sorter (Stanford FACS Facility) purchased using funds from the Parker Institute for Cancer Immunotherapy. Tumor cell suspensions were stained in PBS with Zombie Aqua dye (Biolegend) for viability assessment. This was followed by staining in PBS with 2% FBS in Fc Blocking solution (Biolegend) plus the following antibodies: anti-CD4 (OKT4, Biolegend), anti-CD8 (SK1, Biolegend), anti-CD3 (OKT3, Biolegend), anti-CD45 (H130, Biolegend), anti-CD25 (BC96, Biolegend), anti-PD-1(EH12.2H7, Biolegend), anti-CD137 (4B4-1, BD Biosciences), anti-HLA-DR (L243, Biolegend). $CD3^{+}CD45^{+}AquaZombie^{-}$ cells were index sorted directly into 96-well plates preloaded with 4 uL of capture buffer, snap frozen on dry ice, and stored at −80°C. Ectopic HLA-B*35 was detected with anti-HLA-BC monoclonal antibody (clone B1.23.2, Thermo Fisher Scientific). Transduced Jurkat 76 cells expressing exogenous TCRα/β chains were sorted on a FACSAria Fusion sorter at Stanford or a BD Biosciences Influx High Speed Cell Sorter at the Flow Cytometry Core Facility of the Cancer Institute of New Jersey.

Establishment of T cell specificity groups

The GLIPH2 algorithm was implemented for the establishment of T cell specificity groups using 778,938 distinct CDR3β sequences from the MD Anderson NSCLC dataset (Reuben et al., 2020). Briefly, by comparing with the reference dataset of 273,920 distinct CDR3β sequences (both CD4 and CD8) from 12 healthy individuals, GLIPH2 first discovered clusters of CDR3β sequences sharing either global or local motifs as previously described (Huang et al., 2020). The output of CDR3β clusters with shared sequence motifs is accompanied by multiple statistical measurements to facilitate the calling of high-confidence specificity groups, including biases in *Vβ* gene usage, CDR3β length distribution (relevant only for local motifs), cluster size, HLA allele usage, and clonal expansion. To establish high-confidence specificity groups with the NSCLC dataset, we prioritized TCR specificity groups with at least 3 distinct CDR3β members from a minimum of 3 different patients with significant biases in Vβ gene usage, and CDR3β clonal expansion in comparison with the reference dataset. This led to the discovery of 4,226 specificity groups that formed the basis for further analyses throughout the study.

Classification of TCRs and specificity groups

For CDR3β clonotypes, we included only distinct sequences from each MDACC patient with frequencies above 0.1% in tumors or adjacent lung samples in order to focus on the most expanded TCRs. In Figures 2A and 2B, we compared the abundance (rounded, normalized count) of each distinct TCR in the tumor versus the paired adjacent lung from the same patient. The *p* value for the comparison in abundance between tumor and the adjacent lung were calculated with the poisson.test function in R (alternative = ''two.-sided''). For specificity groups with clonal expansion (n = 4,226), a list of summed frequencies (up to 100%, rounded to integers) of all CDR3β members that belong to each specificity group was first created for both tumor and the adjacent lung from each MDACC patient. Poisson test was then used to calculate the *p* value for the comparison of these summed frequencies in the lists using the poisson.test function (Figures 2A and 2B).

Annotation of specificity groups

To annotate inferred specificity groups from lung cancer patients, we ran a combined GLIPH analysis using both the MD Anderson lung cancer patient CDR3β sequences and publicly available, tetramer-derived CDR3β sequences (Glanville et al., 2017; Shugay et al., 2018; Song et al., 2017). To do so, we first identified tetramer-derived CDR3β sequences that could form TCR specificity groups by running an independent GLIPH analysis with a total 10,051 CDR3β sequences from the tetramer datasets. This led to the formation

of 395 specificity groups containing 1,561 CDR3β sequences. We then combined these 1,561 CDR3β sequences with the 778,938 CDR3β sequences from the MD Anderson lung cancer dataset for the aforementioned GLIPH2 analysis. Any specificity group that includes at least one CDR3β sequence from the tetramer data is considered "annotated" and would be assigned a specificity and HLA restriction according to the associated tetramer sequence(s). Of note, in all cases where multiple tetramer-derived CDR3β sequences were found in a given specificity group, there was only one dominant tetramer-defined specificity/HLA involved.

Validation of HLA restriction inference

For the tetramer-annotated specificity groups mentioned above (n = 71), we validated the inferences of HLA restriction made by the GLIPH2 algorithm against the HLA restriction informed by tetramers. Specificity groups annotated with HLA-A*02 (n = 49 out of 71) or HLA-B*08 (n = 8 out of 71) tetramers were chosen for the validation because they were the most prevalent. To validate a specificity group for enrichment with *HLA-A*02* alleles, we first constructed a contingency table with the number of patients in the specificity group carrying *HLA-A*02* supertype allele(s) and the number of patients without these alleles, number of all NSCLC patients carrying *HLA-A*02* supertype allele(s) (n = 79) and those who do not (n = 98). We then calculated *p* values using the hypergeometric test (phyper in R, lower.tail = FALSE). We reported the numbers of specificity groups significantly enriched with *HLA-A*02* supertype alleles ($p < 0.05$ by the hypergeometric test) as a fraction over the number of specific groups annotated with HLA-A*02 tetramers (n = 18 out of 49). We also reported the numbers of specificity groups significantly enriched with *HLA-A*02* supertype alleles as a fraction over the number of specificity groups annotated with non-HLA-A*02 tetramers (n = 0 out of 22). We repeated this process for the validation of specificity groups enriched with *HLA-B*08* supertype alleles. To identify top-enriched HLA allele(s) for a specificity group (Figures 2C and 2D), hypergeometric test was used to first uncover HLA allele(s) that are significantly enriched (phyper, lower.tail = FALSE). The highest value of fraction (# of patients carrying the allele within a specificity group / all patients within a specificity group) was determined and used to find top-enriched allele(s) with both *p* value < 0.05 and the highest fraction value.

*HLA-A*02:01* specificity group bootstrapping

To estimate the number of *HLA-A*02:01*$^+$ NSCLC patients needed to cover 50% of all *HLA-A*02:01*-enriched specificity groups (n = 77), we carried out a bootstrapping process through random sampling of patients with incremental sampling sizes. First, we established 77 specificity groups (from the 4,226 NSCLC-enriched specificity groups) that were significantly enriched with the *HLA-A*02:01* allele ($p < 0.05$). Bootstrapping was conducted with random sampling (with replacement) of 1 through 160 patients for 100 times. For each sampling event, we tallied the sum of *HLA-A*02:01*-enriched specificity groups found using the CDR3β sequences from the sampled patients (specificity count, Figures 1F and 1G). We then calculated the mean and the standard error of the specificity counts from the bootstrapping process. As an internal control, we repeated the bootstrapping process on the rest of *HLA-A*02:01*$^-$ NSCLC patients. To compare with specificity groups from a healthy cohort, we used 989,816 distinct CDR3β sequences from 304 *HLA-A*02:01*$^+$ and 1,153,600 CDR3β sequences from 362 *HLA-A*02*$^-$ healthy donors' PBMC from a publicly available dataset (Emerson dataset, (Emerson et al., 2017)). To adjust for the differences in sequencing depth (below), 5000 distinct CDR3β sequences (with the highest frequencies) from each healthy donor were included for the GLIPH analysis. To address the influence of clonal expansion on specificity group quantification, we compared the bootstrapping results with the aforementioned *HLA-A*02:01*-enriched specificity groups to an equal number of *HLA-A*02:01*-enriched specificity groups without clonal expansion (n = 77). We used a similar strategy to address how the total number of specificity groups impacted this result. We performed bootstrapping using various enrichment cutoffs for HLA-A*02 enrichment ($p < 0.05$, n = 1,267; $p < 0.025$, n = 319; $p < 0.01$, n = 71 specificity groups). Finally, to address the impact of sequencing depth on specificity group quantification, we down-sampled the total input CDR3β sequences randomly in the bootstrapping process by the indicated proportions (50%, 25%, 12.5%, or 0% down-sampled).

GSEA analysis of the TCGA data

Normalized gene expression data from bulk RNA-Seq analyses of human NSCLC resected tumors and adjacent lungs from the Cancer Genome Atlas (TCGA) were downloaded from the NCI GDC Legacy Archive (n = 1,017 for tumors and n = 110 for adjacent lungs). To conduct gene set enrichment analysis (GSEA) with the TCGA dataset, we first calculated the correlation coefficients between any gene and *TMEM161A* using the Pearson correlation. The sorted gene list based on the correlation coefficient with *TMEM161A* gene expression was then used for GSEA with the Preranked tool (v2.2.2, Broad Institute) and all hallmark gene sets (Subramanian et al., 2005). The signature scores were derived using the gene lists of indicated hallmark signatures with the single-sample GSEA (ssGSEA) method as described previously (Hänzelmann et al., 2013).

FACS sorting of antigen-specific CD8 T cells

Recombinant HLA-A*02 monomer with UV exchangeable peptide were either synthesized as previously described (Altman and Davis, 2003) or purchased commercially (Biolegend). UV peptide exchange was performed over 20 min with 1 mM of peptide in PBS using Strategene UV Stratalinker 2400. Streptavidin conjugated fluorophore was added incrementally the following day for a final 4:1 molar ratio of MHC:streptavidin. Tetramer staining was performed in PBS plus 2% FBS in Fc Blocking solution (Biolegend) at room temperature for 1 h. For peripheral blood samples, cells were subsequently stained with anti-TCRγδ (B1, Biolegend), anti-CD19 (H1B19, Biolegend), anti-CD14 (M5E2, Biolegend), anti-CD3 (OKT3, Biolegend), anti-CD4 (RPA-T4, Biolegend), anti-CD8 (HIT8a, Biolegend), and live/dead near-IR dye (Invitrogen). For tumor samples, cells were stained with anti-CD4 (OKT4, Biolegend), anti-CD8 (HIT8a, Biolegend), anti-CD3 (UCHT1, Biolegend), anti-CD45 (H130, Biolegend).

Single-cell RNA-seq (scRNA-Seq)

Full transcriptomes from FACS sorted T cells at the single-cell level were generated according to the previously reported procedures with some modifications (Picelli et al., 2014). First strand cDNA was then generated with Takara's SMARTScribe Reverse Transcriptase kit according to manufacturer's protocol (Takara Bio). Notable changes from the previously reported Smart-Seq2 RT step includes: 2 mM of dNTP and 2 μM of oligo dT were included in the capture buffer; 1M of Betaine and additional 6 mM $MgCl_2$ were included in the RT reaction buffer. The cDNA samples were then amplified with the KAPA Library Quantification kit for 22 – 25 cycles (Roche). We used 1 (of total 25/well) μL of amplified cDNA for single-cell TCR-sequencing and thus bypassing the RT step as reported previously (Han et al., 2014). To proceed with scRNA-Seq, full-length cDNA samples were first cleaned up with 0.6 – 0.8x volume of precalibrated AMPure XP beads (Beckman Coulter) to exclude DNA fragments smaller than 500 base pairs. We used the automatic liquid handler Biomek FX^P Automated Workstation (Beckman Coulter) in order to eliminate cell-to-cell variabilities. The quality of purified full-length cDNA was validated with the AATI Fragment Analyzer (Agilent). Subsequently, we used the measurements from the Fragment Analyzer in order to normalize the cDNA input with a Mantis liquid handler (Formulatrix). We then consolidated the cDNA samples into a 384-well plate (LVSD) with a Mosquito X1 liquid handler (TTP labtech). After transfer, Illumina sequencing libraries were prepared using a Mosquito HTS liquid handler (TTP labtech). We used only 0.4 uL (of total 23 uL) of cDNA per well to make the full transcriptome libraries with the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1096). We used custom-made i5 and i7 unique 8-bp indexing primers (IDT) to multiplex 384 wells in a single sequencing run. The libraries were amplified on a C1000 Touch Thermal Cycler with 384-Well Reaction Module (Bio-rad). We checked the pooled libraries with the Agilent 2100 Bioanalyzer (Stanford PAN facility) and acquired paired-end sequences (150bp x 2) on a Hiseq 4000 Sequencing System (Illumina) purchased with funds from NIH (S10OD018220) for the Stanford Functional Genomics Facility (SFGF).

Single-cell TCR sequencing (scTCR-seq)

Single T cells were sorted and captured as described above in the method for scRNA-Seq sample preparation. Following first strand cDNA synthesis (Takara) and amplification (Roche), we used 1 uL (of total 25 uL/well) of amplified cDNA for single-cell TCR-sequencing and thus bypassing the RT step as reported previously (Han et al., 2014). Nested PCR was performed with TCRα/β primers carrying multiplexing barcodes that enabled pooled CDR3α/β sequencing in a single Miseq run. Paired sequencing reads were joined, demultiplexed, and mapped to the human TCR references from the international ImMunoGeneTics information system® (IMGT) with custom scripts as reported previously (Han et al., 2014). Paired CDR3αβ sequences from the resected tumor of patient A6 were derived using the Chromium Single-Cell V(D)J kit from the 10x Genomics according to the protocol from the manufacturer. For advanced/metastatic lung cancer patients treated with anti-PD1 therapy, bulk TCR sequencing was performed on pre- and post-treatment PBMCs with Immunoseq assay (Adaptive Biotechnologies, Seattle, WA). Single-cell TCR sequencing was performed on post-treatment samples sorted for $CD38^+HLA\text{-}DR^+$ cells, as described above.

Data analyses of scRNA-Seq results

Sequencing reads were first de-multiplexed and binned into separate FASTQ files that correspond with the full transcriptomes of individual T cells. STAR aligner (2.6.1d) (Dobin et al., 2013) was used to map the reads with default parameters against human genome reference GRCh38 (v21) from the UCSC genome browser. Mapped reads were sorted and indexed with samtools (1.4) (Li et al., 2009). Gene expression was first quantified by counting reads mapped to genes with htseq-count (HTSeq 0.5.4p5) using the following settings:–stranded = no–type = exon–idattr = gene_name–mode = intersection-nonempty (Anders et al., 2015). Unless otherwise stated, all single-cell T cell states were analyzed with Seurat (3.1.4) packages in R using raw read counts. To derive TCR repertoires from the scRNA-Seq results, reads mapped to both the TCRα and TCRβ genes were first reconstructed with the TraCeR algorithm as described previously (Stubbington et al., 2016). The reconstructed DNA sequences were then submitted to the IMGT to call gene segment usage and the CDR3 amino acid sequences through HighV-QUEST.

GLIPH2 analysis on TRACERx data

Raw FASTQ files (tumor, n = 202; adjacent lung, n = 63) with demultiplexed, joined reads of the bulk CDR3β nucleotide sequences from the TRACERx cohort of NSCLC were downloaded from the Short Read Archive as reported (Joshi et al., 2019). The amino acid sequences of CDR3β, V gene usage, and the error-corrected clonal counts were subsequently derived by using the Decombinator scripts established previously (Oakes et al., 2017). To quantify the percentages of tumor-enriched specificity groups shown in Figure 1C, we first conducted joint GLIPH2 analyses with combined CDR3β sequences from the MDACC cohort (n = 778,938) and the bulk CDR3β sequences from the TRACERx cohort (tumor, n = 1,173,806 CDR3β sequences; adjacent lung, n = 247,578 CDR3β sequences). The total percentages (%) of top-20 clonally expanded as well as the rest CDR3β clonotypes that belonged to the 435 tumor-enriched specificity groups were then derived for each tumor (n = 202) and the adjacent lung tissue (n = 63).

Soluble biotinylated TCRα/β synthesis

Soluble TCRα/β chains used for yeast selections were made as described previously (Gee et al., 2018). Briefly, synthetic gene blocks (gBlocks®) of N-terminal truncated TCRα or TCRβ chain V and modified C gene fragments were assembled into the baculoviral pAcGP67a construct (BD Biosciences) with Gibson assembly (New England BioLabs). The final baculoviral plasmid was co-transfected into Sf9 cells (ATCC) with Bestbac 2.0 (Expression systems) with FuGENE® 6 (Promega) to make the crude viral supernatant (P0). Subsequently, viruses were passaged at a dilution of 1:500 in 30-50 mL cultures at a density of 1×10^6 cells/mL to generate

higher titer viruses (P1). To generate the soluble TCRα/β chains, up to 4 L of High Five (Hi5, ThermoFisher Scientific) cells were infected with P1 baculovirus at a dilution of 1:500-1:1000 at a density of 2 × 10^6 cells/mL for a week before protein purification. Recombinant TCRα/β chains were bound with Ni-NTA resin (QIAGEN) in the Hi5 cell media for 3 h at room temperature, washed with 20 mM imidazole in 1X HBS at pH 7.2, and eluded eluted in 200 mM imidazole in 1X HBS at pH 7.2. After buffer exchange to 1X HBS at pH 7.2 with a 30 kDa filter (Millipore-Sigma), purified proteins were biotinylated overnight with birA ligase in the presence of 100 μM biotin, 40 mM Bicine at pH 8.3, 10 mM ATP, and 10mM Magnesium Acetate at 4°C. Biotinylated proteins were purified by size-exclusion chromatography using an AKTAPurifier Superdex 200 column (GE Healthcare) and validated on a SDS-PAGE gel to confirm the stoichiometry and biotinylation with excess streptavidin.

Antigen discovery with the yeast library

To uncover the cognate antigens of the candidate TCRα/β, we used the yeast HLA-A*02 libraries displaying highly diverse peptides of 4 different length (Gee et al., 2018). Briefly, we first expanded 4 separate naive HLA-A*02 libraries carrying distinct lengths of peptides to beyond 10x diversities in SDCAA pH 6.0 before induction of the peptide-HLA-Aga2p composite proteins with SGCAA. Induced libraries were used for affinity-based selection with biotinylated soluble TCRα/β chains coupled to streptavidin-coated magnetic MACS beads (Miltenyi) in the presence of 0.5% bovine serum albumin and 1 mM EDTA to reduce the background. We cultured the selected yeast clones in SDCAA until confluency, then induced confluent cells in SGCAA for 2-3 days before the next round of selection. The selection was repeated four times and then enrichment of cognate antigens was confirmed with Sanger sequencing of 20 colonies. Once confirmed, we prepared the plasmid DNA from 5-10 × 10^7 yeast cells per round of selection by miniprep (Zymoprep II kit, Zymo Research). The peptide coding regions were PCR-amplified with composite oligos with Illumina P5/P7-Truseq indexed adapters and gel purified for pooled sequencing on a Miseq sequencer (2x150 V2 kit).

Lentiviral TCR transduction

TCRα chain, P2A linker, and TCRβ chain fusion gene fragments were purchased from IDT and cloned into MCS of the EF1a-MCS-GFP-PGK-puro lentiviral vector (Glanville et al., 2017; Witwicka et al., 2015). HEK293T cells were plated on a 10-cm dish at a density of 7.5 × 10^6 cells in 10 mL of DMEM the day prior to transfection. 293Ts were co-transfected with 3.3 μg of the lentiviral plasmid, 2.5 μg of the gag-pol plasmid, and 0.83 μg of the VSV-G envelope plasmid pre-mixed with 33 μL of PEI in 120 μL of Opti-MEM (ThermoFisher Scientific). After 24 h, the medium was replenished and viral supernatant was collected 24 and 48 h later. TCR-deficient Jurkat cells (below) were transduced with viral supernatant, TCR expression was assessed by flow cytometry, and TCR-expressing cells were sorted based on the expression of GFP, CD3, and the transduced TCRα/β chains. For lentivirus expressing full-length EntS, LMP2, and FluM1, gene fragments were also purchased from IDT and cloned into MCS of EF1a-MCS-GFP-PGK-puro lentiviral vector. Lentivirus for expressing human *TMEM161A* (NM_017814) was purchased from GeneCopoeia. Lentivirus was produced as described above, and 293T cells stably expressing HLA-A*02 (293A2) were transduced with viral supernatant. Transduced 293A2 cells were sorted based on GFP expression and used for *in vitro* T cell stimulation.

Retroviral TCR transduction

For retroviral-mediated expression of TCR2 in primary T cells, TCRα chain, P2A linker, and TCRβ chain were PCR amplified from the lentiviral vector (described above) and cloned into the MCS of an MSGV1-based retroviral vector (gift from Steve Rosenberg laboratory) using In-Fusion Cloning (Takara). For retroviral-mediated expression of TCR14 in primary T cells, TCRα chain, P2A linker, and TCRβ chain fusion gene fragments were purchased from IDT and cloned into MCS of an MSGV1-based retroviral vector.

Cell culture

The Jurkat 76 T cell line deficient for both TCRα and TCRβ were provided by Dr. Shao-An Xue (Department of Immunology, University of College London).

Jurkat cells and primary T cells were grown in complete RPMI (ThermoFisher) containing 10% FBS, 25 mM HEPES, 290 μg/mL L-glutamine, 100 U/mL penicillin, 100 U/mL streptomycin, 1mM sodium pyruvate, and 1x non-essential amino acids. T2 cells were grown in IMDM (Fisher Scientific) with 20% FBS, 290 μg/mL L-glutamine, 100 U/mL penicillin, 100 U/mL streptomycin. 293T cells stably expressing HLA-A*02 were provided by Dr. Steve Feldman (Stanford School of Medicine) and grown in DMEM (ThermoFisher) with 10% FBS, 290 μg/mL L-glutamine, 100 U/mL penicillin, 100 U/mL streptomycin.

In vitro stimulation of the Jurkat T cells

Jurkat 76 cells expressing the exogenous TCR of interest were sorted on CD3/GFP double-positive populations (Figure S3A) and co-cultured with T2 cells in complete RPMI as detailed above. To find homologous sequences of the identified epitopes, netMHCpan was used to predict the binding affinity of the homologous peptide to a given HLA allele (Jurtz et al., 2017; Reynisson et al., 2020). We used the BLASTP algorithm to perform the search for matching peptides in the UniParc protein database (Johnson et al., 2008). Peptides were dissolved in DMSO at 20 mM stock concentration and diluted to a final concentration of 2 μM. After 18 h of stimulation, cells were washed and stained with anti-CD3 (OKT3, Biolegend), anti-CD69 (FN50, Biolegend), and anti-TCRα/β (IP26, Biolegend) antibodies. Cells were acquired using FACS Fortessa (BD Biosciences) automated high throughput sampler or the Accuri C6 Plus flow cytometer (BD), and data analyzed using FlowJo software (Treestar).

Expression of TCRα/β on primary T cells

T cells were isolated from a leukoreduction system chamber from an HLA-A*02 positive healthy donor from the Stanford institutional blood bank using the RosetteSep human T cell enrichment cocktail (Stem Cell Technologies) and viably stored in liquid nitrogen. For T cell activation, T cells were thawed and stimulated with anti-CD3/CD28 beads (Life Technologies) in the presence of IL-2 (100 IU/mL). On days 1 and 2, activated T cells were retrovirally transduced using Retronectin (Takara) coated plates in media containing 100 IU/mL IL-2. Anti-CD3/CD28 beads were removed on day 3 and media containing IL-2 were replenished once every 2 days. Following 8 days of *in vitro* expansion, T cells were co-cultured with 293A2 cells expressing full-length TMEM161A, EntS, LMP2, FluM1, or GFP alone at a 1:1 ratio. Following 18 h incubation, cells were stained with anti-CD3 (OKT3, Biolegend), anti-CD69 (FN50, Biolegend), anti-TCRα/β (IP26, Biolegend), anti-CD137 (4B4-1, BD Biosciences), and live/dead near-IR dye (Invitrogen). Data were acquired using FACS Fortessa (BD Biosciences) automated high throughput sampler and analyzed using FlowJo software (Treestar).

Binding affinity measurements using BLI

Binding affinity of TCR2 to the indicated pMHC monomers was determined by BLI using an Octet QK instrument (ForteBio). The purified, soluble TCR2 was captured onto amine reactive second-generation (AR2G) biosensors using the amine reactive second-generation reagent kit. The ligand-bound biosensors were dipped into a concentration series (20 μM followed by 4-fold dilutions) of the indicated analytes in PBST (PBS with 0.05% Tween-20) to determine the binding kinetics. A series of unliganded biosensors dipped into the analytes served as controls for referencing. In addition, signals from analyte binding to an irrelevant TCR was used for non-specific binding correction. The traces were processed using ForteBio Data Analysis Software.

In vitro cytotoxicity assay

Primary T cells were isolated from HLA-A*02^{+} healthy donors. Cells were retrovirally transduced with TCR2 as described above. Following 9 days of *in vitro* expansion, cells were co-stained with TMEM9-mer/HLA-A*02 tetramers (APC) and FluM1/A*02 tetramers (PE) and enriched with anti-APC microbeads (Miltenyi Biotech), and enrichment confirmed with analysis on FACS Fortessa. The following day, TMEM9-mer/HLA-A*02 tetramers-enriched T cells were co-cultured with H1395 lung adenocarcinoma cells at a 20:1 ratio in 96-well flat bottom plates for over 120 h. A minimum of triplicate wells were plated for each condition. Plates were imaged every 3 h using the InuCyte ZOOM Live-Cell analysis system (Essen Bioscience). Four images per well at 10x zoom were acquired at each time point. Total integrated GFP intensity per well was recorded and normalized to the starting measurement and plotted over time.

Immunohistochemistry of TMEM161A

TMEM161A staining of paraffin-embedded tissue was performed according to standard procedures by the Stanford Human Pathology/Histology Service Center.

Anti-TMEM161A antibody was stained at 1:50 (abcam ab180954), followed by HRP-conjugated secondary antibody. Tissue was counterstained with hematoxylin. Automated imaging analysis was performed using Fiji imaging processing package (Schindelin et al., 2012).

Whole-exome sequencing

Whole-exome sequencing of tumor DNA and matched germline leukocyte DNA was performed by inputting 75ng of sheared genomic DNA for library preparation with the KAPA HyperPrep Kit (Roche) with modifications to the manufacturer's instruction, as described previously (Hellmann et al., 2020). Library-prepared samples were captured with the SeqCap EZ MedExome Kit (NimbleGen) according to the manufacturer's instructions. Sequencing data were demultiplexed and mapped to hg19 using a custom bioinformatics pipeline, as described previously (Newman et al., 2014). VarScan 2 (Koboldt et al., 2012), Mutect (Cibulskis et al., 2013), and Strelka (Saunders et al., 2012) were used to call variants use default parameters. Variants called by at least two of the approaches were then filtered by requiring: 1) variant allele frequency of at least 2.5%, 2) at least 30X depth in both tumor and germline samples, 3) zero germline reads, and 4) a population allele frequency of less than 0.1% in the Genome Aggregation database (Lek et al., 2016).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless stated otherwise, all statistical analyses performed in finding high-confidence specificity groups with GLIPH2 were Fisher's exact tests using the contingency tables with the CDR3β query set (specificity group) and the reference set (Huang et al., 2020). Poisson test was used to determine the representation bias in comparisons of distinct CDR3β sequences or specificity groups between tumors and adjacent lungs. Hypergeometric test (phyper in R, lower.tail = FALSE) was used to quantify the enrichment of HLA supertype alleles (Figure 1D) and to find top-enriched HLA allele(s) for each specificity group (Figures 2C and 2D). Student's t test was used to assess the results from all *in vitro* assays. Statistical significance was defined as p value < 0.05.

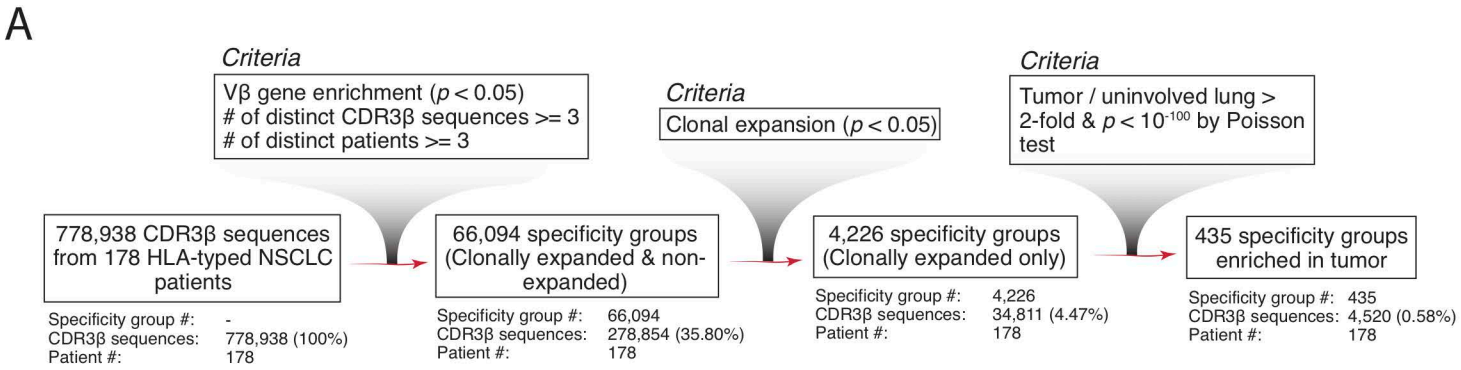
Immunity, Volume *54*

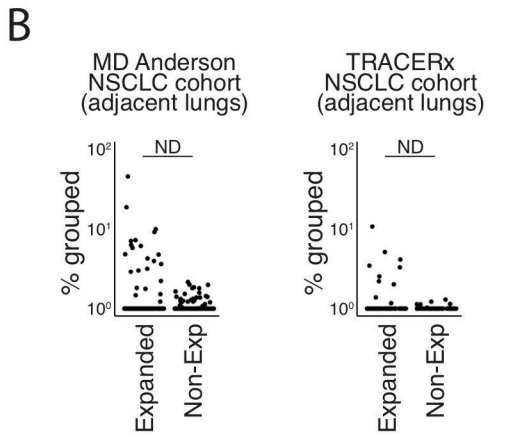
Supplemental information

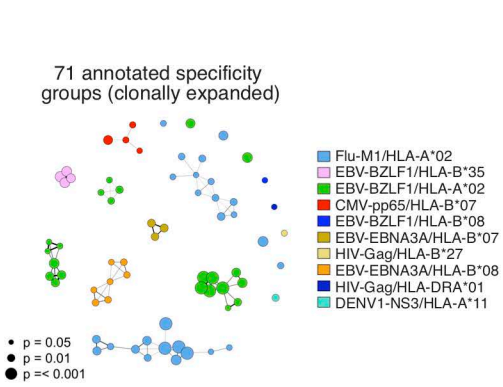
Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery

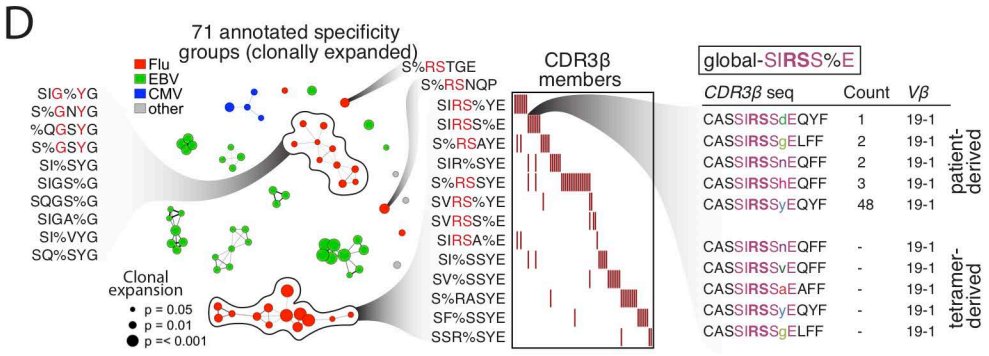
Shin-Heng Chiou, Diane Tseng, Alexandre Reuben, Vamsee Mallajosyula, Irene S. Molina, Stephanie Conley, Julie Wilhelmy, Alana M. McSween, Xinbo Yang, Daisuke Nishimiya, Rahul Sinha, Barzin Y. Nabet, Chunlin Wang, Joseph B. Shrager, Mark F. Berry, Leah Backhus, Natalie S. Lui, Heather A. Wakelee, Joel W. Neal, Sukhmani K. Padda, Gerald J. Berry, Alberto Delaidelli, Poul H. Sorensen, Elena Sotillo, Patrick Tran, Jalen A. Benson, Rebecca Richards, Louai Labanieh, Dorota D. Klysz, David M. Louis, Steven A. Feldman, Maximilian Diehn, Irving L. Weissman, Jianjun Zhang, Ignacio I. Wistuba, P. Andrew Futreal, John V. Heymach, K. Christopher Garcia, Crystal L. Mackall, and Mark M. Davis

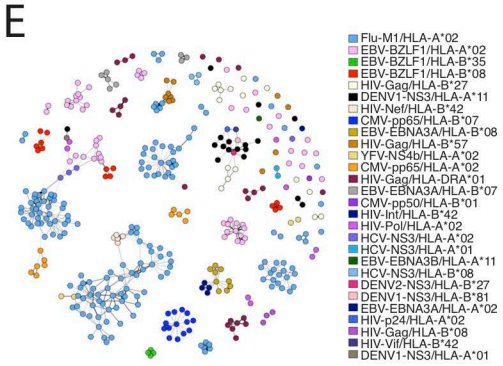
Figure S1

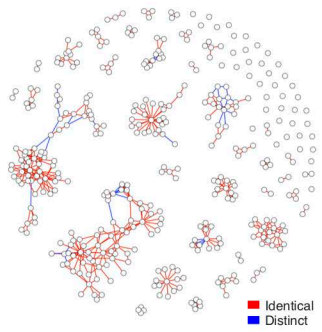

A
Criteria
Vβ gene enrichment (p < 0.05)
# of distinct CDR3β sequences >= 3
# of distinct patients >= 3
Criteria
Clonal expansion (p < 0.05)
Criteria
Tumor / uninvolved lung > 2-fold & p < 10-100 by Poisson test
778,938 CDR3β sequences from 178 HLA-typed NSCLC patients
Specificity group #: -
CDR3β sequences: 778,938 (100%)
Patient #: 178
66,094 specificity groups (Clonally expanded & non-expanded)
Specificity group #: 66,094
CDR3β sequences: 278,854 (35.80%)
Patient #: 178
4,226 specificity groups (Clonally expanded only)
Specificity group #: 4,226
CDR3β sequences: 34,811 (4.47%)
Patient #: 178
435 specificity groups enriched in tumor
Specificity group #: 435
CDR3β sequences: 4,520 (0.58%)
Patient #: 178

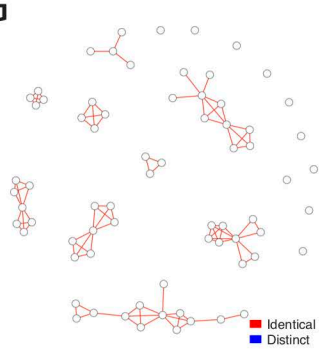

B
MD Anderson NSCLC cohort (adjacent lungs)
TRACERx NSCLC cohort (adjacent lungs)
ND
% grouped
Expanded
Non-Exp

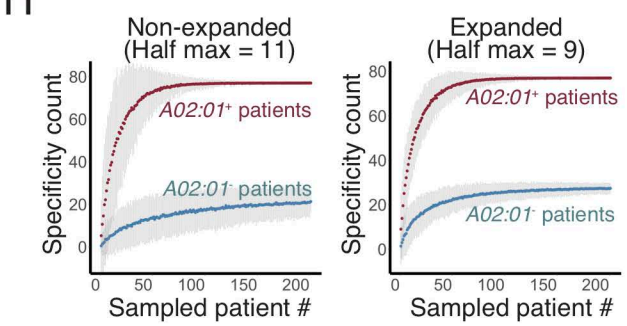

C
71 annotated specificity groups (clonally expanded)
Flu-M1/HLA-A*02
EBV-BZLF1/HLA-B*35
EBV-BZLF1/HLA-A*02
CMV-pp65/HLA-B*07
EBV-BZLF1/HLA-B*08
EBV-EBNA3A/HLA-B*07
HIV-Gag/HLA-B*27
EBV-EBNA3A/HLA-B*08
HIV-Gag/HLA-DRA*01
DENV1-NS3/HLA-A*11
p = 0.05
p = 0.01
p =< 0.001

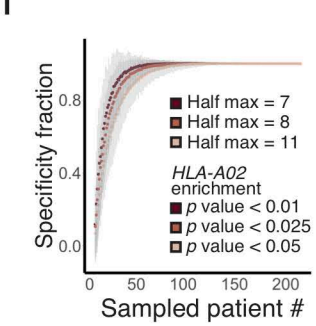

D
71 annotated specificity groups (clonally expanded)
Flu
EBV
CMV
other
SIG%YG
S%GNYG
%QGSYG
S%GSYG
SI%SYG
SIGS%G
SQGS%G
SIGA%G
SI%VYG
SQ%SYG
Clonal expansion
p = 0.05
p = 0.01
p =< 0.001
S%RSTGE
S%RSNQP
SIRS%YE
SIRSS%E
S%RSAYE
SIR%SYE
S%RSSYE
SVRS%YE
SVRSS%E
SIRSA%E
SI%SSYE
SV%SSYE
S%RASYE
SF%SSYE
SSR%SYE
CDR3β members
global-SIRSS%E
CDR3β seq
Count
Vβ
CASSIRSSdEQYF 1 19-1
CASSIRSSgELFF 2 19-1
CASSIRSSnEQFF 2 19-1
CASSIRSShEQFF 3 19-1
CASSIRSSyEQYF 48 19-1
patient-derived
CASSIRSSnEQFF - 19-1
CASSIRSSvEQFF - 19-1
CASSIRSSaEAFF - 19-1
CASSIRSSyEQYF - 19-1
CASSIRSSgELFF - 19-1
tetramer-derived


E
Flu-M1/HLA-A*02
EBV-BZLF1/HLA-A*02
EBV-BZLF1/HLA-B*35
EBV-BZLF1/HLA-B*08
HIV-Gag/HLA-B*27
DENV1-NS3/HLA-A*11
HIV-Nef/HLA-B*42
CMV-pp65/HLA-B*07
EBV-EBNA3A/HLA-B*08
HIV-Gag/HLA-B*57
YFV-NS4b/HLA-A*02
CMV-pp65/HLA-A*02
HIV-Gag/HLA-DRA*01
EBV-EBNA3A/HLA-B*07
CMV-pp50/HLA-B*01
HIV-Int/HLA-B*42
HIV-Pol/HLA-A*02
HCV-NS3/HLA-A*02
HCV-NS3/HLA-A*01
EBV-EBNA3B/HLA-A*11
HCV-NS3/HLA-B*08
DENV2-NS3/HLA-B*27
DENV1-NS3/HLA-B*81
EBV-EBNA3A/HLA-A*02
HIV-p24/HLA-A*02
HIV-Gag/HLA-B*08
HIV-Vif/HLA-B*42
DENV1-NS3/HLA-A*01


F
Identical
Distinct


G
Identical
Distinct


H
Non-expanded (Half max = 11)
Expanded (Half max = 9)
A02:01+ patients
A02:01- patients
Specificity count
Sampled patient #


I
Specificity fraction
Half max = 7
Half max = 8
Half max = 11
HLA-A02 enrichment
p value < 0.01
p value < 0.025
p value < 0.05
Sampled patient #

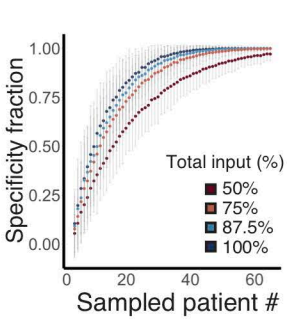

J
Specificity fraction
Total input (%)
50%
75%
87.5%
100%
Sampled patient #

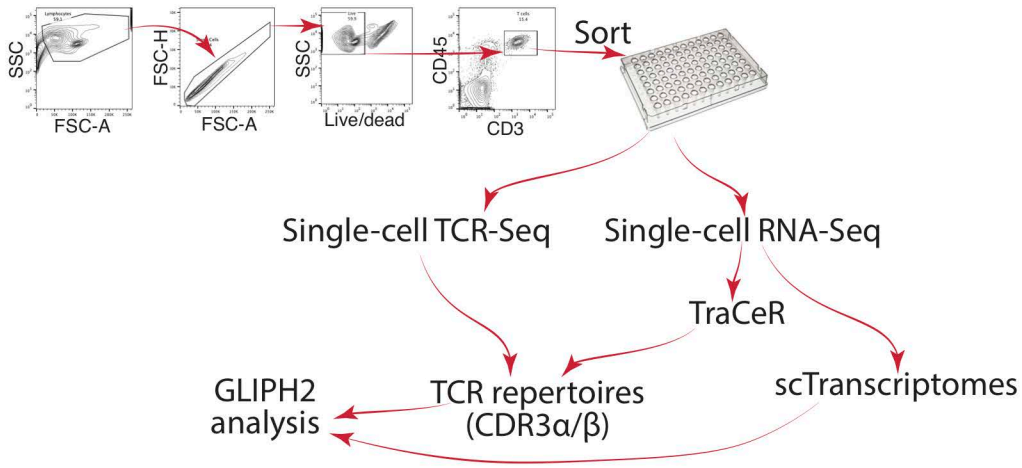
Figure S1, related to Figure 1

(**A**) Specificity inference pipeline. (**B**) Low percentages of TCR clonotypes from adjacent lung tissues are grouped into tumor-enriched specificity groups. The percentages of the top 20 most expanded CDR3β clonotypes from the adjacent lung tissues of patients belonging to the MDACC NSCLC cohort (n=178 samples, left) and the TRACERx cohort (n=63 samples, right) were quantified for those that belonged to the 435 tumor-enriched specificity groups as in Figure 1B (% grouped, $\log_{10}$-converted). The same analysis was performed on the remainder of the CDR3β clonotypes (Non-exp, non-expanded). ND, no statistically significant difference was found. (**C**) The 71 clonally expanded specificity groups annotated and colored with 10 indicated tetramers are shown in the network. (**D**) Left to right, network analysis of 71 clonally expanded specificity groups colored as in Figure 1C is shown; two large Flu-related communities (red) are circled and the CDR3β members of the specificity groups are highlighted with the previously reported short motifs "RS" and "GxY" highlighted in red font; heatmap showing distinct CDR3β members (columns) of the Flu-related (with the "RS" motif) specificity groups (rows) and the levels of shared CDR3β members between specificity groups within the circled community; table showing an example of the "SIRSS%E" specificity group containing the short "RS" motif (bold) that is annotated with 5 Flu-specific tetramer sequences (bottom). The counts of distinct CDR3β members from tumor and the *Vβ* gene usage are shown (top). (**E**) 394 specificity groups annotated with indicated tetramers (key) were organized into distinct communities through shared CDR3β sequence(s) as in Figure 1C. Thickness of edge represents numbers of shared CDR3β sequence(s) between any two connected nodes. (**F**) Community plot as in (**E**). Color of edge represents shared CDR3β sequence(s) between specificity groups with identical (red) or distinct (blue) specificities defined by tetramer-derived sequences (labeled with distinct colors, **E**). 588 of all (n=634) connections (edges) are labeled in red (92.74%). (**G**) 71 clonally expanded specificity groups as in (**C**). Color of edge represents shared CDR3β sequence(s) between specificity groups with identical (red) or distinct (blue) specificities defined by tetramer-derived sequences (**C**). 92 of all (n=92) connections (edges) are labeled in red (100%). (**H**-**J**) TCR specificity group saturation is dependent of the level of clonal expansion, the absolute numbers of specificity groups, as well as the sequencing depth of the repertoires. (**H**) Bootstrapping for quantification of clonally expanded (right, n=77) and non-expanded (left, n=77) *HLA-A*02:01*-enriched specificity groups with CDR3β sequences from either HLA-A02$^+$ (red) or HLA-A02$^-$ (blue) NSCLC patients. Bootstrapping was done by "sampling with replacement" and the X axis represents the number of patients randomly sampled (Sampled patient #, Methods) and the Y axis represents the numbers of specificity groups quantified with a given sampling event. Shades of error bars represent the 3X standard errors derived from 100 sampling events

for a given number of sampled patients. (**I**) Bootstrapping for quantification of *HLA-A*02:01*-enriched specificity groups with varying cutoffs for *HLA-A*02:01* enrichment ($p < 0.05$, n=1267; $p < 0.025$, n=319; $p < 0.01$, n=72). As in (**H**), X axis represents the number of patients randomly sampled (Sampled patient #). Y axis represents the numbers of specificity groups quantified with a given sampling event that are normalized against the number of the total specificity groups used (Specificity fraction). Shades of error bars represent the 3X standard errors derived from 100 sampling events for a given number of sampled patients. (**J**) Bootstrapping for quantification of *HLA-A*02:01*-enriched specificity groups with varying input CDR3β sequencing depth (50, 75, 87.5, or 100% of total input by random down-sampling). As in (**I**), X axis represents the number of patients randomly sampled and Y axis represents the normalized numbers of specificity groups. Shades of error bars represent the 3X standard errors derived from 100 sampling events for a given number of sampled patients.

Figure S2

A



B

Flu-related

Specificity group ID SV%SNQP

TCRβ	Vβ
CASSVaSNQPQHF	TRBV9-1
CASSVdSNQPQHF	TRBV9-1
CASSVeSNQPQHF	TRBV6-8
CASSVfSNQPQHF	TRBV19-1
CASSVgSNQPQHF	TRBV9-1
CASSVkSNQPQHF	TRBV9-1
CASSVlSNQPQHF	TRBV6-8
CASSVnSNQPQHF	TRBV10-2
CASSVsSNQPQHF	TRBV19-1
CASSVtSNQPQHF	TRBV4-2
CASSVySNQPQHF	TRBV19-1

TCR12 clone
α **CATDDSGGFKTIF**
β **CASSVTSNQPQHF**

Specificity group ID SIRS%YE

TCRβ	Vβ
CASSIRStYEQYF	TRBV19-1
CASSIRSgYEQFF	TRBV19-1
CASSIRSgYEQYF	TRBV19-1
CASSIRSdYEQYF	TRBV19-1
CASSIRSeYEQYF	TRBV19-1
CASSIRSaYEQFF	TRBV19-1
CASSIRSaYEQYF	TRBV19-1

TCR13 clone
α **CAGPTGGGSQGNLIF**
β **CASSIRSAYEQYF**

Specificity group ID S%RSTDT

TCRβ	Vβ
CASSkRSTDTQYF	TRBV19-1
CASSdRSTDTQYF	TRBV2-1
CASShRSTDTQYF	TRBV19-1
CASSsRSTDTQYF	TRBV19-1
CASSlRSTDTQYF	TRBV19-1
CASSeRSTDTQYF	TRBV10-2
CASSsRSTDTQYF	TRBV6-8

TCR14 clone
α **CAVTYGGSQGNLIF**
β **CASSSRSTDTQYF**

EBV-related

Specificity group ID RTG%GNT

TCRβ	Vβ
CATRTGgGNTIYF	TRBV6-8
CASRTGpGNTIYF	TRBV9-1
CSARTGtGNTIYF	TRBV29-1
CASRTGdGNTIYF	TRBV9-1
CSARTGvGNTIYF	TRBV20-1
CSVRTGaGNTIYF	TRBV29-1
CAYRTGsGNTIYF	TRBV30-1
CASRTGsGNTIYF	TRBV6-8

TCR15 clone
α **CAEDLNARLMF**
β **CSARTGVGNTIYF**

C

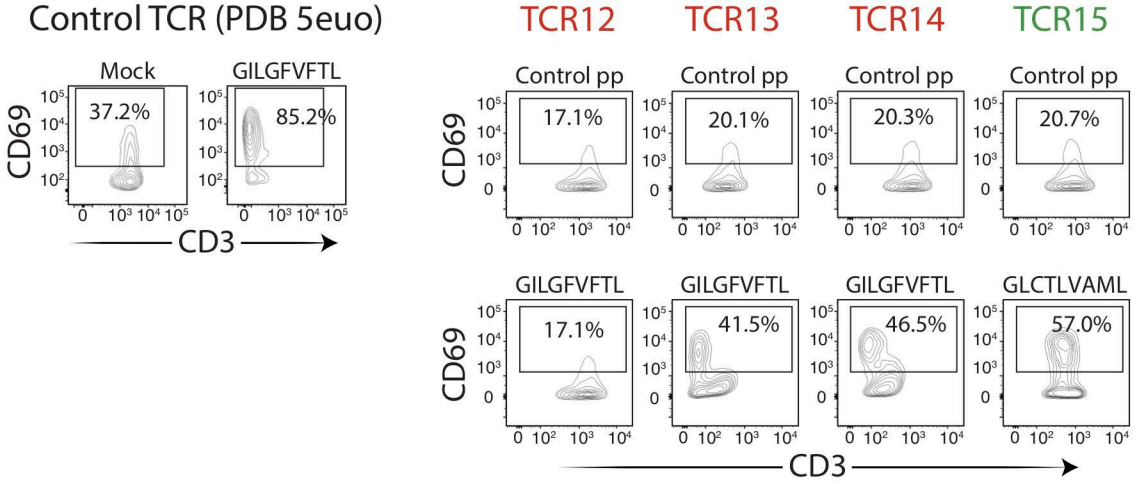


Figure S2, related to Figure 2

(**A**) Schematic of the combined single-cell TCR-Seq and single-cell RNA-Seq (scRNA-seq) procedures. $CD45^{+}CD3^{+}$ T cells were sorted from single-cell suspensions of lung tumor samples from patients with NSCLC at Stanford. Single-cell TCR-Seq was performed using nested multiplexed PCR as previously described [Han *et al*., 2014. Nat. Biotechnol. 32, 684]. scRNA-seq was performed according to previous methods [Picelli *et al*., 2014. Nat. Protoc. 9, 171] with modifications (STAR Methods). TCR repertoires were integrated from the single-cell TCR-Seq pipeline and from the scRNA-seq data with reconstruction using the TraCeR algorithm [Stubbington *et al*., 2016. Nat. Methods 13, 329] for GLIPH2 analysis. scTranscriptomes, single-cell transcriptomes derived from scRNA-seq. (**B**) GLIPH2 inferred clone TCR12, TCR13, and TCR14 to recognize Influenza virus M1 (FluM1) 9mer peptide "GILGFVFTL" in the context of HLA-A*02; clone TCR15 is inferred to recognize EBV BMLF1 9mer peptide "GLCTLVAML" in the context of HLA-A*02. Tables show subsets of CDRβ members for each specificity group. The selected T cell clones with paired CDR3α/β sequences are highlighted in bold in the tables and the CDR3 sequences of both TCRα/β chains are shown at the bottom. All four clones were found in tumors from 2 different NSCLC patients from the Stanford cohort. (**C**) Right, the TCR α/β sequences of the four chosen T cell clonotypes in (**B**) were ectopically expressed in TCR-deficient Jurkat-76 cells and stimulated with T2 ($HLA\text{-}A02^{+}$) cells pulsed with indicated peptides (right, above FACS plots). CD69 expression quantified by FACS is shown. Left, Jurkat cells expressing the control TCRα/β chains (PDB 5euo) previously reported to recognize FluM1 in the context of HLA-A*02 and stimulated with or without (Mock) the 9mer peptide "GILGFVFTL" are shown. Control peptide (pp), $CMV/pp65_{495\text{-}503}$.

Figure S3

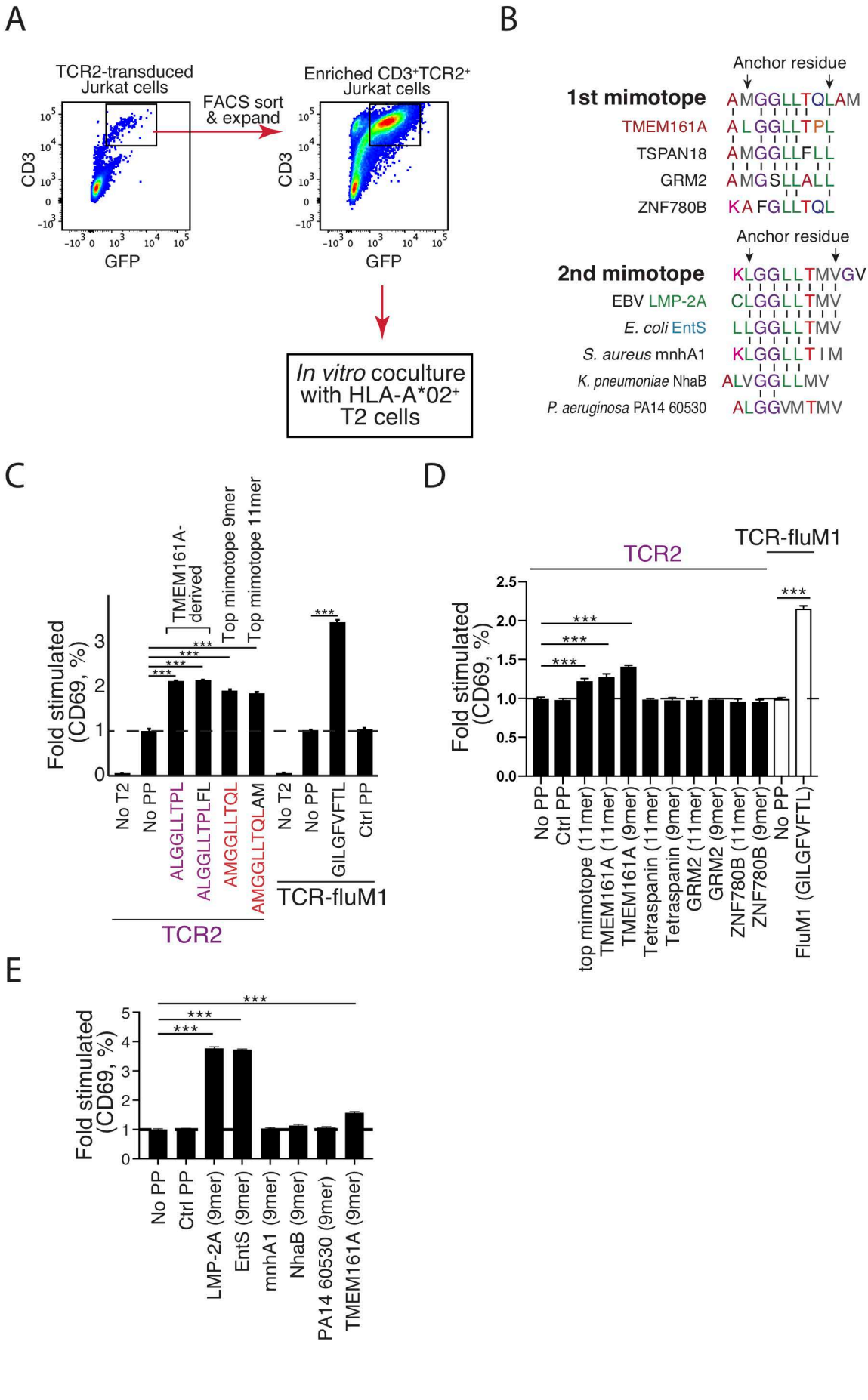
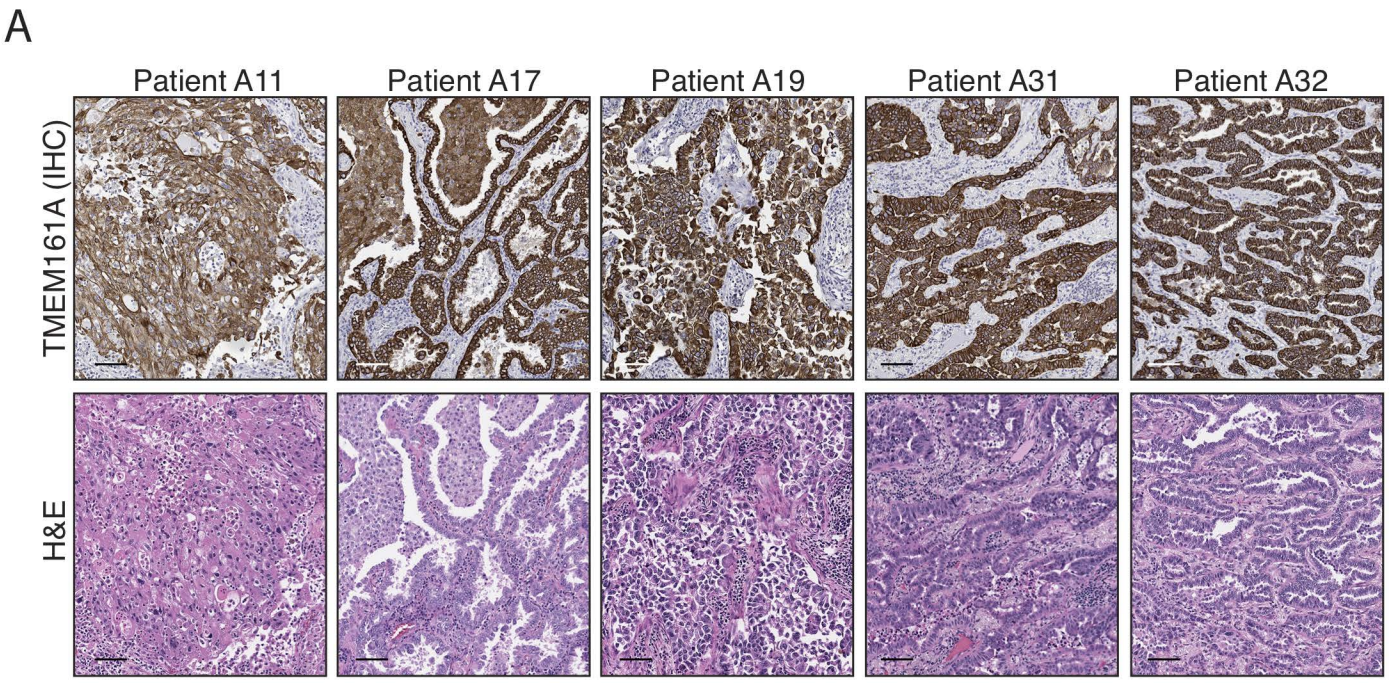
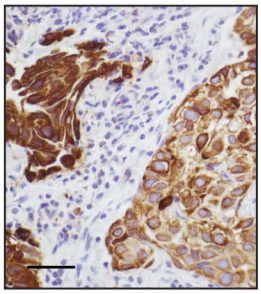

A
TCR2-transduced Jurkat cells
FACS sort & expand
Enriched CD3+TCR2+ Jurkat cells
CD3
GFP
In vitro coculture with HLA-A*02+ T2 cells
B
Anchor residue
1st mimotope
AMGGLLTQLAM
TMEM161A
ALGGLLTPL
TSPAN18
AMGGLLFLL
GRM2
AMGSLLALL
ZNF780B
KAFGLLTQL
Anchor residue
2nd mimotope
KLGGLLTMVGV
EBV LMP-2A
CLGGLLTMV
E. coli EntS
LLGGLLTMV
S. aureus mnhA1
KLGGLLTIM
K. pneumoniae NhaB
ALVGGLLMV
P. aeruginosa PA14 60530
ALGGVMTMV
C
Fold stimulated (CD69, %)
TMEM161A-derived
Top mimotope 9mer
Top mimotope 11mer
No T2
No PP
ALGGLLTPL
ALGGLLTPLFL
AMGGLLTQL
AMGGLLTQLAM
No T2
No PP
GILGFVFTL
Ctrl PP
TCR2
TCR-fluM1
D
TCR2
TCR-fluM1
Fold stimulated (CD69, %)
No PP
Ctrl PP
top mimotope (11mer)
TMEM161A (11mer)
TMEM161A (9mer)
Tetraspanin (11mer)
Tetraspanin (9mer)
GRM2 (11mer)
GRM2 (9mer)
ZNF780B (11mer)
ZNF780B (9mer)
No PP
FluM1 (GILGFVFTL)
E
Fold stimulated (CD69, %)
No PP
Ctrl PP
LMP-2A (9mer)
EntS (9mer)
mnhA1 (9mer)
NhaB (9mer)
PA14 60530 (9mer)
TMEM161A (9mer)

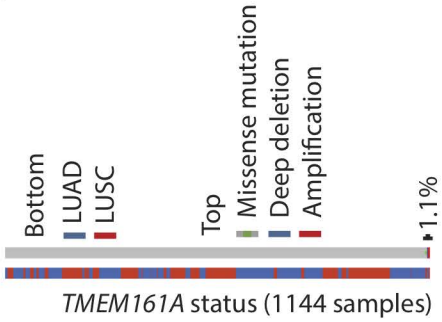
Figure S3, related to Figure 3

(**A**) TCR-deficient Jurkat cells were transduced with lentivirus carrying a composite coding region of TCR2α chain, 2A peptide sequence (2Ap), TCR2β chain, 2Ap, and GFP (Jurkat-TCR2). Transduced cells were subsequently sorted by FACS for the GFP^+CD3^+ population (left) and allowed to expand (top right) and used in *in vitro* stimulation experiments (bottom right). Similar strategies were used to make other stable TCR-Jurkat clones throughout the paper. (**B**) Protein database search showed partial matches of the top 2 mimotopes with candidate coding sequences from various species. All matches were 9mers and predicted to bind HLA-A*02 with high affinities by netMHCpan 4.0. (**C**) Jurkat-TCR2 was stimulated with T2 ($HLA\text{-}A2^+$) cells pulsed with indicated peptides, including the top mimetope from the yeast screen (11mer, AMGGLLTQLAM), the 9mer from the top mimetope predicted to bind HLA-A*02 with high affinity (AMGGLLTQL), and both 9mer/11mer peptides from the TMEM161A coding region (ALGGLLTPL and ALGGLLTPLFL, respectively) with sequence homology to the top mimetope. CD69 upregulation was quantified by FACS. Jurkat cells expressing the control TCRα/β chains (TCR-fluM1, PDB 5euo) and stimulated with the cognate 9mer "GILGFVFTL", the control peptide (Ctrl PP, $CMV/pp65_{495\text{-}503}$), no peptide (No PP), or no co-cultured T2 cells are included as controls. ***, $p < 0.001$ by t test. (**D**,**E**) The endogenous peptides derived from human TMEM161A, LMP-2A of EBV, and EntS of *E. coli* were recognized by TCR2. Jurkat-TCR2 cells were cocultured with $HLA\text{-}A02^+$ T2 cells and pulsed with indicated peptides: Ctrl PP (control peptide), $Flu/M1_{58\text{-}66}$ GILGFVFTL; top mimetope 11mer, AMGGLLTQLAM; TMEM161A 11mer, ALGGLLTPLFL; TMEM161A 9mer, ALGGLLTPL; Tetraspanin 11mer, AMGGLLFLLGF; Tetraspanin 9mer, AMGGLLFLL; GRM2 (Glutamate receptor) 11mer, AMGSLLALLAL; GRM2 9mer, AMGSLLALL; ZNF780B (Zinc finger protein 780B isoform X1) 11mer, KAFGLLTQLAQ; ZNF780B 9mer, KAFGLLTQL in (**D**) and Ctrl PP (control peptide), $Flu/M1_{58\text{-}66}$ GILGFVFTL; LMP-2A 9mer, CLGGLLTMV; EntS 9mer, LLGGLLTMV; mnhA1 9mer, KLGGLLTIM; NhaB 9mer, ALVGGLLMV; PA14 60530 9mer, ALGGVMTMV; TMEM161A 9mer, ALGGLLTPL in (**E**). Jurkat cells expressing the control TCR (TCR-fluM1, PDB 5euo) are stimulated with the cognate 9mer "GILGFVFTL" for comparison.

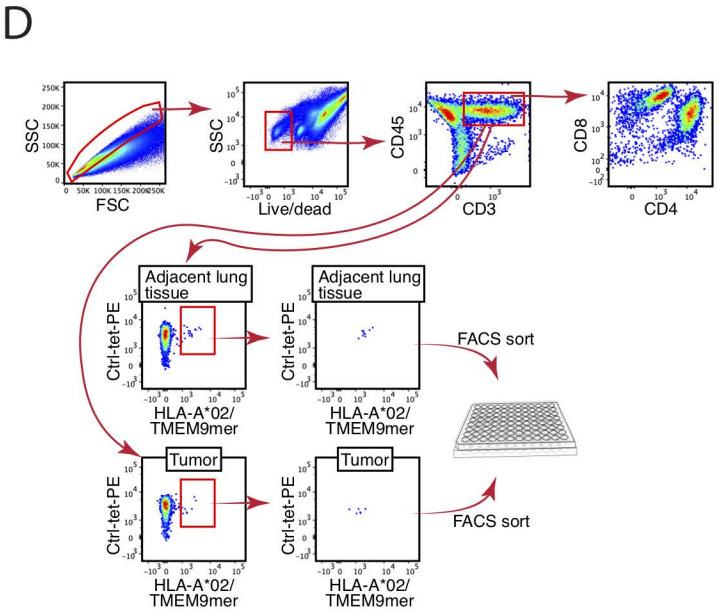
Figure S4

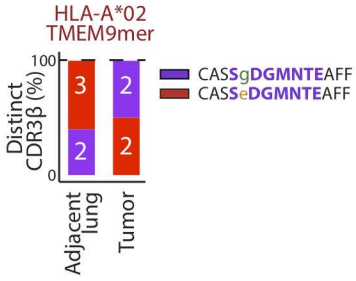






Mutation type	Count
Missense mutation	4
Deep deletion	2
Amplification	6
No alterations	1132





F

S%DGMNTE	LUSC	LUAD
+	21	10
—	16	31

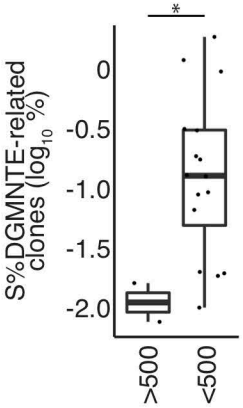
$p = 5.3 \times 10^{-3}$

Figure S4, related to Figure 4

(**A**) TMEM161A is broadly expressed on human NSCLC tumors. Representative images of TMEM161A immunohistochemistry on tumor (top) and hematoxylin & eosin (H&E) staining from adjacent sections (bottom) from Stanford patients A11, A17, A19, A31, and A32. Scale bar, 100 μm. (**B**) Strong and dim TMEM161A staining on a section from A6 patient tumor. Scale bar, 40 μm. (**C**) TMEM161A is a non-mutated tumor antigen. Left, percentages of all types of genetic alterations within the TMEM161A locus defined with whole-genome sequencing from the Cancer Genome Atlas (TCGA) project (pan-lung cancer cases, n=1144) are shown. Right, number of cases with indicated genomic alterations. (**D,E**) TMEM9mer/A02 tetramer^{+} sorted T cells from tumor and the adjacent lung tissue carry the "S%DGMNTE" motif, as predicted by GLIPH2. (**D**) Sorting HLA-A*02/TMEM9mer^{+}CD8^{+} T cells from the resected tumor and the adjacent lung tissue of NSCLC patient A6 through FACS. Single cell suspensions were prepared and stained with anti-CD4, CD8, CD3, and CD45 antibodies, live/dead marker AquaZombie, PE-conjugated tetramer HLA-A*02/viral peptides (CMV-pp65$_{495\text{-}503}$ and EBV-BMLF1$_{280\text{-}288}$), and APC-conjugated tetramer HLA-A*02/TMEM9mer. Specific T cells were sorted onto 96-well plates based on the following criteria: non-doublets, live cells, CD3^{+}CD45^{+}, HLA-A*02/viral peptides^{-}, and HLA-A*02/TMEM9mer^{+}. (**E**) Percentages (%) of distinct CDR3β sequences of tetramer-sorted CD8^{+} T cells as in (**D**) from the adjacent lung tissue and tumor are shown. Numbers in bars represent the counts of sorted cells. (**F**) Numbers of A02^{+} squamous cell lung carcinoma (LUSC) or lung adenocarcinoma (LUAD) patients from the MDACC cohort with or without detected T cells carrying the "S%DGMNTE" CDR3β motif. *p* value = 5.3 x 10^{-3} by Fisher's Exact test.

Figure S5

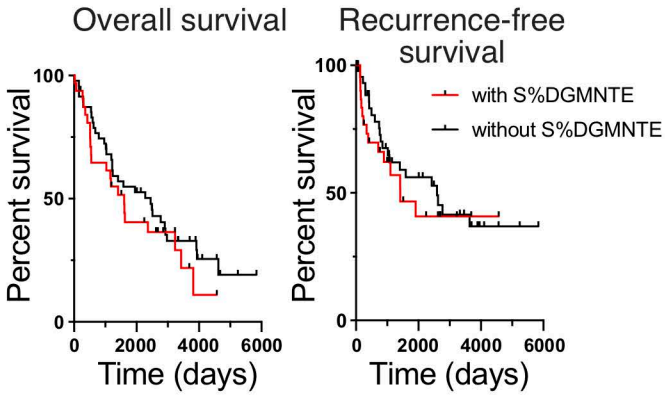
A



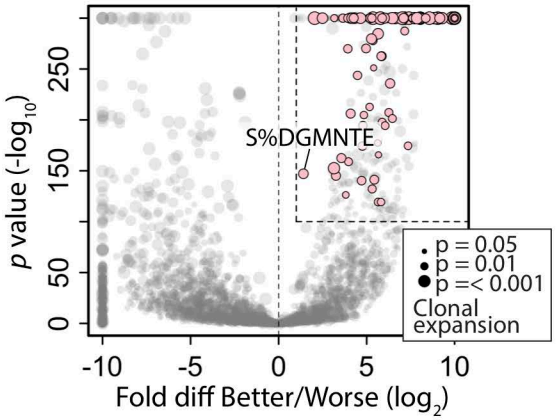
B

	With S%DGMNTE motif	Without S%DGMNTE motif	Fisher's exact test
# Patients	32	47	
Gender			
M	16/32 (50%)	29/47 (62%)	ns
F	16/32 (50%)	18/47 (38%)	ns
Race			
Caucasian	31/32 (97%)	43/47 (91%)	ns
Other	1/32 (3%)	4/47 (9%)	ns
Histology			
Adenocarcinoma	10/32 (31%)	31/47 (66%)	p < 0.01
Squamous cell carcinoma	21/32 (65%)	16/47 (34%)	p < 0.01
Smoking status			
Current smoker	16/32 (50%)	21/47 (45%)	ns
Former smoker	16/32 (50%)	21/47 (45%)	ns
Never smoker	0/32 (0%)	5/47 (11%)	ns
Stage			
I	13/32 (41%)	23/47 (49%)	ns
II	16/32 (50%)	18/47 (38%)	ns
III	3/32 (9%)	6/47 (12%)	ns
Primary tumor size (cm)	4.6	3.8	ns
Adjuvant chemotherapy	12/30 (40%)	14/46 (30%)	ns
Multifocal disease	5/32 (16%)	5/47 (11%)	ns

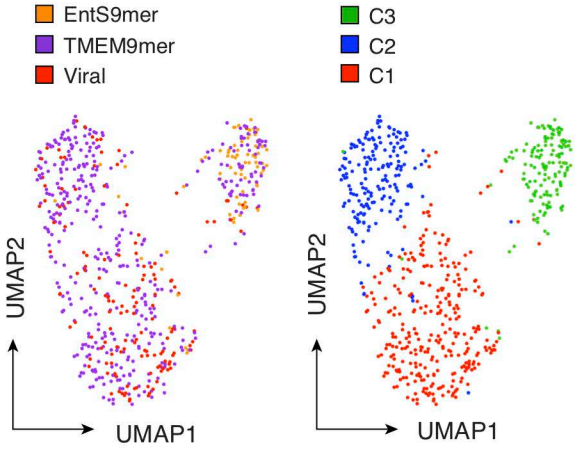
C

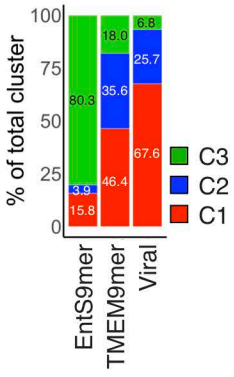


D



E





F

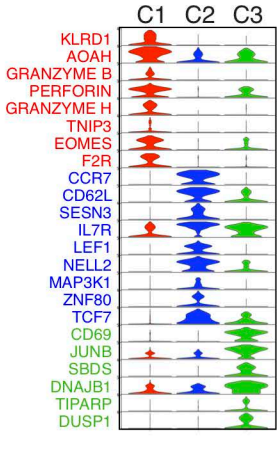
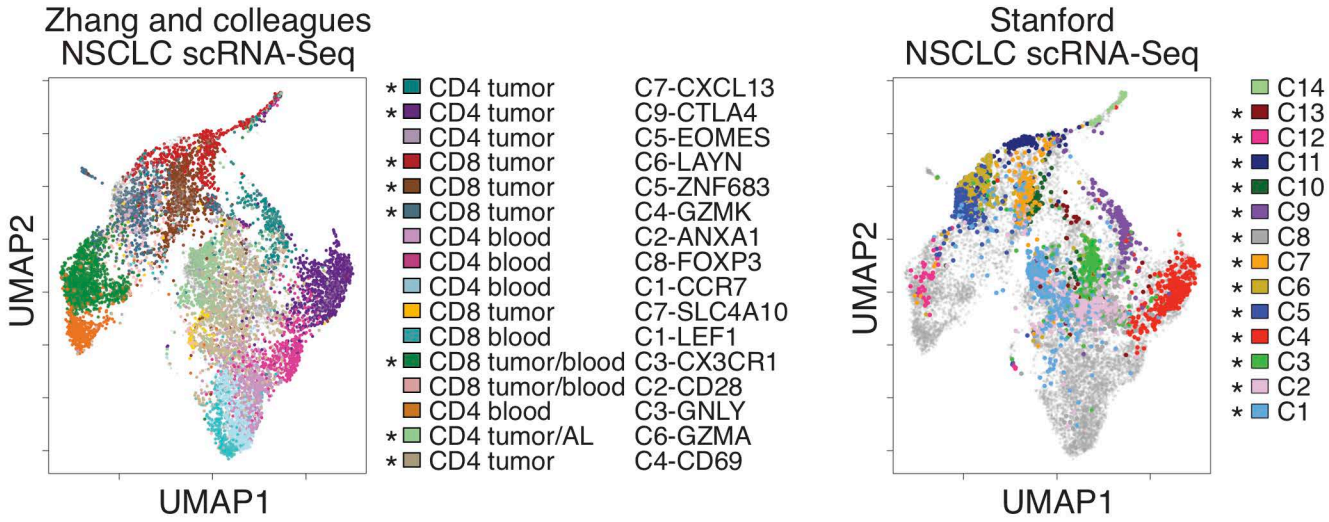


Figure S5, related to Figure 5

(**A**) Total percentage (log_{10}) of T cells carrying the "S%DGMNTE" CDR3β motif for $A02^+$ patients from the MDACC cohorts stratified as having more than 500 (>500) or less than 500 (<500) total mutation counts (n = 34). (**B**) Break down of the clinical data of the MDACC NSCLC patient cohort stratified by detection of CDR3β sequences carrying the "S%DGMNTE" sequence motif. (**C**) There are no differences in overall survival (left) or recurrence-free survival (right) among patients with tumor-infiltrating T cell CDR3β containing the "S%DGMNTE" sequence motif (n = 32) vs patients without tumor-infiltrating T cell CDR3β containing the same motif (n = 49). (**D**) Volcano plot showing the probabilities of clonal biases for the 4,226 clonally expanded specificity groups between the tumors from patients with (worse, n = 88) or without recurrence (better, n = 90) by Poisson test. Tumor-enriched specificity groups that show a significant clonal bias in patients without recurrence are highlighted in pink (n = 146). Specificity group S%DGMNTE is highlighted. (**E**,**F**) A subset of TMEM161A-specific T cells found in healthy donors' peripheral blood reveal effector phenotype. (**E**) Dimension reduction by Uniform Manifold Approximation and Projection (UMAP) of the scRNA-Seq results from the sorted $CD45^+CD8^+CD3^+$ T cells from PBMC with indicated HLA-A*02 tetramers (left) as in Fig 5A identified 3 major cell states (right, # indicates percentage of each cell state). Viral, CD8 T cells sorted with HLA-A*02 tetramers loaded with Flu, EBV, or CMV peptides (Flu-$M1_{58\text{-}66}$, CMV-$pp65_{495\text{-}503}$,and EBV-$BMLF1_{280\text{-}288}$). (**F**) Stacked violin plot showing the differential genes expressed by the identified cell states as in (**E**).

Figure S6

A



B

Cluster ID (Stanford)	Lineage (Stanford)	Cluster ID (Guo)	Lineage (Guo)	Location (Guo)
C1	CD4	C6-GZMA	CD4	T and AL
C2	CD4	C4-CD69	CD4	T
C3	CD4	C4-CD69	CD4	T
C4	CD4	C9-CTLA4	CD4	T
C5	CD8	C4-GZMK	CD8	T
C6	CD8	C4-GZMK	CD8	T
C7	CD8	C5-ZNF683	CD8	T
C8	CD4	C4-CD69	CD4	T
C9	CD4	C7-CXCL13	CD4	T
C10	CD8	C5-ZNF683	CD8	T
C11	CD8	C6-LAYN	CD8	T
C12	CD8	C3-CX3CR1	CD8	*PB and T
C13	CD4	C4-GZMK/C4-CD69	CD8/CD4	T/T
C14	CD4/CD8	-	-	-

C

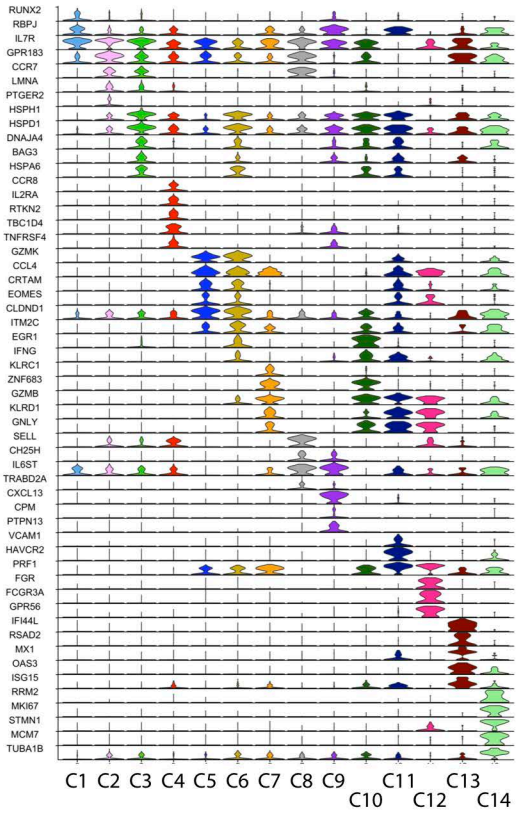
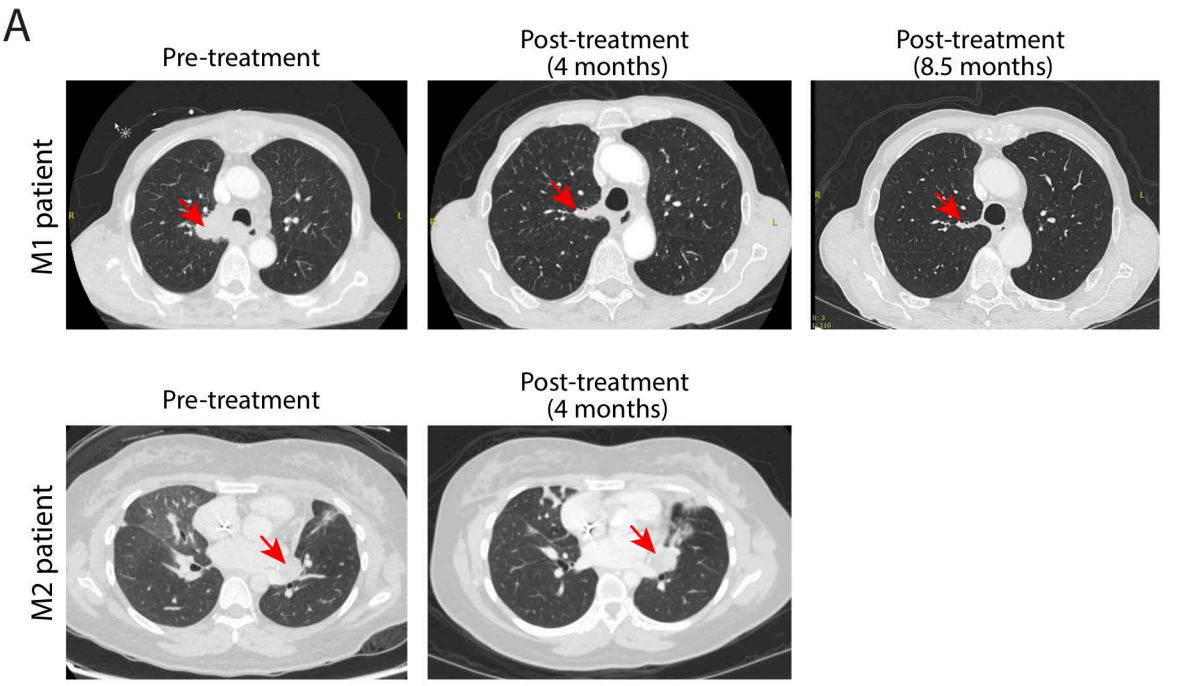
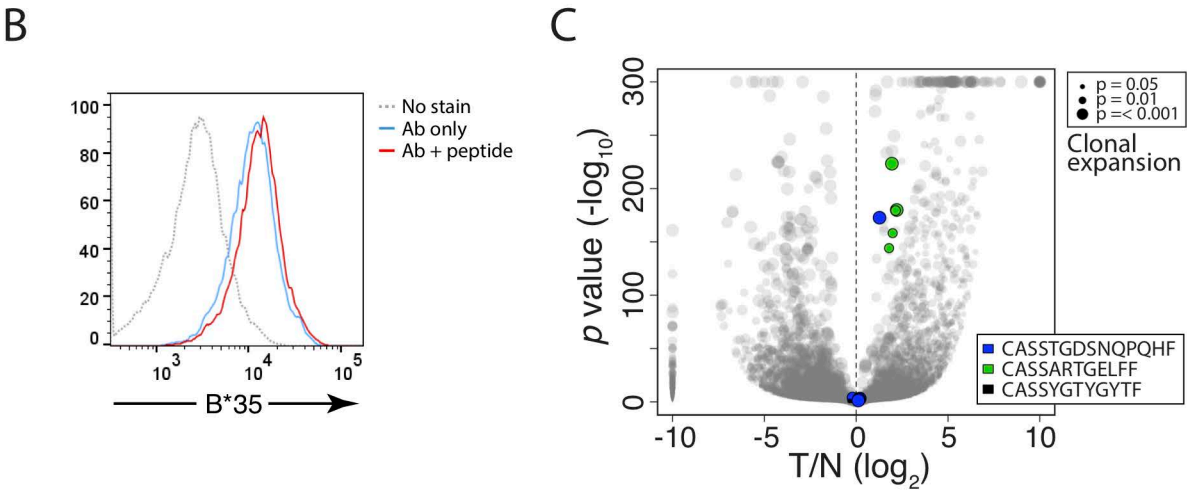


Figure S6, related to Figure 6

(**A**) Dimension reduction by UMAP of the published NSCLC scRNA-Seq results from Zhang and colleagues (Guo *et al*. 2018. Nat. Med. 24, 978). 12,346 sorted T cells from the report by Zhang and colleagues were combined with the 2,950 sorted T cells from the current study (Stanford cohort) for a joint analysis of dimension reduction by UMAP. Cells from Zhang and colleagues are colored according to the identified cell states as reported (left) in comparison with cells from the Stanford cohort colored with the 14 cell states identified in the current study (right). *, cell states identified by Zhang and colleagues (left) that mostly resembled at least one of the cell states identified in the current study (*, right). AL, adjacent lung. (**B**) Cross reference of the 14 cell clusters (C1-C14) identified in the Stanford lung cancer cohort with the clusters reported by Guo *et al*. (Guo *et al*. 2018. Nat. Med. 24, 978). *, CD8 T cells of the C3-CX3CR1 cluster could be found in both the peripheral blood and tumor (Guo *et al*. 2018. Nat. Med. 24, 978). T, tumor; AL, adjacent lung tissue; PB, peripheral blood. (**C**) Selected top differentially expressed genes for each of the identified clusters (n = 14). Expression of indicated genes were shown with the violin plot.

Figure S7


A
Pre-treatment
Post-treatment (4 months)
Post-treatment (8.5 months)
M1 patient
Pre-treatment
Post-treatment (4 months)
M2 patient


B
No stain
Ab only
Ab + peptide
B*35
C
p value (-log10)
T/N (log2)
p = 0.05
p = 0.01
p =< 0.001
Clonal expansion
CASSTGDSNQPQHF
CASSARTGELFF
CASSYGTYGYTF

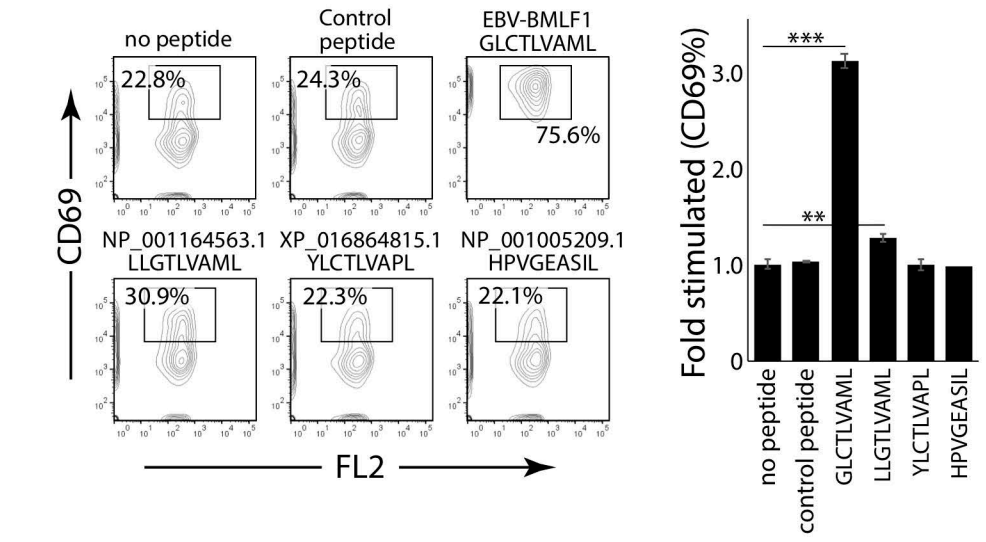

D
no peptide
Control peptide
EBV-BMLF1 GLCTLVAML
22.8%
24.3%
75.6%
NP_001164563.1 LLGTLVAML
XP_016864815.1 YLCTLVAPL
NP_001005209.1 HPVGEASIL
30.9%
22.3%
22.1%
CD69
FL2
Fold stimulated (CD69%)
***
**
no peptide
control peptide
GLCTLVAML
LLGTLVAML
YLCTLVAPL
HPVGEASIL

Figure S7, related to Figure 7

(**A**) CT scan images of pre- and post-treatment from NSCLC patient M1 (top panels) and M2 (bottom panels) treated with anti-PD1 therapy. Tumors are highlighted with red arrowheads. (**B**) T2 (174 x CEM.T2) cells were transduced with lentiviral vector encoding the full-length coding sequence of WT human *HLA-B*35:01*. Cells were selected with puromycin and the surface B*35 expression was quantified by FACS with or without the control peptide "LPFDFTPGY" reported previously (Takamiya *et al.*, 1994. Int Immunol. Vol. 6, 255). (**C**) Volcano plot showing the comparison of the 66,094 shared specificity groups between tumor (T) and the adjacent lung (N) by Poisson test. The y-axis represents the negative $\log_{10}$ converted *p* values of the Poisson tests and the x-axis represents the $\log_2$ converted fold difference between tumor and the adjacent lung (T/N). Dot size represents levels of clonal expansion. Specificity groups annotated with pathogen-related tetramer CDR3β sequences as in Figure 7B (n=11) are highlighted according to the respective CDR3β sequences of the expanded clones (5th column, Figure 7B). (**D**) Left panels, representative FACS plots showing the stimulation of the Jurkat-TCR15 cells with 9mers from the EBV *BMLF1* locus (GLCTLVAML), uniprot NP_001164563.1 (*CLDN2* locus, LLGTLVAML), XP_016864815.1 (*SERINC5* locus, YLCTLVAPL), and NP_001005209.1 (*TMEM198* locus, GLLCGLVAML). Right, results of Jurkat-TCR15 cell stimulation in triplicate. Control peptide: flu M1 "GILGFVFTL". ***, $p < 0.001$; **, $p < 0.01$ by student t test.

Table S1, related to Figure 1.

Patient ID	Lesion type	Gender	Ethnicity	HLA-A	HLA-A	HLA-B	HLA-B	HLA-C	HLA-C
MDA-1150	SCC	F	Caucasian	A*01:01	A*02:01	B*37:01	B*08:01	C*06:02	C*07:01
MDA-1170	SCC	F	Caucasian	A*02:01	A*02:01	B*67:01	B*50:01	C*12:03	C*06:02
MDA-1193	ADC	F	Caucasian	A*26:01	A*24:02	B*07:02	B*07:02	C*05:01	C*07:02
MDA-1206	SCC	M	Caucasian	A*03:01	A*03:01	B*57:01	B*14:02	C*08:01	C*08:02
MDA-1214	SCC	M	Hispanic	A*24:02	A*11:01	B*44:03	B*45:01	C*05:01	C*06:02
MDA-1248	SCC	M	Caucasian	A*31:01	A*02:01	B*35:01	B*07:02	C*07:02	C*04:01
MDA-1264	ADC	M	Caucasian	A*32:01	A*02:01	B*13:02	B*44:02	C*05:01	C*06:02
MDA-1273	SCC	M	Caucasian	A*66:01	A*03:01	B*15:16	B*08:01	C*17:01	C*14:02
MDA-1278	ADC	F	Caucasian	A*02:01	A*02:01	B*18:01	B*44:02	C*07:01	C*12:03
MDA-1283	SCC	M	Caucasian	A*29:02	A*02:51	B*14:02	B*27:05	C*01:02	C*08:02
MDA-1321	SCC	M	Caucasian	A*24:02	A*03:01	B*51:01	B*81:03	C*07:02	C*01:02
MDA-1339	SCC	F	Caucasian	A*03:01	A*01:01	B*42:01	B*07:02	C*07:01	C*07:02
MDA-1346	ADC	F	Caucasian	A*02:60	A*02:03	B*15:01	B*44:09	C*05:01	C*03:03
MDA-1356	ADC	M	Caucasian	A*32:01	A*03:26	B*40:02	B*55:22	C*01:02	C*02:02
MDA-1361	SCC	M	Caucasian	A*25:01	A*02:01	B*40:01	B*44:02	C*03:04	C*05:01
MDA-1366	SCC	M	Caucasian	A*03:26	A*03:04	B*07:02	B*35:01	C*07:02	C*04:01
MDA-1386	SCC	M	Caucasian	A*01:01	A*01:01	B*14:01	B*08:01	C*08:02	C*07:01
MDA-1388	SCC	M	Caucasian	A*29:02	A*03:01	B*35:01	B*44:03	C*16:01	C*04:01
MDA-1392	ADC	F	Caucasian	A*02:01	A*02:01	B*51:01	B*07:02	C*07:02	C*14:02
MDA-1402	SCC	M	African American	A*24:02	A*03:01	B*15:03	B*07:02	C*02:10	C*07:02
MDA-1409	SCC	M	Caucasian	A*24:02	A*01:14	B*08:01	B*55:01	C*01:02	C*07:01
MDA-1413	ADC	M	Caucasian	A*30:04	A*68:01	B*18:01	B*27:05	C*05:01	C*02:02
MDA-1420	SCC	F	Caucasian	A*31:01	A*24:02	B*35:08	B*07:02	C*07:02	C*04:01
MDA-1425	SCC	M	Caucasian	A*25:01	A*25:01	B*08:01	B*18:01	C*12:03	C*07:01
MDA-1429	SCC	M	Caucasian	A*30:04	A*02:01	B*07:02	B*18:01	C*05:01	C*07:02
MDA-1450	ADC	F	Caucasian	A*26:01	A*02:03	B*35:01	B*44:02	C*05:08	C*04:01
MDA-1453	ADC	F	Caucasian	A*02:01	A*03:01	B*07:02	B*44:02	C*05:01	C*07:02
MDA-1455	SCC	M	Caucasian	A*02:01	A*02:01	B*07:02	B*44:09	C*05:01	C*07:02
MDA-1463	SCC	M	Caucasian	A*31:01	A*02:01	B*40:01	B*44:02	C*05:01	C*03:04
MDA-1489	SCC	M	Caucasian	A*31:01	A*01:01	B*08:01	B*18:01	C*05:01	C*07:01
MDA-1514	ADC	F	Caucasian	A*26:01	A*02:01	B*50:01	B*27:05	C*01:02	C*06:02
MDA-1517	SCC	F	Caucasian	A*26:01	A*02:03	B*07:02	B*27:05	C*07:02	C*01:02
MDA-1520	SCC	M	Caucasian	A*29:02	A*29:05	B*51:01	B*51:42	C*15:02	C*05:01
MDA-1530	SCC	M	Caucasian	A*03:01	A*02:01	B*07:02	B*07:02	C*07:02	C*07:02
MDA-1532	SCC	M	Caucasian	A*02:01	A*02:01	B*51:01	B*57:02	C*16:02	C*07:01
MDA-1549	SCC	M	Caucasian	A*11:01	A*24:02	B*15:02	B*15:02	C*08:01	C*08:01
MDA-1550	ADC	F	Caucasian	A*02:02	A*02:01	B*13:02	B*15:01	C*06:02	C*03:03
MDA-1556	SCC	F	Caucasian	A*02:01	A*02:01	B*18:03	B*35:01	C*07:01	C*04:01
MDA-1560	ADC	F	African American	A*03:01	A*03:01	B*07:02	B*57:01	C*06:02	C*07:02
MDA-1570	SCC	M	African American	A*68:02	A*29:02	B*07:02	B*07:02	C*07:02	C*16:01
MDA-1587	SCC	M	Caucasian	A*02:01	A*01:01	B*57:01	B*44:02	C*05:01	C*06:02
MDA-1600	SCC	M	Caucasian	A*11:01	A*02:01	B*35:01	B*44:03	C*02:02	C*04:01
MDA-1612	SCC	F	Caucasian	A*02:266	A*01:01	B*49:01	B*57:01	C*06:02	C*12:03
MDA-1615	SCC	F	African American	A*29:02	A*29:02	B*81:01	B*07:02	C*18:01	C*14:02
MDA-1636	ADC	M	Caucasian	A*32:01	A*11:01	B*15:01	B*14:01	C*08:02	C*03:03
MDA-1649	ADC	M	Caucasian	A*24:02	A*03:01	B*46:01	B*14:02	C*08:02	C*03:13
MDA-1660	SCC	M	Caucasian	A*31:01	A*01:01	B*27:05	B*08:01	C*01:02	C*07:01
MDA-1664	SCC	F	Caucasian	A*26:01	A*11:01	B*38:01	B*52:01	C*12:03	C*12:02
MDA-1708	ADC	M	Caucasian	A*02:05	A*02:05	B*13:02	B*53:01	C*06:02	C*07:18
MDA-1736	ADC	F	Caucasian	A*29:02	A*11:01	B*35:01	B*44:03	C*04:01	C*16:01
MDA-1742	ADC	F	Caucasian	A*01:01	A*24:02	B*15:17	B*35:03	C*07:01	C*12:03
MDA-1743	ADC	M	Caucasian	A*02:68	A*24:02	B*44:02	B*44:03	C*03:03	C*05:01
MDA-1756	SCC	M	Caucasian	A*32:01	A*02:01	B*15:18	B*40:02	C*07:04	C*02:02
MDA-1758	ADC	F	Caucasian	A*02:01	A*01:01	B*51:01	B*08:01	C*07:01	C*15:02
MDA-1769	ADC	M	Caucasian	A*03:01	A*02:01	B*07:02	B*15:01	C*03:03	C*07:02
MDA-1772	SCC	M	Caucasian	A*02:01	A*24:02	B*40:79	B*44:02	C*05:01	C*03:04
MDA-1774	ADC	M	Caucasian	A*02:01	A*02:05	B*49:01	B*15:01	C*03:03	C*07:01
MDA-1775	ADC	F	Caucasian	A*02:01	A*29:02	B*51:01	B*44:03	C*15:02	C*16:01
MDA-1785	SCC	F	Caucasian	A*02:01	A*02:03	B*08:01	B*44:02	C*05:01	C*07:01
MDA-1793	ADC	M	Caucasian	A*24:02	A*24:03	B*46:01	B*51:01	C*14:02	C*01:02
MDA-1818	ADC	F	Caucasian	-	-	B*35:01	B*08:01	C*07:01	C*04:01
MDA-1822	SCC	M	Caucasian	A*02:03	A*02:01	B*41:01	B*44:46	C*07:01	C*05:01
MDA-1823	ADC	M	Caucasian	A*24:03	A*01:01	B*42:01	B*07:02	C*07:01	C*07:02
MDA-1833	ADC	F	Caucasian	A*30:01	A*11:01	B*13:02	B*35:01	C*06:02	C*04:01
MDA-1837	ADC	F	Caucasian	A*03:01	A*03:01	B*35:01	B*07:02	C*07:02	C*04:01
MDA-1845	ADC	F	Caucasian	A*68:02	A*29:02	B*45:01	B*45:01	C*06:02	C*16:01
MDA-1848	ADC	M	Caucasian	A*68:01	A*68:01	B*37:01	B*40:01	C*06:02	C*03:04
MDA-1853	ADC	M	Caucasian	A*03:01	A*03:01	B*07:02	B*07:02	C*07:02	C*07:02
MDA-1872	ADC	M	Caucasian	A*68:01	A*74:01	B*53:01	B*07:02	C*15:05	C*04:01
MDA-1875	ADC	F	Caucasian	A*01:01	A*24:02	B*08:01	B*39:06	C*07:02	C*07:01
MDA-1893	SCC	M	Caucasian	A*02:01	A*02:01	B*57:01	B*35:02	C*04:01	C*06:02
MDA-1896	SCC	M	Caucasian	A*03:26	A*02:01	B*14:02	B*35:03	C*04:01	C*08:02
MDA-1904	ADC	M	Caucasian	A*03:01	A*23:01	B*14:02	B*49:01	C*08:02	C*07:01
MDA-1905	ADC	M	Caucasian	A*02:01	A*02:06	B*07:02	B*07:02	C*03:05	C*07:02
MDA-1906	ADC	M	Caucasian	A*23:01	A*01:01	B*08:01	B*49:01	C*07:01	C*07:01
MDA-1908	SCC	F	Caucasian	A*02:68	A*23:01	B*35:01	B*44:03	C*04:01	C*04:01
MDA-1913	SCC	M	Caucasian	A*02:01	A*29:02	B*44:03	B*18:01	C*07:01	C*16:01
MDA-1914	ADC	M	Caucasian	A*03:01	A*03:01	B*07:02	B*18:01	C*07:02	C*12:03
MDA-1928	SCC	F	Caucasian	A*68:02	A*68:02	B*44:03	B*08:01	C*07:01	C*16:01
MDA-1930	SCC	M	Caucasian	A*32:01	A*32:01	B*44:02	B*41:01	C*17:01	C*05:01
MDA-1937	ADC	M	Caucasian	A*32:01	A*32:01	B*14:02	B*46:01	C*03:03	C*08:02
MDA-1950	SCC	M	Caucasian	A*02:07	A*24:02	B*44:09	B*44:02	C*07:04	C*17:03
MDA-1960	ADC	M	Caucasian	A*31:01	A*29:02	B*08:01	B*07:02	C*07:01	C*07:74
MDA-1971	ADC	M	African American	A*02:01	A*11:01	B*35:01	B*51:01	C*15:02	C*04:01
MDA-1991	SCC	F	Caucasian	A*01:01	A*01:01	B*07:02	B*14:08	C*07:02	C*08:02
MDA-2004	ADC	M	Caucasian	A*03:01	A*03:01	B*39:06	B*35:01	C*07:02	C*04:01
MDA-2016	ADC	F	Caucasian	A*02:01	A*31:01	B*35:03	B*07:02	C*04:01	C*07:02
MDA-2017	ADC	F	Caucasian	A*01:01	A*36:01	B*53:01	B*07:02	C*07:02	C*04:01
MDA-2021	ADC	M	Hispanic	A*34:08	A*33:03	B*35:01	B*35:01	C*06:02	C*16:01
MDA-2022	ADC	M	Caucasian	A*03:01	A*03:01	B*55:02	B*46:01	C*03:03	C*03:04
MDA-2025	ADC	M	Caucasian	A*33:03	A*30:01	B*53:01	B*42:01	C*17:01	C*04:01
MDA-2028	ADC	F	Caucasian	A*11:01	A*11:01	B*15:18	B*44:02	C*05:01	C*08:01
MDA-2031	SCC	M	African American	A*02:01	A*03:01	B*35:01	B*35:01	C*06:02	C*16:01
MDA-2034	SCC	M	Caucasian	A*03:56	A*03:01	B*07:02	B*07:02	C*06:02	C*07:02
MDA-2037	ADC	F	Caucasian	A*32:01	A*24:02	B*49:01	B*35:01	C*07:01	C*04:01
MDA-2040	SCC	M	Caucasian	A*24:02	A*24:02	B*41:01	B*07:02	C*07:01	C*07:02
MDA-2046	ADC	M	Caucasian	A*03:01	A*02:01	B*40:01	B*44:02	C*05:01	C*07:02
MDA-2052	ADC	M	Caucasian	A*02:01	A*68:01	B*38:01	B*15:01	C*12:03	C*04:01

MDA-2054	ADC	F	African American	A*23:01	A*11:01	B*14:02	B*35:01	C*08:02	C*04:01
MDA-2057	SCC	M	Caucasian	A*31:01	A*36:01	B*08:12	B*40:01	C*03:04	C*07:01
MDA-2060	ADC	M	Caucasian	A*01:01	A*01:01	B*08:01	B*08:01	C*06:02	C*07:01
MDA-2061	SCC	F	Caucasian	A*68:01	A*68:01	B*51:01	B*08:01	C*01:02	C*07:01
MDA-2068	SCC	F	Caucasian	A*26:01	A*02:60	B*18:01	B*07:02	C*01:02	C*07:02
MDA-2070	ADC	M	Caucasian	A*03:56	A*03:01	B*51:01	B*35:01	C*04:01	C*15:02
MDA-2080	ADC	F	Caucasian	A*02:01	A*01:01	B*08:01	B*18:01	C*07:01	C*07:01
MDA-2082	ADC	M	Caucasian	A*02:01	A*01:01	B*07:02	B*15:01	C*03:04	C*07:02
MDA-2088	SCC	M	Caucasian	A*11:01	A*02:01	B*55:02	B*55:01	C*05:01	C*03:03
MDA-2090	SCC	M	Caucasian	A*01:01	A*03:56	B*07:02	B*15:17	C*07:01	C*07:01
MDA-2093	ADC	M	African American	A*11:01	A*02:01	B*15:01	B*35:01	C*03:03	C*04:01
MDA-2095	ADC	F	African American	A*02:05	A*03:01	B*50:01	B*07:02	C*06:02	C*07:02
MDA-2097	ADC	M	Caucasian	A*02:05	A*02:51	B*35:02	B*35:01	C*04:01	C*07:18
MDA-2108	SCC	M	Caucasian	A*02:05	A*30:01	B*18:01	B*57:01	C*12:03	C*06:02
MDA-2115	ADC	M	African American	A*03:04	A*03:63	B*51:02	B*51:01	C*15:02	C*03:04
MDA-2118	ADC	F	Asian or Pacific Islander	A*24:02	A*01:01	B*08:01	B*15:01	C*01:02	C*07:01
MDA-2121	SCC	F	Caucasian	A*03:01	A*23:01	B*35:01	B*35:01	C*04:01	C*04:01
MDA-2122	SCC	M	Caucasian	A*31:01	A*31:01	B*27:05	B*40:01	C*03:04	C*02:02
MDA-2139	SCC	M	Caucasian	A*02:01	A*03:26	B*27:02	B*07:02	C*07:02	C*02:02
MDA-2140	SCC	M	Caucasian	A*25:01	A*01:01	B*18:01	B*18:13	C*12:03	C*12:03
MDA-2150	ADC	F	Caucasian	A*03:01	A*03:01	B*27:05	B*07:02	C*07:02	C*01:02
MDA-2152	SCC	M	Caucasian	A*68:01	A*68:01	B*35:01	B*44:02	C*02:02	C*04:01
MDA-2155	SCC	M	Caucasian	A*02:01	A*02:01	B*08:01	B*14:02	C*08:02	C*07:01
MDA-2161	SCC	M	Caucasian	A*25:01	A*02:01	B*18:01	B*44:02	C*05:01	C*12:03
MDA-2175	ADC	M	Caucasian	A*02:01	A*03:01	B*07:02	B*15:01	C*03:04	C*07:02
MDA-2176	ADC	M	Caucasian	A*02:01	A*03:01	B*07:02	B*18:01	C*07:02	C*12:13
MDA-2195	SCC	F	Caucasian	A*33:01	A*03:01	B*14:02	B*08:01	C*08:02	C*07:01
MDA-2214	ADC	F	Caucasian	A*26:01	A*26:01	B*38:01	B*57:01	C*06:02	C*12:03
MDA-2221	ADC	M	Caucasian	A*24:02	A*02:01	B*08:01	B*14:02	C*08:02	C*07:01
MDA-2226	SCC	M	Caucasian	A*31:01	A*02:01	B*35:03	B*35:03	C*04:01	C*04:01
MDA-2230	ADC	M	Caucasian	A*30:02	A*03:01	B*53:01	B*07:02	C*07:19	C*07:01
MDA-2237	ADC	F	Caucasian	A*11:05	A*33:01	B*14:02	B*14:02	C*08:02	C*08:02
MDA-2240	ADC	F	Caucasian	A*26:01	A*29:02	B*27:05	B*44:03	C*01:02	C*16:01
MDA-2243	ADC	M	African American	A*03:01	A*24:02	B*27:05	B*07:02	C*07:02	C*01:02
MDA-2244	ADC	M	Caucasian	A*02:01	A*03:01	B*15:01	B*07:02	C*07:02	C*07:39
MDA-2254	ADC	M	Caucasian	A*26:01	A*11:01	B*51:01	B*15:01	C*15:02	C*04:01
MDA-2262	SCC	M	Caucasian	A*03:01	A*03:01	B*08:01	B*07:02	C*07:02	C*07:01
MDA-2264	SCC	F	Caucasian	A*02:03	A*02:03	B*07:02	B*07:02	C*06:02	C*07:02
MDA-2265	SCC	F	Caucasian	A*03:01	A*03:01	B*07:02	B*18:01	C*05:01	C*07:02
MDA-2271	ADC	F	Caucasian	A*11:01	A*11:01	B*08:01	B*44:02	C*05:01	C*07:01
MDA-2276	ADC	F	Caucasian	A*01:01	A*02:01	B*44:02	B*44:02	C*05:01	C*05:01
MDA-2286	ADC	M	Caucasian	A*24:02	A*02:03	B*40:01	B*51:01	C*03:04	C*01:02
MDA-2292	ADC	M	Caucasian	A*32:01	A*03:01	B*07:02	B*55:01	C*03:03	C*03:03
MDA-2294	ADC	F	Caucasian	A*03:01	A*03:01	B*51:01	B*08:01	C*01:02	C*07:01
MDA-2300	SCC	M	Caucasian	A*24:02	A*24:03	B*08:01	B*08:01	C*07:01	C*07:01
MDA-2307	ADC	M	Caucasian	A*02:01	A*01:01	B*15:01	B*51:01	C*16:02	C*03:04
MDA-2315	ADC	F	Caucasian	A*02:01	A*11:01	B*35:01	B*44:02	C*05:01	C*04:01
MDA-2316	ADC	M	African American	A*03:01	A*03:01	B*08:01	B*35:01	C*07:01	C*04:01
MDA-2318	ADC	F	Caucasian	A*30:26	A*03:01	B*39:01	B*13:02	C*06:02	C*07:02
MDA-2320	ADC	F	Caucasian	A*02:01	A*02:01	B*15:01	B*44:09	C*03:03	C*05:01
MDA-2325	SCC	F	Caucasian	A*26:01	A*24:10	B*18:01	B*13:02	C*06:02	C*07:01
MDA-2333	ADC	M	Caucasian	A*32:01	A*24:02	B*07:02	B*07:02	C*07:02	C*02:02
MDA-2335	ADC	F	Caucasian	A*25:01	A*25:01	B*08:01	B*57:01	C*06:02	C*07:01
MDA-2337	ADC	F	Caucasian	A*24:02	A*02:01	B*15:01	B*15:01	C*03:03	C*03:04
MDA-2343	ADC	M	Caucasian	A*68:02	A*30:02	B*15:10	B*35:01	C*06:02	C*03:02
MDA-2344	SCC	F	Caucasian	A*11:01	A*02:03	B*08:01	B*18:01	C*07:01	C*07:01
MDA-2345	ADC	M	Caucasian	A*01:01	A*01:01	B*08:01	B*08:01	C*07:01	C*07:01
MDA-2348	ADC	F	Caucasian	A*02:01	A*02:01	B*07:02	B*07:02	C*05:01	C*07:02
MDA-2352	ADC	F	Caucasian	A*03:01	A*03:01	B*51:01	B*07:02	C*07:02	C*14:02
MDA-3203	ADC	M	Caucasian	A*02:01	A*11:01	B*35:01	B*07:02	C*04:01	C*07:02
MDA-3212	ADC	F	Caucasian	A*30:01	A*02:01	B*13:02	B*15:01	C*03:61	C*06:02
MDA-3214	ADC	F	Caucasian	A*02:01	A*02:89	B*50:01	B*38:01	C*04:01	C*12:03
MDA-3234	SCC	F	Caucasian	A*03:01	A*02:01	B*51:01	B*35:01	C*02:02	C*04:01
MDA-3237	SCC	F	Caucasian	A*24:02	A*24:02	B*15:07	B*40:01	C*03:03	C*03:04
MDA-3238	ADC	M	African American	A*33:03	A*30:02	B*53:01	B*15:10	C*03:04	C*04:01
MDA-3242	ADC	M	Caucasian	A*25:01	A*25:01	B*15:17	B*18:01	C*07:01	C*12:03
MDA-3253	SCC	F	Caucasian	A*03:01	A*02:01	B*07:02	B*07:02	C*07:02	C*07:02
MDA-3259	ADC	F	Caucasian	A*01:01	A*01:01	B*08:01	B*07:02	C*07:01	C*07:01
MDA-3260	ADC	M	Caucasian	A*29:02	A*29:02	B*44:03	B*57:01	C*16:04	C*06:02
MDA-3273	ADC	M	Asian or Pacific Islander	A*24:02	A*02:01	B*40:06	B*40:01	C*15:02	C*08:01
MDA-3290	SCC	F	Caucasian	A*02:251	A*24:02	B*07:02	B*27:05	C*01:02	C*07:02
MDA-4309	ADC	F	Caucasian	A*02:01	A*02:01	B*27:07	B*08:01	C*01:02	C*07:01
MDA-4347	ADC	F	Caucasian	A*03:01	A*03:02	B*08:01	B*15:01	C*03:04	C*07:39
MDA-4361	ADC	M	Hispanic	A*68:03	A*11:01	B*39:05	B*40:138	C*03:04	C*07:02
MDA-4363	ADCA/SCC	F	Caucasian	A*11:01	A*02:01	B*35:01	B*57:01	C*06:02	C*04:01
MDA-4378	ADC	F	Caucasian	A*29:02	A*11:01	B*51:01	B*44:03	C*15:02	C*16:01
MDA-4382	SCC	M	Hispanic	A*31:01	A*11:01	B*52:01	B*39:06	C*07:02	C*12:02
MDA-4383	ADC	M	Caucasian	A*02:01	A*02:01	B*40:01	B*57:01	C*06:02	C*03:04
MDA-4385	ADC	F	Caucasian	A*26:01	A*26:01	B*37:01	B*08:01	C*06:02	C*07:01
MDA-4386	ADC	M	Caucasian	A*03:01	A*03:01	B*08:01	B*18:01	C*07:01	C*07:01

HLA class I genotypes and clinical attributes for the MDACC NSCLC cohort (Reuben *et al*., 2020 Nat Commun). ADC, adenocarcinoma; SCC, squamous cell carcinoma.

Table S4, related to Figure 2.

Patient ID	scTCR-seq	scRNA-Seq	Histology (tumor)	Histology (adjacent lung tissue)	Exome seq	Age	Gender	HLA-A*02	Smoking (pk yr)	Stage	Histology
A2	yes					65	M	N	50	IB	SCC
A3	yes					71	M	Y	50	IB	SCC
A4	yes		yes			49	M	Y	0	IA	adeno
A5	yes		yes	yes		65	F	N	0	IB	adeno
A6*	yes		yes	yes		73	F	Y	45	IB	SCC
A7			yes	yes		64	M	Y	118	IB	adeno
A8			yes	yes		71	F	Y	60	IB	adeno
A9			yes			77	M	N	45	IB	adeno
A10			yes	yes		82	F	N	0	IIA	adeno
A11	yes	yes	yes	yes	yes	76	F	Y	50	IIIA	adeno
A14			yes	yes		62	F	Y	10	IA	adeno
A15	yes	yes	yes	yes	yes	71	F	N	3	IB	adeno
A16	yes	yes	yes	yes	yes	68	F	N	20	IB	adeno
A17	yes	yes	yes	yes	yes	65	F	Y	0	IA	adeno
A18	yes	yes	yes	yes		65	M	N	30	IB	adeno
A19	yes	yes	yes	yes	yes	70	F	N	37	IIA	adeno
A21			yes	yes		31	M	N	0	IIIA	adeno
A28	yes	yes	yes	yes	yes	64	M	N	24	IA	adeno
A29	yes	yes	yes	yes	yes	64	F	Y	0	IA	adeno
A31	yes	yes	yes	yes	yes	68	F	Y	0	IIIB	adeno
A32	yes	yes	yes	yes	yes	49	F	N	0	IB	adeno
Total # with yes	15	10	19	17	9	-	-	-	-	-	-

Data availability and clinical attributes of the Stanford NSCLC cohort. Single-cell TCR-seq was performed on 15 patients, single-cell RNA-seq was performed on 10 patients, histology was performed on tumors from 19 patients and the adjacent lung tissues from 17 patients, and 9 NSCLC patients have tumors assessed with whole-exome sequencing (yes, data available). A6*, the paired TCR CDR3α/β sequences from the tumor of patient A6 were derived by using the Chromium Single Cell Immune Profiling Solution for human T cell repertoires from the 10x Genomics. M, male; F, female; N, HLA-A*02^{-}; Y, HLA-A*02^{+};adeno, adenocarcinoma; SCC, squamous cell carcinoma.

Table S5, related to Figure 3.

Rank	Peptide seq (Naive)	#	Peptide seq (Round 1)	#	Peptide seq (Round 2)	#	Peptide seq (Round 3)	#	Peptide seq (Round 4)	#
1	SMQGLTRPPRV	43	VMRQLACSRRL	25	RLHVAPGQRRM	149	AMGGLLTQLAM	14764	AMGGLLTQLAM	70220
2	AMGGLLTQLAM	30	FLQNGRSSERM	24	RLREAAPILPM	108	EMRVASAMWGM	1205	KLGGLLTMVGV	23704
3	ALRTPRKKQTL	17	RLQSRKLPPLL	22	PMDTHLPTSVL	89	GMVEDICSTLL	712	GMVEDICSTLL	18950
4	CLTQSSQEWTL	16	TLARTHHTRRL	20	AMQERLGTRQV	88	RLHEQVPEVTL	557	EMRVASAMWGM	11431
5	KLDSANAQAIV	13	HMHRPRPSGRL	19	WLTAALPSCGV	88	HLRVKEHRHGL	469	RLHEQVPEVTL	3165
6	PMLTRSPLSRM	10	DLLVQKITRHL	18	CLPWQKTESPL	81	HMHHHNEHNRM	468	KMKMKHHTRPL	2760
7	WMRQMEPSGRL	10	RLNELRLAKGL	18	RLTRTQGSETL	78	RLHVAPGQRRM	315	HMHHHNEHNRM	1229
8	GMLASLAARQV	10	RLHKAPRPAPL	18	NMEALGHCHRL	77	PLACAKRRGGM	280	ALNLPSSPPSL	1019
9	EMTLPQTAKWL	9	DLQSTLEAWAV	17	RMEQVATARVL	73	NLKLMGIWIEM	268	KMPHRPKHKVV	987
10	WLSLAAPPPAL	9	RMSSRPRSLLL	17	RLAEPSHWIRL	67	ALNLPSSPPSL	264	HLRVKEHRHGL	968
11	RLWQRAPVTRL	9	GMRTRGPWRHL	17	RLHVRPGPRPL	67	NLRHRHPSRFL	262	HMHEHRVKGKV	841
12	GLGIAPALTDL	9	HMTGKHRTHMV	17	MMRTGPLLGLL	67	SMPKRGTQVTM	261	RLHVRPGPRPL	834
13	RMPTRRPTKRM	9	AMHRNSRWHSL	16	RMLTRSPRLAL	67	HMHEHRVKGKV	260	WLETPTTPLLL	818
14	PLAHTPPRRTL	9	ALSARLPRQAL	16	HMSRTHAPGQL	67	RLHVRPGPRPL	259	PLACAKRRGGM	573
15	TLMRLHAGVVL	9	RLHAPTHHHSL	16	MMSCPAQEPCL	67	ALREKPRLPRL	257	NLRHRHPSRFL	459
16	RMGPLALSPSM	9	ALSRARLSRPV	16	TMKTQGQETSM	66	RLPARHPPEAL	255	PLTSARKKPRM	396
17	AMPNPRVREHL	9	RLPPLPQTWPL	15	RLAHPRARQQV	65	RLPEPPPARPL	249	RLHGRPTRWNL	382
18	AMLTLTRGQQM	9	PLTPTPRPGSM	15	RMSGKPQQHPL	65	RLLAKVPQGTL	248	HMPRHRWAARV	359
19	PLPPPLPRPPM	9	PLLATARTTWV	15	NLLPWRPTKLM	63	QLRLASTPRGL	247	SMGGLLTQLAM	350
20	RMTMTARRRPV	9	ALEASWEWTGM	15	VMRTPRRPVPL	63	QMTARGRVHLL	234	RMPAAARKRGV	326
21	GLLQLMMSLAV	9	VLPALHAPGPL	15	KLAAGKQGLAM	63	RMACGQHPGPL	218	RLPARHPPEAL	314
22	RLPLEPHRQCM	8	ELARSHRAWSM	14	RLRAHQHMFRV	61	RMPAAARKRGV	217	HMRGGRHSRLM	302
23	LLASPTGVTKM	8	EMTCQLGPKPL	14	RLPKPPQVPAL	61	RMLTRSPRLAL	216	KMKSHHLRWGL	299
24	RLQVLPNSPCV	8	GLPPSAARKVL	14	NLKLMGIWIEM	60	KMHQHGQHRQL	214	PMEPSPCSTGV	293
25	KLHPNTCAKGM	8	TLKSHHHWGQL	14	PLRPPSLPTKM	60	RLHGRPTRWNL	212	KMHQHGQHRQL	286
26	TLSPSTGTVSM	8	RLIPWRTTLLL	14	VLGWRDEPGTV	60	LLPLAPPRRGV	193	HLMGTRGHRCL	282
27	RLCARMWNQSV	8	RMPPTPAPKGM	14	EMLTGAVGTPL	60	HLHSHARTPRM	190	ALREKPRLPRL	268
28	AMRPPSMATRL	8	HMPRDQNRVAL	14	PMAIHHPSSQL	59	KLGGLLTMVGV	188	HLRLRHHHREL	267
29	ALPLRPARLSL	8	PLHGRQLLRAL	14	TLQEQTARTRM	59	QLSGAHHHRAL	187	RMPRPPIPSPL	263
30	RLRWIMLRTAL	8	TLMKRPGLRGL	14	RMHARMTSLKL	59	RLHHPHQLRPM	178	RLPRPHMLPRL	260
31	CLESLCRAMQL	8	LLPPTTPPHAV	14	AMPRAPRSSWV	59	HLMGTRGHRCL	175	HMARPGPHRHL	256
32	PLPLTQRTPAV	8	RLARSPQQHSL	14	GLSGPCPQPAL	58	RMPAQRVAQAM	170	HLHPHHPPGAL	247
33	NLRQREGTRIL	7	RMDHMPWHVRL	14	RLQMSSRRHHV	58	KLNGQVSERML	169	HMHRPRPSGRL	241
34	RLPSLERWPRL	7	EMPMPPAAQPL	14	RMTPCLQAEAL	58	TLTPAQPPVLL	168	HMQHRGTHTKV	229
35	GMWARSVRASL	7	HLSKPRRRHEL	14	TMPPPGLWQAV	58	LMCTMAAASRL	168	HMPKPRTRVPL	208
36	ALRGVRRTRVL	7	ALCLASREVLM	14	RLAPPRLPHIV	57	RLPTEMGTPPL	167	HLHSHARTPRM	206
37	TLLVSTLRGCL	7	TMQHGAQQTAL	13	RLPAPHRPETV	57	MLKRPRTPRQV	165	MLFDGLPPLGV	202
38	RLSQSHRPRPL	7	TMRTCTERTPM	13	RLMWAKQEMLL	57	KMKMKHHTRPL	164	RLHVAPGQRRM	193
39	WMRVWREPVRM	7	EMAPHAMLERV	13	RLSRPPLPTGV	56	HLRLRHHHREL	162	SMPKRGTQVTM	186
40	QLSAPLAPRSL	7	NLTWLPTATRL	13	RLPMHPPALAV	56	PLSAWTATASL	162	RMPPPPPRLAL	184
41	LLRGQRLEHLL	7	RLHQPPRPLPL	13	RLPEPPGPAAV	56	LLQEEHRHPTV	162	RLHHRVVLPPL	179
42	CLLTTTPPTAL	7	RLDKVPRPGPL	13	RMARHRPGLQL	56	RMPRPPIPSPL	160	QMTARGRVHLL	177
43	QLTRWRHSKGM	7	GLQRVTGHRPL	13	RMACGQHPGPL	56	RLHKSRPWRSV	160	RLPLLPAPRLV	177
44	HLPWLPLHGRM	7	KLAPPNPRPPL	13	KLFSGKSGQVL	55	RLPARCRLGAV	160	RLPEPPGPAAV	172
45	PMAFRGWHTWM	7	HMRRPPPVRLL	13	RLPARPGPLRV	55	RMPPSLQSGCL	159	TLAPPPAPRPL	169
46	TMHGPRSRRAL	7	ELPGCAPQCLM	13	GMSMGPSQCRL	55	RLRPGSNAVLM	153	TLPLHQPPKLV	169
47	SLAWRAGAMDM	7	PLGPPRLRLPL	13	DMTSRLVRWAM	54	HMPRHRWAARV	153	HLARPPPPARL	164
48	RLLNPKGTSAM	7	SLPSAPRHGNL	13	ELRRAEELRNV	54	ALAPTRATRRV	151	RMRAFHPHTRL	164
49	RLSKSLRHTCM	7	PMPAPPLTLPL	13	QLRLASTPRGL	54	DLLPQPWARGM	149	RMRFHSPHKQL	162
50	AMMQSLPAPGM	7	VLLPPLESPPL	13	KMRRTGTLLDV	54	PMVSGEHRGAV	149	AMARPPPRPAL	161

Summary of results from each round of affinity selection on a 11mer yeast HLA-A*02 library with TCR2. Numbers (#) indicate counts of the top-50 mimotopes. Top-two mimotopes from the round 4 screen were highlighted.

Table S6, related to Figure 4.

Sample ID	Gene ID	Unique mut ID	Var class	Var type	AA var	coverage	Ref depth	Alt depth	af	dbsnp	exac	counts_cosmic	filter
A6	ABCA13	chr7:48313276:C>G	exonic	nonsense	S1338X	74	64	10	0.144	FALSE	FALSE	0	PASS
A6	ABCC11	chr16:48201528:G>T	exonic	nonsynonymous_snv	T1312K	143	119	24	0.172	FALSE	FALSE	0	PASS
A6	ADAM7	chr8:24346811:G>A	exonic	nonsynonymous_snv	D411N	71	55	16	0.233	FALSE	FALSE	0	PASS
A6	ADGRB3	chr6:69653730:G>C	exonic	nonsynonymous_snv	V347L	74	64	10	0.145	FALSE	FALSE	0	PASS
A6	AKAP9	chr7:91707139:G>A	exonic	nonsynonymous_snv	E2291K	110	96	14	0.133	FALSE	FALSE	0	PASS
A6	ALS2	chr2:202622283:C>A	exonic	nonsynonymous_snv	S438I	92	76	16	0.18	FALSE	FALSE	0	PASS
A6	ANK2	chr4:114274311:G>C	exonic	nonsynonymous_snv	D1513H	85	68	17	0.206	FALSE	FALSE	0	PASS
A6	ANKFN1	chr17:54520251:G>T	exonic	nonsynonymous_snv	W355C	86	75	11	0.136	FALSE	FALSE	0	PASS
A6	ANKLE2	chr12:133327337:C>T	exonic	nonsynonymous_snv	E247K	265	200	65	0.246	FALSE	FALSE	0	PASS
A6	ANKS1B	chr12:99640454:T>G	exonic	nonsynonymous_snv	K649Q	83	72	11	0.14	FALSE	FALSE	0	PASS
A6	ARHGEF17	chr11:73021562:A>C	exonic	nonsynonymous_snv	T627P	360	290	70	0.196	FALSE	FALSE	0	PASS
A6	ARL2	chr11:64785853:G>A	exonic	nonsynonymous_snv	G28E	318	265	53	0.168	FALSE	FALSE	0	PASS
A6	ASB9	chrX:15266932:C>T	exonic	nonsynonymous_snv	G232S	204	167	37	0.184	FALSE	FALSE	0	PASS
A6	ASIC2	chr17:32483281:G>T	exonic	nonsynonymous_snv	L91I	513	426	87	0.17	FALSE	FALSE	0	PASS
A6	ATP13A5	chr3:193048947:T>C	exonic	nonsynonymous_snv	M476V	115	71	44	0.385	FALSE	FALSE	0	PASS
A6	ATP4B	chr13:114312422:C>T	exonic	nonsynonymous_snv	R13H	685	631	54	0.079	TRUE	TRUE	0	PASS
A6	ATP7A	chrX:77296175:G>A	exonic	nonsynonymous_snv	E1249K	235	211	24	0.104	FALSE	FALSE	0	PASS
A6	AUTS2	chr7:69364300:G>A	exonic	nonsynonymous_snv	R113H	80	52	28	0.353	TRUE	TRUE	3	PASS
A6	BAI3	chr6:69653730:G>C	exonic	nonsynonymous_snv	V347L	74	64	10	0.135	FALSE	FALSE	0	PASS
A6	BCOR	chrX:39933652:G>C	exonic	nonsynonymous_snv	P316R	654	593	61	0.094	FALSE	FALSE	0	PASS
A6	BLVRA	chr7:43830961:A>T	exonic	nonsynonymous_snv	Y83F	73	61	12	0.173	FALSE	FALSE	0	PASS
A6	C1orf173	chr1:75065441:C>T	exonic	nonsynonymous_snv	R555H	67	58	9	0.134	TRUE	TRUE	1	PASS
A6	C20orf141	chr20:2795930:C>A	exonic	nonsynonymous_snv	R34S	498	421	77	0.155	FALSE	FALSE	0	PASS
A6	C20orf203	chr20:31238353:C>A	exonic	nonsynonymous_snv	A156S	742	622	120	0.162	FALSE	FALSE	0	PASS
A6	CACNA1A	chr19:13409626:C>G	exonic	nonsynonymous_snv	E941Q	697	576	121	0.174	FALSE	FALSE	0	PASS
A6	CAST	chr5:96031574:C>G	exonic	nonsynonymous_snv	S58C	54	37	17	0.32	FALSE	FALSE	0	PASS
A6	CC2D2A	chr4:15504120:G>C	exonic	nonsynonymous_snv	E106Q	232	177	55	0.239	FALSE	FALSE	0	PASS
A6	CCDC159	chr19:11460855:T>A	exonic	nonsynonymous_snv	I69N	149	135	14	0.098	FALSE	FALSE	0	PASS
A6	CCDC22	chrX:49104195:C>A	exonic	nonsynonymous_snv	A353D	1037	946	91	0.088	FALSE	FALSE	0	PASS
A6	CCNJ	chr10:97817990:G>A	exonic	nonsynonymous_snv	E371K	120	93	27	0.229	FALSE	FALSE	1	PASS
A6	CEP128	chr14:81251610:C>T	exonic	nonsynonymous_snv	E614K	109	75	34	0.315	FALSE	FALSE	0	PASS
A6	CEP95	chr17:62533725:T>C	exonic	nonsynonymous_snv	L765S	63	51	12	0.2	FALSE	FALSE	0	PASS
A6	CHIC1	chrX:72783256:G>C	exonic	nonsynonymous_snv	E46Q	338	270	68	0.203	FALSE	FALSE	0	PASS
A6	CLGN	chr4:141317335:T>C	exonic	nonsynonymous_snv	I303M	149	120	29	0.197	FALSE	FALSE	0	PASS
A6	CNBD1	chr8:88249177:T>C	exonic	nonsynonymous_snv	L203P	67	48	19	0.288	FALSE	FALSE	0	PASS
A6	COL22A1	chr8:139618637:G>C	exonic	nonsynonymous_snv	P1364R	69	55	14	0.211	FALSE	FALSE	0	PASS
A6	COL4A1	chr13:110835623:C>A	splicing	splice	G633V	151	109	42	0.278	FALSE	FALSE	0	PASS
A6	COL9A1	chr6:70965071:G>T	exonic	nonsynonymous_snv	P509H	61	55	6	0.109	FALSE	FALSE	0	PASS
A6	CTNNA2	chr2:80816447:C>A	exonic	nonsynonymous_snv	P676T	114	88	26	0.232	FALSE	FALSE	0	PASS
A6	DAW1	chr2:228767798:G>C	exonic	nonsynonymous_snv	Q192H	175	149	26	0.152	FALSE	FALSE	0	PASS
A6	DCAF8L1	chrX:27999212:T>A	exonic	nonsynonymous_snv	E80D	457	372	85	0.187	FALSE	FALSE	0	PASS
A6	DECR1	chr8:91049143:A>G	exonic	nonsynonymous_snv	K205E	58	49	9	0.166	FALSE	FALSE	0	PASS
A6	DIRAS3	chr1:68512702:C>G	exonic	nonsynonymous_snv	K93N	179	152	27	0.154	FALSE	FALSE	0	PASS
A6	DLC1	chr8:13356569:G>A	exonic	nonsynonymous_snv	R338C	89	79	10	0.12	TRUE	TRUE	2	PASS
A6	DNA2	chr10:70192032:G>T	exonic	nonsynonymous_snv	Q602K	89	78	11	0.131	FALSE	FALSE	0	PASS
A6	DNAH3	chr16:21069431:T>A	exonic	nonsynonymous_snv	Q1300H	53	46	7	0.145	FALSE	FALSE	0	PASS
A6	DNAH6	chr2:85035598:C>A	exonic	nonsynonymous_snv	L3891M	81	73	8	0.108	FALSE	FALSE	0	PASS
A6	DUSP27	chr1:167095602:G>A	exonic	nonsynonymous_snv	G412R	199	141	58	0.294	FALSE	FALSE	0	PASS
A6	DYNC1I1	chr7:95614255:G>C	exonic	nonsynonymous_snv	D254H	102	61	41	0.404	FALSE	FALSE	0	PASS
A6	EDNRB	chr13:78493532:C>A	exonic	nonsynonymous_snv	E73D	157	118	39	0.252	FALSE	FALSE	0	PASS
A6	EHMT1	chr9:140708956:A>T	exonic	nonsynonymous_snv	D1085V	346	276	70	0.203	FALSE	FALSE	0	PASS
A6	EMD	chrX:153609138:C>T	exonic	nonsynonymous_snv	S142F	295	235	60	0.205	FALSE	FALSE	0	PASS
A6	ERICH3	chr1:75065441:C>T	exonic	nonsynonymous_snv	R555H	67	58	9	0.1343	TRUE	TRUE	1	PASS
A6	FAM181B	chr11:82443751:G>A	exonic	nonsynonymous_snv	R341C	348	308	40	0.117	FALSE	FALSE	0	PASS
A6	FAT2	chr5:150886965:C>A	exonic	nonsynonymous_snv	R4089S	192	142	50	0.263	FALSE	FALSE	0	PASS
A6	FAT3	chr11:92086057:A>G	exonic	nonsynonymous_snv	N260S	130	118	12	0.097	FALSE	FALSE	0	PASS
A6	FBXO4	chr5:41927145:C>G	exonic	nonsynonymous_snv	L74V	130	116	14	0.113	FALSE	FALSE	0	PASS
A6	FGD1	chrX:54472725:G>T	exonic	nonsynonymous_snv	F901L	527	481	46	0.087	FALSE	FALSE	0	PASS
A6	FGFR2	chr10:123324084:G>A	exonic	nonsynonymous_snv	S129L	70	56	14	0.207	FALSE	FALSE	0	PASS
A6	FICD	chr12:108913075:C>A	exonic	nonsense	Y400X	249	207	42	0.171	FALSE	FALSE	0	PASS
A6	FIG4	chr6:110064446:C>T	exonic	nonsynonymous_snv	S337L	51	44	7	0.15	FALSE	FALSE	0	PASS
A6	FLT4	chr5:180041175:C>T	exonic	nonsynonymous_snv	R1075Q	682	543	139	0.204	FALSE	FALSE	2	PASS
A6	FPR1	chr19:52249356:G>T	exonic	nonsynonymous_snv	P298T	235	177	58	0.249	FALSE	FALSE	0	PASS
A6	FREM1	chr9:14859371:G>C	exonic	nonsynonymous_snv	F147L	184	156	28	0.156	TRUE	TRUE	0	PASS
A6	FRK	chr6:116325135:G>C	exonic	nonsense	S124X	69	57	12	0.181	FALSE	FALSE	0	PASS
A6	GALNT13	chr2:154996951:C>T	exonic	nonsense	Q82X	82	59	23	0.285	FALSE	FALSE	0	PASS
A6	GLB1L	chr2:220107905:C>G	exonic	nonsynonymous_snv	L68F	193	158	35	0.184	FALSE	FALSE	0	PASS
A6	GNA12	chr7:2883750:C>T	exonic	nonsynonymous_snv	E16K	40	24	16	0.405	FALSE	FALSE	0	PASS
A6	GOLGA6L2	chr15:23686407:C>A	exonic	nonsynonymous_snv	R405S	737	644	93	0.127	FALSE	FALSE	0	PASS

A6	GOLGB1	chr3:121414947:C>T	exonic	nonsynonymous_snv	E1470K	81	56	25	0.312	FALSE	FALSE	0	PASS
A6	GPC3	chrX:132795781:C>G	exonic	nonsynonymous_snv	D487H	80	64	16	0.207	FALSE	FALSE	0	PASS
A6	GRM3	chr7:86416222:G>T	exonic	nonsynonymous_snv	V372F	482	422	60	0.125	FALSE	FALSE	0	PASS
A6	GSK3B	chr3:119631590:C>G	exonic	nonsynonymous_snv	E226Q	48	43	5	0.119	FALSE	FALSE	0	PASS
A6	HMGCS1	chr5:43298731:C>T	exonic	nonsynonymous_snv	E113K	108	93	15	0.144	FALSE	FALSE	0	PASS
A6	HPN	chr19:35556161:C>G	exonic	nonsynonymous_snv	I273M	232	143	89	0.385	FALSE	FALSE	0	PASS
A6	HTATSF1	chrX:135593097:G>C	exonic	nonsynonymous_snv	R398T	178	141	37	0.21	FALSE	FALSE	0	PASS
A6	IFT140	chr16:1634385:C>G	exonic	nonsynonymous_snv	V398L	209	173	36	0.175	TRUE	FALSE	0	PASS
A6	IKBIP	chr12:99038439:C>G	exonic	nonsynonymous_snv	G14A	1117	887	230	0.206	FALSE	FALSE	0	PASS
A6	ITPR3	chr6:33655009:G>A	exonic	nonsynonymous_snv	V2028I	293	240	53	0.183	TRUE	TRUE	0	PASS
A6	KAT8	chr16:31141434:G>C	exonic	nonsynonymous_snv	E290Q	228	204	24	0.109	FALSE	FALSE	1	PASS
A6	KIAA1191	chr5:175774788:G>A	exonic	nonsense	R245X	130	107	23	0.182	FALSE	FALSE	1	PASS
A6	KIAA1324L	chr7:86554885:G>A	exonic	nonsynonymous_snv	S339F	79	69	10	0.135	FALSE	FALSE	1	PASS
A6	KIAA1524	chr3:108300342:C>G	exonic	nonsynonymous_snv	E157Q	93	61	32	0.347	FALSE	FALSE	0	PASS
A6	KIF20A	chr5:137518944:G>A	exonic	nonsynonymous_snv	E307K	101	83	18	0.185	FALSE	FALSE	0	PASS
A6	KIF25	chr6:168431475:C>G	exonic	nonsynonymous_snv	Q39E	108	83	25	0.236	TRUE	TRUE	0	PASS
A6	KIF26B	chr1:245850140:G>A	exonic	nonsynonymous_snv	M1285I	548	495	53	0.097	FALSE	FALSE	0	PASS
A6	KIF4A	chrX:69561692:C>G	exonic	nonsynonymous_snv	Q393E	94	76	18	0.196	FALSE	FALSE	1	PASS
A6	KLHL15	chrX:24024710:G>T	exonic	nonsynonymous_snv	T34N	141	111	30	0.216	FALSE	FALSE	0	PASS
A6	KLRG1	chr12:9162105:A>T	exonic	nonsynonymous_snv	H102L	103	80	23	0.228	FALSE	FALSE	0	PASS
A6	KNTC1	chr12:123019308:C>A	exonic	nonsense	S76X	92	82	10	0.116	FALSE	FALSE	0	PASS
A6	LDOC1	chrX:140270961:G>T	exonic	nonsynonymous_snv	F82L	591	470	121	0.205	FALSE	FALSE	0	PASS
A6	LRTM1	chr3:54958718:T>A	exonic	nonsense	K178X	91	62	29	0.324	TRUE	TRUE	0	PASS
A6	LYST	chr1:235894365:T>G	exonic	nonsynonymous_snv	K2972Q	169	132	37	0.221	FALSE	FALSE	0	PASS
A6	MAGEA1	chrX:152482674:G>T	exonic	nonsynonymous_snv	L113M	313	255	58	0.187	FALSE	FALSE	0	PASS
A6	MAGEA10	chrX:151303932:G>T	exonic	nonsynonymous_snv	S54Y	254	216	38	0.152	FALSE	FALSE	0	PASS
A6	MAGI1	chr3:65342735:C>A	exonic	nonsynonymous_snv	R1236L	261	213	48	0.186	FALSE	FALSE	0	PASS
A6	MAP2	chr2:210594676:G>T	exonic	nonsynonymous_snv	G1753V	55	48	7	0.139	FALSE	FALSE	0	PASS
A6	MAP3K4	chr6:161514830:G>T	exonic	nonsynonymous_snv	C578F	72	62	10	0.149	FALSE	FALSE	0	PASS
A6	MGAT5B	chr17:74922812:C>T	exonic	nonsynonymous_snv	P431S	292	263	29	0.101	TRUE	TRUE	4	PASS
A6	MORF4L1	chr15:79172879:A>G	exonic	nonsynonymous_snv	K38R	50	41	9	0.192	FALSE	FALSE	0	PASS
A6	MROH2B	chr5:41042265:C>T	exonic	nonsynonymous_snv	D628N	118	99	19	0.166	FALSE	FALSE	0	PASS
A6	MS4A18	chr11:60506423:G>T	exonic	nonsynonymous_snv	A193S	129	117	12	0.098	FALSE	FALSE	0	PASS
A6	MTMR3	chr22:30384487:G>C	exonic	nonsynonymous_snv	E77Q	94	79	15	0.166	FALSE	FALSE	0	PASS
A6	MTTP	chr4:100503214:C>G	exonic	nonsynonymous_snv	P72A	70	51	19	0.277	FALSE	FALSE	0	PASS
A6	MTUS2	chr13:30077281:G>A	exonic	nonsynonymous_snv	V1360I	244	196	48	0.199	TRUE	TRUE	1	PASS
A6	MYEF2	chr15:48450201:G>A	exonic	nonsynonymous_snv	P325L	55	42	13	0.244	TRUE	TRUE	0	PASS
A6	NCAN	chr19:19338481:G>T	exonic	nonsynonymous_snv	L684F	339	307	32	0.095	FALSE	FALSE	0	PASS
A6	NDUFB9	chr8:125551440:G>T	exonic	nonsynonymous_snv	A5S	440	401	39	0.089	FALSE	FALSE	0	PASS
A6	NECAB1	chr8:91953091:C>A	exonic	nonsense	Y275X	152	124	28	0.188	FALSE	FALSE	0	PASS
A6	NFAT5	chr16:69718837:C>T	exonic	nonsynonymous_snv	P486S	69	57	12	0.183	TRUE	TRUE	0	PASS
A6	NFE2L2	chr2:178096280:T>G	exonic	nonsynonymous_snv	S351R	142	112	30	0.215	FALSE	FALSE	0	PASS
A6	NHLRC4	chr16:618291:A>G	exonic	nonsynonymous_snv	I82V	583	470	113	0.194	FALSE	FALSE	0	PASS
A6	NLRP4	chr19:56369450:G>T	exonic	nonsynonymous_snv	D231Y	200	145	55	0.277	FALSE	FALSE	0	PASS
A6	NMUR1	chr2:232389993:C>A	exonic	nonsynonymous_snv	G348C	403	327	76	0.19	FALSE	FALSE	0	PASS
A6	NOTCH2	chr1:120497715:C>T	exonic	nonsynonymous_snv	E723K	187	158	29	0.158	TRUE	TRUE	2	PASS
A6	NR0B1	chrX:30327047:A>G	exonic	nonsynonymous_snv	I145T	959	872	87	0.091	FALSE	FALSE	0	PASS
A6	NRP2	chr2:206562303:G>T	exonic	nonsynonymous_snv	A37S	236	216	20	0.087	FALSE	FALSE	0	PASS
A6	NRXN3	chr14:78709656:G>T	exonic	nonsynonymous_snv	D74Y	222	173	49	0.222	FALSE	FALSE	0	PASS
A6	OBSCN	chr1:228560380:G>A	exonic	nonsynonymous_snv	E8258K	548	445	103	0.188	TRUE	TRUE	0	PASS
A6	PCDHB15	chr5:140625549:G>C	exonic	nonsynonymous_snv	E135Q	226	165	61	0.272	FALSE	FALSE	0	PASS
A6	PCDHGB2	chr5:140741957:A>T	exonic	nonsynonymous_snv	Y752F	315	229	86	0.274	FALSE	FALSE	0	PASS
A6	PCDHGB7	chr5:140798837:A>G	exonic	nonsynonymous_snv	I471V	466	377	89	0.192	FALSE	FALSE	0	PASS
A6	PDE5A	chr4:120488257:C>T	exonic	nonsynonymous_snv	G240S	54	41	13	0.249	TRUE	TRUE	0	PASS
A6	PDGFRA	chr4:55138618:C>T	exonic	nonsynonymous_snv	T457M	203	180	23	0.116	TRUE	TRUE	2	PASS
A6	PER1	chr17:8045134:G>A	exonic	nonsynonymous_snv	L1197F	227	195	32	0.143	FALSE	FALSE	0	PASS
A6	PGAM4	chrX:77225005:G>C	exonic	nonsynonymous_snv	A44G	384	316	68	0.176	FALSE	FALSE	0	PASS
A6	PHKA1	chrX:71932728:C>G	exonic	nonsynonymous_snv	V44L	242	209	33	0.138	FALSE	FALSE	0	PASS
A6	PIEZO2	chr18:10789214:G>C	exonic	nonsynonymous_snv	L678V	276	177	99	0.359	FALSE	FALSE	0	PASS
A6	PKN2	chr1:89271206:A>G	exonic	nonsynonymous_snv	M510V	74	61	13	0.184	FALSE	FALSE	0	PASS
A6	PLA2G4A	chr1:186916078:A>G	exonic	nonsynonymous_snv	M417V	39	25	14	0.366	FALSE	FALSE	0	PASS
A6	PLCH1	chr3:155199800:C>G	exonic	nonsynonymous_snv	D1338H	244	199	45	0.186	FALSE	FALSE	0	PASS
A6	PLXNA2	chr1:208219335:G>T	exonic	nonsynonymous_snv	S1128Y	234	214	20	0.088	FALSE	FALSE	1	PASS
A6	PPP1R13L	chr19:45885927:T>C	exonic	nonsynonymous_snv	Y769C	547	375	172	0.315	TRUE	TRUE	0	PASS
A6	PPP2CB	chr8:30655241:C>G	exonic	nonsynonymous_snv	L114F	78	62	16	0.212	FALSE	FALSE	0	PASS
A6	PRKCB	chr16:24192133:A>G	exonic	nonsynonymous_snv	M473V	57	50	7	0.135	FALSE	FALSE	0	PASS
A6	PRPS2	chrX:12837682:G>T	exonic	nonsynonymous_snv	R196M	254	198	56	0.222	FALSE	FALSE	0	PASS
A6	PRR23A	chr3:138724393:C>A	exonic	nonsynonymous_snv	V240L	783	670	113	0.145	FALSE	FALSE	0	PASS
A6	PRSS56	chr2:233388226:G>T	exonic	nonsynonymous_snv	G317V	567	441	126	0.222	FALSE	FALSE	0	PASS
A6	PTPRF	chr1:44084976:C>T	exonic	nonsynonymous_snv	T1285I	91	80	11	0.128	FALSE	FALSE	0	PASS
A6	PTPRO	chr12:15661514:G>A	exonic	nonsynonymous_snv	G426E	170	151	19	0.115	FALSE	FALSE	0	PASS
A6	QSER1	chr11:32956207:G>A	exonic	nonsynonymous_snv	E1006K	104	81	23	0.225	FALSE	FALSE	0	PASS
A6	RANBP17	chr5:170725797:G>A	exonic	nonsynonymous_snv	D1068N	85	59	26	0.31	FALSE	FALSE	0	PASS
A6	RBFOX3	chr17:77102861:C>G	exonic	nonsynonymous_snv	E78Q	583	465	118	0.203	FALSE	FALSE	0	PASS
A6	ROBO1	chr3:78685196:C>G	exonic	nonsynonymous_snv	E1034Q	93	63	30	0.326	FALSE	FALSE	0	PASS
A6	RORC	chr1:151787559:C>T	exonic	nonsynonymous_snv	R214K	223	142	81	0.365	FALSE	FALSE	0	PASS

A6	RP11-321N4.5	chr6:86322638:C>A	splicing	splice	G85G	68	57	11	0.162	FALSE	FALSE	0	PASS
A6	RP11-399J13.3	chr11:64785853:G>A	exonic	nonsynonymous_snv	G28E	324	271	53	0.164	FALSE	FALSE	0	PASS
A6	RTCB	chr22:32792234:C>T	exonic	nonsynonymous_snv	A273T	77	65	12	0.164	FALSE	FALSE	0	PASS
A6	RTP2	chr3:187416515:T>C	exonic	nonsynonymous_snv	H150R	883	685	198	0.225	FALSE	FALSE	0	PASS
A6	SACS	chr13:23911873:C>T	exonic	nonsynonymous_snv	E2048K	107	92	15	0.146	FALSE	FALSE	1	PASS
A6	SC5D	chr11:121177887:T>C	exonic	nonsynonymous_snv	L189S	116	102	14	0.126	TRUE	TRUE	0	PASS
A6	SHISA8	chr22:42306502:C>A	exonic	nonsynonymous_snv	G241C	530	461	69	0.131	FALSE	FALSE	0	PASS
A6	SLC10A2	chr13:103703691:A>T	exonic	nonsynonymous_snv	L226Q	132	83	49	0.373	TRUE	TRUE	0	PASS
A6	SLC12A8	chr3:124854600:C>A	exonic	nonsense	E217X	192	143	49	0.257	TRUE	TRUE	0	PASS
A6	SLC14A1	chr18:43319130:G>T	exonic	nonsynonymous_snv	L222F	93	67	26	0.284	FALSE	FALSE	0	PASS
A6	SLC16A2	chrX:73740853:G>C	exonic	nonsynonymous_snv	M153I	188	141	47	0.252	FALSE	FALSE	0	PASS
A6	SLC2A13	chr12:40441855:C>G	exonic	nonsynonymous_snv	W238C	74	64	10	0.143	FALSE	FALSE	1	PASS
A6	SLC40A1	chr2:190428865:T>A	exonic	nonsynonymous_snv	T283S	198	179	19	0.099	FALSE	FALSE	0	PASS
A6	SLC4A10	chr2:162627561:G>C	exonic	nonsynonymous_snv	E54Q	35	29	6	0.188	FALSE	FALSE	0	PASS
A6	SLC5A7	chr2:108627285:G>C	exonic	nonsynonymous_snv	E571Q	61	49	12	0.206	FALSE	FALSE	0	PASS
A6	SORCS1	chr10:108536359:T>C	exonic	nonsynonymous_snv	Y273C	66	59	7	0.118	TRUE	TRUE	0	PASS
A6	SPANXN1	chrX:144337210:G>C	exonic	nonsynonymous_snv	R32T	131	107	24	0.187	FALSE	FALSE	0	PASS
A6	SPEN	chr1:16260921:C>T	exonic	nonsynonymous_snv	A2729V	209	174	35	0.17	TRUE	TRUE	0	PASS
A6	SRRM1	chr1:24981357:C>T	exonic	nonsynonymous_snv	S351L	89	81	8	0.098	FALSE	FALSE	2	PASS
A6	ST3GAL4	chr11:126277484:G>T	exonic	nonsynonymous_snv	R120L	398	356	42	0.106	FALSE	FALSE	0	PASS
A6	SUGP1	chr19:19389597:C>T	exonic	nonsynonymous_snv	E513K	215	175	40	0.189	FALSE	FALSE	0	PASS
A6	SYNCRIP	chr6:86322638:C>A	exonic	nonsense	G551X	68	57	11	0.17	FALSE	FALSE	0	PASS
A6	SYNE1	chr6:152675975:T>C	exonic	nonsynonymous_snv	Y3582C	126	111	15	0.125	FALSE	FALSE	0	PASS
A6	TACR3	chr4:104512737:C>A	exonic	nonsynonymous_snv	W331L	82	66	16	0.202	FALSE	FALSE	0	PASS
A6	TAS2R13	chr12:11061142:C>G	exonic	nonsynonymous_snv	W252C	100	68	32	0.323	FALSE	FALSE	0	PASS
A6	TBC1D23	chr3:100038044:C>T	exonic	nonsynonymous_snv	P607S	115	104	11	0.099	FALSE	FALSE	0	PASS
A6	TBX1	chr22:19753930:G>T	exonic	nonsynonymous_snv	R343L	297	254	43	0.147	FALSE	FALSE	0	PASS
A6	TECTA	chr11:121028698:C>T	exonic	nonsynonymous_snv	S1485F	359	288	71	0.199	FALSE	FALSE	1	PASS
A6	TMEM200C	chr18:5890672:T>A	exonic	nonsynonymous_snv	Y464F	428	274	154	0.361	FALSE	FALSE	0	PASS
A6	TMEM239	chr20:2795930:C>A	exonic	nonsynonymous_snv	R34S	502	424	78	0.156	FALSE	FALSE	0	PASS
A6	TP53	chr17:7577082:C>A	exonic	nonsense	E247X	109	90	19	0.18	FALSE	FALSE	37	PASS
A6	TPO	chr2:1499952:G>A	exonic	nonsynonymous_snv	R733K	542	493	49	0.091	FALSE	FALSE	0	PASS
A6	TRPM8	chr2:234839319:G>A	exonic	nonsynonymous_snv	V42M	44	32	12	0.281	FALSE	FALSE	0	PASS
A6	TRRAP	chr7:98602936:G>A	exonic	nonsynonymous_snv	R3559K	450	378	72	0.16	FALSE	FALSE	0	PASS
A6	TTN	chr2:179456135:G>T	exonic	nonsynonymous_snv	P20106H	125	94	31	0.252	FALSE	FALSE	1	PASS
A6	TUBGCP6	chr22:50659585:G>A	exonic	nonsynonymous_snv	S1068L	696	590	106	0.153	FALSE	FALSE	0	PASS
A6	URAD	chr13:28552584:C>A	exonic	nonsense	E61X	148	121	27	0.186	FALSE	FALSE	0	PASS
A6	USH2A	chr1:216595521:G>T	exonic	nonsynonymous_snv	P53Q	141	114	27	0.194	FALSE	FALSE	0	PASS
A6	UTP20	chr12:101767513:G>C	exonic	nonsynonymous_snv	D2367H	138	124	14	0.106	TRUE	TRUE	0	PASS
A6	VAX1	chr10:118896040:G>T	exonic	nonsense	Y124X	465	377	88	0.19	TRUE	TRUE	0	PASS
A6	VMA21	chrX:150565812:C>T	exonic	nonsynonymous_snv	A11V	264	223	41	0.157	FALSE	FALSE	0	PASS
A6	VPS13C	chr15:62170909:C>T	exonic	nonsynonymous_snv	E3347K	40	27	13	0.332	TRUE	FALSE	0	PASS
A6	VWA5A	chr11:123988993:G>T	exonic	nonsynonymous_snv	G115V	128	97	31	0.246	TRUE	TRUE	0	PASS
A6	WDR24	chr16:735301:C>A	exonic	nonsynonymous_snv	V659F	884	695	189	0.214	FALSE	FALSE	0	PASS
A6	XIRP2	chr2:167992549:G>T	exonic	nonsynonymous_snv	S180I	64	53	11	0.181	FALSE	FALSE	0	PASS
A6	XKR7	chr20:30584398:C>A	exonic	nonsynonymous_snv	P293Q	660	564	96	0.146	FALSE	FALSE	0	PASS
A6	YJEFN3	chr19:19640173:G>T	exonic	nonsynonymous_snv	R20S	707	577	130	0.184	FALSE	FALSE	0	PASS
A6	ZIC1	chr3:147131169:A>T	exonic	nonsynonymous_snv	Q392L	297	234	63	0.213	FALSE	FALSE	0	PASS
A6	ZNF197	chr3:44671013:G>C	exonic	nonsynonymous_snv	D123H	139	106	33	0.237	FALSE	FALSE	0	PASS
A6	ZNF197;ZNF660-ZNF197	chr3:44671013:G>C	exonic	nonsynonymous_snv	D123H	138	105	33	0.242	FALSE	FALSE	0	PASS
A6	ZNF367	chr9:99180195:G>C	exonic	nonsynonymous_snv	I40M	298	241	57	0.193	FALSE	FALSE	0	PASS
A6	ZNF383	chr19:37726574:A>G	exonic	nonsynonymous_snv	M45V	162	89	73	0.451	TRUE	TRUE	0	PASS
A6	ZNF41	chrX:47307429:T>A	exonic	nonsynonymous_snv	E580D	99	85	14	0.148	FALSE	FALSE	0	PASS
A6	ZNF768	chr16:30536686:G>A	exonic	nonsynonymous_snv	R259W	327	286	41	0.127	FALSE	FALSE	1	PASS

Exome-seq result of the resected tumor from Stanford NSCLC patient A6.