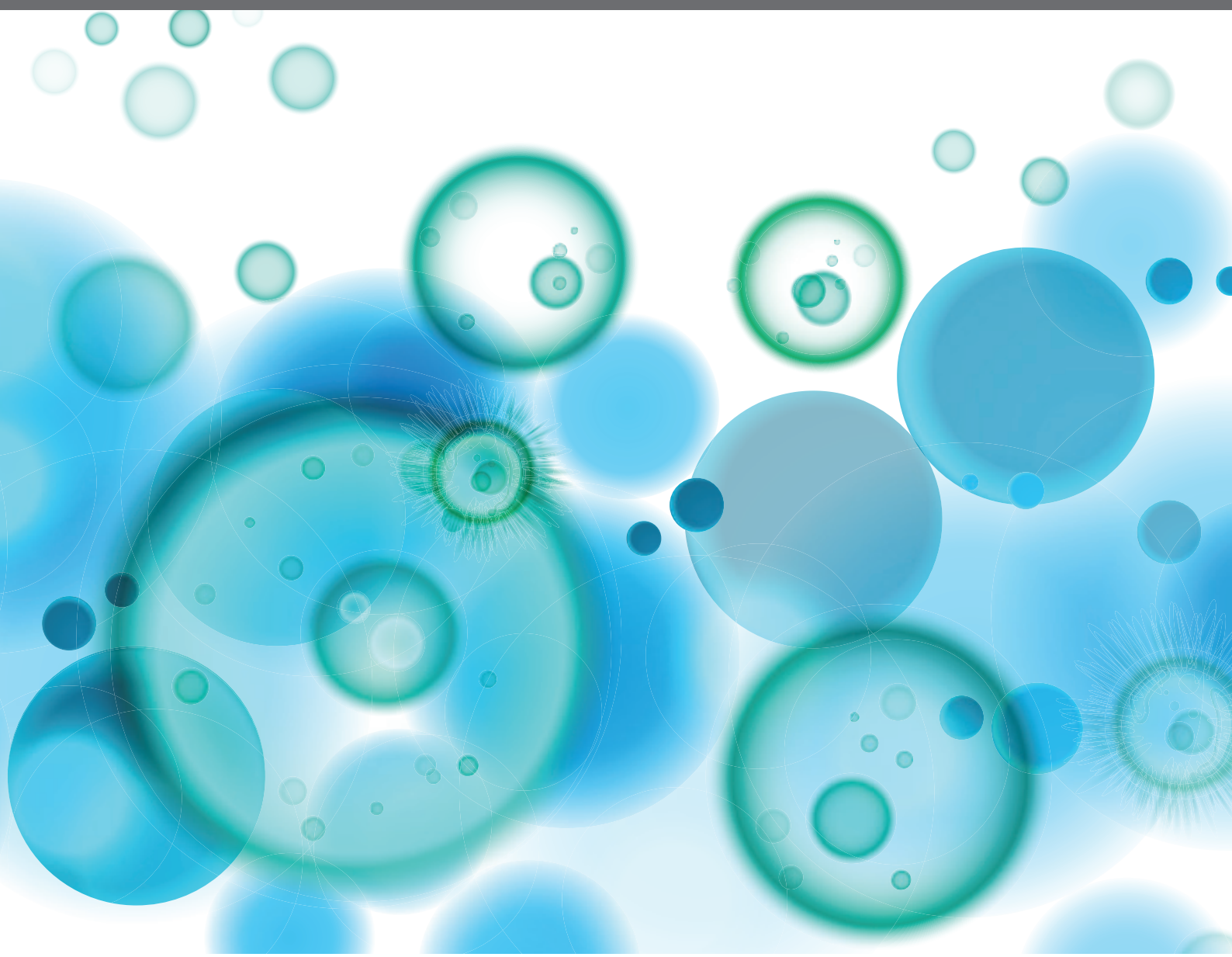
FROM THE COMPUTER TO THE CLINIC: METHODS FOR THE PREDICTION, VALIDATION AND CHARACTERIZATION OF TUMOR NEO-ANTIGENS FOR APPLICATIONS IN CANCER IMMUNOTHERAPY

EDITED BY: Nikolaos G. Sgourakis and John M. Maris
PUBLISHED IN: Frontiers in Immunology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-159-5
DOI 10.3389/978-2-88974-159-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

FROM THE COMPUTER TO THE CLINIC: METHODS FOR THE PREDICTION, VALIDATION AND CHARACTERIZATION OF TUMOR NEO-ANTIGENS FOR APPLICATIONS IN CANCER IMMUNOTHERAPY

Topic Editors:
Nikolaos G. Sgourakis, University of Pennsylvania, United States
John M. Maris, University of Pennsylvania, United States

Citation: Sgourakis, N. G., Maris, J. M., eds. (2022). From the Computer to the Clinic: Methods for the Prediction, Validation and Characterization of Tumor Neo-Antigens for Applications in Cancer Immunotherapy.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88974-159-5

Table of Contents

frontiers in Immunology

ORIGINAL RESEARCH
published: 15 July 2019
doi: 10.3389/fimmu.2019.01573

Prognostic Significance of Potential Immune Checkpoint Member HHLA2 in Human Tumors: A Comprehensive Analysis

*Ben Wang[1], Zhujie Ran[2], Mengmeng Liu[3] and Yunsheng Ou[1]**

[1] Department of Orthopedics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2] School of Public Health and Community Medicine, Chongqing Medical University, Chongqing, China, [3] Graduated School of Anhui University of Traditional Chinese Medicine, Hefei, China

Immunological checkpoint inhibitors have been immensely successfully applied in the treatment of cancer, however, a portion of tumor patients can't benefit from checkpoint therapy. The low PD-1/CTLA-4 positive rate and involvement of multiple immunosuppressive pathways are thought to be one of the reasons for treatment failure in non-responding patients. A new immune checkpoint molecule, HHLA2, which was widely expressed in PD-1 negative human tumors, may be a promising target for the improvement of recent immune therapy. Yet, the prognostic value and transcriptional regulatory mechanisms of HHLA2 remains unclear. In this study, we aimed to evaluate the prognostic value and transcriptional regulation mechanism of HHLA2 according to clinical and experimental data from multiple databases, including cBioPortal, TCGA, Cistrome, TIMER, Oncomine, Kaplan-Meier, GeneXplain. It was found that the expression of HHLA2 was significantly elevated in renal tumors, and significantly decreased in colorectal tumors. Pan-cancer survival analysis indicates that HHLA2 was an independent prognostic factor in 9/20 of human cancers. Especially in renal clear cell carcinoma (P = 3.0E-7). Through plotting survival curve in Kaplan-Meier Plotter, it was found that hypomethylation of HHLA2 DNA was a favorable prognostic factor for KIRC patients. Yet, the copy number variant of HHLA2 was not significantly correlated with the overall survival of KIRC patients. Finally, by analyzing the motif of HHLA2 co-expression genes, we identified 15 transcription factors that may be involved in the regulation of the HHLA2 co-expression network. Among these transcription factors, BATF in B lymphocyte and SMAD in monocyte were confirmed to be able to directly bind to HHLA2 DNA according to chip-seq experimental data from Cistrome database.

Keywords: HHLA2, immune checkpoint, B7, PD-1, expression profiling

OPEN ACCESS

Edited by:
John M. Maris,
University of Pennsylvania,
United States

Reviewed by:
Daniel Olive,
Aix Marseille Université, France
Thorbald Van Hall,
Leiden University, Netherlands

***Correspondence:**
Yunsheng Ou
ouyunsheng2001@163.com

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *02 April 2019*
Accepted: *24 June 2019*
Published: *15 July 2019*

Citation:
Wang B, Ran Z, Liu M and Ou Y (2019) Prognostic Significance of Potential Immune Checkpoint Member HHLA2 in Human Tumors: A Comprehensive Analysis. Front. Immunol. 10:1573. doi: 10.3389/fimmu.2019.01573

INTRODUCTION

Recently, immunotherapy has shown remarkable therapeutic effects in anti-tumor therapy. Among them, the immune checkpoint inhibitors blocking the immunosuppressive receptor of T cells is an effective solution (1). In the opinion of recent studies, the interaction between the B7 and CD28 family plays a central role in regulating T lymphocyte function. The B7-1/B7-2 molecule

on the surface of the cell membrane of antigen presenting cell (APC) binds to the CD28 molecule on the surface of T lymphocyte, providing an initial costimulatory signal for the activation of T lymphocyte. After this costimulatory signal, coinhibitory molecule (such as CTLA-4 and PD-1) of CD28 family on the surface of T lymphocyte binds to B7-1/B7-2 to inhibit T lymphocyte activation (2, 3), this interaction mediated T lymphocyte co-stimulation and co-suppression lay the foundation of the regulation of anti-tumor immune responses (4, 5).

Over the past decade, a series of studies have revealed the important immune function of other molecules belonging to B7 and CD28 families, including B7h/ICOS (6), PD-L1/PD-L2/PD-1 (7), B7-H3 (8). Among them, PD-1/PD-L1 inhibitors have achieved great success in clinical trials (9), which has rapidly changed the treatment landscape for non-small-cell lung cancer (NSCLC) (10). Two FDA-approved PD-L1 inhibitors are being evaluated for the first-line treatment of NSCLC (11–14). However, since PD-1 is only expressed in part of NSCLC, finding new broader expressed immune checkpoint will be important for improving the response rate of immunotherapy (15).

Human endogenous retro virus-H Long repeat-associating 2 (HHLA2) is a newly discovered immune checkpoint molecule belonging to the B7-CD28 family (16, 17). Previous studies confirmed that it participates in the regulation of T-lymphocyte function, but previous studies didn't reach an agreement on the function of HHLA2. The first reported high-quality research concluded that HHLA2 (B7-H5) is a costimulatory molecule that acts to promote T lymphocyte proliferation and secret related cytokines by binding to CD28H receptors on T lymphocyte (16). Subsequent research provides a different conclusion of HHLA2 function, Zhao et al. (17) reported that HHLA2 may be a co-inhibitory ligand for T lymphocyte, which inhibits the anti-tumor function of T lymphocyte. In addition, the heterogeneity of HHLA2 function was also observed in the co-culture experiment, the T lymphocytes from different donors showed a different response to the HHLA2 protein in cytokine production (17). This contradictory phenomenon is interesting, but also confusing for the further researches on HHLA2 (18–21).

To contribute to the understanding of these discrepancies, we investigated the expression profiling and prognostic value of HHLA2 in human cancer according to multiple public databases and investigated what transcription factor may be associated with the dysregulation of HHLA2 in KIRC, our finding may be helpful for the further study on HHLA2.

MATERIALS AND METHODS

Expression Profiling of HHLA2 in Human Cancers

Two large-scale databases (GEPIA, Oncomine) was used to explore the expression pattern of HHLA2 in human cancers. Gene expression profiling interactive analysis (GEPIA) (http://gepia.cancer-pku.cn) is an interactive web server for analyzing the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from TCGA and GTEx projects (22). The Oncomine datasets (23) (http://www.oncomine.org) is a web application for bioinformatics services that includes 715 independent data sets and 86,733 samples that provide greater scale, high quality, consistency, a standardized analytical method for gene expression profile analysis.

Pan-Cancer Survival Analysis

The Kaplan-Meier plotter (24) (http://kmplot.com//analysis) is capable of assessing the prognostic effect of 54,675 genes on using 10,461 cancer samples. To investigate the prognostic role of HHLA2 in human cancers, the Kaplan-Meier Plotter was used to determine the prognostic significance. The forest plot was constructed by RevMan software (25). UCSC Xena browser (26, 27) was used for evaluating the correlation between expression level of HHLA2 mRNA, methylation status, copy number variant of HHLA2 DNA and the overall survival time in kidney clear cell carcinoma.

To validate the prognostic role of HHLA2 in KIRC, we carefully conducted searches in public data platforms including the GEO database and the ArrayExpress database. The datasets meeting following criteria were used as validation cohorts: Datasets with survival associated clinical information. Datasets from human tumor samples. The genechip used in the datasets includes the probe of HHLA2. After detailed and careful searching, two independent clinical cohorts (GSE40435, GSE22541) containing tumor staging information were used for validation.

Potential Transcription Regulatory Mechanism of HHLA2

The cBioPortal for Cancer Genomics (http://www.cbioportal.org) provides a Web resource analyzing, visualizing genomics data. The cBioPortal was used for extracting co-expression genes of HHLA2 in KIRC (28, 29). The geneXplain platform (http://genexplain.com/transfac) is a tool for a broad range of bioinformatics applications. Motif enrichment function in this platform was used for the identification of potential transcription factors of co-expression gene network of HHLA2 with the TRANSFAC database (30–32). TFmapper (33) is used to identify whether transcription factors bind directly to the HHLA2 promoter according to cistrome database. The Cistrome data browser (34, 35) integrates human and mouse cis-regulatory experiment information from ChIP-seq, DNase-seq, and ATAC-seq chromatin analysis. The miRWalk (36, 37), miRDB (38, 39), TargetScan (40) are comprehensive online resource for miRNA and target predictions. They were used to identify potential microRNAs. The common part of three predictions results was regarded as potential microRNAs targeting on HHLA2 mRNA.

Functional Annotation of Co-expression Gene Network of HHLA2

Kyoto Encyclopedia of Genes and Genomes (41) (KEGG) and gene ontology (42) (GO) is a commonly used database of functional annotations for gene list. MetaScape is an excellent integrated analytics platform that integrates multiple annotation datasets. The functional enrichment analysis of this study was

applied by this platform to understand this function of HHLA2 co-expression genes (43).

RESULT

The Expression Level of HHLA2 Was Significantly Elevated in Renal Cancers and Decreased in Colorectal Cancers

The Cancer Genome Atlas (TCGA) datasets is a comprehensive database containing 11,000 patient samples. GEPIA website (TCGA data) and SAGE (**Figure 1**) were used to draw a body map to show the expression distribution of HHLA2 in human tissue (**Figure 2A**). Then, the pan-cancer expression profiling of HHLA2 was visualized according to TCGA data by TIMER platform (**Figure 2C**). It was found that HHLA2 is widely expressed in multiple human tissues, and it abounds in renal, colorectal tissue. Next, we investigated whether the expression of HHLA2 is elevated in human cancer compared with normal tissue. Using Oncomine (23) database($P < 10E-4$, fold Change > 2), which contains 715 datasets and 86,733 samples, we found that HHLA2 is over-expressed in the 7/134 human tumor datasets, and under-expressed in 10/134 datasets (**Figure 2B**). The most common over-expression cancer type in this study is renal cancer, the most common under-expression cancer type is colorectal cancers.

Pan-Cancer Survival Analysis of HHLA2 mRNA Expression Revealed a Unique Prognostic Role of HHLA2 in Renal Clear Cell Carcinoma

RNA-Seq data of 7,462 samples from Kaplan Meier plotter datasets (24) was used for pan-cancer survival analysis. The significant prognostic significance of HHLA2 was determined in 10/20 tumors in this pan-cancer analysis (**Supplementary Table 1**). Overview of these results was shown in **Figure 3**, there is an obvious heterogeneity between different cancer types.

Our result shows that it is hard to determine whether the prognostic role of HHLA2 is favorable or unfavorable. However, in the result of pan-cancer survival analysis, we observed a unique and evident prognostic role of HHLA2 in kidney clear cell carcinoma comparing to other human cancers.

We further extended these results using two independent KIRC cohorts from the GEO database (GSE40435, GSE22541). All of these cohorts showed that the expression of HHLA2 was higher in low pathological grade than the high pathological grade (**Figure 4**). This was consistent with our result that overexpression of HHLA2 was a protective factor for KIRC patients.

Combing with the result of expression profiling analysis, HHLA2 abounds in human normal and malignant tissue, however only differentially expressed in some tumor types (KIRC, COAD etc.) (**Supplementary Table 1**), the subsequent studies will focus on the unique role of HHLA2 in renal clear cell carcinoma.

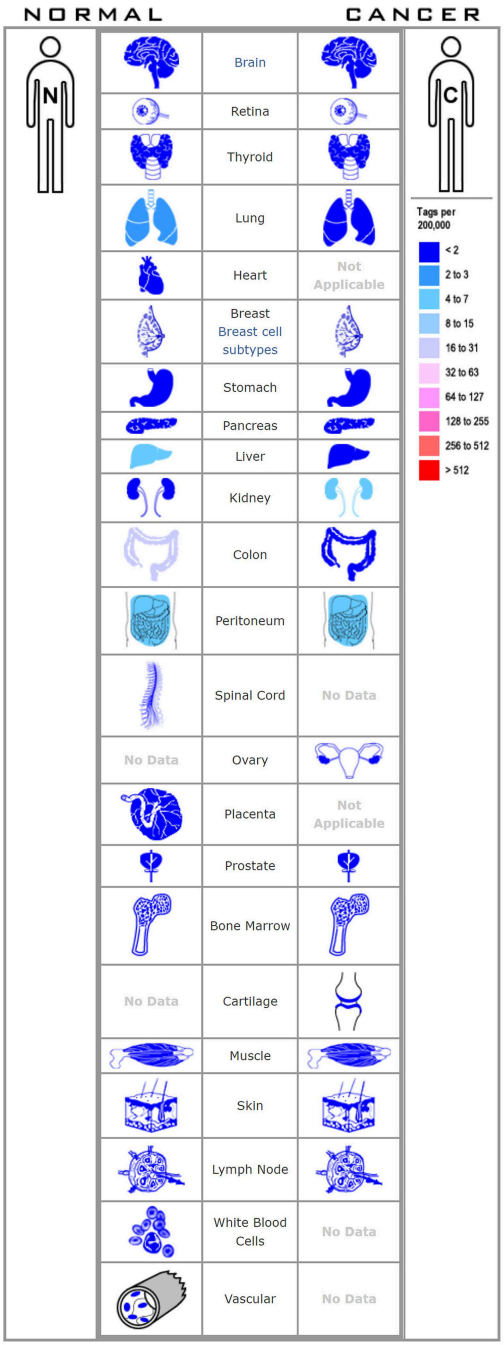


FIGURE 1 | The figure shows the distribution of HHLA2 in various human tissues.

Survival Analysis for Kidney Clear Cell Carcinoma (KIRC) Patients Revealed Methylation Status of HHLA2 DNA Was Associated With the Prognosis of KIRC Patients

The previous studies showed that copy number variant of HHLA2 DNA may be involved in the dysregulation of HHLA2

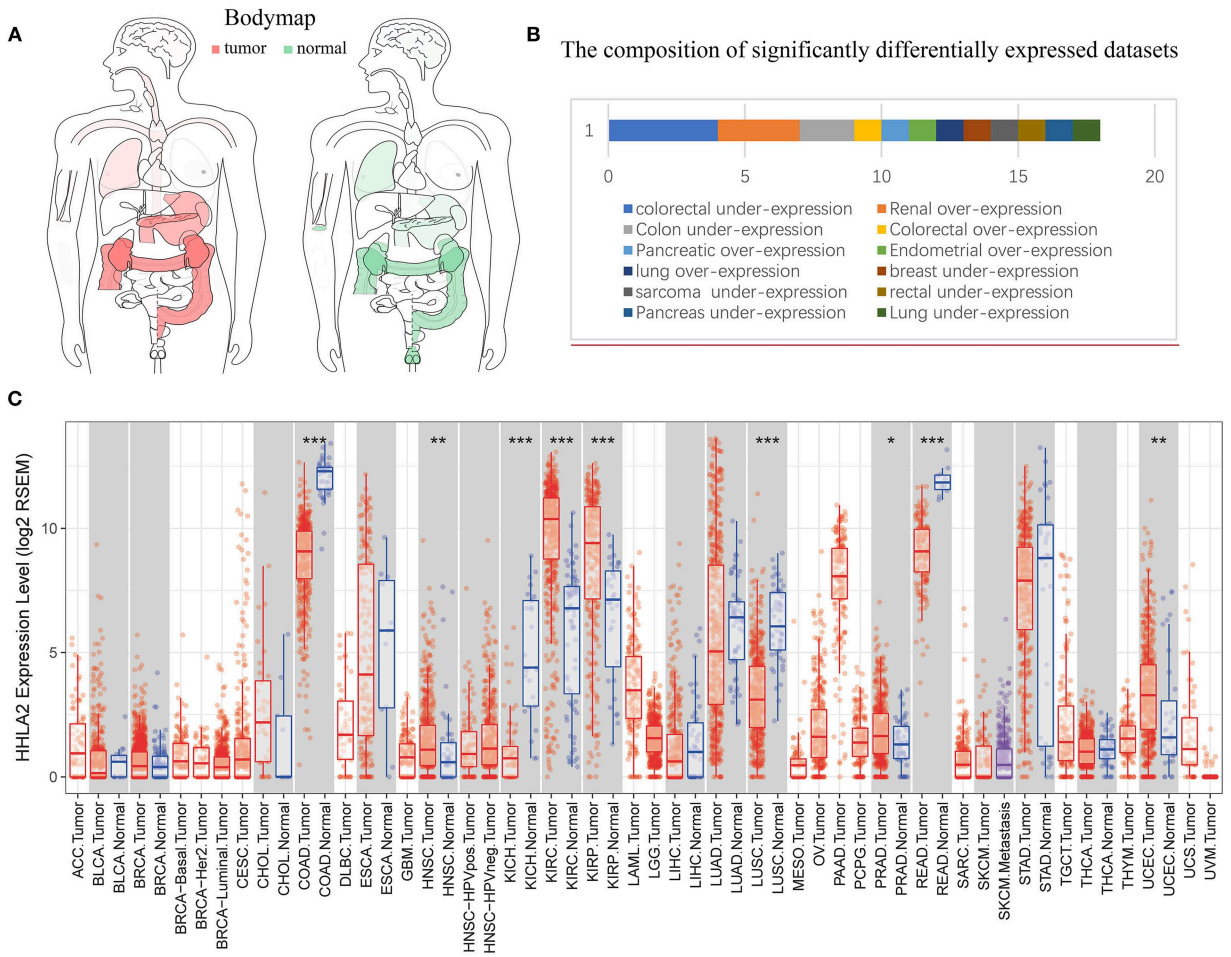


FIGURE 2 | Pan-cancer expression profiling analysis of HHLA2. **(A)** Interactive body map of HHLA2 mRNA expression constructed by GEPIA (TCGA), the darker color corresponds to the higher gene expression level. Red means the median expression of HHLA2 from tumor samples, and the green is from normal samples. **(B)** The length of the bar corresponds to the number of studies show significant over or under expression of HHLA2 vs. normal tissue (the number of significantly differently expressed GEO clinical cohorts). **(C)** The boxplot shows the pan-cancer expression profiling of HHLA2 in human cancers. The below row refers to the standard abbreviations of tumor in TCGA. The color refers to the tumor (red) or normal (blue). *P*-value Significant Codes: 0 ≤ ***< 0.001 ≤ **< 0.01 ≤ *< 0.05 ≤. < 0.1.

expression level in breast cancer (44). So, we will investigate whether DNA variant of HHLA2 have a similar effect on KIRC patients. Our result shows that the expression of HHLA2 was correlated with the methylation status of HHLA2 DNA in KIRC, but not copy number variant (**Figure 5A**). Then, we investigate whether DNA change was correlated with the prognosis of KIRC. Survival analysis showed that the hypomethylation of HHLA2 DNA and overexpression of HHLA2 mRNA were favorable prognostic factors for KIRC patients. However, there was no significant correlation between HHLA2 copy number variant and patient prognosis (**Figures 5B–D**).

Functional Annotation of Co-expression Gene Network of HHLA2

Co-expression gene analysis is a systematic approach for analyzing the potential regulatory pattern of the complex system (45). Co-expression genes of renal clear cell carcinoma were extracted from cBioPortal and Gene expression omnibus (GEO) datasets (accession ID: GSE2109) (16). The co-expression gene of HHLA2 with the absolute value of the Pearson correlation coefficient > 0.4 was used to draw Venn plot (**Figure 6A**). The intersection of these gene lists was regarded as the potential co-expression gene of HHLA2 (**Figure 6A**, **Supplementary Table 2**). To investigate the potential function of co-expression genes, MetaScape (43) was used for functional enrichment. The most statistically significant terms were shown in **Figure 6B**. The enrichment result shows that the SLC-mediated transmembrane transport might be associated with the co-expression genes of HHLA2.

Transcription and Post-transcription Regulation of HHLA2 Co-expression Genes

Taken together, the above finding indicated HHLA2 is significantly over-expressed in KIRC, this dysregulation

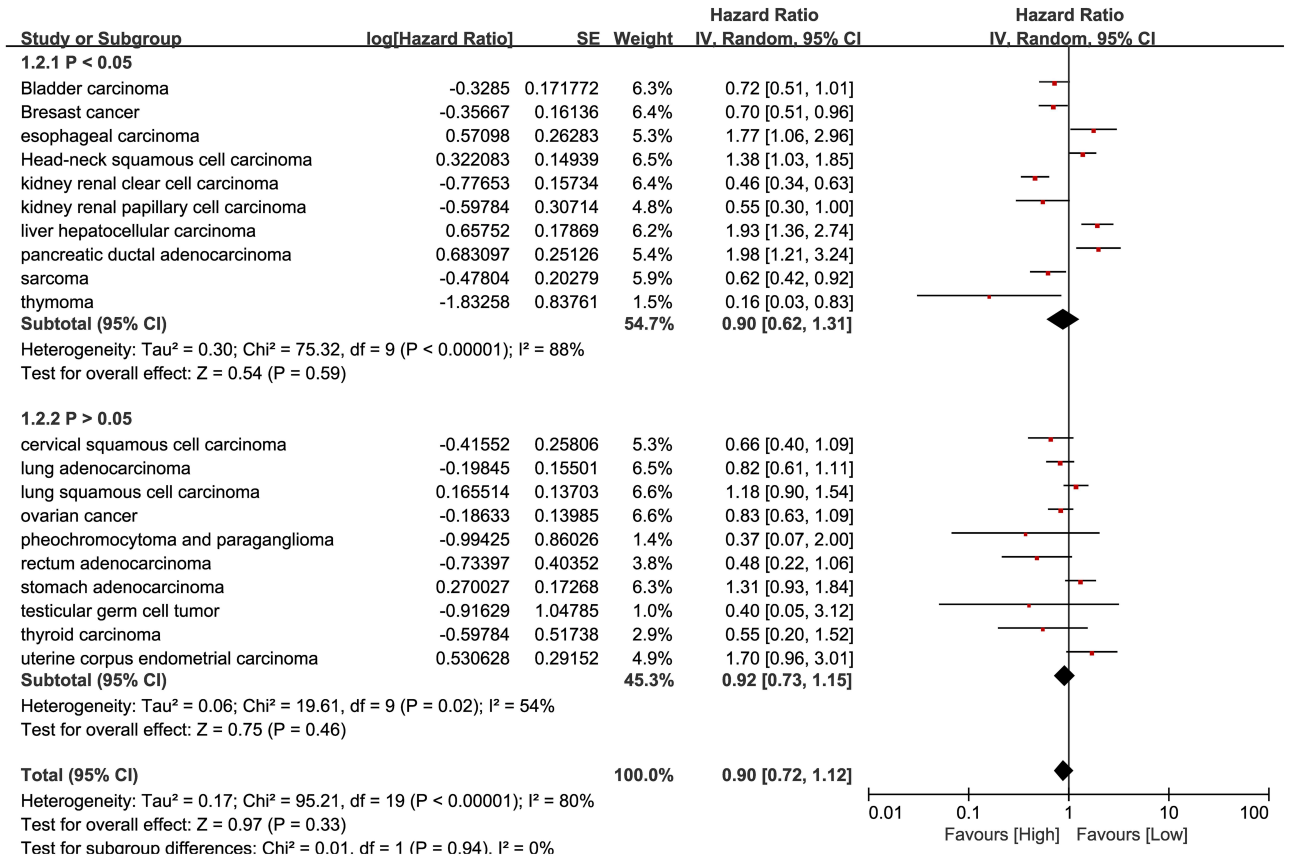


FIGURE 3 | Forest plot shows the result of pan-cancer survival analysis. The upper subgroups are composed of studies whose *P*-value is <0.05, the below subgroups are more than 0.05.

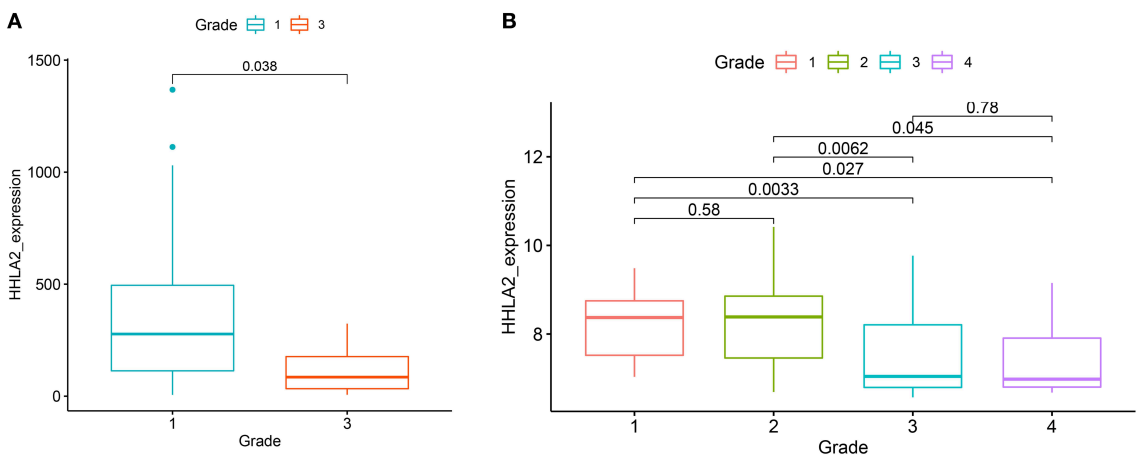


FIGURE 4 | The boxplot shows the correlation between the expression of HHLA2 and the pathological grade in KIRC (left:from GSE22541, right: GSE40435). The label at the top of the picture corresponds to the pathological stage of the patient.

may be derived from post-transcription or transcription regulation. So, we investigate what transcription factors might play an upstream regulation role of the HHLA2 co-expression network. Using the geneXplain platform, 15 transcription factors were identified (**Supplementary Table 3**).

Next, we further investigated whether the identified transcription factors can directly bind to the HHLA2 DNA.

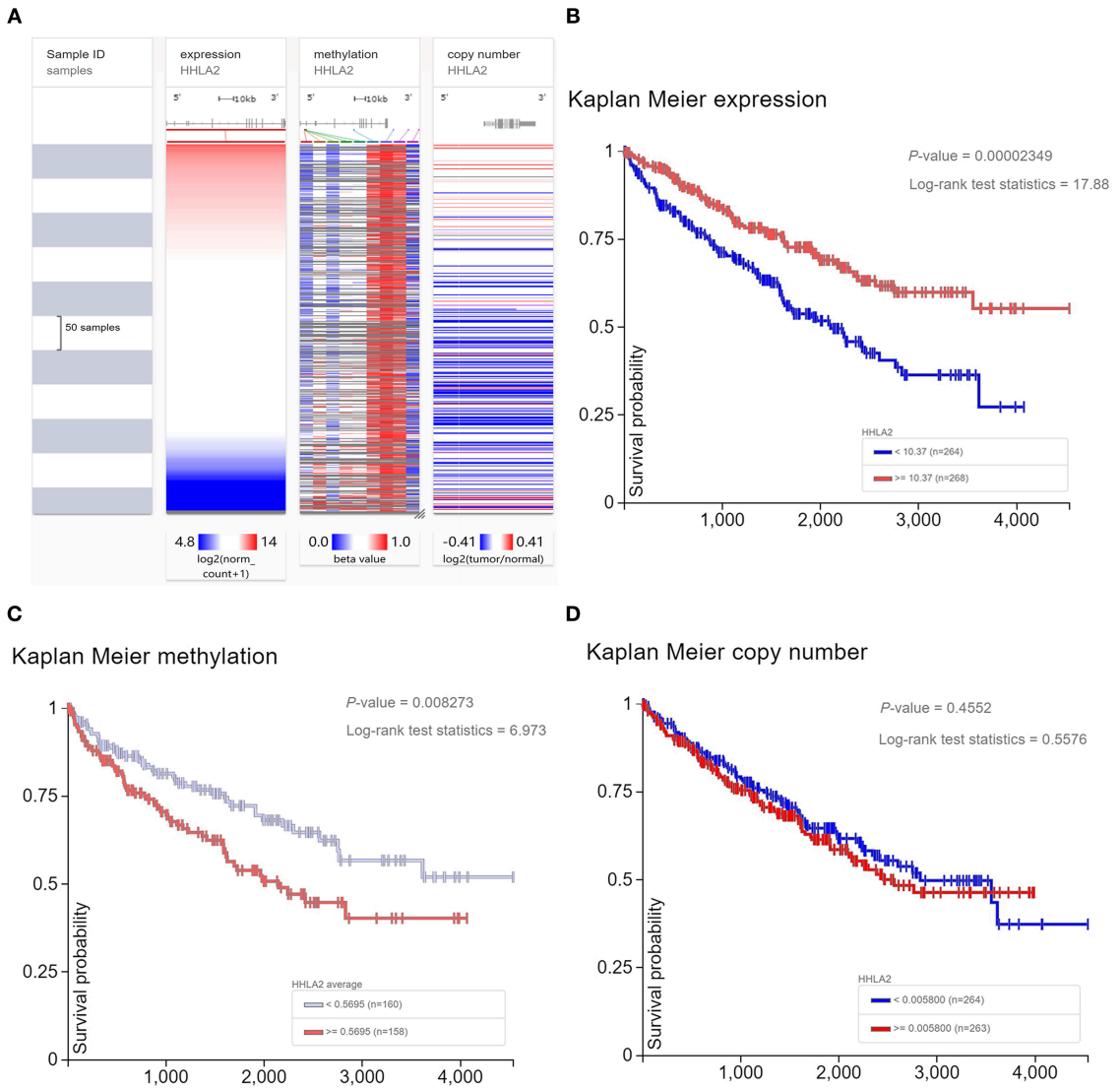


FIGURE 5 | The Heatmap and Kaplan Meier curves of expression level, methylation status, copy number variant of HHLA2. **(A)** The heatmap shows the expression level, methylation status, copy number variant of HHLA2 in the TCGA database, determined by UCSC Xena. **(B–D)** Kaplan Meier plot shows the over-expression and hypomethylation of HHLA2 is favorable prognostic factor for KIRC patients.

Chip-seq is a commonly used experimental technique for studying protein-DNA interactions (46). Cistrome data browser is a comprehensive database that integrates human and mouse transcription factors ChIp-seq, DNase-seq, and ATAC-seq experiment information. We used Cistrome data's Chip-seq data to confirm that BATF, GATA3, HSF1, HOXC9, SMAD1 can directly bind to HHLA2 DNA.

The evidence that BATF can bind HHLA2 DNA is derived from B Lymphocyte and Lymphoblastoid, SMAD1 is derived from Monocyte and Haematopoietic Progenitor Cell, and GATA3 is found in normal organs such as breast and brain. HOXC9 is found in brain and HSF1 is found in the breast. Detailed results can be found in the **Supplementary Table 4**. Because previous studies suggested that HHLA2 is expressed in monocyte and B lymphocytes (34, 35), so we speculate that BATF is involved in the regulation of HHLA2 expression in B lymphocytes and that SMAD1 is in monocytes.

In addition, using miRDB (38, 39), mirWalk (36, 37), and TargetScan (40) databases, we also investigated what microRNAs were involved in the post-transcription regulation of HHLA2.Using strict screening criteria (TargetScan: context++ score percentile > 98, context++ score < -0.4, miRDB: Score > 70, mirWalk: P-value < 0.01), two microRNAs (hsa-miR-3116 hsa-miR-6870-5p) that might bind to HHLA2 mRNA were identified (**Supplementary Figure 1**).

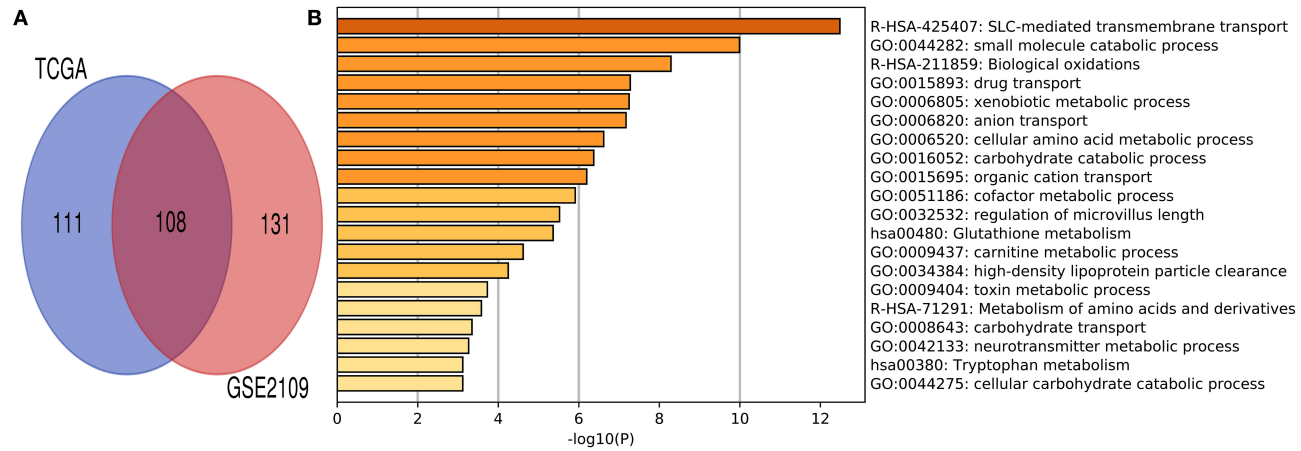


FIGURE 6 | Venn plot and Metascape enrichment result of co-expression genes of HHLA2. **(A)** The intersection of Venn plot is considered to be co-expressed genes of HHLA2. **(B)** The bar plot of HHLA2 co-expression gene enrichment term.

DISCUSSION

HHLA2 is a newly identified gene of the B7/CD28 family. Previous studies have reported that it was widely expressed in patients with PD-1-negative NSCLC, which suggests HHLA2 might be promising immunotherapy target for tumor patients who do not response to PD-1 related therapy (15). Our results are consistent with the previous reports that HHLA2 is widely expressed in a number of human tumors including kidney, Colon et al. Overall, according to expression profiling analysis of Oncomine and TCGA, HHLA2 expression was significantly elevated in most of the kidney cancers associated studys and decreased in colorectal cancer datasets.

In the terms of prognostic role, since the over-expression of immune checkpoint stimulator was always associated with a better survival time of tumor patients, these two opposite conclusions about HHLA2 function make the exact prognostic role of HHLA2 still uncertain.

Our study indicated that HHLA2 was a significant prognostic factor in a portion of tumors, but obvious heterogeneity prognostic value was observed on different kinds of tumors, part of them are protective and others are unfavorable or no significant prognostic factors.

Interestingly, a unique prognostic role of HHLA2 was observed in renal clear cell carcinoma compared with other tumor types. In a further study with KIRC as a case, we found that overexpressed or hypomethylation of HHLA2 is a favorable prognostic factor for KIRC, however, there was no significant correlation between copy number variant of HHLA2 and the prognosis of KIRC patients. Our finding was not consistent with previous report that elevated expression of HHLA2 may be associated with abnormal copy-number variant of HHLA2 DNA in basal breast cancer (44). We are unable to reconcile this discrepancy due to the lack of information on the primary antibody applied in that study.

In general, HHLA2, as a newly identified immune checkpoint molecule, little is known about the prognostic significance of HHLA2, recent publishing studies were still on "preliminary research stage" and "clinical validation stage" (47, 48), further evidence from *in vivo* and *in vitro* experimental evidence is needed for the understanding of the immune function of HHLA2, before this experimental evidence, the prognostic role should be discussed in detail, this was directly associated with the next research direction of HHLA2 and whether it is a promising prognostic marker or a new immunotherapeutic target.

According to our results, we proposed some inspiring insights and hypothesis for the research of HHLA2.

I) Our result revealed that the prognostic significance of HHLA2 was with slight evidence in portion of tumors. Although some of the previous studies reported significant prognostic value of HHLA2 in a variety of tumors (47, 48), due to the lack of blind method, relatively small sample size and subjective quantitative method, high risk of bias may exist in part of the previous reports. Therefore, in the future, more rigorous researches and meta-analysis with high-qualified evidence are needed to determine the prognostic value of HHLA2, especially in KIRC, and whether it is correlated with tumor types.

II) Our results showed that HHLA2 was a favorable prognostic marker for KIRC with strong evidence, however, the favorable role was not always observed in previous studies. Our finding was consistent with the first biological experimental report about the function of HHLA2 (16), but was contrary to part of previous researches (47, 48), so, why is the favorable prognostic value of HHLA2 unique and evident for KIRC? Further studies which are focused on the correlation between characteristic of KIRC and the function of HHLA2 will contribute to the understanding of the heterogeneous immune function of HHLA2.

III) Gene expression dysregulation always corresponds to an important biological function in most cases. According to our result, HHLA2 played an evident protective prognostic role and was significantly overexpressed in KIRC, the biological mechanism behind this phenomenon is needed to note, further *in vivo* and *in vitro* experiment focused on this phenomenon may be helpful to understand how HHLA2

is interacting with tumor cells. In addition to the function of HHLA2 in KIRC, the potential mechanism involved in dysregulated expression of HHLA2 is also noteworthy, our result provided a possible explanation for this question, inconsistent with the previous report on breast cancer (44), the epigenetic modification but not copy number variant may be responsible for the dysregulated expression of HHLA2 in KIRC.

Our study is the first to report the potential transcriptional regulation mechanism of HHLA2 in bioinformatic view. Our result showed that BATF in B lymphocyte and SMAD in monocytes might be responsible for the dysregulation of HHLA2 in KIRC. Further studies of the BATF/HHLA2 axis and the SMAD/HHLA2 axis may help to further understand the role of HHLA2 in the human immune system.

DATA AVAILABILITY

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

ML and ZR drafted the paper. BW analyzed the data. YO designed the study and rephrased the paper. The paper was approved by all authors.

FUNDING

The National Natural Science Foundation of China (81572634).

ACKNOWLEDGMENTS

Thanks for FangdongdeMiao in the process of writing this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.01573/full#supplementary-material

REFERENCES

1. Anagnostou V, Smith KN, Forde PM, Niknafs N, Bhattacharya R, White J, et al. Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer. *Cancer Discov*. (2017) 3:264–76. doi: 10.1158/1538-7445.AM2017-NG01
2. Boussiotis VA. Molecular and biochemical aspects of the PD-1 checkpoint pathway. *N Engl J Med*. (2016) 375:1767–78. doi: 10.1056/NEJMra1514296
3. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol*. (2008) 36:677–704. doi: 10.1146/annurev.immunol.26.021607.090331
4. Liu WR, Shipp MA. Signaling pathways and immune evasion mechanisms in classical Hodgkin lymphoma. *Blood*. (2017) 330:2265–70. doi: 10.1182/blood-2017-06-781989
5. Steinbach A, Riemer AB. Immune evasion mechanisms of human papillomavirus: An update. *Int J Cancer*. (2018) 342:224–9. doi: 10.1002/ijc.31027
6. Yoshinaga SK, Whoriskey JS, Khare SD, Sarmiento U, Guo J, Horan T, et al. T-cell co-stimulation through B7RP-1 and ICOS. *Nature*. (1999) 302:827–32. doi: 10.1038/45582
7. Kryczek I, Lin H, Wei S, Green M, Zou W. Relevance of host and tumor PD-L1 in PD-L1 pathway blockade. *Int J Infect Dis*. (2018) 300(1 Suppl.):56.25. doi: 10.1016/j.ijid.2017.01.028
8. Aung PP, Parra ER, Barua S, Sui D, Ning J, Mino B, et al. B7-H3 expression in Merkel cell carcinoma-associated endothelial cells correlates with locally aggressive primary tumor features and increased vascular density. *Clin Cancer Res*. (2019). doi: 10.1158/1078-0432.CCR-18-2355
9. Baumeister SH, Freeman GJ, Dranoff G, Sharpe AH. Coinhibitory pathways in immunotherapy for cancer. *Annu Rev Immunol*. (2016) 34:539–73. doi: 10.1146/annurev-immunol-032414-112049
10. Büttner R, Gosney JR, Skov BG, Adam J, Motoi N, Bloom KJ, et al. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non–small-cell lung cancer. *J Clin Oncol*. (2017) 35:3867–3876. doi: 10.1200/JCO.2017.74.7642
11. Antonia SJ, Kim S-W, Spira AI, Ahn M-J, Ou S-HI, Stjepanovic, et al. Safety and clinical activity of durvalumab (MEDI4736), an anti-PD-L1 antibody, in treatment-naïve patients with advanced non–small-cell lung cancer. *J Clin Oncol*. (2016) 34(15_suppl.):9029. doi: 10.1200/JCO.2016.34.15_suppl.9029
12. Gulley JL, Rajan A, Spigel DR, Iannotti N, Chandler J, Wong DJL, et al. Avelumab for patients with previously treated metastatic or recurrent non-small-cell lung cancer (JAVELIN Solid Tumor): dose-expansion cohort of a multicentre, open-label, phase 1b trial. *Lancet Oncol*. (2017) 38:599–610. doi: 10.1016/S1470-2045(17)30240-1
13. Jerusalem G, Chen F, Spigel D, Iannotti N, Mcclay E, Redfern C, et al. OA03.03 JAVELIN solid tumor: safety and Clinical activity of avelumab (Anti-PD-L1) as first-line treatment in Patients with Advanced NSCLC. *J Thor Oncol*. (2017) 32:S252. doi: 10.1016/j.jtho.2016.11.240
14. Planchard D, Yokoi T, McCleod MJ, Fischer JR, Kim YC, Ballas M, et al. A Phase III study of durvalumab (MEDI4736) with or without tremelimumab for previously treated patients with advanced NSCLC: rationale and protocol design of the ARCTIC study. *Clin Lung Cancer*. (2016) 37:232–6.e1. doi: 10.1016/j.cllc.2016.03.003
15. Cheng H, Borczuk A, Janakiram M, Ren X, Lin J, Assal A, et al. Wide expression and significance of alternative immune checkpoint molecules, B7x and HHLA2, in PD-L1-negative human lung cancers. *Clin Cancer Res*. (2018) 34:1954–64. doi: 10.1158/1078-0432.CCR-17-2924
16. Zhu Y, Yao S, Iliopoulou BP, Han X, Augustine MM, Xu H, et al. B7-H5 costimulates human T cells via CD28H. *Nat Commun*. (2013) 3:2043. doi: 10.1038/ncomms3043
17. Zhao R, Chinai JM, Buhl S, Scandiuzzi L, Ray A, Jeon H, et al. HHLA2 is a member of the B7 family and inhibits human CD4 and CD8 T-cell function. *Proc Natl Acad Sci USA*. (2013) 310:9879–84. doi: 10.1073/pnas.1303524110
18. Faget J, Sisirak V, Blay J-Y, Caux C, Bendriss-Vermare N, Ménétrier-Caux C. ICOS is associated with poor prognosis in breast cancer as it promotes the amplification of immunosuppressive CD4(+) T cells by plasmacytoid dendritic cells. *Oncoimmunology*. (2013) 3:e23185. doi: 10.4161/onci.23185
19. Jie HB, Schuler PJ, Lee SC, Srivastava RM, Argiris A, Ferrone S, et al. CTLA-4⁺ regulatory T cells increased in cetuximab-treated head and neck cancer patients suppress NK cell cytotoxicity and correlate with poor prognosis. *Cancer Res*. (2015) 35:2200–10. doi: 10.1158/0008-5472.CAN-14-2788
20. Parry RV, Chemnitz JM, Frauwirth KA, Lanfranco AR, Braunstein I, Kobayashi SV, et al. CTLA-4 and PD-1 receptors inhibit T-cell activation by distinct mechanisms. *Mol Cell Biol*. (2005) 35:9543–53. doi: 10.1128/MCB.25.21.9543-9553.2005
21. Zhang L, Wang L, Shahzad KA, Xu T, Wan X, Pei W, et al. Paracrine release of IL-2 and anti-CTLA-4 enhances the ability of artificial polymer antigen-presenting cells to expand antigen-specific T cells and inhibit tumor

growth in a mouse model. *Cancer Immunol Immunother*. (2017) 36:1229–41. doi: 10.1007/s00262-017-2016-9
22. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. (2017) 35:W98–102. doi: 10.1093/nar/gkx247
23. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia*. (2007) 3:166–80. doi: 10.1593/neo.07112
24. Menyhart O, Nagy A, Gyorffy B. Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. *R Soc Open Sci*. (2018) 3:181006. doi: 10.1098/rsos.181006
25. Beller E, Clark J, Tsafnat G, Adams C, Diehl H, Lund H, et al. Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). *Syst Rev*. (2018) 3:77. doi: 10.1186/s13643-018-0740-7
26. Cui X, Yi Q, Jing X, Huang Y, Tian J, Long C, et al. Mining prognostic significance of MEG3 in human breast cancer using bioinformatics analysis. *Cell Physiol Biochem*. (2018) 30:41–51. doi: 10.1159/000493956
27. Boyiadzis MM, Kirkwood JM, Marshall JL, Pritchard CC, Azad NS, Gulley JL. Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease. *J Immunother Cancer*. (2018) 3:35. doi: 10.1186/s40425-018-0342-x
28. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. (2012) 3:401–4. doi: 10.1158/2159-8290.CD-12-0095
29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. (2013) 3:pl1. doi: 10.1126/scisignal.2004088
30. Koschmann J, Bhar A, Stegmaier P, Kel AE, Wingender E. "Upstream Analysis": An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays*. (2015) 3:270–86. doi: 10.3390/microarrays4020270
31. Orekhov AN, Oishi Y, Nikiforov NG, Zhelankin AV, Dubrovsky L, Sobenin IA, et al. Modified Ldl particles activate inflammatory pathways in monocyte-derived macrophages: transcriptome analysis. *Curr Pharm Des*. (2018) 24:3143–51. doi: 10.2174/1381612824666180911120039
32. Stegmaier P, Voss N, Meier T, Kel A, Wingender E, Borlak J. Advanced computational biology methods identify molecular switches for malignancy in an EGF mouse model of liver cancer. *PLoS ONE*. (2011) 3:e17738. doi: 10.1371/journal.pone.0017738
33. Zeng J, Li G. A tool for searching putative factors regulating gene expression using ChIP-seq data. *Int J Biol Sci*. (2018) 34:1724–31. doi: 10.7150/ijbs.28850
34. Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, et al. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res*. (2016) 35:D658–62. doi: 10.1093/nar/gkw983
35. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res*. (2018) 37:D729–35. doi: 10.1093/nar/gky1094
36. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods*. (2015) 32:697. doi: 10.1038/nmeth.3485
37. Dweep H, Sticht C, Pandey P, Gretz N. miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform*. (2011) 34:839–47. doi: 10.1016/j.jbi.2011.05.002
38. Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. Genome Biol. (2019) 20:18. doi: 10.1186/s13059-019-1629-z
39. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res*. (2014) 33:D146–52. doi: 10.1093/nar/gku1104
40. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. (2015) 4:e05005. doi: 10.7554/eLife.05005
41. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. (2019) 37:D590–5. doi: 10.1093/nar/gky962
42. Makabe T, Arai E, Hirano T, Ito N, Fukamachi Y, Takahashi Y, et al. Genome-wide DNA methylation profile of early-onset endometrial cancer: its correlation with genetic aberrations and comparison with late-onset endometrial cancer. *Carcinogenesis*. (2019) bgz046. doi: 10.1093/carcin/bgz046
43. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, et al. Meta- and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. *Cell Host Microbe*. (2015) 38:723–35. doi: 10.1016/j.chom.2015.11.002
44. Janakiram M, Chinai JM, Fineberg S, Fiser A, Montagna C, Medavarapu R, et al. Expression, clinical significance, and receptor identification of the newest B7 family member HHLA2 protein. *Clin Cancer Res*. (2015) 31:2359–66. doi: 10.1158/1078-0432.CCR-14-1495
45. Villa-Vialaneix N, Liaubet L, Laurent T, Cherel P, Gamot A, SanCristobal M. The structure of a gene co-expression network reveals biological functions underlying eQTLs. *PLoS ONE*. (2013) 3:e60045. doi: 10.1371/journal.pone.0060045
46. Chen X, Bhadauria V, Ma B. ChIP-Seq: a powerful tool for studying protein-DNA interactions in plants. *Curr Issues Mol Biol*. (2018) 37:171–80. doi: 10.21775/cimb.027.171
47. Jing CY, Fu Y-P, Yi Y, Zhang M-X, Zheng S-S, Huang J-L, et al. HHLA2 in intrahepatic cholangiocarcinoma: an immune checkpoint with prognostic significance and wider expression compared with PD-L1. *J Immunother Cancer*. (2019) 3:77. doi: 10.1186/s40425-019-0554-8
48. Lin G, Ye H, Wang J, Chen S, Chen X, Zhang C. Immune checkpoint human endogenous retrovirus-H long terminal repeat-associating protein 2 is upregulated and independently predicts unfavorable prognosis in bladder urothelial carcinoma. *Nephron*. (2019) 341:256–64. doi: 10.1159/000495887

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 28 August 2019
doi: 10.3389/fimmu.2019.02047

Structure Based Prediction of Neoantigen Immunogenicity

Timothy P. Riley†‡, Grant L. J. Keller‡, Angela R. Smith, Lauren M. Davancaze, Alyssa G. Arbuiso, Jason R. Devlin and Brian M. Baker**

Department of Chemistry and Biochemistry and the Harper Cancer Research Institute, University of Notre Dame, Notre Dame, IN, United States

OPEN ACCESS

Edited by:
Nikolaos G. Sgourakis, University of California, Santa Cruz, United States

Reviewed by:
Philip Bradley, Fred Hutchinson Cancer Research Center, United States
Angelika B. Riemer, German Cancer Research Center (DKFZ), Germany

***Correspondence:**
Timothy P. Riley
triley368@gmail.com
Brian M. Baker
brian-baker@nd.edu

†Present address:
Timothy P. Riley, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA, United States

‡*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *06 May 2019*
Accepted: *13 August 2019*
Published: *28 August 2019*

Citation:
Riley TP, Keller GLJ, Smith AR, Davancaze LM, Arbuiso AG, Devlin JR and Baker BM (2019) Structure Based Prediction of Neoantigen Immunogenicity. Front. Immunol. 10:2047. doi: 10.3389/fimmu.2019.02047

The development of immunological therapies that incorporate peptide antigens presented to T cells by MHC proteins is a long sought-after goal, particularly for cancer, where mutated neoantigens are being explored as personalized cancer vaccines. Although neoantigens can be identified through sequencing, bioinformatics and mass spectrometry, identifying those which are immunogenic and able to promote tumor rejection remains a significant challenge. Here we examined the potential of high-resolution structural modeling followed by energetic scoring of structural features for predicting neoantigen immunogenicity. After developing a strategy to rapidly and accurately model nonameric peptides bound to the common class I MHC protein HLA-A2, we trained a neural network on structural features that influence T cell receptor (TCR) and peptide binding energies. The resulting structurally-parameterized neural network outperformed methods that do not incorporate explicit structural or energetic properties in predicting $CD8^+$ T cell responses of HLA-A2 presented nonameric peptides, while also providing insight into the underlying structural and biophysical mechanisms governing immunogenicity. Our proof-of-concept study demonstrates the potential for structure-based immunogenicity predictions in the development of personalized peptide-based vaccines.

Keywords: structure, neoantigen, peptide, MHC, modeling, Rosetta, neural network, personalized vaccines

INTRODUCTION

The development of immunological therapies that incorporate peptide antigens presented to T cells by major histocompatibility complex (MHC) proteins is a long sought-after goal, particularly for cancer. Although early cancer vaccine studies relying on non-mutated shared antigens were disappointing (1), advances in sequencing and bioinformatics have led to the identification of "neoantigens" with non-synonymous mutations that differentiate tumors from healthy tissues [reviewed in (2)]. Vaccination using such neoantigens has in some cases led to promising outcomes (3, 4), and neoantigens are now also being explored as a means to improve the safety, specificity, and efficacy of other immunotherapies. A significant challenge remains, however, in identifying those mutated peptides that are immunogenic and can thus promote anti-tumor immune responses.

Following sequencing, potential neoantigens have been identified via bioinformatic approaches that predict processing and presentation by MHC proteins, and more recently, via mass spectrometry (5, 6). Mutations at anchor residues which improve the binding of a peptide to an MHC protein have been associated with immunogenicity and tumor rejection (7–9). In these cases, T cells not eliminated by negative selection may exist that efficiently recognize

the neoantigen; indeed, in viruses, recent findings suggest that for peptides presented by class I MHC proteins, peptide binding affinity is the best predictor of immunogenicity (10). However, in the more common instances in which mutations occur outside of anchor residues and do not strongly impact peptide-MHC binding, T cells that efficiently recognize the wild-type peptide will have been deleted or otherwise tolerized. In these instances, an immunogenic neoantigen must possess structural and physical properties distinct enough to promote efficient recognition by T cells that ignore the wild-type peptide (i.e., the single mutation must result in a peptide that is "sufficiently different" from its wild-type counterpart to overcome self-tolerance).

However, it is becoming increasingly understood that, even after taking tolerance mechanisms into account, not all well-presented peptides are strongly immunogenic (11, 12), suggesting the existence of peptide features that influence T cell recognition independently of peptide processing and presentation. For example, recent work suggests that immunogenic peptides are enriched in hydrophobic (including aromatic) amino acids at positions often contacted by T cell receptors (TCRs) (13, 14). Efforts at incorporating features that influence T cell recognition into neoantigen prediction tools are in development (14–18), and these complement well-developed tools for predicting MHC binding (19–23). The immune epitope database (IEDB) and NetTepi servers, for example, incorporate positional enrichment of hydrophobic amino acids into class I MHC immunogenicity prediction tools (13, 18, 24). Other physicochemical features that have been considered include amino acid charge and size, wild-type and mutant sequence divergence, and sequence entropy (15, 17).

Despite these advances, the mechanisms by which physicochemical features of peptides influence TCR binding have not been widely considered. For example, enrichment in hydrophobic amino acids at potential TCR contact sites for immunogenic peptides can be interpreted in the context of protein biophysics: burial of exposed hydrophobic surface promotes protein binding through the hydrophobic effect, which is almost universally favorable, requiring only that a hydrophobic group dock into another hydrophobic environment (25–27). Burying charges, on the other hand, requires overcoming energetically expensive desolvation penalties and thus high structural precision between atoms of opposing charge (28–30). Because of this, TCRs with architectures that permit precise charge complementarity will occur less frequently than those that can accommodate a hydrophobic (or aromatic) group (31). This leads to the prediction that neoantigens whose mutations replace centrally located, charged amino acids with hydrophobic or aromatic amino acids will be immunogenic, as some studies have indeed reported (32). Likewise, introduction of charges can reduce immunogenicity, as has also been reported (33) and commonly seen in studies of T cell specificity using peptide libraries [e.g., (34)].

Yet for a peptide bound to an MHC protein, the impact of features such as exposed hydrophobic surface and charges are determined by the peptide's conformation within the binding groove, as well as the size and position of the various amino acid side chains. Thus, efforts to predict peptide immunogenicity should be strengthened by approaches that account for the structural properties of peptide/MHC complexes (8, 16, 35).

Here, we explored how considering structurally-determined physical features can improve efforts at predicting peptide immunogenicity. We developed a rapid procedure for accurately modeling large numbers of peptides bound to the common class I MHC protein HLA-A*0201 (HLA-A2), which was applied to a curated dataset incorporating thousands of immunogenic, non-immunogenic, and non-HLA-A2 binding peptides. We included non-HLA-A2 binding peptides as we aimed to capture both peptide binding to the MHC protein as well as TCR binding to the peptide/MHC complex, as both contribute to immunogenicity and both are governed by structural and physicochemical features. Indeed, strong TCR binding can compensate for weak peptide-MHC binding and *vice versa* (11, 12, 36); considering both elements together permits capturing the impact of both. We then trained a neural network on energetic features that are encoded not by peptide sequence, but by the modeled three-dimensional structures of the peptide/HLA-A2 complexes. The network recovered known features of immunogenic peptides such as enrichment in hydrophobicity, and, as assessed by the ability to predict $CD8^+$ T cell responses, against the training data outperformed other models and prediction tools based only on sequence characteristics. Deployed against a set of HLA-A2-presented nonameric neoantigens, the network not only permitted predictions of immunogenicity, but yielded testable hypotheses about how the mutations influenced immunogenicity. From this proof-of-concept study we identify clear avenues for improvement and scale up.

RESULTS

Development and Performance of a Rapid Peptide/MHC Modeling Strategy

To develop a rapid structural modeling strategy, we compiled a list of peptide/MHC structures within the Protein Data Bank (PDB). We restricted our analysis to high resolution, HLA-A2 structures presenting nonamers with good electron density. We focused on nonamers as these are the most represented in the PDB and relatively constrained in class I MHC peptide binding grooves. Additionally, nonameric, HLA-A2 data are the most represented in immunological databases. As we intended to emphasize structural differences emerging from amino acid mutations, we further narrowed our database by pairing each peptide/HLA-A2 complex with at least one other in which the peptide differed by only a single amino acid, either as a substitution or transposition. Our final database contained 53 structures presenting distinct peptide epitopes (**Table S1**).

To simulate a realistic setting where many peptides need to be evaluated, we prioritized modeling speed over complexity. As has been noted previously (37), nonameric peptides bound to class I MHC proteins adopt relatively conserved backbone conformations. We therefore modeled each complex in our database by threading the desired peptide sequence into our template HLA-A2 structures, followed by Monte-Carlo-based conformational sampling and energy minimization for side

chains and the peptide backbones utilizing Rosetta (38, 39). This approach required approximately 10 min per model on 2016-vintage CPU hardware. We considered three different templates to compare the effect starting coordinates had on model accuracy: HLA-A2 presenting the HTLV1 Tax_{11-19} peptide (PDF 1DUZ; peptide sequence LLFGYPVYV) (40), the $MART1_{27-35}$ tumor antigen (PDB 3QFD; peptide sequence AAGIGILTV) (41), and a *Toxoplasma gondii* epitope (PDB 5FA3; peptide sequence GLLPELPAV) (42). These three structures were chosen based on their resolution (<1.9 Å) and variations around the nonameric backbone conformation. The modeling procedure performed similarly with all three templates, yielding full atom root mean square deviations (RMSD values) between 1.86 and 2.08 Å, and Cα RMSD values between 0.87 and 1.15 Å (**Figure 1A**; **Table S1**). Other approaches to model peptides in class I MHC binding grooves have incorporated docking, molecular dynamics simulations, protein threading, or combinations of these methods. These other methods have reported Cα or full-atom RMSD values between model and experiment within the approximate range of 1–2 Å (8, 16, 37, 43–50). Our approach thus compares favorably with or even outperforms other efforts.

Of the three templates considered, the models generated from 3QFD were the closest to the crystal structures; the average backbone RMSD of models derived from 3QFD was significantly lower ($p = 0.0006$ and 0.018 when compared to results with the 1DUZ and 5FA3 templates, respectively) (**Figure 1A**). Results from the 3QFD template were thus used for all further comparisons and modeling efforts.

The greatest discrepancy between modeled and actual structures was an unusual register-shifted peptide (LAGIGILTV) which, compared to the native peptide (AAGIGILTV), left the p1 (or "A") pocket of the HLA-A2 molecule empty in the crystal structure, resulting in the nonameric peptide adopting a decameric configuration (41) (**Figure 1B**). Our modeling procedure was not able to sample such dramatic conformational shifts, and thus the model of this peptide resembled more traditional nonameric peptide/MHC structures.

Given recent attention on the role of exposed surface features in the immunogenicity of MHC-presented peptides, we asked how our modeling procedure recovered peptide hydrophobic solvent accessible surface area (hSASA). After comparing models and structures, the Pearson correlation coefficient between predicted and experimental hSASA was 0.66 (**Figure 1C**). Our rapid modeling procedure thus provides a good approximation of peptide structural properties within the binding groove of HLA-A2 and the changes that occur upon mutation.

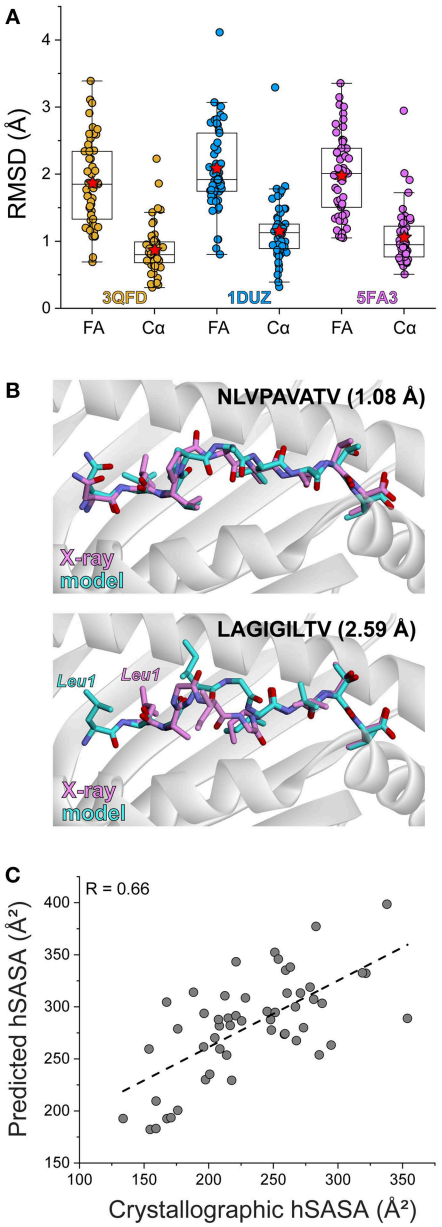


FIGURE 1 | Rapid structural modeling for peptide/HLA-A2 complexes. **(A)** Modeling performance as indicated by peptide Cα and full atom (FA) RMSD values for modeled vs. crystallized peptide/HLA-A2 complexes. Boxes illustrate the 1st and 3rd quartiles, with a horizontal line at the median. Whiskers show 1.5 times the interquartile range. Red stars indicate mean values. Results for three templates (3QFD, 1DUZ, and 5FA3) are shown. The 3QFD template performed the best and was used for all further modeling efforts. **(B)** Structural images of representative models and their corresponding structures. The top shows the model of NLVPAVATV, which superimposes on the crystal structure with a full atom RMSD of 1.08 Å. The bottom shows the model of LAGIGILTV, which superimposes on the structure with a full atom RMSD of 2.59 Å. For the latter case, the leucine at position 1 forces the nonameric peptide to bind in a register-shifted decameric configuration, with the p1 leucine in the B rather than A pocket. Our modeling procedure did not permit such drastic conformational sampling. **(C)** Correlation between exposed peptide hydrophobic surface area in the models vs. the crystallographic structures. The two sets of data correlate with a R value of 0.66.

Experimental Test of the Peptide/MHC Modeling Strategy

To further test the rapid modeling procedure, we crystallized and determined the three-dimensional structures of three new peptide/HLA-A2 complexes. The peptide ILNAMIAKI is a melanoma neoantigen identified in a recent study (51). We determined the structure of ILNAMIAKI bound to HLA-A2, as well as those of the corresponding wild-type peptide ILNAMITKI and another single amino acid variant, ILNAMIVKI (**Table 1**).

TABLE 1 | X-ray data collection and refinement statistics.

	ILNAMIAKI/HLA-A2	ILNAMITKI/HLA-A2	ILNAMIVKI/HLA-A2
Data collection			
Resolution (Å)*	30.89–2.15 (2.23–2.15)	41.82–1.90 (1.97–1.90)	31.57–1.79 (1.84–1.79)
Space group	P1	P1	P 21 21 21
Unit cell dimensions (Å)	50.68, 63.67, 75.18	58.43, 84.11, 85.36	49.56, 74.64, 122.85
Unit cell angles (°)	81.47, 75.85, 77.23	90.01, 90.06, 90.02	90, 90, 90
Unique reflections*	47,203 (4,655)	123,624 (11,774)	42,624 (2,553)
R-merge	0.205 (0.684)	0.096 (0.290)	0.167 (0.203)
I/σ^*	5.7 (1.3)	10.8 (4.3)	22.2 (2.6)
Data completeness*	97.9% (96.3%)	96.7% (92.6%)	97.6% (87%)
Refinement			
R-work, R-free	0.19, 0.23	0.16, 0.19	0.17, 0.20
R-free test set	4711 (10.00%)	12476 (10.1%)	1955 (4.59%)
Wilson B-factor ($Å^2$)	24.0	15.1	22.6
Total number of atoms	6,853	14,303	3,626
Bond lengths RMSD (Å)	0.005	0.003	0.008
Bond angles RMSD (°)	1.02	0.661	0.927
Ramachandran (favored, allowed, outlier)	98%, 2%, 0%	98%, 2%, 0%	98%, 2%, 0%
PDB ID Code	6PTB	6PTE	6OPD

**Values in parentheses are statistics for the highest resolution shells.*

We also subjected the three complexes to the modeling procedure described above. In the structures, the peptides all adopt the typical nonameric conformation, with a bulge initiating at Asn3 and continuing through Ile6. There are no systematic changes in response to the differences at position 7. The sidechains of Met5 and Lys8 extend away from the peptide backbone, with some variations in torsion angles across the three structures (seen as well in the multiple copies in the asymmetric units for the structures with ILNAMITKI and ILNAMIAKI) (**Figure 2A**). The models compared well with the crystallographic structures. One discrepancy was found at the backbone of Ala4, which impacted the geometry of the subsequent Met5 side chain. Nonetheless, the position and extension of the Met5 side chain were well captured, as was the similarly extended Lys8 side chain (**Figure 2B**). The peptide Cα RMSD values between structures and models were between 0.7 and 0.8 Å, and the full atom RMSD values were between 1.7 and 2.0 Å. These values are consistent with the results found when comparing modeled to previously determined experimental structures (**Figure 1A**) and confirm that our modeling scheme can reproduce major structural features of peptide/HLA-A2 complexes.

Collecting a Peptide Dataset to Relate Peptide Structural Features to $CD8^+$ T Cell Responses

To test whether consideration of structural features could lead to improved immunogenicity predictions, we developed a peptide database that contains immunogenic and non-immunogenic peptides. We again emphasized nonameric, HLA-A2 restricted peptides for consistency with our modeling strategy and as data for HLA-A2-presented nonamers is most represented in various immunological databases.

While the IEDB has records for immunogenic peptides, it contains limited data on peptides that are poorly immunogenic yet still well-presented by MHC proteins. To account for such peptides, we relied on lists of peptides identified via proteomic analyses of human HeLa cells (52, 53), yielding a dataset of 2756 nonameric, HLA-A2-presented self-peptides. While this dataset will necessarily include peptides that would be efficiently recognized by TCRs, we rationalized it would be dominated by peptides that bind well to the MHC protein but are not well-recognized (i.e., in a host, peptides that might pass positive selection but not would not drive negative selection). To this set of self-peptides, we added 155 well-characterized immunogenic peptides listed in the IEDB, selected by filtering for HLA-A2-presented human nonamers with IFN-γ ELISPOT response frequencies of 50 or higher in order to minimize false positives. The immunogenic peptide dataset primarily included epitopes from viral sources, although humans and other organisms were also represented (**Table S2**). The dataset included multiple amino acid variants of various peptides, which we rationalized would be important when aiming to predict the immunogenicity of mutant peptides and their wild-type counterparts.

We completed our dataset by adding 1044 HLA-A2-incompatible peptides selected from IEDB training sets. Incorporating non-HLA-A2 binding peptides ensured that our efforts addressed both TCR and MHC binding, as both directly contribute to immunogenicity and are dependent upon structure-determined energetic features. Accounting for both TCR and MHC binding together is necessary for predicting immunogenicity, as a peptide that binds weakly to an MHC protein could still prove immunogenic by possessing optimal features for TCR binding and *vice versa* (11, 12, 36). Moreover, peptide mutations can influence both TCR and MHC binding simultaneously, as seen with differential T cell recognition of some "anchor fixed" shared tumor antigens (41, 54).

Amino acid distributions for the immunogenic, HeLa, and HLA-A2-incompatible peptides are shown in **Figure 3A**. To further ask if our dataset reflected previously noted distinctions

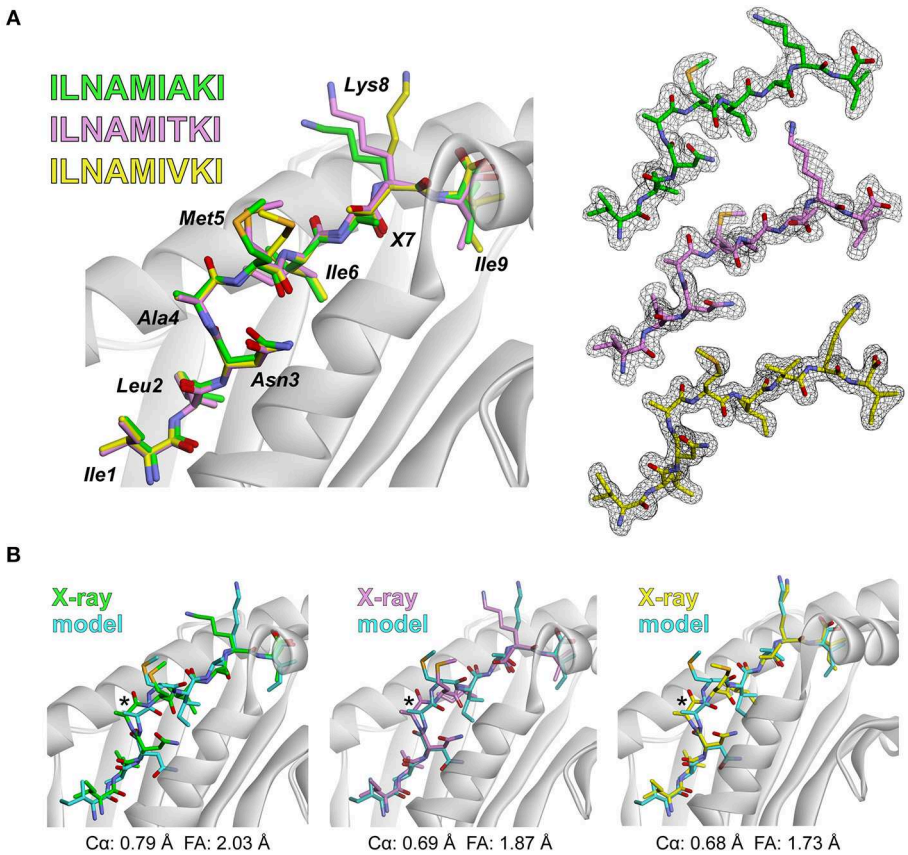


FIGURE 2 | Experimental test of the peptide/MHC modeling strategy. **(A)** Structures of the ILNAMIAKI, ILNAMITKI, and ILNAMIVKI peptides in the binding groove of HLA-A2 as determined by X-ray crystallography. The peptides are colored green, pink, or yellow as indicated; this color scheme is maintained throughout the figure. The left panel shows all three structures superimposed. Peptide amino acids are indicated, with "X" used to indicate the various amino acids at position 7. The right panel shows $2F_O$-F_C electron density maps contoured at 1σ for peptides from each peptide/HLA-A2 structure. **(B)** Comparison of the models of each peptide/HLA-A2 complex with the crystallographic structures. For each panel, the backbone at Ala4 is highlighted with an asterisk and the Cα and full atom (FA) RMSD values for each structure/model pair are indicated.

between immunogenic and non-immunogenic peptides, we evaluated the hydrophobicity of the peptides in the immunogenic and HeLa self-peptide pools. Using the Wimley-White interface hydropathy index (55), we determined the mean hydrophobicity for each peptide position in the two pools. Comparing the results for the two showed that certain positions across the peptides were statistically more likely to be more hydrophobic in the immunogenic than the HeLa self-peptide pool, with the most pronounced differences at positions 4, 7, and 8 (**Figure 3B**). These results, including the distinctiveness of positions 4, 7, and 8, are consistent with previous observations (13, 14) and support the conclusion that our peptide pools appropriately encompass both immunogenic and non-immunogenic peptides.

A Neural Network to Predict Immunogenicity From Structure-Derived Parameters Outperforms Other Approaches

Using our structural modeling procedure and the database of peptides, we next constructed an artificial neural network to predict the immunogenicity of nonameric peptides bound to HLA-A2, relying on structural and energetic features determined from three-dimensional models as the network inputs. Using our rapid modeling scheme, we first generated structural models of all 3,955 peptide/HLA-A2 complexes. To describe the conformation-dependent physical properties of the peptides in the binding groove, we used the 18 terms in the Talaris2014 energy function to evaluate the energy of each modeled peptide/HLA-A2 complex (39, 56). The terms, listed in **Table S3** and described in Alford et al. (56), account for features such as energies of attraction, repulsion, and solvation; energies of side chain and backbone hydrogen bonds; and energies and probabilities of side chain and backbone conformations. We also selected nine terms from the same energy function for all nine positions in the peptide, choosing terms that emphasized atomic-level features and avoiding those descriptive of particular amino acids (e.g., tyrosine planarity). To the nine amino-acid level terms, we also added total and hydrophobic solvent accessible surface areas. Overall, 117 terms that describe each modeled peptide/HLA-A2 complex were used as network inputs.

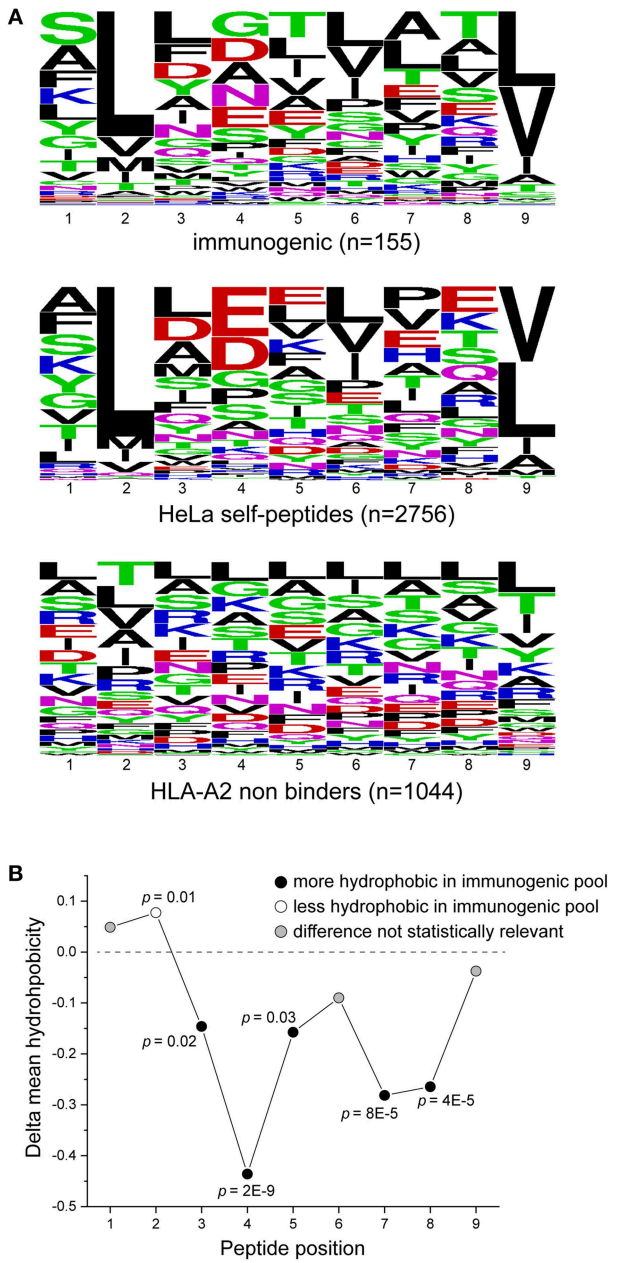


FIGURE 3 | Characteristics of peptides in the training set. **(A)** Sequence logos of immunogenic peptides (top), HeLa self-peptides (middle), and HLA-A2 non-binding peptides (bottom). **(B)** Comparison of the hydrophobicity of each peptide position in the immunogenic and self-peptide datasets (shown as the difference between the immunogenic and self-peptide datasets) as determined using the Wimley-White hydropathy index. Values less than zero (below dashed line) indicate greater hydrophobicity in the immunogenic dataset. *P*-values are indicated where the differences are statistically significant.

As with previous efforts in predicting immunogenicity, we used a binary classification system for each peptide in our dataset, classifying peptides identified from the IEDB as immunogenic (score of 1) and the HeLa and non-HLA-A2 binding peptides as non-immunogenic (score of 0). The network output is thus a score, from zero to one, indicating the degree of confidence in immunogenicity.

In developing the neural network, we used a nested 5-fold cross-validation procedure that eliminated redundant terms. The final model consisted of the 18 terms for the entire peptide/MHC complex and seven for each amino acid in the peptide, yielding 81 terms for network inputs, with five hidden neurons and two constant bias nodes (**Figure 4**; **Table S3**). The average cross-validated area under the curve (AUC) in a receiver operating characteristic (ROC) plot was 0.69 (the AUC values in the ROC plots predict the probability that the neural network will more favorably score an immunogenic peptide compared to a non-immunogenic peptide). After training with the entire dataset, the final neural network classified all peptides used with a total AUC of 0.73 (**Figure 5A**).

To assess the added value of structural and energetic information, we developed a control neural network trained on the same 3,955 peptides but encoded by a sparse matrix that considered only peptide sequence. In this head-to-head comparison, the structurally-parameterized network outperformed the sequence-only network (AUC of 0.73 vs. 0.61), demonstrating conclusively that incorporating structural and energetic features improves predictions compared to considering sequence alone. As a further control, we developed another sequence-based network that considered only amino acid hydropathy values. The hydropathy network outperformed the sequence-only network (AUC of 0.64 vs. 0.61), but still did not match the performance of the structurally-parameterized network.

For comparison to other tools, we evaluated the same 3,955 peptides with the IEDB immunogenicity and the NetTepi immunogenicity prediction tools (13, 18, 24). IEDB classified the peptides with an AUC of 0.58, whereas NetTepi yielded a value of 0.54. Although not designed to predict immunogenicity, peptide-MHC binding predictions are often used in this fashion when assessing putative neoantigens, largely due to experience from viral antigens (10). Consistent with earlier findings (7), predictions using NetMHCpan 4.0 (19) did not perform well, yielding AUC values of 0.51 in affinity mode and 0.46 in ligand-likelihood mode. Overall then, whether compared with a simpler sequence-based neural network, a network capturing solely amino acid hydrophobicity, existing sequence-based immunogenicity tools, or predictions of peptide-HLA-A2 binding affinity, the structure-based network performed the best in predicting immunogenicity.

Significance of Structure-Derived Energetic Network Inputs for Classifying Immunogenicity

Although interpreting the weights of inputs used within a neural network is difficult due to the complexity and non-linear nature of the models, the weights of structural features used within the model can provide clues to their contributions in the evaluation of immunogenicity. For MHC binding, the structure-based network considered the impact of anchor residues 2 and 9 by assessing terms such as

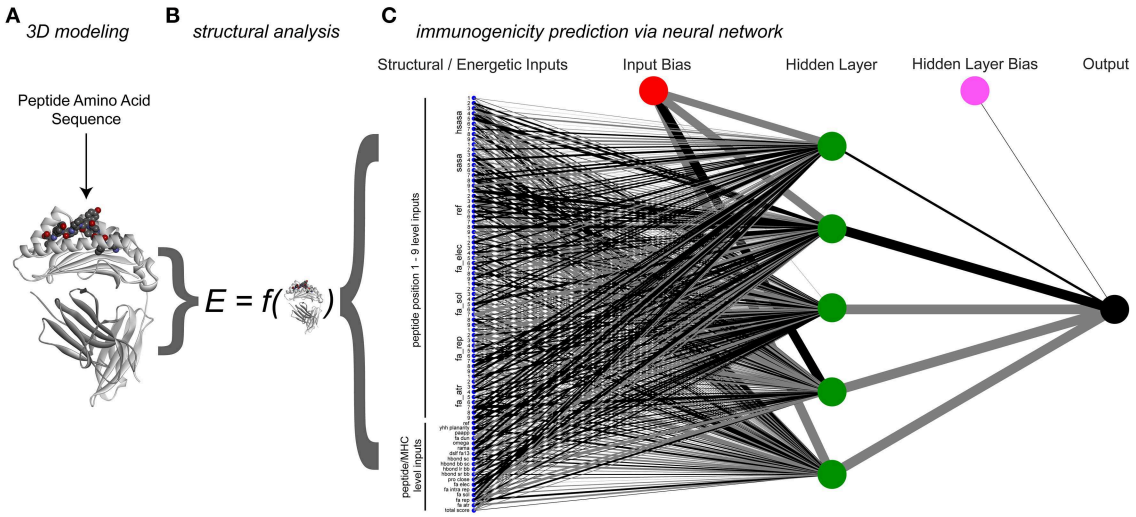


FIGURE 4 | Process and architecture of the structure-based immunogenicity neural network. **(A)** The process begins with a peptide sequence, which is used to generate a model of the peptide/HLA-A2 three-dimensional structure. **(B)** Analysis of the modeled structure yields energetic and topographical information, which are used as inputs for the structure-based immunogenicity neural network. **(C)** Trained neural network architecture, with 81 structure-derived inputs shown on the left (seven for each peptide position, 18 for the overall complex). A single hidden layer is present with five hidden neurons, along with two constant bias nodes. Black lines give positive weights, gray lines negative weights, with line width indicating weight magnitude.

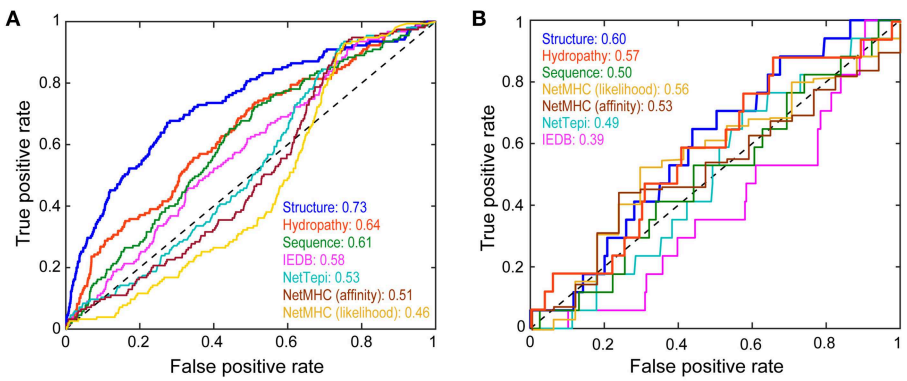


FIGURE 5 | Performance of the structure-based neural network in categorizing peptide immunogenicity. **(A)** Performance of the structure-based neural network compared to control sequence-only and hydropathy-only neural networks, the IEDB and NetTepi immunogenicity prediction servers, and NetMHCpan 4.0 in evaluating the training data as demonstrated by a receiver operating characteristic curve. The area under the curve (AUC) for each approach gives the probability that the approach will more favorably score an immunogenic peptide than a non-immunogenic peptide. The structure-based neural network performed the best. **(B)** Against a neoantigen dataset of 291 nonameric peptides not used for training the structure-based network performed less favorably but still outperformed the other approaches.

favorable van der Waals interactions at these positions in order to quantify how compatible an epitope was with HLA-A2. The network also focused on the interactions surrounding peptide position 3, likely considering peptide-MHC interactions in this tightly packed region of the HLA-A2 binding groove.

Consistent with the hypothesis that solvent exposed residues provide information regarding peptide immunogenicity by promoting TCR binding, the network emphasized hydrophobic SASA. Notably, the weights for hydrophobic SASA and hydrophobic solvation energy values at positions 5, 7, and 8 were in the top 10% of all weights in the neural network. These positions are typically "TCR facing" in HLA-A2-presented nonameric peptides. Indeed, in the structural models used for evaluating the modeling, positions 5, 7, and 8 had high degrees of solvent exposure, and crystallographic structures of TCRs bound to nonameric peptide/HLA-A2 complexes show that these positions on average bury more than 80% of their exposed surface upon receptor binding (**Figure S1**).

One notable result from our analysis was that, excluding the non-HLA-A2 binding peptides, the average computed energies of the immunogenic complexes (as determined by the Talaris2014 total energy score used in the structural modeling) was higher than the non-immunogenic complexes. Although

the difference was small (average of −560 Rosetta energy units for immunogenic complexes vs. −562 for non-immunogenic complexes), the energy reflects the entire peptide/MHC complex, of which the peptide is only approximately 2% by mass. Scoring only the peptides (in the context of the binding groove) recapitulated this trend (average of 11 Rosetta energy units for immunogenic peptides vs. 9.6 for non-immunogenic), and the difference was statistically significant ($p = 0.0017$). We believe this to be an indicator of how structure and energy can influence the immunogenicity of neoantigens: amino acid substitutions that impart a higher energy onto a peptide/MHC (for example, by removing exposed charges and/or increasing exposed hydrophobic surface area) yield ligands that have more energy to release upon TCR binding, translating into stronger binding affinities.

As a separate test of this hypothesis, we computed the total and hydrophobic SASA for the models of the immunogenic and non-immunogenic peptides (again excluding the non-HLA-A2 binding peptides). Although the difference in total SASA was insignificant, the exposed hydrophobic solvent accessible surface area of the immunogenic peptides was higher than the non-immunogenic peptides (244 Å^2 vs. 224 Å^2; $p = 9 \times 10^{-7}$). Exposing hydrophobic surface to water raises free energy via the hydrophobic effect, with widely used estimates relating 1 Å^2 of exposed hydrophobic surface to 25–50 cal/mol in free energy (57–59). The surface area analysis is consistent with the results from the Rosetta scoring and supports our interpretation that immunogenicity can arise from peptide substitutions that yield higher energies and subsequently stronger TCR binding affinities.

Testing Performance on Data Not Used in Training

Our structure-based neural network outperformed sequence-based tools when classifying the training data. Ideally, large test sets of neoantigens would be available for further evaluation. Unfortunately, the number of well-categorized neoantigens is still small, and further reduced by our restriction on nonamers presented by HLA-A2. A recent survey identified ∼1,400 potential neoantigens (60), 291 of which were nonameric peptides presented by HLA-A2 (**Table S4**). Of these, 17 were reported as immunogenic. Although the numbers are small, in evaluating these peptides, the structure-based neural network outperformed sequence-based approaches when considering the impact of a mutation on immunogenicity (**Figure 5B**). Performance was only marginally favorable (AUC of 0.60), but again the dataset is small, and these epitopes are not curated or vetted to the same extent as those recorded in the IEDB.

Evaluation of Select Neoantigens and Their Wild-Type Counterparts

To illustrate how structural information can help inform the determination of immunogenicity and provide hypotheses for testing and improving our approach, we examined structural models of mutant peptides and their wild-type counterparts, choosing select epitopes that could demonstrate the principles encoded by our structure-based assessments as well as highlight areas for improvement.

The LIIPFIHLI epitope was identified in a study of heterologous T cell recognition of melanoma neoantigens, and incorporates a cysteine to phenylalanine substitution at position 5 (32). The structural models show the position 5 side chain to be almost fully extended, with the phenylalanine mutant resulting in the exposure of an additional 90 Å^2 of hydrophobic surface area, which could promote stronger TCR binding due to the hydrophobic effect (**Figure 6A**). The structure-based neural network indeed predicted the mutation would improve immunogenicity, with an increase in score of 0.14.

ALGALTVWL was identified in a study of neoantigens in breast cancer (61). This epitope replaces an unfavorable histidine in the primary HLA-A2 anchor residue at position 9 with a preferred leucine, and thus improves peptide binding to the MHC protein. Demonstrating how our approach captures not only TCR binding to the peptide/MHC complex but also peptide binding to MHC, the neural network correctly predicted the leucine variant to be more immunogenic than the wild-type peptide with the position 9 histidine, with an increase in the immunogenicity score of 0.19. Structurally, the model predicts the leucine at position 5 and tryptophan at position 8 to be solvent exposed, likely responsible for forming interactions with neoantigen-specific TCRs (**Figure 6B**).

QLMQLIEPA was identified in a large-scale study of neoantigens in melanoma (62). The peptide substitutes a glutamate for a glycine at position seven of the peptide. The neural network predicted the mutated peptide would be less immunogenic than the wild-type peptide, with a change in score of −0.16. Indeed, no T cell responses were identified with QLMQLIEPA. The modeling indicates the new glutamate would be fully exposed (**Figure 6C**), which as discussed above would likely require close charge complementarity by an incoming TCR, consistent with the reduced likelihood of immunogenicity.

Lastly, ALIDLSSGL was identified in the same study as QLMQLIEPA (62). The peptide incorporates a leucine to a proline substitution at position 5 of the peptide. The neural network predicted the neoantigen to be more immunogenic than the wild-type peptide; however, no T cell responses were identified with the peptide. This false positive may be due to a conformational impact of the proline mutation: the structural modeling predicts that both mutant and wild-type peptides adopt very similar pathways through the HLA-A2 binding groove, with a minor impact on the position of Asp4 (**Figure 6D**). However, a proline substitution could impact the peptide backbone in a fashion not captured by our modeling scheme, possibly indicating a need for more exhaustive conformational sampling in the structural modeling as discussed below.

DISCUSSION

Identifying immunogenic peptides in cellular immunity remains a challenge, particularly in the development of personalized "neoantigen" vaccines based on individual tumor genomes. Here, we tested the hypothesis that immunogenicity predictions

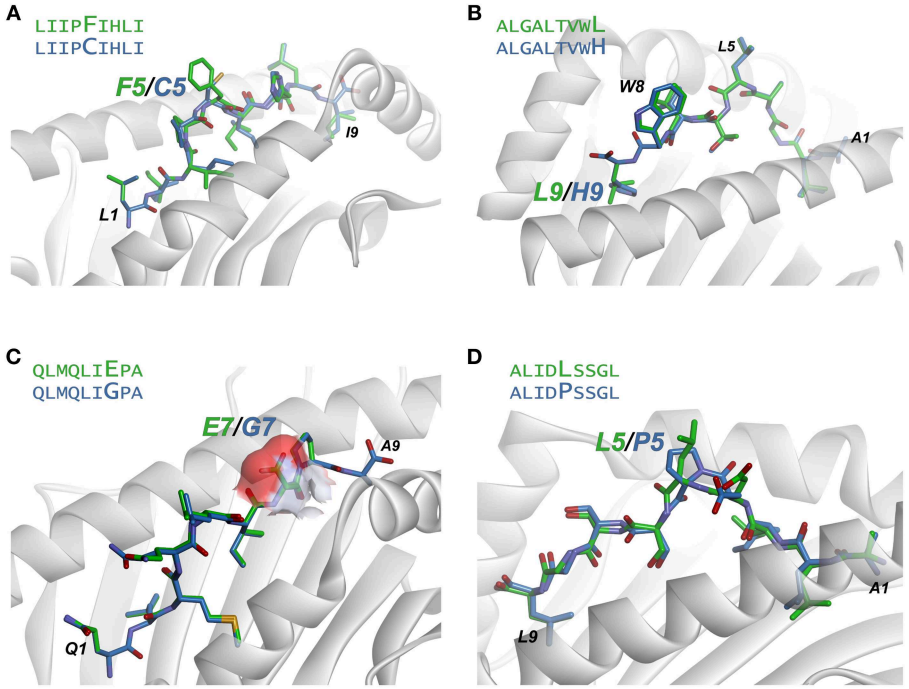


FIGURE 6 | Examination of modeled structures of select neoantigens and their wild-type counterparts. **(A)** The neoantigen LIIPFIHLI substitutes a phenylalanine for a cysteine at position 5. The position 5 side chain is predicted to extend from the top of a bulge in the peptide, and the mutation results in an increase in exposed hydrophobic surface of 90 Å^2. **(B)** The neoantigen ALGALTVWL substitutes a leucine for a histidine at position 9, "fixing" the second primary anchor residue and improving peptide binding to HLA-A2. The leucine at position 5 and tryptophan at position 8 are predicted to extend up from the peptide backbone to form interactions with T cell receptors. **(C)** The neoantigen QLMQLIEPA substitutes a glutamate for a glycine at position 7. The glutamate 7 side chain in the mutant peptide is predicted to be fully exposed, increasing exposed charged surface area as indicated by the surface area representation for the position 7 side chain. **(D)** The neoantigen ALIDLSSGL replaces a proline with a leucine at position 5 of the peptide. No conformational consequences are predicted for the mutation.

for peptides presented by class I MHC proteins could be improved by considering features encoded by the structure of the peptide/MHC complex, rather than by features of the peptide amino acid sequence alone. Our hypothesis was predicated on the notion that immunogenicity is influenced by both peptide binding to the MHC protein as well as TCR binding to the peptide/MHC complex. Prediction methods for the former are well-developed, whereas prediction methods for the latter are in their infancy. Some peptide features that possibly promote TCR binding have been identified (13–17), but we suggest these and other important features are best interpreted, and ultimately predicted, by examining structures and their physicochemical properties and energies. Identifying these features and the magnitudes required are particularly important for tumor neoantigens, as the bar for establishing "difference from self" for neoantigens is higher than for antigens from viruses or other pathogens due to the various tolerance mechanisms that limit self-reactivity.

To explore our hypothesis, we developed a rapid and accurate procedure for modeling the structures of nonamers bound to the class I MHC protein HLA-A2. Our procedure performed well-compared to previously published methods and was suitably rapid for use with large peptide databases. Following this, we assembled a database of immunogenic and non-immunogenic peptides, including in the latter HLA-A2-incompatible peptides predicted to be very weak binders. We used the structural modeling procedure to model the nearly 4,000 peptide/HLA-A2 complexes in this database. We then trained an artificial neural network for predicting immunogenicity that relied on energetic features determined by the structures. As potential terms to be incorporated, we included features such as van der Waals interactions, hydrophobic solvation, Coulombic potentials, hydrogen bond energies, side chain rotamer energies, as well as solvent accessible surface areas. We designed our approach to capture both peptide binding to the MHC protein as well as TCR binding to the peptide/MHC, as both are determined by structural and physicochemical features, and in determining immunogenicity, strong TCR binding can compensate for weak peptide-MHC binding and *vice versa* (11, 12, 36).

Our structural approach outperformed other prediction tools, including comparisons with sequence-based neural networks trained on the same peptide datasets. Our model also outperformed publicly available immunogenicity prediction tools, as well as predictions based on peptide-MHC binding affinity. Beyond immunogenicity prediction, an important added benefit of our approach is the availability of structural models to aid in interpreting results. The utility of these models is found when comparing results for mutant peptides and their wild-type

counterparts, as the models can indicate the types of structural alterations that impact TCR recognition of a mutated vs. wild-type peptide/MHC complex. In cases in which the mutation improves immunogenicity by enhancing peptide binding to the MHC protein, the models still provide indications about which peptide features may be important for TCR recognition.

A key observation that emerged from our analysis is that, excluding non-HLA-A2 binders, immunogenic peptides possessed higher total energies than non-immunogenic complexes. We believe this result is an important indicator of how structure and energy can influence immunogenicity beyond simply enhancing peptide affinity for the MHC protein: amino acid substitutions that impart a higher energy onto a peptide/MHC complex yield ligands that have more energy to release upon TCR binding, thus translating into stronger TCR binding affinities. Higher energy would, for example, be imparted by removing exposed charges or increasing exposed hydrophobic surface area. Given a compatible TCR, this high energy would be released upon binding, contributing to a favorable TCR-peptide/MHC binding free energy, i.e., better TCR binding affinity. From this interpretation, neoantigens have a greater likelihood of being immunogenic not simply when they are chemically or structurally "different" from a corresponding wild-type peptide, but different in ways that promote strong TCR binding. In addition to helping explain immunogenicity, this interpretation connects immunogenicity to well-understood physical attributes of biomolecular recognition, and is consistent with studies on the composition and structural properties of protein-protein interfaces and how they differ from other protein surfaces, as well as the properties of "hot spots" in protein-protein interfaces (63, 64).

Although we consider our results promising, improvements are necessary before a structural modeling/energetic scoring methodology can be widely deployed. Our rapid modeling procedure, while matching or even exceeding the performance of previous approaches, did not capture all the observed structural changes that occur in response to peptide modification. We expect that incorporating more exhaustive conformational sampling will yield superior models. Although this will increase computational time, this impact will be offset by ongoing improvements in computing hardware and sampling methods. Growth in the number of crystallographic structures of peptides bound to class I MHC proteins, particularly structures of closely related peptide pairs, will help benchmark the accuracy of structural modeling. Our study was limited to nonamers and HLA-A2, primarily because of the large amount of structural and immunological data available for nonamer/HLA-A2 complexes. In the absence of more structural data, extension to peptides of other lengths and other HLA haplotypes will call for even more sophisticated modeling.

Another area for methodological improvements is in the energy functions and other terms used to evaluate peptide/MHC models. We relied upon an energy function and set of terms frequently used in the analysis and design of protein structures. As with modeling procedures, more complex means to assess protein structures and energies are available, and these undergo regular refinements. Incorporation of additional or more sophisticated energetic terms (e.g., electrostatic surface potentials, more accurate approaches to computing solvation energies, and consideration of changes in peptide flexibility that occur upon mutation) could thus also be explored.

Attention should also be focused toward generating datasets that can be used to train models aiming to predict peptide immunogenicity (65). Epitopes with verified, strong immune responses can be found in the IEDB as we relied upon here, and efforts such as the Cancer Antigenic Peptide Database aim to tabulate immunogenic neoantigens (66). However, experimentally validated immunogenic neoantigens remain rare, and some efforts aimed at building CTL epitope databases prioritize peptide processing and MHC presentation over T cell recognition (67). Additionally, prediction tools also require knowledge of non-immunogenic epitopes, particularly those that bind well to MHC proteins yet do not favor TCR binding. We relied on a list of self-peptides which we hypothesized would be dominated by such epitopes, but to some extent would also include ones that are well-recognized by TCRs. This would include not only self-peptides that would drive negative selection, but as our peptide list was derived from immortalized HeLa cells (52, 53), it also likely includes various HPV epitopes and associated neoantigens. The fact that we recovered previously identified positional differences in hydrophobicity between immunogenic and non-immunogenic peptides suggests that the influence of such peptides in our self-dataset is small. However, a better accounting of peptides (ideally derived from healthy tissues) which bind well to class I MHC proteins yet do not promote strong immunogenicity when tested across multiple T cell populations is needed.

Lastly, our prediction efforts were centered on the ability to elicit strong $CD8^+$ T cell responses. Eliciting $CD8^+$ T cell responses remains a goal of peptide-based vaccine efforts for both pathogens and cancer (68), and $CD8^+$ T cell activity is associated with antitumor immunity (69). However, additional factors will influence successful peptide-based vaccines. A deeper understanding of the strengths and types of responses associated with successful antitumor immunity will further allow prediction efforts to improve.

In conclusion, we have explored the potential for large-scale structure-based modeling and energetic scoring for predicting peptide immunogenicity, with an emphasis on cancer neoantigens. Our approach outperformed other approaches and although it is a proof-of-concept, the avenues for improvement are clear and actionable. The structural modeling allows for insights into immunogenicity lacking from other approaches. Furthermore, because it is fully atomistic, the approach can grow to incorporate complexities not addressable via sequence considerations alone, such as those arising from peptides incorporating post translational modifications or non-standard amino acids.

MATERIALS AND METHODS

Structural Modeling of HLA-A2 Presented Peptides

Structural modeling of peptide/HLA-A2 complexes was performed with PyRosetta using the Talaris2014 energy function (38, 39). The desired peptide sequence was computationally introduced into HLA-A2, using PDB IDs 1DUZ, 3QFD, and 5FA3 as templates (40–42). This was followed by 50 Monte Carlo-based simulated annealing sidechain and peptide backbone minimization steps using the LoopMover_Refine_CCD protocol, generating 10 independent decoys per peptide from each starting template. The large number of resulting packing operations introduced some minor variability when scoring the models. Therefore, the unweighted score terms for the three lowest scoring trajectories were averaged and used for neural network inputs. Solvent accessible surface area calculations were performed with PyRosetta and Discovery Studio using a 1.4 Å radius probe. The modeling procedure is available as **Supplementary File 1**.

Dataset Collection

Our structural database for evaluating modeling strategies consisted of high resolution (<3.0 Å) nonameric peptide/HLA-A2 structures within the PDB. Structures in this dataset were selected for strong electron density as determined by visual inspection using Coot for calculating $2F_o$-F_c density maps (70). Our final database contained 53 structures presenting different peptide epitopes.

The neural network training set contained 3,955 nonameric peptides collected from published sources. For self-peptides categorized as non-immunogenic we used lists of peptides identified via mass spectrometry analysis of human HeLa cells transfected with soluble HLA-A2 (52, 53). HLA-A2 incompatible nonamers were obtained from IEDB training sets (24). Immunogenic peptides were selected from IEDB to ensure quality of data and minimize false positives by selecting only HLA-A2-restricted nonamers with a positive IFNγ ELISpot with a response frequency starting at 50. The test dataset was derived from a recent review of neoantigens (60), again selecting only nonamers presented by HLA-A2 for evaluation, resulting in a dataset consisting of 291 candidate neoantigens.

ARTIFICIAL NEURAL NETWORK TRAINING

Two-layer feed-forward networks were trained with the scaled conjugate gradient back-propagation trainscg tool in MATLAB 2017b. Training and evaluation of neural network architectures was performed using a nested 5-fold cross-validation procedure (23). The peptides in the training dataset were randomly split into five sets of training, validation, and test data. The splitting was performed such that all sets had approximately the same distribution of non-binding, self, and immunogenic peptides. With the binary classification of immunogenic or non-immunogenic (with non-immunogenic incorporating self and non-binding peptides), the training data were used to perform feed-forward and back propagation. The validation set defined the stopping criteria for the network training, and the test set evaluated performance via AUC. Sets were rotated to ensure each was used in training, validation, and testing. To maintain an equal distribution of classifiers and eliminate bias for non-immunogenic peptides, immunogenic peptides in the training sets, but not testing or validation sets, were randomly oversampled.

The structure-based neural network architecture used was a conventional feed-forward network with an input layer containing 80–117 neurons, one hidden layer with 1–10 neurons, and a single neuron output layer. The neurons in the input layer describe structural and structure-derived energetic features of the nine amino acids in the peptide sequence, with each amino acid represented by up to 11 neurons. The remaining 18 neurons describe global structural and structure-derived energetic features of the entire peptide/HLA-A2 complex. The structural and energetic features were those that comprise the Talaris2014 energy function (39) or derived from the structure as listed in **Table S3** [described in (56)]. For each of the five training and test sets, a series of network trainings were performed each with a different number of hidden neurons (2, 3, 4, 6, 8, and 10) and a different number of input neurons. Finally, a single network with the highest test performance was selected.

For control networks that considered peptide sequence or amino acid hydropathy values, we encoded peptide sequences in 20 × 9 sparse matrices encoding peptide sequence or 1 × 9 matrices containing Wimley-White hydropathy values (55) corresponding to the amino acid at each position. These matrices were used to train networks of the same architecture (except they relied on 180 or 9 input nodes) subject to the same cross validation procedure.

Protein Crystallization and Structure Determination

Purified complexes of ILNAMIVKI, ILNAMIAKI, and ILNAMITKI with HLA-A2 were generated by refolding recombinant heavy chain and β_2-microglobulin from bacterially-produced inclusion bodies according to standard procedures (71), followed by purification using anion exchange and size-exclusion chromatography. Peptides were synthesized commercially by AAPTEC at >90% purity. Crystals of the ILNAMIVKI complex were grown by hanging-drop vapor diffusion at 23°C in 20% PEG 8000, 100 mM MES pH 6.5, 200 mM magnesium acetate. Crystals of the ILNAMITKI complex were grown at 23°C in 20% PEG 3350, 100 mM HEPES pH 7.5. Crystals of the ILNAMIAKI complex were grown at 4°C in 15% PEG 3350, 100 mM MES pH 6.5. Crystals were harvested and cryoprotected in ∼15% glycerol and ∼85% mother liquor and then immediately frozen in liquid nitrogen.

Data for the ILNAMIVKI complex were collected at the 22ID beamline at the Advanced Photon Source at Argonne National

Laboratories. Data for ILNAMITKI and ILNAMIAKI complexes were collected at the 23ID-D beamline at the Advanced Photon Source. For the ILNAMIVKI and ILNAMITKI structures, data integration and scaling were performed using the HKL2000 suite. Integration and scaling of the ILNAMIAKI data were performed with DIALS (72). The structures were solved by molecular replacement using Phaser in PHENIX (73), with PDB 3PWL used as a search model for ILNAMIVKI and ILNAMIAKI (74), while 1TVH was used as a search model for ILNAMITKI (75). Peptides were deleted from search models prior to molecular replacement. Multiple steps of restrained refinement were performed using PHENIX Refine (76). Evaluation of models and fitting to maps were performed using Coot (70). MolProbity was used to evaluate structures during and after refinement (77).

DATA AVAILABILITY

The experimental structural data generated for this study can be found in the Protein Data Bank under accession numbers 6OPD, 6PTB, and 6PTE.

AUTHOR CONTRIBUTIONS

TR developed the modeling, scoring, and prediction approaches. GK analyzed data and performed control analyses. JD analyzed data and compared to published results. GK, AS, LD, and AA performed X-ray crystallography. BB oversaw and directed the research. The manuscript was written and edited by all authors.

FUNDING

Supported by NIH grant R35GM118166 to BB and a Notre Dame Research Like a Champion award to LD. TR and JD were supported by NIH grants TR001107 and TR001108. GK was supported by NIH grant TR002529.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02047/full#supplementary-material

REFERENCES

1. Rosenberg SA, Yang JC, Restifo NP. Cancer immunotherapy: moving beyond current vaccines. *Nat Med.* (2004) 10:909–15. doi: 10.1038/nm1100
2. Bräunlein E, Krackhardt AM. Identification and characterization of neoantigens as well as respective immune responses in cancer patients. *Front. Immunol.* (2017) 8:1702. doi: 10.3389/fimmu.2017.01702
3. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*. (2017) 547:217–21. doi: 10.1038/nature22991
4. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. (2017) 547:222–6. doi: 10.1038/nature23003
5. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun*. (2016) 7:13404. doi: 10.1038/ncomms13404
6. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity.* (2017) 46:315–26. doi: 10.1016/j.immuni.2017.02.007
7. Duan F, Duitama J, Al Seesi S, Ayres CM, Corcelli SA, Pawashe AP, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J Exp Med.* (2014) 211:2231–48. doi: 10.1084/jem.20141308
8. Toor JS, Rao AA, McShan AC, Yarmarkovich M, Nerli S, Yamaguchi K, et al. A recurrent mutation in anaplastic lymphoma kinase with distinct neoepitope conformations. *Front Immunol*. (2018) 9:99. doi: 10.3389/fimmu.2018.00099
9. Fritsch EF, Rajasagi M, Ott PA, Brusic V, Hacohen N, Wu CJ. HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunol Res.* (2014) 2:522–9. doi: 10.1158/2326-6066.CIR-13-0227
10. Croft NP, Smith SA, Pickering J, Sidney J, Peters B, Faridi P, et al. Most viral peptides displayed by class I MHC on infected cells are immunogenic. *Proc Natl Acad Sci USA*. (2019) 116:3112–7. doi: 10.1073/pnas.1815239116
11. Feltkamp MCW, Vierboom MPM, Kast WM, Melief CJM. Efficient MHC class I-peptide binding is required but does not ensure MHC class I-restricted immunogenicity. *Mol Immunol.* (1994) 31:1391–401. doi: 10.1016/0161-5890(94)90155-4
12. Ochoa-Garay J, McKinney DM, Kochounian HH, McMillan M. The ability of peptides to induce cytotoxic T cells *in vitro* does not strongly correlate with their affinity for the H-2Ld molecule: implications for vaccine design and immunotherapy. *Mol Immunol.* (1997) 34:273–81. doi: 10.1016/S0161-5890(97)00019-9
13. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. (2013) 9:e1003266. doi: 10.1371/journal.pcbi.1003266
14. Chowell D, Krishna S, Becker PD, Cocita C, Shu J, Tan X, et al. TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes. *Proc Natl Acad Sci USA*. (2015) 112:E1754–E62. doi: 10.1073/pnas.1500973112
15. Tung C-W, Ziehm M, Kämper A, Kohlbacher O, Ho S-Y. POPISK: T-cell reactivity prediction using support vector machines and string kernels. *BMC Bioinform*. (2011) 12:446. doi: 10.1186/1471-2105-12-446
16. Pang Y-P, Elsbernd LR, Block MS, Markovic SN. Peptide-binding groove contraction linked to the lack of T cell response: using complex structure and energy to identify neoantigens. *ImmunoHorizons.* (2018) 2:216–25. doi: 10.4049/immunohorizons.1800048
17. Kim S, Kim HS, Kim E, Lee MG, Shin EC, Paik S, et al. Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Ann Oncol*. (2018) 29:1030–6. doi: 10.1093/annonc/mdy022
18. Trolle T, Nielsen M. NetTepi: an integrated method for the prediction of T cell epitopes. *Immunogenetics.* (2014) 66:449–56. doi: 10.1007/s00251-014-0779-0
19. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol*. (2017) 199:3360–8. doi: 10.4049/jimmunol.1700893
20. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. (2016) 32:511–7. doi: 10.1093/bioinformatics/btv639
21. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics.* (2012) 64:177–86. doi: 10.1007/s00251-011-0579-8
22. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics.* (1999) 50:213–9. doi: 10.1007/s002510050595

23. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*. (2003) 12:1007–17. doi: 10.1110/ps.0239403
24. Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front Immunol*. (2017) 8:278. doi: 10.3389/fimmu.2017.00278
25. Dill KA. Dominant forces in protein folding. *Biochemistry*. (1990) 29:7133–55. doi: 10.1021/bi00483a001
26. Kauzmann W. Some factors in the interpretation of protein denaturation. *Adv Protein Chem*. (1959) 14:1–63. doi: 10.1016/S0065-3233(08)60608-7
27. Southall NT, Dill KA, Haymet ADJ. A view of the hydrophobic effect. *J Phys Chem B*. (2002) 106:521–33. doi: 10.1021/jp015514e
28. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol*. (2000) 10:153–9. doi: 10.1016/S0959-440X(00)00065-8
29. Bosshard HR, Marti DN, Jelesarov I. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J Mol Recognit*. (2004) 17:1–16. doi: 10.1002/jmr.657
30. Blevins SJ, Pierce BG, Singh NK, Riley TP, Wang Y, Spear TT, et al. How structural adaptability exists alongside HLA-A2 bias in the human αβ TCR repertoire. *Proc Natl Acad Sci USA*. (2016) 113:E1276–E85. doi: 10.1073/pnas.1522069113
31. Singh NK, Riley TP, Baker SCB, Borrman T, Weng Z, Baker BM. Emerging concepts in TCR specificity: rationalizing and (Maybe) predicting outcomes. *J Immunol*. (2017) 199:2203–13. doi: 10.4049/jimmunol.1700744
32. Strønen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science*. (2016) 352:1337–41. doi: 10.1126/science.aaf2288
33. Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC. Four A6–TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity*. (1999) 11:45–56. doi: 10.1016/S1074-7613(00)80080-1
34. Hausmann S, Biddison WE, Smith KJ, Ding YH, Garboczi DN, Utz U, et al. Peptide recognition by two HLA-A2/Tax11–19–specific T cell clones in relationship to their MHC/peptide/TCR crystal structures. *J Immunol*. (1999) 162:5389–97.
35. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*. (2014) 515:572–6. doi: 10.1038/nature14001
36. Bowerman NA, Crofts TS, Chlewicki L, Do P, Baker BM, Christopher Garcia K, et al. Engineering the binding properties of the T cell receptor:peptide:MHC ternary complex that governs T cell activity. *Mol Immunol*. (2009) 46:3000–8. doi: 10.1016/j.molimm.2009.06.012
37. Schueler-Furman O, Elber R, Margalit H. Knowledge-based structure prediction of MHC class I bound peptides: a study of 23 complexes. *Fold Design*. (1998) 3:549–64. doi: 10.1016/S1359-0278(98)00070-4
38. Chaudhury S, Lyskov S, Gray JJ. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*. (2010) 26:689–91. doi: 10.1093/bioinformatics/btq007
39. O'Meara MJ, Leaver-Fay A, Tyka MD, Stein A, Houlihan K, DiMaio F, et al. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with rosetta. *J Chem Theory Comput*. (2015) 11:609–22. doi: 10.1021/ct500864r
40. Khan AR, Baker BM, Ghosh P, Biddison WE, Wiley DC. The structure and stability of an HLA-A*0201/octameric tax peptide complex with an empty conserved peptide-N-terminal binding site. *J Immunol*. (2000) 164:6398–405. doi: 10.4049/jimmunol.164.12.6398
41. Borbulevych OY, Insaidoo FK, Baxter TK, Powell DJ, Jr., Johnson LA, Restifo NP, et al. Structures of MART-1(26/27–35) Peptide/HLA-A2 complexes reveal a remarkable disconnect between antigen structural homology and T cell recognition. *J Mol Biol*. (2007) 372:1123–36. doi: 10.1016/j.jmb.2007.07.025
42. Remesh SG, Andreatta M, Ying G, Kaever T, Nielsen M, McMurtrey C, et al. Unconventional peptide presentation by major histocompatibility complex (MHC) class I Allele HLA-A*02:01: BREAKING CONFINEMENT. *J Biol Chem*. (2017) 292:5262–70. doi: 10.1074/jbc.M117.776542
43. Tong JC, Tan TW, Ranganathan S. Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci Publ Protein Soc*. (2004) 13:2523–32. doi: 10.1110/ps.04631204
44. Raveh B, London N, Schueler-Furman O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins Struc Funct Bioinform*. (2010) 78:2029–40. doi: 10.1002/prot.22716
45. Rosenfeld R, Zheng Q, Vajda S, DeLisi C. Computing the structure of bound peptides: application to antigen recognition by class I major histocompatibility complex receptors. *J Mol Biol*. (1993) 234:515–21. doi: 10.1006/jmbi.1993.1607
46. Sezerman U, Vajda S, DeLisi C. Free energy mapping of class I MHC molecules and structural determination of bound peptides. *Protein Sci A Public Protein Soc*. (1996) 5:1272–81. doi: 10.1002/pro.5560050706
47. Rognan D, Lauemøller SL, Holm A, Buus S, Tschinke V. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J Medic Chem*. (1999) 42:4650–8. doi: 10.1021/jm9910775
48. Park M-S, Park SY, Miller KR, Collins EJ, Lee HY. Accurate structure prediction of peptide–MHC complexes for identifying highly immunogenic antigens. *Mol Immunol*. (2013) 56:81–90. doi: 10.1016/j.molimm.2013.04.011
49. Bui HH, Schiewe AJ, Von Grafenstein H, Haworth IS. Structural prediction of peptides binding to MHC class I molecules. *Proteins Struc Funct Genet*. (2006) 63:43–52. doi: 10.1002/prot.20870
50. Fagerberg T, Cerottini J-C, Michielin O. Structural prediction of peptides bound to MHC class I. *J Mol Biol*. (2006) 356:521–46. doi: 10.1016/j.jmb.2005.11.059
51. Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med*. (2013) 19:747. doi: 10.1038/nm.3161
52. Trolle T, McMurtrey CP, Sidney J, Bardet W, Osborn SC, Kaever T, et al. The length distribution of class I–restricted T cell epitopes is determined by both peptide supply and MHC allele–specific binding preference. *J Immunol*. (2016) 196:1480–7. doi: 10.4049/jimmunol.1501721
53. McMurtrey C, Trolle T, Sansom T, Remesh SG, Kaever T, Bardet W, et al. *Toxoplasma gondii* peptide ligands open the gate of the HLA class I binding groove. *eLife*. (2016) 5:e12556. doi: 10.7554/eLife.12556
54. Madura F, Rizkallah PJ, Holland CJ, Fuller A, Bulek A, Godkin AJ, et al. Structural basis for ineffective T-cell responses to MHC anchor residue-improved "heteroclitic" peptides. *Eur J Immunol*. (2015) 45:584–91. doi: 10.1002/eji.201445114
55. White SH, Wimley WC. Membrane protein folding and stability: physical principles. *Ann Rev Biophys Biomol Struc*. (1999) 28:319–65. doi: 10.1146/annurev.biophys.28.1.319
56. Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, et al. The rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput*. (2017) 13:3031–48. doi: 10.1021/acs.jctc.7b00125
57. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*. (1991) 11:281–96. doi: 10.1002/prot.340110407
58. Sharp K, Nicholls A, Fine R, Honig B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science*. (1991) 252:106–9. doi: 10.1126/science.2011744
59. Sharp KA, Nicholls A, Friedman R, Honig B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry*. (1991) 30:9686–97. doi: 10.1021/bi00104a017
60. Bjerregaard A-M, Nielsen M, Jurtz V, Barra CM, Hadrup SR, Szallasi Z, et al. An analysis of natural T cell responses to predicted tumor neoepitopes. *Front Immunol*. (2017) 8:1566. doi: 10.3389/fimmu.2017.01566
61. Zhang X, Kim S, Hundal J, Herndon JM, Li S, Petti AA, et al. Breast cancer neoantigens can induce $CD8^+$ T-cell responses and antitumor immunity. *Cancer Immunol Res*. (2017) 5:516–23. doi: 10.1158/2326-6066.CIR-16-0264
62. Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, et al. Isolation of neoantigen-specific T cells from

tumor and peripheral lymphocytes. *J Clin Invest.* (2015) 125:3981–91. doi: 10.1172/JCI82416
63. Birtalan S, Fisher RD, Sidhu SS. The functional capacity of the natural amino acids for molecular recognition. *Mol BioSyst.* (2010) 6:1186–94. doi: 10.1039/b927393j
64. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol.* (1998) 280:1–9. doi: 10.1006/jmbi.1998.1843
65. Eklund AC, Szallasi Z. Computational prediction of neoantigens: do we need more data or new approaches? *Ann Oncol.* (2018) 29:799–801. doi: 10.1093/annonc/mdy070
66. Vigneron N, Stroobant V, Van den Eynde BJ, van der Bruggen P. Database of T cell-defined human tumor antigens: the 2013 update. *Cancer Immunity.* (2013) 13:15. Available online at: https://cancerimmunolres.aacrjournals.org/content/canimmarch/13/3/15
67. Llano A, Williams A, Overa A, Silva-Arrieta S, Brander C. Best-characterized HIV-1 CTL epitopes: the 2013 update. In: Yusim K, Korber B, Brander C, Barouch D, de Boer R, Haynes BF, Koup R, Moore JP, Walker BD, editors. *HIV Molecular Immunology 2013.* Los Alamos, NM: Theoretical Biology and Biophysics Group; Los Alamos National Laboratory (2013). p. 3–19.
68. Cosma G, Eisenlohr L. CD8+ T-cell responses in vaccination: reconsidering targets and function in the context of chronic antigen stimulation [version 1; referees: 2 approved]. *F1000Res.* (2018) 7:508. doi: 10.12688/f1000research.14115.1
69. Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent advances in targeting CD8 T-cell immunity for more effective cancer immunotherapy. *Front Immunol.* (2018) 9:14. doi: 10.3389/fimmu.2018.00014
70. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr Section D.* (2010) 66:486–501. doi: 10.1107/S0907444910007493
71. Davis-Harrison RL, Armstrong KM, Baker BM. Two different T cell receptors use different thermodynamic strategies to recognize the same peptide/MHC ligand. *J Mol Biol.* (2005) 346:533–50. doi: 10.1016/j.jmb.2004.11.063
72. Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr Section D.* (2018) 74:85–97. doi: 10.1107/S2059798317017235
73. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Echols N, Headd JJ, et al. The Phenix software for automated determination of macromolecular structures. *Methods.* (2011) 55:94–106. doi: 10.1016/j.ymeth.2011.07.005
74. Borbulevych OY, Piepenbrink KH, Baker BM. Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics. *J Immunol.* (2011) 186:2950–8. doi: 10.4049/jimmunol.1003150
75. Borbulevych OY, Baxter TK, Yu Z, Restifo NP, Baker BM. Increased immunogenicity of an anchor-modified tumor-associated antigen is due to the enhanced stability of the peptide/MHC complex: implications for vaccine design. *J Immunol.* (2005) 174:4812–20. doi: 10.4049/jimmunol.174.8.4812
76. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr Section D.* (2012) 68:352–67. doi: 10.1107/S0907444912001308
77. Chen VB, Arendall WB, III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Section D.* (2010) 66:12–21. doi: 10.1107/S0907444909042073

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 05 November 2019
doi: 10.3389/fimmu.2019.02472

The Immunogenicity and Anti-tumor Efficacy of a Rationally Designed Neoantigen Vaccine for B16F10 Mouse Melanoma

Yan Zhang[1], *Zhibing Lin*[1], *Yuhua Wan*[1], *Huaman Cai*[1], *Li Deng*[2] *and Rongxiu Li*[1,2,3]*

[1] State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [2] Shanghai HyCharm Inc., Shanghai, China, [3] Engineering Research Center of Cell and Therapeutic Antibody, Ministry of Education, Shanghai, China

Tumor neoantigens are ideal targets for cancer immunotherapy as they are recognized by host immune system as foreigners and can elicit tumor-specific immune responses. However, existing strategies utilizing RNA or long peptides for the neoantigen vaccines render limited immune responses since only 20–30% of neoantigens predicted *in silico* to bind MHC I molecules are capable of eliciting immune responses with the majority of responding T cells are $CD4^+$. Therefore, it warrants further exploration to enhance neoantigen-specific $CD8^+$ T cells responses. Since neoantigens are naturally weak antigens, we asked whether foreign T help epitopes could enhance their immunogenicity. In present study, we chose 4 weak B16F10 neoantigens as vaccine targets, and fused them to the transmembrane domain of diphtheria toxin, namely DTT-neoAg. Strikingly, the vaccine elicited anti-tumor $CD8^+$ T cells responses and enhanced tumor infiltration of both T cells and NK cells. Impressively, DTT-neoAg vaccine significantly deterred tumor growth with the inhibition rate reached 88% in the preventive model and 100% in the therapeutic model at low dose of tumor challenge. Furthermore, after second challenge with higher dose of tumor cells, 33.3% of the immunized mice remained tumor-free for 6 months in the therapeutic model. Because DTT is a non-toxic domain of diphtheria toxin, it may be not of great concern in terms of safety as a Th epitope provider. Thus, the fusion strategy employed by this study may become a feasible and powerful approach for development of personalized cancer vaccines.

Keywords: cancer vaccine, immune response, tumor neoantigen, B16F10 melanoma, helper T cell, cytotoxic T lymphocytes

OPEN ACCESS

Edited by:
John M. Maris,
University of Pennsylvania,
United States

Reviewed by:
María Marcela Barrio,
Fundación Cáncer, Argentina
Aparna Rao,
University of Pittsburgh, United States

***Correspondence:**
Rongxiu Li
rxli@sjtu.edu.cn

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *20 May 2019*
Accepted: *03 October 2019*
Published: *05 November 2019*

Citation:
Zhang Y, Lin Z, Wan Y, Cai H, Deng L and Li R (2019) The Immunogenicity and Anti-tumor Efficacy of a Rationally Designed Neoantigen Vaccine for B16F10 Mouse Melanoma. Front. Immunol. 10:2472. doi: 10.3389/fimmu.2019.02472

INTRODUCTION

Genomic mutations altering signal transduction pathways that control cell proliferation or apoptosis can cause development of cancers (1). The mutated proteins can give rise to novel antigens, so called neoantigens when they are processed and presented to T cells (2). Neoantigen-specific T cells are found in tumors as well as in peripheral blood of cancer patients (3), and they are the principle mechanism that underlies clinical responses to many standard treatments and immunotherapeutic interventions including checkpoint blockade (4) and adoptive T cell transfer (5).

Tumor neoantigens are attractive targets for cancer vaccine design (6). Both preclinical (7–11) and early phase clinical studies (12–14) found that neoantigen-based poly-epitope vaccines can substantially expand the tumor-specific T cell pools, and steer the immune system to the selective destruction of cancers with limited off-target toxicities, which leads to cancer regression and long-term tumor-free survival.

With the advance of next generation sequencing technology, non-synonymous mutations can be identified by whole exome sequencing, and their expression can be determined by RNA-Seq (8). Nevertheless, to choose the right neoantigen epitopes for the vaccine design is not an easy task at all as over 90 percent of neoantigens are patient-specific (2) and multiple neoantigens are needed to address clonal heterogeneity of tumors (15). In addition, only a limited fraction of non-synonymous mutations can naturally lead to activation of neoantigen specific $CD4^+$ (16, 17) or $CD8^+$ (18–20) T cells. These cells are detectable within tumor-infiltrating lymphocytes or in peripheral blood.

Candidate mutation peptides with high affinity to MHC I molecules can be identified with high confidence by sensitive computation algorithms (21). However, only 20–30% of MHC I neo-antigen peptides can stimulate T cell responses by vaccination. Surprisingly, over 90% of the immunogenic MHC I peptides elicit $CD4^+$ T cell responses (7, 8, 22). Although it has been confirmed that some of the neoantigen-specific $CD4^+$ T cells are able to kill tumor cells, the majority of tumor-specific killer T cells identified in patients have been of $CD8^+$ T cells origin (23).

In this study, we asked whether fusion of low-immunogenic neoantigens to DTT, a membrane translocation domain of diphtheria toxin, could enhance antigen-specific immune responses, in particular, $CD8^+$ cytotoxic T cell responses. DTT has been shown to enhance immune responses to self-molecules (24, 25). Indeed, we found MHC I-binding mutation peptides of B16F10 melanoma that failed to elicit cytotoxic T cell responses become highly immunogenic when they are fused to DTT. In addition, we show that the fusion antigens can elicit tumor-specific $CD8^+$ cytotoxic T lymphocytes when formulated with CpG and Alum adjuvants, and enhance $CD8^+$ T cells and NK cells' infiltrating into tumor. This strategy would significantly expand the pool of candidate neoantigens, improve the effectiveness of the neoantigen vaccines and reduce the cost of patient-oriented vaccine design.

MATERIALS AND METHODS

Mice, Mouse Melanoma Cell Line, Adjuvants

C57BL/6 mice (female, 6–8 weeks old, average weight 20 g) were purchased from Slaccas Laboratory Animal Inc. (Shanghai, China), and were housed in a climate controlled facility. All animal studies were performed in accordance with the guidelines approved by the Institutional Animal Care and Use Committee of Shanghai Jiao Tong University. Mouse melanoma cell line B16F10 was purchased from the Cell Bank of Chinese Academy of Sciences (Shanghai, China), and cultured in DMEM (Gibco, USA) supplemented with 10% FBS (Gibco, USA), 100 U/mL penicillin and 100 μg/mL streptomycin (Gibco, USA) at 37°C under a humidified atmosphere of 5% CO_2. Aluminum hydroxide gel (Alum) was purchased from Invitrogen (Invitrogen, USA). CpG ODN1826 (TCCATGACGTTCCTGACGTT) was synthesized by HuaGene (Shanghai, China). All chemical reagents were of analytical grade.

Neoantigen Selection

Four neoantigens were selected for this study based on a recent report from the laboratory of Sahin (7). These neoantigens are expressed in B16F10, and can bind to MHC I, but they are not immunogenic. For validation their existence in our working cell line, the mutated gene fragments were amplified from B16F10 genomic DNA, subjected to Sanger sequencing, and the cognate mutations were confirmed (data not shown).

Expression Vector Construction

Four neoantigens, each contains 27 amino acids with the mutant positioned in the center. The neoantigens are fused in tandem via SG linkers. The DNA fragment encoding the neoantigen fusion was chemically synthesized and cloned into pUC57 (Huagene, Shanghai, China). DTT DNA fragment encoding the amino acid residues 202–378 of diphtheria toxin was previously described (25). The DTT fragment and the neoantigen fusion fragment were connected by GGGGSGGGGS linker sequence with DTT at N-terminal. The corresponding DNA fragment was generated by overlapping PCR with primers listed in **Supplementary Table 1**. DTT was amplified from pUC19-DTT with primers DTT-F and DTT-R. Neoantigen fusion gene was amplified from pUC57-neoAg with primers neoAg-F and neoAg-R. DTT-neoantigen fusion was generated by PCR with DTT and neoantigen fusion fragments as templates, DTT-F and neoAg-R as primers. The resulting recombinant DTT-neoAg fragment was double-digested with *BamH I* and *Xho I*, and cloned into pGEX-6p-1. DTT-wtAg construct was generated in the same way as described above except that the mutant residues are replaced with wild type residues. The neoantigen fragment was also fused to the C-terminal of CTB, the resulting recombinant protein CTB-neoAg was used as an ELISA coating antigen for detection of antibodies against neoantigens in mouse sera vaccinated with DTT-neoAg.

Protein Expression and Purification

pGEX-DTT-neoAg and pGEX-DTT-wtAg were transformed into *E. coli* BL21(DE3), respectively. A single colony was inoculated into 3 mL LB media with 50 μg/mL ampicillin, cultured overnight at 37°C. The culture was expanded into 500 mL of LB media until OD600 nm reached 0.6. IPTG (isopropyl-ß-D-thiogalactoside) was then added to a final concentration of 0.5 mM. The culture was incubated at 16°C for 24 h. The cells were harvested by centrifugation, and the cell pellets were resuspended in phosphate-buffered saline (PBS) and lysed by sonication (60 cycles of 5 s on ice). The lysate was subject to centrifugation at 12 000 × g for 30 min at 4°C, and the supernatant was applied to GST affinity columns. GST tag was removed by PreScission protease cleavage at 4°C for 20 h,

in 50 mM Tris-HCl, 140 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol, pH 7.4. The protein samples were analyzed by 15% ExpressPlus PAGE gels (GenScript, Nanjing, China).

Mice Immunization

Female C57BL/6 mice (6–8 weeks of age, 5–10 per group) were injected subcutaneously into the lateral flank with 30 μg DTT-neoAg or DTT-wtAg, 300 μg Alum, and 30 μg CpG, formulated in 200 μL PBS. The control group of mice were administered with 300 μg Alum and 30 μg CpG. Each mouse received three injections at one- or 2-week intervals. Blood samples were drawn from orbital sinus 1 week after injection.

ELISA for Antibody Detection

To detect antibodies against neoAg, ELISA plates were incubated overnight at 4°C with 100 ng CTB-neoAg, or DTT, or CTB-wtAg in 100 μL sodium carbonate buffer, pH 9.6. The non-specific binding sites were blocked with 100 μL 3% skim milk in PBS + 0.05% Tween 20 at room temperature for 1 h. Subsequently, the mouse sera with indicated dilution were added to the wells and incubated for 1 h at room temperature. The bound antibodies were detected using goat anti-mouse IgG-HRP, or goat anti-mouse IgG1-HRP, or IgG2a-HRP, or IgG2b-HRP, or IgG3-HRP, or IgM-HRP (1:5,000 dilutions, Shanghai Immune Biotech Co. Ltd., Shanghai, China) using 3,3′,5,5′-tetramethylbenzidine (TMB, TIANGEN, Shanghai, China) as substrate. The absorbance at 450 nm was measured by EnSpire 2300 ELISA reader (PerkinElmer, Waltham, MA, USA). The antibody titers are defined as the reciprocals of the highest dilution yielding an optical density of 0.2 or greater than that of pre-immune mouse sera.

Tumor Challenge

For the prophylactic tumor model with DTT-neoantigen vaccination, C57BL/6 mice (n = 5–6, 6–8-weeks old) were subcutaneously immunized on day 0, 12, 24. 9 days after the third immunization, the mice were subcutaneously (s.c.) injected with 1×10^5 B16F10 cells in 100 μL of PBS into the right flank. In prophylactic tumor model with neoantigen peptides (neoAg-pep) or DTT-wtAg vaccination, C57BL/6 mice (n = 6–8, 6–8-weeks old) were subcutaneously immunized on day 0, 10, 20. 7 days after the third immunization, the mice were subcutaneously (s.c.) injected with 1×10^5 B16F10 cells in 100 μL of PBS into the right flank.

For the therapeutic tumor model, 6–8 weeks old C57BL/6 mice (n = 5–10) were first s.c. inoculated with 2.5×10^4 B16F10 cells in 100 μL of PBS. Then they were administered with indicated vaccines 7 days after tumor cell injection and boosted twice at 1-week intervals. Animal appearance and behavior were monitored on a daily basis. Tumor sizes were measured every 2–3 days by calipers, and calculated using equation: volume = [(length) × (width)^2]/2 in mm^3. Ninety days after the first tumor challenge, the tumor-free mice were re-challenged s.c. with 7.5×10^4 B16F10 cells on the left flank and monitored for tumor growth. Seven days after the rechallenge, the mice were injected s.c. with indicated vaccines three times at 1-week intervals. Mice were sacrificed when tumor volumes reached 2,000 mm^3 and recorded as death.

T Cells Proliferation and the Subset Analysis

Single cell suspensions were prepared with immunized mouse spleens and treated with ACK lysis buffer (0.15 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM disodium EDTA, pH 7.2). The cells were plated in a 96-well flat plate at 1×10^5 cells/well in 100 μL of RPMI-1640 medium supplemented with 10% FCS, and stimulated with 30 μg/mL CTB-neoAg protein or 5 μg/mL Con A. After 72 h, the culture was treated with CCK-8 solution following manufacture's instruction (YEASEN, Shanghai, China). The optical density (OD) of cells at wavelength 450 nm was measured with OD at 650 nm as a reference. Stimulation index (SI) was calculated as the ratio of optical density (OD) 450 nm of stimulated cells to that of unstimulated cells.

To detect $CD8^+$ T cells, freshly isolated spleen lymphocytes were stimulated with 30 μg/mL CTB-neoAg protein for 72 h. Unstimulated cells were used as negative control. The lymphocytes were stained with anti-mouse CD3ε antibody-FITC and anti-mouse CD8 antibody-PerCP-Cy5.5, and analyzed using a FACS Calibur instrument (Beckman Coulter, USA).

For $CD19^+$ B cell analysis, the freshly isolated spleen lymphocytes were stained with anti-mouse CD19 antibody-FITC. To measure $Fop3^+/CD4^+$ ratio, tumor-infiltrating lymphocytes or spleen lymphocytes were isolated from DTT-neoAg-treated mice or PBS-treated mice and stained with anti-mouse CD3ε antibody-APC (BD Biosciences) and CD4 antibody-FITC (BD Biosciences). Tumor infiltrating leucocytes were prepared from subcutaneous B16F10 tumors when tumor volumes reached 1500 mm^3. Tumor tissues were ground and filtered through a 70-μm cell strainer. The TILs are purified by a tumor infiltrating lymphocytes separation solution according to the manufacturer's protocol (Beijing Solarbio Science and Technology Co., Beijing, China). Then intracellular $FoxP3^+$ staining was performed according to the manufacturer's protocol (Mouse Foxp3 Buffer Set, BD). The samples were analyzed using a FACS Calibur instrument (Beckman Coulter, USA). All experiments were repeated three times and the average values were calculated.

Intracellular Cytokine Staining

The mutation-specific IFN-γ^+ T cells were detected by intracellular cytokine staining (ICS), and analyzed by flow cytometry. Bone marrow-derived dendritic cells (BMDCs) were obtained by culturing bone marrow cells of naïve C57BL/6 mice in RPMI-1640 medium containing 10% FBS, GM-CSF (20 ng/mL) and IL-4 (20 ng/mL) (Sino Biological, Beijing, China) for 6 days as previously described (26). BMDCs were loaded with or without 50 μg/mL DTT-neoAg overnight.

Splenocytes harvested from DTT-neoAg-immunized mice were incubated for 5 min at room temperature in ACK Lysing Buffer (0.15 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM disodium EDTA, pH 7.2) and then washed in RPMI-1640 medium (Gibco-BRL) with 10% fetal bovine serum (FBS). The splenocytes (3×10^6) were co-incubated with DTT-neoAg-loaded BMDCs (3×10^5) for 48 h while splenocytes co-incubated with BMDCs

served as a negative control. Cells were further incubated for 6 h at 37°C in the presence of Brefeldin A (3.0 μg/mL, Sino Biological, Beijing, China). The splenocytes were treated with phorbol 12-myristate 13-acetate (PMA, 81 nM) and Ionomycin (1.34 μM, Thermo Fisher Scientific) served as a positive control. The cells were then harvested and stained with anti-CD3ε antibody-PerCP-Cy 5.5 (BD Biosciences), anti-CD4 antibody-FITC (BD Biosciences), and anti-CD8 antibody-PE (BD Biosciences) followed by anti-IFN-γ antibody-APC (BD Biosciences) staining after Cytofix/Cytoperm treatment according to the manufacturer's protocol (BD Biosciences). Subsequently, the samples were analyzed on a CytoFLEX flow cytometer (Beckman Coulter), and the data were analyzed using FlowJo software.

Cytotoxicity Assay

Cytotoxic T lymphocyte assay was performed as previously described (27). Briefly, spleen lymphocytes or TILs stimulated with 30 μg/mL CTB-neoAg protein for 72 h in the presence of 20 units/mL rhIL-2 (ExCell Bio, Shanghai, China) were used as effector cells. The stimulated lymphocytes were co-cultured with B16F10 (target cells) at different effector to target ratios (50:1, 20:1, and 10:1) for 4 h at 37°C. Fifty microliter of the culture supernatant was used to assess cytotoxic activity by CytoTox96 non-radioactive cytotoxicity assay kit (Promega, Madison, WI, USA) according to the manufacturer's instructions.

IFN-γ Assay

Splenocytes were isolated from immunized mice, stimulated with 30 μg/mL CTB-neoAg in 24-well culture plates (5×10^5/well, in 500 μL media) at 37°C with 5% CO_2. The culture media were collected at indicated time points. The IFN-γ concentrations were measured by ELISA kit (R&D Systems, USA).

Quantitative PCR Assay

Total tumor RNA were prepared using TRIzol reagent (Invitrogen). cDNA were reverse-transcribed using PrimeScript™ RT reagent Kit with gDNA Eraser (Takara, Dalian, China). Real-time PCR was performed using the TB Green™ Premix Ex Taq™ II according to the manufacture's protocol (Takara, Dalian, China). The mRNA levels of IFN-γ, IL-12, IL-4, and IL-10 in each sample were calculated using the $2^{-\Delta\Delta ct}$ method. The primer sequences are listed in **Supplementary Table 2**.

Immunohistochemistry Analysis

Immunohistochemistry was performed as described previously (8). Tumor tissues, livers, kidneys, or lungs were fixed in 4% paraformaldehyde and embedded in paraffin. The paraffin-embedded tissues were cut into 4 μm-slides, stained with hematoxylin and eosin (H&E) or rat anti-mouse CD8 antibody (Santa Cruz Biotechnology Inc., Dallas, TX, USA) or anti-mouse NK1.1 antibody (Bio Legend, USA) according to the manufacture's protocol. Immunohistochemical detection was performed following the instruction of MaxVision™ kit (Fuzhou Maixin Biological Inc., Fuzhou, China). Five different fields for each slice were selected under 200 × magnification to quantify the percentages of $CD8^+$ cells by Image J.

Statistical Analyses

All data were analyzed using GraphPad Prism 7.0 software and were presented as mean ± standard deviation (SD). The statistical significance between two groups was analyzed by a two-tailed, unpaired Student's *T*-test. The survival curves were generated using the Kaplan-Meier method and the statistical significance between two groups were analyzed using the log-rank test. *indicates $P < 0.05$, **indicates $P < 0.01$, ***indicates $P < 0.001$, ****indicates $P < 0.0001$, and ns indicates no significant difference.

RESULT

The Design of Vaccine Immunogen DTT-neoAg

Activation of effector $CD4^+$ T cells is a hallmark of sustained and protective immunity induced by vaccination (28). For neoantigen-based cancer vaccine, antigen-specific $CD4^+$ T cell effectors can lead to tumor regression through direct cytotoxic mechanisms (29) or activation of macrophages (30). Due to immune tolerance, self-molecule proteins are not recognized by T cell receptors under normal physiological conditions (31). This self-tolerance can be bleached by insertion or fusion of a foreign helper T cell epitope into the self-molecules (32). In our previous studies, we found insertion of a neutralizing epitope peptide of TNF α or EGFR into a membrane translocation domain of Diphtheria toxin, named DTT, can induce TNFα- or EGFR-specific antibody responses by virtual of its several universal Th epitopes (25, 33). When formulated with Th1-inducing adjuvants such as poly I/C or CpG, EGFR-specific cytotoxic T lymphocytes were activated which conferred antitumor immunity in mouse tumor models (33). Moreover, fusion of self-protein domains to DTT also induced antibody responses toward self-proteins such as EGFR, TNF α, VEGF, FXI (24, 25, 33, 34). In this study, we applied DTT fusion strategy to neoantigen-based vaccine design, and tested the immunogenicity of those tumor neoantigens predicted *in silico* to possess good affinity to MHC I molecules, but experimentally failed to elicit any tumor-specific cytotoxic T cell responses.

We chose four MHC I-binding, non-immunogenic mutation peptides identified in a previous report by Sahin (7) as a part of vaccine immunogen. These neoantigen peptides are encoded by genes of Pi4k2b, Ddb1, Pcdhga11, Atp11a, and the corresponding mutations are verified in our laboratory B16F10 cell line. Indeed, these 27aa neoantigen peptides failed to elicit anti-tumor immune response when formulated with Alum and CpG (**Figures S1B,C**).The mutation peptides are linked together by GS linkers (7) and fused to the C-terminal of DTT as shown in **Figure 1A**. The recombinant immunogen is named DTT-neoAg.

An *E. coli* expression vector was constructed for DTT-NeoAg. For control experiments, an expression vector of corresponding wild-type peptide antigens, named as DTT-wtAg, was also constructed. The sequences of neoantigens and their wild-types were provided in the **Supplementary Table 3**. For detection

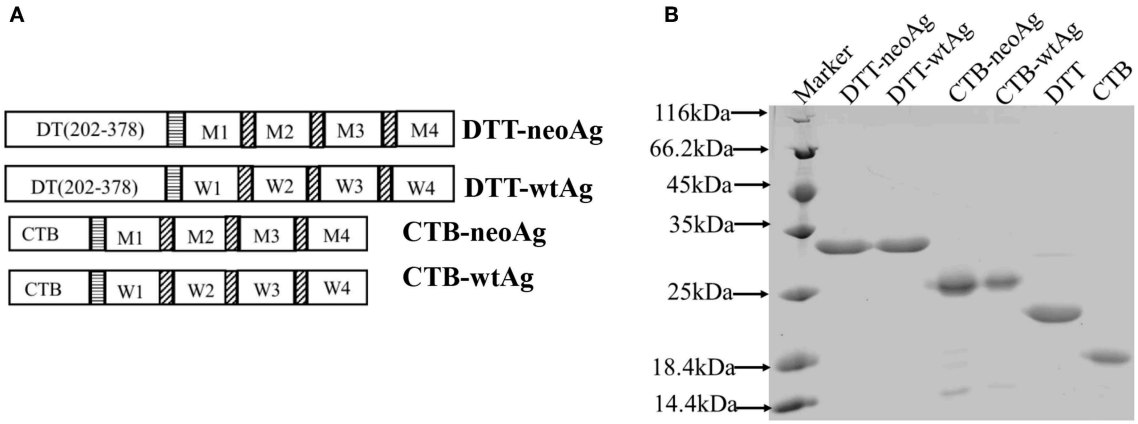


FIGURE 1 | Design and expression of DTT adjuvant-based neoantigen vaccine. **(A)** A schematic diagram of DTT-neoAg and DTT-wtAg. M1, M2, M3, and M4 are the mutation peptide of Pi4k2b, Ddb1, Pcdhga11 and Atp11a, respectively. W1, W2, W3, and W4 are the corresponding wild type peptides. Each mutation peptide is 27 amino acid residues in length with the mutated residue positioned in the central (position 14). The horizontal stripe box stands for the linker sequence (GGGGSGGGGS) and the diagonal stripe box stands for the linker sequence (GS). **(B)** 15% SDS-PAGE analysis of the purified recombinant proteins.

of neoAg-specific antibody responses by ELISA and *in vitro* expansion of neoAg-specific T cells, an expression vector was constructed for neoAg fused to C-terminal of Cholera Toxin B chain, named as CTB-neoAg.

The DTT-neoAg and DTT-wtAg were expressed in *E. coli* as soluble GST-tagged recombinant proteins. After GST-affinty purification, the GST tag was removed by PSP cleavage and GST affinity column chromatography (**Figure 1B**).

DTT-neoAg Induces Neoantigen-Specific Antibody and Antitumor Cellular Immune Responses in Mouse

We immunized C57BL/6 mice with DTT-neoAg formulated with Alum adjuvant and CpG (**Figure 2A**). The control mice were administered with PBS + Alum + CpG. One week after the third injection, ELISA was performed to measure the antibody responses against neoAg. We observed robust antibody responses in DTT-neoAg treated mice (**Figure 2B**), but not in PBS treated mice, with the antibody titers at the levels of 6×10^5 (**Figure 2C**).

Because antibody responses depend on B cells, we tested whether the vaccine can increase the number of B lymphocytes in spleen. Seven days after the third immunization, the splenocytes of DTT-neoAg or PBS-treated mice were labeled with CD19 antibody, and analyzed by flow cytometry. We found that the percentage of $CD19^+$ lymphocytes in DTT-neoAg-treated mice was 47.47 ± 1.04%, while in the PBS-treated mice was 38.87 ± 1.82% (**Figure 2D**). Therefore, the B cells of DTT-neoAg-immunized mice were increased significantly in comparison with those of the PBS group mice.

We further analyzed the antibody subclasses in DTT-neoAg-treated mouse sera (**Figure 2E**), and found that the subclasses of IgGs were primarily IgG1 and IgG2b. The other subclasses including IgM, IgG2a, and IgG3 were at below detection levels. As the ratio of IgG1/IgG2b was >1, we conclude that immune responses induced by DTT-NeoAg is biased toward Th2 type.

To examine the cellular immune responses of DTT-neoAg vaccination, we stimulated the splenocytes of each group of mice with CTB-neoAg *in vitro* for 72 h, and measured the cell proliferation by CCK8 kit. Significant cell proliferation was observed in the splenocytes of DTT-neoAg-treated mice (**Figure 2F**) with the average stimulation index 3 times that of PBS-treated ones (**Figure 2G**). The cell proliferation was similar when both groups of the splenocytes were stimulated with Con A (**Figure 2G**). These data indicated that DTT-neoAg vaccine can induce mutation-specific T cell memory responses.

We next evaluated the cytotoxicity of DTT-neoAg-treated splenocytes by LDH release assay after *in vitro* stimulation with CTB-neoAg. At effector-target ratios of 50:1 and 20:1, the percentages of cell lysis were 49.65 ± 4.19 and 17.29 ± 9.33%, respectively (**Figure 2H**). For the splenocytes of PBS-treated mice, the percentages of cell lysis were 11.62 ± 6.06 and 6.64 ± 4.33%, respectively. These data indicated that DTT-neoAg vaccination elicit robust cytotoxic immune responses against B16F10.

DTT-neoAg Vaccination Significantly Inhibits Tumor Growth in a Prophylactic Mouse Tumor Model

Further to assess whether the vaccine-elicited immune responses render tumor growth inhibition, we challenged the vaccinated mice subcutaneously with 1×10^5 B16F10 in the right flank 9 days after the third immunization (**Figure 3A**). While tumors were palpable in all of mice treated with PBS + adjuvant on day 10 after tumor challenge, the DTT-neoAg-treated mice were all tumor-free. Tumors were detected in 4 of 6 mice treated with DTT-neoAg on day 16 after the tumor challenge. More strikingly, one DTT-neoAg-treated mouse was tumor-free until 30 days after tumor inoculation. Furthermore, the tumors in DTT-neoAg-treated mice grew much slower than those of PBS-treated ones. On day 18, the average tumor volumes of PBS-treated mice

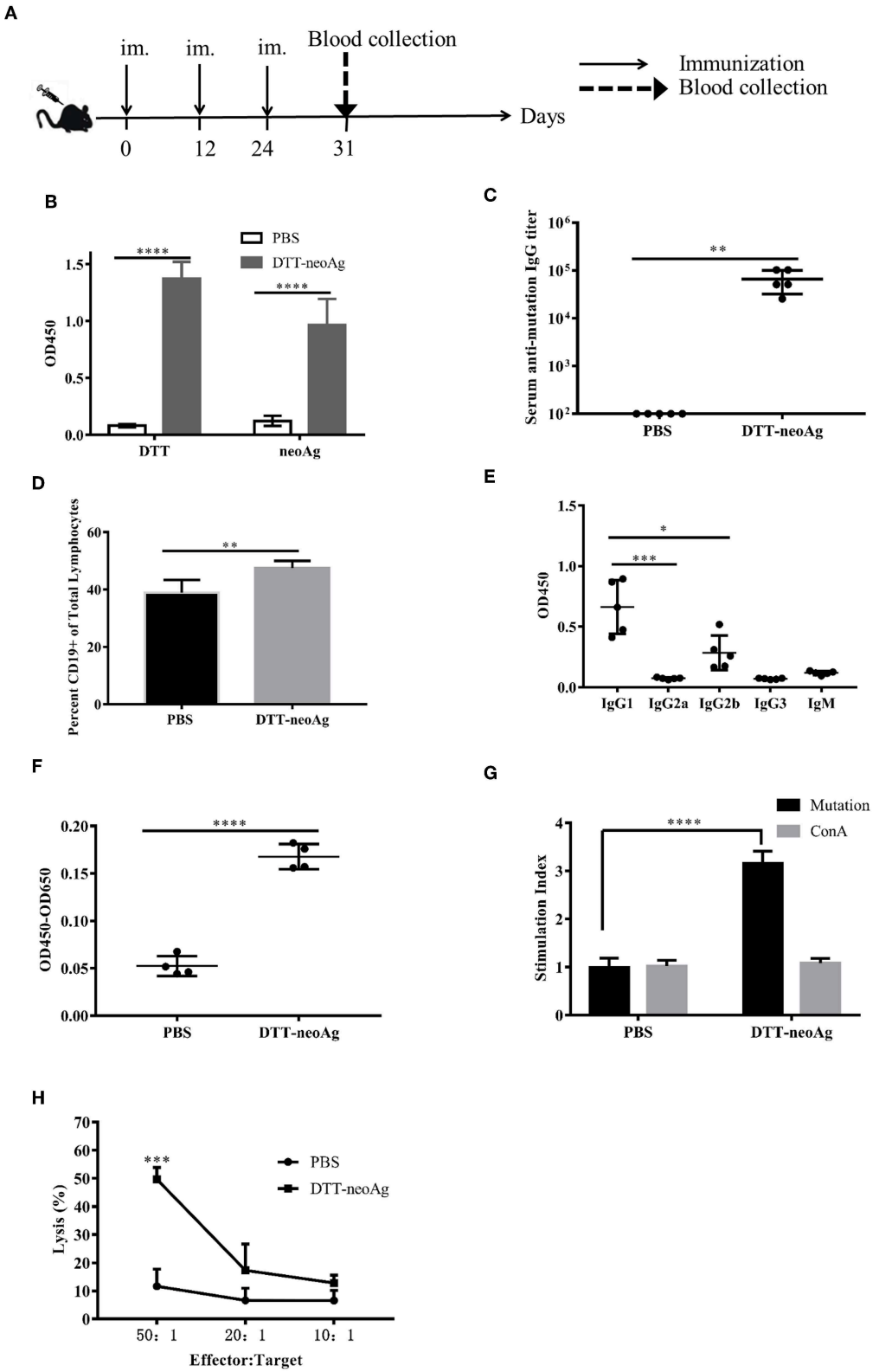


FIGURE 2 | DTT-neoAg induces neoantigen-specific antibodies and antitumor cellular immune responses in mouse. The statistical significances were determined by Student's T test. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$. **(A)** A schematic diagram of the experimental design. **(B–D)** Female C57BL/6 mice ($n = 5$) were immunized subcutaneously on day 0, 12, 24 with DTT-neoAg vaccine. One week after the third immunization, mouse sera were collected by orbital blood sampling,

(Continued)

FIGURE 2 | and the antibody responses were analyzed by ELISA. **(B)** ELISA with sera in 1:100 dilution and with indicated coating antigens. **(C)** The antibody titers against the neoantigen. **(D)** The percentage of $CD19^+$ B cells in the lymphocytes of immunized mice. Seven days after the third immunization, the splenocytes of the vaccinated mice were stained with anti-CD19-FITC, and the percentage of $CD19^+$ B cells in the lymphocytes was analyzed by flow cytometry. **(E)** The antibody subclass analyses. The sera were 1:100 diluted. **(F–H)** DTT-neoAg vaccine induced cellular immune response. Six to eight weeks old female C57BL/6 mice ($n = 3$) were immunized with DTT-neoAg or PBS formulated with Alum + CpG, respectively, at 2-week intervals. Seven days after the third immunization, the splenocytes were stimulated with CTB-neoAg or PBS for 72 h. Cell proliferation **(F)** and stimulation indices **(G)** were measured by CCK8 kit. **(H)** The splenocytes from DTT-neoAg-treated mice or PBS-treated mice were stimulated with CTB-neoAg (30 μg/mL) for 72 h, and were used as effector cells, then co-cultured with target cells B16F10 at indicated ratios for 4 h at 37°C. The percentage of cell lysis was determined by LDH assay.

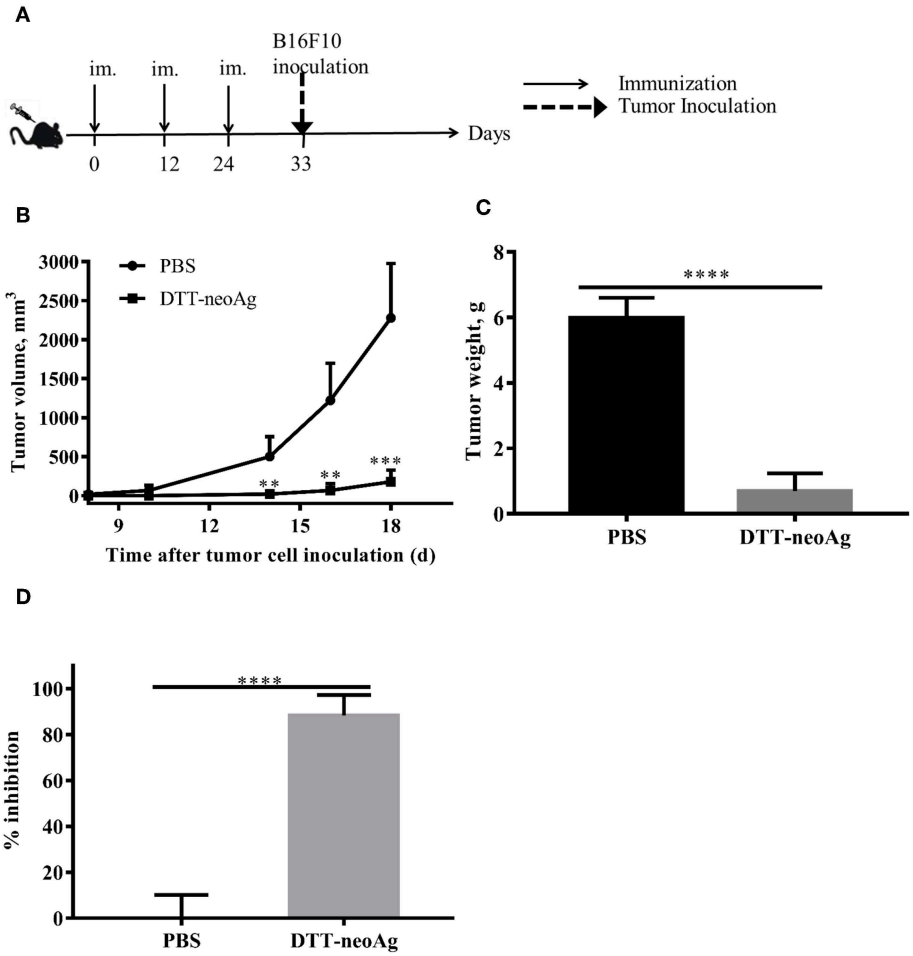


FIGURE 3 | Anti-tumor effects of DTT-neoAg vaccination in the prophylactic mouse melanoma model. The data are means ± SEM, ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$, Student's T-test. **(A)** The treatment schedule. C57BL/6 mice ($n = 5$) were immunized with DTT-neoAg or PBS formulated with Alum + CpG on day 0, day 12 and 24. Nine days after the third immunization, 1×10^5 B16 F10 cells were s.c. administered into the right flank of the mice. **(B)** The tumor growth curves. **(C)** On day 22 after the tumor challenge, tumors weights were measured. **(D)** Tumor inhibition rate of PBS-treated or DTT-neoAg-treated mice.

was 2278 ± 349.2 mm^3 while that of DTT-neoAg was 177.7 ± 66.53 mm^3, a growth reduction by more than 80% (**Figure 3B**). On day 22, the average tumor weight was 0.7 ± 0.27 g for DTT-neoAg treated-mice in contrast with that of PBS-treated ones which was 6 ± 0.27 g (**Figure 3C**). Therefore, DTT-neoAg vaccine reduced the tumor growth by 88% (**Figure 3C**).

In order to prove that DTT is necessary to elicit the immune response against neoAg, and only an immune response specific to neoAg can inhibit tumor, we s.c. injected 1×10^5 B16F10 in the right flank of mice, 7 days after the third immunization with neoAg-pep + adjuvant or DTT-wtAg + adjuvant, or PBS + adjuvant (**Figure S1A**). Seven days after the challenge, tumors were detected in all of the PBS-treated mice, neoAg-pep-treated mice and DTT-wtAg-treated mice. Meanwhile, the tumor volumes of PBS-treated mice, neoAg-pep-treated mice and DTT-wtAg-treated mice were 12.2 ± 2.74, 11.31± 3.24, and 16.81 ± 2.97 mm^3, respectively. On day 17, the average tumor volumes of PBS group, neoAg-pep group, and DTT-wtAg were 1801 ± 378.8, 1514 ± 251.2, and 1439 ± 133.6 mm^3, respectively. Overall, there was no significant difference in the growth curves of each group

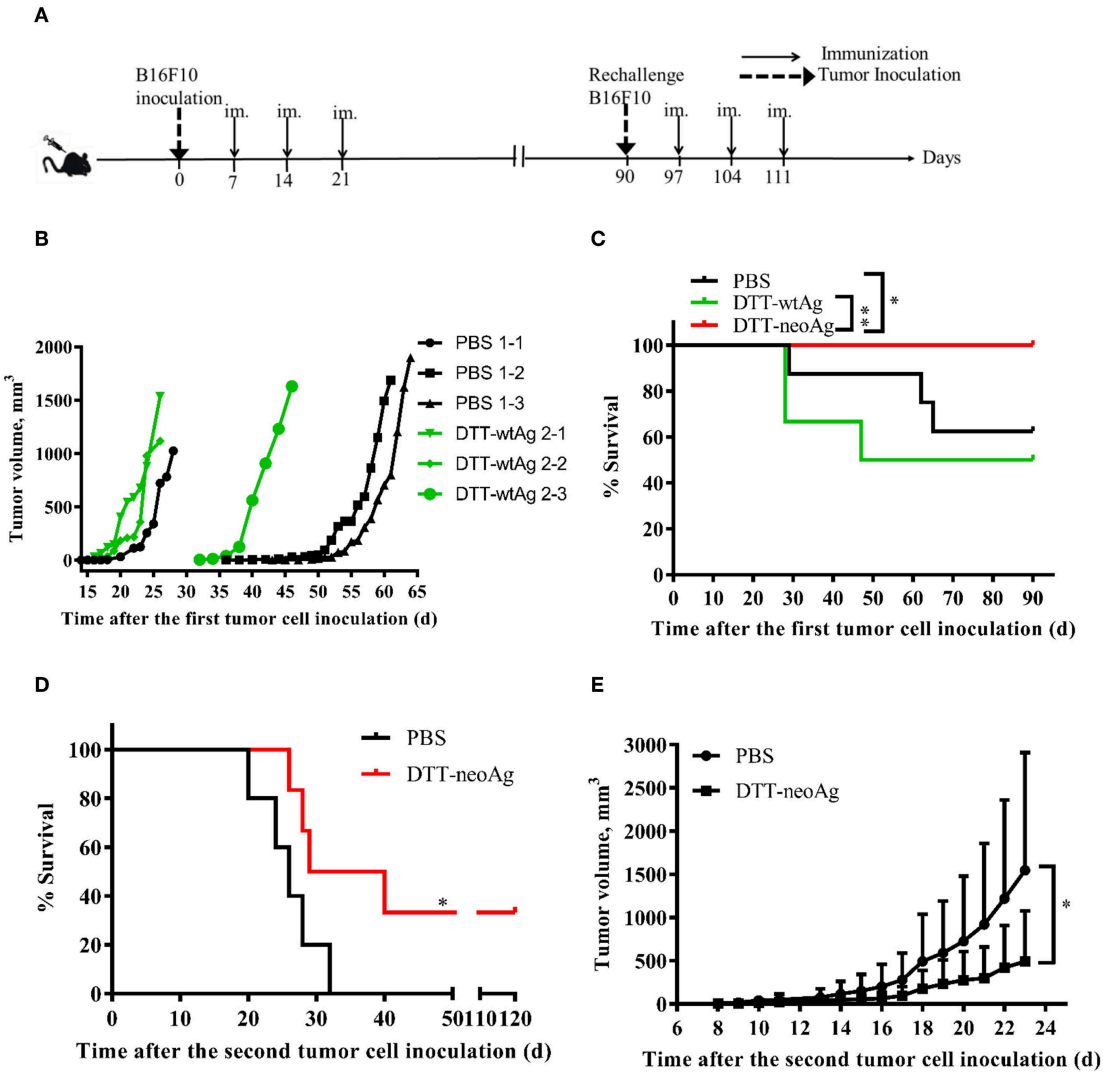


FIGURE 4 | Anti-tumor effects of DTT-neoAg vaccination in the therapeutic mouse melanoma model. The data are means ± SEM, ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$, Student's T-test. **(A)** The treatment schedule. **(B,C)** C57BL/6 mice ($n = 6$–13) were s.c. administered with 2.5×10^4 B16F10 cells into the right flank of the mice. Seven days after challenge, mice were immunized with DTT-neoAg, DTT-wtAg, or PBS formulated with Alum + CpG three times at 1-week intervals. **(B)** The tumor growth curves for individual mouse. Green line: DTT-wtAg-treated mice, and black line: PBS-treated mice. **(C)** The Kaplan-Meier survival plot. The statistical significance was determined by Log-rank test. *$P < 0.05$. **(D,E)** On day 90, the tumor-free mice in PBS-treated group and neoAg-treated group were re-challenged with 7.5×10^4 B16F10 cells in the left flank. One week after the re-challenge, mice were immunized with PBS or neoAg formulated with Alum + CpG for 3 times at weekly intervals. **(D)** The Kaplan-Meier survival plot. The statistical significance was determined by Log-rank test. *$P < 0.05$. **(E)** The tumor growth curves.

(**Figure S1B**). The median survival rates of the three groups were all 19 days (**Figure S1C**). Therefore, both neoAg-pep and DTT-wtAg failed to inhibit tumor growth in the prophylactic mouse tumor model.

DTT-neoAg Vaccination Inhibits Tumor Growth in the Therapeutic Mouse Tumor Model

B16F10 melanoma cell line is highly aggressive when implanted into a syngeneic host (35). Therefore, to examine therapeutic efficacy of DTT-neoAg vaccine, we s.c. injected a low dose of 2.5×10^4 B16F10 cells in the right flanks of mice to allow sufficient time for the vaccine to induce anti-tumor immunity. Seven days after the tumor challenge, mice were administered with the indicated vaccines and boosted twice at 1 week intervals (**Figure 4A**). Tumors were detected in some of the PBS-treated mice as early as 18 days after tumor challenge (**Figure 4B**). On day 43, 37.8% of PBS-treated mice developed tumors. Tumors were detected in 16.7% of DTT-wtAg-treated mice on 15 days, 3 days earlier than those of PBS group of mice (**Figure 4B**), and on day 32, 50% of the mice had

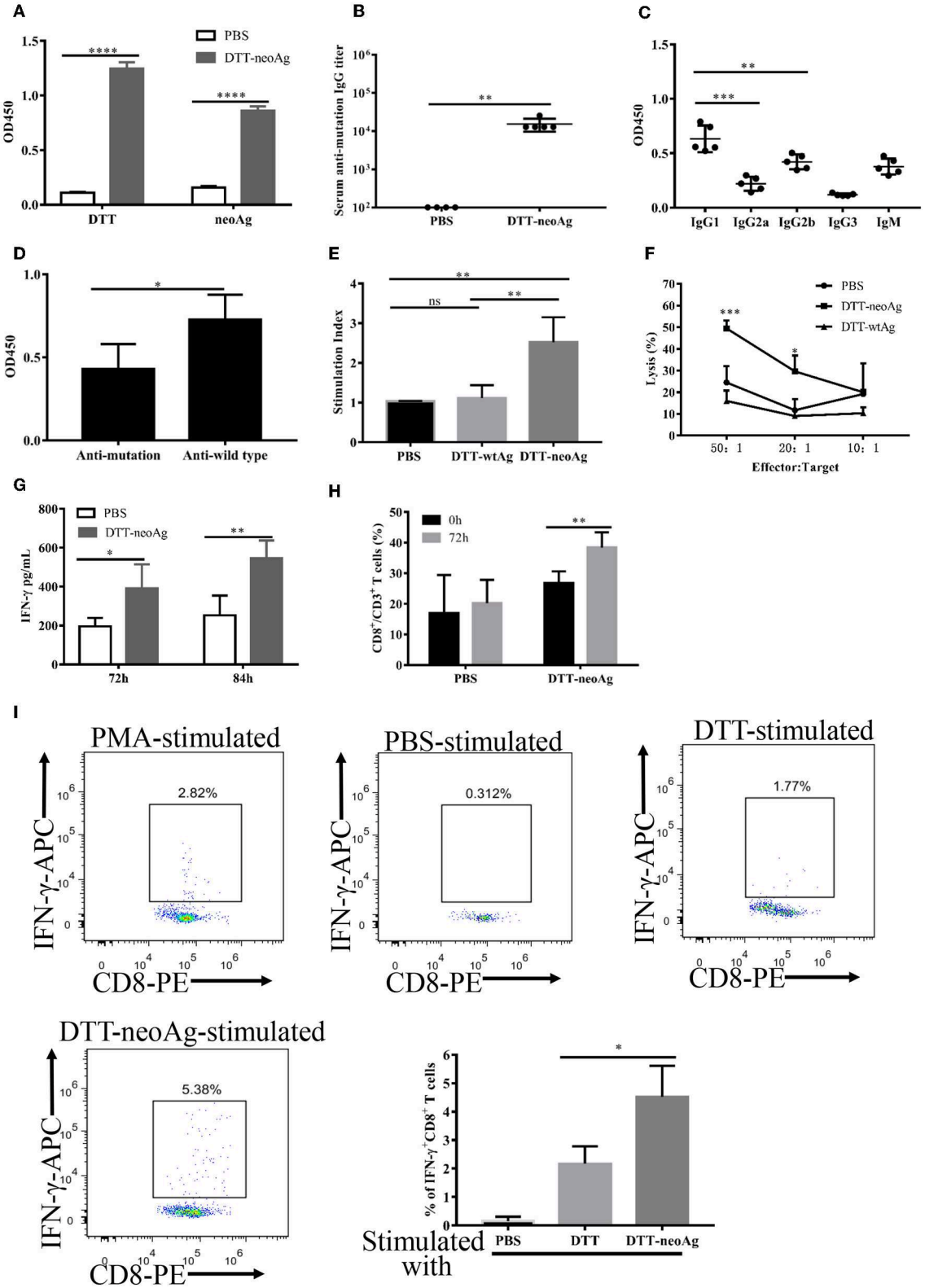


FIGURE 5 | Antibody responses and cellular immune responses elicited by DTT-neoAg vaccination in tumor challenged mice. The statistical significance was determined by Student's T-test. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. **(A–D)** C57BL/6 mice ($n = 6$–13) were administered s.c. with 2.5×10^4 B16 F10 cells into the right flank of the mice. Seven days after tumor challenge, mice were immunized three times at weekly intervals with DTT-neoAg, DTT-wtAg, or PBS, formulated with

(Continued)

FIGURE 5 | Alum + CpG. **(A)** The mouse sera were collected on day 7 after the third immunization. The antibody responses were measured by ELISA with the sera 1:100 diluted and with the indicated coating antigens. **(B)** The anti-neoAg antibody titers were measured by ELISA. **(C)** Analysis of anti-neoAg antibody subclasses. **(D)** Anti-sera from DTT-wtAg-treated mice crossreact with neoAg. The sera were collected from mice immunized with DTT-wtAg after the third immunization, and the antibody specificity was analyzed by ELISA with sera 1:100 diluted and CTB-wtAg or CTB-neoAg as a coating antigen. **(E)** The splenocytes from indicated mice were stimulated *in vitro* with CTB-neoAg for 3 days. The stimulation indices was measured by CCK8 assay. **(F)** The cytotoxic T lymphocytes induced by DTT-neoAg vaccination. The splenocytes from DTT-neoAg-treated, or DTT-wtAg-treated, or PBS-treated mice were stimulated with CTB-neoAg for 3 days, and used as effector cells. B16F10 cells were used as target cells. The splenocytes and B16F10 cells were co-cultured at indicated ratios for 4 h at 37°C. The percentage of cell lysis was measured by LDH assay. Asterisks indicate statistically significant differences between DTT-neoAg-treated and DTT-wtAg-treated mice. **(G)** INF-γ production by *in vitro* neoAg-stimulated splenocytes of DTT-neoAg vaccinated mice. The splenocytes were stimulated with CTB-neoAg for 72 or 84 h, and the INF-γ in the culture supernatant was analyzed using a mouse IFN-γ ELISA kit. **(H)** $CD8^+$ memory T cells induced by DTT-neoAg vaccination. The splenocytes of the vaccinated mice were *in vitro* stimulated with CTB-neoAg for 72 h, stained with anti-CD3ε-FITC anti-CD8-PerCP-Cy5.5, and the percentage of $CD8^+$ T cells in $CD3^+$ T cells was analyzed by flow cytometry. **(I)** The splenocytes of DTT-neoAg-treated mice were incubated with antigen-loaded DC for 48 h, then Golgi-stop was added; 6 h later, the cells were harvested and stained with anti-CD3ε-PerCP-Cy5.5, anti-CD4- FITC, anti-CD8-PE Ab. After permeabilization, intracellular cytokines were stained with anti-IFN-γ-APC antibody and analyzed by flow cytometry.

tumors. In contrast, all DTT-neoAg-treated mice were tumor-free on day 90, and the overall survival of the DTT-neoAg-immunized group was significantly prolonged when compared to two control groups. The survival rates for PBS and DTT-wtAg groups were similar (**Figure 4C**). Moreover, mice of the control groups appeared weak, unresponsive, less active, and arched since day 14 while DTT-neoAg-treated mice behaved normal. These data suggest that DTT-neoAg induce therapeutic and mutation-specific antitumor immunity.

We next challenged the tumor-free mice with a higher dose of B16F10 cells to evaluate therapeutic efficacy of DTT-neoAg vaccine. The tumor-free, PBS-treated or DTT-neoAg-treated mice were challenged with 7.5×10^4 B16F10 cells on the left flanks on day 90. These mice were vaccinated three more times 7 days after the second tumor challenge as indicated in **Figure 4A**. Strikingly, 33.33% of the mice treated with DTT-neoAg were tumor-free even on day 120, whereas all the mice in the PBS-treated group developed tumors, and died within 32 days (**Figure 4D**). The average survival days of the tumor-bearing, DTT-neoAg-treated mice was 34.5, which is 8.5 days longer than that of PBS-treated group (**Figure 4D**). Consistently, the tumor growth was markedly reduced in DTT-neoAg-immunized mice when compared with that of PBS-immunized mice (**Figure 4E**). Taken together, these data indicated that DTT-neoAg vaccine confer therapeutic benefit to tumor bearing mice.

DTT-neoAg Vaccination Induces Humoral and Cellular Immune Responses in Tumor Challenged Mice

Since DTT-neoAg vaccination provided efficient tumor control in the therapeutic tumor model, we assessed the antibody responses and cellular immune responses in tumor bearing mice elicited by the vaccine. We measured antibodies against DTT and neoAg by ELISA, and found that both anti-DTT and anti-neoAg antibodies were robust with antibody titers up to 1.5×10^5 (**Figures 5A,B**). As the ratios of IgG1 to IgG2a were >1, the induced immunity was biased toward Th2 type (**Figure 5C**). Compared to the sera from mice treated with DTT-neoAg before the tumor challenge, more IgG2a (12% vs. 6%) and less IgG1 (37% vs. 54%) were generated when DTT-neoAg vaccine was applied after the tumor challenge, suggesting that immune response shifted toward Th1 type of immunity in therapeutic setting (**Figure S2A**). However, the class switch was not complete in mice of the therapeutic tumor model as the IgM subclass was present in the mouse sera.

In mice treated with DTT-wtAg vaccine, robust antibody responses against wtAg were observed with weak cross activities to neoAg (**Figure 5D**). This result is consistent with a T7 phage-based vaccine targeting multiple neoepitopes including Atp11a mutations (36). Since DTT-wtAg vaccination failed to inhibit tumor growth, the antitumor immunity induced by DTT-neoAg was mutation-specific.

We then evaluated the cellular immune responses induced by DTT-neoAg vaccine. Thus, the splenocytes were stimulated with CTB-neoAg for 72 h, the cell proliferations as well as the cytotoxic activities were measured. As shown in **Figure 5E**, significant cell proliferations were observed only in splenocytes of DTT-neoAg-treated mice. The stimulation index was almost 3 times that of DTT-wtAg or PBS treated mice, which indicated that DTT-neoAg induced mutation-specific cellular responses. Again, significant cytotoxic activities were observed only in DTT-neoAg treated mice (**Figure 5F**). The percentage of neoAg-specific cell lysis was 49.46 ± 2.07% at effector/target ratio of 50:1, and 29.58 ± 4.27% at lower ratio of 20:1 (**Figure 5F**).

Further to examine whether the cytotoxic activities were conferred by $CD8^+$ T cells, we first measured the amount of INF-γ in the culture media of the splenocytes stimulated with CTB-neoAg, and found that DTT-neoAg-vaccinated mice splenocytes secret twice amount of INF-γ as compared to that of PBS treated mice splenocytes (**Figure 5G**). In addition, the percentage of $CD8^+$ $CD3^+$ lymphocytes in DTT-neoAg-treated mice was increased from 26.8 ± 3.81% to 38.42 ± 4.94% by CTB-neoAg stimulation (**Figure 5H**). Next, we labeled the splenocytes of DTT-neoAg-treated mice stimulated *in vitro* by DTT-neoAg with CD8 antibodies, and found that ∼4.5% of the $CD8^+$ T cell population expressed IFN-γ, which was 2-fold higher than that of the same splenocytes stimulated with control DTT (**Figure 5I**), indicating that neoantigen-specific CTLs were induced by the vaccine. In contrast with RNA and long-peptide neoantigen vaccines in preclinical and early phase clinical studies showing that immune responses were predominantly of $CD4^+$ T cells (7, 8), DTT fusion strategy significantly enhanced immunogenicity of neoantigens as well as the CTL responses.

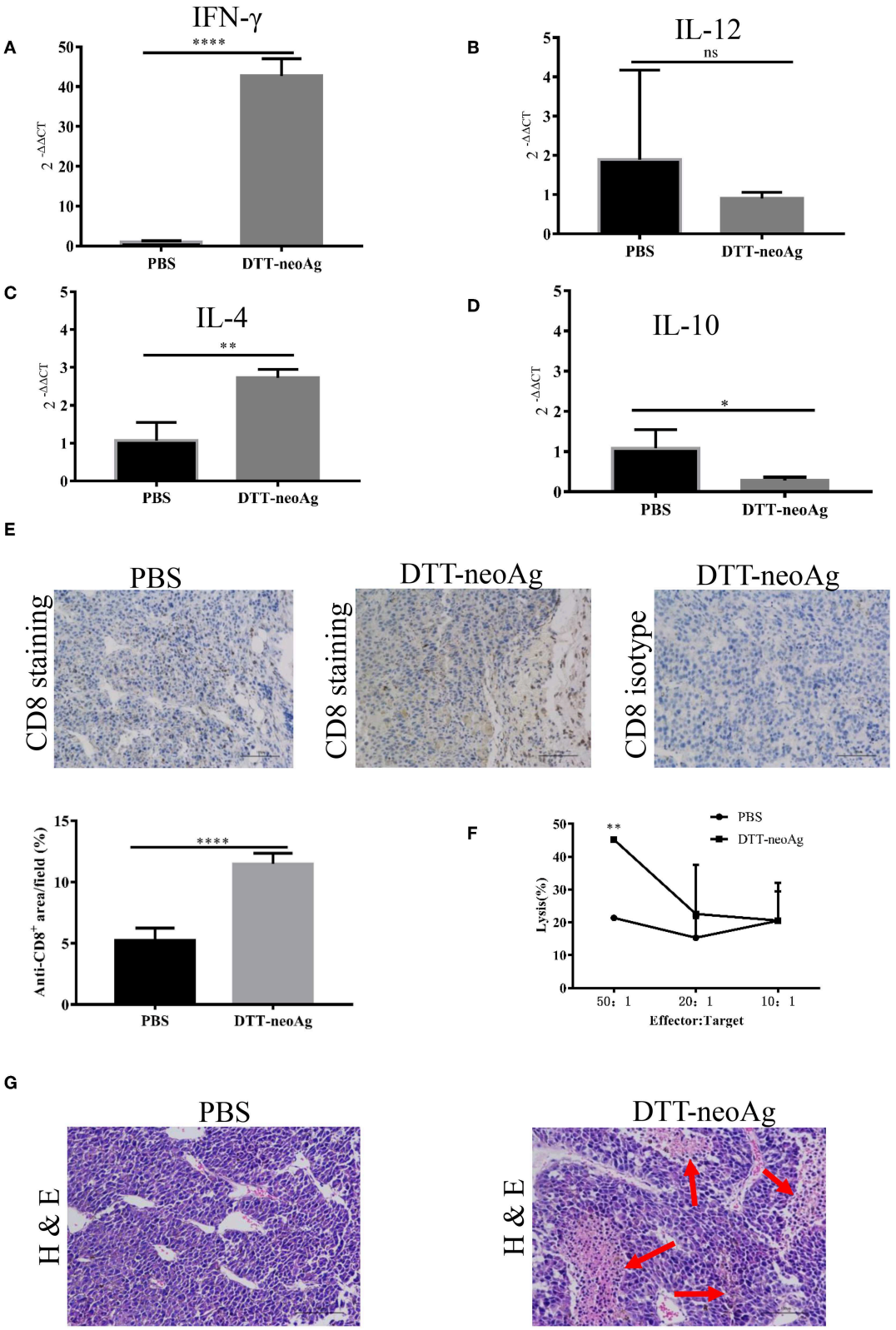


FIGURE 6 | DTT-neoAg vaccination increases tumor necrosis and enhances CD8$^+$ T cells infiltration. The statistical significances were determined by Student's T-test. ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. Frozen sections of B16F10 tumor tissues isolated from mice immunized with DTT-neoAg or PBS as in **Figure 3D**. **(A–D)** The relative expression levels of IFN-γ, IL-12, IL-4, IL-10 in the tumor tissues were determined by RT-PCR, respectively. **(E,G)** B16F10 tumors from

(Continued)

FIGURE 6 | indicated mice ($n = 3$) were collected and sectioned when the volumes reached 2,000 mm^3. **(E)** The tissues sections were stained with anti-CD8 antibody. The density of $CD8^+$ T cells were quantified by Image J. Numbers in the panel indicate average values of three samples per group, quantified by Image J. **(F)** The TILs from DTT-neoAg-treated mice or PBS-treated mice were stimulated with CTB-neoAg (30 μg/mL) for 72 h, and were used as effector cells, then co-cultured with target cells B16F10 at indicated ratios for 4 h at 37°C. The percentage of cell lysis was determined by LDH assay. **(G)** H&E staining of B16F10 tumor sections from indicated mice. Red arrows indicate the necrosis area, scale bar length is 100 μm.

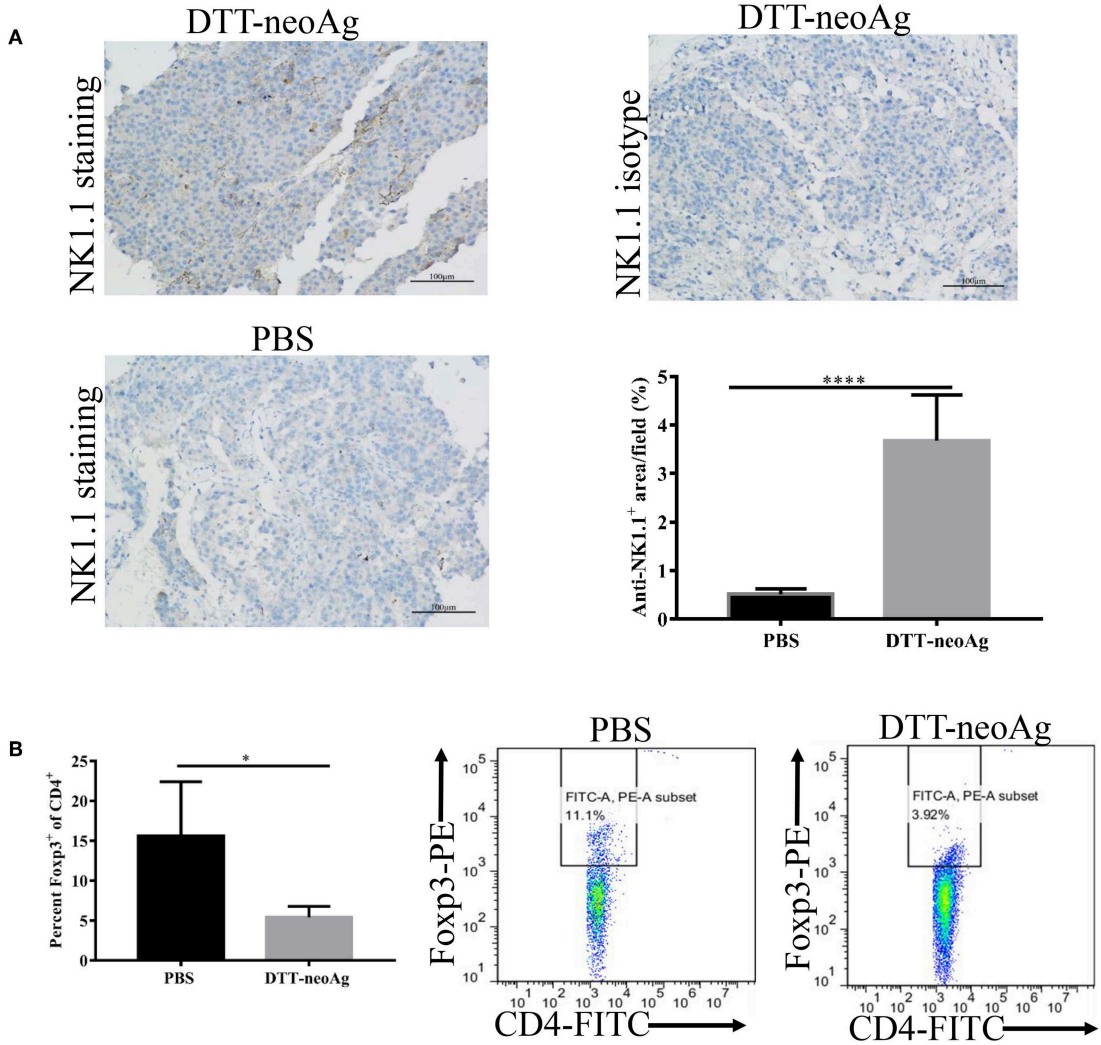


FIGURE 7 | DTT-neoAg vaccination enhances $NK1.1^+$ cell infiltration and reduces $Foxp3^+/CD4^+$ ratio in TILs and spleens. The statistical significances were determined by Student's T-test. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$. **(A)** B16F10 tumors from indicated mice ($n = 3$) were collected and sectioned when the tumor volumes reached 1,500 mm^3. The tissues sections were stained with anti-NK1.1 antibody. The density of $NK1.1^+$ cells were quantified by Image J. Numbers in the panel indicate average values of three samples per group, quantified by Image J. **(B)** The splenocytes from DTT-neoAg or PBS vaccinated mice were isolated and stained with anti-CD3ε-APC, anti-CD4- FITC. After permeabilization, cells were stained with anti-Foxp3-PE antibody and analyzed by flow cytometry.

DTT-neoAg Vaccination Enhances Intra-tumor Th1 Immunity and Causes Tumor Necrosis

To characterize the immune responses in the tumor tissues, we measured the cytokine levels by RT-PCR and $CD8^+$ T cell infiltration by immunohistochemistry analyses. As shown in **Figures 6A–D**, tumor tissues from DTT-neoAg-treated mice expressed markedly higher levels of IFN-γ and IL-4 than PBS-treated mice while the levels of IL-12 and IL-10 were similar. The levels of INF-γ in DTT-neoAg-treated mice were increased by 40-fold in comparison with those of PBS-treated mice (**Figure 6A**), and the levels of IL-4 were increased by 2.7-fold (**Figure 6C**). Both IL-12 (**Figure 6B**) and IL-10 (**Figure 6D**) were detected, although the levels were low and show no significant difference between the experimental groups.

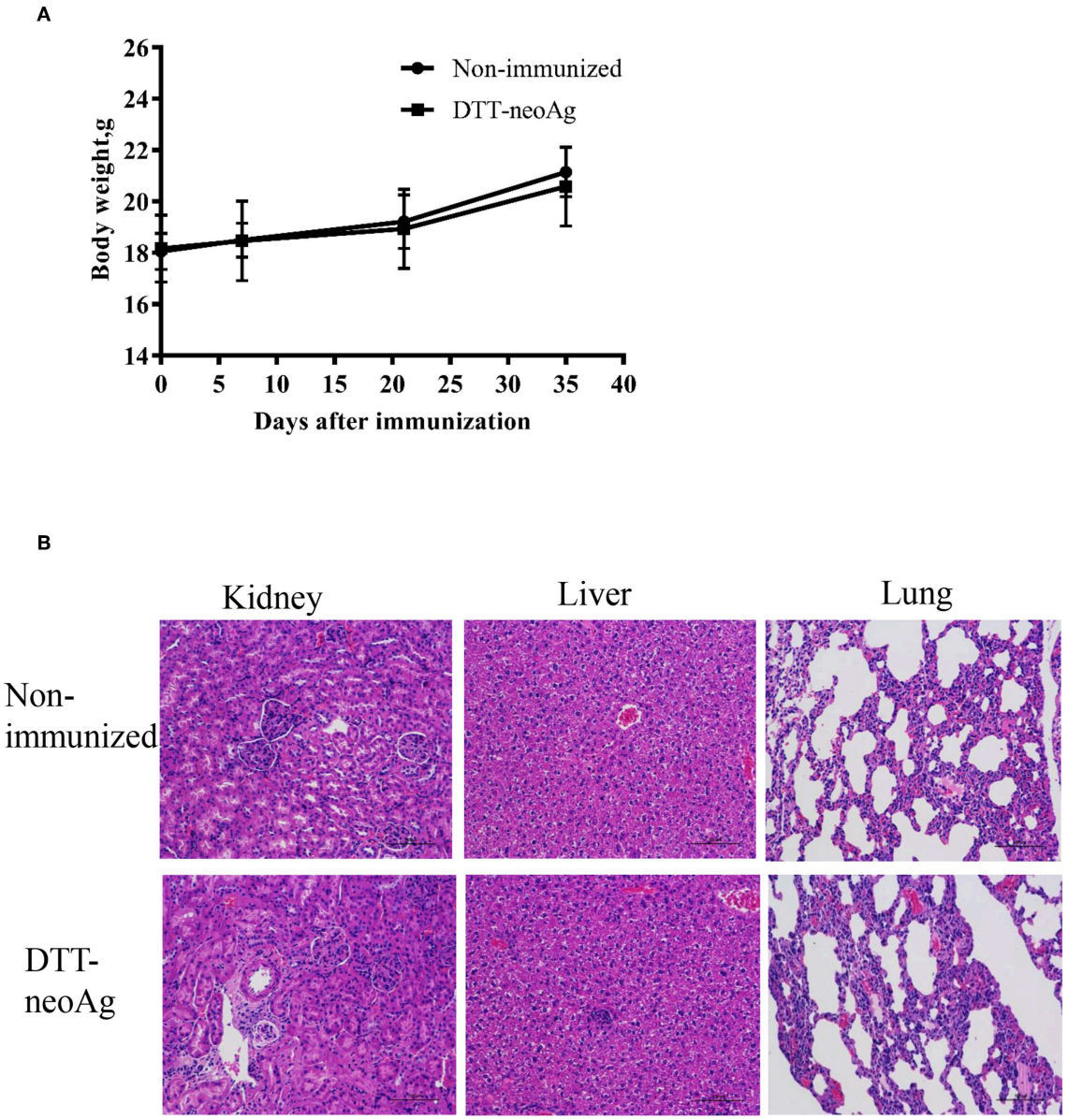


FIGURE 8 | Toxicity evaluation of DTT-neoAg. **(A)** The body weight of the two groups were measured at 7 days after each immunization. **(B)** Morphology evaluation by H&E. Two hundred and ten days after the third immunization, the kidneys, livers, and lungs of DTT-neoAg-immunized mice were collected and analyzed by H&E.

Furthermore, the percentage of $CD8^+$ T cells in the tumor lesions of DTT-neoAg immunized mice was 11.48 ± 0.35%, a 2-fold increase as compared with that of the PBS-treated mice (5.25 ± 0.41%) (**Figure 6E**).

To verify whether tumor-infiltrating lymphocytes have anti-tumor cytotoxic activity, we isolated TILs from tumor tissues of DTT-neoAg-treated or PBS-treated mice. The TILs were stimulated with CTB-neoAg for 72 h, and the cytotoxic activities were measured. Indeed, significant cytotoxic activities were observed in TILs of DTT-neoAg-treated mice. The percentage of neoAg-specific cell lysis was 45.21 ± 0.67% at effector/target ratio of 50:1, but of the PBS group was only 21.37 ± 0.61% at ratio of 50:1 (**Figure 6F**). In addition, extensive necrosis and hemorrhage were observed in tumor sections of DTT-neoAg-immunized mice (**Figure 6G**), but not in those of the PBS-treated mice. These data indicated that DTT-neoAg induced a mix of Th1 and Th2 immunity in tumors.

DTT-neoAg Vaccination Promotes NK Cells Infiltration into Tumor and Reduces $Foxp3^+/CD4^+$ Ratio in TILs and Spleens

Further to assess the effect of the DTT-neoAg vaccination on cellular immune composition of TILs, we measured the

percentage of NK cells in the tumor tissues and FoxP3$^+$/CD4$^+$ ratio in the TILs. The percentage of NK1.1$^+$ cells in the tumor lesions of DTT-neoAg immunized mice was 3.67 ± 0.38%, a 7-fold increase as compared with that of the PBS-treated mice (0.51 ± 0.04%) (**Figure 7A**).

The FoxP3$^+$/CD4$^+$ ratio in the TILs of DTT-neoAg-vaccinated mice was lower than that of PBS-treated mice (**Figure S3A**), although they were statistically not significant. The ratio of Foxp3$^+$/CD4$^+$ (5.43 ± 0.59%) in the spleens of mice treated with DTT-neoAg was markedly decreased as compared to that of PBS-treated mice (15.62 ± 3.39%) (**Figure 7B**). These data indicate that DTT-neoAg vaccine can alter cellular immune composition of the tumor microenvironments as well as the spleens.

Toxicity Evaluation of DTT-neoAg Vaccine

We monitored the appearance and behavior of the immunized mice for 7 months since the beginning of the experiments, and found no significant difference between the DTT-neoAg-immunized and non-immunized mice in terms of weight loss, ruffing of fur, and feeding behavior (**Figure 8A**). Furthermore, no significant changes were observed in histologies of kidneys, livers, and lungs (**Figure 8B**) based on H&E staining of the tissues thereof, indicating that the vaccine was safe at least in a short term after treatment. Longer term of safety concern warrants further assessment.

DISCUSSION

Self-antigens shared by cancer patients have been used in the development of cancer vaccine, yet the clinical outcomes were disappointing (31). One reason is that the self-antigens usually elicit low immune responses as they are subjected to central tolerance (15). Another reason is that they are usually expressed in normal tissues although at very low levels. Therefore, the antigens are not entirely specific to the tumor (15). So far, only one therapeutic vaccine (sipuleucel-T) has been approved by the US Food and Drug Administration for the treatment of prostate cancer (37). Recently, cancer vaccines targeting multiple tumor neoantigens have demonstrated remarkable therapeutic outcomes by virtue of higher specificity (12, 13). However, the majority of tumor neoantigens are weak immunogens (2). To achieve sufficient therapeutic benefits, a large number of neoantigens are required for the vaccine design (12), which limit their application to those cancers having high mutation loads. In this study, we show that fusion of neoantigens to a Th epitope carrier protein, DTT, can enhance the immunogenicity of weak neoantigens as well as antitumor efficacy thereof. The effects of tumor regression conferred by the fusion vaccine were comparable to those of highly immunogenic neoantigen vaccines (7–10). Therefore, our strategy significantly expand the pool of effective vaccine targets, which should benefit cancer patients with intermediate/low mutation burdens such as Ovary cancer, Prostate cancer, low grade Glioma, Chronic lymphocytic leukemia, and Acute myeloid leukemia (2, 38, 39).

CTL responses play pivotal roles in the success of antitumor immunity by vaccination. Current neoantigen-based cancer vaccines confer therapeutic benefits largely through neoantigen-specific CD4$^+$ T cells (8). Various strategies have been attempted to improve CTL responses such as nanomaterial-based as well as DNA-based vaccination. Nanomaterial vaccines promote antigen uptake and prolong the presentation time on APC (9, 10, 39), while DNA vaccines promote antigen processing and presentation (11, 40). Nevertheless, these strategies have not been successfully used for the weak neoantigens. DTT possesses several T help epitopes which enhance both antibody and CTL responses (41). It is well-recognized that interactions of the B cells with activated CD4$^+$ Th cells is required for induction of high affinity antibody responses (42). The priming, expansion, and memory formation and survival of CD8$^+$ cytotoxic T lymphocytes are known to require CD4$^+$ T cells which help increase expression level of antigen-presentation machinery, costimulatory molecules and cytokines by antigen-presenting cells (APCs) (43–45). We show here that DTT stimulate neoantigen-specific CTL responses, and confer therapeutic benefits in a mouse melanoma model. Of note, the Th epitopes of DTT are recognized by over 70% of human population (41), which indicates that the fusion strategy would benefit a large population of cancer patients.

Diphtheria toxoid and CRM197, a catalytic inactive mutant of Diphtheria toxin have long been used as a vaccine carrier and adjuvant in clinical applications with excellent safety records. Since DTT is a non-toxic domain of the toxin molecule, its safety would not be of great concern as a vaccine adjuvant. It has been shown that pre-immunization of diphtheria toxoid promotes DC migration and recruitment to lymph nodes thereby improves the anti-tumor effect of vaccines (46). Since many people have previously been immunized with DT, DTT-containing vaccine would rapidly induce a memory recall response of CD4$^+$ T cells (47). These could help neoAg rapidly elicit antigen specific CTL responses. Our data demonstrated that DTT-neoAg design is easier and more feasible for clinical cancer immunotherapy than current neoantigen-based vaccine approaches.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/**Supplementary Files**.

ETHICS STATEMENT

All experiments involving mice were performed according to Shanghai Jiao Tong University Experimental Animal Center guidelines. The experimental protocol involving animals were reviewed and approved by Shanghai Jiao Tong University Laboratory Animal Ethics and Use Committee.

AUTHOR CONTRIBUTIONS

YZ, LD, and RL conceived the experiments. YZ, ZL, YW, and HC performed experiments. YZ and LD wrote the manuscript.

FUNDING

This work was supported by the National key research and development program of China (2017YFC0909002), and the National Science and Technology Major Projects of China (Major Infectious Diseases 2017ZX10201301-003-004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02472/full#supplementary-material

REFERENCES

1. Vogelstein B, Kinzler KW. The path to cancer—three strikes and you're out. *N Engl J Med*. (2015) 373:1895–98. doi: 10.1056/NEJMp1508811
2. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. (2015) 348:69–74. doi: 10.1126/science.aaa4971
3. Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, et al. Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J Clin Invest*. (2015) 125:3981–91. doi: 10.1172/JCI82416
4. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. (2015) 348:124–8. doi: 10.1126/science.aaa1348
5. Strønen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science*. (2016) 352:1337–41. doi: 10.1126/science.aaf2288
6. Delamarre L, Mellman I, Yadav M. Neo approaches to cancer vaccines. *Science*. (2015) 348:760–1. doi: 10.1126/science.aab3465
7. Castle JC, Kreiter S, Diekmann J, Lower M, van de Roemer N, de Graaf J, et al. Exploiting the mutanome for tumor vaccination. *Cancer Res*. (2012) 72:1081–91. doi: 10.1158/0008-5472.CAN-11-3722
8. Kreiter S, Vormehr M, van de Roemer N, Diken M, Lower M, Diekmann J, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature*. (2015) 520:692–6. doi: 10.1038/nature14426
9. Li AW, Sobral MC, Badrinath S, Choi Y, Graveline A, Stafford AG, et al. A facile approach to enhance antigen response for personalized cancer vaccination. *Nat Mater*. (2018) 17:528–34. doi: 10.1038/s41563-018-0028-2
10. Kuai R, Ochyl LJ, Bahjat KS, Schwendeman A, Moon JJ. Designer vaccine nanodiscs for personalized cancer immunotherapy. *Nat Mater*. (2017) 16:489–96. doi: 10.1038/nmat4822
11. Aurisicchio L, Salvatori E, Lione L, Bandini S, Pallocca M, Maggio R, et al. Poly-specific neoantigen-targeted cancer vaccines delay patient derived tumor growth. *J Exp Clin Cancer Res*. (2019) 38:78. doi: 10.1186/s13046-019-1084-4
12. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*. (2017) 547:217–21. doi: 10.1038/nature22991
13. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. (2017) 547:222–6. doi: 10.1038/nature23003
14. Hu Z, Leet D, Sarkizova S, Holden R, Sun J, Klaeger S, et al. Abstract A010: personalized neoantigen-targeting vaccines for high-risk melanoma generate epitope spreading. *Cancer Immunol Res*. (2019) 7 (Suppl. 2):AACR. doi: 10.1158/2326-6074.CRICIMTEATIAACR18-A010
15. Hu Z, Ott PA, Wu CJ. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat Rev Immunol*. (2018) 18:168–82. doi: 10.1038/nri.2017.131
16. Linnemann C, van Buuren MM, Bies L, Verdegaal EM, Schotte R, Calis JJ, et al. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med*. (2015) 21:81–5. doi: 10.1038/nm.3773
17. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, Dudley ME, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science*. (2014) 344:641–5. doi: 10.1126/science.1251102
18. Robbins PF, Lu YC, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med*. (2013) 19:747–52. doi: 10.1038/nm.3161
19. Tran E, Ahmadzadeh M, Lu YC, Gros A, Turcotte S, Robbins PF, et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science*. (2015) 350:1387–90. doi: 10.1126/science.aad1253
20. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med*. (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
21. Liu XS, Mardis ER. Applications of immunogenomics to cancer. *Cell*. (2017) 168:600–12. doi: 10.1016/j.cell.2017.01.014
22. Bekri S, Uduman M, Gruenstein D, Mei AH-C, Tung K, Rodney-Sandy R, et al. Neoantigen synthetic peptide vaccine for multiple myeloma elicits T cell immunity in a pre-clinical model. *Am Soc Hematology*. (2017) 130 (Suppl 1):1868.
23. Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent advances in targeting CD8 T-cell immunity for more effective cancer immunotherapy. *Front Immunol*. (2018) 9:14. doi: 10.3389/fimmu.2018.00014
24. Xu A, Zhang L, Chen Y, Lin Z, Li R. Immunogenicity and efficacy of a rationally designed vaccine against vascular endothelial growth factor in mouse solid tumor models. *Cancer Immunol Immunother*. (2017) 66:181–92. doi: 10.1007/s00262-016-1928-0
25. Zhang L, Wang J, Xu A, Zhong C, Lu W, Deng L, et al. A rationally designed TNF-alpha epitope-scaffold immunogen induces sustained antibody response and alleviates collagen-induced arthritis in mice. *PLoS ONE*. (2016) 11:e0163080. doi: 10.1371/journal.pone.0163080
26. Lutz MB, Kukutsch N, Ogilvie ALJ, Rossner S, Koch F, Romani N, et al. An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow. *J Immunol Methods*. (1999) 223:77–92. doi: 10.1016/S0022-1759(98)00204-X
27. Xie X, Zhou W, Hu Y, Chen Y, Zhang H, Li Y. A dual-function epidermal growth factor receptor pathway substrate 8 (Eps8)-derived peptide exhibits a potent cytotoxic T lymphocyte-activating effect and a specific inhibitory activity. *Cell Death Dis*. (2018) 9:379. doi: 10.1038/s41419-018-0420-5
28. Kennedy R, Celis E. Multiple roles for CD4+ T cells in anti-tumor immune responses. *Immunol Rev*. (2008) 222:129–44. doi: 10.1111/j.1600-065X.2008.00616.x
29. Fang M, Siciliano NA, Hersperger AR, Roscoe F, Hu A, Ma X, et al. Perforin-dependent CD4+ T-cell cytotoxicity contributes to control a murine poxvirus infection. *Proc Natl Acad Sci USA*. (2012) 109:9983–8. doi: 10.1073/pnas.1202143109
30. Lauritzsen G, Bogen B. The role of idiotype-specific, CD4+ T cells in tumor resistance against major histocompatibility complex class II molecule negative plasmacytoma cells. *Cell Immunol*. (1993) 148:177–88. doi: 10.1006/cimm.1993.1100
31. Yarchoan M, Johnson BA III, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment antitumour immunity. *Nat Rev Cancer*. (2017) 17:209–22. doi: 10.1038/nrc.2016.154
32. Shimizu K, Thomas EK, Giedlin M, Mule JJ. Enhancement of tumor lysate- and peptide-pulsed dendritic cell-based vaccines by the addition of foreign helper protein. *Cancer Res*. (2001) 61:2618–24.
33. Cheng C, Deng L, Li R. The immunogenicity and anti-tumor efficacy of a rationally designed EGFR vaccine. *Cell Physiol Biochem*. (2018) 46:46–56. doi: 10.1159/000488408
34. Zhong C, Zhang L, Chen L, Deng L, Li R. Coagulation factor XI vaccination: an alternative strategy to prevent thrombosis. *J Thromb Haemost*. (2017) 15:122–30. doi: 10.1111/jth.13561
35. Overwijk WW, Restifo NP. B16 as a mouse model for human melanoma. *Curr Protoc Immunol*. (2001) Chapter 20:Unit 20.21. doi: 10.1002/0471142735.im2001s39

36. Shukla GS, Sun YJ, Pero SC, Sholler GS, Krag DN. Immunization with tumor neoantigens displayed on T7 phage nanoparticles elicits plasma antibody and vaccine-draining lymph node B cell responses. *J Immunol Methods*. (2018) 460:51–62. doi: 10.1016/j.jim.2018.06.009
37. Cheever MA, Higano CS. PROVENGE (Sipuleucel-T) in prostate cancer: the first FDA-approved therapeutic cancer vaccine. *Clin Cancer Res*. (2011) 17:3520–6. doi: 10.1158/1078-0432.CCR-10-3126
38. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science*. (2013) 339:1546–58. doi: 10.1126/science.1235122
39. Martin SD, Brown SD, Wick DA, Nielsen JS, Kroeger DR, Twumasi-Boateng K, et al. Low mutation burden in ovarian cancer may limit the utility of neoantigen-targeted vaccines. *PLoS ONE*. (2016) 11:e0155189. doi: 10.1371/journal.pone.0155189
40. Duperret EK, Perales-Puchalt A, Stoltz R, G HH, Mandloi N, Barlow J, et al. A synthetic DNA, multi-neoantigen vaccine drives predominately MHC class I CD8(+) T-cell responses, impacting tumor challenge. *Cancer Immunol Res*. (2019) 7:174–82. doi: 10.1158/2326-6066.CIR-18-0283
41. Diethelm-Okita BM, Okita DK, Banaszak L, Conti-Fine BM. Universal epitopes for human CD4+ cells on tetanus and diphtheria toxins. *J Infect Dis*. (2000) 181:1001–9. doi: 10.1086/315324
42. Steinaa L, Rasmussen PB, Gautam A, Mouritsen S. Breaking B-cell tolerance and CTL tolerance in three OVA-transgenic mouse strains expressing different levels of OVA. *Scand J Immunol*. (2008) 67:113–20. doi: 10.1111/j.1365-3083.2007.02045.x
43. Steinaa L, Rasmussen PB, Wegener AM, Sonderbye L, Leach DR, Rygaard J, et al. Linked foreign T-cell help activates self-reactive CTL and inhibits tumor growth. *J Immunol*. (2005) 175:329–34. doi: 10.4049/jimmunol.175.1.329
44. Melssen M, Slingluff CL Jr. Vaccines targeting helper T cells for cancer immunotherapy. *Curr Opin Immunol*. (2017) 47:85–92. doi: 10.1016/j.coi.2017.07.004
45. Shedlock DJ, Shen H. Requirement for CD4 T cell help in generating functional CD8 T cell memory. *Science*. (2003) 300:337–9. doi: 10.1126/science.1082305
46. Mitchell DA, Batich KA, Gunn MD, Huang MN, Sanchez-Perez L, Nair SK, et al. Tetanus toxoid and CCL3 improve dendritic cell vaccines in mice and glioblastoma patients. *Nature*. (2015) 519:366–9. doi: 10.1038/nature14320
47. Fraser CC, Altreuter DH, Ilyinskii P, Pittet L, LaMothe RA, Keegan M, et al. Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. *Vaccine*. (2014) 32:2896–903. doi: 10.1016/j.vaccine.2014.02.024

Conflict of Interest: LD and RL were employed by company Shanghai HyCharm Inc., Shanghai, China. YZ and RL have a patent "一种靶向癌症变异肽的免疫制剂及应用" (Immunological preparations and application targeting cancer mutant peptides) pending.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 15 November 2019
doi: 10.3389/fimmu.2019.02572

An *in silico—in vitro* Pipeline Identifying an HLA-A*02:01$^+$ KRAS G12V$^+$ Spliced Epitope Candidate for a Broad Tumor-Immune Response in Cancer Patients

Michele Mishto[1,2], Artem Mansurkhodzhaev[3], Ge Ying[4], Aruna Bitra[4], Robert A. Cordfunke[5], Sarah Henze[3], Debdas Paul[3], John Sidney[6], Henning Urlaub[7,8], Jacques Neefjes[9], Alessandro Sette[6,10], Dirk M. Zajonc[11,12] and Juliane Liepe[3,13]**

[1] Centre for Inflammation Biology and Cancer Immunology (CIBCI) & Peter Gorer Department of Immunobiology, King's College London, London, United Kingdom, [2] Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health, Institut für Biochemie, Berlin, Germany, [3] Quantitative and Systems Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany, [4] Division of Immune Regulation, La Jolla Institute for Allergy and Immunology, La Jolla, CA, United States, [5] Department of Immunohematology and Bloodbank, Leiden University Medical Center LUMC, Leiden, Netherlands, [6] Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA, United States, [7] Bioanalytical Mass Spectrometry Group, Max-Planck-Institute for Biophysical Chemistry, Goettingen, Germany, [8] Institut for Clinical Chemistry, University Medical Center Goettingen Bioanalytics, Goettingen, Germany, [9] Department of Cell and Chemical Biology, Oncode Institute, Leiden University Medical Center LUMC, Leiden, Netherlands, [10] Department of Medicine, University of California, San Diego, San Diego, CA, United States, [11] Division of Immune Regulation, La Jolla Institute for Immunology, La Jolla, CA, United States, [12] Department of Internal Medicine, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium, [13] Department of Life Sciences, Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London, United Kingdom*

OPEN ACCESS

Edited by:
John M. Maris, University of Pennsylvania, United States

Reviewed by:
Graham Robert Leggatt, University of Queensland, Australia
Jan Joseph Melenhorst, University of Pennsylvania, United States

****Correspondence:***
Michele Mishto
michele.mishto@kcl.ac.uk
Juliane Liepe
jliepe@mpibpc.mpg.de

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *07 August 2019*
Accepted: *16 October 2019*
Published: *15 November 2019*

Citation:
*Mishto M, Mansurkhodzhaev A, Ying G, Bitra A, Cordfunke RA, Henze S, Paul D, Sidney J, Urlaub H, Neefjes J, Sette A, Zajonc DM and Liepe J (2019) An in silico—in vitro Pipeline Identifying an HLA-A*02:01$^+$ KRAS G12V$^+$ Spliced Epitope Candidate for a Broad Tumor-Immune Response in Cancer Patients. Front. Immunol. 10:2572. doi: 10.3389/fimmu.2019.02572*

Targeting CD8$^+$ T cells to recurrent tumor-specific mutations can profoundly contribute to cancer treatment. Some of these mutations are potential tumor antigens although they can be displayed by non-spliced epitopes only in a few patients, because of the low affinity of the mutated non-spliced peptides for the predominant HLA class I alleles. Here, we describe a pipeline that uses the large sequence variety of proteasome-generated spliced peptides and identifies spliced epitope candidates, which carry the mutations and bind the predominant HLA-I alleles with high affinity. They could be used in adoptive T cell therapy and other anti-cancer immunotherapies for large cohorts of cancer patients. As a proof of principle, the application of this pipeline led to the identification of a KRAS G12V mutation-carrying spliced epitope candidate, which is produced by proteasomes, transported by TAPs and efficiently presented by the most prevalent HLA class I molecules, HLA-A*02:01 complexes.

Keywords: proteasome, peptide splicing, adoptive T cell therapy targets, antigen presentation, cancer epitopes, KRAS, tumor immunology

INTRODUCTION

Activating CD8$^+$ T cells to recurrent tumor-specific mutations is one of a number of cutting-edge strategies to treat cancer. It can be achieved by immunotherapy approaches such as adoptive T cell therapy (ATT), peptide vaccination and dendritic cell (DC) vaccination. Neoepitopes that carry cancer recurrent mutations and efficiently bind common Human Leukocyte Antigen class I (HLA-I) variants are ideal targets for tumor immunology vaccination of large cohorts of patients.

Peptide epitopes are generally produced by proteasomes, which are the final effectors of the ubiquitin-proteasome system (1). Epitope production is the first step of the antigen processing and presentation (APP) pathway, which accounts for the epitope translocating into the endoplasmic reticulum (ER) lumen through mediation by transporters associated with antigen processing (TAPs), binding to the peptide loading complex, trimming by exopeptidases, binding to the HLA-I complex, and transport to the cell surface for recognition by cytotoxic T lymphocytes (CTLs) (2).

There are several proteasome isoforms that can be involved in APP. The most active proteasome isoform is a large (26S) protease consisting of a 20S proteasome core coupled to one or two 19S regulatory complexes. The 19S subunit contains the ubiquitin recognition and removal system as well as an unfolding activity, the 20S form is the actual protease. The 20S proteasome is constituted of four rings; two α rings at the apexes; and two β rings forming the central chamber. Each ring has seven distinct subunits. Each β ring carries three catalytic (i.e., β1, β2, and β5) subunits, which have distinct preferences for peptide sequence motifs (1). Human cells can express different isoforms of catalytic subunits, which are incorporated in distinct proteasome isoforms. Standard proteasomes (s-proteasomes) contain β1, β2, and β5 subunits. Immunoproteasomes (i-proteasomes) contain β1i, β2i, and β5i subunits and are constitutively present in immune cells, such as mature DCs, as well as in cells exposed to an inflammatory milieu (3). Tumors express various intermediate-type proteasome isoforms, in which standard- and immuno-subunits are assembled in one proteasome complex (4, 5). 20S proteasome is also functional alone in cells or coupled to other regulatory subunits such as PA28 αβ (3, 6, 7).

Proteasomes can break proteins and release the peptide fragments or re-ligate them, thereby forming "spliced peptides," which have sequences that do not recapitulate the linear sequence of the parental protein (3, 8, 9). Spliced peptides may represent a sizeable portion of the peptide pool bound to HLA-I molecules—i.e., the HLA-I immunopeptidome—of tumor and non-tumor human cell lines (10–12). This hypothesis is currently a matter of debate since different analytical approaches obtained discordant results (10–16). According to our previous analysis, however, several antigens displayed by HLA-I immunopeptidomes are represented only by spliced peptides. The antigens represented by spliced peptides seem to be longer, more hydrophobic and more basic than those represented by non-spliced peptides. Antigens that are represented by both spliced and non-spliced peptides show antigenic hot spots, i.e., antigenic areas from which both spliced and non-spliced peptides derive (12). HLA-I-bound spliced peptides are generally less abundant than non-spliced peptides (10–12, 17). Proteasome-generated spliced epitopes can trigger specific CTL responses *ex vivo*/*in vivo* against tumor- and type 1 diabetes-associated antigens (17, 18) as well as pathogens (19). ATT targeting a spliced epitope successfully treated a melanoma patient (20, 21).

Proteasome-catalyzed peptide splicing (PCPS; see **Figure 1A**) can occur by combining two non-contiguous sequences of the same molecule (*cis* PCPS) or of two distinct molecules (*trans* PCPS). The latter seems to be efficiently catalyzed *in vitro* by purified proteasomes (22–24) and may constitute a large portion of the HLA-I associated spliced immunopeptidomes (11).

Although the role of spliced peptides in central tolerance still has to be investigated, the theoretically large sequence variability of spliced peptides makes them attractive for anti-cancer immunotherapy (25). Indeed, some of the most recurrent driver mutations in tumors often cannot be efficiently represented by canonical non-spliced peptides bound to the predominant HLA-I variants because of antigen sequence restrictions. On the contrary, they might be represented on the cell surface by tumor-specific spliced epitopes. Therefore, the identification of tumor antigen-specific spliced epitopes might represent a unique opportunity to treat a large cohort of patients.

We here present a pipeline combining *in silico* and *in vitro* approaches. It successfully identifies tumor-specific spliced and non-spliced epitope candidates, which can be further validated as targets for anti-cancer immunotherapies, as illustrated by the HLA-A*02:01^{+} KRAS G12V^{+} spliced epitope candidate here described (**Figure 1B**).

MATERIALS AND METHODS

Antigen Selection and Spliced Epitope Candidate Identification

To rank antigens that are over-represented in HLA-I immunopeptidomes by spliced and non-spliced peptides according to their protein characteristics, we generated a simple model based on the following characteristics: length, hydrophobicity index, isoelectric point, and instability index.

We first calculated these characteristics for all proteins in the Uniprot Reference human proteome database. Next, we analyzed the previously published HLA-I immunopeptidomes of GR-LCL, HCC1143 and HCT116 cell lines (12). All identified spliced and non-spliced peptides were mapped to their antigen(s) of origin, thereby determining a set of represented antigens. Among those antigens not represented in the MS-detected HLA-I immunopeptidomes, there are likely many antigens that would be represented if we considered larger HLA-I immunopeptidome datasets. For this reason, we compared the characteristics of the represented antigens to the characteristics of all proteins (control set). The aim was to determine a combination of the four selected protein characteristics that has the largest difference between the represented antigen set and the control set.

We define the feature sum (f) as: $f = \sum p_i c_i$, where c is the vector of the four selected features (length, hydrophobicity index, isoelectric point and instability index) and p is the vector of factors ranging from −1 to 1. A factor of −1 would favor presentation, while a factor of 1 would disfavor presentation. We use Bayesian inference in a Markov Chain Monte Carlo scheme to determine the factors that result in a distribution of f for represented antigens (F_1), which is most different to the distribution of f for control proteins (F_0). The difference between the densities $F_{diff} = F_1$-F_0 indicates which f is favoring (positive values) or disfavoring (negative values) protein representation. Next, we calculated $F_{diff}(f)$ for a set of candidate antigens (BRAF, KRAS, HRAS, NRAS, TP53, CDK4, IDH1, TYR) using

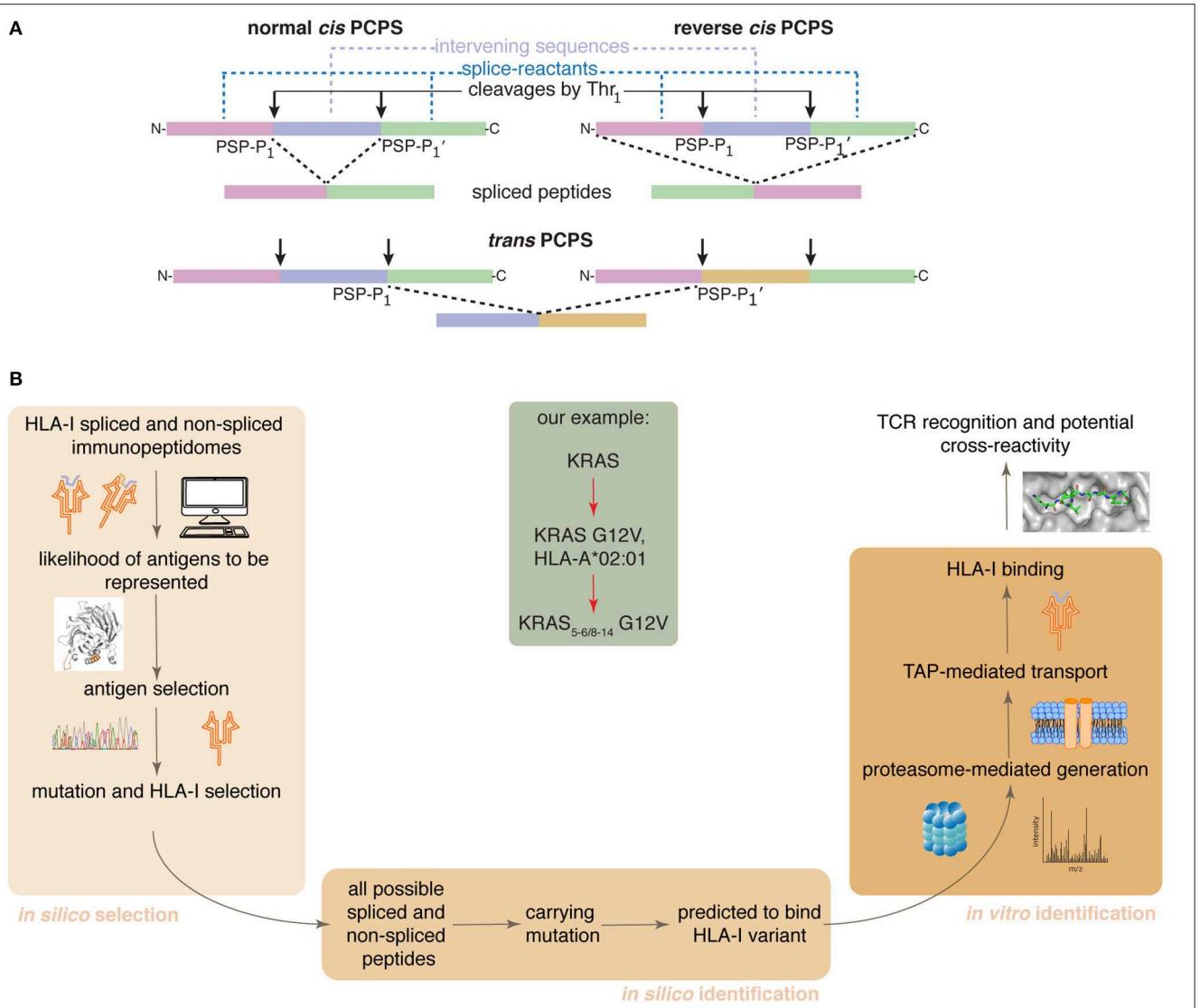


FIGURE 1 | Proteasome-catalyzed peptide splicing and the *in silico/in vitro* pipeline to identify tumor-specific spliced epitope candidates. **(A)** Proteasome-generated spliced peptides can be formed by: (i) *cis* PCPS, when the two splice-reactants derive from the same polypeptide molecule, the ligation can occur in normal order, i.e., following the orientation from N- to C-terminus of the parental protein (normal *cis* PCPS), or in reverse order (reverse *cis* PCPS); (ii) *trans* PCPS, when the two splice-reactants originate from two distinct protein molecules or two distinct proteins. PSP-P1 and PSP-P1' are the C-terminal residue of the N-terminal splice-reactant and the N-terminal residue of the C-terminal splice-reactant, respectively. The splice-reactants are separated by the intervening sequence. **(B)** Representation of the steps of the *in silico – in vitro* pipeline proposed here. Through its application, we identified an HLA-A*02:01^{+} KRAS G12V^{+} spliced epitope candidate (center gray frame).

the Kolmogorov-Smirnov distance. To do so, we sampled (N = 1,000) from the posterior distribution of factors resulting in distributions of F_{diff} for each candidate antigen. Antigens with the highest F_{diff} have characteristics that lead to more likely representation of those antigens in HLA-I immunopeptidomes.

Peptide Synthesis and Proteasome Purification

All peptides were synthesized using Fmoc solid phase chemistry (**Supplementary Table 2**). 20S proteasome was purified from peripheral blood as follows: (i) 10 ml peripheral blood was homogenized, lysed and centrifuged; (ii) the supernatant was fractionated by ammonium sulfate precipitation (35% and then 75%); (iii) the latter pellet was fractioned by chromatography on DEAE-Sephacel; (iv) the selected fractions were separated by 10–40% sucrose gradient and followed by (v) anion exchange chromatography on Mono Q in an Akta-FPLC system; (vi) the selected fractions (2–4 mL) were further purified by DEAE-Affi-gel-blue chromatography. In each of the (ii–vi) steps, the fractions were monitored by degradation assays of standard short fluorogenic substrate Suc-LLVY-AMC. Proteasome concentration was measured by Bradford staining

and verified by Coomassie staining of an SDS-Page gel, as shown elsewhere (26). The purity of the preparation using this protocol has been previously shown (27).

In vitro Digestions and MS Measurements

Synthetic polypeptides (40 μM) were digested by 3 μg 20S proteasomes in 100 μl TEAD buffer for different time points (0–20 h) at 37°C, as previously described (27). We performed three independent experiments, each of them measured either 3 times (for the 0–4 h kinetics) or 2 times (for 20 h digestions) by mass spectrometry (MS).

The identification of target peptide products was carried out by targeted MS using a mass to charge ratio (m/z) inclusion list. The inclusion list was comprised of all theoretically possible 8–12 mer spliced and non-spliced peptide products derived from $KRAS_{2-35}$ G12V synthetic polypeptide substrate, which carried the G12V mutation and were predicted to bind HLA-A*02:01 complex with IC50 $\leq$ 100 nM (see below). The same principle was applied for the peptide products derived from the wild type $KRAS_{2-35}$ G12 synthetic polypeptide substrate (**Supplementary Table 1**). To this end, 20 h *in vitro* digestions with 20S proteasomes were measured by Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific). Prior to measurement, the samples were diluted with the loading buffer (2% acetonitrile, 0.05% Trifluoroacetic acid) containing human insulin (Sigma-Aldrich) to a final substrate concentration of 25 μM and insulin concentration of 2 μM. Insulin was used as a coating polymer to prevent binding of peptides to the glass vials used for measurements and to improve reproducibility between technical replicates. Eight μl of those dilutions (corresponding to 200 pmol of substrate initially present in the sample) were injected. Samples were loaded and separated by a nanoflow HPLC (RSLC Ultimate 3000) on an Easy-spray C18 nano column (30 cm length, 75 μm internal diameter) coupled on-line to a nano-electrospray ionization Fusion Lumos mass spectrometer (Thermo Fisher Scientific). Peptides were eluted with a linear gradient of 5–55% buffer B (80% ACN, 0.1% formic acid) over 88 min at 50°C at a flow rate of 300 nl/min. The instrument was programmed within Xcalibur 3.1.66.10 to acquire MS data in a Data Dependent Acquisition mode using Top 20 precursor ions. We acquired one full-scan MS spectrum at a resolution of 120,000 with an automatic gain control (AGC) target value of 1,000,000 ions and a scan range of 300–1,600 m/z with maximum injection time set to 50 ms and intensity threshold set to 50,000. The MS/MS fragmentation was conducted using HCD collision energy (35%) with an orbitrap resolution of 30,000 at 1.4 m/z isolation window with Fixed First Mass set to 105 m/z. The AGC target value was set up at 100,000 with a maximum injection time of 128 ms. A dynamic exclusion of 30 s and 1–7 included charged states were defined within this method.

In vitro proteasome-mediated digestion kinetics (0–4 h) and the 20 h digestions were measured by LC-MS/MS as follows: Prior to measurement, samples were diluted with the loading buffer and insulin as described above. Eight μl (i.e., 200 pmol substrate) of those dilutions were loaded. Samples were loaded and separated by a nanoflow HPLC (RSLC Ultimate 3000) on an Easy-spray C18 nano column (30 cm length, 75 μm internal diameter; Dr. Maisch) coupled on-line to a nano-electrospray ionization Q Exactive Hybrid-Quadrupol-Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptides were eluted with a linear gradient of 5–55% buffer B (80% ACN, 0.1% formic acid) over 88 min at 50°C at a flow rate of 300 nl/min. The instrument was programmed within Xcalibur 3.1.66.10 to acquire MS data in a Data Dependent Acquisition mode using Top 20 precursor ions. We acquired one full-scan MS spectrum at a resolution of 70,000 with an automatic gain control (AGC) target value of 1,000,000 ions and a scan range of 350∼1,600 m/z. The MS/MS fragmentation was conducted using HCD collision energy (30%) with an Orbitrap resolution of 35,000 at 2 m/z isolation window with Fixed First Mass set to 110 m/z. The AGC target value was set up at 100,000 with a maximum injection time of 128 ms. For Data Dependent Scans the minimum AGC target value and the Intensity threshold were set to 2,600–20,000 accordingly. A dynamic exclusion of 25 s and 1–6 included charged states were defined within this method.

Spliced and Non-spliced Peptide Identification and Quantification As Well as Computation of SCS-P1 and PSP-P1

Peptides were identified using the Mascot version 2.6.1 (Matrix Science) search engine. Mass spectra were searched against a customized database that includes all theoretically possible spliced and non-spliced peptides (28). M oxidation, N-terminal acetylation and NQ deamidation were set as variable modification. For the peptide identification in the Orbitrap Q Exactive measurements, we set as mass tolerances for MS and MS/MS 6 ppm and 20 ppm, respectively. For the peptide identification in the Fusion Lumos measurements, we set as mass tolerances for MS and MS/MS 5 ppm and 0.03 Da, respectively.

Peptide hits were filtered using an ion score cut-off of 20, a *q*-value cut off of 0.05 and a delta score between two spliced peptide hits or between a top scoring spliced peptide and a lower scoring non-spliced peptide of 30% (12). Mascot Distiller's label-free quantification toolbox was used to automatically extract MS ion peak areas of all identified peptides for all five time points (0–4 h) and all three technical replicates simultaneously. Biological replicates were processed separately. The resulting peptide kinetics were filtered for peptide synthesis artifacts and non-reproducible peptide kinetics between technical replicates. Furthermore, peptides that showed unrealistic generation kinetic behavior (such as alternating MS ion peak areas between consecutive time points) were removed. In the final analysis, only peptides that were detected and quantified in two biological replicates were considered.

$KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V generation kinetics were manually quantified by extraction of an ion chromatogram (XIC) corresponding to the peptides monoisotopic peaks, using instrument precursor tolerance and retention time information (from the identified peptides in the 20 h digestions) via Mascot Distiller, followed by determination of the area under the peak at each time point in the kinetics series.

Absolute peptide quantification was carried out through the application of the method QPuB on detected MS ion peak areas for each peptide product, as described elsewhere (see Data availability section). In the specific case of the two epitope candidates $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V, we computed their amount using a titration curve of the cognate synthetic peptides since their amount was too low to be estimated with high confidence using QPuB. Synthetic peptide concentration for titration ranged from 0 to 10 pmol injected. Each titration sample was measured twice and right after measuring *in vitro* digestion samples.

SCS-P1 (site specific cleavage strength after amino acid residue P1) and PSP-P1 (frequency of peptide splicing catalyzed using the C-terminus of the N-terminal splice-reactant as splicing site) were calculated based on the absolute amount of each product (resulting from QPuB) identified in the proteasome-catalyzed digestions (23). Briefly, for each time point and each amino acid in the substrate, the sum over all product (non-spliced and spliced peptides) amount that have the corresponding substrate amino acid at their C-terminus has been computed and normalized, so that they add up to 100%, resulting in SCS-P1. For each time point and each amino acid in the substrate, the sum over all spliced peptide amount that have the corresponding substrate amino acid at their C-terminus of the N-terminal splice-reactant was computed and normalized, so that they add up to 100%, resulting in PSP-P1.

TAP Assay

The transport efficiency of target peptides (**Supplementary Table 2**) into the ER lumen mediated by TAPs was carried out as previously described (29) although some modifications were introduced. These include the use of a fluorescent peptide tracer and the use of microsomes rather than Streptolysin O permeabilized cells. Briefly, peptides were dissolved in DMSO and different concentrations were distributed in a final volume of 10 µl DMSO. At the same time, a mixture of 10 mM ATP, 100 mM Tris-HCl pH7.5 and 5 mM $MgCl_2$ and fluorescent tracer peptide was prepared. 60 µl of this mixture was added to the 10 µl competing peptide mixture to a final volume of 70 µl. This was prewarmed to 37°C and 30 µl of pre-warmed microsomes were added. Microsomes were derived from LCL721 cells, as previously described (29).

The mixture was incubated for 20 min at 37°C followed by cell lysis with lysis mixture (0.5% TX100, 5 mM $MgCl_2$ in 100 mM Tris-HCl pH7.5) at 4°C. After at least 30 min incubation at 4°C, DNA was pelleted and the supernatant transferred to a new vial including ConA-beads. After at least 30 min incubation on ice, cells were washed four times with lysis mixture and the last time with 100 mM Tris-HCl pH7.5 before transfer to 96 wells plates (Corning) followed by fluorescence measurements in a plate reader. To detect background signals, a sample without competing peptide and ATP was included and fully performed as described above. This signal was subtracted from the detected signal. The curves were normalized to the highest value set at 100% and EC_{50} values were calculated.

HLA-I–Peptide Binding Affinity Prediction and Measurement

The binding affinity between theoretical spliced and non-spliced peptides and HLA-A*02:01 was predicted using the NetMHCPan 3.0 algorithm (30). We restricted the prediction to 8–12 mer peptides and imposed an IC_{50} cut-off of 100 nM. The binding affinity between the synthetic peptides and HLA-A*02:01 complexes was measured using purified HLA-I molecules, as described elsewhere (10).

HLA-I-Peptide Crystal Structure and Analysis

The ectodomains of HLA-A*02:01 (residues 21–274) and human β_2-microglobulin ($h\beta_2m$) (residues 1–99) were expressed in *Escherichia coli* BL21 DE3 cells as inclusion bodies after 4 h induction with 1 mm isopropyl 1-thio-d-galactopyranoside (IPTG) at A_{600} of 0.6. Cells were harvested by centrifugation (5,000 × *g* for 20 min), resuspended in lysis buffer (100 mM Tris-HCl, pH7.0, 5 mM EDTA, 5 mM DTT, 0.5 mM PMSF), and broken through a microfluidizer (Microfluidics). Inclusion bodies were collected from cell lysate (50,000 × *g* for 30 min at 4°C), washed 3 times in 100 mM Tris-HCl, pH7.0, 5 mM EDTA, 5 mM DTT, 2 M urea, 2% (w/v) Triton X-100 plus 1 time in 100 mM Tris-HCl, pH7.0, 5 mM EDTA, 2 mM DTT), and finally dissolved in 50 mM Tris-HCl, pH7.0, 5 mM EDTA, 2 mM DTT, 6 M guanidine HCl) for the following refolding.

Refolding was performed in a 100 ml system. Briefly, 1.2 mg of $h\beta_2m$ was loaded dropwise into refolding buffer (0.1 M Tris-HCl, pH8.0, 2 mM EDTA, 400 mM l-arginine, 5 mM oxidized glutathione, 5 mM reduced glutathione) and stirred for 1 h at 4°C. Then, 6 mg of HLA-A*02:01 mixed with 1.2 mg of individual peptide (**Supplementary Table 2**) was added dropwise into the refolding system and stirred at 4°C for up to 72 h. The refolding system was concentrated to 0.5 mL for size exclusion chromatography using a Superdex S200 Increase 10/300 GL column in 20 mM Tris-HCl pH7.5, 150 mM NaCl. Fractions containing refolded HLA-A*02:01-β_2m-peptide complexes were pooled and concentrated to 5–10 mg/ml for subsequent crystallization.

Thick plate-like or 3-dimensional crystals of HLA-A*02:01-β_2m-peptide complexes were obtained by setting drop vapor diffusion at 1:1–1.5 ratio with 30% PEG 4000, 0.1 M Tris-HCl, pH8.5, 0.2 M lithium sulfate at room temperature after 3 days. The crystals were flash frozen in crystallization solution plus glycerol (25% v/v) using liquid nitrogen.

Diffraction data for HLA-A*02:01-β_2m-peptides were collected remotely at beam line 9.2 at the Stanford Synchrotron Radiation Light source and processed to 1.40–1.55 Å resolution, using HKL2000. Phases were obtained by molecular replacement with Phaser MR in ccp4 using the protein coordinates from a former HLA-A*02:01-β_2m-peptide structure (Protein Data Bank code 5ENW) (31). The model was built with COOT (32) and refined with REFMAC5 (33). Data collection and refinement statistics are shown in **Supplementary Table 3**.

Statistical Analysis

If not stated otherwise, all statistical tests have been done in R and differences in distributions have been tested using the Kolmogorov-Smirnov test.

Dataset and Software Availability

A summary of the RAW files of the LC-MS/MS measurements of the *in vitro* digestions accessible via repository is reported in the following Mendeley dataset: http://doi.org/10.17632/63rj3xczmb.1.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (34) partner repository with the dataset identifier PXD015580.

The HLA-I immunopeptidome elution MS files used in the first step of the pipeline are available at the PRIDE repository with the dataset identifier PXD000394 (files: 20120321_EXQ1_MiBa_SA_HCC1143_1.raw, 20120321_EXQ1_MiBa_SA_HCC1143_2.raw, 20120322_EXQ1_MiBa_SA_HCC1143_1_A.raw, 20120515_EXQ3_MiBa_SA_HCT116_mHLA-1.raw, 20120515_EXQ3_MiBa_SA_HCT116_mHLA-2.raw, 20120617_EXQ0_MiBa_SA_HCT116_1_mHLA_2hr.raw, 20120617_EXQ0_MiBa_SA_HCT116_2_mHLA_2hr.raw) and at the Datadryad.org archive (doi: 10.5061/dryad.r984n) and were generated by Bassani et al. (35) and Mommen et al. (36).

QPuB software is available at GitHub (https://github.com/QuantSysBio/QPuB).

RESULTS

Prioritization of KRAS as Antigen Over-represented on Cell Surface

Antigens represented by spliced peptides in HLA-I immunopeptidomes tend to be preferentially long, hydrophobic and basic, thereby suggesting that the chemical and physical characteristics of antigens can impinge upon spliced peptide generation and presentation (12, 24). To select suitable antigens from which a spliced epitope candidate might be derived, we first investigated which combination of protein features may result in more likely potential over-representation in HLA-I immunopeptidomes by spliced and non-spliced peptides.

Accordingly, we used a previously published HLA-I spliced and non-spliced immunopeptidome database (12), which includes 13,666 unique non-spliced and 1,318 unique spliced peptides, as well as 7,328 represented antigens. With this dataset, we developed a simple model based on protein length, hydrophobicity, isoelectric point and instability index to determine the possible over-representation of a given antigen by spliced and non-spliced peptides in HLA-I immunopeptidomes. These characteristics were previously described to influence the probability of observing peptides of a protein being presented in HLA-I immunopeptidomes determined by antigen gene expression level and antigen abundance as key determinants for efficient presentation (12, 37). However, we here opted to focus on protein intrinsic characteristics that are conserved independently of cell types and cell status, to obtain a model for antigen selection that can be generalized.

Combining these four selected protein characteristics yielded a distribution of known represented antigens, which can be compared to the feature distribution of all proteins (**Figure 2A**). Proteins with feature values that show a higher density for represented antigens compared to all proteins are more likely to be favored for antigen presentation than those proteins with feature values that show a lower density for represented antigens compared to all proteins (**Figure 2A**). We therefore aimed to find a combination of features that maximizes the difference between the two distributions. We defined a model calculating a feature sum $\sum p_i\, c_i$, where p_i are factors ranging from -1 (favoring representation) to 1 (disfavoring representation) and c_i are the protein characteristic values.

Using Bayesian inference, we estimated the factors that provide the largest distance between the resulting feature sum distributions for represented antigens compared to all proteins. We found that protein length favors representation. On the contrary, very hydrophobic or instable proteins are disfavored during representation. The isoelectric point appeared to have minor influence (**Figure 2B**).

As proof-of-principle, we focused our analysis on a series of major tumor antigens - BRAF, KRAS, HRAS, NRAS, TP53, CDK4, IDH1, TYR - which all carry recurrent oncogenic mutations. We calculated the feature sums for those eight antigens and determined corresponding density differences for each of those feature sums, which allowed us to rank the candidate antigens (**Figure 2C**). Among them, HRAS and KRAS are the two antigens that are most likely over-represented in HLA-I immunopeptidomes as compared to the whole proteome (**Figure 2C**).

Prioritization of KRAS G12V neoantigen as Source of Potentially Antigenic HLA-A*02:01-Bound Tumor-specific Spliced Peptides

HRAS and KRAS are two GTPases that function as molecular switches in regulatory pathways responsible for proliferation and survival. In particular, KRAS is frequently mutated in cancers with an average of 22% cancers carrying a KRAS mutation, a frequency that rises to 33–61% in colorectal cancer and pancreatic adenocarcinoma (38). The mutations often occur in the KRAS G12 and G13 residues, which impairs the KRAS GTPase activity and renders the mutants persistently in the GTP-bound active form, thereby promoting tumorigenesis and tumor malignancy (39). KRAS G12/13 is a driver in tumors and in combination with its high frequency in cancers makes KRAS an ideal target for immunotherapies. For example, ATT using multiple T cell Receptors (TCRs), which recognize HLA-A*02:01^{+} spliced epitopes carrying KRAS mutations, could treat around 30% of pancreatic adenocarcinoma patients and a large portion of colorectal carcinoma patients. The efficacy of ATT has been demonstrated in a xenograft mouse model (40) and a metastatic colorectal cancer patient (41) by targeting non-spliced epitopes carrying KRAS G12D mutations and presented by HLA-A*11:01 or -C*08:02 molecules, respectively.

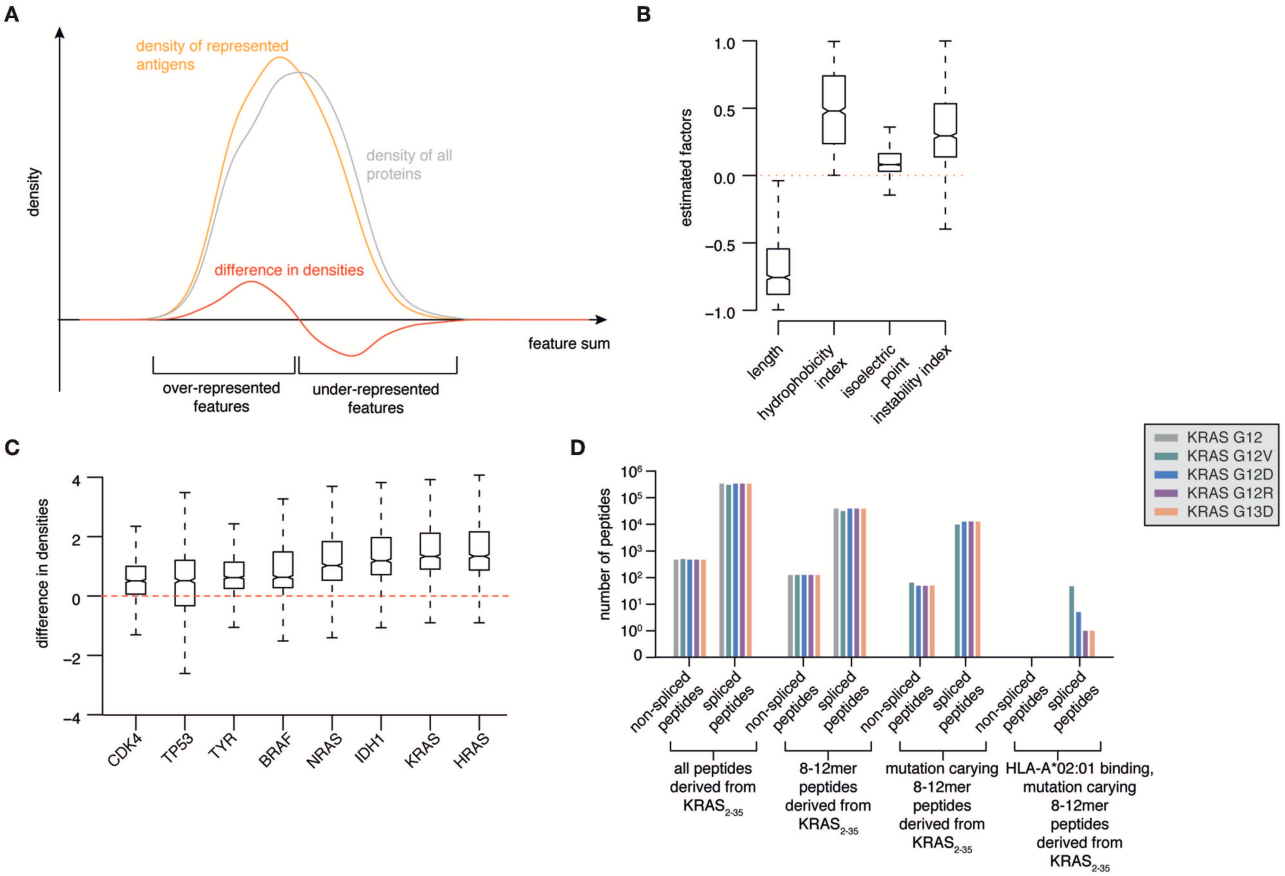


FIGURE 2 | Prioritization of KRAS G12V as source of potentially antigenic HLA-A*02:01-bound spliced peptides. **(A)** Illustration of a model to select antigens with features more likely to be over-represented in HLA-I immunopeptidomes. Protein length, hydrophobicity, isoelectric point and instability are used to calculate a protein feature sum for all represented antigens. The resulting distribution (orange line) is compared to the feature sum distribution based on all human proteins (gray line). The difference between these two distributions (red line) indicates which features favor and disfavor the representation of antigens in HLA-I immunopeptidomes. **(B)** Marginal posterior distributions of the estimated factors that maximize the difference between feature sum distribution of represented antigens vs. all proteins. **(C)** Ranking of selected tumor-associated antigens BRAF, KRAS, HRAS, NRAS, TP53, CDK4, IDH1, TYR as antigens with features over-represented in HLA-I immunopeptidomes. **(D)** Progressive reduction of the theoretical number of spliced and non-spliced epitope candidates, thereby narrowing down to those carrying KRAS G12V/D/R or G13D mutations (*in silico* derived from the $KRAS_{2-35}$ sequence) and predicted to bind HLA-A*02:01 with $IC_{50} \leq 100$ nM.

All these features define KRAS as an attractive tumor antigen to be further investigated using our pipeline. We investigated *in silico* the sequence surrounding residues 12 and 13 of KRAS wild type, G12V/D/R and G13D antigens. All spliced and non-spliced peptides that could be theoretically generated were computed. From this list, we removed all peptide candidates not carrying the target mutations, as well as those candidates shorter than 8 residues or longer than 12 residues, which is the length range most often observed in HLA-I immunopeptidomes (**Figure 2D**). Since the HLA-A*02:01 allele is the predominant allele in Caucasian populations, we predicted the binding affinity of this HLA-I variant to the remaining peptides using the NetMHCPan 3.0 algorithm (30). Finally, we filtered out all peptides that were predicted to bind with $IC_{50} > 100$ nM. None of the non-spliced epitope candidates passed this step, whilst 54 spliced epitope candidates had the required features. Among them, 47 can theoretically carry the KRAS G12V mutation (**Figure 2D**).

Identification of KRAS G12V$^+$ Spliced Epitope Candidates Generated by Proteasomes

The majority of the HLA-I-restricted epitopes are produced by proteasomes. Their production can be verified through *in vitro* digestion of synthetic polypeptides by 20S proteasomes, as measured by MS. Because of the high frequency of putative HLA-A*02:01$^+$ spliced epitope candidates carrying the KRAS G12V mutation (KRAS G12V$^+$), we focused on this mutation and digested the synthetic $KRAS_{2-35}$ wild type and G12V polypeptides with 20S standard proteasomes for 20 h. The digestions were measured by targeted MS, which used a m/z inclusion list of target spliced epitope candidates identified in the previous pipeline step (**Supplementary Table 1**), and confirmed that one spliced epitope candidate, $KRAS_{5-6/8-14}$ G12V [KL][VVGA**V**GV], is generated by proteasomes under these conditions (**Figure 3**). This spliced peptide could be

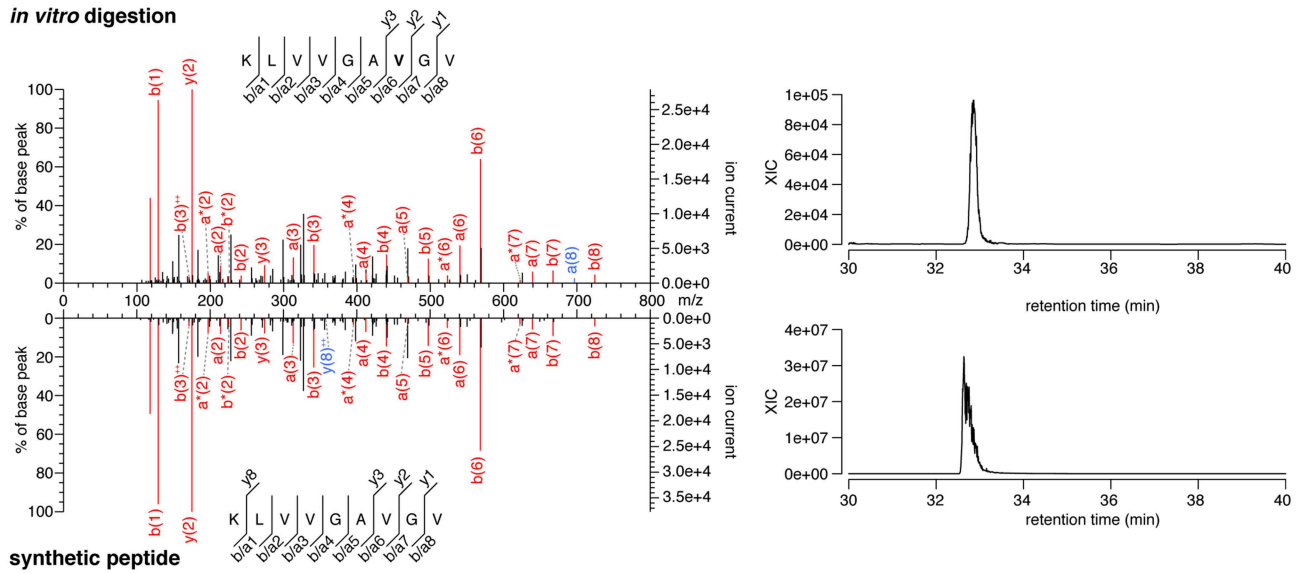


FIGURE 3 | MS/MS spectra of the $KRAS_{5-6/8-14}$ G12V spliced epitope candidate. MS/MS spectrum of the peptide $KRAS_{5-6/8-14}$ G12V [KL][VVGA**V**GV] identified in the *in vitro* digestion of the synthetic polypeptide $KRAS_{2-35}$ G12V and the MS/MS spectrum of the cognate synthetic peptide (left panels). The peptide sequence is shown with the corresponding b-, a- and y-ions identified. The G12V mutation is depicted in bold. In the spectra, assigned peaks for b-, a-, and y-ions are reported in color. Ion neutral loss of ammonia is symbolized by *. Red marked peaks are assigned in both *in vitro* digestion detected MS/MS spectrum and synthetic peptide MS/MS spectrum, whereas blue marked peaks are assigned only in one of the two spectra. The extracted ion chromatogram for the peptide identified in the *in vitro* digestion and the synthetic counterpart is plotted and indicates matching retention times for both peptides (right panels).

generated by the removal of one of the three V residues in the KRAS sequence, i.e., it could be reported as $KRAS_{5-6/8-14}$ G12V, $KRAS_{5-7/9-14}$ G12V, or $KRAS_{5-8/10-14}$ G12V. We will refer to it as $KRAS_{5-6/8-14}$ G12V for the sake of simplicity.

Notably, this spliced peptide is not present in the reaction at $t = 0$ and the 20 h reaction containing the synthetic polypeptide substrate in absence of proteasomes (data not shown). The cognate spliced peptide $KRAS_{5-6/8-14}$ G12 [KL][VVGA**G**GV] is not produced by 20S proteasomes whilst processing the synthetic wild type $KRAS_{2-35}$ polypeptide.

In the $KRAS_{2-35}$ G12V polypeptide digestion, we also identified the non-spliced epitope candidate $KRAS_{5-14}$ G12V [KLVVVGA**V**GV] (**Supplementary Figure 1**). The spontaneous response of peripheral blood mononuclear cells (PBMCs) of pancreatic adenocarcinoma patients against this latter epitope candidate was previously described (42). Although this peptide was filtered out in the early steps of our pipeline because it has a predicted HLA-A*02:01 binding affinity $IC_{50} > 100$ nM, we compared this epitope candidate to the $KRAS_{5-6/8-14}$ G12V spliced epitope candidate in the next validation steps.

Spliced Peptide and $KRAS_{5-6/8-14}$ G12V Spliced Epitope Candidate Production Kinetics by Proteasomes

To be a robust epitope candidate, a peptide should be produced *in vitro* by proteasomes in a detectable amount and with consistent kinetics. Correspondence between *in vitro* experiments carried out with purified 20S proteasomes and *in cellulo* and *in vivo* experiments has been demonstrated in various studies investigating both viral and tumor epitopes (4, 5, 9, 17, 19, 21, 43–50). Therefore, we performed digestion kinetics (0–4 h) of the synthetic $KRAS_{2-35}$ wild type and G12V polypeptides with 20S standard proteasomes. The samples were measured by MS to identify all digestion products (via MS/MS). Quantification of peptides was performed using QPuB, a method that uses detected MS ion peak areas to estimate the absolute amount of each spliced and non-spliced peptide products (see Data Availability section), and by comparison with synthetic peptide titration for the two epitope candidates.

In the synthetic $KRAS_{2-35}$ G12V polypeptide digestion, we identified and successfully quantified 131 peptide products. 65.6% were non-spliced, 31.3% *cis* spliced and 3.1% *trans* spliced peptides (**Figure 4A**). The length distribution of the non-spliced, *cis* spliced and *trans* spliced peptides did not significantly differ and its median was 10 amino acid residues (**Figure 4B**). N- and C-terminal splice-reactants had a median length of 7 and 3 amino acid residues, respectively (**Figure 4C**). The intervening sequences of *cis* spliced peptides had a median length of 5 amino acid residues (**Figure 4C**). From the quantitative point of view, *cis* and *trans* spliced peptides represent proximately 17.0 and 0.1% of the peptide abundance, respectively (**Figure 4D**). On average, a *trans* spliced peptide is less abundant than a *cis* spliced peptide, which is less abundant than an average non-spliced peptide (**Figure 4E**).

Through the application of QPuB to the synthetic $KRAS_{2-35}$ wild type and G12V polypeptide digestions, we could also compute how frequently proteasomes cleaved the substrate after

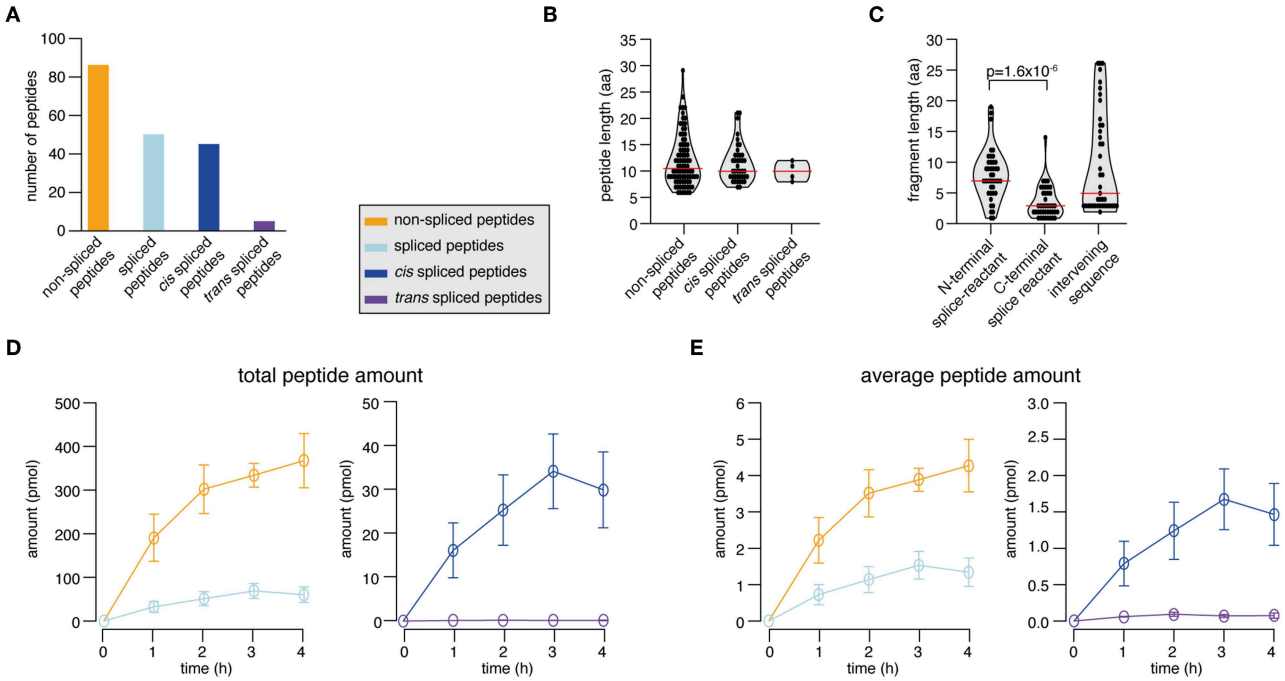


FIGURE 4 | Spliced peptide characteristics and kinetics in $KRAS_{2-35}$ G12V polypeptide degradation. **(A–D)** Results of the analysis of *in vitro* digestions of the synthetic $KRAS_{2-35}$ G12V polypeptide substrate by 20S proteasomes (two biological replicates each measured three times). Spliced and non-spliced peptide products were identified by MS and quantified by applying QPuB. Only peptides identified in both biological replicates with reproducible kinetics have been analyzed. **(A)** Number of non-spliced, *cis* spliced, and *trans* spliced peptides identified in the reactions. **(B)** Length distribution of non-spliced peptides, *cis* spliced and *trans* spliced peptides. **(C)** Length distribution of N- and C-terminal splice-reactants of *cis* spliced peptides as well as of their intervening sequences. **(D)** Total amount of spliced and non-spliced, as well as *cis* spliced and *trans* spliced peptides quantified by applying QPuB to *in vitro* kinetics. **(E)** Abundance of an average spliced and non-spliced, as well as *cis* spliced or *trans* spliced, peptides in the *in vitro* kinetics over the digestion time. In **(B,C)**, violin plots indicate the fragment length distribution. Red lines indicate the median. When statistically significant, *p*-values are reported.

each of its individual residues (substrate cleavage strength, i.e., SCS-P1) or used each residue for the PCPS reaction (proteasome-generated spliced peptide P1 positions, i.e., PSP-P1). From this analysis we confirmed our previous hypothesis (23), whereby proteasomes splice at sites at which the substrates are less frequently cleaved at (and *vice versa*), as emerged by comparing SCS-P1 and PSP-P1 (**Figures 5A,B**).

The quantitative analysis of the $KRAS_{2-35}$ synthetic substrate degradation (**Figure 5C**) also showed that the $KRAS_{5-6/8-14}$ G12V spliced epitope candidate is produced in amounts smaller than the average amount of spliced peptides (**Figures 4E, 5D** shall be compared).

$KRAS_{5-6/8-14}$ G12V Spliced Epitope Candidate Is a TAP Substrate and Efficiently Binds Hla-A*02:01

The production of a peptide by proteasomes is not sufficient alone to incur presentation on the cell surface. There are several other steps in the APP pathway that can direct the peptide fate, such as peptide transport into the ER lumen mediated by TAPs and peptide binding to HLA-A*02:01 complex. We studied the $KRAS_{5-6/8-14}$ G12V [KL][VVGA**V**GV] spliced epitope candidate in comparison with the $KRAS_{5-14}$ G12V [KLVVVGA**V**GV] non-spliced epitope candidate. We also extended the study to four control peptides. Two peptides - peptide #1 YLVVVGAVGV and peptide #2 KLVVVAVGV - shared a large portion of the $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V epitope candidate sequences (**Figure 6A**). The other two control peptides are unrelated peptides predicted to bind to the HLA-A*02:01 complex (peptides #3 FLHEDLEKI and #4 FLHEDTEKI; see **Supplementary Table 2**).

To quantify the efficiency of peptide transport into the ER lumen by TAP, we measured the competition between the target peptides and a fluorescent reference peptide for TAP-dependent translocation into free microsomes. The reference peptide has an N-linked glycosylation consensus sequence and peptide glycosylation is used to monitor entry into ER microsomes and as an isolation handle (29). $KRAS_{5-6/8-14}$ G12V epitope candidate is efficiently transported by TAPs as the $KRAS_{5-14}$ G12V non-spliced epitope candidate is. Peptide #1, which has a K to Y substitution at position 1 as compared to $KRAS_{5-14}$ G12V peptide, is transported by TAP as efficiently as the non-spliced epitope candidate. In contrast, peptide #2, which has the removal of residue G in position 6 as compared to $KRAS_{5-14}$ G12V peptide and a G to V substitution at position 5 as compared to $KRAS_{5-6/8-14}$ G12V peptide, is not competing with the reference peptide and thus ignored by TAP. This suggests a role

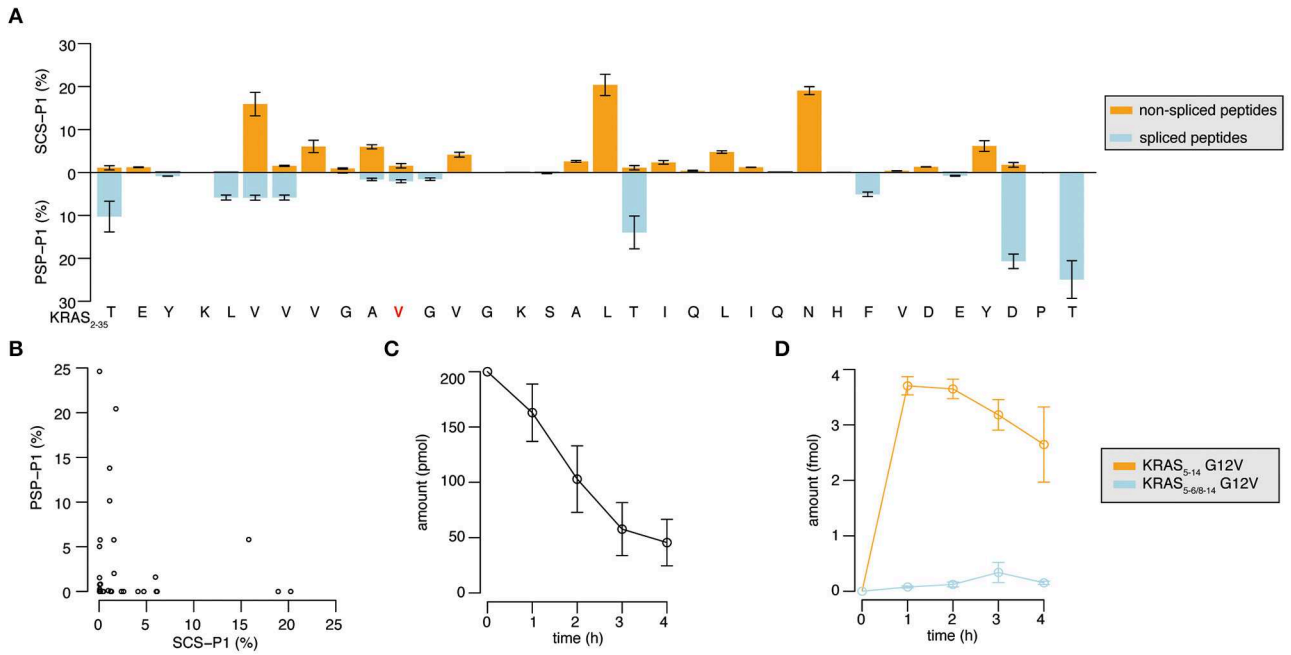


FIGURE 5 | Substrate cleavage- and splicing-site preferences and generation kinetics of the $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V epitope candidates. **(A)** Relative usage of the substrate sites for cleavage (SCS-P1), and splicing (PSP-P1) in the *in vitro* digestion kinetics of the synthetic $KRAS_{2-35}$ G12V polypeptide substrate by 20S proteasomes (two biological replicates each measured three times). Mean and *SD* of biological replicates (bars) are shown. **(B)** Scatter plot of SCS-P1 and PSP-P1, which depicts the absence of direct correlation between splicing and cleavage frequencies. In **(A,B)** spliced and non-spliced peptide products were identified by MS/MS in the *in vitro* digestion kinetics of synthetic polypeptide $KRAS_{2-35}$ G12V with 20S proteasomes, and were quantified by applying QPuB based on their MS ion peak area. SCS-P1 and PSP-P1 were computed using the average amount of all time points of each peptide product. **(C,D)** Abundance of the synthetic polypeptide substrate $KRAS_{2-35}$ **(C)** as well as the $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V epitope candidates **(D)** in the *in vitro* digestions with 20S proteasomes (three biological replicates each measured 3–4 times). Spliced and non-spliced peptide products were identified by MS/MS and quantified based on their MS ion peak area, using titration of synthetic peptides as reference.

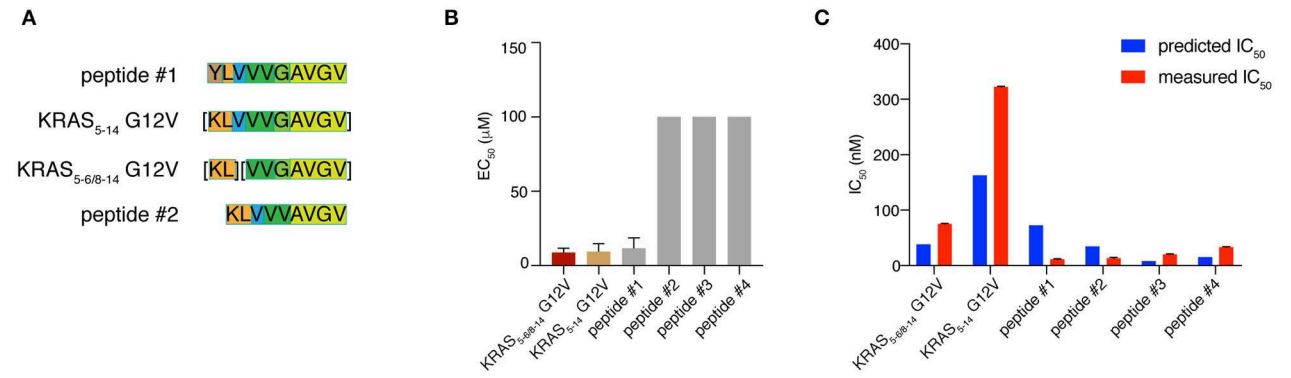


FIGURE 6 | $KRAS_{5-6/8-14}$ G12V epitope candidate is efficiently transported by TAPs and strongly binds HLA-A*02:01 complex. **(A)** Sequence comparison between $KRAS_{5-6/8-14}$ G12V [KL][VVGA**V**GV] and $KRAS_{5-14}$ G12V [KLVVVGA**V**GV] epitope candidates as well as their modified versions (peptides #1 YLVVVGAVGV and #2 KLVVVAVGV). Common sequences among peptides are color-coded. **(B)** Transport efficiency into the ER lumen mediated by TAPs of $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V epitope candidates, their modified versions (peptides #1 and #2) and two unrelated peptides (peptides #3 and #4; **Supplementary Table 2**). The EC_{50} was computed using a competing peptide as reference. We here report the EC_{50} values obtained upon subtracting the peptide transport in absence of ATP. **(C)** Predicted and measured binding affinities of the peptides to the HLA-A*02:01 complex. Binding affinity prediction was carried out with the NetMHCPan 3.0 algorithm. In **(B,C)** mean and SD of biological replicates (bars) are shown.

of the residue G at the center of the peptides in TAP-mediated transport. The other two control peptides are not substrates for TAP (**Figure 6B**).

Once a peptide arrives in the ER lumen, its binding affinity to the specific HLA-I molecule determines whether it will ultimately be presented. Therefore, we measured the binding affinity of the

same six peptides previously tested in the TAP assay and the HLA-A*02:01 complex in a cell-free system utilizing purified HLA-I molecules. The $KRAS_{5-6/8-14}$ G12V epitope candidate was confirmed to efficiently bind the HLA-A*02:01 complex, in contrast to the $KRAS_{5-14}$ G12V non-spliced epitope candidate, which had a measured IC_{50} larger than 300 nM (and larger than predicted). The measured and predicted IC_{50} of the control peptides was quite similar and the peptides appear to be good binders (**Figure 6C**).

Conformation of $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V Epitope Candidates Within HLA-A*02:01 cleft

Once a peptide is bound to HLA-I complexes and presented at the cell surface, it can be recognized by TCRs of $CD8^+$ T cells. The conformation of the peptide in the HLA-I groove is paramount not only for HLA-I-peptide affinity and stability, but also for the TCR-HLA-I-peptide interaction. To study this aspect, we individually refolded and crystallized HLA-A*02:01 with spliced epitope candidate $KRAS_{5-6/8-14}$ G12V, non-spliced epitope candidate $KRAS_{5-14}$ G12V, as well as two control peptides (peptides #1 and #2), in which either the N-terminal residue or one of the central residues was substituted, as compared to the epitope candidates (**Figures 7A–K**). The crystal structures of these individual HLA-I-peptide complexes were determined at resolutions ranging from 1.4 to 1.58 Å by molecular replacement using PDB ID 5ENW as a search model (**Supplementary Table 3**). The global superposition of all these four peptide-HLA-I complexes in the peptide binding groove reveals a similar binding orientation with a root mean square deviation value (rmsd) of 0.124 Å. The electron densities for all four peptides are also well-defined over the entire peptide length.

Some structural differences in individual peptide binding were observed when comparing the 9 mer peptides ($KRAS_{5-6/8-14}$ G12V [KL][VVGA**V**GV] and peptide #2 KLVVVAVGV) with the 10 mer peptides ($KRAS_{5-14}$ G12V [KLVVVGA**V**GV] and peptide #1 YLVVVGAVGV). Comparison of peptides with a same length generally only reveals a single amino acid change in a similar orientation or the addition of a side chain, e.g., when V replaces G in peptide #2 as compared to $KRAS_{5-14}$ G12V peptide (**Figures 7A–D**). Specifically, while the N-terminal and C-terminal ends of all four peptides match perfectly, structural superposition reveals conformations in the middle portions of peptide $KRAS_{5-6/8-14}$ G12V and peptide #2 unique from the remaining two peptide ligands ($KRAS_{5-14}$ G12V and peptide #1). In the latter two cases, the middle portions of the peptides containing P4, P5, and P6 residues bulge out of the binding pockets to accommodate both peptide ends inside the peptide-binding groove of the HLA-I molecule (**Figures 7C,D,G,H**).

We next evaluated at the detailed interactions between HLA-A*02:01 and individual peptides. Throughout these interfaces, extensive hydrophobic and hydrogen bonding networking is seen with the majority of peptide residues participating in the contact with HLA-I residues Y7, F9, M45, E63, K66, V67, H70, T73, T80, L81, Y84, Y99, Y116, T143, K146, W147, V152, Y159, W167, Y171 (**Figures 7E–H**).

In the HLA-I-peptide #1 complex, the N-terminal P1 Y residue makes hydrophobic contact with T163, which is missing in all the other three peptide complexes and may explain the significantly higher binding affinity of 10 mer peptide #1 compared to 10 mer $KRAS_{5-14}$ G12V peptide (**Figure 7H**). The A and F pockets forming the peptide binding groove of HLA-I are mostly composed of hydrophobic residues and some 12 polar and 21 van der Waals contacts were, throughout the peptide length in all complexes, observed between the peptide moiety and HLA-A*02:01.

While the HLA-I interaction interface seems to be conserved in both spliced peptide $KRAS_{5-6/8-14}$ G12V and non-spliced peptide $KRAS_{5-14}$ G12V, the binding affinity of $KRAS_{5-6/8-14}$ G12V peptide toward HLA-A*02:01 is higher compared to the $KRAS_{5-14}$ G12V peptide (**Figure 6C**). Hence, to understand this differential affinity of these peptides for HLA-A*02:01, we compared the crystal structures of $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V peptides and modified variants (peptide #1 and #2) bound to HLA-A*02:01 complexes in more detail (**Figures 7I–K**). The superpositions of either $KRAS_{5-6/8-14}$ G12V peptide and peptide #2 or $KRAS_{5-14}$ G12V peptide and peptide #1 do not show any relevant differences (**Figures 7I,J**).

In contrast, although the structural superposition of $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V peptides bound to HLA-A*02:01 molecules reveals a similar type of HLA-I interaction network at their N-terminal and C-terminal regions, their structural arrangements deviate in their middle portions. Due to this, the spliced peptide $KRAS_{5-6/8-14}$ G12V makes several unique interactions with HLA-A*02:01. Firstly, in the structure of HLA-A*02:01 complexed with spliced peptide $KRAS_{5-6/8-14}$ G12V, the P6 A residue makes both hydrogen bonding and van der Waals contacts with the side chain of T73 residue of HLA-I, whereas the P6 G residue of $KRAS_{5-14}$ G12V is not in contact with HLA-A*02:01 and its P7 A residue maintains only hydrophobic interactions with T73.

Another difference between both these complexes is at their C-termini. In the HLA-I-$KRAS_{5-14}$ G12V peptide complex, the HLA-A*02:01 K146 residue adopts a different orientation, due to which it interacts with only the terminal PΩ residue. In the HLA-I-$KRAS_{5-6/8-14}$ G12V complex, the amino group of K146 forms a hydrogen bond with both the carbonyl oxygen of the PΩ-1 residue and the terminal PΩ residue. Furthermore, though $KRAS_{5-14}$ G12V is longer [10 amino acids, compared to $KRAS_{5-6/8-14}$ G12V (9 amino acids)] and reorganizes its central region in the HLA binding groove, this structural rearrangement does not favor any additional contacts with HLA-A*02:01. From our structural analysis, we can predict that the higher affinity of the spliced $KRAS_{5-6/8-14}$ G12V peptide, compared to $KRAS_{5-14}$ G12V peptide, might be due to these two additional hydrogen bonding contacts between the spliced peptides P6 A residue and the T73 of HLA-A*02:01, as well as the spliced peptides PΩ-1 residue and the K146 residue of the HLA-A*02:01 molecule (**Figure 7K**).

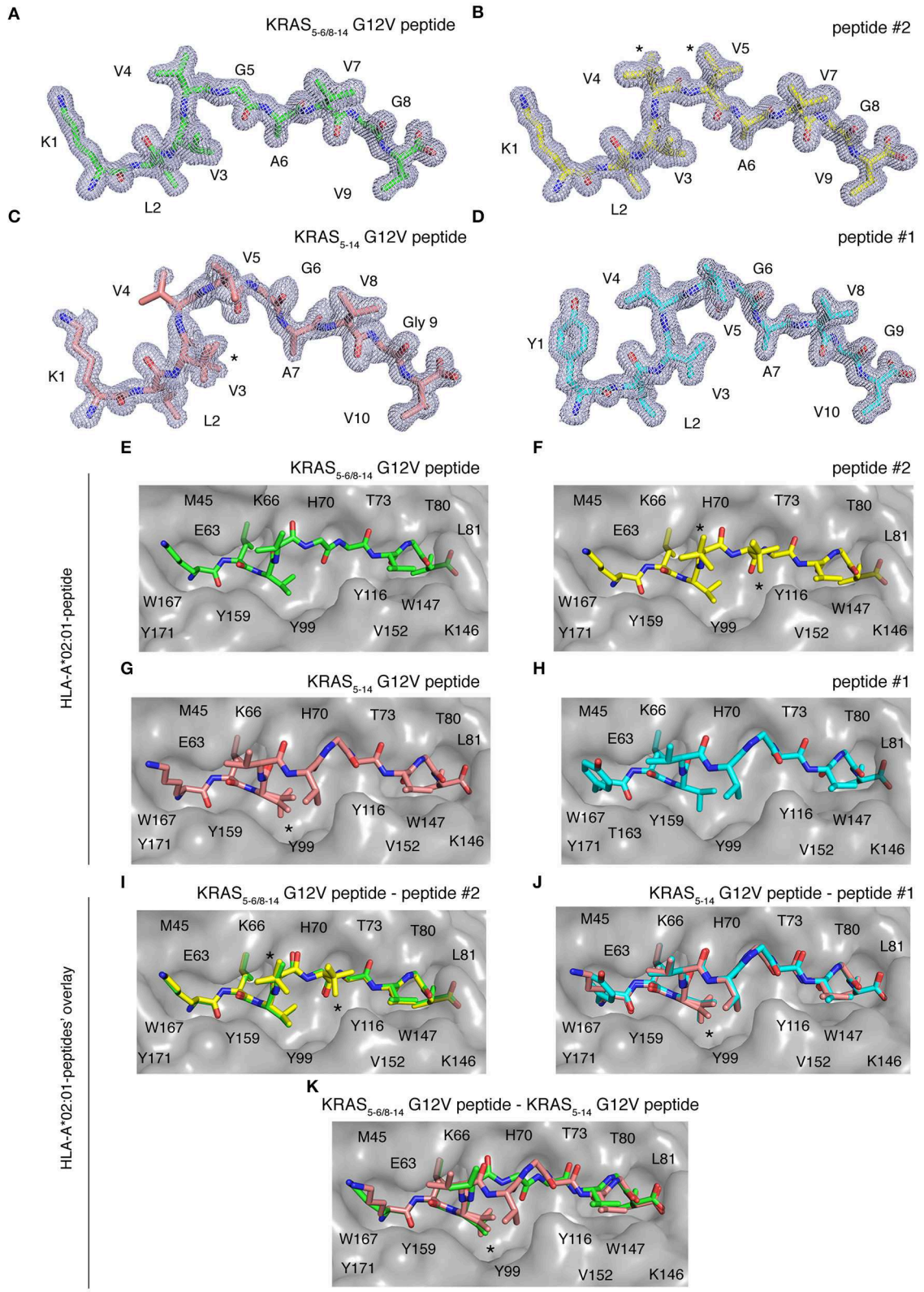


FIGURE 7 | HLA-A*02:01-KRAS G12V peptides binding mode. Binding mode of spliced epitope candidate $KRAS_{5-6/8-14}$ G12V [KL][VVGAVGV], non-spliced epitope candidate $KRAS_{5-14}$ G12V [KLVVVGAVGV], peptides #1 YLVVVGAVGV and #2 KLVVVAVGV to HLA-A*02:01 complex. **(A–D)** 2Fo-Fc electron density map contoured at 1σ for $KRAS_{5-6/8-14}$ G12V peptide **(A)**, peptide #2 **(B)**, non-spliced peptide $KRAS_{5-14}$ G12V **(C)**, and peptide #1 **(D)**. **(E–H)** Binding of spliced peptide

(Continued)

FIGURE 7 | $KRAS_{5-6/8-14}$ G12V (**E**; green sticks), peptide #2 (**F**; yellow sticks), non-spliced peptide $KRAS_{5-14}$ G12V (**G**; brown sticks), and peptide #1 (**H**; cyan sticks) to HLA-A*02:01 protein (gray molecular surface). **(I–K)** Overlay of $KRAS_{5-6/8-14}$ G12V and peptide #2 binding to HLA-A*02:01 **(I)**, $KRAS_{5-14}$ G12V peptide and peptide #1 **(J)**, as well as $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V peptides **(K)** binding to HLA-A*02:01 molecule. All peptides are shown as sticks and in **(A–D)**, peptide residues are labeled with one-letter amino acid codes. In **(E–K)**, the residues of HLA-A*02:01 that are extended in the peptide binding interface are labeled with single-letter amino acid codes. In **(B,F,I)**, * indicates the alternate conformations for residues V4 and V5 of peptide #2. In **(C,G,J–K)**, *indicates the alternate conformations for residues V3 of $KRAS_{5-14}$ G12V peptide.

Potential Recognition of $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V Epitope Candidates Within HLA-A*02:01 Cleft by Different TCRs

Once the peptide binds to a HLA-I molecule, it gets displayed for TCR recognition, which can then induce effective immune responses. Using structure as a tool, we tried to determine the mode of TCR-HLA-I-peptide interaction. Our evaluation of HLA-I-$KRAS_{5-6/8-14}$ G12V and HLA-I-$KRAS_{5-14}$ G12V peptide complexes provides a link to the potential cross recognition by a given $CD8^+$ T cell clone. The middle portion of both spliced and non-spliced epitope candidates containing P4 and P5 residues does not make ample contacts with the HLA-A*02:01 molecule but has limited flexibility in the crystal structure, otherwise this would not have been solved in the structure (**Figures 8A,B**). In the HLA-A*02:01-$KRAS_{5-14}$ G12V peptide complex, the middle portion that bulges out from the binding groove makes it more accessible for TCR recognition (**Figure 8B**). The side chain of P4 V and P5 V residues are facing in an upward direction and can be easily accommodated into the binding pocket located over the central peptide, formed by the most structurally diverse CDR loops, CDR3α and CDR3β of the TCR. Similarly, in the HLA-I-$KRAS_{5-6/8-14}$ G12V complex, the side chain of the P4 V residue is available to mediate hydrophobic contact for TCR recognition. Also, in both complexes there is a possibility of hydrogen bonding interactions between the main chain carbonyl and amide groups of P4 and P5 residues with the TCR (**Figures 8A–C**). Hence, our analysis of the crystal structure suggests that both the spliced $KRAS_{5-6/8-14}$ G12V and the non-spliced $KRAS_{5-14}$ G12V epitope candidates can be contacted by the same TCR at their P4 site, thereby promoting cross reactivity.

On the other hand, even though some potential TCR cross reactivity exists toward both spliced and non-spliced epitope candidates, the structural superposition of both peptide complexes revealed deviation in their peptide conformation at the region where the TCR interaction is expected to happen (**Figure 8C**). Hence, depending on the direction that the TCR encounters in the HLA-I-peptide complex, there might be a definite possibility of having TCRs that exhibit preference or exclusive binding toward either the non-spliced or the spliced epitope, rather than recognizing both of them. As the $KRAS_{5-6/8-14}$ G12V possess three peptide residues, P4 V, P5 V, and P6 G, that can mediate both hydrogen bonding and van der Waals contacts with a TCR, whilst the spliced peptide contains only P4 V and P5 G residues; hence, we can speculate that TCRs more likely will have selectivity and specificity for one of the two epitope candidates.

DISCUSSION

Epitope discovery is an essential first step for antigen-targeted immunotherapies against cancer, infection and some autoimmune diseases. In the last decade, several studies proposed strategies to achieve this, especially in light of anti-cancer immunotherapies. The majority of these studies identify epitope candidates in HLA-I immunopeptidomes eluted from cells. Although this strategy guarantees that the identified epitope is presented at the cell surface, it cannot include all targetable epitopes because of its relative low sensitivity (51). TCRs can still be considered more sensitive than MS-based methods and can sense even a few epitope molecules bound to HLA-I molecules to trigger cytotoxic responses. There are several examples of epitopes that were not identified by analytical methods based on HLA-I immunopeptidomes of cells, but were well-recognized by specific CTLs. The pipeline that we proposed here tries to circumvent this problem. While starting from a large number of theoretical epitope candidates, the pipeline narrows them down to a few selected candidates step by step. One of the advantages of our strategy is that its sequential steps could be exchanged and adapted to the specific requirements of a given application. For instance, in this study we developed a model to rank antigens by their potential over-representation in HLA-I immunopeptidomes considering four protein features, without including any cell-specific assays, such as transcriptome or intracellular proteome analysis. If such data was available, our pipeline could use more complex algorithms, such as that published by Pearson et al. (37), and likely reach a more in-depth antigen selection.

The same principle of flexible structure and interchangeable steps could be applied to the "*in vitro* selection" section of our pipeline. In this study, we tested *in vitro* three steps of the HLA-I APP pathway: proteasome-mediated generation; TAP-mediated transport into the ER lumen; and the efficient binding to the selected HLA-I variant. While some epitopes may be presented by HLA-I molecules in a proteasome- and TAP-independent fashion, the majority of HLA-I-restricted epitopes depends on these two steps.

Efficient binding to the selected HLA-I molecule is, on the contrary, mandatory. However, although a threshold of 500 nM would capture ∼85% of all HLA-I-bound peptides (52, 53), it is still an open question what the optimal IC_{50} threshold is to define a "good epitope target" for ATT. The most determining factor could be the off-rate of peptide binding, a feature that we likely determine indirectly via IC_{50}, because poor and good peptides have been reported to have similar on-rates but different off-rates (54).

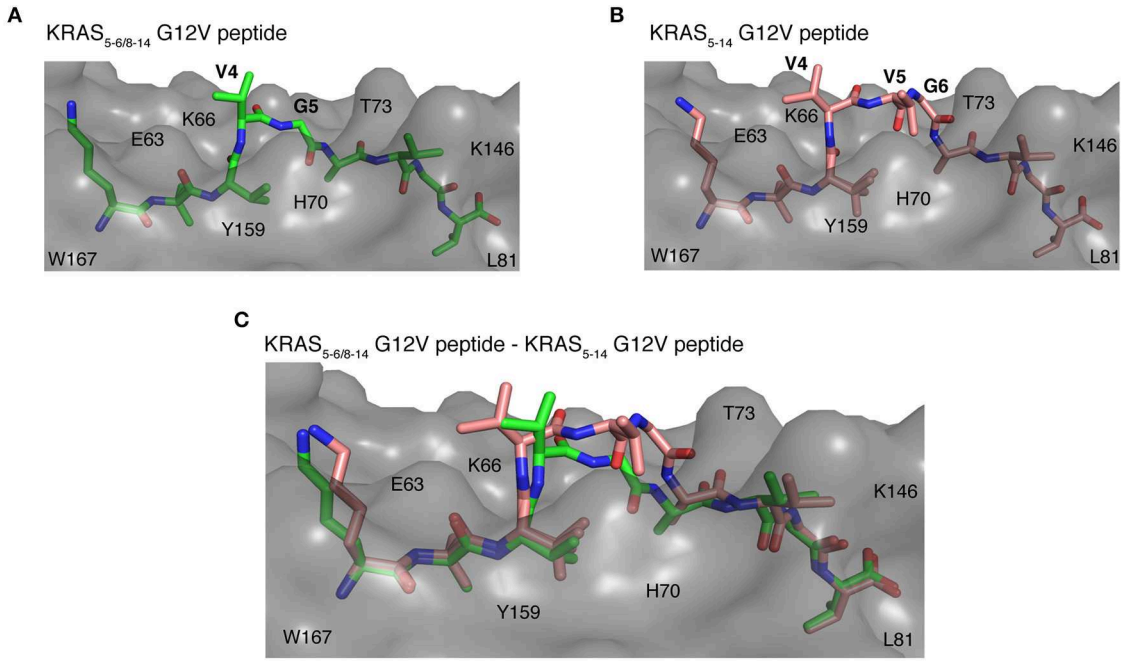


FIGURE 8 | Potential recognition of $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V epitope candidates within HLA-A*02:01 cleft by different TCRs. **(A,B)** Side view of binding mode of spliced epitope candidate $KRAS_{5-6/8-14}$ G12V [KL][VVGAVGV] (**A**; green sticks), non-spliced epitope candidate $KRAS_{5-14}$ G12V [KLVVVGAVGV] (**B**; brown sticks) into the HLA-A*02:01 binding grove displaying the peptide residues exposed for TCR recognition. **(C)** Structural superposition of $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V peptides complexed with HLA-A*02:01 molecule revealing the conformational deviation at their middle region where TCR interaction is expected to happen. In all three panels, the HLA-A*02:01 protein is shown as gray molecular surface and peptide residues as sticks. In **(A,B)**, peptide residues exposed for TCR recognition are labeled. In all panels, some of the residues of HLA-A*02:01 that are extended in the peptide binding interface are labeled.

Of course, the "*in vitro* selection" section of our pipeline could enlist other APP steps such as tapasin-dependency, cytosolic peptidase and ERAP trimming, etc. (2, 55), which could be selectively chosen based on tumor features and the known APP pathway of the target antigen.

Our pipeline identifies epitope candidates which shall further be validated by isolating specific CTL clones and their TCRs and using them to confirm that the epitope candidates are produced *in cellulo* and eventually *in vivo*. There are several strategies to this end. For example, the $KRAS_{5-6/8-14}$ spliced epitope candidate here identified has been validated in collaboration with Blankenstein et al. (56). Specific TCRs have been isolated from humanized AB*ab*DII mice (57), cloned into expressing vectors and transduced into human PBMCs. Transduced human $CD8^+$ T cells selectively recognized the $KRAS_{5-6/8-14}$ spliced epitope. They also recognized human cancer cell lines expressing KRAS G12V antigen and HLA-A*02:01 complex and release IFNγ. They do not recognize a cancer cell line expressing the wild type KRAS G12 protein and the HLA-A*02:01 complex. These outcomes validated the $KRAS_{5-6/8-14}$ spliced epitope candidate as a genuine epitope (56). We tried to identify the $KRAS_{5-6/8-14}$ G12V spliced and the $KRAS_{5-14}$ G12V non-spliced epitope candidates through the MS measurement of the HLA-I immunopeptidomes of the SW480 pancreatic adenocarcinoma cell line, which expresses the HLA-A*02:01 complex and the KRAS G12V mutated protein (42). Despite CTL clones could recognize both epitope candidates presented by cancer cell lines (42, 56), none of the two peptides was identified in the SW480-derived HLA-I immunopeptidomes (data not shown), thereby confirming the usefulness of the pipeline described here.

Another advantage of our pipeline is its ability to select and identify proteasome-generated spliced epitope candidates, which we and other groups have found to represent a sizeable pool of immunologically relevant epitopes, especially within the framework of anti-cancer immunotherapies (3, 11, 12, 20, 21, 44, 58). The PCPS reaction was shown to generate a large number and a significant amount of spliced peptides in the *in vitro* processing of $KRAS_{2-35}$ neoantigen by 20S proteasomes. This suggests that we might have previously underestimated PCPS frequency in the *in vitro* proteasome digestions, likely due to the low MS sensitivity available at that time (23).

The benefit of including these unconventional epitopes in our pipeline is evident. The recurrent KRAS G12V mutation can be efficiently presented by HLA-A*02:01 complexes only through spliced peptides. TCRs specific to this spliced epitope candidate could be used to treat around 15–20% of colorectal cancer and pancreatic adenocarcinoma by ATT. According to our analysis, the G12V mutation promotes not only the binding affinity of the $KRAS_{5-6/8-14}$ G12V spliced epitope candidate to HLA-A*02:01 complex, but also the splicing reaction at that site, since we did not identify the $KRAS_{5-6/8-14}$ G12 spliced peptide in the *in vitro* digestion of wild type $KRAS_{2-35}$

G12 by 20S proteasomes. The $KRAS_{5-6/8-14}$ G12V sequence [KL][VVGA**V**GV] cannot be generated by any other human protein by peptide hydrolysis or *cis* peptide splicing (data not shown), thereby defining it as a unique neoepitope. In the $KRAS_{2-35}$ G12V polypeptide digestion, we also identified the non-spliced epitope candidate $KRAS_{5-14}$ G12V, which was shown to be recognized by PBMCs of pancreatic adenocarcinoma patients (42). For a cancer-targeted strategy, it would be informative to perform *in vitro* digestions using proteasome isoforms recapitulating those present in the target cancer, since they vary from tumor to tumor with implications for the quantity (and perhaps the quality) of peptide produced (4, 5, 27). Both $KRAS_{5-6/8-14}$ G12V and $KRAS_{5-14}$ G12V epitope candidates are efficiently transported by TAP into the ER lumen, which is then no bottleneck. The non-spliced epitope candidate, however, binds HLA-A*02:01 less efficiently than the spliced epitope candidate with an IC_{50} larger than 300 nM, which could be higher than the binding affinity currently suggested for immunodominant epitopes. According to analysis of the HLA-A*02:01-peptide crystal structures, the two epitope candidates differ in the region exposed to TCRs. Therefore, we would expect that unique $CD8^+$ T cell clones could recognize these, although cross-reactive TCRs cannot be excluded. In the case of cross-reactivity, the immunodominance of the spliced epitope over the non-spliced epitope might be favored by the higher binding affinity to HLA-A*02:01 complex. We do not have enough information about the other steps of their APP pathways, including production by cancer-associated proteasome isoforms, to conclude about presentation in cancer patients. Yet, our pipeline allows identification of potential new neoepitopes derived from peptide splicing that are unique for a driver in oncogenesis, KRAS G12V. Such epitopes could be critical in new vaccination approaches for the related tumors.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are either included in the article/**Supplementary Material** or deposited in online repositories, as clarified in the Materials and Methods section.

AUTHOR CONTRIBUTIONS

MM and JL developed the whole project, performed and supervised the data analysis and data generation, and wrote the manuscript. DZ supervised the crystal structure analysis and wrote the cognate text. JS, JN, and AS supervised part of the analysis, wrote, and edited the text. AM, GY, AB, and RC performed and analyzed part of the experiments and proofread the manuscript. SH and DP carried out QPuB analysis and proofread the manuscript. HU supervised MS measurement and optimization.

FUNDING

This study has been in part supported by: (i) NIH to AS (R21Al134127), (ii) Cancer Research UK King's Health Partners Center at King's College London (Development Fund 2018), KCL-Monash Collaborative Award 2018 and MPI-BPC collaboration agreement 2018 to MM; the initial design of the study was conceptualized and the spliced epitope was identified by MM, while appointed at Charité – Universitätsmedizin Berlin with a contract financially supported by the BIH grant awarded to P.M. Kloetzel (BIH, CRG1-TP1), and by JL, whilst she was supported by NC3Rs through a David Sainsbury Fellowship (NC/K001949/1) at Imperial College London. SH and AM are supported by the International Max Planck Research School (IMPRS) for Genome Science.

ACKNOWLEDGMENTS

We thank P. Henklein, P. Kunert, B. Brecht-Jachan (Charité Berlin), and the Peptide Synthesis Facility of the Charité, as well as O. Jahn (Max Planck Institute for Experimental Medicine) for peptide synthesis. We thank W. Scott (KCL) for proofreading the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02572/full#supplementary-material

REFERENCES

1. Weissman AM, Shabek N, Ciechanover A. The predator becomes the prey: regulating the ubiquitin system by ubiquitylation and degradation. *Nat Rev Mol Cell Biol*. (2011) 12:605–20. doi: 10.1038/nrm3173
2. Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol*. (2016) 37:724–37. doi: 10.1016/j.it.2016.08.010
3. Mishto M, Liepe J. Post-Translational peptide splicing and T cell responses. *Trends Immunol*. (2017) 38:904–15. doi: 10.1016/j.it.2017.07.011
4. Guillaume B, Chapiro J, Stroobant V, Colau D, Van Holle B, Parvizi G, et al. Two abundant proteasome subtypes that uniquely process some antigens presented by HLA class I molecules. *Proc Natl Acad Sci USA*. (2010) 107:18599–604. doi: 10.1073/pnas.1009778107
5. Guillaume B, Stroobant V, Bousquet-Dubouch MP, Colau D, Chapiro J, Parmentier N, et al. Analysis of the processing of seven human tumor antigens by intermediate proteasomes. *J Immunol*. (2012) 189:3538–47. doi: 10.4049/jimmunol.1103213
6. Solomon H, Brauning B, Fainer I, Ben-Nissan G, Rabani S, Goldfinger N, et al. Post-translational regulation of p53 function through 20S proteasome-mediated cleavage. *Cell Death Differ*. (2017) 24:2187–98. doi: 10.1038/cdd.2017.139
7. Olshina MA, Ben-Nissan G, Sharon M. Functional regulation of proteins by 20S proteasome proteolytic processing. *Cell Cycle*. (2018) 17:393–4. doi: 10.1080/15384101.2017.1414682
8. Hanada K, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature*. (2004) 427:252–6. doi: 10.1038/nature02240
9. Vigneron N, Stroobant V, Chapiro J, Ooms A, Degiovanni G, Morel S, et al. An antigenic peptide produced by peptide splicing in the proteasome. *Science*. (2004) 304:587–90. doi: 10.1126/science.1095522

10. Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A, et al. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science*. (2016) 354:354–8. doi: 10.1126/science.aaf4384
11. Faridi P, Li C, Ramarathinam SH, Vivian JP, Illing PT, Mifsud NA, et al. A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands. *Sci Immunol*. (2018) 3:eaar3947. doi: 10.1126/sciimmunol.aar3947
12. Liepe J, Sidney J, Lorenz FKM, Sette A, Mishto M. Mapping the MHC class I-spliced immunopeptidome of cancer cells. *Cancer Immunol Res*. (2019) 7:62–76. doi: 10.1158/2326-6066.CIR-18-0424
13. Mylonas R, Beer I, Iseli C, Chong C, Pak HS, Gfeller D, et al. Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *Mol Cell Proteomics*. (2018) 17:2347–57. doi: 10.1074/mcp.RA118.000877
14. Rolfs Z, Solntsev SK, Shortreed MR, Frey BL, Smith LM. Global identification of post-translationally spliced peptides with neo-fusion. *J Proteome Res*. (2018) 18, 349–58. doi: 10.1021/acs.jproteome.8b00651
15. Faridi P, Li C, Ramarathinam SH, Illing PT, Mifsud NA, Ayala R, et al. Response to Comment on "A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands". *Sci Immunol*. (2019) 4:eaaw8457. doi: 10.1126/sciimmunol.aaw8457
16. Rolfs Z, Muller M, Shortreed MR, Smith LM, Bassani-Sternberg M. Comment on "A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands". *Sci Immunol*. (2019) 4:eaaw1622. doi: 10.1126/sciimmunol.aaw1622
17. Ebstein F, Textoris-Taube K, Keller C, Golnik R, Vigneron N, Van Den Eynde BJ, et al. Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. *Sci Rep*. (2016) 6:24032. doi: 10.1038/srep24032
18. Gonzalez-Duque S, Azoury ME, Colli ML, Afonso G, Turatsinze JV, Nigi L, et al. Conventional and neo-antigenic peptides presented by beta cells are targeted by circulating naive CD8+ T cells in type 1 diabetic and healthy donors. *Cell Metab*. (2018) 28:946–60.e6. doi: 10.1016/j.cmet.2018.07.007
19. Platteel ACM, Liepe J, Textoris-Taube K, Keller C, Henklein P, Schalkwijk HH, et al. Multi-level strategy for identifying proteasome-catalyzed spliced epitopes targeted by CD8+ T cells during bacterial infection. *Cell Rep*. (2017) 20:1242–53. doi: 10.1016/j.celrep.2017.07.026
20. Robbins PF, El-Gamil M, Kawakami Y, Stevens E, Yannelli JR, Rosenberg SA. Recognition of tyrosinase by tumor-infiltrating lymphocytes from a patient responding to immunotherapy. *Cancer Res*. (1994) 54:3124–6.
21. Dalet A, Robbins PF, Stroobant V, Vigneron N, Li YF, El-Gamil M, et al. An antigenic peptide produced by reverse splicing and double asparagine deamidation. *Proc Natl Acad Sci USA*. (2011) 108:E323–31. doi: 10.1073/pnas.1101892108
22. Dalet A, Vigneron N, Stroobant V, Hanada K, Van Den Eynde BJ. Splicing of distant Peptide fragments occurs in the proteasome by transpeptidation and produces the spliced antigenic peptide derived from fibroblast growth factor-5. *J Immunol*. (2010) 184:3016–24. doi: 10.4049/jimmunol.0901277
23. Mishto M, Goede A, Taube KT, Keller C, Janek K, Henklein P, et al. Driving forces of proteasome-catalyzed peptide splicing in yeast and humans. *Mol Cell Proteomics*. (2012) 11:1008–23. doi: 10.1074/mcp.M112.020164
24. Berkers CR, De Jong A, Schuurman KG, Linnemann C, Meiring HD, Janssen L, et al. Definition of proteasomal peptide splicing rules for high-efficiency spliced peptide presentation by MHC class I molecules. *J Immunol*. (2015) 195:4085–95. doi: 10.4049/jimmunol.1402455
25. Liepe J, Ovaa H, Mishto M. Why do proteases mess up with antigen presentation by re-shuffling antigen sequences? *Curr Opin Immunol*. (2018) 52:81–6. doi: 10.1016/j.coi.2018.04.016
26. Mishto M, Raza ML, De Biase D, Ravizza T, Vasuri F, Martucci M, et al. The immunoproteasome beta5i subunit is a key contributor to ictogenesis in a rat model of chronic epilepsy. *Brain Behav Immun*. (2015) 49:188–96. doi: 10.1016/j.bbi.2015.05.007
27. Mishto M, Liepe J, Textoris-Taube K, Keller C, Henklein P, Weberruss M, et al. Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation. *Eur J Immunol*. (2014) 44:3508–21. doi: 10.1002/eji.201444902
28. Liepe J, Mishto M, Textoris-Taube K, Janek K, Keller C, Henklein P, et al. The 20S proteasome splicing activity discovered by spliceMet. *PLOS Comput Biol*. (2010) 6:e1000830. doi: 10.1371/journal.pcbi.1000830
29. Neefjes JJ, Momburg F, Hammerling GJ. Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter. *Science*. (1993) 261:769–71. doi: 10.1126/science.8342042
30. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med*. (2016) 8:33. doi: 10.1186/s13073-016-0288-x
31. Remesh SG, Andreatta M, Ying G, Kaever T, Nielsen M, Mcmurtrey C, et al. Unconventional peptide presentation by major histocompatibility complex (MHC) class I Allele HLA-A*02:01: BREAKING CONFINEMENT. *J Biol Chem*. (2017) 292:5262–70. doi: 10.1074/jbc.M117.776542
32. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. (2010) 66:486–501. doi: 10.1107/S0907444910007493
33. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. (1997) 53:240–55. doi: 10.1107/S0907444996012255
34. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res*. (2019) 47:D442–50. doi: 10.1093/nar/gky1106
35. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics*. (2015) 14:658–73. doi: 10.1074/mcp.M114.042812
36. Mommen GP, Frese CK, Meiring HD, Van Gaans-Van Den Brink J, De Jong AP, Van Els CA, et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc Natl Acad Sci USA*. (2014) 111:4507–12. doi: 10.1073/pnas.1321458111
37. Pearson H, Daouda T, Granados DP, Durette C, Bonneil E, Courcelles M, et al. MHC class I-associated peptides derive from selective regions of the human genome. *J Clin Invest*. (2016) 126:4690–701. doi: 10.1172/JCI88590
38. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res*. (2012) 72:2457–67. doi: 10.1158/0008-5472.CAN-11-2612
39. Shin SM, Choi DK, Jung K, Bae J, Kim JS, Park SW, et al. Antibody targeting intracellular oncogenic Ras mutants exerts anti-tumour effects after systemic administration. *Nat Commun*. (2017) 8:15090. doi: 10.1038/ncomms15090
40. Wang QJ, Yu Z, Griffith K, Hanada K, Restifo NP, Yang JC. Identification of T-cell receptors targeting KRAS-mutated human tumors. *Cancer Immunol Res*. (2016) 4:204–14. doi: 10.1158/2326-6066.CIR-15-0188
41. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med*. (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
42. Kubuschok B, Neumann F, Breit R, Sester M, Schormann C, Wagner C, et al. Naturally occurring T-cell response against mutated p21 ras oncoprotein in pancreatic cancer. *Clin Cancer Res*. (2006) 12:1365–72. doi: 10.1158/1078-0432.CCR-05-1672
43. Chapiro J, Claverol S, Piette F, Ma W, Stroobant V, Guillaume B, et al. Destructive cleavage of antigenic peptides either by the immunoproteasome or by the standard proteasome results in differential antigen presentation. *J Immunol*. (2006) 176:1053–61. doi: 10.4049/jimmunol.176.2.1053
44. Warren EH, Vigneron NJ, Gavin MA, Coulie PG, Stroobant V, Dalet A, et al. An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science*. (2006) 313:1444–7. doi: 10.1126/science.1130660
45. Deol P, Zaiss DM, Monaco JJ, Sijts AJ. Rates of processing determine the immunogenicity of immunoproteasome-generated epitopes. *J Immunol*. (2007) 178:7557–62. doi: 10.4049/jimmunol.178.12.7557
46. Tenzer S, Wee E, Burgevin A, Stewart-Jones G, Friis L, Lamberth K, et al. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. *Nat Immunol*. (2009) 10:636–46. doi: 10.1038/ni.1728
47. Dalet A, Stroobant V, Vigneron N, Van Den Eynde BJ. Differences in the production of spliced antigenic peptides by the standard proteasome and the immunoproteasome. *Eur J Immunol*. (2011) 41:39–46. doi: 10.1002/eji.201040750
48. Zanker D, Waithman J, Yewdell JW, Chen W. Mixed proteasomes function to increase viral peptide diversity and broaden antiviral CD8+ T cell responses. *J Immunol*. (2013) 191:52–9. doi: 10.4049/jimmunol.1300802

49. Michaux A, Larrieu P, Stroobant V, Fonteneau JF, Jotereau F, Van Den Eynde BJ, et al. A spliced antigenic peptide comprising a single spliced amino acid is produced in the proteasome by reverse splicing of a longer peptide fragment followed by trimming. *J Immunol*. (2014) 192:1962–71. doi: 10.4049/jimmunol.1302032
50. Platteel AC, Mishto M, Textoris-Taube K, Keller C, Liepe J, Busch DH, et al. CD8 T cells of Listeria monocytogenes-infected mice recognize both linear and spliced proteasome products. *Eur J Immunol*. (2016) 44:3543–59. doi: 10.1002/eji.201545989
51. Loffler MW, Mohr C, Bichmann L, Freudenmann LK, Walzer M, Schroeder CM, et al. Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med*. (2019) 11:28. doi: 10.1186/s13073-019-0636-8
52. Sette A, Vitiello A, Reherman B, Fowler P, Nayersina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol*. (1994) 153:5586–92.
53. Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J Immunol*. (2013) 191:5831–9. doi: 10.4049/jimmunol.1302101
54. Garstka MA, Fish A, Celie PH, Joosten RP, Janssen GM, Berlin I, et al. The first step of peptide selection in antigen presentation by MHC class I molecules. *Proc Natl Acad Sci USA*. (2015) 112:1505–10. doi: 10.1073/pnas.1416543112
55. Boulanger DSM, Eccleston RC, Phillips A, Coveney PV, Elliott T, Dalchau N. A mechanistic model for predicting cell surface presentation of competing peptides by MHC class I molecules. *Front Immunol*. (2018) 9:1538. doi: 10.3389/fimmu.2018.01538
56. Blankenstein T, Willimsky G, Liepe J, Kloetzel PM, Mishto M, Beier C. T cell receptors for tumor specific proteasome splice variants and uses thereof. *European Patent* (2019). PCT/EP2019/050027. Pub. No. WO/2019/129892
57. Li LP, Lampert JC, Chen X, Leitao C, Popovic J, Muller W, et al. Transgenic mice with a diverse human T cell antigen receptor repertoire. *Nat Med*. (2010) 16:1029–34. doi: 10.1038/nm.2197
58. Bonnet D, Warren EH, Greenberg PD, Dick JE, Riddell SR. CD8(+) minor histocompatibility antigen-specific cytotoxic T lymphocyte clones eliminate human acute myeloid leukemia stem cells. *Proc Natl Acad Sci USA*. (1999) 96:8639–44. doi: 10.1073/pnas.96.15.8639

Conflict of Interest: MM and JL are co-inventors of the spliced epitope and specific TCRs protected by the patent PCT/EP2019/050027.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 29 November 2019
doi: 10.3389/fimmu.2019.02799

Neoantigen Fitness Model Predicts Lower Immune Recognition of Cutaneous Squamous Cell Carcinomas Than Actinic Keratoses

*Elizabeth S. Borden[1,2], Paul Kang[3], Heini M. Natri[2], Tanya N. Phung[2], Melissa A. Wilson[2], Kenneth H. Buetow[2] and Karen Taraszka Hastings[1]**

[1] Department of Basic Medical Sciences, College of Medicine Phoenix, University of Arizona, Phoenix, AZ, United States, [2] School of Life Sciences, Arizona State University, Tempe, AZ, United States, [3] Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Phoenix, AZ, United States

OPEN ACCESS

Edited by:
Nikolaos G. Sgourakis, University of California, Santa Cruz, United States

Reviewed by:
Arya Biragyn, National Institute on Aging (NIA), United States
Nahum Puebla-Osorio, University of Texas MD Anderson Cancer Center, United States

***Correspondence:**
Karen Taraszka Hastings
khasting@email.arizona.edu

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: 01 August 2019
Accepted: 14 November 2019
Published: 29 November 2019

Citation:
Borden ES, Kang P, Natri HM, Phung TN, Wilson MA, Buetow KH and Hastings KT (2019) Neoantigen Fitness Model Predicts Lower Immune Recognition of Cutaneous Squamous Cell Carcinomas Than Actinic Keratoses. Front. Immunol. 10:2799. doi: 10.3389/fimmu.2019.02799

A low percentage of actinic keratoses progress to develop into cutaneous squamous cell carcinoma. The immune mechanisms that successfully control or eliminate the majority of actinic keratoses and the mechanisms of immune escape by invasive squamous cell carcinoma are not well-understood. Here, we took a systematic approach to evaluate the neoantigens present in actinic keratosis and cutaneous squamous cell carcinoma specimens. We compared the number of mutations, the number of neoantigens predicted to bind MHC class I, and the number of neoantigens that are predicted to bind MHC class I and be recognized by a T cell receptor in actinic keratoses and cutaneous squamous cell carcinomas. We also considered the relative binding strengths to both MHC class I and the T cell receptor in a fitness cost model that allows for a comparison of the immune recognition potential of the neoantigens in actinic keratosis and cutaneous squamous cell carcinoma samples. The fitness cost was subsequently adjusted by the expression rates of the neoantigens to examine the role of neoantigen expression in tumor immune evasion. Our analyses indicate that, while the number of mutations and neoantigens are not significantly different between actinic keratoses and cutaneous squamous cell carcinomas, the predicted immune recognition of the neoantigen with the highest expression-adjusted fitness cost is lower for cutaneous squamous cell carcinomas compared with actinic keratoses. These findings suggest a role for the down-regulation of expression of highly immunogenic neoantigens in the immune escape of cutaneous squamous cell carcinomas. Furthermore, these findings highlight the importance of incorporating additional factors, such as the quality and expression of the neoantigens, rather than focusing solely on tumor mutational burden, in assessing immune recognition potential.

Keywords: neoantigen, actinic keratosis, cutaneous squamous cell carcinoma, MHC class I, T cell receptor, cancer, immunoediting

INTRODUCTION

In the Medicare fee-for-service population, there were over one million cutaneous squamous cell carcinomas (cuSCCs) diagnosed in 2012 in the United States, and the incidence is increasing (1). While most cuSCC tumors are successfully treated with excision, ∼4% of cuSCC patients develop metastases and 2% die from cuSCC (2). Especially in sun-rich states, the high incidence of cuSCC, coupled with the risk of metastasis and death, results in similar estimates of death from cuSCC as melanoma (3). Immunotherapy using immune checkpoint inhibition with the drug cemiplimab has recently been FDA-approved for the treatment of locally advanced, unresectable, and metastatic cuSCC; however, only ∼50% of cuSCC patients respond to cemiplimab treatment (4). Therefore, it is critical to improve our understanding of the immune recognition of cuSCC, in order to advance the prevention and treatment options for this disease.

Actinic keratoses (AKs) are generally considered to result from cumulative ultra-violet light-induced DNA mutations, and a small percentage of these precursor lesions progress to invasive cuSCC over time (5). However, despite ongoing research, there is not yet a clear understanding of what allows some AKs to progress to cuSCC. Lesion progression involves mutations in the epithelial cells which allow malignant transformation. In addition, mutations in the tumor cells generate neoantigens which may be recognized by the naturally-occurring or therapeutically-induced immune response, and thus the immune response modulates tumor development. The interaction between cancer and the immune system is explained by the cancer immunoediting model, which has three phases (6). In the elimination phase, the immune system destroys the developing tumor before the tumor becomes clinically apparent. The elimination phase can result in complete elimination or residual cancer cell variants that resist elimination and enter the equilibrium phase. In the equilibrium phase, the immune response controls tumor growth. Editing of the immunogenicity of the tumor occurs in the equilibrium phase as a consequence of selective pressure from the immune response. Tumor cell variants that are no longer recognized by the immune system enter the escape phase and manifest as clinically apparent tumors. Examples of immune escape include the loss of tumor antigens or the loss of the ability to present the tumor antigens on MHC class I. The immunoediting process has recently been demonstrated in early stage, untreated non-small cell lung cancer (7). Tumors with intact MHC class I had a significant decrease in expressed neoantigens compared with non-neoantigenic, somatic mutations, and only tumors with intact MHC class I and immune cell infiltration exhibited a decrease in expressed neoantigens (7).

Cancer immunoediting is also observed in response to therapeutically-induced immune responses from immunotherapy with immune checkpoint inhibitors. Immune checkpoint receptors are expressed on T cells after activation and function as part of a homeostatic mechanism to turn off T cell responses (8). While immune checkpoint receptor-ligand interactions are helpful in constricting T cell responses after an infection is cleared, some tumors co-opt this mechanism to avoid eliciting a T cell response. Immune checkpoint inhibition therapy blocks this inhibitory signal and improves immune-mediated tumor destruction. Several studies have demonstrated an association of higher tumor mutational burden with improved response to immune checkpoint inhibition in melanoma and non-small cell lung cancer (9–13). A subset of these studies also analyzed the number of neoantigens and found that the number of neoantigens was also positively correlated with the response to immunotherapy (9, 11–13). However, these results have not been consistent across all cancers. Multiple myeloma patients were found to have an inverse relationship between progression-free survival and the tumor mutational burden and neoantigens (14). These opposing results suggest that there are likely additional factors influencing the relationship between neoantigens and the response to immunotherapy.

One possible factor suggested by McGranahan et al. is the homogeneity of the tumor neoantigens. Their group found that in lung cancers, in addition to the number of neoantigens, the degree of homogeneity of the tumor was highly associated with the survival of the patient (9). They found that the more homogenous the neoantigen presence was in the tumor, the greater the patient survival (9). Another factor is defined by Łuksza et al. as the "fitness cost of the tumor," which is directly proportional to MHC class I binding and T-cell receptor (TCR) recognition potential (15). An increased fitness cost was found to highly correlate with improved response to immunotherapies, suggesting that, in addition to the number of mutations and neoantigens, the strength of these neoantigens is important in predicting the immune response (15). This finding is corroborated by Anagnostou et al. who demonstrated the changes in the neoantigens found in tumors before and after immunotherapy (16). In this study, the overall number of neoantigens increased from the original tumor to the post-treatment tumor; however, the strength of retained and gained neoantigens was less than the strength of the neoantigens that were lost (16). These results again suggest that the ability of a tumor to grow in the face of a competent immune system requires changes not only in the number of neoantigens but in their quality.

There have been a few analyses of the differences in tumor mutational burden between pre-cancerous and cancerous lesions (17–19). Two of these analyses compared the tumor mutational burden between Barrett's esophagus and esophageal adenocarcinomas and found that the tumor mutational burden increased from the pre-cancerous to the cancerous lesion in paired samples (17, 18). A single study investigated the tumor mutational burden in AKs and cuSCCs (19), but, to our knowledge, there are no reports of the number of neoantigens or the fitness costs of these neoantigens in AK or cuSCC. In order to address these gaps, we have compared the neoantigen burden and fitness cost between AKs and cuSCCs. Understanding what allows the immune system to keep AKs in equilibrium while it fails to do so with cuSCCs may help to explain which AKs progress to cuSCCs.

TABLE 1 | Summary of the samples available from each patient.

Patient ID	1	2	3	4	5	6	8	10	12
WES normal skin	X	X	X	X	X	X	X	X	
WES saliva	x			x	x	x	x	x	X
WES AK	X	X	X	X	X			X	X
WES cuSCC		X	X	X, X	X	X	X		
RNAseq normal skin		x	x	x	x	x	x	x	
RNAseq AK	x	X	X	x	X	x	x	x	x
RNAseq cuSCC		X	X	x, x	X	X	x	x	x

X denotes any samples used in our analyses, and x denotes samples available that were not used.

METHODS

Datasets

Whole exome sequencing (WES) data with an average coverage of 135 × ± 22 (mean ± s.d.) (Illumina Hi-Seq) and RNAseq data average of 64 million reads per sample (Illumina Hi-Seq) from Chitsazzadeh et al. was used for this analysis (19). Datasets from 7 AK samples and 7 cuSCC samples from 9 different patients were kindly provided by Dr. Ken Tsai and Dr. Kimal Pajapakshe. Eight out of the 9 patients also had a normal skin sample which was used for comparison. For the remaining patient (patient 12), a saliva sample was used in place of the normal skin. These samples and the patients they come from are summarized in **Table 1**. Note that the AK and cuSCC samples from an individual patient are separate lesions.

Neoepitope Prediction

WES and RNAseq data files for normal skin, AK, and cuSCC samples were obtained as FASTQ files. WES and RNAseq FASTQ files underwent quality control using FastQC (20) and were trimmed for quality using Trimmomatic (21) IlluminaClip with the following parameters: seed_mismatches = 2, palindrome_clip_threshold = 30, simple_clip_threshold = 10, leading = 10, trailing = 10, winsize = 4, winqual = 15. Trimmed WES reads were mapped to the GRCh38.p7 reference genome using HISAT2 v. 2.1.0 (22). SAM files were converted to BAM and coordinate sorted using SAMtools v. 1.4 (23). WES BAM files were processed according to Broad Institute GATK (Genome Analysis Toolkit) best practices (24–26). Read groups were added with Picard Toolkit's AddOrReplaceReadGroups and optical duplicates marked with Picard Toolkit's MarkDuplicates (v.2.18.1, http://broadinstitute.github.io/picard/). Base quality scores were recalibrated with GATK (v.4.0.3.0) BaseRecalibrator. The BAM files were then converted to pileup format using Samtools 1.3.1 (23). To determine the neoantigens present in the AKs compared to the cuSCCs, first, high confidence single nucleotide polymorphisms (SNPs) and indels were called using VarScan2 version 2.3.9 with minimum coverage of 10, minimum variant allele frequency of 0.08, and somatic *p*-value of 0.05 (27). Somatic mutations were separated from those SNPs that fell within 1 bp of an indel position, as these were likely to be false positives due to alignment errors. The variants were annotated using the Variant Effect Predictor tool from Ensembl version 90.9 (28, 29). Finally, peptides were identified and prioritized from these variants using pVACtools version 3.0.5 (30, 31). For every variation in amino acid, all possible peptides were generated in which the changed amino acid was included at every position in a sequence. Sequences of 21 amino acids were considered. The non-mutated sequence corresponding to each of these possible neoepitopes was also extracted for comparison sake. These steps were done as outlined in the EpitopeHunter pipeline from Narang et al. (32).

HLA Typing

HLA typing was completed for three major MHC class I genes (HLA-A, -B, and -C) using POLYSOLVER (POLYmorphic loci reSOLVER) version 1.0 (33). POLYSOLVER aligns reads from the WES data in the HLA region of each sample and then aligns these regions to a library of all known HLA alleles. Then, a Bayesian classification approach is used to determine the two alleles for each gene for each patient.

Predicting MHC Class I Binding Epitopes

Prediction of the potential epitopes that would effectively bind to MHC class I was completed with the NetMHCpan server version 4.0 (34). Neoantigens of only 9 amino acids were considered. Scores were calculated for both the mutated peptides and their wild-type counterparts. Each was scored based on its dissociation constant to each of the alleles predicted by POLYSOLVER. The top binding potential for each neoantigen was selected independently for the wild-type and mutant peptides. These binding probabilities were then used to determine an "amplitude" (A) using the methods outlined by Łuksza et al. (15). Only neoantigens with a maximum dissociation constant of 500 nM were considered, and the ratio of the dissociation constants (K_d) for the wild-type (WT) compared to the mutant (MT) peptides was calculated as shown here:

$$A = K_d^{WT} / K_d^{MT}$$

For wild-type peptides with exceptionally high predicted dissociation constants, an adjustment is made to account for the lack of accuracy of netMHCpan at predicting dissociation constants above the range in which it was trained. This adjustment is described in detail in the paper of Łuksza et al. and will therefore not be discussed further here (15).

Predicting Neoantigen TCR Recognition

Prediction of TCR recognition potential, R, was calculated as described by Łuksza et al. (15). A BLOSUM62 similarity matrix was used to assess the sequence similarity between a neoantigen and peptide sequences that are known T cell antigens from the Immune Epitope Database (IEDB) (35). The same set of IEDB sequences was used as was optimized by Łuksza et al. for evaluation of neoantigens in melanoma and small cell lung cancer (15). The sequence similarity was then used in place of binding energies, and the TCR recognition potential was calculated as:

$$R = Z(k)^{-1} \sum_{e \in IEDB} \exp[-k(a - |s, e|)]$$

Where |s,e| is the sequence similarity, a is the horizontal displacement of the binding curve, and k sets the steepness of the curve at a. Based on the model fit by Łuksza et al., the parameters a and k were set to 26 and 4.87, respectively (15). These parameters were optimized for both melanoma and lung cancer patients and shown to have consistent predictive value in both cancers. This gives us confidence that they can likewise be used with predictive value for cuSCC without the need to introduce a new training set. Finally, Z(k) is the partition function over the unbound state and all bound states, calculated as follows:

$$Z(k) = 1 + \sum_{e \in IEDB} \exp[-k(a - |s, e|)]$$

Fitness Cost Calculation

Given the TCR recognition potential (R) and the MHC-binding amplitude (A), the fitness cost (F) of the neoantigens was calculated as described by Łuksza et al. (15):

$$F = A \times R$$

Rather than taking the negative of this value as was done by Łuksza et al., we have left this value as a positive and will refer to it as the fitness cost. The greater the fitness cost, the more immune recognition and immune-mediated destruction we predict for the tumor. We have chosen to look at both the average and the maximum fitness cost for the neoantigens in each tumor to avoid presupposing that the single strongest neoantigen is of sole importance in the tumor recognizability.

Recognizing that to be visible to the immune system, neoantigens must be expressed, the fitness cost was then adjusted for the magnitude of expression of the given neoantigen. Transcriptome assembly and read count quantifications were completed with Salmon version 0.11.3 (36) normalized to reads per kilobase of transcript per million mapped reads (RPKM). Using RNAseq data, the expression of each neoantigen with non-zero fitness cost was calculated as a fraction of the total read counts for these neoantigens. This fraction was multiplied by the amplitude and TCR recognition potential to generate an adjusted fitness cost.

Statistical Analysis

Statistical Analyses were performed using STATA version 14 (STATAcorp; College Station, TX). *P*-value calculations were done using a Wilcoxon Rank Sum calculation. Correlation between variables were calculated as Spearman's Rho correlation coefficients.

HLA Mutations and Expression

To determine the role of HLA mutations and expression on the neoantigen profile of the tumor samples, somatic mutations were identified within the coding regions for HLA-A, B, and C. The mutations identified were then compared to the regions encoding the peptide binding groove. The regions encoding the peptide binding groove are found on exons 2 and 3. HLA coding regions and peptide binding groove regions were determined by the NCBI Gene database (37). MHC class I pathway members, TAP1, TAP2, or B2M were also interrogated for mutations. RPKM-normalized RNAseq expression levels of HLA-A, B, and C as well as TAP1, TAP2, and B2M were fit with a linear regression against the adjusted maximum fitness cost. The mutation's effect on the expression of the gene was evaluated by calculating the average expression and the standard deviation of the specific gene across all samples. For the mutations found in HLA-B for patient 2 cuSCC and in HLA-C for patient 5 AK, the expression value was compared to the average expression of the respective HLA gene. A value within two standard deviations from the mean was considered unaffected by the mutation.

Cell Enrichment Type Analysis

We used the xCell web service to deconvolute the diverse cell populations present in the cuSCC and AK samples (38). This program uses a gene signature-based method to estimate the proportions of different cell types in a bulk-RNA-sequence sample. RPKM gene expression data was used to estimate the fraction of the tumor sample made up by different cell types. In this work, the specific focus was on the immune-related cell types. Linear regressions were fit to the adjusted maximum fitness cost against the expression level of each of the following: T cell subsets (CD4 T cells, CD8 T cells, CD8 central memory T cells, CD8 effector memory T cells, CD8 naïve T cells, regulatory T cells), dendritic cell populations (immature DCs and conventional DCs), and natural killer (NK) cells. A Bonferroni correction was completed on the *p*-values from these correlations to adjust for the number of comparisons.

RESULTS

Analysis of Mutation and Neoantigen Counts

Somatic mutations and neoantigens were identified and filtered using the methods described. As previously shown by Chitsazzadeh et al. the mutational burden varied widely over both the AK and cuSCC samples (19). For cuSCCs there was an average of 2,861 mutations with a range of 389–11,504 and for AKs there was an average of 424 mutations with a range of 346–1,697 (**Table 2**, **Figure 1A**). While the maximum number of mutations was higher for the cuSCC samples than the AK samples, there was no statistically significant difference in the average number of mutations ($p = 0.3379$).

The number of binding neoantigens was defined as those neoantigens predicted by NetMHCpan to have a dissociation constant for MHC class I of <500 nM. AK and cuSCC samples both had large ranges of binding neoantigens (88–442 and 45–3,600, respectively), but there was no statistically significant difference in the average number observed for the two populations (**Table 2**, **Figure 1B**, $p = 0.336$). These data show that neoantigens predicted to bind MHC class I are present in both AK and cuSCC samples.

The number of neoantigens can be further refined by selecting only those with a non-zero predicted TCR recognition potential. We will refer to these as "immunogenic neoantigens." Again, a large range was observed (21–112 for AKs, 8–965 for cuSCCs), but no statistically significant difference was present (**Table 2**,

TABLE 2 | Summary of the number of somatic mutations, 21 base pair peptides (predicted from only non-synonymous mutations), 9 base pair binding neoantigens, and immunogenic neoantigens for each patient sample.

Sample	Somatic mutations	Peptides (21 mer)	Possible neoantigens (9 mer)	Binding neoantigens	Immunogenic neoantigens
Patient 1 AK	50	2	26	0	0
Patient 2 AK	1,697	1,191	15,480	442	112
Patient 2 cuSCC	2,540	2,185	28,403	190	41
Patient 3 AK	617	317	4,121	85	21
Patient 3 cuSCC	5,385	2,941	38,231	1,297	339
Patient 4 AK	75	2	27	2	0
Patient 4 cuSCC1	83	12	156	2	1
Patient 4 cuSCC2	47	1	13	0	0
Patient 5 AK	346	197	2,561	88	28
Patient 5 cuSCC	11,504	7,678	99,799	3,600	965
Patient 6 cuSCC	389	245	3,185	45	8
Patient 8 cuSCC	82	12	157	7	0
Patient 10 AK	133	24	312	2	0
Patient 12 AK	53	0	0	0	0

Binding neoantigens are those neoantigens with <500 nM dissociation constant from MHC class I, and immunogenic neoantigens are those neoantigens with a non-zero TCR-binding potential.

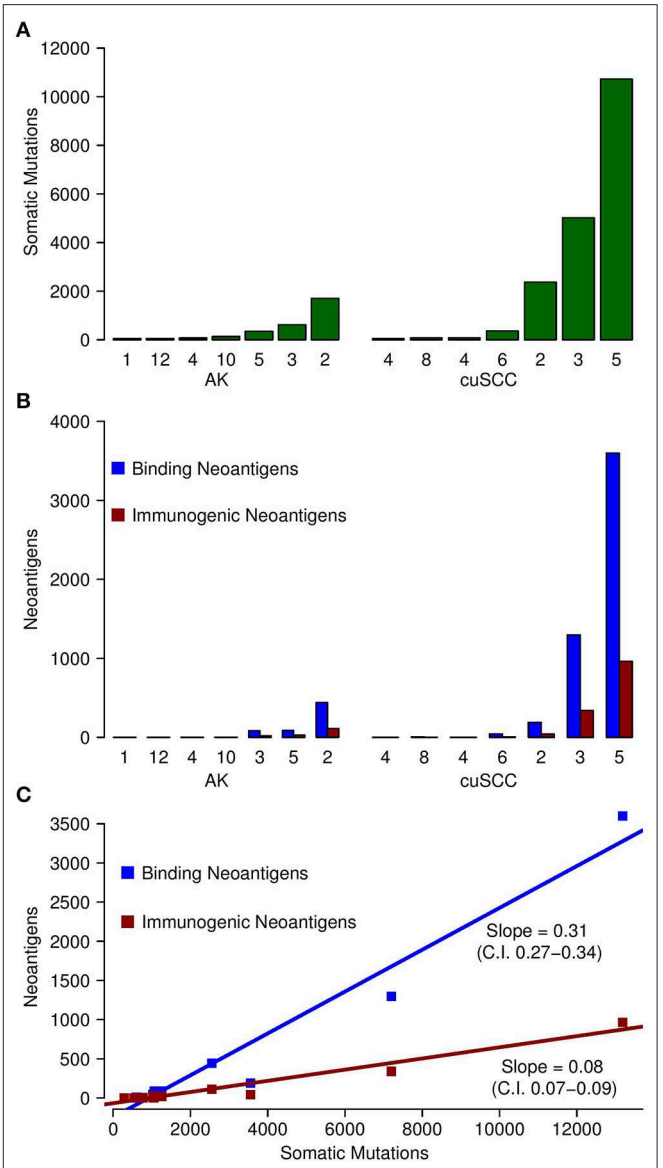


FIGURE 1 | (A) Comparison of the number of somatic mutations in each of the samples for AK and cuSCC. **(B)** In blue, comparison of the number of neoantigens predicted to have an MHC class I dissociation constant of <500 nM (termed Binding Neoantigens) in AK and cuSCC. In red, comparison of the number of neoantigens predicted to have an MHC class I dissociation constant of <500 nM and a non-zero TCR recognition potential (termed Immunogenic Neoantigens) in AK and cuSCC. **(C)** Correlation between the number of somatic mutations and the number of binding neoantigens (blue) and immunogenic neoantigens (red). Each data point represents an AK or cuSCC sample. Numbers for AK and cuSCC indicate the patient sample from **Table 1** used for the analyses. Note that the AK and cuSCC samples from an individual patient are separate lesions.

Figure 1B, $p = 0.3185$). These data show that the number of neoantigens can be further refined by incorporating predicted T cell recognition to yield a smaller subset of neoantigens, which are predicted to be immunogenic.

As shown in **Figure 1C**, there was a strong correlation between the somatic mutations and both the binding and immunogenic neoantigens (Spearman's rho correlation coefficients of 96 and 92%, respectively). However, these correlations represent different associations. The slope of the somatic mutation to binding neoantigen correlation was 0.31 (95% CI of 0.27–0.34), and the slope of the somatic mutation to immunogenic neoantigen correlation was 0.08 (95% CI of 0.07–0.09). Given that there was no overlap for the 95% confidence intervals, we can conclude that the difference in these slopes was statistically significant. While both the number of binding and immunogenic neoantigens increase with an increase in the number of somatic mutations, the increase in the number of immunogenic neoantigens is less than the increase in the number of binding neoantigens. These data shown that the immunogenic neoantigens are a much smaller subset than those with predicted MHC binding capacity.

Analysis of Neoantigen Fitness Cost

To continue our comparison of the neoantigens present in AKs compared to those in cuSCCs, we analyzed only those

samples with more than one immunogenic neoantigen (AKs from patients 2, 3, and 5 and cuSCCs from patients 2, 3, 5, and 6) as this enabled us to look at the comparative strengths of the neoantigens in each population. Note that patient 4 had one immunogenic neoantigen in one of the cuSCC samples. However, this patient was not included in the subsequent analyses as the TCR binding and fitness cost of this neoantigen were 1.77×10^{-4} and 2.36×10^{-5}, respectively. These values were 2 orders of magnitude smaller than the averages for all the other samples included, and thus, this sample was excluded.

To examine the percent of neoantigens recognized by the immune system in the AK and cuSCC samples comparatively, we took the ratio of the number of immunogenic neoantigens to the number of binding neoantigens. The ratios were similar for both AKs and cuSCCs (averages of 0.273 and 0.231, respectively, **Figure 2A**). However, if a more stringent criteria for the binding of the TCR was applied where the binding to the TCR must be $\geq$0.01, the average for AKs was 0.0290 and the average for the cuSCCs was 0.0147, which was a statistically significant difference ($p = 0.033$, **Figure 2B**). Since a greater percentage of the mutations in AKs are theoretically visible to the immune system, there may be greater immune recognition of these lesions despite the lack of difference in their mutational rates.

Next, we considered the relative MHC class I binding potential for the neoantigens in AKs compared to cuSCCs. The AK samples had a slightly higher average MHC class I dissociation constant (195 nM) than did the cuSCCs (179 nM), which was statistically significant (**Figure 3A**, $p = 0.0497$). This indicates that, on average, the neoantigens from AKs were predicted to bind MHC class I with lower affinity than those from cuSCC. Considering this factor alone, it would seem to suggest that the AKs should be less visible to the immune system than the cuSCCs. However, as T cells that recognize self-antigens are eliminated during T cell development, Łuksza et al. propose that if the binding of the wild-type peptide is high, it is likely that the T cell that would recognize this antigen has already been deleted (15). To adjust for this, the amplitude was calculated as the ratio of the dissociation constant for the wild-type peptide:MHC to the dissociation constant for the mutant peptide:MHC. As explained by Łuksza et al., the amplitude reflects the relative probability that a neoantigen will be bound to MHC class I times the relative probability that the corresponding wild-type peptide will not be bound (15). In all cases except AK #5, the amplitude is <1.0, indicating that on average the wild-type peptides were predicted to bind MHC class I with greater affinity that the mutated peptides or neoantigens. The average amplitudes for AK and cuSCC were similar (**Figure 3B**, $p = 0.4795$), suggesting that any difference in the MHC binding potential of the two groups is eliminated when the relative binding of the wild-type peptide is accounted for. Finally, the average TCR recognition potential was 0.121 for AK and 0.078 for cuSCC (**Figure 3C**, $p = 0.0771$). While the trend is toward a higher TCR recognition potential for AK than cuSCC, at our sample size this variable is unable to definitively differentiate the two groups.

The fitness cost of the neoantigens was defined as the amplitude times the TCR recognition potential. The fitness cost predicts the immune recognition potential of the neoantigen, such that an increased fitness cost is predicted to have greater immune-mediated destruction. The average fitness cost for AK was 0.109, while the average for the cuSCC was 0.048; this difference was not statistically significant ($p = 0.288$, **Figure 4A**). However, given that T cell-mediated immune responses are mounted to one or a few "immunodominant" antigenic peptides, another important factor to consider is the neoantigen with the highest fitness cost. The maximum fitness cost had a very large range for the cuSCCs (0.604–15.309) compared to a smaller range (2.791–6.941) for AKs (**Figure 4B**). Despite the lack of statistically significant difference between the maximum fitness

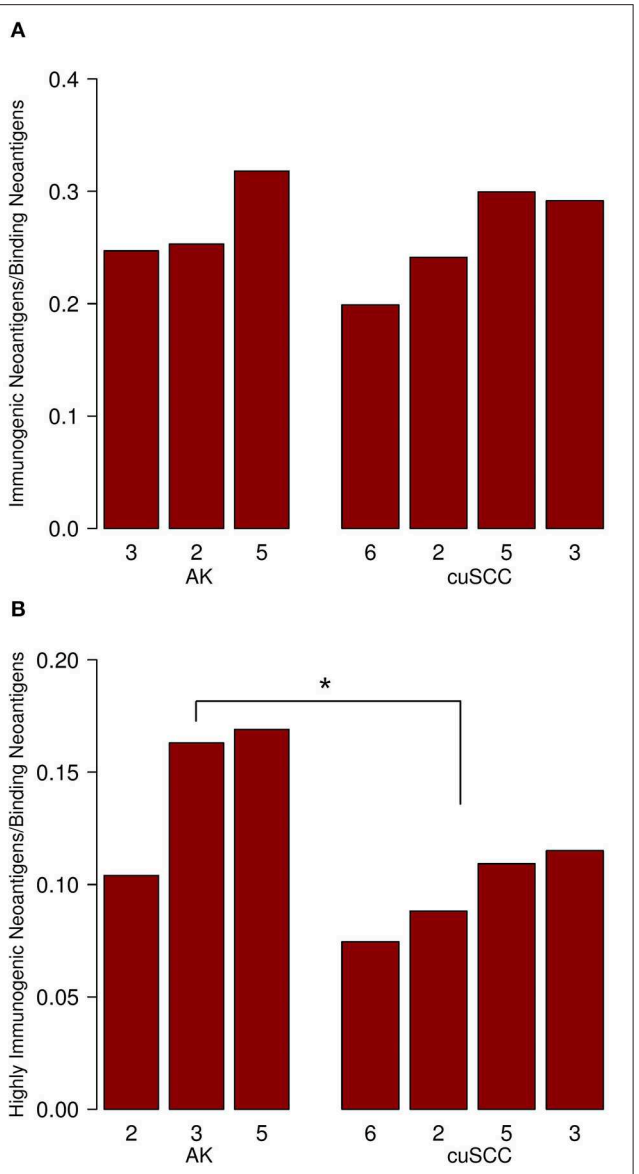


FIGURE 2 | (A) Ratio of immunogenic neoantigens to binding neoantigens for AK and cuSCC. **(B)** Ratio of those immunogenic neoantigens with a TCR recognition potential of $\geq$0.01 to the binding neoantigens. Numbers for AK and cuSCC indicate the patient sample from **Table 1** used for the analyses. *$p < 0.05$.

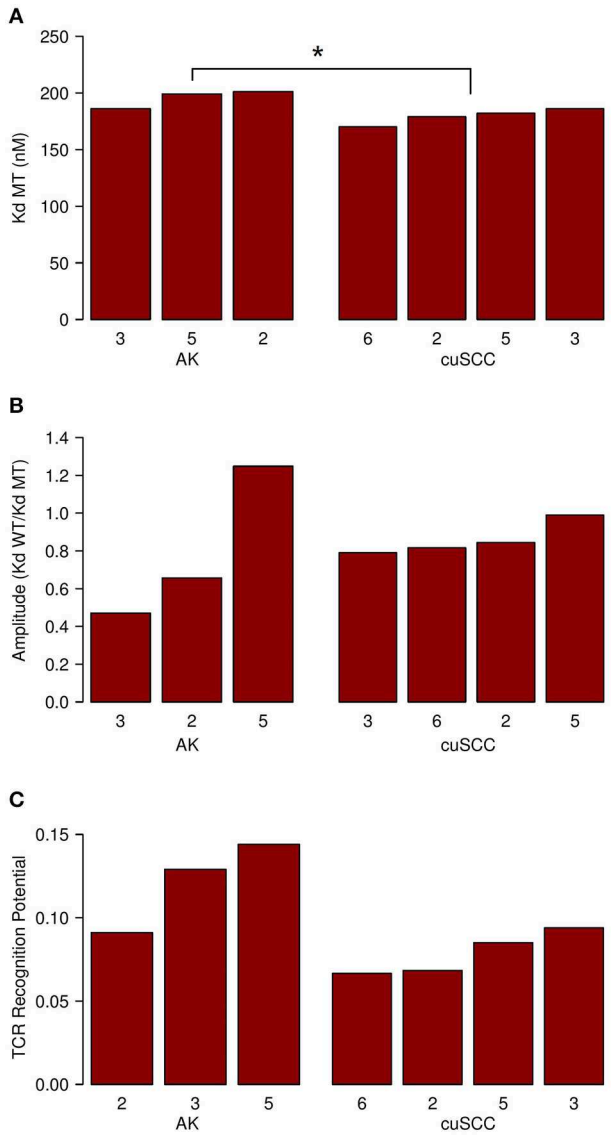


FIGURE 3 | (A) Average dissociation constant (K_d) for mutant (MT) peptide:MHC in AK and cuSCC. **(B)** Average amplitude of MHC binding calculated as the ratio of the dissociation constant (K_d) for wild-type (WT) peptide:MHC to the dissociation constant (K_d) of the mutant (MT) peptide:MHC in AK and cuSCC. **(C)** Average TCR recognition potential for AK and cuSCC. Numbers for AK and cuSCC indicate the patient sample from **Table 1** used for the analyses. *$p < 0.05$.

costs of AKs and cuSCCs ($p = 1.0$), it is notable that some of the cuSCCs had neoantigens with a greater fitness cost than those of the AKs.

In an attempt to explain why some cuSCCs had neoantigens with a high fitness cost, suggesting that these cuSCCs had neoantigens with high immune recognition potential yet escaped immune-mediated destruction, we adjusted the fitness cost values based on the expression of the neoantigens using the RPKM-normalized RNAseq data. There was a trend for an increased adjusted average fitness cost in AKs (0.00344) compared with cuSCCs (0.00023) (**Figure 4C**, $p = 0.08$). When evaluating the neoantigen with the maximum fitness cost after adjusting for expression, adjusted maximum fitness cost for AKs (0.22) was significantly increased compared with cuSCCs (0.014) (**Figure 4D**, $p = 0.03$). Taken together these data demonstrate that while cuSCCs have neoantigens with immune recognition potential, these neoantigens are not expressed, resulting in a higher adjusted maximum fitness cost in AKs compared to cuSCC.

Analysis of HLA Mutations and Immune Infiltration

We analyzed HLA-A, B, and C for both mutations and expression levels because recognition of neoantigens by the immune system requires functional MHC class I proteins to be made and expressed within the cell. Mutational analysis revealed two somatic mutations within HLA coding regions for the HLA-B gene in the cuSCC sample from patient 2 and the HLA-C gene in the AK sample from patient 5. However, neither of these mutations were within the peptide binding groove, which suggests that these mutations do not alter peptide binding. Additional analysis was performed to examine whether the mutations affected the expression level. Expression of HLA-B for patient 2 cuSCC was within one standard deviation of the mean expression and expression of HLA-C for patient 5 AK was within two standard deviations of the mean, suggesting that these mutations do not alter expression. Mutations in members of the MHC class I pathway (TAP1, TAP2, and B2M) can also lead to the loss of cell surface MHC class I expression and immune escape (39). Therefore, we interrogated TAP1, TAP2, and B2M genes for mutations, but no mutations in these genes were identified. We used a linear regression analysis to determine if the rate of expression of HLA (A, B, and C), TAP1, TAP2, or B2M correlated with the adjusted maximum fitness cost. No significant correlation was found between the adjusted maximum fitness cost and the normalized expression of any of the six proteins analyzed (data not shown). This result suggests that there was no meaningful impact of mutations in or expression of these proteins on the fitness cost of the tumor.

Analysis was also done to compare the level of immune infiltration (as determined by xCell analysis) to the neoantigen strength. Linear regressions were fit to the adjusted maximum fitness cost against the expression level of each of the following: T cell subsets (CD4 T cells, CD8 T cells, CD8 central memory T cells, CD8 effector memory T cells, CD8 naïve T cells, regulatory T cells), dendritic cell populations (immature DCs and conventional DC) and NK cells. There were trends for increased CD4 and CD8 T cell signatures in the tumor with increased adjusted maximum fitness cost (data not shown). However, no statistically significant associations were identified, which may in part be due to the small sample size.

DISCUSSION

Given that a low percentage of AKs progress to cuSCCs, it is important to understand what allows some of these

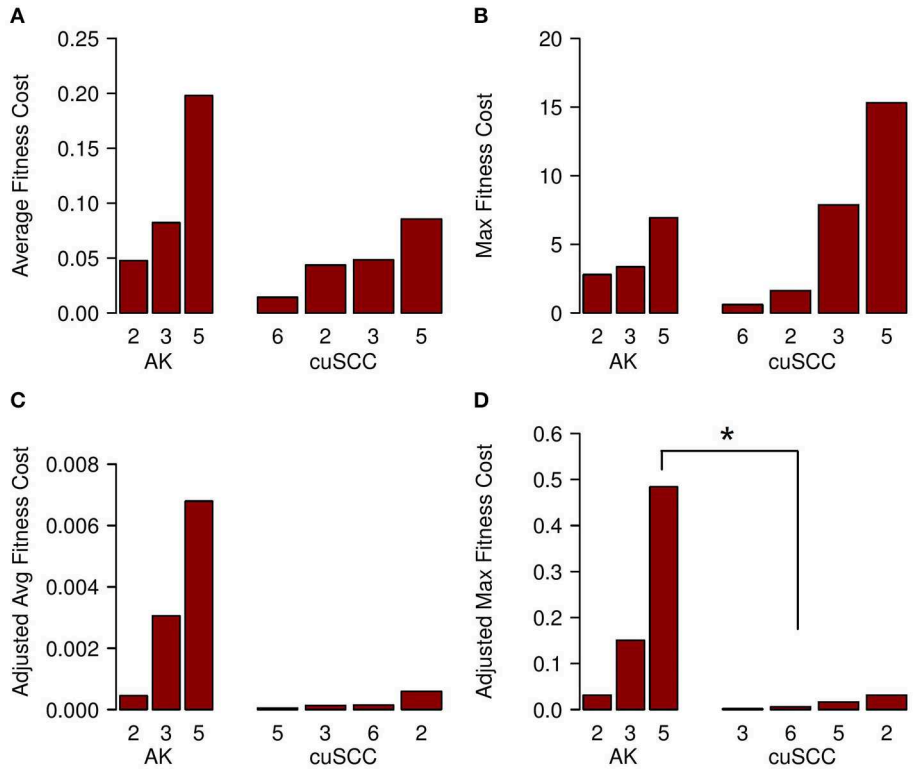


FIGURE 4 | (A) Average fitness cost for AK compared to cuSCC. Fitness cost is defined as the MHC-binding amplitude multiplied by the TCR recognition potential. **(B)** Maximum (Max) fitness cost for AK compared to cuSCC. Maximum fitness cost is defined as the highest calculated fitness cost across all neoantigens for each lesion. **(C)** Adjusted average (Avg) fitness cost for AK compared to cuSCC. Adjusted fitness cost is defined as the MHC-binding amplitude multiplied by the TCR recognition potential multiplied by the ratio of RPKM expression for the individual neoantigen to the sum of the RPKM expression of all neoantigens. **(D)** Maximum adjusted fitness cost for AK compared to cuSCC. Maximum adjusted fitness cost is defined as the highest calculated fitness cost after adjustment for RNA expression. Numbers for AK and cuSCC indicate the patient sample from **Table 1** used for the analyses. *$p < 0.05$.

precursor lesions to escape immune surveillance and become invasive cuSCC, while others AKs remain in equilibrium or are eliminated. An improved understanding of immunoediting in AKs and cuSCCs may aid clinical decision making and assist the development of preventive or treatment strategies. Chitsazzadeh et al. found, and we confirmed, that the maximum number of somatic mutations is greater in the cuSCCs than the AKs (though the large range prevents a statistically significant difference in the average number) (19). This finding is also consistent with work showing that esophageal adenocarcinomas have higher rates of mutations than the precursor Barrett's esophagus lesions (17, 18). To our knowledge, our study is the first to identify neoantigens in AKs and cuSCCs and compare the quality of neoantigens between precursor and invasive, cancerous lesions. A limitation of our analyses was the small sample size; unfortunately, no other publicly-accessible WES and RNAseq datasets were found for AK and cuSCC samples in PubMed, NCBI dbGaP or Google Scholar. We identified binding neoantigens (based on predicted MHC binding) and immunogenic neoantigens (based on predicted MHC binding and TCR recognition potential) in both AKs and cuSCC with similar findings to the number of somatic mutations. The maximum number of both binding and immunogenic neoantigens is higher in cuSCCs than AKs, but there was not a statistically significant difference in the average number. These findings indicate that there must be some factor other than the number of mutations and neoantigens that influences the immune-evasion potential of the tumor.

There was variation in the number of mutations and neoantigens across the different samples. This magnitude of variation is consistent with a prior study that has shown large differences in somatic mutations even in those tumors with high mutational burdens (40). Our results are also consistent with previous work reporting a small number of mutations that form binding neoantigens (41). For a mutation to lead to a neoantigen it must be in a coding region of the DNA, change the amino acid sequence, and meaningfully bind the MHC molecule. Because of these criteria, only a small number of the mutations accumulated in a tumor will create binding neoantigens and an even smaller number of these will be able to be recognized by the immune system as immunogenic neoantigens (as is reflected in **Table 2**).

One significant difference observed between the cuSCCs and the AKs is the percentage of their neoantigens that are highly recognized by TCRs. The cuSCCs on average had a lower percentage of binding neoantigens that are also predicted to be

highly recognized by the TCRs. Thus, the presence of a greater numbers of neoantigens in cuSCCs may not directly correlate to the visibility of that lesion to the immune system. Rather, the increased percentage of neoantigens that are recognized in AKs compared to cuSCCs indicates that the TCR recognition of the neoantigens predicted to bind MHC may be an important factor in determining the success of immunosurveillance. This finding corroborates the work of Anagnostou et al. who demonstrated that, after immunotherapy, the number of neoantigens increased, but the strength of those neoantigens decreased (16).

When considered separately, neither predicted MHC binding nor predicted TCR recognition of neoantigens appears to account for a difference in immune evasion of cuSCC compared with AK. The predicted affinity of neoantigens binding MHC in cuSCC is slightly higher on average than in AK, which would suggest that the immune system should be able to recognize the cuSCCs more readily. However, this difference is eliminated when the relative binding of the corresponding wild-type peptide is considered. This suggests that the negative selection of self-reactive T cells is one important factor to consider in determining the ability of the immune system to recognize the tumors. We observed a trend for a difference in TCR recognition potential between AKs and cuSCCs ($p = 0.0771$), suggesting that, perhaps, with a larger sample size, this might prove to be a statistically significant factor in determining immune recognition. However, with our current sample size, it is not possible to determine if TCR recognition potential alone is able to explain the difference in immune evasion between the pre-cancerous and cancerous lesions.

Predicted MHC binding and TCR recognition potential were then combined into an overall fitness cost, as has been done by Łuksza et al. (15), but the average and maximum fitness costs were found to have no statistically significant difference between AKs and cuSCCs. Surprisingly, the range of the maximum fitness costs included higher values for the cuSCCs than the AKs. This result seemed to contradict the expected result that the cuSCCs should have enhanced immune evasion mechanisms than the AKs, as they have escaped immune recognition to develop into invasive cancers. We then incorporated RNAseq expression data into our calculation since a neoantigen of any strength cannot be recognized by the immune system unless it is appreciably expressed by the cell. When the expression data was incorporated, the maximum fitness cost for AKs was significantly increased compared with cuSCCs, and there was a trend for increased average fitness cost for AKs compared with cuSCCs. These results suggest that, although the cuSCCs accumulate many potentially immunogenic neoantigens (many of which are stronger than the immunogenic neoantigens of the AKs), these neoantigens are not appreciably expressed. This leads us to hypothesize that down-regulating the expression of these strong neoantigens is a mechanism of immune evasion for cuSCCs that may allow them to escape the immune system whereas the AKs fail to do so. Consistent with this hypothesis, Rosenthal et al. demonstrated that there are higher rates of gene suppression in those genes with mutations that formed neoantigens compared to genes with non-neoantigenic mutations (7). The work of Matsushita et al. also corroborates our finding by showing that a strong neoantigen is downregulated in those tumors which succeed in growing in an immunocompromised host compared to an immunodeficient host (42). Their work also corroborates the suggestion that it is the maximum strength neoantigen that is of most importance in determining the immune detection of the tumor.

By identifying and comparing the quality of neoantigens in AKs and cuSCCs, we show that cuSCCs have lower rates of predicted TCR recognition of neoantigens that bind MHC class I than do AKs. Additionally, when expression is considered, many of the strong neoantigens from cuSCCs are insufficiently expressed, causing the maximum fitness cost (and by extension, immune recognition) to be lower for cuSCCs than for AKs. Our findings shed light on why tumor mutational burden alone may not accurately predict the anti-tumor immune response in treatment with immune checkpoint inhibition (9–13). Our results also suggest a role for the down-regulation of highly immunogenic neoantigens in the escape of cuSCCs from immune recognition.

DATA AVAILABILITY STATEMENT

The RNAseq data analyzed in this study was obtained from NCBI/GEO with SuperSeries accession code GSE84194 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84194) and the WES data was obtained from Dr. Kenneth Tsai. Requests to access the WES datasets should be directed to Dr. Kenneth Tsai, kenneth.tsai@moffitt.org.

AUTHOR CONTRIBUTIONS

EB, KB, and KH contributed conception and design of the study. EB performed the calculations and wrote the first draft of the manuscript. PK completed the statistical calculations. HN, TP, and MW provided the neoantigen identification pipeline and technical assistance in running the calculations. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work was supported in part by the ASU Center for Evolution and Medicine postdoctoral fellowship and the Marcia and Frank Carlucci Charitable Foundation postdoctoral award from the Prevent Cancer Foundation (HN), the Springboard Initiative from the University of Arizona College of Medicine Phoenix (KH), and the University of Arizona College of Medicine—Phoenix M.D./Ph.D. Program (EB).

ACKNOWLEDGMENTS

We thank Dr. Kenneth Tsai and Dr. Kimal Pajapakshe for kindly providing the data for this analysis. We thank PK of the Biostatistics and Study Design Core at the University of Arizona College of Medicine—Phoenix for providing statistical analysis services and thank Dr. Chengcheng Hu for providing guidance to PK.

REFERENCES

1. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence estimate of non-melanoma skin cancer (keratinocyte carcinomas) in the US population, 2012. *JAMA Dermatol*. (2015) 151:1081. doi: 10.1001/jamadermatol.2015.1187
2. Schmults CD, Karia PS, Carter JB, Han J, Qureshi AA. Factors predictive of recurrence and death from cutaneous squamous cell carcinoma: a 10-year, single-institution cohort study. *JAMA Dermatol*. (2013) 149:541. doi: 10.1001/jamadermatol.2013.2139
3. Karia PS, Han J, Schmults CD. Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012. *J Am Acad Dermatol*. (2013) 68:957–66. doi: 10.1016/j.jaad.2012.11.037
4. Migden MR, Rischin D, Schmults CD, Guminski A, Hauschild A, Lewis KD, et al. PD-1 blockade with cemiplimab in advanced cutaneous squamous-cell carcinoma. *N Engl J Med*. (2018) 379:341–51. doi: 10.1056/NEJMoa1805131
5. Fernandez Figueras MT. From actinic keratosis to squamous cell carcinoma: pathophysiology revisited. *J Eur Acad Dermatology Venereol*. (2017) 31:5–7. doi: 10.1111/jdv.14151
6. Ward JP, Gubin MM, Schreiber RD. The role of neoantigens in naturally occurring and therapeutically induced immune responses to cancer. *Adv Immunol*. (2016) 130:25–74. doi: 10.1016/bs.ai.2016.01.001
7. Rosenthal R, Cadieux E, Salgado R, Bakir MA, Moore DA, Hiley CT, et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature*. (2019) 567:479–85. doi: 10.1038/s41586-019-1032-7
8. Rausch MP, Hastings KT. *Immune Checkpoint Inhibitors in the Treatment of Melanoma: From Basic Science to Clinical Application*. Brisbane, QLD: Codon Publications (2017). doi: 10.15586/codon.cutaneousmelanoma.2017.ch9
9. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. (2016) 14:1463–9. doi: 10.1126/science.aaf1490
10. Hellmann MD, Callahan MK, Awad MM, Calvo E, Ascierto PA, Atmaca A, et al. Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer. *Cancer Cell*. (2018) 33:853–61. doi: 10.1016/j.ccell.2018.04.001
11. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*. (2015) 348:124–8. doi: 10.1126/science.aaa1348
12. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. (2015) 350:207–11. doi: 10.1126/science.aad0095
13. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. (2014) 371:2189–99. doi: 10.1056/NEJMoa1406498
14. Miller A, Asmann Y, Cattaneo L, Braggio E, Keats J, Auclair D, et al. High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. *Blood Cancer J*. (2017) 7:e614. doi: 10.1038/bcj.2017.94
15. Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A neoantigen fitness model predicts tumor response to checkpoint blockade immunotherapy. *Nature*. (2017) 551:517–20. doi: 10.1038/nature24473
16. Anagnostou V, Smith KN, Forde PM, Niknafs N, Bhattacharya R, White J, et al. Evolution of neoantigen landscape during immune checkpoint blockade in non-small cell lung cancer. *Cancer Discov*. (2017) 7:264–76. doi: 10.1158/2159-8290.CD-16-0828
17. Ross-Innes CS, Becq J, Warren A, Cheetham RK, Northen H, O'Donovan M, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat Genet*. (2015) 47:1038–46. doi: 10.1038/ng.3357
18. Stachler MD, Camarda ND, Deitrick C, Kim A, Agoston AT, Odze RD, et al. Detection of mutations in barrett's esophagus before progression to high-grade dysplasia or adenocarcinoma. *Gastroenterology*. (2018) 155:156–67. doi: 10.1053/j.gastro.2018.03.047
19. Chitsazzadeh V, Coarfa C, Drummond JA, Nguyen T, Joseph A, Chilukuri S, et al. Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates. *Nat Commun*. (2016) 7:1–17. doi: 10.1038/ncomms12601
20. Andrews S. *FastQC A Quality Control tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
22. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. (2015) 12:357–60. doi: 10.1038/nmeth.3317
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. (2010) 20:1297–303. doi: 10.1101/gr.107524.110
25. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. (2011) 43:491–8. doi: 10.1038/ng.806
26. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma*. (2013) 43:1–33. doi: 10.1002/0471250953.bi1110s43
27. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. (2012) 22:568–76. doi: 10.1101/gr.129684.111
28. Zuehlsdorff TJ, Isborn CM. Ensembl 2018. *Nucleic Acids Res*. (2018) 46:D754–61. doi: 10.1093/nar/gkx1098
29. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol*. (2016) 17:122. doi: 10.1186/s13059-016-0974-4
30. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, et al. pVAC-Seq: a genome-guided *in silico* approach to identifying tumor neoantigens. *Genome Med*. (2016) 8:11. doi: 10.1186/s13073-016-0264-5
31. Hundal J, Kiwala S, McMichael J, Miller CA, Wollam AT, Xia H, et al. pVACtools: a computational toolkit to identify and visualize cancer neoantigens. *bioRxiv*. (2019) 2019:501817. doi: 10.1101/501817
32. Narang P, Chen M, Sharma AA, Anderson KS, Wilson MA. The neoepitope landscape of breast cancer: implications for immunotherapy. *BMC Cancer*. (2019) 19:200. doi: 10.1186/s12885-019-5402-1
33. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol*. (2015) 33:1152–8. doi: 10.1038/nbt.3344
34. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol*. (2017) 199:3360–8. doi: 10.4049/jimmunol.1700893
35. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res*. (2015) 43:D405–12. doi: 10.1093/nar/gku938
36. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. (2017) 14:417–9. doi: 10.1038/nmeth.4197
37. Gene. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; https://www.ncbi.nlm.nih.gov/gene/. Published 2004. Accessed October 5, 2019.
38. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. (2017) 18:220. doi: 10.1186/s13059-017-1349-1
39. Wang Z, Margulies L, Hicklin DJ, Ferrone S. Molecular and functional phenotypes of melanoma cells with abnormalities in HLA class I antigen expression. *Tissue Antigens*. (1996) 47:382–90. doi: 10.1111/j.1399-0039.1996.tb02573.x

40. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. (2013) 500:415–21. doi: 10.1038/nature12477
41. Wood MA, Paralkar M, Paralkar MP, Nguyen A, Struck AJ, Ellrott K, et al. Population-level distribution and putative immunogenicity of cancer neoepitopes. *BMC Cancer*. (2018) 18:414. doi: 10.1186/s12885-018-4325-6
42. Matsushita H, Vesely MD, Koboldt DC, Rickert CG, Uppaluri R, Magrini VJ, et al. Cancer exome analysis reveals a T cell dependent mechanism of cancer immunoediting. *Nature*. (2012) 482:400–4. doi: 10.1038/nature10755

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 11 December 2019
doi: 10.3389/fimmu.2019.02766

Cancer Neoepitopes for Immunotherapy: Discordance Between Tumor-Infiltrating T Cell Reactivity and Tumor MHC Peptidome Display

Stina L. Wickström[1†], Tanja Lövgren[1,2†], Michael Volkmar[3], Bruce Reinhold[4,5], Jonathan S. Duke-Cohan[4,5], Laura Hartmann[1], Janina Rebmann[3], Anja Mueller[6], Jeroen Melief[1], Roeltje Maas[1], Maarten Ligtenberg[1], Johan Hansson[1], Rienk Offringa[3], Barbara Seliger[6], Isabel Poschke[3,7], Ellis L. Reinherz[4,5] and Rolf Kiessling[1]*

[1] Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden, [2] Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden, [3] Division of Molecular Oncology of Gastrointestinal Tumors, German Cancer Research Center (DKFZ), Heidelberg, Germany, [4] Laboratory of Immunobiology, Dana-Farber Cancer Institute, Boston, MA, United States, [5] Department of Medical Oncology, Dana-Farber Cancer Institute and Department of Medicine, Harvard Medical School, Boston, MA, United States, [6] Institute of Medical Immunology, Martin Luther University Halle-Wittenberg, Halle, Germany, [7] DKTK Immune Monitoring Unit, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

OPEN ACCESS

Edited by:
Nikolaos G. Sgourakis, University of California, Santa Cruz, United States

Reviewed by:
Pin Wu, Zhejiang University, China
Joan Font-Burgada, Fox Chase Cancer Center, United States

***Correspondence:**
Stina L. Wickström
stina.wickstrom@ki.se

[†] These authors share first authorship

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: 30 August 2019
Accepted: 12 November 2019
Published: 11 December 2019

Citation:
Wickström SL, Lövgren T, Volkmar M, Reinhold B, Duke-Cohan JS, Hartmann L, Rebmann J, Mueller A, Melief J, Maas R, Ligtenberg M, Hansson J, Offringa R, Seliger B, Poschke I, Reinherz EL and Kiessling R (2019) Cancer Neoepitopes for Immunotherapy: Discordance Between Tumor-Infiltrating T Cell Reactivity and Tumor MHC Peptidome Display. Front. Immunol. 10:2766. doi: 10.3389/fimmu.2019.02766

Tumor-infiltrating lymphocytes (TIL) are considered enriched for T cells recognizing shared tumor antigens or mutation-derived neoepitopes. We performed exome sequencing and HLA-A*02:01 epitope prediction from tumor cell lines from two HLA-A2-positive melanoma patients whose TIL displayed strong tumor reactivity. The potential neoepitopes were screened for recognition using autologous TIL by immunological assays and presentation on tumor major histocompatibility complex class I (MHC-I) molecules by Poisson detection mass spectrometry (MS). TIL from the patients recognized 5/181 and 3/49 of the predicted neoepitopes, respectively. MS screening detected 3/181 neoepitopes on tumor MHC-I from the first patient but only one was also among those recognized by TIL. Consequently, TIL enriched for neoepitope specificity failed to recognize tumor cells, despite being activated by peptides. For the second patient, only after IFN-γ treatment of the tumor cells was one of 49 predicted neoepitopes detected by MS, and this coincided with recognition by TIL sorted for the same specificity. Importantly, specific T cells could be expanded from patient and donor peripheral blood mononuclear cells (PBMC) for all neoepitopes recognized by TIL and/or detected on tumor MHC-I. In summary, stimulating the appropriate inflammatory environment within tumors may promote neoepitope MHC presentation while expanding T cells in blood may circumvent lack of specific TIL. The discordance in detection between physical and functional methods revealed here can be rationalized and used to improve neoantigen-targeted T cell immunotherapy.

Keywords: TIL, tumor, Immune peptidome, neoepitopes, mass spectrometry, immunotherapy

INTRODUCTION

Tumor-infiltrating lymphocytes (TIL) in patients with metastatic malignant melanoma are thought to be enriched for T cells that can recognize antigens expressed by the patient's tumor. In line with this, therapy with autologous TIL, expanded to large numbers *ex vivo* and reinfused to melanoma patients, can induce long-lasting clinical responses in a large proportion (40–70%) of patients (1). Different categories of tumor-associated antigens (TAA) are recognized by TIL, and initial efforts focused on broadly expressed TAA shared between patients. Such TAA include both differentiation antigens that are found in the normal melanocytic counterparts and aberrantly expressed antigens such as cancer-testis antigens that are normally expressed only in immune privileged sites. Therapeutic approaches with T cells transduced with T cell receptors (TCR) recognizing these types of shared TAA, exemplified by NY-ESO-1, MART-1, gp100, and MAGE-A3, have resulted in clinical regressions of metastatic lesions in a limited number of treated patients, sometimes with severe side effects caused by cross-reactivity to normal tissues (2, 3).

Recently, the focus of the research field has shifted toward tumor-specific antigens associated with somatic mutations (neoantigens/neoepitopes), which are in the majority of cases unique for each patient. This development has been spurred by advancements in next-generation sequencing (NGS) techniques that have made it possible to almost routinely identify all tumor-associated mutations, including both shared mutations in driver genes (e.g., Ras, p53) and patient-unique passenger mutations. Passenger mutations are not part of oncogenesis, but tend to accumulate during tumor progression especially in tumors caused by UV or carcinogen exposure, typically exemplified by melanomas, and lung cancers.

Neoepitopes resulting from mutations are attractive cancer immunotherapy targets. The mutation is not present during the selection in the thymus and thus exempt from central tolerance. Thus, neoepitopes are seen as "foreign" non-self. In addition, the mutations are truly tumor-specific and there is less risk for ON-target, OFF-tumor side effects although cross-reactivities to epitopes in other proteins can probably occur. Several lines of evidence have indicated that neoepitope frequency can be decisive in determining the capacity of patient's T cells to reject their tumors. Thus, an association between mutational load and clinical outcome in patients treated with antibodies blocking the checkpoint molecules CTLA4 and PD-1 has been described (4, 5). In addition, a connection between clinical efficacy of TIL adoptive cell therapy (ACT) and the presence of T cells specific for tumor-derived mutations in the infused TIL has been suggested (6, 7). Furthermore, ACT performed with TIL enriched for neoepitope-specific T cells has resulted in successful clinical outcomes (8, 9).

In this study, we used two peptide libraries containing *in silico*-predicted T cell neoepitopes derived from whole exome sequencing data of early passage tumor cell lines from two HLA-A2*02:01-positive melanoma patients. The predicted neoepitopes were screened for their ability to activate autologous TIL in functional assays and for their presence on MHC-I by mass spectrometry (MS). This combined approach revealed a significant discordance between the immunologic and physical detection methods, with neoepitopes recognized by autologous TIL not being detected by MS, and vice versa. Here, this discrepancy is examined. Peptide recovery and detection sensitivity for MS are characterized in addition to TIL functional assays including assessment of specificity, avidity, and activation capacity, as well as neoepitope immunogenicity. Our results highlight the difficulties to be faced when aiming to target tumors with neoepitope-specific, T cell-based immunotherapy and suggest strategies on how to improve such therapy.

MATERIALS AND METHODS

Patients

Patient ANRU is a male born in 1975 who was operated in November 2014 for stage III axillary lymph node metastatic melanoma from which a tumor line and TIL cells were isolated. He had a relapse with CNS metastases in 2015, for which he was operated and received local radiotherapy. Since then, he is tumor free and has received no systemic therapy. Patient KADA is a male born in 1938 who was operated in 2011 for stage III axillary lymph node metastasis from which a tumor line and TIL cells were isolated. He has had no systemic treatment and has since then remained recurrence free. The protocol for patient participation was approved by the local Ethics Committee (Dno. 2011/143-32/1 and 2015/1862-32) and the Institutional Review Board. Both patients signed a written informed consent in accordance with the Declaration of Helsinki.

Cells and Tissues

Original SK-OV-3, HLA-A2*02:01-transfected SK-OV-3 and T2 cell lines were grown in RPMI supplemented with FCS (10%), penicillin (100 U/ml; LifeTechnologies), and streptomycin (100 μg/ml; LifeTechnologies). PBMC were prepared from healthy blood donor buffy coats by Ficoll-Hypaque (GE Healthcare) density-gradient centrifugation, according to the manufacturer's instructions.

Monocytes and CD8+ T cells were isolated using CD14+ or CD8+ magnetic-bead-based isolation (Miltenyi Biotec), respectively, according to the manufacturer's instructions. Monocytes were matured into dendritic cells (DC) using a two-step protocol as described previously (10).

Resected tumors from two HLA-A2*02:01-positive melanoma patients (acronym ANRU and KADA) were used for generation of tumor cell lines and expansion of TIL. Tumor cell lines were established by mechanical dissociation of tumor tissue by cutting and grinding through a 70-μm cell strainer (Corning). Tumor cells were cultured in RPMI (LifeTechnologies) supplemented with FCS (20%), penicillin (100 U/ml; LifeTechnologies), and streptomycin (100 μg/ml; LifeTechnologies). Tumor cells were monitored frequently for growth and medium was changed/cultures were expanded when necessary. Where indicated, tumor cells were first cultured for 24 h using standard culture conditions and were thereafter exposed to IFN-γ (25 ng/ml; R&D Systems) for 72 h.

TIL were expanded as described by Poschke et al. (11), by first stimulating expansion from tumor fragments with IL-2 alone and thereafter performing a rapid expansion protocol (REP) by stimulation with IL-2 and anti-CD3 antibodies in the presence of irradiated PBMC as feeder cells. All TIL and T cells, including co-cultures, were cultured in CellGro® plus human AB-serum (2%; The Blood Bank, Karolinska University Hospital).

Exome Sequencing Data Analysis

See **Supplementary Materials and Methods**.

Expression Analysis of Mutated Genes and Alleles

To coarsely estimate expression of mutated genes in the absence of whole transcriptome data for the tumor cell lines, publicly available RNA-seq profiles of seven melanoma cell lines [GSE46817 (12, 13)] and averaged reads per kilo base per million mapped reads (RPKM) expression values were collected and analyzed. Genes with mean RPKM >1 and a low standard deviation were considered expressed.

To verify transcription of selected genes and mutated/wild-type alleles, total RNA from ANRU and KADA tumor cell lines was isolated using the RNeasy Mini kit (Qiagen) and quality and quantity were measured on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). To confirm expression of the mutated alleles, isolated RNA was treated with DNase (Thermo Fisher Scientific, Waltham, MA USA) and converted to cDNA using SuperScript™ III reverse transcriptase (RT; Thermo Fisher Scientific) and oligo$(dT)_{15}$ primers. For each sample, a control without the reverse transcriptase was included to detect possible contaminating genomic DNA. The selected genes were amplified using MyTaq DNA polymerase (Bioline, London, UK) and PCR primers listed in **Supplementary Table 4**. The PCR products were sent for Sanger sequencing (Eurofins MWG GmbH, Ebersberg, Germany) and allelic expression was assessed from the resulting electropherograms.

To quantify total expression of selected genes and the common melanoma tumor-associated antigens MART-1 and gp100, relative to housekeeping genes GAPDH and β-actin [primers from Radonić et al. (14)], quantitative PCR (qPCR) was performed using the LightCycler® 480 SYBR Green I Master and a Roche LightCycler® 480. Amplification efficiency was verified with serial dilutions of template cDNA. All samples were amplified in triplicate and resulting Cp values were averaged.

Peptides and Peptide Libraries

Neoepitopes with a predicted affinity of <50 nM and mean RPKM of at least 1 were ordered as crude micro-scale peptide libraries (JPT Peptide Technologies). The KADA library contained 181 peptides and the ANRU library contained 49 peptides. Neoepitope peptides found to activate TIL or that were found on tumor MHC-I were thereafter ordered in larger scale and higher purity, as were corresponding wild-type peptides.

Analysis of MHC-I Presentation Machinery

The analysis of MHC presentation machinery was performed as previously described (15) by Western blot staining for peptide processing components (TAP1 and 2, Tapasin, MHC-I heavy chain, LMP2 and 10, and β2m) in untreated and IFN-γ-treated tumor cells, confirmed also by quantitative PCR.

Isolation of Peptide-HLA Class I Complexes From Melanoma Cell Lines

See **Supplementary Materials and Methods.**

Immune Peptidome Analysis by MS

All Poisson detection MS methodology and liquid chromatography (LC) data independent acquisition (DIA) MS methodologies have been described in detail previously (16–18). In brief, mass spectra were collected on a quadrupole-TOF (Sciex 6600+) instrument in a DIA format. The m/z region 400–680 was split into 11 minimally overlapping windows of variable width designed to transmit equal ion fluxes with MHC-I immune peptidomes. MS data were collected in a series of a single full-range MS spectrum followed with 11 MS/MS spectra for each transmitted window. The MS/MS spectra were compared with reference patterns obtained from synthetic peptides using an algorithm based on the theory of sampling a Poisson process (18). High LC-MS sensitivity was promoted using electrospray ionization with 20 μm ID alkane modified polystyrene-divinylbenzene monolithic columns [fabricated in-house (19)] at flow rates of roughly 10 nl/min. Elution positions of the synthetic peptides relative to shared endogenous immune peptides using the same column configuration were also determined, and this provides a restrictive map for the elution positions of the neoepitope candidates in the tumor DIA MS data (**Supplementary Figure 5**).

TIL Functional Assessment

Recognition of tumor cells or neoepitope-pulsed antigen-presenting cells (APC; SK-OV-3) by TIL was assessed in co-cultures using a TIL:tumor cell/APC ratio of 5:1 in 96-well U-bottom plates. Recognition of neoepitope peptides by TIL was first tested using pools of 5 (ANRU) or 10 (KADA) peptides. For pools that activated TIL, each peptide was tested individually. Known shared TAA epitopes (MART-1: ELAGIGILTV, MAGE-A4: GVYDGREHTV, MAGE-A10: GLYDGMEHL, gp100: IMDQVPFSV, tyrosinase: YMDGTMSQV, NY-ESO-1: SLLMWITQV) or viral epitopes (CMV pp65 antigen: NLVPMVATV, HCV NS3 antigen: KLVALGINAV, HIV p17 antigen: SLYNTVATL, Influenza M1 antigen: GILGFVFTL) were included as controls.

Peptide recognition was analyzed by pulsing of HLA-A2-transfected SK-OV-3 cells with 10 μg/ml total peptide concentration (resulting in 2 or 1 μg/ml of each peptide in the 5- and 10-peptide pools, respectively) in PBS for 1 h at 37°C before washing and co-culture. Original untransfected SK-OV-3 were included as negative controls. For KADA, due to too high background stimulation by SK-OV-3-A2+ cells, peptide pools (10 μg/ml) were added directly to KADA TIL, which then served as APC themselves. Thereafter, single peptides from pools that activated KADA TIL were analyzed using HLA-A2-transfected SK-OV-3 as APC, as described. Where indicated, MHC-I was blocked by pre-incubating tumor cells or APC for

30 min at 37°C with 20 μg/ml anti-HLA-ABC antibody (clone W6/32, BioLegend) or anti-HLA-A2 antibody (clone BB7.2, AbD Serotec) before addition of TIL. CD3/CD28 Dynabeads (LifeTechnologies) were used as positive control according to the manufacturer's instructions.

Readouts for TIL activation were degranulation measured as surface expression of CD107a and cytokine production as measured by intracellular or secreted IFN-γ by FACS or ELISA, respectively. Experiments aiming to determine T cell functional avidity were performed by titrating peptides in eight steps of 10-fold dilutions from 100 μg/ml directly on TIL.

In experiments where CD107a and dextramer staining was to be performed, co-cultures were incubated for 5 h before being harvested for staining. In experiments with CD107a staining but without dextramer staining, GolgiPlug™ and GolgiStop™ (BD Bioscience) were added after 2 h co-culture, and cells were harvested after an additional 4 h co-culture.

In experiments where IFN-γ ELISA analysis of supernatants was to be performed, co-cultures were incubated for 24 h.

Flow Cytometry

All antibodies and other flow cytometry reagents were used according to the manufacturer's instructions, unless otherwise stated. All had been titrated for optimal signal-to-noise ratio and staining was, unless specified differently, performed in PBS supplemented with 0.1% albumin. Data for all flow cytometry were acquired on a NovoCyte (ACEA Biosciences) or a BD LSR II (BD Biosciences) and analyzed using FlowJo Software (TreeStar) as geometric MFI or percent positive cells compared to the parent population.

Staining was performed for T cells specific for known shared tumor-associated antigen T cell epitopes (MART-1, NY-ESO-1, MAGE-A3, Tyrosinase, gp100, and MAGE-A1; Melanoma Dextramer® Collection 1 kit, Immudex) or for neoepitope-specific T cells (custom-ordered PE-labeled neoepitope/HLA-A2*02:01 dextramers, Immudex). Cell surface expression was analyzed for CD8 (clone SK1, APC-Cy7, BioLegend), CD3 (clone UCHT1, PE-Cy7, BioLegend), MHC-I (clone W6/32, APC, BioLegend), and HLA-A2 (clone BB7.2, PE, BioLegend). All staining protocols included a dead cell marker (staining in PBS only; LIVE/DEAD® fixable Aqua Dead cell stain, InVitrogen).

Detection of activated T cells was performed by staining with CD107a antibody (clone H4A3, FITC, BioLegend), which was added to stimulated TIL/T cell cultures at experiment setup (20).

When dextramer staining was performed, dextramers were always added first, then CD8 staining and last dead cell labeling. In experiments where intracellular staining was performed, cells were stained for dead cells, then CD3 and CD8, before fixation and permeabilization using CytoPerm/CytoFix™ (BD Biosciences) and intracellular staining for IFN-γ (clone 4S.B3, PE, Biolegend).

HLA-A2 Stabilization Assay

HLA-A2 stabilization assays were performed using T2 cells that were harvested, washed, resuspended in serum-free RPMI and thereafter seeded at 200,000 cells/well in 96 U-bottom plates (TPP®). Peptides were added in serial dilution of 1.5–100 μg/ml and incubated overnight in room temperature and then an additional 2.5 h at 37°C. The cells were then harvested and stained for HLA-A2 as described above. HLA-A2 stabilization data were normalized according to the formula (gMFI (peptide) – gMFI (no peptide))/gMFI (no peptide).

Generation of Neoepitope-Specific TIL Lines or Stimulation of Neoepitope-Specific Cells From PBMC

TIL were sorted for neoepitope specific T cells by labeling with custom PE-labeled neoepitope dextramers, as described. The dextramer-stained cells were enriched by MACS by binding to anti-PE microbeads (Miltenyi Biotec), following the manufacturer's instructions. Both enriched and depleted populations were immediately subjected to a REP to expand TIL. As control, unsorted TIL were subjected to REP in parallel.

To expand neoepitope-specific cells from patient or HLA-A2+ donor blood, DC were loaded with neoepitope peptides as described for SK-OV-3 cells above. The DC were co-cultured with autologous CD8+ T cells in a 1:5 ratio for 14 days in CellGro® supplemented with 20 IU/ml IL-2 (Proleukine, Novartis).

Immunoassays

ELISA for IFN-γ (MabTech) was performed according to the manufacturer's instructions. Standard curves were plotted as four-parameter sigmoidal curves and unknowns were calculated and plotted using GraphPad Prism (GraphPad).

RESULTS

Recognition of Tumor Cells and Common Melanoma TAA by TIL

Tumor cell lines and corresponding TIL were generated from two HLA-A*02:01-positive melanoma patients (KADA and ANRU). Both TIL recognized autologous tumor cells and responded with degranulation and production of IFN-γ, which was measured by FACS as increased cell surface CD107a and intracellular IFN-γ (**Supplementary Figure 1A**) or by ELISA as secreted IFN-γ in supernatants [**Figure 1A** (KADA) and **Figure 1D** (ANRU)]. The activation was partly decreased by blocking HLA-A2 (BB7.2 mAb) and more pronouncedly by total blocking of MHC-I (W6/32 mAb) on tumor cells (**Supplementary Figure 1A**).

Next, we evaluated whether activation of TIL was due to recognition of some of the TAA commonly expressed by melanoma [**Figure 1A** (KADA) and **Figure 1D** (ANRU)] by co-culture of TIL and peptide-pulsed HLA-A2-transfected SK-OV-3 target cells. KADA TIL were only activated by a control peptide derived from the Influenza M1 matrix protein, while ANRU TIL were activated by peptides from both MART-1 and, although weaker, gp100. This specificity of ANRU TIL for MART-1 and weakly for gp100 was confirmed by FACS analysis using a panel of dextramers for common melanoma antigens (MART-1, NY-ESO-1, MAGE-A3, Tyrosinase, gp100, and MAGE-A1; **Supplementary Figure 1B**). Of note, both these peptides were tested in their modified form, optimized for better HLA-A2

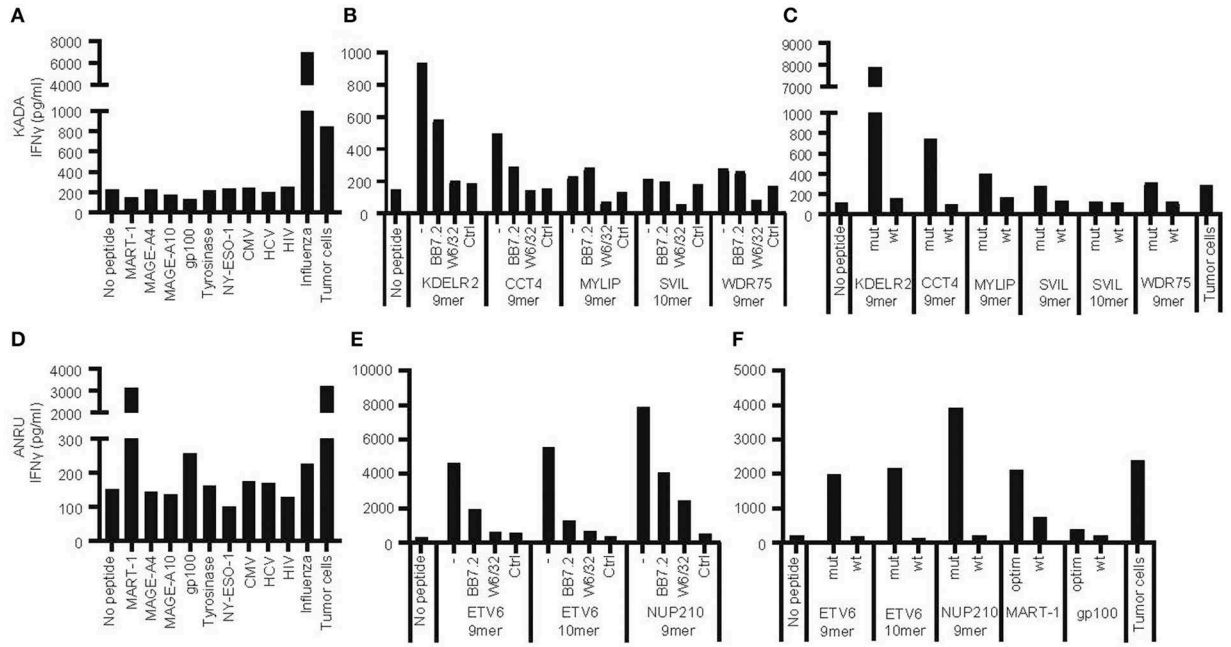


FIGURE 1 | Recognition of autologous tumor cells or peptides from shared tumor-associated antigens, from mutated genes, or from viruses by tumor-infiltrating lymphocytes. Tumor-infiltrating lymphocytes (TIL) and primary tumor cell lines were expanded from tumors of patients KADA and ANRU. The ability of TIL to recognize corresponding tumor cells or HLA-A2-transfected SK-OV-3 target cells pulsed with peptides from common tumor-associated antigens or viruses (**A**, KADA; **D**, ANRU), or HLA-A2-transfected SK-OV-3 cells pulsed with neoepitope peptides derived from mutated genes in tumor cells, in the absence or presence of anti-HLA-A2 (BB7.4) or -MHC-I (W6/32) blocking monoclonal antibodies (**B**, KADA; **E**, ANRU; only recognized peptides shown), or HLA-A2-transfected SK-OV-3 pulsed with neoepitope peptides compared to corresponding wild-type peptides (**C**, KADA; **F**, ANRU), and respond with IFN-γ secretion was assessed after 24 h by ELISA. Unpulsed HLA-A2-transfected SK-OV-3 target cells were used as a no peptide control and pulsed original SK-OV-3 cells (Ctrl) were used as a no HLA-A2 control as indicated.

TABLE 1 | Selection of mutated peptides predicted to bind HLA-A2 with high affinity from whole exome sequences of tumor cell lines.

Mutation type	Prediction category	Patient	
		KADA	ANRU
SNV	Non-synon. mutations	2554	323
	Predicted HLA-A2 peptides	1713	182
	Ordered HLA-A2 peptides	165	37
DNV,TNV	Non-synon. mutations	120	9
	Predicted HLA-A2 peptides	37	7
	Ordered HLA-A2 peptides	12	0
InDels	Non-synon. mutations	571	611
	Predicted HLA-A2 peptides	211	177
	Ordered HLA-A2 peptides	4	12
Total	Non-synon. mutations	3244	943
	Predicted HLA-A2 peptides	1961	366
	Ordered HLA-A2 peptides	181	49

SNV, single-nucleotide variant; DNV, double-nucleotide variant; TNV, triple-nucleotide variant; InDels, insertion/deletions.

binding, while the native peptides have a much lower affinity for HLA-A2 (**Table 2**) and also a decreased ability to activate T cells (**Figure 1F**).

Identification of Somatic Mutations in Tumor Cells

Exome sequencing was performed for tumor cell lines and normal tissue (PBMC) from patients KADA and ANRU. A large number of somatic mutations including single-, double-, and triple-nucleotide variants as well as insertions and deletions were found for both tumor cell lines (**Supplementary Table 1**). Many of these were found to be non-silent with potential to result in neoepitopes (**Table 1**). The amino acid sequences encoded by the mutated alleles, plus 9 preceding and tailing residues, were fed into the NetMHCpan 2.8 algorithm to predict HLA-A*02:01-binding epitopes allowing for 9- or 10-mers as output. The net result was 1961 candidate neoepitopes for KADA and 366 for ANRU (**Table 1**) with a predicted affinity of $K_d \leq 1\,\mu M$ for HLA-A2*02:01. The number of peptides to screen for ability to activate TIL was further restricted by including only peptides that had the highest predicted affinity of $K_d \leq 50$ nM for HLA-A2*02:01 and that originated from proteins that are expressed by melanoma cells (mRNA expression RPKM $\geq$1 in a transcriptome dataset of 7 melanoma cell lines, data not shown). This rendered 181 and 49 peptides that were used to stimulate KADA and ANRU TIL, respectively.

Recognition of Mutated Neoepitope Peptides by TIL

TIL were screened for induction of IFN-γ secretion against the 181 and 49 selected mutated peptides in pools of either 10 (KADA) or 5 (ANRU), and individual peptides from positive pools were subsequently identified. For KADA, this resulted in the identification of five peptides, containing mutations from genes KDELR2, CCT4, MYLIP, SVIL, and WDR75, which

were able to activate autologous TIL (**Figure 1B**, **Table 2**). All were 9-mer peptides, except SVIL that was a 10-mer. However, the corresponding 9-mer SVIL peptide, which was excluded due to predicted HLA-A2 binding affinity just below cutoff criteria (**Table 2**), was ordered later and shown to active TIL even better than the 10-mer (**Figure 1C**), pointing out the limitation of MHC-binding predictions. For ANRU, three peptides containing mutations that activated autologous TIL were identified (**Figure 1E**, **Table 2**). These peptides were 9- and 10-mers containing the same mutation from the gene ETV6 and a 9-mer containing a mutation from the gene NUP210. For both donors, the responses were decreased by blocking HLA-A2 or MHC-I on target cells.

Next, we interrogated whether TIL could distinguish between mutated and corresponding wild-type peptides. Of note, none of the wild-type peptides were able to activate TIL [**Figure 1C** (KADA) and **Figure 1F** (ANRU)], even if the predicted binding of the mutated and the wild-type peptide was very similar in most cases (**Table 2**). The predictions were confirmed by HLA-A2 stabilization assays performed on T2 cells (**Supplementary Figure 2**). Thus, recognition of neoepitopes is highly specific, and tolerance to wild-type antigens has not been broken. Furthermore, the discrimination between mutated and corresponding wild-type peptides from KADA and ANRU resides mainly on the TCR side of the MHC/peptide/TCR interaction.

Presentation of Mutated Peptides on Tumor Cell MHC-I by MS

Immune peptidomes for each of the tumor lines were obtained from affinity-purified (W6/32) peptide-HLA class I complexes. Employing Poisson LC-DIAMS, the immune peptidome MS/MS spectra were compared to the fragmentation patterns and elution positions for the 181 and 49 synthetic peptide candidate neoepitopes (**Table 1**).

For KADA, patterns were obtained for 136 of the 181 synthetic peptides with the 45 unobserved peptides either containing cysteine residues ($n = 23$) or being very hydrophobic ($n = 11$) or unobserved for undetermined reasons ($n = 11$). Two peptides from the predicted KADA neoepitopes could be detected (AGPS and ENC1; **Figures 2A,C**). The cysteine-containing CCT4 neoepitope recognized by TIL was detected using nanospray MS3 with Poisson detection (**Figure 2E**) (18). The KDELR2 neoepitope recognized by TIL was too hydrophobic for either our LC-MS or nanospray analysis and could not be analyzed. It is noteworthy that corresponding wild-type epitopes could also be detected for ENC1 (**Figure 2B**) and CCT4 (**Figure 2F**), but not so for AGPS (**Figure 2D**), which was calculated to be an HLA-A*0201 non-binder (**Table 2**, **Supplementary Figure 2A**).

For ANRU, fragmentation patterns were obtained for 38 of the 49 peptides in the ordered library with two of the unobserved peptides being rich for cysteines, seven being very hydrophobic, and two being undetected for unknown reasons. However, none of the candidate neoepitopes, or MART-1 could be detected on ANRU tumor cell MHC-I by MS (data not shown). The ETV6 10-mer was among those epitopes that were too hydrophobic to be detected.

A high-purity isotope-labeled MYLIP peptide was added to the KADA sample to determine if the failure to detect TIL-activating peptides by MS reflected insufficient sensitivity. To address the potential for sample handling losses prior to adding the quantitation peptide, HLA class I complexes were tracked by native Western blots throughout the affinity isolation procedure and shown to be efficiently captured (**Supplementary Figure 3A**). Spiking 100 attomoles of isotope-labeled mutant MYLIP neoepitope into peptide–HLA complexes from 500,000 KADA cells generated unambiguous Poisson detection (**Supplementary Figure 3B**) with a peak amplitude of 1,300 counts *per second* (cps). The reference MS and MS/MS spectra for the natural isotope MYLIP neoepitope can be scaled by the ion signal and elution profile and reinserted into the MS data, showing that the MYLIP peptide would be readily detected with only 10 copies (eight attomoles) per cell (**Supplementary Figure 3C**). This result argues against low sensitivity as a reason for failure to detect this peptide on the KADA tumor cell line. Also, the inability to detect the MART-1 peptide by MS was investigated by quantitative spiking a high purity but unlabeled sample of the native MART-1 9mer, AAGIGILTV. This MART-1 peptide, reported to be naturally processed and presented on HLA-A2, was added at 10 attomoles and gave similar detection sensitivity as the MYLIP peptide (**Supplementary Figure 3D**). However, AAGIGILTV is not predicted to bind HLA-A*02:01 well (**Table 2**) and has been shown to generate an unstable complex with a short half-life (21). If MART-1 complexes are generated at a high rate, they could activate TIL but may not survive the isolation procedure required for MS analysis.

To confirm that the genes containing the mutated peptides that were recognized by TIL, but not presented on the tumor cells, were indeed expressed by the tumor cells, mRNA levels of each of the genes (total expression of mutated and germline alleles) were compared to β-actin (**Table 2**). All of the mutated proteins that were recognized by TIL or that were presented by tumor cell MHC-I were clearly expressed. Furthermore, the RNA level of the mutated compared to the germline allele was analyzed, and as expected, both the mutated and the wild-type alleles were expressed (**Table 2**, **Supplementary Table 2**). Of note, however, the expression levels of all mutated genes were substantially lower than those of MART-1 and gp100 in ANRU cells, whose levels were more than 71-fold and 591-fold, respectively, higher than that of β-actin.

Assessment of the Frequency and Avidity of Neoepitope-Specific T Cells in TIL

To detect and measure the frequency of neoepitope-specific TIL within the total TIL population, neoepitope-specific PE-conjugated peptide/MHC dextramers were custom ordered. In KADA TIL (**Figure 3A**), stained populations could clearly be detected for KDLER2, MYLIP, and SVIL epitopes. Staining with CCT4 dextramers was very weak and, for WDR75 dextramers, virtually absent. Furthermore, dextramers were produced also for

TABLE 2 | Analysis of predicted HLA-A2 binding mutated or tumor associated peptides that trigger activation of tumor-infiltrating lymphocytes and/or are detected in pMHC complexes by MS.

Patient	Mutated epitope	Mutated position	Mutated sequence	Wild type sequence	Predicted HLA-A2 binding by mutated epitope (nM)	Predicted HLA-A2 binding by wild type epitope (nM)	Expression of mutant allele	Expression of gene compared to β-actin (fold)	Detection of mutant epitope on tumor cells by MS	Detection of wild type epitope on tumor cells by MS	Expression of gene compared to β-actin after IFNγ treatment (fold)	Detection of mutant epitope on tumor cells after IFNγ treatment	Detection of wild type epitope on tumor cells after IFNγ treatment
KADA	AGPS 9mer[#]	2	ALWDRVVDL	APWDRVVDL	18.12	18335.02	65%	3.28	Yes	No	4.14	Yes	No
	ENC1 9mer[#]	3	YLSELLQTV	YLPELLQTV	2.25	3.1	58%	0.72	Yes	Yes	1.47	Yes	Yes
	KDELR2 9mer	4	ILWIFSIYL	ILWTFSIYL	16.96	4.95	67%	8.07	ND	ND	11.39	ND	ND
	CCT4 9mer	1	FLLDSCTKL	SLLDSCTKL	4.31	23.02v	48%	12.45	Yes	Yes	22.79	Yes	Yes
	MYLIP 9mer	2	RLDAVLMEV	RPDAVLMEV	5.75	6090.18	52%	0.16	No	No	0.36	No	No
	SVIL 9mer	5	YLTDKDFEF	YLTDEDFEF	75.01	156.56	35%	0.62	No	No	0.43	No	No
	SVIL 10mer	5	YLTDKDFEFA	YLTDEDFEFA	13.53	14.11	35%	0.62	No	No	0.43	No	No
	WDR75 9mer	1	FMFVNSLLL	SMFVNSLLL	7.36	62.76	37%	5.18	No	No	9.67	No	No
	Flu M1 9mer	NA[£]	NA	GILGFVFTL	NA	15.03	NA	NA	NA	NA	NA	NA	NA
ANRU	ETV6 9mer	1	VLWDYVYQL	LLWDYVYQL	2.24	2.16	55%	0.92	No	No	3.68	Yes	Yes
	ETV6 10mer	1	VLWDYVYQLL	LLWDYVYQLL	4.8	4.34	55%	0.92	ND	ND	3.68	ND	ND
	NUP210 9mer	8	AIDAALTFV	AIDAALTSV	17.01	34.36	42%	1.78	No	No	2.17	No	No
	MART 10mer	2*	ELAGIGILTV	EAAGIGILTV	375.16	7627.98	NA	71.4	NA	No	28.91	NA	No
	MART 9mer	NA¤	NA	AAGIGILTV	NA	3448.53	NA	71.4	NA	No	28.91	NA	No
	gp100 9mer	2*	IMDQVPFSV	ITDQVPFSV	5.47	188.19	NA	591	NA	ND	175	NA	ND

[#]*Undetected by TIL activation.*

[£]*Viral peptide without mutations.*

[¤]*Unmutated peptide.*

[*]*Heteroclitic peptide with optimized aminoacid at mentioned position.*

NA, Not applicable; ND, Not determined.

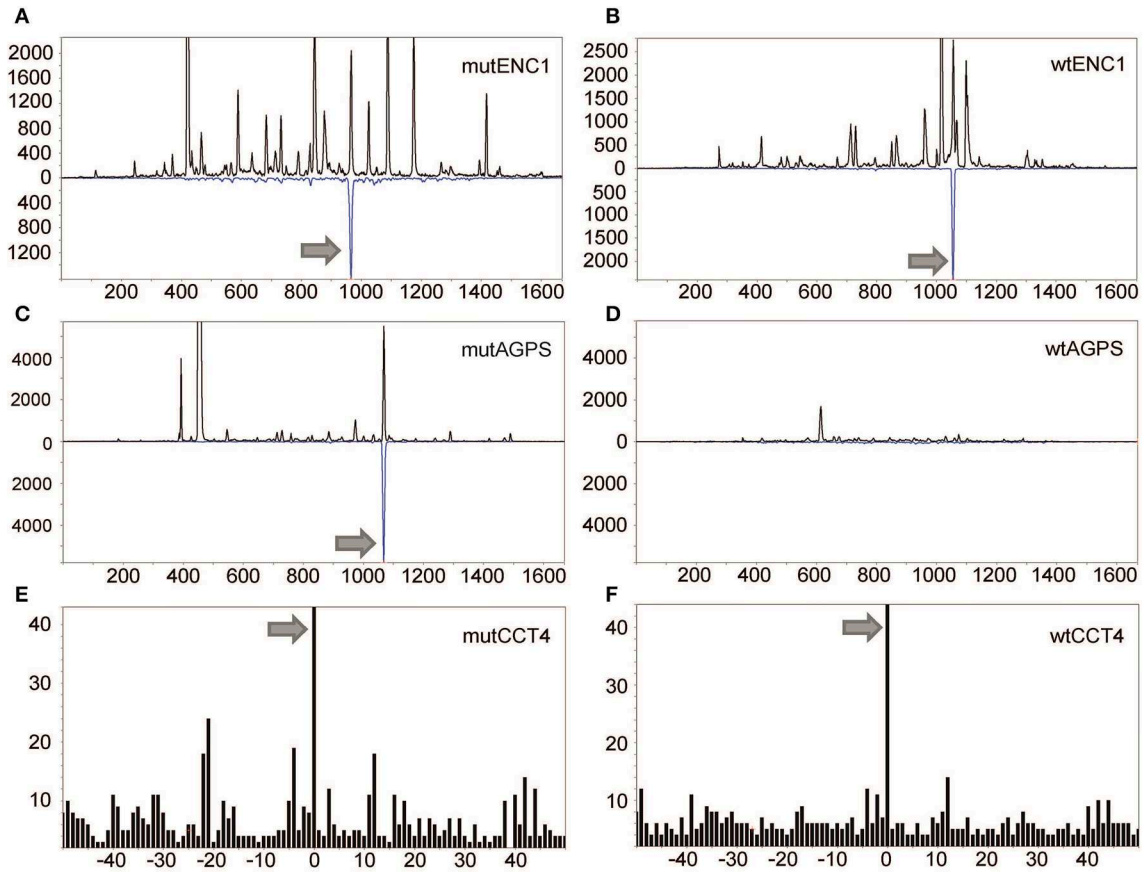


FIGURE 2 | Detection of mutated peptides presented on the tumor cell surface. LC-DIAMS Poisson detection plots for neoepitopes from mutated **(A)** and wild-type **(B)** ENC1 and mutated **(C)** and wild-type **(D)** AGPS from 500,000 cells. Top black traces are extracted ion chromatograms for *m/z* of the doubly charged precursor ion in units of counts *per second*. The inverted traces (blue) are Poisson chromatograms showing the number of events, scaled 10-fold (as a convenience in plotting), that can be embedded at fixed cutoff probability in the MS/MS spectrum of the DIA window containing the precursor *m/z* (17). Nanospray MS3 Poisson detection of cysteine-containing neoantigen peptide CCT4 **(E)** or the corresponding wild-type peptide **(F)** was performed from 1.5 million cells, as marked with a 0-offset Poisson peak (18).

AGPS and ENC1, the neoepitopes detected on KADA cell MHC-I exclusively by MS. In line with the TIL stimulation results, the AGPS dextramer did not bind TIL at all, while ENC1 showed a weak level of staining, impossible to separate from background staining. In ANRU TIL, on the other hand (**Figure 3A**), all the custom dextramer stainings (ETV6 9- and 10-mer, NUP210) resulted in well-defined populations, comparable to the staining seen with the MART-1 dextramer (**Supplementary Figure 1A**).

The weak CCT4 dextramer staining indicated that the interaction of the specific T cells with the MHC/peptide complex was of lower avidity. In addition, functional avidity of the specific T cells was assessed by titrating the peptides directly onto TIL and measuring the activation as IFN-γ secretion [**Supplementary Figure 3E** (ANRU) and **Supplementary Figure 3F** (KADA) and **Supplementary Table 3**]. These titration curves indicate that the activation of specific T cells by MHC-bound CCT4, KDLR2, SVIL 10-mer, ETV6 10-mer, and wild-type MART-1 peptides is of lower functional avidity that requires micromolar concentrations of peptide for activation.

Ability of Neoepitope-Specific T Cells to Recognize Autologous Tumor Cells

We next aimed to test the capacity of TIL-derived neoepitope-specific T cells to recognize the autologous tumors from which they were derived. We first attempted to enrich neoepitope-specific T cells from TIL by sorting neoepitope dextramer-stained TIL using anti-PE-coupled magnetic beads. The selected cells were subjected to a round of rapid expansion, as were unsorted TIL as a control. For KADA, only the population specific for KDELR2 and SVIL epitope could be effectively enriched by this approach (79 and 13% dextramer-stained cells in the sorted TIL vs. 1.1 and 0.3% in the unsorted, respectively). Although some enrichment was achieved also for the other neoepitope-specific T cell populations from KADA, the enriched TIL populations still contained a high proportion of dextramer-negative cells, probably due to the low starting frequency of

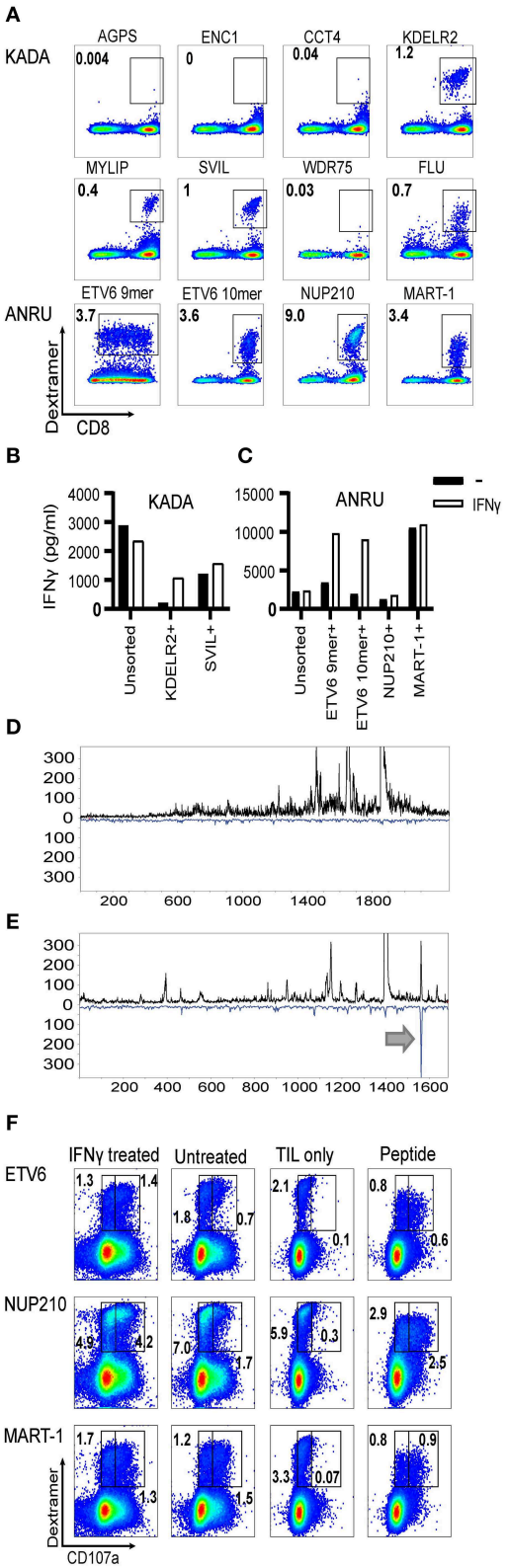


FIGURE 3 | Frequencies of neoantigen-specific T cells in TIL and assessment of their ability to recognize autologous tumor cells. HLA-A2/mutated peptide *(Continued)*

FIGURE 3 | dextramers were produced for peptides found to activate TIL and/or found to be presented on tumor cell MHC-I. The dextramers were used to stain tumor-infiltrating lymphocytes **(A)**, and anti-PE beads were used to enrich for stained cells followed by a rapid-expansion protocol. Thereafter, recognition of KADA **(B)** and ANRU **(C)** tumor cells by the sorted TIL was assessed by IFN-γ ELISA with or without IFN-γ pretreatment of the tumor. Only the specificities that could be significantly enriched by the sorting are shown. Unsorted cells were used as a control. The tumor cells were analyzed for neoepitope expression by MS and presented as LC-DIAMS Poisson detection plots for mutant ETV6 in untreated **(D)** and IFN-γ-treated ANRU tumor cells **(E)** with the arrow in the latter indicating detection. The total TIL population (**F**. ANRU) was co-cultured with untreated or IFN-γ pretreated autologous tumor cells, or with the different neoepitope peptides, analyzed by dextramer staining for the same epitope and for cell surface CD107a. As negative control, TIL alone were used and stained as described above. Dot plots are gated on lymphocytes/singlets/live cells and frequency indicates % dextramer$^+$ out of CD8$^+$ cells **(A)** and gated in the same way plus on CD8$^+$ and frequency indicates % dextramer$^+$CD07a$^-$ or dextramer$^+$CD107a$^+$ of CD8$^+$ cells **(F)**.

the dextramer-targeted TIL (<0.2%). For ANRU, enrichment could be performed for all neoepitope dextramers with 16–34% (NUP210:28%, ETV6 9-mer: 16%, ETV6 10-mer: 34%) dextramer-stained cells in sorted TIL vs. 2.1–4.8% (NUP210: 5.1%, ETV6 9-mer: 2.2%, ETV6 10-mer: 2.3%) in unsorted TIL. Sorted TIL were co-cultured with autologous tumor cells and analyzed for tumor recognition. All enriched neoepitope-specific T cells were also functionally enriched as they recognized their corresponding peptide with increased efficiency (data not shown). However, disappointingly, none of the enriched neoepitope-specific populations recognized tumor cells better than unsorted TIL [**Figure 3B** (KADA) and **Figure 3C** (ANRU)]. These results were in line with the MS results. However, MART-1-specific TIL were successfully enriched (63% dextramer-stained cells in sorted TIL vs. 3.6% in unsorted TIL) and recognized the autologous tumor better than unsorted TIL (**Figure 3C**). Nevertheless, the MART-1 epitope was not detected by MS on tumor MHC-I. The reason for this may be the short half-life of the MART-1/MHC complex as already discussed.

Analysis of Neoepitope Presentation on MHC-I and Activation of Specific TIL After IFN-γ Treatment of Tumor Cells

We next asked if the lack of neoepitope presentation by tumor MHC-I and the concurrent inability of neoepitope-specific TIL to recognize these cells could be due to an inefficient peptide processing and presentation machinery. IFN-γ-treated KADA and ANRU tumor cells showed markedly increased expression of MHC-I surface antigens by FACS [**Supplementary Figure 4A** (KADA) and **Supplementary Figure 4B** (ANRU)], and components of the peptide presentation machinery (APM) by Western blot (TAP1 and 2, Tapasin, MHC-I heavy chain, LMP2 and 10 and β2m; **Supplementary Figure 4C**) and RT-qPCR (data not shown). Therefore, we investigated if pre-treatment of tumor cells with IFN-γ would affect the expression of the mutated genes or the presentation of the neoepitopes on MHC-I. The mRNA expression levels in untreated vs. IFN-γ-treated cells did not display any dramatic upregulation

of any of the neoepitopes (**Table 2**) but interestingly there was a clear decrease of both MART-1 and gp100 expression after IFN-γ treatment. Although most of the neoepitopes were still not detected by MS, the 9-mer ETV6, which was not present on untreated ANRU cells (**Figure 3D**), could be detected by MS analysis on ANRU tumor cell MHC-I after IFN-γ treatment (**Figure 3E**). The same cytokine-inducible upregulation was true for the ETV6 9-mer wild-type peptide (**Table 2**). The ETV6 10-mer could, as mentioned before, not be analyzed by MS due to their hydrophobic nature.

Next, the ability of neoepitope-specific TIL to recognize IFN-γ-treated tumor cells was assessed. This was first interrogated using the dextramer-sorted neoepitope-specific TIL populations. For KADA, there was an increased recognition of IFN-γ-treated tumor cells compared to untreated tumor cells by TIL enriched by KDELR2 dextramers (**Figure 3B**), which, as mentioned, could not be analyzed by MS due to its hydrophobicity. For ANRU, in line with the MS results, the same was true for TIL enriched by ETV6 9- or 10-mer dextramers, while the MART-1-sorted TIL recognized untreated and IFN-γ-treated tumor cells to the same high extent (**Figure 3C**). Since the sorting of neoepitope specific TIL was less efficient for several epitopes, we next investigated the recognition of IFN-γ-treated tumor cells starting from the total TIL population. Unsorted ANRU TIL, which has a high frequency of neoepitope specific T cells, were co-cultured with IFN-γ-treated and -untreated tumor cells (**Figure 3A**). Neoepitope-specific T cells were identified with dextramers after co-culture. The activation was analyzed as degranulation by measuring CD107a surface expression and by comparing the brightness of the dextramer staining, since activation is known to lead to decreased levels of the TCR on the cell surface. In line with the MS-results, ETV6 dextramer-stained cells expressed CD107a and displayed decreased levels of dextramer-staining when cultured with IFN-γ-treated tumor cells (**Figure 3F**). In addition, and in contrast to both the MS results and the results with the dextramer-sorted cells, in this experiment, NUP210-dextramer-positive cells recognized the tumor demonstrated by tumor-induced activation of degranulation (**Figure 3F**).

To better characterize the relation between MS sensitivity and TIL response, IFN-γ-treated ANRU cells were loaded with NUP210 peptide at a lower concentration than was required to activate IFN-γ production by TIL (**Supplementary Figures 3G,H**). MS sensitivity was substantially greater than biological assay as detailed therein.

Generation of Neoepitope-Specific T Cells From Blood

Since we found two neoepitopes, from mutated AGPS and ENC1, that were expressed on ANRU tumor MHC-I, but that were not recognized by TIL, we wanted to determine whether these epitopes were selectively non-immunogenic for the patient, broadly non-immunogenic, or, alternatively, if we could expand neoepitope-specific T cells from healthy donors. To test these possibilities, CD8+ T cells derived from blood from patient KADA (**Figure 4A**) or three healthy HLA-A2+ donors (**Figures 4B–D**) or from patient ANRU (**Figure 4E**) and three other HLA-A2+ healthy donors (**Figures 4F–H**) were stimulated with autologous monocyte-derived DC pulsed with peptides. The CD8+ T cells were screened for ability to recognize each peptide that had either been found to activate TIL or to be presented on the tumor cell MHC-I. Of note, for KADA, all of the peptide-stimulated PBMC cultures had detectable neoepitope specific CD8+ T cells, shown by dextramer-positive populations. For KDELR2 and WDR75, a high frequency of strongly stained cells was detected, and for AGPS, ENC1, MYLIP, and SVIL, the staining was strong and distinct, although the frequencies of stained cells were low. For CCT4, however, few cells were stained and the staining was also weak, suggesting that cells had expanded but were of lower affinity, similarly to the situation in KADA TIL. In addition, in the healthy donors, most of the neoepitope-specific T cell populations could be expanded in at least one donor, except for WDR75 (**Figures 4B–D** and data not shown). In contrast, in mock stimulated (unpulsed DC) cultures, there was no staining with any of the dextramers for any of the individuals. For ANRU, all the peptides expanded specific cells in ANRU PBMC, resulting in clearly defined dextramer-stained populations. In all three healthy donors, both NUP210 and MART-1 stimulations resulted in clear dextramer-positive populations (**Figures 4F–H**).

Thus, all the mutated peptides were found to be immunogenic, even for T cells derived from the blood of both donors (KADA and ANRU) from which the autologous tumor cell lines were derived. Importantly, the blood was drawn from these patients long after the tumor was removed, and they were both cancer free at the time. Therefore, it is not surprising that there were no tumor-specific cells found in mock-stimulated blood. Of particular importance, however, even the MS-defined neoepitopes that were unable to activate the autologous KADA TIL (AGPS and ENC1) could expand T cells from both KADA PBMC and from HLA-A2+ donor PBMC.

We also investigated if the neoepitope specific CD8+ T cells, derived from ANRU PBMC or from the three healthy HLA-A2+ donors, could recognize the ANRU tumor cells. Therefore, the stimulated CD8+ (DC pulsed with peptides) were re-stimulated with ANRU tumor cells and recognition was measured by FACS as increased CD107a expression/degranulation [**Figure 4I** (ANRU) and **Figures 4J–L** (healthy donors)]. A portion of the ANRU neoepitope and MART-1-specific CD8+ T cells recognized the autologous tumor observed by dextramer and CD107a double-positive cells. All three healthy donors' CD8+ T cells stimulated either with DC pulsed with the NUP210 or the MART-1 epitope could also recognize the ANRU tumor. As a positive control, for each neoepitope, stimulated CD8+ cells were re-stimulated with the peptide, which resulted in activation of CD8+ T cells for each epitope (data not shown).

DISCUSSION

Infiltrating CD8+ T cells mediate the predominant immune response in malignant melanoma (22). Accordingly, infusion of in vitro-expanded TIL which are dominated by CD8+

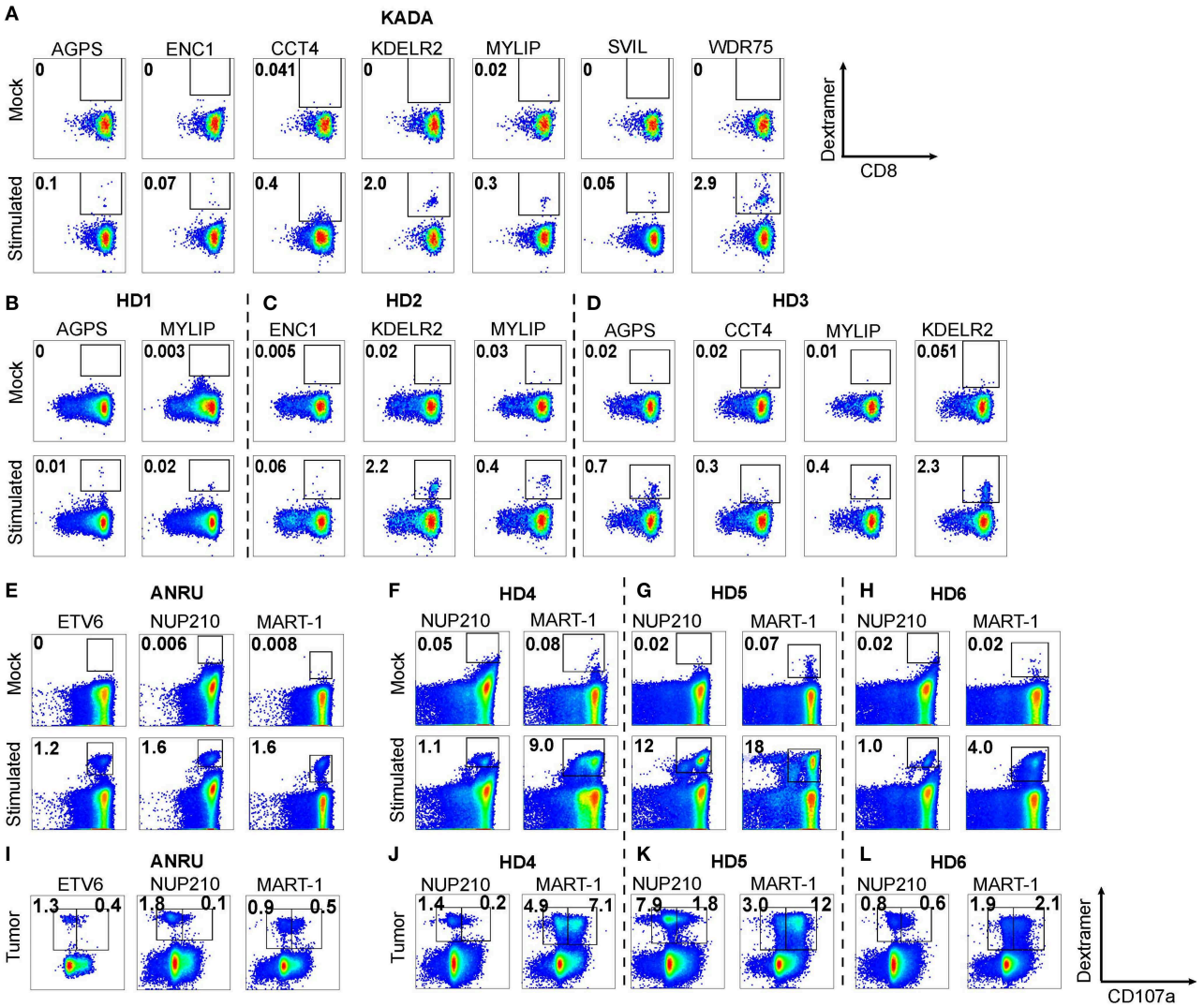


FIGURE 4 | Expansion of neoepitope peptide-specific T cells from PBMC. Dendritic cells from KADA **(A)** or ANRU **(E)** were loaded with their respective TIL-activating and/or tumor-presented peptides and used to stimulate autologous CD8 T cells. The same experiment was performed using three healthy donors for each peptide set (**B–D**, KADA neoepitopes; **F–H**, ANRU neoepitopes and MART-1). After 10 days of expansion, the cells were stained with corresponding HLA-A2/peptide dextramers and analyzed by flow cytometry. Dot plots are gated on lymphocytes/singlets/live cells and frequency indicates %dextramer$^+$ out of CD8$^+$ cells. For ANRU epitopes, the function of the neoepitope specific T cells was assessed by re-stimulation with ANRU tumor cells and evaluated by CD107a staining (**I**, ANRU; **J–L**, healthy donors). Only positive stainings are shown. Dot plots are gated on lymphocytes/singlets/live cells/CD8$^+$ cells and frequency indicates % dextramer$^+$CD07a$^-$ or dextramer$^+$CD107a$^+$ of CD8$^+$ cells.

T cells commonly produces complete and long-lasting regressions of metastatic lesions (1). There is, however, limited information on the precise specificity of the tumor-derived T cells that mediate tumor rejection in melanoma patients, and the extent to which important tumor epitopes are derived from broadly expressed shared tumor antigens or from private mutated tumor epitopes. We followed two parallel but distinct strategies to discover immunogenic neoeptitopes: one based on peptide reactivity of T cells isolated from the patient's tumor, the other on physical detection of putative neoepitopes presented on the surface of the tumor cells.

We performed whole exome sequencing of early passage tumor cell lines originated from two stage III/IV metastatic melanoma patients, to identify mutated epitopes based on *in silico* predicted HLA-A2 binding and expected expression in melanoma. Similar to others (23), we focused on MHC-I HLA-A02:01 restricted epitopes, motivated by more developed and precise algorithms for predicting T-cell epitopes for MHC-I and in particular for HLA-A02:01. In addition, Swedish patients with advanced melanoma have a higher prevalence for HLA-A02:01, in particular in those patients with a poor prognosis (24).

The TIL were derived from two long-term survivors (KADA and ANRU), indicating an ongoing immune response. Their

TIL demonstrated strong reactivity against autologous tumor lines, providing us with an efficient tool for screening the predicted neoepitope library, in a similar fashion to what has been performed before by others (23). This resulted in the identification of a total of eight neoantigen epitopes (five from KADA and three from ANRU) with the ability to activate IFN-γ release and degranulation by autologous TIL. The finding that only 8 out of 230 mutated neoepitopes predicted to bind MHC-I were recognized by the patients' TIL confirms similar observations from others (23, 25).

One might expect that the presence in TIL of T cell clones specific for these eight epitopes, all originating from proteins that were highly expressed at the mRNA level in the tumor cells, would signify that those same epitopes should be presented on MHC-I on the surface of the tumor cells. This was, however, not the case, as shown by MS analysis of those peptides that were actually presented in association with MHC-I on tumor cells. In fact, the 10,000-fold greater rate of cellular protein turnover vs. pMHC turnover mandates that few, if any, peptides derived from a given protein are expressed on the cell surface complexed to MHC-I (26).

Conventionally, discrepancies between T cell recognition and MS detection are attributed to poor MS sensitivity. The lack of false negatives associated with the data-independent acquisition format (see **Figures 2**, **3** and **Supplementary Figure 3**), the tracking of MHC-I complexes during the affinity isolation, the calibration of MS sensitivity into the single attomole (10^{-18} mole) level by the addition of internal standards, and the recovery and MS detection peptides, loaded onto tumor cells at concentrations below TIL recognition, are however hard to reconcile with neoepitope detection failing due to MS sensitivity. This said, peptide–MHC-I complex instability and the reduced sensitivity with cysteine-containing or very hydrophobic peptides are limitations of our current MS methodology. The approach of identifying peptides that activate T cells by functional assays and in parallel detecting peptides presented on tumor cell MHC-I by MS provided us with a unique methodological comparison. Several important conclusions can be drawn from this comparison, pointing at strengths and limitations of each of these two methods.

First, we can conclude that our MS approach allows detection of three HLA-A2 restricted neoepitopes presented on the melanoma cells (AGPS, ENC1, CCT4; **Table 2**). For one of these (CCT4 from the KADA tumor), we were able to show recognition by the autologous TIL and expansion of specific T cells from patient peripheral blood. In spite of this, we were unable to specifically sort out the T cell clones recognizing the neoepitope CCT4, most likely explained by poor binding to the peptide/MHC dextramer resulting in difficulties in enriching this T cell population, leaving us with no possibility to prove the expected tumor recognition by these CCT4 specific T cells.

For the two other epitopes identified by MS as presented on KADA tumor MHC-I (AGPS, ENC1), we could unequivocally demonstrate that these epitopes were not recognized by the autologous TIL. Since these epitopes were demonstrated to be immunogenic after culturing PBMC with peptide-loaded DCs, one explanation for the absence of reactivity is a low level of antigen. We expect that DC cross-presentation of phagocytosed cellular debris cannot display a substantial fraction of MHC-I-binding peptides from the tumor's full proteome and the likelihood of cross-presenting DC activating neoepitope-specific CD8 T cells decreases as the abundance of the mutant source protein decreases. When abundance is under the threshold for priming and expansion by cross-presenting DCs in draining lymphoid tissues, the full repertoire of circulating and lymphoid-resident T cells may not be activated, expanded, and deployed for neoepitope recognition. Lack of epitope recognition may also reflect immunodominance, where immune responses target only a few antigenic peptides of the many displayed, thereby curtailing natural responses against non-dominant epitopes and/or be a consequence of ineffective cross-presentation of the epitopes (27–29). There is no obvious dominance of KADA TIL responses, however. Peripheral tolerance mechanisms mediated by MDSC and regulatory T cells in melanoma patients may also limit the ability of TIL to respond to these antigens (30). Heterogeneity of neoantigen expression, resulting in T cells reactive with individual, sub-clonal mutations (presence in only a subset of tumor cells) and not with clonal mutations (presence in all tumor cells) may also explain the absence of reactivity against certain epitopes (31).

Immunogenicity of a neoepitope has been reported to arise both due to changes in anchor residues (32) and TCR contacts (33). In our study, the majority of neoepitopes that could activate TIL had not gained binding affinity to HLA-A2 compared to the corresponding wild-type peptide; only for the AGPS and MYLIP peptides from the KADA tumor was an increased binding observed, as confirmed by HLA-A2 stabilization assays. Also, the corresponding wild-type epitope could not activate TIL, indicating that tolerance to the wild-type antigen has not been broken. These findings are therefore in line with those of Fritsch and collaborators (34), who found mutations located in the TCR-facing residues of the neoepitopes rather than the anchor residues when analyzing 40 neoepitopes of human cancers that induced immune responses associated with regression or long-term disease stability.

As expected (15), IFN-γ resulted in increased levels of MHC-I on the tumor surface and of various components of the MHC-I APM. In addition, MHC presentation of a neoepitope derived from the protein ETV6 was detected by MS in IFN-γ-treated, but not in untreated ANRU tumor cells. This may be explained either by the observed total increase in MHC-I, or alternatively and more likely by an altered peptide repertoire induced by IFN-γ-mediated increase in the expression of the immune proteasome (LMP2 and 10) which we found to have a markedly enhanced expression. In line with the ETV6 peptide being presented, we could also detect an enhanced recognition of the tumor cells by dextramer-sorted TIL specific for this peptide compared to unsorted TIL. Such a cytokine-inducible epitope display is consistent with the importance of an intact antigen presentation and IFN-γ signature for immunotherapy to be efficient (35). Also, IFN-γ-treated KADA tumor cells, but not the untreated tumor cells, showed increased stimulation of TIL enriched for T cells specific for the KDELR2 epitope. We were, however, unable to confirm with MS the MHC-I presentation of the KDELR2 epitope

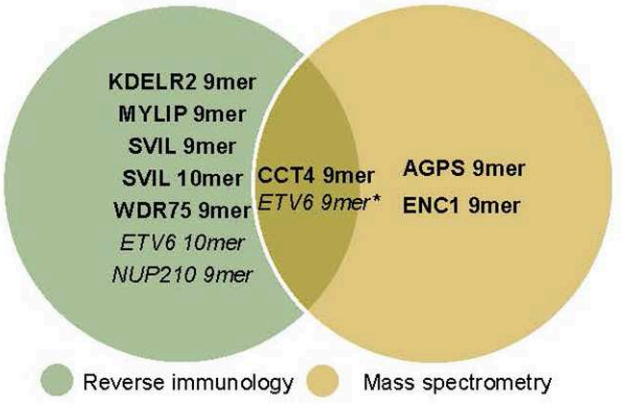


FIGURE 5 | Neoepitopes detected by reverse immunology vs. mass spectrometry. With reverse immunology TIL reactivity against 5/181 (**KADA**) and 3/49 (*ANRU*) of the predicted neoepitopes were observed. With mass spectrometry 3/136 (**KADA**, not all peptides could be analyzed) and 1*/38 (*ANRU*, not all peptides could be analyzed) of the predicted neoepitopes were detected on the cell surface. **KADA** epitopes are in Bold, *ANRU* epitopes are in Italic. *Detected after IFNγ treatment of the tumor.

on IFN-γ-treated tumor cells due to the hydrophobic nature of this epitope. IFN-γ-treated tumor cells were also recognized by TIL specific for the neoepitope NUP210, but the activation only resulted in degranulation, not in IFN-γ production.

Several other possible explanations may account for the discrepancies between TIL recognition of tumor cells and MS detection of the peptides on tumor MHC-I, which are summarized in **Figure 5**. A better understanding of the basis of these divergent results are essential for future developments of optimal methods for clinical application. Since TIL were blocked by MHC-I antibody, our data argue strongly against non-specific activity and favor the interpretation that TIL are activated by an epitope/MHC-I complex. One possibility is that part of the results could be explained by cross-reactivity of TCR on the TIL with unidentified peptides on the surface of the tumor cells. The high concentrations (μM) of peptides used for screening in the absence of external bioforces that occur *in vivo* and normally tune TCR recognition, fostering antigen discrimination and low copy number pMHC activation of T cells, encourage such TCR cross-reactivity (36, 37).

Another likely explanation for discrepancies between TIL screening result and the MS method lies in limitation in terms of sensitivity for our MS method for known physicochemical challenging peptides and very weak binders. This limitation is likely to explain our difficulties to detect by MS two of the TIL detected epitopes (KDELR2 and MART-1), and limits the usefulness of the MS detection method for such peptides. That said, those outliers are readily flagged as problematic prior to MS analysis. MART-1 is of particular interest, since this epitope is extensively used as a prototype for tumor antigen, also in clinical trials with TCR-modified T cells (38). While MART-1 transcriptional rates are extremely high, the wtMART-1 peptide has orders of magnitude lower affinity for HLA-A2 than the heteroclitic MART-1 counterpart and an unusual TCR binding to MART-1 peptide/HLA-A2 complexes has been documented (39). The latter will foster TCR cross-reactivity, likely also accounting for discordance in TAA heteroclitic MART-1 binding to TIL and their stimulation in the absence of MS detection of native MART-1 on the ANRU tumor cells.

The advantage of the MS approach is that a positive identification can give unequivocal proof for the MHC-I presentation on the patient's tumor, even when the patient's TIL are not available or are not reactive. This is exemplified by the AGPS and ENC1 neoepitopes that were detected by MS as being presented on the tumor lines. When analyzed for immunogenicity by stimulating T cells from PBMC with autologous monocyte-derived DC pulsed with peptide, our results clearly showed that specific T cell populations could be expanded from patient (KADA and ANRU) or normal donor PBMC using autologous monocyte-derived DC. In addition, a population of these specific T cells could recognize and respond with degranulation upon re-stimulation. That only a part of the expanded cells responded could be due to lower affinity of neoepitope-specific T cells derived from blood (40, 41). Notwithstanding, these findings have important therapeutic implications, and clinical applications where MS detected neoepitope-specific T cells are expanded from blood by either *in vitro* stimulation or by vaccination hold great promise for clinical developments. In our ongoing clinical trial (NCT01946373), we are applying a combination of ACT with TIL and a tumor vaccine composed of autologous tumor lysate pulsed monocytic DCs. The results above have motivated us to consider extending our clinical trial to involve ACT with autologous TIL or peripheral blood enriched for neoepitope-specific T cells, followed by a boost of this neoepitope-specific response by a DC tumor vaccine derived from the same mutated epitopes. This type of approach has recently been spurred by results from others, demonstrating T cell activation following administration of DC-, peptide-, or RNA-based tumor antigens to cancer patients (42, 43).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Regionala etikprövingsnämnden i Stockholm, Sweden. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SW and TL designed the experiments involving neoantigen and tumor cell line recognition by TIL. BR and ER designed the experiments regarding analysis using mass spectrometry. BS has designed experiments regarding analysis of APM components. MV, IP, and RO planned and coordinated next-generation sequencing. MV performed NGS data analysis and prediction

of neoepitopes. SW, TL, MV, BR, JD-C, LH, JR, AM, JM, and RM has performed laboratory work, generated data, and performed data analysis. ML and JH have been involved in discussions. JH recruited the patients. SW, TL, BR, ER, and RK writing of the manuscript. ER and RK have supervised the study.

FUNDING

The study was supported by grants to RK from the Swedish Cancer Society CAN2016/315, the Cancer Society in Stockholm 164073, the Swedish Medical Research Council 2016-01414, Stockholm City Council Project Grant 201700452, and the Knut and Alice Wallenberg Foundations (KAW2015.0063 and 2013.0093), and a grant to the Reinherz laboratory from the Dunkin' Donuts/NDCP Drives Cancer Breakthroughs program. RO is a KH-Bauer-Stiftung-endowed professor at the German Cancer Research Center (DKFZ), Heidelberg.

ACKNOWLEDGMENTS

We thank the High Throughput Sequencing unit of the Genomics & Proteomics Core Facility, German Cancer Research Center (DKFZ), for providing excellent exome sequencing services and the HUSAR team (W610) for providing bioinformatics infrastructure and support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02766/full#supplementary-material

REFERENCES

1. Feldman SA, Assadipour Y, Kriley I, Goff SL, Rosenberg SA. Adoptive cell therapy–tumor-infiltrating lymphocytes, T-cell receptors, and chimeric antigen receptors. *Semin Oncol*. (2015) 42:626–39. doi: 10.1053/j.seminoncol.2015.05.005
2. Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, et al. Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science*. (2006) 314:126–9. doi: 10.1126/science.1129003
3. Morgan RA, Chinnasamy N, Abate-Daga D, Gros A, Robbins PF, Zheng Z, et al. Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. *J Immunother*. (2013) 36:133–51. doi: 10.1097/CJI.0b013e3182829903
4. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. (2015) 372:2509–20. doi: 10.1056/NEJMoa1500596
5. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. (2014) 371:2189–99. doi: 10.1056/NEJMoa1406498
6. Huang J, El-Gamil M, Dudley ME, Li YF, Rosenberg SA, Robbins PF. T cells associated with tumor regression recognize frameshifted products of the CDKN2A tumor suppressor gene locus and a mutated HLA class I gene product. *J Immunol*. (2004) 172:6057–64. doi: 10.4049/jimmunol.172.10.6057
7. Zhou J, Dudley ME, Rosenberg SA, Robbins PF. Persistence of multiple tumor-specific T-cell clones is associated with complete tumor regression in a melanoma patient receiving adoptive cell transfer therapy. *J Immunother*. (2005) 28:53–62. doi: 10.1097/00002371-200501000-00007
8. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med*. (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
9. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, Dudley ME, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science*. (2014) 344:641–5. doi: 10.1126/science.1251102
10. Lövgren T, Sarhan D, Truxová I, Choudhary B, Maas R, Melief J, et al. Enhanced stimulation of human tumor-specific T cells by dendritic cells matured in the presence of interferon-gamma and multiple toll-like receptor agonists. *Cancer Immunol Immunother*. (2017) 66:1333–44. doi: 10.1007/s00262-017-2029-4
11. Poschke I, Lövgren T, Adamson L, Nyström M, Andersson E, Hansson J, et al. A phase I clinical trial combining dendritic cell vaccination with adoptive T cell transfer in patients with stage IV melanoma. *Cancer Immunol Immunother*. (2014) 63:1061–71. doi: 10.1007/s00262-014-1575-2
12. Pawlikowski JS, McBryan T, van Tuyn J, Drotar ME, Hewitt RN, Maier AB, et al. Wnt signaling potentiates nevogenesis. *Proc Natl Acad Sci USA*. (2013) 110:16009–14. doi: 10.1073/pnas.1303491110
13. Capell BC, Drake AM, Zhu J, Shah PP, Dou Z, Dorsey J, et al. MLL1 is essential for the senescence-associated secretory phenotype. *Genes Dev*. (2016) 30:321–36. doi: 10.1101/gad.271882.115
14. Radonić A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun*. (2004) 313:856–62. doi: 10.1016/j.bbrc.2003.11.177
15. Respa A, Bukur J, Ferrone S, Pawelec G, Zhao Y, Wang E, et al. Association of IFN-gamma signal transduction defects with impaired HLA class I antigen processing in melanoma cell lines. *Clin Cancer Res*. (2011) 17:2668–78. doi: 10.1158/1078-0432.CCR-10-2114
16. Keskin DB, Reinhold B, Lee SY, Zhang G, Lank S, O'Connor DH, et al. Direct identification of an HPV-16 tumor antigen from cervical cancer biopsy specimens. *Front Immunol*. (2011) 2:75. doi: 10.3389/fimmu.2011.00075
17. Keskin DB, Reinhold BB, Zhang GL, Ivanov AR, Karger BL, Reinherz EL. Physical detection of influenza A epitopes identifies a stealth subset on human lung epithelium evading natural CD8 immunity. *Proc Natl Acad Sci USA*. (2015) 112:2151–6. doi: 10.1073/pnas.1423482112
18. Reinhold B, Keskin DB, Reinherz EL. Molecular detection of targeted major histocompatibility complex I-bound peptides using a probabilistic measure and nanospray MS3 on a hybrid quadrupole-linear ion trap. *Anal Chem*. (2010) 82:9090–9. doi: 10.1021/ac102387t
19. Ivanov AR, Zang L, Karger BL. Low-attomole electrospray ionization MS and MS/MS analysis of protein tryptic digests using 20-microm-i.d. polystyrene-divinylbenzene monolithic capillary columns. *Anal Chem*. (2003) 75:5306–16. doi: 10.1021/ac030163g
20. Betts MR, Brenchley JM, Price DA, De Rosa SC, Douek DC, Roederer M, et al. Sensitive and viable identification of antigen-specific CD8+ T cells by a flow cytometric assay for degranulation. *J Immunol Methods*. (2003) 281:65–78. doi: 10.1016/S0022-1759(03)00265-5
21. Valmori D, Fonteneau JF, Lizana CM, Gervois N, Liénard D, Rimoldi D, et al. Enhanced generation of specific tumor-reactive CTL in vitro by selected Melan-A/MART-1 immunodominant peptide analogues. *J Immunol*. (1998) 160:1750–8. doi: 10.1016/S0165-2478(97)85892-0
22. Spranger S, Spaapen RM, Zha Y, Williams J, Meng Y, Ha TT, et al. Up-regulation of PD-L1, IDO, and T(regs) in the melanoma tumor microenvironment is driven by CD8(+) T cells. *Sci Transl Med*. (2013) 5:200ra116. doi: 10.1126/scitranslmed.3006504
23. Robbins PF, Lu YC, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med*. (2013) 19:747–52. doi: 10.1038/nm.3161
24. De Petris L, Bergfeldt K, Hising C, Lundqvist A, Tholander B, Pisa P, et al. Correlation between HLA-A2 gene frequency, latitude, ovarian and prostate cancer mortality rates. *Med Oncol*. (2004) 21:49–52. doi: 10.1385/MO:21:1:49

25. Strønen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science*. (2016) 352:1337–41. doi: 10.1126/science.aaf2288
26. Yewdell JW. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. *Trends Immunol*. (2011) 32:548–58. doi: 10.1016/j.it.2011.08.001
27. Bräunlein E, Krackhardt AM. Identification and characterization of neoantigens as well as respective immune responses in cancer patients. *Front Immunol*. (2017) 8:1702. doi: 10.3389/fimmu.2017.01702
28. Akram A, Inman RD. Immunodominance: a pivotal principle in host response to viral infections. *Clin Immunol*. (2012) 143:99–115. doi: 10.1016/j.clim.2012.01.015
29. Kastenmuller W, Gasteiger G, Gronau JH, Baier R, Ljapoci R, Busch DH, et al. Cross-competition of CD8+ T cells shapes the immunodominance hierarchy during boost vaccination. *J Exp Med*. (2007) 204:2187–98. doi: 10.1084/jem.20070489
30. Mao Y, Poschke I, Wennerberg E, Pico de Coaña Y, Egyhazi Brage S, Schultz I, et al. Melanoma-educated CD14+ cells acquire a myeloid-derived suppressor cell phenotype through COX-2-dependent mechanisms. *Cancer Res*. (2013) 73:3877–87. doi: 10.1158/0008-5472.CAN-12-4115
31. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. (2016) 351:1463–69. doi: 10.1126/science.aaf1490
32. Duan F, Duitama J, Al Seesi S, Ayres CM, Corcelli SA, Pawashe AP, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J Exp Med*. (2014) 211:2231–48. doi: 10.1084/jem.20141308
33. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*. (2014) 515:572–6. doi: 10.1038/nature14001
34. Fritsch EF, Rajasagi M, Ott PA, Brusic V, Hacohen N, Wu CJ. HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunol Res*. (2014) 2:522–9. doi: 10.1158/2326-6066.CIR-13-0227
35. Sade-Feldman M, Jiao YJ, Chen JH, Rooney MS, Barzily-Rokni M, Eliane JP, et al. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat Commun*. (2017) 8:1136. doi: 10.1038/s41467-017-01062-w
36. Feng Y, Reinherz EL, Lang MJ. alphabeta T Cell Receptor Mechanosensing Forces out Serial Engagement. *Trends Immunol*. (2018) 39:596–609. doi: 10.1016/j.it.2018.05.005
37. Feng Y, Brazin KN, Kobayashi E, Mallis RJ, Reinherz EL, Lang MJ. Mechanosensing drives acuity of alphabeta T-cell recognition. *Proc Natl Acad Sci USA*. (2017) 114:E8204–13. doi: 10.1073/pnas.1703559114
38. Romero P, Valmori D, Pittet MJ, Zippelius A, Rimoldi D, Lévy F, et al. Antigenicity and immunogenicity of Melan-A/MART-1 derived peptides as targets for tumor reactive CTL in human melanoma. *Immunol Rev*. (2002) 188:81–96. doi: 10.1034/j.1600-065X.2002.18808.x
39. Borbulevych OY, Insaidoo FK, Baxter TK, Powell DJ, Johnson LA, Restifo NP, et al. Structures of MART-126/27–35 Peptide/HLA-A2 complexes reveal a remarkable disconnect between antigen structural homology and T cell recognition. *J Mol Biol*. (2007) 372:1123–36. doi: 10.1016/j.jmb.2007.07.025
40. Bobisse S, Genolet R, Roberti A, Tanyi JL, Racle J, Stevenson BJ, et al. Sensitive and frequent identification of high avidity neo-epitope specific CD8 (+) T cells in immunotherapy-naive ovarian cancer. *Nat Commun*. (2018) 9:1092. doi: 10.1038/s41467-018-03301-0
41. Tanyi JL, Bobisse S, Ophir E, Tuyaerts S, Roberti A, Genolet R, et al. Personalized cancer vaccine effectively mobilizes antitumor T cell immunity in ovarian cancer. *Sci Transl Med*. (2018) 10:eaao5931. doi: 10.1126/scitranslmed.aao5931
42. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*. (2017) 547:217–21. doi: 10.1038/nature22991
43. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. (2017) 547:222–6. doi: 10.1038/nature23003

Conflict of Interest: RK is a board member/consultant for Clinical Laser Thermia Systems AB and is also a Scientific Advisor for Anocca AB and Phion Pharmaceutics and receives research grants from these 2 companies. He has also been paid for organizing courses in Immunotherapy for BMS Sweden.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 24 January 2020
doi: 10.3389/fimmu.2020.00027

Computational Prediction and Validation of Tumor-Associated Neoantigens

Vladimir Roudko[1,2], *Benjamin Greenbaum*[2,3,4] *and Nina Bhardwaj*[1*]

[1] *Department of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai Hospital, New York, NY, United States,* [2] *Center for Computational Immunology, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai Hospital, New York, NY, United States,* [3] *Department of Pathology, Icahn School of Medicine at Mount Sinai Hospital, New York, NY, United States,* [4] *Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai Hospital, New York, NY, United States*

Tumor progression is typically accompanied by an accumulation of driver and passenger somatic mutations. A handful of those mutations occur in protein coding genes which introduce non-synonymous polymorphisms. Certain substitutions may give rise to novel, tumor-associated antigens or neoantigens, presentable by cancer cells to the host adaptive immune system. As antigen recognition is the core of an effective immune response, the identification of patient tumor specific antigens derived from transformed cells is of importance for immunotherapeutic approaches. Recent technological advances in DNA sequencing of tumor genomes, advances in gene expression analysis, algorithm development for antigen predictions and methods for T-cell receptor (TCR) repertoire sequencing have facilitated the selection of candidate immunogenic neoantigens. In this regard, multiple research groups have reported encouraging results of neoantigen-based cancer vaccines that generate tumor antigen specific immune responses, both in mouse models and clinical trials. Additionally, both the quantity and quality of neoantigens has been shown to have predictive value for clinical outcomes in checkpoint-blockade immunotherapy in certain tumor types. Neoantigen recognition by vaccination or through adoptive T cell therapy may have unprecedented potential to advance cancer immunotherapy in combination with other approaches. In our review we discuss three parameters regarding neoantigens: computational methods for epitope prediction, experimental methods for epitope immunogenicity validation and future directions for improvement of those methods. Within each section, we will describe the advantages and limitations of existing methods as well as highlight pressing fundamental problems to be addressed.

Keywords: neoantigen, TCR, WES, HLA-allele, MHC-I epitope

OPEN ACCESS

Edited by:
Peter Brossart, University of Bonn, Germany

Reviewed by:
Zsolt Sebestyen, University Medical Center Utrecht, Netherlands
Juliane Sarah Walz, Tübingen University Hospital, Germany

***Correspondence:**
Nina Bhardwaj
nina.bhardwaj@mssm.edu

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *23 August 2019*
Accepted: *08 January 2020*
Published: *24 January 2020*

Citation:
Roudko V, Greenbaum B and Bhardwaj N (2020) Computational Prediction and Validation of Tumor-Associated Neoantigens. Front. Immunol. 11:27. doi: 10.3389/fimmu.2020.00027

INTRODUCTION

Successful targeting of immune checkpoints including cytotoxic T lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1) has achieved durable regressions in a wide range of human cancers (referred to as checkpoint blockade). They include melanoma (1, 2), renal cell carcinoma (3), lung (4), bladder (5), and ovarian cancers (6), and microsatellite

unstable malignancies (7, 8). Despite different mechanisms of action, both approaches have resulted in the activation and proliferation of tumor-reactive T cells (9). T cells recognize peptides presented on the major histocompatibility complex (MHC) of tumor cells. Tumor specific antigens which arise due to mutations in coding regions are collectively referred to as "neoantigens." Neoantigens have a diversity of properties. They can differ from their wild type sequences by SNV, relative expression levels in the tumor, MHC affinity, differential recognition by T-cell receptors (TCRs) and elicitation of heightened cytotoxic and cytokine responses. Theoretically, T cells recognizing neoantigens may have not been deleted or tolerized so they have the potential to become primed. Moreover, unlike tumor-associated antigens (TAA) that are shared between tumor cells and normal tissue (e.g., Melan A/MART-1) neoantigens have a selective potential to elicit tumor exclusive T cell responses which makes them key elements for inclusion in cancer vaccines and as the basis for adoptive T cell transfer approaches (10–14). Indeed, initial attempts to target overexpressed TAA have met limited success in clinical trials potentially due to central and peripheral tolerance mechanisms which removes high-affinity TCRs that would otherwise potently recognize these TAA (15, 16). Unleashing immune responses against tumor-specific clonal mutations can achieve tumor regression through recognition by antigen-specific T cells (17–22). Furthermore, as a tumor's mutational landscape evolves with ongoing immunotherapy, the immune system may accommodate by changing the specificity of infiltrating T cell clones (23–25). Efficient approaches to identify and characterize immunogenic tumor neoantigens are central for these types of therapies.

Thus, far, MHC-I affinity is the only parameter which can be predicted with some reliability using neoantigen peptide and patient HLA allele sequences *in silico*, by using several computational tools. Our group recently proposed the concept of "neoantigen quality" (26, 27). This concept combines biophysical, chemical and computationally inferred properties of a neoantigen that make it more likely to induce a productive immune response against the tumor. These properties may include affinity of a neoantigen to MHC, avidity of the peptide-MHC complex to the recognizing TCR, type of T cells responding to the neoantigen and sequence similarity to known highly immunogenic epitopes (**Figure 1**). Recent studies from our group have shown that this parameter is a critical aspect in segregating responders to checkpoint therapy, but is not usually considered in algorithms of neoantigen prediction.

T cells are primed by antigen presenting cells (APC) that have taken up tumor antigens and processed them into smaller peptides that are eventually presented on MHC class I and II molecules (**Figure 1**). Intracellularly, antigens arise from proteins targeted for degradation by the 80S proteasome. Peptides of 9-12 amino acid residues in length are transported from the cytosol by specialized protein machinery (transporter associated with antigen presentation, TAP) and loaded on MHC-I molecules within the endoplasmic reticulum (28–30). Alternatively, antigens can arise from extracellular sources; captured necrotic or apoptotic cells and other vesicles that are cross-presented on professional APC such as dendritic cells (DC) (30, 31). As a tumor grows, tissue resident and migratory DC subsets capture tumor cell debris and convey them to draining lymph nodes (32–34). There, APC prime naïve T cells and educate them to recognize the harvested antigens (35, 36). Depending on the APC subset, nature of antigen and type of processing pathway, different responses can be achieved, either CD4+ T cell responses (Th1, Th2, Th17, and Treg) or cytotoxic CD8+ T cell responses (37, 38). The majority of APC prime naïve CD4+ T cells through MHC-II presented peptides, while the cross-presenting XCR1+ DC subset uniquely primes naïve CD8+ T cells (39, 40). The latter appear to be essential for successful immunotherapy regimens (41–43). After priming, reactive and expanded T cells can infiltrate the cancer site and eliminate these cells. Overall, proper antigen selection, processing, and T cell priming are at the heart of successful immune responses.

With the recognition that neoantigens can be a significant pool of tumor derived antigens depending upon the underlying mutational status of the tumor, the field has turned its attention to developing and optimizing neoantigen targeted immune therapies. There are generally two approaches:

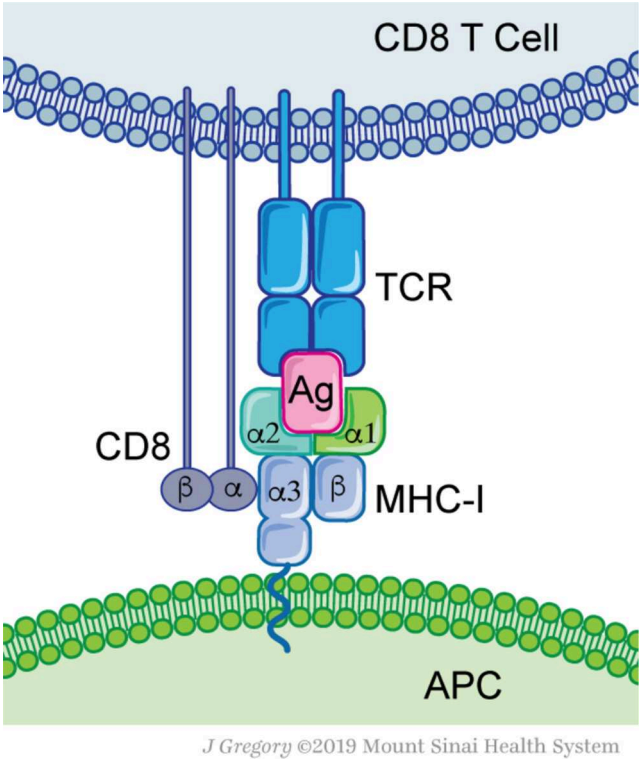


FIGURE 1 | Molecular basis for antigen recognition. Antigen-presenting cells (APC) express MHC-I complex that contains an antigenic peptide (Ag) with its groove. MHC-I consist of two proteins, a conserved β_2-microglobulin and a variable α-chain. The MHC-I-Ag complex is recognized by the T-cell receptor (TCR). Each TCR defines a clonal T cell population. Additional interactions, such as the CD8 protein—MHC-I, are not essential for Ag recognition, but are required for efficient T cell activation.

neoantigen vaccines and neoantigen targeted adoptive T cell therapies. Several clinical trials have been completed and/or are currently ongoing to enhance tumor-specific responses through neoantigen vaccination to induce expansion of neoantigen-specific CD4+ and CD8+ T cells. Vaccination techniques employ different neoantigen formulations such as peptides (13, 44, 45) combined with different adjuvants (46–49), mRNA (50, 51), DNA or as expressed in viral or bacterial vectors. Another interesting approach targets neoantigens to specific receptors expressed specifically by the cross-presenting APC cDC1. Here antigens are fused in frame to antibodies targeting the XCR1 receptor (52). Thus, far, vaccine-based approaches have demonstrated successful immunization of patients (53, 54), although CD4+ as opposed to CD8+ T cells are preferentially generated. Cell-based vaccination is also under investigation. *In vitro* expanded, neoantigen-pulsed dendritic cells have been evaluated for autologous injection in patients (46, 55–58) confirming immunogenicity (57, 58). Another approach focuses on the adoptive T cell transfer of expanded T cells purified from the patient's tumor or peripheral blood mononuclear cells (PBMC) either non-specifically or through selection *in vitro*. Some strategies have successfully utilized neoantigen-specific CD4+ and CD8+ cytotoxic T cells to eradicate solid tumors (59, 60). In another example, autologous T cell transfer of CD8+ T cells, specific to clonal neoantigens derived from cancer driver mutations, e.g., KRAS, has led to nearly complete tumor regression (61).

Mutation burden, neoantigen burden and quality have been demonstrated to be predictive for outcome of checkpoint blockade (26, 62–70). A few studies have highlighted the importance of neoantigens in shaping tumor evolution during immunotherapy with antibodies that target checkpoint molecules such as CTLA-4 and PD-1 (71, 72). However, neoantigen prediction approaches are not aligned with some utilizing solely gene expression (73, 74) or combining transcriptomics with genomics (75). The successful characterization of immunogenic neoantigens is critical to optimizing approaches that target these key epitopes. In this review we critically discuss current tools and methods for their selection (76).

THE LANDSCAPE OF NEOANTIGENS

Neoantigens arise from multiple genetic and epigenetic aberrations (**Figure 2**). Well-characterized sources of neoantigens are somatic missense and indel mutations, or other genomic rearrangements, such as gene fusions. Frameshift neoantigens may prove to be more immunogenic than missense ones due to the lack of similarity to sequences in the human coding genome and are currently under active investigation (77, 78). Neoantigens derived from gene fusions have recently passed the immunogenicity test (12), and may be of special significance when mutational burden is low. Correct detection of somatic mutations is essential to identify neoantigens

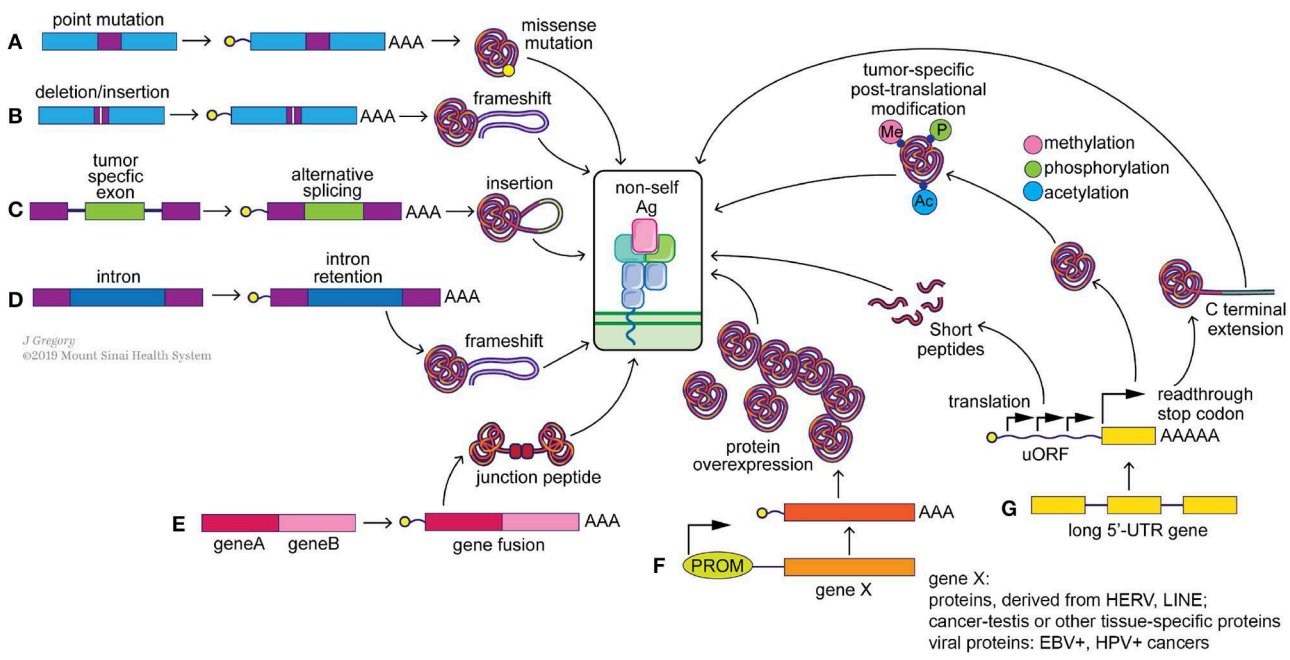


FIGURE 2 | Potential sources of "non-self" tumor neoantigens. Genomic alterations such as point mutations **(A)**, indels **(B)** and gene fusions **(E)** can result in the generation of missense and frameshift neoantigens. Splicing aberrations such as the retention of cancer-specific exons **(C)** and introns **(D)** can also lead to frameshift neoantigens. Epigenetic changes can alter the expression levels of immunogenic genomically-encoded proteins **(F)** including viral proteins from integrated chronically-infected cells (EBV, HPV), cancer-testis antigens (e.g., MAGE-A4) and proteins derived from LINE and HERV elements. Post-transcriptional changes **(G)**, including translation of upstream open reading frames (uORF), stop codon readthrough and protein modifications, such as methylation, phosphorylation and acetylation, can generate tumor specific neoantigens as well.

incorporated within these alterations (17, 79, 80). Neoantigens can also arise from transcriptome-based aberrations, including cancer-specific gene overexpression, alternative exon splicing, intron retention, premature transcription ending, readthrough the stop-codon by ribosomes and from upstream open reading frames (uORF) (**Figure 2**). Virus-induced cancers, e.g., HPV+, EBV+, generate strong immune responses due to presentation of viral antigens (81) and as such can be considered as cancer-specific antigens. Transcript-specific changes in exon usage (82, 83), intron retention (84), and transcription end usage were recently shown to produce cancer-specific neoantigens. Translation-based neoantigens, originating from uORF regions, cryptic short ORFs in non-coding RNAs still await their discovery on a pan-cancer level. Whole genome sequencing, deep RNAseq gene expression analysis, whole-cell and MHC-eluate mass-spectrometry will be necessary for a determination of the complete landscape of such neoantigens (85). Finally, cancer-specific post-translational protein modifications, e.g., phosphorylation, acetylation, methylation, citrullination, and etc., can be a potential source of neoantigens as well (86, 87). Aberrant over activity of protein kinases, histone acetylases, and methylases is well-known in multiple cancers. This can result in frequent modifications of non-natural protein targets or cancer-specific proteins, which may in turn produce immunogenic, tumor-specific neoantigens (88, 89) (**Figure 2**). It is important to point out that T cells with the capacity to recognize these modified antigens likely have escaped central tolerance and thus represent a large pool of T cell clones that could be harnessed to attack cancer cells. Technological advances in mass-spectrometry peptide detection from cancer MHC-I eluates will be essential for neoantigen discovery of this class (90).

ON A COMPUTATIONAL HUNT FOR NEOANTIGENS

Somatic Mutation Calling

Despite the broad range of potential sources of neoantigens in cancer cells, the process of selection of genomically encoded antigens that are of immunological significance remains to be well-established. Many computational pipelines have been developed to predict neoantigens from cancer genomes (91, 92). A joint effort referred to as the Tumor Neoantigen Selection Alliance (TESLA; supported by the Parker Institute for Cancer Immunotherapy and the Cancer Research Institute) to find the right predictive algorithms for targeting neoantigens (based upon NSVs) through large scale validation is ongoing. At this time, a "typical" neoantigen pipeline includes the following steps:

- Whole exome or genome sequencing (WES or WGS) of tumor and matched normal DNA samples by Illumina short read sequencing platform.
- Quality control of sequencing reads.
- Alignment to the reference genome.
- Base quality recalibration and indel realignment.
- Comparison of normal and tumor alignments to call somatic mutations.
- Conversion of coding DNA somatic mutations to corresponding mutated peptide sequences.
- HLA-allele typing.
- Assessment of HLA-allele and mutated epitope (9–11mer) affinity to call neoantigens.
- Expression analysis of putative neoantigens, e.g., RNAseq, when possible.

Multiple tools exist to check the quality of sequencing reads with the most commonly used being FastQC (93, 94). Alternative tools are included in the Genome Analysis Toolkit (GATK) bundle. To perform read alignment, Novoalign (95), BWA (96), bowtie (97), STAR (98) are the most favored aligners. For a typical WES (or WGS) dataset BWA is a commonly used aligner. Base quality recalibration and indel realignment around clusters of putative somatic mutations are both integral tools of GATK and have been shown to reduce the false positive rates of mutation calling (99). Collectively, these "pre-processing" steps output aligned, cleaned, equilibrated *.bam files of tumor and matched normal samples. These matched datasets are fed to a combination of mutation callers to predict somatic mutations in tumor samples. A wide range of somatic mutation callers exist to date, such as Mutect (100), Varscan2 (101), VarDict (102), SomaticSniper (103), Strelka, and FastD (104). Many comparative studies have been performed to call mutations (105–108) (**Figure 3A**). Some key observations are noted below.

- A combination of multiple algorithms vs. a single mutation caller significantly lowers the false positive rate (108–110).
- Calling somatic mutations from additional sequencing such as of RNAseq of the same tumor sample and determining overlap may help to reduce the false positive rate. However, it may increase the rate of false negatives due to transient gene expression and variable read coverage (111).
- PCR-free WES protocols [KAPA HyperPrep Kit (112)] produce less bias in tumor allele frequencies but achieve it at the expense of reduced total genome loci coverage and of lowered total power of somatic mutation calling (109).
- Exome capture kits (Agilent SureSelect, NimbleGen SeqCap, Illumina TrueSeq, and Illumina Nextera) introduce sequencing coverage biases due to differences in capture probes. This makes it potentially hard to compare final mutation calls obtained from different WES kits of the same sample of DNA, resulting in increased false negative rates for somatic mutation calls (113–116).
- Sequencing read coverage drops significantly in GC-rich regions, decreasing the sensitivity of tumor allele detection in these loci. Correcting for GC-bias may help to rescue certain mutations and improve tumor allele frequency estimations (114).
- Maintaining high tumor purity of the sample before DNA sequencing is essential. High levels of normal DNA "contamination" decreases sensitivity of tumor mutation calling (105, 109, 114).
- The quality of the sample is important, e.g., fresh tissue samples are better than FFPE. It is highly advisable to avoid excessive sample handling known to introduce

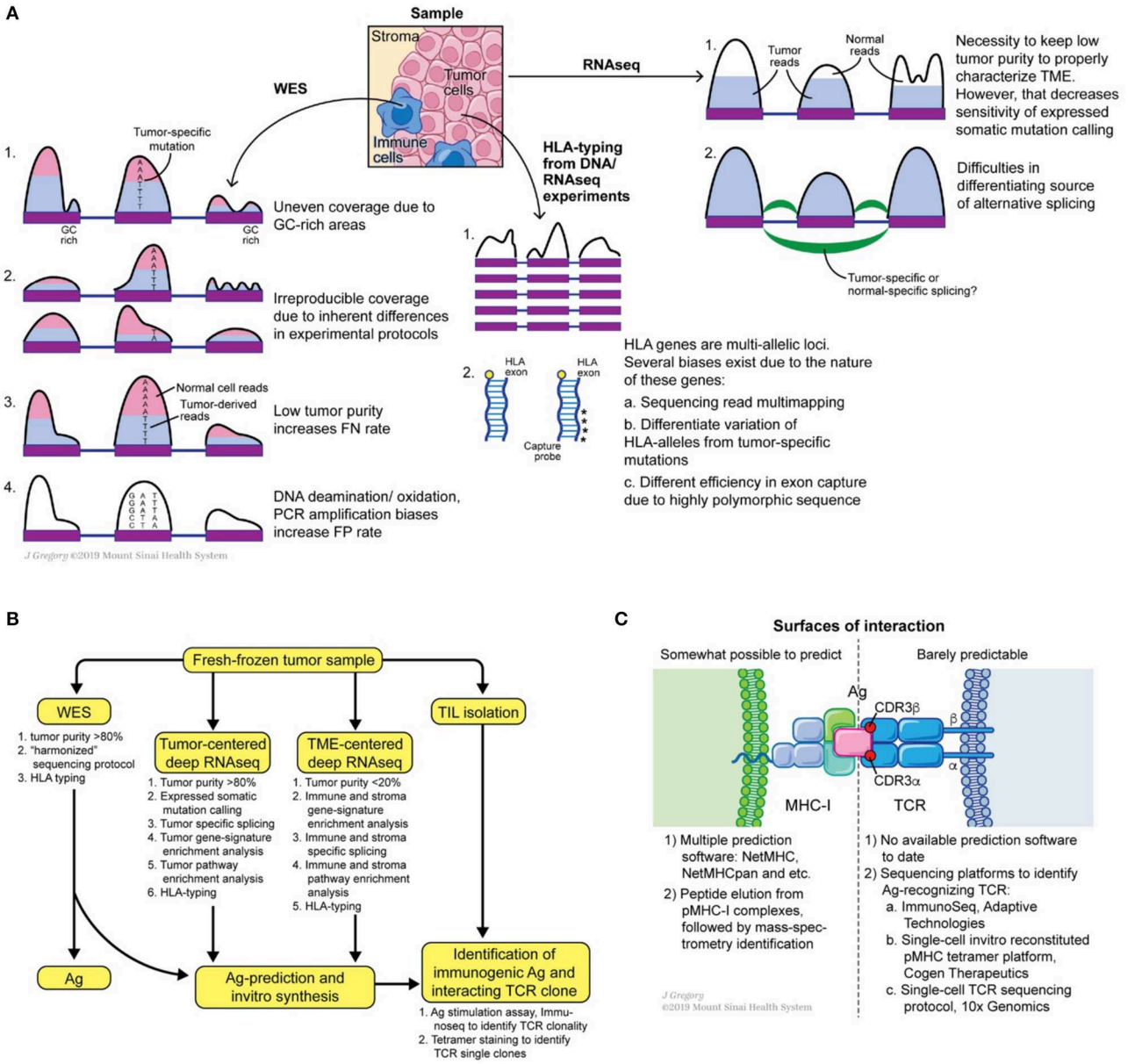


FIGURE 3 | Potential sources of somatic mutation irreproducibility **(A)**, suggested experimental design for comprehensive annotation of tumor mutational burden and tumor microenvironment, **(B)** and possible roads for innovation at the MHC-I antigen TCR interfaces **(C)**.

random DNA mutations, e.g., adenine/cytosine deamination, guanine oxidation, which can impact the results of Illumina sequencing. Otherwise the somatic mutation false positive rates increase (105, 109).

Overall, using fresh or **fresh-frozen** samples with **high tumor purity** (**>80%**), sticking to one **WES protocol**, introducing a **low** number of **PCR cycles**, following **GATK pre-processing** recommendations, and applying **several somatic mutation callers** can benefit the generation of a reproducible, "harmonized" lists of somatic mutations (117). Calling somatic mutations from **RNAseq** of samples with **high tumor content** (**>80%**) can further refine the list of **expressed** mutations. Consensus on unified somatic calling pipelines will be essential to improve the overall prediction of neoantigens and detection of the shared ones.

HLA-Allele Typing

The next step in a neoantigen calling pipeline is HLA-allele typing. CD8+ T cells see antigens presented on the MHC-I complex, which is composed of conserved β_2-microglobulin and a variable α-chain. The latter subunit is highly polymorphic and

encoded within the HLA gene, which is represented by three loci on human chromosome 6: HLA-A, HLA-B, and HLA-C. Thus, HLA allele assignment consists of the gene name (A, B, or C) followed by a set of digits separated by colons: the first two digits specify serological activity (A*01, B*03, etc.) and the second two digits indicate protein sequence (A*01:05, B*03:05, etc.). Due to the high level of polymorphism of each gene, precise HLA-allele typing at protein level resolution from WES and RNAseq reads is a complicated task (118). Multiple tools were developed to address this problem such as PHLAT (119), seq2HLA (120), Optitype (121), Polysolver (122), HLAMatchmaker (123), HLAreporter (124), HLAforest (125), HLAminer (126), xHLA (127). Each tool differs in its performance, utilized set of input parameters and analyzed sequencing dataset (RNAseq or DNAseq). Comparative studies were performed and showed that Optitype has both greater specificity and selectivity (128). However, it is important to keep in mind that the quality of WES/RNAseq is critical for any successful HLA typing. Indeed, due to the highly polymorphic nature of HLA genes, WES capture kits vary in the capturing efficiency of DNA from those regions (**Figure 3A**). This technical variability in capturing clearly affects downstream results in allele determination. Thus, careful examination of WES/RNAseq read coverage in HLA gene regions is imperative for making optimal predictions.

Prediction of Neoantigen HLA-Allele Interactions

In the final step, the researcher performs predictions of tumor antigenic epitope- HLA-allele interactions to identify neoantigens from the total pool of mutated peptides. Several tools and programs which are undergoing constant modification, are dedicated to this problem; NetMHC-pan being the one most widely used. NetMHC utilizes a combination of several artificial neural-networks (ANNs) to predict peptide affinity to selected HLA alleles. Initially NetMHC was trained on viral antigens from IEDB [https://www.iedb.org, (129–134)], therefore rendering a bias toward the selection of viral-like epitopes. Despite its general popularity, users should always keep in mind the biases these classification methods can introduce. For example, viral epitopes were originally described for the most frequent HLA-alleles, e.g., HLA*A02:01, HLA*B07:02. Thus, netMHC based predictions for tumor epitopes are a priori better for highly frequent HLA-alleles than for low-frequency HLA-alleles. One way to overcome this issue is to improve predictions by training the algorithm on peptides eluted from MHC complexes of mono-allelic cancer cell lines and identified by mass-spectrometry analysis (135). However, mass-spectrometry itself has limited ability to detect all possible eluted antigens, thus the false negative rate can be high (90, 136, 137). Data from mass spectrometry analyses indicates that only a small fraction of neoepitopes is presented on the cell surface, likely due to a combination of such systematic biases and real biophysical effects in the processing machinery (138–140). Taken together, there is an urgent need for novel, unbiased methods to generate MHC-I complexes for every HLA-allele with broadly diversified antigen sequences in order to design novel classification tools (**Figure 3C**).

Apart from class I epitopes, class II restricted neoantigens are receiving increased interest. Class II neoantigens are those epitopes presented by the MHC-II complex and recognized by CD4+ T cells. Despite the recognition that MHC-II is significant for tumor neoantigen presentation and priming of CD4+ T cells (141) and for immunotherapy outcomes (142), the accuracy and precision of MHC-II epitope predictions are poor when compared to class I (143). The main difficulties with designing such classification tools are associated with the "openness" of the peptide-binding groove of HLA class II, which permits binding of a highly degenerate set of peptides, and therefore increasing the size of datasets needed for accurate machine learning-based model training. However, these obstacles provide an opportunity for more creative efforts to develop algorithms to predict such neoantigens.

Identification of Immunogenic Neoantigen-Reactive T Cells

Not every neoantigen presented on MHC-I complexes will have the capacity to induce CD8+ T cell responses (79). What defines neoantigen immunogenicity? Conventionally, an immunogenic neoantigen must prime and stimulate T cells efficiently. This occurs through (i) interaction of the neoantigen-MHC-I complex with a TCR on one or several T cell clones, and (ii) induction of T cell priming. This process generally results in either TNF-α, IFN-γ, or double TNF-α and IFN-γ cytokine responses, IL-2 release and T cell proliferation, and the acquisition of cytolytic activity in the case of CD8+ T cells. As reviewed above, vaccination has led to the priming and expansion of neoantigen-specific T cells in humans. These responses can be enumerated through assays which measure production of cytokines upon re-exposure to peptides (through ELISA type assays or intracellular staining) or binding to synthetic tetrameric or dextrameic complexes of peptide-MHC (pMHC) molecules. The latter method relies on *in vitro* folding of the MHC-I complex (144, 145) with peptide or UV-cleavable substrate (146) which is later exchanged for the peptide of interest (147).

Neoantigen-specific T cells with effector function have been identified within PBMC following vaccination or even after spontaneous induction (148), tumor infiltrating lymphocytes (149) and can even be differentiated from progenitors through *in vitro* priming approaches (150). A concerted effort is being made to expand potent neoantigen-reactive T cells for the purpose of adoptive cell therapy or to identify high avidity neoantigen-reactive TCRs which can be modified and transduced into a primary T cells. For example, to overcome thymic negative selection, which decreases TCR diversity *in vivo* (151), humanized mice can be used to select the most-optimal neoantigen-reactive TCRs (152). Tetramer-purified, neoantigen-reactive T cell clones can also be expanded from these sources or human blood or TILs in single-cell fashion and their TCRs sequenced. The selected TCRs can be used for recombinant TCR reconstitution (153) and characterization *in vitro*, for additional modification to improve TCR avidity and stability (129, 130) and then adapted for adoptive T cell transfer using a cancer patient's own T cells (131). In this regard tetramer staining can be applied to identify neoantigen-specific TCRs in a high-throughput manner (132). One discovery platform

generates *in vitro* translated, DNA-barcoded pMHC complexes from a chemically synthesized DNA library (133). Once tetramer-positive T cells are purified, their interacting TCRs and DNA-barcoded antigens are identified through single-cell sequencing. Moreover, the same platform can be repurposed to characterize all possible peptide specificities for each HLA-allele of MHC-I and MHC-II complexes. Indeed, the ability to (i) start from a randomized DNA library of putative epitopes and (ii) characterize folding potential of produced pMHC complexes in large scale could yield invaluable information to train novel classification algorithms. Despite the obvious advantage of tetramer staining in identifying neoantigen-reactive T cells, this tool provides limited information on the functional status of purified T cells and their cytotoxic capacity (134). The recent development of T-scan screening technology holds promise to overcome this issue (154). Likewise, a recently developed method referred to as imPACT Isolation Technology identifies pre-existing T cell clones that recognize tumor neoantigens (155). Such approaches lay the foundation for multi-group collaborations to synthesize neoantigen-specific T cells for personalized adoptive T cell therapies (155).

Collectively, the identification of immunogenic neoantigens is a multi-step process that requires significant time, cost and labor to accomplish. Personalized neoantigen-based immunotherapies suffer from such drawbacks, sometimes requiring up to 3 months to manufacture the a short list of "best" candidates (156). A potential solution to this pipeline problem is to target shared neoantigens, that are highly recurrent, clonal, and broadly immunogenic across cancer patients. However, whether such immunogenic shared antigens are sufficiently available across broad cancer types remains to be determined. Prioritizing such antigens whenever possible is important, as any "off-the shelf" strategies that can be developed will significantly reduce the cost and increase the efficiency of neoantigen-specific cancer immunotherapies.

CONCLUDING REMARKS

We review the available tools for the computational prediction and experimental validation of tumor-associated neoantigens, discussing approaches for somatic mutation detection, HLA allele typing, and prediction of peptide-MHC interactions. We have made an effort to highlight the biases associated with particular approaches and suggest possible ways to minimize their influence. We also outline technologies for identifying immunogenic neoantigens. Future developments that could improve these strategies are suggested in **Figure 3**. Firstly, harmonization of somatic mutation calling can improve reproducibility across different platforms and sequencing centers. Secondly, *in vitro* assays for folding and characterization of pMHC complexes starting from randomized peptide libraries can improve existing prediction tools. Applying the same approach for peptide-MHC-II complexes may also improve MHC-II classification tools (157). Finally, single-cell identification of TCR-antigen interacting pairs will provide information on the principles of TCR-neoantigen interactions, making it possible to develop predictive methods for this type of interaction (158). The latter will be an invaluable tool for immunogenic neoantigen selection for vaccine designs, refining immunotherapy outcome predictions, or selecting the most avid TCR for adoptive recombinant T cell therapies. We believe the field of neoantigen-based immunotherapies of cancer is undergoing a major renaissance. Equipped with powerful sequencing technologies, sensitive computational tools for neoantigen discovery and efficient high-throughput platforms for characterization of their immunogenicity, scientists will have the potential to bring novel disruptive immunotherapies to the clinic to definitely improve outcomes of cancer patients.

AUTHOR CONTRIBUTIONS

VR wrote the manuscript. VR, BG, and NB reviewed and revised the manuscript.

FUNDING

NB has National Institutes of Health grants R01CA201189, R01CA180913, and R01AI081848 and receives support from the Melanoma Research Alliance, Cancer Research Institute, Leukemia & Lymphoma Society and NYSTEM. BG has National Institutes of Health grants 7R01AI081848-04 and 1P30CA196521-01; BG has Stand Up To Cancer-National Science Foundation-Lustgarten Foundation Convergence Dream Team Grant sponsored by Stand Up to Cancer, the Lustgarten Foundation, the V Foundation and the National Science Foundation grant NSF 1545935; BG is the Pershing Square Sohn Prize-Mark Foundation Fellow supported by funding from The Mark Foundation for Cancer Research.

ACKNOWLEDGMENTS

We thank the members of NB and BG labs for critical assessment of the manuscript.

REFERENCES

1. Hodi S, O'Day S, McDermott D, Weber R, Harris A, Johnson DW, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med.* (2010) 363:609–19. doi: 10.1056/NEJMoa1003466
2. Topalian S, Hodi S, Brahmer J, Gettinger S, Smith DC, Mcdermott DF, et al. Safety, Activity, and immune correlates of anti–PD-1 antibody in cancer. *N Engl J Med.* (2012) 366:2443–54. doi: 10.1056/NEJMoa1200690
3. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med.* (2015) 373:1803–13. doi: 10.1056/NEJMoa1510665
4. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med.* (2015) 373:123–35. doi: 10.1056/NEJMoa1504627

5. Powles T, Eder JP, Fine GD, Braiteh FS, Loriot Y, Cruz C, et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature.* (2014) 515:558–62. doi: 10.1038/nature13904
6. Hamanishi J, Mandai M, Ikeda T, Minami M, Kawaguchi A, Murayama T, et al. Safety and antitumor activity of anti-PD-1 antibody, nivolumab, in patients with platinum-resistant ovarian cancer. *J Clin Oncol.* (2015) 33:4015–22. doi: 10.1200/JCO.2015.62.3397
7. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in tumors with mismatch-repair deficiency. *N Engl J Med.* (2015) 372:2509–20. doi: 10.1200/jco.2015.33.15_suppl.lba100
8. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science.* (2017) 413:409–13. doi: 10.1126/science.aan6733
9. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature.* (2014) 515:568–71. doi: 10.1038/nature13954
10. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science.* (2015) 348:69–74. doi: 10.1126/science.aaa4971
11. De Vries J, Figdor C. Immunotherapy: cancer vaccine triggers antiviral-type defences. *Nature.* (2016) 534:329–31. doi: 10.1038/nature18443
12. Yang W, Lee K, Srivastava RM, Kuo F, Krishna C, Chowell D, et al. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat Med.* (2019) 25:1–9. doi: 10.1038/s41591-019-0434-2
13. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature.* (2019) 565:234–9. doi: 10.1038/s41586-018-0792-9
14. Tran E, Robbins PF, Rosenberg SA. Final "common pathway" of human cancer immunotherapy: targeting random somatic mutations. *Nat Immunol.* (2017) 18:255–62. doi: 10.1038/ni.3682
15. Coulie PG, Van Den Eynde BJ, Van Der Bruggen P, Boon T. Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nat Rev Cancer.* (2014) 14:135–46. doi: 10.1038/nrc3670
16. Stone JD, Harris DT, Kranz DM. TCR affinity for p/MHC formed by tumor antigens that are self-proteins: impact on efficacy and toxicity. *Curr Opin Immunol.* (2015) 33:16–22. doi: 10.1016/j.coi.2015.01.003
17. Efremova M, Finotello F, Rieder D, Trajanoski Z, Gutenberg-universität J, Gutenberg-universität J. Neoantigens generated by individual mutations and their role in cancer immunity and immunotherapy. *Front Immunol.* (2017) 8:1679. doi: 10.3389/fimmu.2017.01679
18. Mcgranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science.* (2016) 351:1463–9. doi: 10.1126/science.aaf1490
19. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell.* (2016) 165:35–44. doi: 10.1016/j.cell.2016.02.065
20. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science.* (2015) 350:207–11. doi: 10.1126/science.aad0095
21. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* (2015) 348:124–8. doi: 10.1126/science.aaa1348
22. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep.* (2016) 15:857–65. doi: 10.1016/j.celrep.2016.03.075
23. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell.* (2017) 171:934–49.e15. doi: 10.1016/j.cell.2017.09.028
24. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer.* (2019) 19:133–50. doi: 10.1038/s41568-019-0116-x
25. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med.* (2019) 25:1251–9. doi: 10.1101/648899
26. Luksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature.* (2017) 551:517–20. doi: 10.1038/nature24473
27. Richman LP, Vonderheide RH, Rech AJ. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. *Cell Syst.* (2019) 9:375–82.e4. doi: 10.1016/j.cels.2019.08.009
28. Santambrogio L, Berendam SJ, Engelhard VH. The antigen processing and presentation machinery in lymphatic endothelial cells. *Front Immunol.* (2019) 10:1033. doi: 10.3389/fimmu.2019.01033
29. Mellman I, Steinman RM. Dendritic cells: specialized and regulated antigen processing machines. *Cell.* (2001) 106:255–8. doi: 10.1016/S0092-8674(01)00449-4
30. Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol.* (2013) 31:443–73. doi: 10.1146/annurev-immunol-032712-095910
31. Cruz FM, Colbert JD, Merino E, Kriegsman BA, Rock KL. The biology and underlying mechanisms of cross-presentation of exogenous antigens on MHC-I molecules. *Annu Rev Immunol.* (2017) 35:149–76. doi: 10.1146/annurev-immunol-041015-055254
32. Boltjes A, van Wijk F. Human dendritic cell functional specialization in steady-state and inflammation. *Front Immunol.* (2014) 5:131. doi: 10.3389/fimmu.2014.00131
33. Dalod M, Chelbi R, Malissen B, Lawrence T. Dendritic cell maturation: functional specialization through signaling specificity and transcriptional programming. *EMBO J.* (2014) 33:1104–16. doi: 10.1002/embj.201488027
34. Worah K, Mathan TSM, Vu Manh TP, Keerthikumar S, Schreibelt G, Tel J, et al. Proteomics of human dendritic cell subsets reveals subset-specific surface markers and differential inflammasome function. *Cell Rep.* (2016) 16:2953–66. doi: 10.1016/j.celrep.2016.08.023
35. Stockwin LH, McGonagle D, Martin IG, Blair GE. Dendritic cells: immunological sentinels with a central role in health and disease. *Immunol Cell Biol.* (2000) 78:91–102. doi: 10.1046/j.1440-1711.2000.00888.x
36. Hancock DG, Guy TV, Shklovskaya E, Fazekas de St Groth B. Experimental models to investigate the function of dendritic cell subsets: challenges and implications. *Clin Exp Immunol.* (2013) 171:147–54. doi: 10.1111/cei.12027
37. Gutiérrez-Martínez E, Planès R, Anselmi G, Reynolds M, Menezes S, Adiko AC, et al. Cross-presentation of cell-associated antigens by MHC class I in dendritic cell subsets. *Front Immunol.* (2015) 6:363. doi: 10.3389/fimmu.2015.00363
38. Merad M, Sathe P, Helft J, Miller J, Mortha A. The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu Rev Immunol.* (2013) 31:563–604. doi: 10.1146/annurev-immunol-020711-074950
39. Jongbloed SL, Kassianos AJ, McDonald KJ, Clark GJ, Ju X, Angel CE, et al. Human CD141+ (BDCA-3)+ dendritic cells (DCs) represent a unique myeloid DC subset that cross-presents necrotic cell antigens. *J Exp Med.* (2010) 207:1247–60. doi: 10.1084/jem.20092140
40. Xu MM, Pu Y, Han D, Shi Y, Cao X, Liang H, et al. Dendritic cells but not macrophages sense tumor mitochondrial DNA for cross-priming through signal regulatory protein α signaling. *Immunity.* (2017) 47:363–73.e5. doi: 10.1016/j.immuni.2017.07.016
41. Broz ML, Binnewies M, Boldajipour B, Nelson AE, Pollack JL, Erle DJ, et al. Dissecting the tumor myeloid compartment reveals rare activating antigen-presenting cells critical for T cell immunity. *Cancer Cell.* (2014) 26:638–52. doi: 10.1016/j.ccell.2014.09.007
42. Hammerich L, Marron TU, Upadhyay R, Svensson-arvelund J, Dhainaut M, Hussein S, et al. Systemic clinical tumor regressions and potentiation of PD1 blockade with *in situ* vaccination. *Nat Med.* (2019) 25:814–24. doi: 10.1038/s41591-019-0410-x
43. Salmon H, Remark R, Gnjatic S, Merad M. Host tissue determinants of tumour immunity. *Nat Rev Cancer.* (2019) 19:215–27.
44. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature.* (2017) 547:217–21. doi: 10.1038/nature22991
45. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanović S, Gouttefangeas C, et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature.* (2019) 565:240–5. doi: 10.1038/s41586-018-0810-y

46. Temizoz B, Kuroda E, Ishii KJ. Vaccine adjuvants as potential cancer immunotherapeutics. *Int Immunol.* (2016) 28:329–38. doi: 10.1093/intimm/dxw015
47. Kranz LM, Diken M, Haas H, Kreiter S, Loquai C, Reuter KC, et al. Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature.* (2016) 534:396–401. doi: 10.1038/nature18300
48. Klinman DM. Immunotherapeutic uses of CpG oligodeoxynucleotides. *Nat Rev Immunol.* (2004) 4:249–59. doi: 10.1038/nri1329
49. Kinkead HL, Jaffee EM, Zaidi N, Kinkead HL, Hopkins A, Lutz E, et al. Combining STING-based neoantigen-targeted vaccine with checkpoint modulators enhances antitumor immunity in murine pancreatic cancer. *JCI Insight.* (2018) 3:122857. doi: 10.1172/jci.insight.122857
50. Scheel B, Teufel R, Probst J, Carralot J-P, Geginat J, Radsak M, et al. Toll-like receptor-dependent activation of several human blood cell types by protamine-condensed mRNA. *Eur J Immunol.* (2005) 35:1557–66. doi: 10.1002/eji.200425656
51. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature.* (2017) 547:222–6. doi: 10.1038/nature23003
52. Tullett KM, Leal Rojas IM, Minoda Y, Tan PS, Zhang J-G, Smith C, et al. Targeting CLEC9A delivers antigen to human CD141(+) DC for CD4(+) and CD8(+)T cell recognition. *JCI Insight.* (2016) 1:e87102. doi: 10.1172/jci.insight.87102
53. Linette GP, Carreno BM. Neoantigen vaccines pass the immunogenicity test. *Trends Mol Med.* (2017) 23:869–71. doi: 10.1016/j.molmed.2017.08.007
54. Branca MA. Rekindling cancer vaccines. *Nat Biotechnol.* (2016) 34:1019–24. doi: 10.1038/nbt.3690
55. Boudreau JE, Bonehill A, Thielemans K, Wan Y. Engineering dendritic cells to enhance cancer immunotherapy. *Mol Ther.* (2011) 19:841–53. doi: 10.1038/mt.2011.57
56. Selmi A, Vascotto F, Kautz-Neu K, Türeci Ö, Sahin U, von Stebut E, et al. Uptake of synthetic naked RNA by skin-resident dendritic cells via macropinocytosis allows antigen expression and induction of T-cell responses in mice. *Cancer Immunol Immunother.* (2016) 65:1075–83. doi: 10.1007/s00262-016-1869-7
57. Binnewies M, Mujal AM, Pollack JL, Ye CJ, Roberts EW, Krummel MF, et al. Unleashing type-2 dendritic cells to drive protective antitumor CD4+ T cell immunity. *Cell.* (2019) 177:556–71.e16. doi: 10.1016/j.cell.2019.02.005
58. Carreno BM, Magrini V, Becker-Hapak M, Hundal J, Ly A, Linette GP. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science.* (2015) 348:803–8. doi: 10.1126/science.aaa3828
59. Emerson R, Chapuis AG, Desmarais C, Lai IP, Roberts IM, Yee C, et al. Tracking the fate and origin of clinically relevant adoptively transferred CD8 + T cells *in vivo*. *Sci Immunol.* (2017) 2:eaal2568. doi: 10.1126/sciimmunol.aal2568
60. Yamaguchi N, Winter CM, Wu MF, Kanno Y, Yamaguchi A, Seo M, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science.* (2014) 9:641–6. doi: 10.1126/science.1251102
61. Tran E, Robbins PF, Lu Y-C, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med.* (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
62. Balachandran VP, Luksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature.* (2017) 551:S12–6. doi: 10.1038/nature24462
63. Zhang J, Caruso FP, Sa JK, Justesen S, Nam D, Sims P, et al. The combination of neoantigen quality and T lymphocyte infiltrates identifies glioblastomas with the longest survival. *Commun Biol.* (2019) 2:135. doi: 10.1038/s42003-019-0369-7
64. Wood MA, Paralkar M, Paralkar MP, Nguyen A, Struck AJ, Ellrott K, et al. Population-level distribution and putative immunogenicity of cancer neoepitopes. *BMC Cancer.* (2018) 18:414. doi: 10.1186/s12885-018-4325-6
65. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* (2017) 18:1009–21. doi: 10.1016/S1470-2045(17)30516-8
66. Spranger S, Luke JJ, Bao R, Zha Y, Hernandez KM, Li Y, et al. Density of immunogenic antigens does not explain the presence or absence of the T-cell-inflamed tumor microenvironment in melanoma. *Proc Natl Acad Sci USA.* (2016) 113:E7759–68. doi: 10.1073/pnas.1609376113
67. Hellmann M. Mutation burden, neoantigens, and response to T cell checkpoint blockade. *TAT Conf Proc.* (2016).
68. Charoentong P, Angelova M, Charoentong P, Finotello F, Angelova M, Mayer C, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade: cell reports. *Cell Rep.* (2017) 18:248–62. doi: 10.1016/j.celrep.2016.12.019
69. Blank C, Haanen J, Ribas A, Schumacher T. The cancer immunogram. *Science.* (2016) 352:658–60. doi: 10.1126/science.aaf2834
70. Bräunlein E, Krackhardt AM. Identification and characterization of neoantigens as well as respective immune responses in cancer patients. *Front Immunol.* (2017) 8:1702. doi: 10.3389/fimmu.2017.01702
71. Rosenthal R, Cadieux EL, Salgado R, Bakir M Al, Moore DA, Hiley CT, et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature.* (2019) 567:479–85. doi: 10.1038/s41586-019-1032-7
72. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, Haar J van de, Engin HB, et al. MHC-I genotype restricts the oncogenic mutational landscape. *Cell.* (2017) 171:1–12.e15. doi: 10.1016/j.cell.2017.09.050
73. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med.* (2018) 24:1545–9. doi: 10.1038/s41591-018-0247-8
74. Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D, et al. Integrative analyses of colorectal cancer show immunoscore is a stronger predictor of patient survival than microsatellite instability. *Immunity.* (2016) 44:698–711. doi: 10.1016/j.immuni.2016.02.025
75. Senft D, Leiserson MDM, Ruppin E, Ronai ZA. Precision oncology: the road ahead. *Trends Mol Med.* (2017) 23:874–98. doi: 10.1016/j.molmed.2017.08.003
76. Snyder A, Chan TA. Immunogenic peptide discovery in cancer genomes. *Curr Opin Genet Dev.* (2015) 30:7–16. doi: 10.1016/j.gde.2014.12.003
77. Roudko V, Bozkus CC, Orfanelli T, Blank SV. Widespread immunogenic poly-epitope frameshift mutations in microsatellite unstable tumors. *bioRvix.* (2019) 1–53. doi: 10.1101/662262
78. Schwitalle Y, Kloor M, Eiermann S, Linnebacher M, Kienle P, Knaebel HP, et al. Immune response against frameshift-induced neopeptides in HNPCC patients and healthy HNPCC mutation carriers. *Gastroenterology.* (2008) 134:988–97. doi: 10.1053/j.gastro.2008.01.015
79. Editorial. The problem with neoantigen prediction. *Nat Biotechnol.* (2017) 35:97. doi: 10.1038/nbt.3800
80. Robbins PF, Lu YC, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med.* (2013) 19:747–52. doi: 10.1038/nm.3161
81. Rosato PC, Wijeyesinghe S, Stolley JM, Nelson CE, Davis RL, Manlove LS, et al. Virus-specific memory T cells populate tumors and can be repurposed for tumor immunotherapy. *Nat Commun.* (2019) 10:567. doi: 10.1038/s41467-019-08534-1
82. Hoyos LE, Abdel-Wahab O. Cancer-specific splicing changes and the potential for splicing-derived neoantigens. *Cancer Cell.* (2018) 34:181–3. doi: 10.1016/j.ccell.2018.07.008
83. Kahles A, Lehmann K Van, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell.* (2018) 34:211–24.e6. doi: 10.1016/j.ccell.2018.07.001
84. Smart A, Margolis C, Pimentel H, He MX, Miao D, Adeegbe D, et al. Intron retention as a novel source of cancer neoantigens. *bioRxiv.* (2018). doi: 10.1101/309450
85. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature.* (2014) 515:572–6. doi: 10.1038/nature14001

86. Doyle HA, Mamula MJ. Post-translational protein modifications in antigen recognition and autoimmunity. *Trends Immunol.* (2001) 22:443–9. doi: 10.1016/S1471-4906(01)01976-7
87. Engelhard VH, Altrich-Vanlith M, Ostankovitch M, Zarling AL. Post-translational modifications of naturally processed MHC-binding epitopes. *Curr Opin Immunol.* (2006) 18:92–7. doi: 10.1016/j.coi.2005.11.015
88. Malaker SA, Penny SA, Steadman LG, Myers PT, Loke JC, Raghavan M, et al. Identification of glycopeptides as posttranslationally modified neoantigens in leukemia. *Cancer Immunol. Res.* (2017) 5:376–84. doi: 10.1158/2326-6066.CIR-16-0280
89. Raposo B, Merky P, Lundqvist C, Yamada H, Urbonaviciute V, Niaudet C, et al. T cells specific for post-translational modifications escape intrathymic tolerance induction. *Nat Commun.* (2018) 9:353. doi: 10.1038/s41467-017-02763-y
90. Purcell AW. Mass spectrometry – based identification of MHC-bound peptides for immunopeptidomics. *Nat Protoc.* (2019) 14:1687–707. doi: 10.1038/s41596-019-0133-y
91. Meena N, Mathur P, Medicherla KM, Suravakhala P. A bioinformatics pipeline for whole exome sequencing: overview of the processing and steps from raw data to downstream analysis. *bioRvix.* (2017). doi: 10.1101/201145
92. Hintzsche JD, Robinson WA, Tan AC. A survey of computational tools to analyze and interpret whole exome sequencing data. *Int J Genomics.* (2016) 2016:7983236. doi: 10.1155/2016/7983236
93. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front Genet.* (2013) 4:288. doi: 10.3389/fgene.2013.00288
94. Wingett SW, Andrews S. FastQ screen: a tool for multi-genome mapping and quality control. *F1000Research.* (2018) 7:1338. doi: 10.12688/f1000research.15931.1
95. Company list. *Novoalign Manual* (2014).
96. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* (2010) 26:589–95. doi: 10.1093/bioinformatics/btp698
97. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinforma.* (2010) Chapter 11:Unit 11.7. doi: 10.1002/0471250953.bi1107s32
98. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
99. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* (2011) 43:491–8. doi: 10.1038/ng.806
100. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* (2013) 31:213–9. doi: 10.1038/nbt.2514
101. Koboldt DC, Larson DE, Wilson RK. Using varscan 2 for germline variant calling and somatic mutation detection. *Curr Protoc Bioinforma.* (2013) 44:15.4.1–17. doi: 10.1002/0471250953.bi1504s44
102. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* (2016) 75:4864. doi: 10.1158/1538-7445.AM2015-4864
103. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. Somaticsniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics.* (2012) 28:311–7. doi: 10.1093/bioinformatics/btr665
104. Xu F, Wang W, Wang P, Jun Li M, Chung Sham P, Wang J. A fast and accurate SNP detection algorithm for next-generation sequencing data. *Nat Commun.* (2012) 3:1258. doi: 10.1038/ncomms2256
105. Jennings LJ, Arcila ME, Corless C, Kamel-reid S, Lubin IM, Pfeifer J, et al. Guidelines for validation of next-generation sequencing e based oncology panels A joint consensus recommendation of the association for molecular pathology and college of american pathologists. *J Mol Diagnostics.* (2017) 19:341–65. doi: 10.1016/j.jmoldx.2017.01.011
106. Clark M, Chen R, Lam H, Karczewski K, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol.* (2014) 29:908–14. doi: 10.1038/nbt.1975
107. Denroche RE, Mullen L, Timms L, Beck T, Yung CK, Stein L, et al. A cancer cell-line titration series for evaluating somatic classification. *BMC Res Notes.* (2015) 8:823. doi: 10.1186/s13104-015-1803-7
108. Warden CD, Adamson AW, Neuhausen SL, Wu X. Detailed comparison of two popular variant calling packages for exome and targeted exon studies. *PeerJ.* (2014) 2:e600. doi: 10.7717/peerj.600
109. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat Commun.* (2015) 6:10001. doi: 10.1038/ncomms10001
110. Ewing A. *Mutation Calling: TCGA Benchmark 4* (2012).
111. Sivachenko A. Comparison and validation of somatic mutation callers. (2011).
112. Meyer J, Whitehorn H, Kits H, Adapters KD, Beads KP. *KAPA HyperPrep Kits offer a flexible, high-efficiency library preparation solution for PCR-free human whole-genome sequencing*. Roche (2018).
113. Wang VG, Kim H, Chuang JH. Whole-exome sequencing capture kit biases yield false negative mutation calls in TCGA cohorts. *PLoS ONE.* (2018) 13:e0204912. doi: 10.1371/journal.pone.0204912
114. Wang Q, Shashikant CS, Jensen M, Altman NS, Girirajan S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep.* (2017) 7:885. doi: 10.1038/s41598-017-01005-x
115. Chilamakuri CSR, Lorenz S, Madoui M, Vodák D, Sun J, Hovig E, et al. Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics.* (2014) 15:449. doi: 10.1186/1471-2164-15-449
116. Shigemizu D, Momozawa Y, Abe T, Morizono T, Boroevich KA, Takata S, et al. Performance comparison of four commercial human whole-exome capture platforms. *Sci Rep.* (2015) 5:1–8. doi: 10.1038/srep12742
117. Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome sequencing : current and future perspectives. *G3.* (2015) 5:1543–50. doi: 10.1534/g3.115.018564
118. Lazaro A, Tu B, Yang R, Xiao Y, Kariyawasam K, Ng J, et al. Human leukocyte antigen (HLA) typing by DNA sequencing. *Methods Mol Biol.* (2013) 1034:161–95. doi: 10.1007/978-1-62703-493-7_9
119. Bai Y, Ni M, Cooper B, Wei Y, Fury W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics.* (2014) 15:1–16. doi: 10.1186/1471-2164-15-325
120. Boegel S, Löwer M, Schäfer M, Bukur T, Graaf J De, Boisguérin V, et al. HLA typing from RNA-Seq sequence reads. *Genome Med.* (2013) 4:102. doi: 10.1186/gm403
121. Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType : precision HLA typing from next-generation sequencing data. *Bioinformatics.* (2014) 30:3310–6. doi: 10.1093/bioinformatics/btu548
122. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Philip M, Steelman S, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol.* (2016) 33:1152–8. doi: 10.1038/nbt.3344
123. Duquesnoy J. HLAMatchmaker : a molecularly based algorithm for histocompatibility determination. *Hum Immunol.* (2002) 63:339–52. doi: 10.1016/S0198-8859(02)00382-8
124. Huang Y, Yang J, Ying D, Zhang Y, Shotelersuk V, Hirankarn N, et al. HLAreporter: a tool for HLA typing from next generation sequencing data. *Genome Med.* (2015) 7:1–12. doi: 10.1186/s13073-015-0145-3
125. Kim HJ, Pourmand N. HLA haplotyping from RNA-seq data using hierarchical read weighting. *PLoS ONE.* (2013) 8:e67885. doi: 10.1371/journal.pone.0067885
126. Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, et al. Derivation of HLA types from shotgun sequence datasets. *Genome Med.* (2012) 4:95. doi: 10.1186/gm396
127. Xie C, Xuan Z, Wong M, Piper J, Long T, Kirkness EF, et al. Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. *PNAS.* (2017) 114:8059–64. doi: 10.1073/pnas.1707945114
128. Kiyotani K, Mai TH, Nakamura Y. Comparison of exome-based HLA class I genotyping tools: identification of platform-specific genotyping errors. *J Hum Genet.* (2016) 62:1–9. doi: 10.1038/jhg.2016.141
129. Kuball J, Hauptrock B, Malina V, Antunes E, Voss R-H, Wolfl M, et al. Increasing functional avidity of TCR-redirected T cells by removing defined N -glycosylation sites in the TCR constant domain. *J Exp Med.* (2009) 206:463–75. doi: 10.1084/jem.20082487

130. Schmitt TM, Stromnes IM, Chapuis AG, Greenberg PD. New strategies in engineering T-cell receptor gene-modified T cells to more effectively target malignancies. *Clin Cancer Res.* (2015) 21:5191–7. doi: 10.1158/1078-0432.CCR-15-0860
131. Klebanoff CA, Rosenberg SA, Restifo NP. Prospects for gene-engineered T cell immunotherapy for solid cancers. *Nat Med.* (2016) 22:26–36. doi: 10.1038/nm.4015
132. Zhang S-Q, Ma K-Y, Schonnesen AA, Zhang M, He C, Sun E, et al. High-throughput determination of the antigen specificities of T cell receptors in single cells. *bioRxiv.* (2018) 457069. doi: 10.1101/457069
133. Cogen Therapeutics. Available online at: http://cogentherapeutics.com
134. Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, McAffee D, et al. Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell.* (2018) 174:672–87.e27. doi: 10.1016/j.cell.2018.06.017
135. Abelin JG, Keskin DB, Sarkizova S, Hacohen N, Rooney MS, Carr SA, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity.* (2017) 46:315–26. doi: 10.1016/j.immuni.2017.02.007
136. Olaf R, Falk K, Deres K, Schild H, Norda M, Metzger J, et al. Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. *Nature.* (1990) 348:252–4. doi: 10.1038/348252a0
137. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, et al. Characterization of peptides bound to the Class I MHC molecule HLA-A2.1 by Mass Spectrometry. *Science.* (1992) 255:1–4. doi: 10.1126/science.1546328
138. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun.* (2016) 7:13404. doi: 10.1038/ncomms13404
139. Newey A, Griffiths B, Michaux J, Pak HS, Stevenson BJ, Woolston A, et al. Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK- inhibitor treatment. *J Immunother Cancer.* (2019) 8:1–15. doi: 10.1186/s40425-019-0769-8
140. Chong C, Müller M, Pak H, Harnett D, Huber F, Grun D, et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *bioRxiv.* (2019) 758680.
141. Axelrod ML, Cook RS, Johnson DB, Balko JM. Biological consequences of MHC-II expression by tumor cells in cancer. *Clin Cancer Res.* (2019) 25:2392–402. doi: 10.1158/1078-0432.CCR-18-3200
142. Alspach E, Lussier DM, Miceli AP, Kizhvatov I, DuPage M, Luoma AM, et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature.* (2019) 574:696–701. doi: 10.1038/s41586-019-1671-8
143. Nielsen M, Lund O, Buus S, Lundegaard C. MHC Class II epitope predictive algorithms. *Immunology.* (2010) 130:319–28. doi: 10.1111/j.1365-2567.2010.03268.x
144. Garboczi DN, Hung DT, Wiley DONC. HLA-A2-peptide complexes: Refolding and crystallization of molecules expressed in *Eschericlhia coli* and complexed with single antigenic peptides. *PNAS.* (1992) 89:3429–33. doi: 10.1073/pnas.89.8.3429
145. Fahnestock ML, Tamir I, Narhi L, Bjorkman PJ. Thermal stability comparison of purified empty and peptide-filled forms of a class I MHC molecule. *Science.* (1992) 258:1658–62. doi: 10.1126/science.1360705
146. Toebes M, Coccoris M, Bins A, Rodenko B, Gomez R, Nieuwkoop NJ, et al. Design and use of conditional MHC class I ligands. *Nat Med.* (2006) 12:246–51. doi: 10.1038/nm1360
147. Rodenko B, Toebes M, Hadrup SR, Esch WJE Van, Molenaar AM, Schumacher TNM, et al. Generation of peptide-MHCclass I complexes through UV-mediated ligand exchange. *Nat Protoc.* (2006) 1:1120–32. doi: 10.1038/nprot.2006.121
148. Ali M, Giannakopoulou E, Böschen M, Strønen E, Yang W, Toebes M, et al. Induction of neoantigen-reactive T cells from healthy donors. *Nat Protoc.* (2019) 14:1926–43 doi: 10.1038/s41596-019-0170-6
149. Gee MH, Han A, Lofgren SM, Beausang JF, Mendoza JL, Birnbaum ME, et al. Antigen identification for orphan T cell receptors expressed on tumor-infiltrating lymphocytes. *Cell.* (2017) 172:549–56.e16. doi: 10.1016/j.cell.2017.11.043
150. Schmitt TM, Aggen DH, Ishida-Tsubota K, Ochsenreither S, Kranz DM, Greenberg PD. Generation of higher affinity T cell receptors by antigen-driven differentiation of progenitor t cells *in vitro*. *Nat Biotechnol.* (2017) 35:1188–95. doi: 10.1038/nbt.4004
151. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see and don't see. *Nat Rev Immunol.* (2014) 14:377–91. doi: 10.1038/nri3667
152. Obenaus M, Leitão C, Leisegang M, Chen X, Gavvovidis I, Van Der Bruggen P, et al. Identification of human T-cell receptors with optimal affinity to cancer antigens using antigen-negative humanized mice. *Nat Biotechnol.* (2015) 33:402–7. doi: 10.1038/nbt.3147
153. Hu Z, Anandappa AJ, Sun J, Kim J, Leet DE, David J, et al. A cloning and expression system to probe T cell receptor specificity and assess functional avidity to neoantigens. *Blood.* (2018) 132:1911–21. doi: 10.1182/blood-2018-04-843763
154. Wang CI, Wucherpfennig KW, Lyerly HK, Elledge SJ, Kula T, Dezfulian MH, et al. T-scan: a genome-wide method for the systematic discovery of T cell epitopes. *Cell.* (2019) 178:1016–28.e13. doi: 10.1016/j.cell.2019.07.009
155. Baker RG, Hoos AX, Adam SJ, Wholley D, Doroshow J, Lowy D, et al. The partnership for accelerating cancer therapies. *Cancer J.* (2018) 24:111–4. doi: 10.1097/PPO.0000000000000321
156. Yamamoto TN, Kishton RJ, Restifo NP. Developing neoantigen-targeted T cell–based treatments for solid tumors. *Nat Med.* (2019) 25:1488–99. doi: 10.1038/s41591-019-0596-y
157. Vollers SS, Stern LJ. Class II major histocompatibility complex tetramer staining: progress, problems, and prospects. *Immunology.* (2008) 123:305–13. doi: 10.1111/j.1365-2567.2007.02801.x
158. Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature.* (2017) 547:89–93. doi: 10.1038/nature22383

Conflict of Interest: NB receives research support or reagents from Novocure, Celldex, the Ludwig Institute for Cell Research, Genentech, Oncovir, and Regeneron, and is on the advisory boards of Neon, Tempest, Checkpoint Sciences, Curevac, Primevax, Novartis, Array BioPharma, Roche, and Avidea. NB receives grant support from and serves on the advisory board of the Parker Institute for Cancer Immunotherapy.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

HYPOTHESIS AND THEORY
published: 06 March 2020
doi: 10.3389/fimmu.2020.00069

Immunogenicity and Immune Silence in Human Cancer

*Mark Yarmarkovich[1,2], Alvin Farrel[1], Artemio Sison III[3], Moreno di Marco[4], Pichai Raman[3,5], Joshua L. Parris[2], Dimitrios Monos[2,6], Hongzhe Lee[2], Stefan Stevanovic[4] and John M. Maris[1,2]**

[1] Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA, United States, [2] Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States, [3] Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, United States, [4] Department of Immunology, University of Tübingen, Tübingen, Germany, [5] The Center for Data Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, United States, [6] Department of Pathology and Lab Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, United States

Despite recent advances in cancer immunotherapy, the process of immunoediting early in tumorigenesis remains obscure. Here, we employ a mathematical model that utilizes the Cancer Genome Atlas (TCGA) data to elucidate the contribution of individual mutations and HLA alleles to the immunoediting process. We find that common cancer mutations including BRAF-V600E and KRAS-G12D are predicted to bind none of the common HLA alleles, and are thus "immunogenically silent" in the human population. We identify regions of proteins that are not presented by HLA at a population scale, coinciding with frequently mutated hotspots in cancer, and other protein regions broadly presented across the population in which few mutations occur. We also find that 9/29 common HLA alleles contribute disproportionately to the immunoediting of early oncogenic mutations. These data provide insights into immune evasion of common driver mutations and a molecular basis for the association of particular HLA genotypes with cancer susceptibility.

Keywords: immunoediting, immune evasion, neoantigens, HLA, cancer susceptibility, KRAS, BRAF, TP53

OPEN ACCESS

Edited by:
Peter Brossart, University of Bonn, Germany

Reviewed by:
Kerry S. Campbell, Fox Chase Cancer Center, United States
Catherine Sautes-Fridman, INSERM U1138 Centre de Recherche des Cordeliers, France

***Correspondence:**
John M. Maris
maris@chop.edu

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *13 May 2019*
Accepted: *10 January 2020*
Published: *06 March 2020*

Citation:
Yarmarkovich M, Farrel A, Sison A III, di Marco M, Raman P, Parris JL, Monos D, Lee H, Stevanovic S and Maris JM (2020) Immunogenicity and Immune Silence in Human Cancer. Front. Immunol. 11:69. doi: 10.3389/fimmu.2020.00069

INTRODUCTION

The immune system is thought to play a dual role in carcinogenesis (1–3). First, when a proper immune response is mounted, the immune system is capable of eliminating neoplastic cells arising from early tumor-initiating events (immunoediting). In contrast, the immune system can initiate signaling of wound healing pathways that can help foster an environment conducive to tumorigenesis. The human leukocyte antigen (HLA) proteins present a snapshot of all nucleated cell's proteomes on the cell surface for surveillance by T cells. While an individual harbors six distinct HLA Class I alleles (A, B, and C), a total of 13,145 unique Class I alleles have been characterized to date at these highly polymorphic loci (4). Presentation of processed pathogen-derived peptide by at least one of these HLA alleles is a prerequisite for the initiation of an adaptive immune response. Each HLA allele possesses the ability to present a distinct set of peptides to the immune system, based on the biophysical properties within the peptide binding groove which restrict specificity to a limited set of available peptides. Peptide binding is largely dictated by two HLA-facing anchor residues, which are restricted to a few amino acids at these positions (5).

Recently, algorithms such as NetMHC have allowed the prediction of binding affinity of peptide sequences to specific HLA alleles, resulting in the correct prediction of >75% of binders, with positive predictive values in the range of 90–95% (6–10).

Presented neoantigens can be divided into two distinct classes: group 1 resulting from mutations in the TCR-facing residues and correspondingly less likely to change the binding affinity of the peptide/HLA complex, and group 2 resulting from the anchor residues of the peptide, and thus presenting a longer sequence of novel polypeptides to the immune system, as compared to single-residue alterations in group one antigens (5). A properly mediated interaction between the HLA protein, presented peptide, and T cells serves to maintain the genomic integrity of the organism by eliminating cells harboring foreign genetic material both from external pathogens and those arising from somatic mutations. The tumor immunoediting theory predicts that early pathogenic events giving rise to precancerous cell growths can be eliminated by the adaptive immune system unless cancer cells evolve the ability to escape this selective pressure (11).

While it has become increasingly appreciated that the adaptive immune system has the potential to play a significant role in the elimination of existing tumors, its role in the clearance of cancer cells during early initiating events has remained difficult to study. Though it has been well demonstrated that immunosuppression in humans is linked to an increased incidence of cancer (12–15), it has remained difficult to quantify early immunoediting events, and to attribute the clearance of precancerous lesions in immunocompetent individuals to clearance of tumor-derived neoantigens, as opposed to other mechanisms such as the elimination of cells harboring cancer-inducing viruses.

Here, using the large cohort of patients characterized by the TCGA, we created an immunogenicity map of neoantigens resulting from 125 consensus cancer-related genes then employ a mathematical model to estimate early immunoediting events by quantifying the underrepresentation of specific neoantigens and their potential HLA pairs as a metric for prior clearance of these neoantigens by the immune system. We also employ an HLA presentation score to characterize population-scale HLA presentation along the span of individual proteins to uncover both "protected" (peptides derived from protein regions highly represented across the population by HLA) and "unprotected" regions (those not presented by common HLAs) within proteins commonly mutated in cancer. Further, we introduce an HLA-centric metric of immunoediting, allowing the modeling of the degree of immunoediting in early tumor initiating events that generate strongly immunogenic antigens presented by MHC Class I at a population scale occurring across various point mutations, histotypes, patients, and HLA alleles. We have released a companion web application that can be used to explore immunogenicity of tumor neoantigens across the population and HLA presentation along the span of individual proteins for the identification of shared tumor neoantigens and tumor vaccine design (http://reslnmaris01.research.chop.edu:3838/shinyNAP/).

RESULTS

HLA Immunogenicity Map Shows That Common Mutations Generate Peptides That Are Immunogenically Silent

To generate a map of immunogenicity across all frequently observed single nucleotide mutations in cancer driver genes, we filtered DNA sequencing data from the TCGA (7,300 subjects representing 33 cancer histologies) for all observed mutations harbored in 125 consensus oncogenes and tumor suppressor genes [(16), **Table S1**], resulting in 26,361 unique variants (**Figure S1**). For each variant we then generated 17 mer amino acid sequences to cover all possible 9 mer peptides resulting from amino acid sequences flanking the mutated site (9 potential variants per mutation), and calculated binding affinity using NetMHC-4.0 across 84 common HLA Class I alleles (8), with the alleles studied estimated to represent at least one allele in 99.4% of the US population (calculated based on the ethnicity-adjusted allele frequency from the Bone Marrow Registry) (17). This analysis generated ∼237,000 potential neoantigens, each with a predicted MHC binding affinity across all available HLA alleles, resulting in $\sim 20e^6$ binding affinities. From these data we aggregated all neoantigens arising from each individual mutation and filtered for neoantigens classified as strong binders (≤0.5% rank binding for its HLA alleles), thereby identifying those neoantigens most likely to induce an immune response resulting in their immunoediting and elimination during early tumorigenesis (**Figure 1**, **Table S2**; STAR methods in **Supplementary Material**). We find that 211,852 of 2,214,324 (9.6%) aggregated mutant/HLA pairs derived from cancer driver proteins are predicted to be strong binders.

Upon analyzing HLA binding by neoantigens derived from each mutant variant, we observed that while the majority of mutants produced binders across multiple HLA alleles, a subset of mutations was predicted to not produce any strong binders across all of the 84 HLA alleles studied, here defined as immunogenically silent mutations. A total of 1,806 putative neoepitopes (6.85% of characterized variants) were predicted not to bind any of the 84 HLA alleles with high affinity (**Figure 2A**, **Table S2**), including the common mutations BRAF V600E and KRAS G12D, extending the observations of Marty et al. (18). For each variant, we calculated the probability of a TCGA patient carrying at least one allele capable of binding that variant (median neoantigen presented across 26.5% of the population), and found that the most common variants in the TCGA were enriched in those mutations predicted to bind HLA less frequently across the population (**Figure 2A**, **Table S2**, $p = 0.00035$), suggesting that mutations generating neoantigens capable of being presented with high affinity by multiple, or more common HLA alleles, have a higher likelihood of being eliminated through immunoediting, and thus are underrepresented in the TCGA. The highest ranking of these, MAP3K1 D3727Y, is predicted to bind MHC with high affinity in 94.5% of the population, producing various epitopes with strong binding affinity to 50 of 84 HLA alleles tested, and this mutation is only observed once in the entire TCGA dataset. Analysis across all variants deriving the most commonly mutated gene, *TP53,* reveals that common variants are highly enriched in loci that are not capable of generating neoantigens (**Figure 2B**,

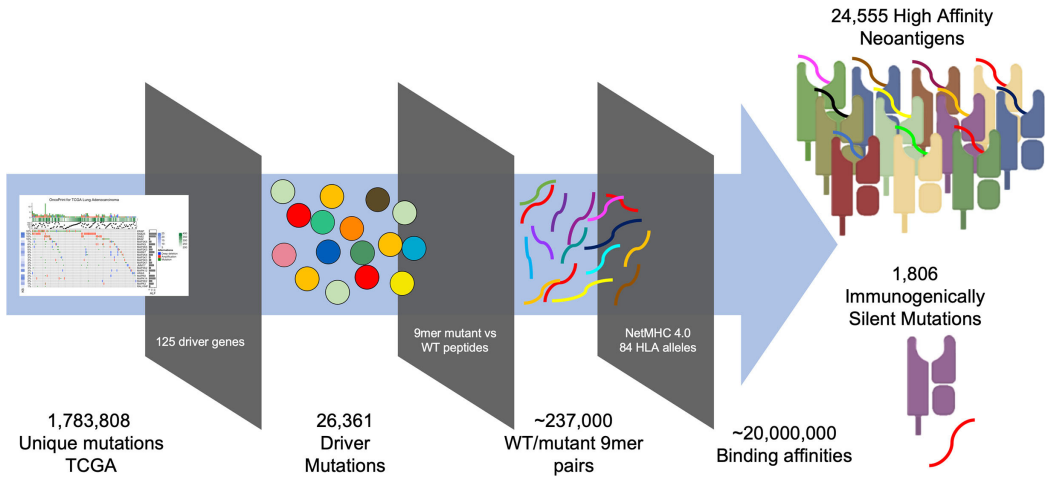


FIGURE 1 | Pipeline for generating immunogenicity map. TCGA mutations imputed by MuSE and Somatic Sniper were filtered for those arising from 125 characterized driver mutations (16). 17 mer peptides sequences were generated for all possible 26,361 variants arising from (covering nine potential epitopes for each variant) each variant/WT pair (~237,000 pairs) and analyzed for MHC binding affinity using NetMHC 4.0 across 84 HLA alleles resulting in $20e^6$ binding affinities. Peptide affinities were aggregated for all peptides arising from each variant and filtered for those that are strong binders (<0.5% rank), revealing 24,555 strong binders and 1,806 immunogenically silent mutations.

$p = 0.008$). Furthermore, our analysis reveals a wide range in the breadth of tumor neoantigens restricted to specific HLA alleles, ranging from 3.5% of neoantigens arising from driver genes bound by HLA-B*37:01 to 18.6% of neoantigens bound by HLA-B*83:01 (**Figure 2C**).

Having identified immunogenically silent neoepitopes, we hypothesized that HLA alleles evolved to preferentially present particular protein motifs, which may leave other regions of cancer proteins unprotected. We applied the algorithm used to generate the immunogenicity map, calculating the presentation scores of 9 mers starting at each amino acid along the span of the entire protein across 84 HLA alleles and calculating the percentage of the population predicted to present a given peptide. We then determined the combined population-wide presentation scores of the neighboring eight amino acids in either direction to calculate the regional immunity of the protein to represent all 9 mers centered at each amino acid position. We generated immune presentation maps of TP53, PI3KCA, and BRAF, and mapped this onto the frequency of point mutations at each amino acid (**Figures 3A–C**). We find that common hot-spot mutations frequently occur in protein domains that score low in population-scale HLA presentation, while domains that are widely protected rarely harbor recurring mutations, suggesting that mutations in unprotected regions are enriched in cancer due in part to consequent immune evasion. We also mapped the presentation score of neoantigens in TP53 onto the regional HLA presentation score (**Figure 3D**), observing a strong overlap between presentation scores for individual neoantigens and the corresponding wild-type protein scores, suggesting that regional scores are a good predictor of the immunogenicity of neoantigens, particularly group 1 neoantigens that do not alter the HLA restriction of the peptide.

In addition to understanding the immune evasion of proteins arising from unprotected domains of proteins, we sought to apply regional HLA presentation to identifying shared tumor epitopes derived from clinically-relevant oncogenes that can be broadly therapeutically applicable across the widest population of patients. We performed mass spectrometry on 16 neuroblastoma tumors to characterize the ligandome and test the predictive ability of the HLA regional scoring across the span of a protein. We mapped the regional presentation score of the most highly represented protein in the neuroblastoma ligandome, NPY (29 MHC Class I peptides detected in 16 neuroblastomas), finding a highly significant concordance between the empirically detected peptides and those regions of the protein expected to be highly presented (**Figure 3E**; $p = 0.000011$), and find no peptides in the ligandome derived from the signal peptide region (aa 1–28) which is cleaved from the full-length pro-NPY protein. Based on the high degree of presentation across the NPY protein across 68/84 HLA alleles, its high level of differential expression (**Figure S6**), and its role in promoting tumor growth (19), we prostulate that NPY is a promising candidate for vaccination strategies. Surprisingly, we find that despite the elevated population presentation score in the highly presented regions, none of the peptides presented in these regions are predicted to bind to HLA-A*02:01, highlighting the utility of a population-scale analysis of HLA presentation in identifying broadly presented epitopes that may be overlooked due to lack of presentation by the most common HLA alleles. We next searched the neuroblastoma immunopeptidomics dataset we created for peptides derived from the *MYCN* oncogene, a major cancer driver in neuroblastoma, finding only a single peptide (KATEYVHSL) presented on the relatively rare HLA-C*16:01 allele representing <5% of the population (**Figure 3F**). Applying the HLA protein scoring map, we find that this peptide is predicted to bind strongly to 10/84 HLA alleles, representing 31.9% of the population (ranking 15th of 456 peptides in population binding score), and suggesting that this peptide can

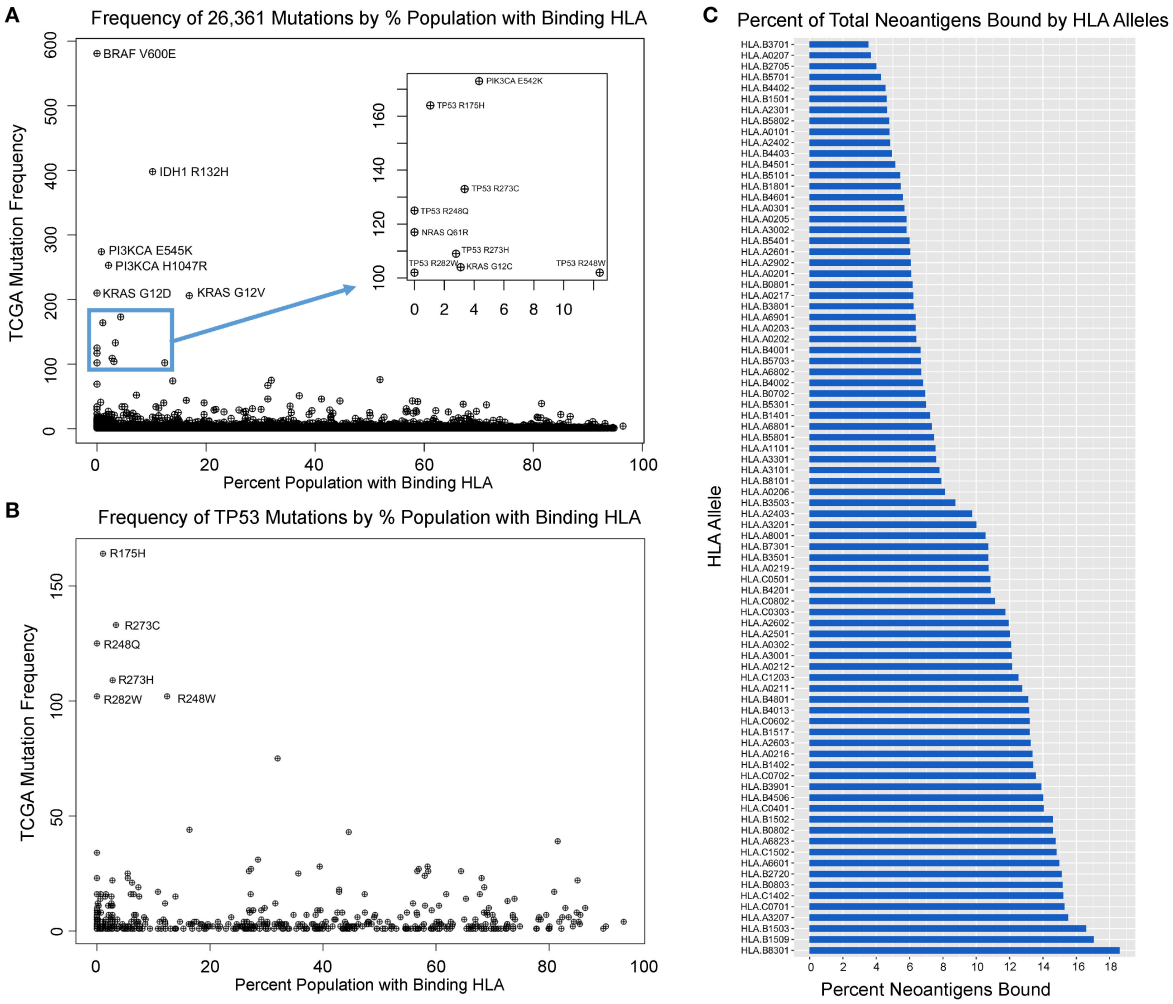


FIGURE 2 | Mutation frequency in TCGA inversely correlates with population-scale HLA binding. **(A)** Mutation frequency for 26,361 TCGA variants compared to the proportion of the population predicted to bind a given neoantigens derived from that mutation reveals immunogenically silent mutations in 1,806 variants (6.85% of mutation in the TCGA, including BRAF V600E and KRAS G12D highlighted) and enrichment of common driver mutations in those which are predicted to not bind any HLA alleles ($p = 0.0004$). **(B)** TP53 mutations are enriched in those that are immunogenically silent or bind HLA alleles in a small portion of the population, and mutations with high probability of binding at least one allele in a given patient are significantly underrepresented in the TCGA ($p = 0.008$). **(C)** Proportion of characterized neoantigens bound by individual HLA alleles reveals broad range in diversity of binding from 3.5% of neoantigens arising from driver genes bound by HLA-B*37:01 to 18.6% bound by HLA-B*83:01.

have significantly broader application as a therapeutic target in this pediatric cancer population. This peptide overlaps with the previously reported immunogenic HLA-A*02:01 peptide VILKKATEYV (20), suggesting that immunization using this region of MYCN may have wider implications beyond HLA-A*02:01 patients. Using our analysis, we find that the highest scoring MYCN peptide (TVRPKNAAL) has predicted binding to 9 HLA alleles, representing 58.1% of the population, and we expect analysis of more neuroblastoma tumor specimens will validate this prediction. We further analyzed regional scores across 17 and 33 mers, we find that these regions are predicted to generate peptides binding to 73.1 and 85.4% of the population, respectively (**Figure 3F**). We suggest that these tools can be used to design and prioritize more broadly applicable therapeutic targets and vaccines for cancer, particularly when paired with ligandomics data (21). Analyses of population-scale presentation along the span of individual proteins and of specific neoantigens is available through the Shiny-NAP web application (http://reslnmaris01.research.chop.edu:3838/shinyNAP/).

HLA Allelic Immunogenicity and Cancer Susceptibility

We hypothesized that specific HLA alleles capable of generating strongly bound neoantigens would be underrepresented in the population of TCGA patients harboring those variants due to early immune based elimination of neoplastic cells. To validate this prediction, we inferred HLA haplotypes from individual patient sequencing data using the PHLAT HLA typing algorithm (22), resulting in 563 unique HLA alleles characterized across the TCGA. To assure predication accuracy, we directly genotyped 15

unique HLA allele calls with the lowest confidence predictions using DNA from cancer cell lines and showed 100% concordance (**Table S3**). We then used the immunogenicity map to determine neoantigen binding across all TCGA driver mutations as defined above (**Table S4**). To estimate the degree of immunoediting across each HLA allele, we used the immunogenicity map to generate a list of all predicted strong binders at each HLA allele having greater than 5% population frequency in the TCGA (29/84 alleles), subset the unique set of patients harboring these mutations, and compared the frequency of their HLA alleles to predicted in the TCGA population from the Bone Marrow Registry data derived largely from younger adults (**Figure S2**). Performing this analysis across 29 HLA alleles, we calculated the magnitude of immunoediting for each allele by the underrepresentation of that allele in the subset of patients harboring strong neoantigens for that allele, as compared with the predicted TCGA population frequency calculated from the bone marrow registry data (**Figure 4A**). As a metric of immunoediting by each allele, we calculated the proportion of observed HLA frequencies in the neoantigen-harboring cases compared to the population frequency, in which a proportion of 0 would represent perfect immunoediting of all early neoantigens and 1 would represent no immunoediting (**Figure 4B**; STAR methods in **Supplementary Material**).

We observe a wide range of immunoediting across the 29 HLA alleles, the highest of which was HLA-A*68:01 with a 44% underrepresentation in the subset of patients harboring predicted binders. We also found that many common HLA alleles appear to not contribute significantly to immunoediting in cancer. Nine of 29 HLA alleles were significantly associated with a protective effect against early neoantigens (**Figure 4**, **Table S2**; $p = 7.7e^{-6}$-0.05, FDR = 0.0001–15.9%). Our modeling suggests that individual HLA alleles have differential ability to initiate an immune response capable of clearing early oncogenes, and across differential breadths of variants. We generated an immunoediting score to determine the degree to which each HLA allele was protective, factoring both breadth and magnitude (% neoantigens bound * % allele underrepresentation), and determined HLA-A*68:01 to score the highest in protectivity against mutations in cancer driver

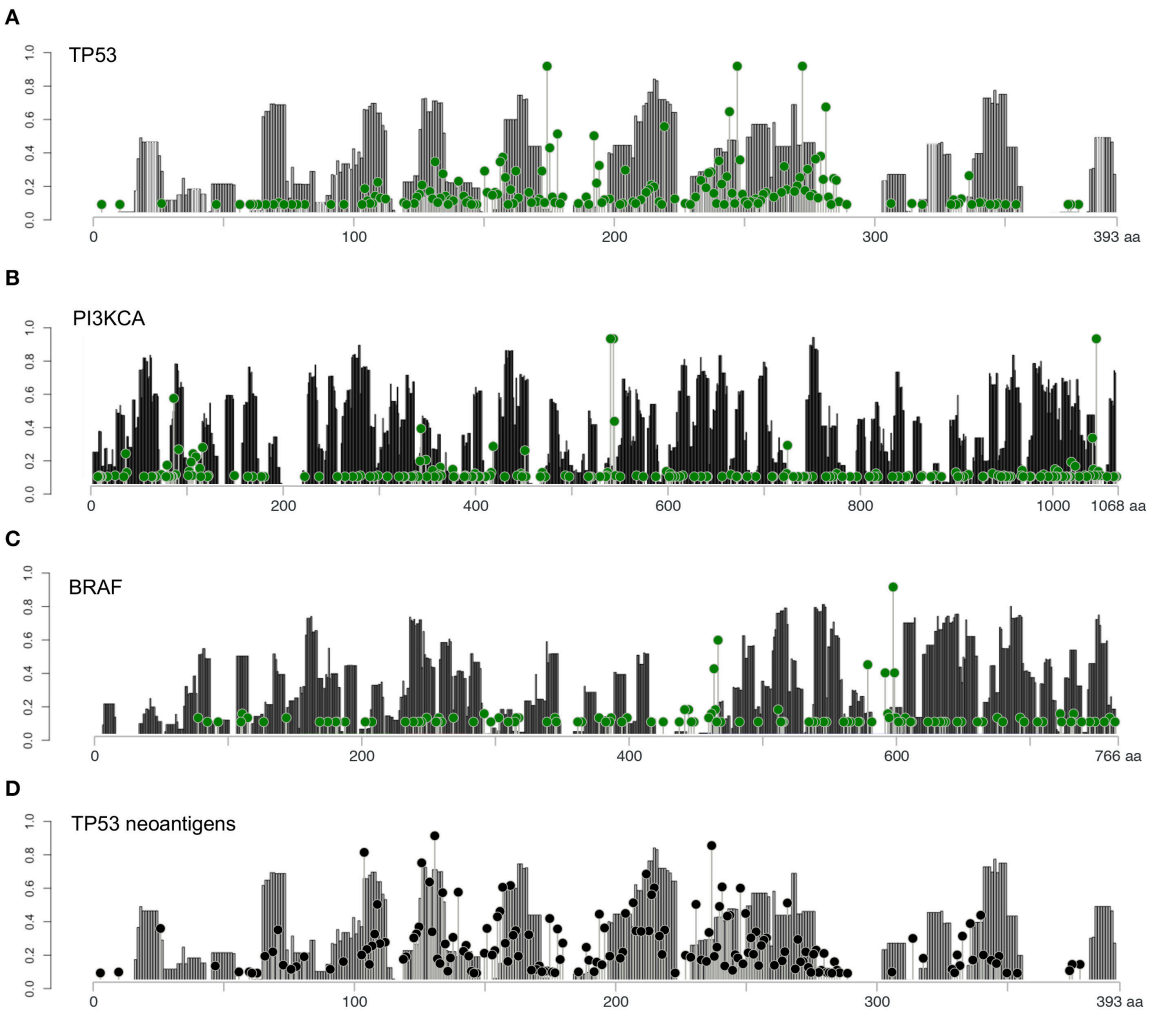


FIGURE 3 | Continued

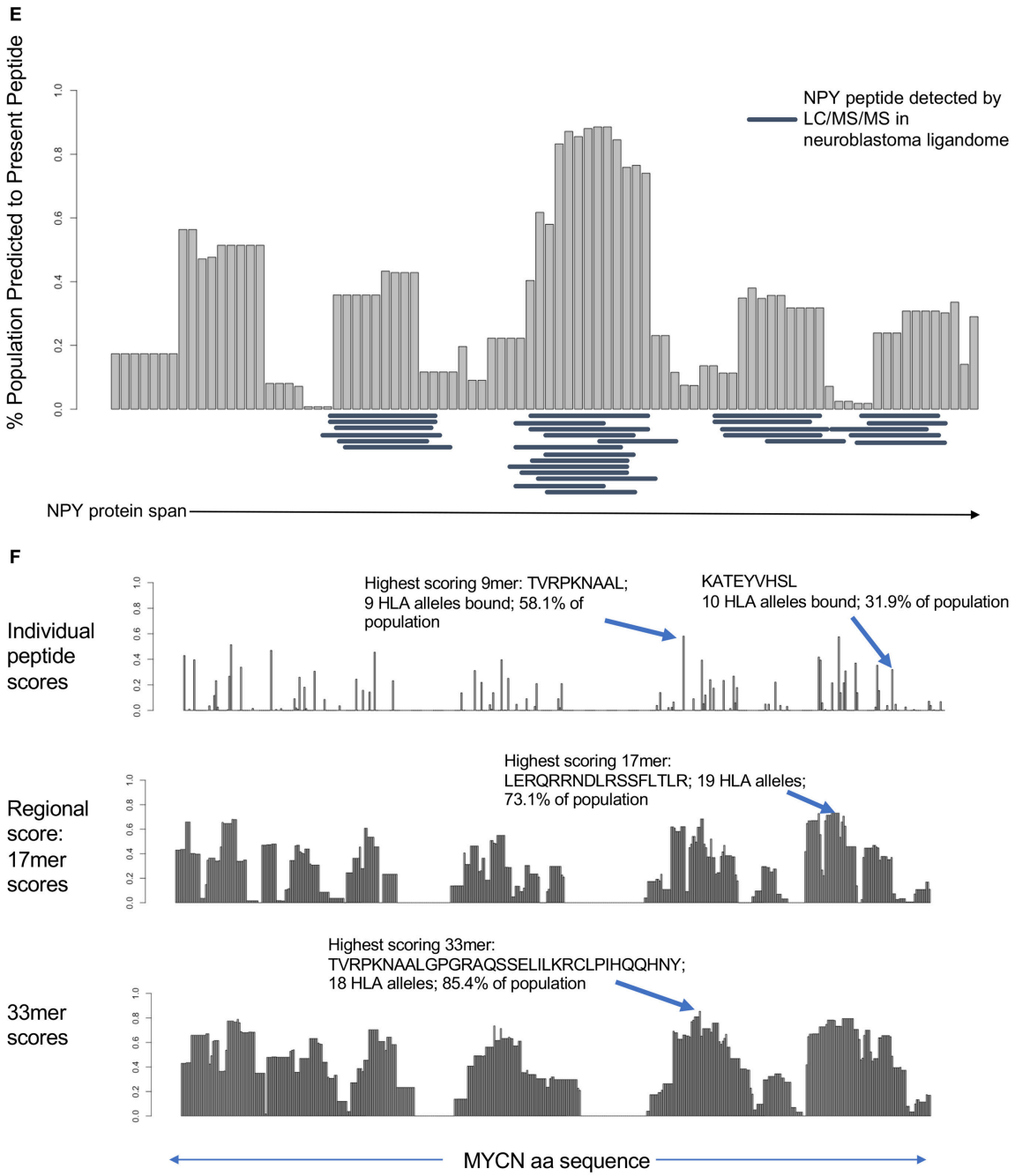


Figure 3 | HLA presentation score by protein region reveals unprotected domains and protein regions broadly applicable as cancer vaccine candidates. **(A–C)** Regional HLA presentation scores determined by calculating the fraction of the population capable of presenting a 9 mer epitope centered at each amino acid location in p53, PI3K, and BRAF (gray bars). Each protein contains domains that are presented on HLA in the majority of the population and other domains that are unprotected. Mutation frequency at each location overlaid (green lollipop plot) reveals that many common mutations are found in unprotected regions of the protein. **(D)** Overlay of mutated neoantigen presentation score with regional presentation score for p53 shows strong correlation between regional protection of WT protein (gray bars) with protection against mutated neoantigens (black lollipop plot). **(E)** Twenty nine peptides detected by ligandomics in 16 neuroblastoma tumors were mapped onto the HLA population presentation scores. Empirically detected peptides were highly enriched in high-scoring regions of the protein ($p = 0.000011$). **(F)** Analysis of MYCN HLA presentation across the span of the protein. Analysis of individual peptides (top) reveals the most highly presented peptide derived from MYCN, TVRPKNAAL, to be presented on 9 HLA alleles, representing 58.1% of the population. KATEYVHSL peptide, detected by ligandomics is predicted to be presented on 10 total HLA alleles (31.9% of the population). Analysis of 17 mer regions (middle) reveals a peptide LERQRRNDLRSSFLTLR generating peptides predicted to bind 19 HLA alleles (73.1% of the population). Analysis of 33 mer regions reveals the highest scoring peptide TVRPKNAALGPGRAQSSELILKRCLPIHQQHNY presented on 18 HLA alleles in 85.4% of the population, suggesting these as promising regions of the MYCN protein for broadly applicable vaccination.

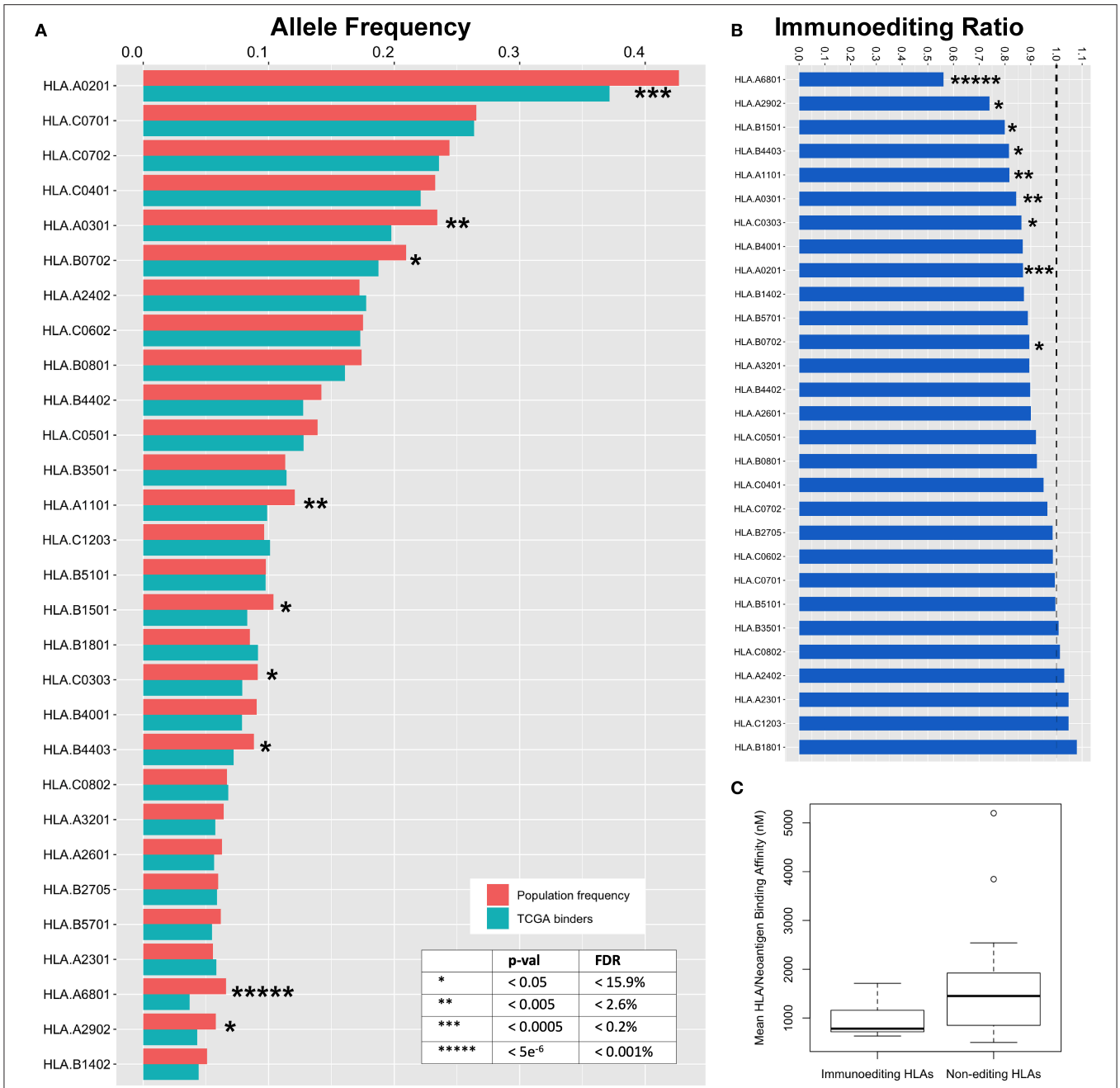


Figure 4 | Immunoediting across HLA alleles. **(A)** Expected TCGA frequency for each allele as calculated from the Bone Marrow Registry and adjusted for ethnicity (red) compared to observed HLA allele frequency in subset of the TCGA population harboring strong neoantigen/HLA binding pairs (blue). To calculate observed HLA frequencies, predicted mutant binders are used to filter TCGA data and HLA frequencies are deduced from the subset of unique individuals. Decreased frequency in TCGA binders compared to expected frequency is a surrogate metric for early immunoediting events that are responsible for having cleared early tumors and are thus underrepresented in TCGA data. **(B)** Ratio of observed:predicted binders. Area below dotted line represents immunoediting frequency for each allele. Zero represents complete immunoediting while 1 represents no contribution of immunoediting by a particular HLA allele. **(C)** Binding affinity between HLA alleles found to participate in immunoediting of cancer neoantigens is stronger than non-contributing alleles ($p = 0.015$).

genes (**Figure S3A**, **Table S2**). We postulated that the binding affinity between those HLA alleles and tumor neoantigens may contribute to immunoediting. By comparing the nine HLA alleles that contribute to immunoediting to the 21 alleles that were found to not contribute significantly to immunoediting, we find that the mean binding among neoantigens in the alleles contributing to immunoediting is 982.8 nM, significantly stronger than 1,612.6 nM in non-contributing alleles (**Figure 4C**,

$p = 0.015$), suggesting that differential affinity for peptides among HLA may contribute to their ability to initiate an anti-tumor response against neoantigens. Furthermore, we scored the contribution of these alleles at a population level as a metric for how many neoantigens these alleles are editing at a population scale by calculating the product of the immunoediting score with the frequency of the HLA allele in the US population (**Figure S3B**, **Table S2**), with HLA-A*02:01 emerging as the most significant contributor to population-scale immunoediting.

Immunoediting in TCGA Across Patients, Tumors, and Point Mutations

We next focused on immunogenically silent mutations defined above and showed that these are enriched in the TCGA dataset, comprising 9.1% of observed mutations, while only representing 6.85% of all characterized variants ($p = 2e^{-16}$). Indeed, we find that 1,974 of 7,300 patients (27%) in the TCGA harbor at least one immunogenically silent mutation. In addition to immunoediting contributions of HLA alleles and evasion through immunogenically silent mutations, we explored editing events across individual patients, tumor types, and specific point mutations. To model the degree of immunoediting across these variables, we calculated the expected frequency with which a given patient, tumor, or point mutation, should harbor at least one HLA allele capable of binding the subset of strongly binding neoantigens in that group, calculated the collective probabilities of the subpopulation generating neoantigens, and compared the observed frequencies to the predicted frequency of binding to at least one HLA allele. To simplify the modeling of predicted allele frequencies, we treated HLA allele frequencies as independent events and disregarded linkage between HLA alleles, thus slightly underestimating the expected frequencies and biasing our prediction of immunoediting toward a more conservative estimate. These analyses lend insight into tumor cell extrinsic mechanisms for recurrent hot spot mutations in cancer (**Table S5**). For example, while KRAS G12D is immunogenically silent and is predicted to bind no common HLA allele with high enough affinity to allow for recognition by T cells, the KRAS G12A mutation is predicted to generate strong neoantigens with four HLA alleles (HLA-A*02:05, HLA-A*02:06, HLA-A*69:01, and HLA-C*03:03). KRAS G12A scored as the 2nd most significantly underrepresented driver SNV of the 26,361 mutations analyzed in the TCGA and was not observed in any patient tumor sample with any of the four predicted HLA binders in its 35 occurrences (4.3 predicted neoantigen/HLA pairs, **Table S5**, $p = 0.01$).

We also examined immunoediting contributions of group 1 and group 2 neoantigens (as defined above) in HLA-A*02:01 by comparing neoantigens arising from mutations in the anchor positions at residues two and nine to those mutations outside of these residues. We found no significant difference in underrepresentation between group 1 and 2 neoantigens (**Figure S4**), suggesting that both types of neoantigens participate in immunoediting in the context of HLA-A*02:01. Analyzing immunoediting across individual patients, we found a subset of patients with disproportionate degrees of immunoediting. The most significant of these, uterine cancer patient TCGA-E6-A1LX, harbored 3.4-fold fewer immunogenic neoantigens than predicted ($p = 3.1e^{-10}$). Interestingly, this patient also harbored 12 immunogenically silent mutations (ranking 3rd highest in immunogenically silent mutations out of 7300 TCGA patients, **Table S6**), suggesting that this patient's tumor had been subjected to significant immunoediting of secondary mutations, yet was being driven largely by immunogenically silent mutations. Although the combined cases of uterine cancer taken together are the least significantly immunoedited histology, a particular subset of these patients are highly enriched among individuals with the highest degree of immunoediting (five out of the top 10 immunoedited individuals in the TCGA), highlighting these as interesting case studies for the mechanisms underlying enhanced early immunoediting and eventual immune escape. In our analysis across histologies, we find that glioblastoma is the most significantly immunoedited in early tumor formation (**Figure S5**; $p = 0.008$), in line with recent evidence that these tumors do not arise in immune privileged sites (23, 24). Overall, these data illustrate differential degrees of immunoediting and immune evasion across TCGA patients, histologies, and variants, suggesting that an enrichment in immunogenically silent mutations may be driving the evolution of tumors in otherwise immunocompetent individuals.

DISCUSSION

Here we describe a model for quantifying immunoediting during early tumorigenesis that provides insight into immunologic contributions to recurrent somatic mutation hotspots observed in human cancer, as well as immunologic contributions to cancer susceptibility. The model described herein employs orthogonal methods to recent studies in demonstrating evidence of immunoediting in the TCGA cohort (18, 25). Using an HLA-based hypothesis, we converge on the conclusion that common driver mutations evade the immune system and provide a population-scale HLA-centric basis for their overrepresentation in human cancer. In each of these studies, the immunoediting process is demonstrated to be imperfect, which can in part be explained by the false-positive predicted peptides included in the analysis which dilute the contribution of peptides actually presented, but also raises questions about disparities in immune responses. Here we provide methods that can be employed to elucidate immunoediting across HLA alleles, patients, individual variants, and other genomic or clinical features. We think that our HLA-centric population-scale model provides a baseline of comparison against which we can estimate the degree of immunoediting, with several examples of disparities across these features highlighted in this manuscript.

This is the first report that we are aware of to map known driver neoantigens across common HLA alleles, and quite strikingly the most recurrent hotspot mutations in human cancer are predicted to bind no common HLA allele with high enough affinity to engage the adaptive immune system, highlighting an immunologic mechanism underlying the evolutionary advantage

of common mutations in addition to their oncogenic potency. This is also the first report of which we are aware of to quantify the contributions of individual HLA alleles to the immunoediting process in cancer, revealing a high disparity in immune protection against cancer across the HLA alleles. Our data suggest that the ability of individual HLA alleles to bind cancer neoantigens with high affinity is strongly associated with its ability to contribute to cancer immunoediting. Though our data support the cancer immunoediting theory, we and others show that the immunoediting process leads to incomplete elimination of neoantigens arising from early driver mutations. Our findings suggest that a significant difference in p/MHC binding affinities contributes to the disparity in immunoediting among HLA alleles, but further studies will be necessary to corroborate these findings and uncover additional mechanisms contributing to these disparities. While we demonstrate that a significant degree of immune evasion may be attributed to immunogenically silent mutations, the absence of complete immunoediting in patients not harboring immunogenically silent mutations may be attributed to factors including tumor intrinsic immune evasion, downregulation of MHC, lack of T cell response of the proper magnitude and quality, poor TCR repertoire, exclusion of T cells from tissue, or peripheral tolerance. We believe that our model can be coupled with genomic surrogates of these features to interrogate these variables in future studies using tumor genomic data.

Here we show that not all HLA alleles are found to be significantly protective against the neoantigens that they present. We postulate that alleles not found to significantly participate in immunoediting may induce sublethal T cell responses or possess biophysical and/or geometric properties that confer suboptimal interactions with germline-encoded binding regions of the TCR. Our findings that specific regions of cancer driver proteins are unprotected by HLA presentation combined with the disparity in binding across HLA alleles raises the question of whether HLA alleles have evolved to confer protectivity against particular viral domains that coincide with motifs found in cancer proteins, and whether the unpresented areas remain unprotected due to lack of evolutionary pressure on these motifs. These results also raise the question of whether HLA presentation of group 2 neoantigens is associated with mutational signatures arising from particular groups of DNA damage that generate variants with more favorable binding properties in the anchor residues (26). We have made available the tools for other investigators to test these and other hypotheses (http://reslnmaris01.research.chop.edu:3838/shinyNAP/).

We also present a tool for mapping the presentation scores across the span of any given protein in the population. We find a highly significant concordance between the peptides empirically detected in the combined ligandome of 16 neuroblastoma tumors carrying various HLA alleles, and the regions of the NPY protein predicted to be most highly presented by HLA across the population. Based on these results paired with the high level of differential expression, and its role in promoting tumor growth, we suggest that NPY is a promising candidate for vaccine development for neuroblastoma patients. Using this tool, we also suggest that current vaccination strategies used against MYCN may have broader application across the population. As efforts are being made to develop HLA-agnostic CAR T cells, specific to only the peptide and not the MHC, we believe these tools will be useful in identifying broader segments of the cancer population likely to benefit from these immunotherapies.

As access to genomic data from cancer patients continues to expand, and the peptide/MHC binding and T cell epitope prediction tools improve, this model will benefit from additional statistical power in stratifying subsets of the patient population based on molecular features occurring in smaller subpopulations. Despite the fact that our model predicted no HLA alleles binding to neoantigens derived from the KRAS G12D mutation, it was recently reported by Tran et al. (27) that KRAS G12D neoantigen GADGVGKSA is able to mediate a T cell response in the context of HLA-C*08:02. This antigen is predicted to be a weak HLA binder (15,390 nM), highlighting the fact that T cell epitopes are not always predicted using this algorithm, particularly on rare alleles for which there are limited training data, and that new methods will help identify neoantigens with non-canonical motifs (28). Here, we restricted our analysis of immunoediting by CD8 T cells through MHC class I presentation of 9 mer antigens only, and did not account for immunoediting that may be triggered by other Class I antigens derived from additional structural variants or non-canonical antigens, peptides of varying lengths, Class II antigens, or the activities of the innate immune system from NK cells or macrophages, as we were focused on maintaining statistical power by using common HLA alleles and the most common Class I peptides for which we were most confident in being contributors to early tumorigenesis.

We find that the highest statistically significant immunoediting takes place in glioblastoma, whereas, taken together, sarcomas, pancreatic tumors, ovarian, adrenocortical tumors and lymphomas show no significant evidence of immunoediting. Given that the immunogenically silent KRAS G12D mutation is pathognomonic of pancreatic cancer, our findings may help explain the lack of efficacy of treatments such as checkpoint inhibitors in pancreatic cancer (29) and the lack of immunoediting observed in our analysis, as these tumors are driven largely by an immunogenically silent mutation. We suggest that our methods could ultimately be used to inform the stratification of groups of patients most likely to respond to immunotherapies such as checkpoint inhibitors based on patient HLA and antigen immunogenicity, and to prioritize shared antigens for vaccine development or HLA-agnostic CAR strategies. This model can also be used to predict how an individual's HLA profile can determine the types of mutations most likely to develop or be protected against.

Using the model of immunogenicity described herein, it may be possible to infer physical properties of neoantigens that elicit high immunoediting as compared to other neoantigens that are presented but not eliminated by the immune system, study the contributions of various molecular pathways across tumor types and across individual patients that contribute to variable degrees of immunoediting, and as a basis for exploring the mechanisms by which specific HLA alleles may contribute to cancer protection and predisposition. With alleles

such as HLA-A*68:01 emerging as disproportionately high in their immunoediting score, we will be interested to see whether contributions of HLA alleles to early immunoediting will translate to improved abilities to induce T cell responses against tumor neoantigens in patients, and whether such alleles are associated with improved outcomes in patients treated with modern immunotherapies. We suggest that the immunogenicity map, HLA typing data, and immunoediting model contained herein will facilitate investigation into neoantigen immunogenicity at the level of HLA alleles, mutations, patients, histologies, and aid in prioritization of shared tumor epitopes for therapeutic development, and further our mechanistic understanding of immune evasion in tumor evolution.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the TCGA and Analyses of population-scale presentation along the span of individual proteins and of specific neoantigens is available through the Shiny-NAP web application (http://reslnmaris01.research.chop.edu:3838/shinyNAP/).

AUTHOR CONTRIBUTIONS

MY and JM: conceptualization, writing, and supervision. MY, PR, MM, SS, DM, HL, and JM: methodology. MY, AF, AS, PR, and JP: software. MY, JP and HL: formal analysis. MY and JP: investigation. MY and JM: funding acquisition.

FUNDING

This work was supported by a St. Baldrick's-Stand Up To Cancer Pediatric Dream Team Translational Research Grant (SU2C-AACR-DT2727) and the Biden Cancer Moonshot (NCI Grant U54 CA232568). Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. This work was also supported by NIH R35 CA220500 and the Giulio D'Angio Endowed Chair.

ACKNOWLEDGMENTS

We thank Erik Lehnert and Michele Mattioni from Seven Bridges Genomics for kindly implementing the HLA typing pipeline on their server. We thank John Warrington for his in help generating figures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.00069/full#supplementary-material

Table S1 | Driver oncogenes and tumor suppressor genes. List of 125 cancer driver genes implicated in carcinogenesis, including oncogenes and tumor suppressor genes that regulate cell fate, cell survival, and genome maintenance (16).

Table S2 | Immunogenicity map of TCGA. Map of all characterized variants derived from 125 cancer driver genes (**Table S1**) and their binding score across each of 84 characterized HLA alleles. Binding score reported as the strongest rank score from the aggregated pool of all peptides resulting from each variant to particular HLA allele, with lower scores representing higher binding affinity relative to binders of each HLA allele.

Table S3 | Validation of HLA genotyping. Comparison of HLA predictions across 15 HLA alleles as inferred from PHLAT algorithm from exome sequencing data of three neuroblastoma cell lines compared to clinical genotyping performed using next generation sequencing of amplified HLA loci.

Table S4 | TCGA mutations and patient HLA types. List of variants used to analyze immunoediting across patients, variants, and histologies. Table populated with all mutations in driver genes from the TCGA with matched HLA typing inferred from PHLAT.

Table S5 | Immunoedited variants from TCGA. List of variants most underrepresented when measured with population of patients harboring HLA alleles predicted to bind neoantigens derived from variant ($p \leq 0.05$). Frequency of mutation is number of occurrences of mutation in **Table S4**. Percent of population with binders is the probability of a TCGA subject harboring an HLA allele capable of binding a neoepitope derived from the particular variant. Observed mutation is frequency calculated from the number of patients with at least one HLA allele from the set of those capable of binding the variant.

Table S6 | Immunoedited subjects from TCGA. List of subjects with highest degrees of immunoediting in the TCGA ($p \leq 0.05$). Expected binders calculated by summing the probability of all individual variants in each patient being bound to an HLA allele in the TCGA. Observed binders is the summed number of variant/HLA pairs that generate at least one epitope across each variant. Observed/expected represents the degree of underrepresentation of presented neoantigens in each patient (0 being perfect immunoediting). Despite being ranked the lowest in significance for immunoediting, uterine cancer represents 5 of the top 10 patients with the most significant degrees of immunoediting. The most significantly immunoedited subject also ranks 3 of 7,300 in number of immunogenically silent mutations.

Figure S1 | Pipeline for inferring HLA type from TCGA and comparing to predicted frequencies. BAM files for individual patients were converted to FASTQ and processed using PHLAT to determine HLA type. HLA frequencies in TCGA were determined using ethnicity-specific allele populations from Bone Marrow Registry and compared to observed frequencies in TCGA. Patient HLA and mutation data were combined to determine number of neoantigens in each individual, allowing the comparison of predicted HLA frequencies to ethnicity-adjusted HLA frequencies in the TCGA across individuals, mutations, and tumor histologies.

Figure S2 | Workflow for modeling immunoediting for individual HLA alleles (example show for HLA-A*02:01). All strong neoantigens predicted to bind given HLA are aggregated and used to filter the TCGA dataset. Resulting mutations are filtered for unique patients to remove patients harboring multiple binders to a single allele. Frequency of unique patients harboring at least one strong neoantigen binding to predicted HLA allele compared to ethnicity-adjusted predicted value for TCGA frequency to determine level of immunoediting by specific HLA allele.

Figure S3 | HLA allele immunoediting scores and population editing scores. Immunoediting scores represent overall ability of HLA alleles to edit mutations, accounting for the repertoire of antigens they are able to bind and the level of editing that they exhibit for that subset of antigens (calculated by % neoantigens bound by allele * % underrepresentation of HLA allele), with HLA-A*68:01 scoring highest in immunoediting of neoantigens arising from mutations in early driver genes. Immunoediting population score is used to estimate the total immunoediting contribution of HLA alleles across the US population (calculated by the product of the immunoediting score with the US HLA allele frequency). **(A)** Immunoediting scores in HLA alleles shown to be statistically significant. **(B)** Population immunoediting scores in HLA alleles shown to be statistically significant.

Figure S4 | Immunoediting of group 1 and group 2 neoantigens. Neoantigens resulting from group 1 neoantigens (those with neoantigens occurring from

mutation at positions outside of anchor residues) were compared with group 2 neoantigens (mutations occurring at anchor residues 2 and 9) in HLA-A*02:01. No significant difference in underrepresentation was found between groups 1 and 2 in HLA-A*02:01.

Figure S5 | Immunoediting by cancer histology. Combined observed binding neoantigens compared to expected. Zero represents complete immunoediting while one represents no contribution of immunoediting by a particular HLA allele. Glioblastoma is the only significantly immunoedited histology in this analysis ($p = 0.008$). Uterine cancer is the least significantly immunoedited tumor, though is highly enriched in individuals exhibiting high degrees of immunoediting (4/8 of the most significantly edited patients in the TCGA, **Table S6**).

Figure S6 | NPY is highly differentially expressed in neuroblastoma and is a promising target for vaccination. RNA-sequencing data from 153 neuroblastoma tumors in TARGET (first column) compared to 1,643 normal tissues from GTEx compiled by organ (subsequent columns) reveals high expression of NPY in neuroblastoma as compared to normal tissues.

REFERENCES

1. Ichim CV. Revisiting immunosurveillance and immunostimulation: implications for cancer immunotherapy. *J Transl Med.* (2005) 3:8. doi: 10.1186/1479-5876-3-8
2. de Sousa JR, Sotto MN, Simões Quaresma JA. Leprosy as a complex infection: breakdown of the Th1 and Th2 immune paradigm in the immunopathogenesis of the disease. *Front Immunol.* (2017) 8:1635. doi: 10.3389/fimmu.2017.01635
3. Prehn RT. The immune reaction as a stimulator of tumor growth. *Science.* (1972) 176:170–1. doi: 10.1126/science.176.4031.170
4. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* (2015) 43:D423–31. doi: 10.1093/nar/gku1161
5. Fritsch EF, Rajasagi M, Ott PA, Brusic V, Hacohen N, Wu CJ. HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunol Res.* (2014) 2:522–9. doi: 10.1158/2326-6066.CIR-13-0227
6. He Y, Rappuoli R, De Groot AS, Chen RT. Emerging vaccine informatics. *J Biomed Biotechnol.* (2010) 2010:218590. doi: 10.1155/2010/218590
7. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE.* (2007) 2:e796. doi: 10.1371/journal.pone.0000796
8. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics.* (2016) 32:511–7. doi: 10.1093/bioinformatics/btv639
9. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.* (2008) 36:W509–12. doi: 10.1093/nar/gkn202
10. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics.* (1999) 50:213–9. doi: 10.1007/s002510050595
11. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity.* (2004) 21:137–48. doi: 10.1016/j.immuni.2004.07.017
12. Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. *Lancet.* (2007) 370:59–67. doi: 10.1016/S0140-6736(07)61050-2
13. Gallagher MP, Kelly PJ, Jardine M, Perkovic V, Cass A, Craig JC, et al. Long-term cancer risk of immunosuppressive regimens after kidney transplantation. *J Am Soc Nephrol.* (2010) 21:852–8. doi: 10.1681/ASN.2009101043
14. Penn I, Starzl TE. Immunosuppression and cancer. *Transpl Proc.* (1973) 5:943–7
15. Coley WB II. Contribution to the knowledge of sarcoma. *Ann Surg.* (1891) 14:199–220. doi: 10.1097/00000658-189112000-00015
16. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science.* (2013) 339:1546–58. doi: 10.1126/science.1235122
17. Gragert L, Madbouly A, Freeman J, Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Human Immunol.* (2013) 74:1313–20. doi: 10.1016/j.humimm.2013.06.025
18. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, et al. MHC-I genotype restricts the oncogenic mutational landscape. *Cell.* (2017) 171:1272–83.e1215. doi: 10.1016/j.cell.2017.09.050
19. Tilan J, Kitlinska J. Neuropeptide Y (NPY) in tumor growth and progression: lessons learned from pediatric oncology. *Neuropeptides.* (2016) 55:55–66. doi: 10.1016/j.npep.2015.10.005
20. Himoudi N, Yan M, Papanastasiou A, Anderson J. MYCN as a target for cancer immunotherapy. *Cancer Immunol Immunother.* (2008) 57:693–700. doi: 10.1007/s00262-007-0409-x
21. Kowalewski DJ, Stevanovic S. Biochemical large-scale identification of MHC class I ligands. *Methods Mol Biol.* (2013) 960:145–57. doi: 10.1007/978-1-62703-218-6_12
22. Bai Y, Ni M, Cooper B, Wei Y, Fury W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genom.* (2014) 15:325. doi: 10.1186/1471-2164-15-325
23. Louveau A, Harris TH, Kipnis J. Revisiting the mechanisms of CNS immune privilege. *Trends Immunol.* (2015) 36:569–77. doi: 10.1016/j.it.2015.08.006
24. Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD, et al. Structural and functional features of central nervous system lymphatic vessels. *Nature.* (2015) 523:337. doi: 10.1038/nature14432
25. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* (2015) 160:48–61. doi: 10.1016/j.cell.2014.12.033
26. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* (2013) 500:415–21. doi: 10.1038/nature12477
27. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med.* (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
28. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity.* (2017) 46:315–26. doi: 10.1016/j.immuni.2017.02.007
29. Winograd R, Byrne KT, Evans RA, Odorizzi PM, Meyer ARL, Bajor DL, et al. Induction of T-cell immunity overcomes complete resistance to PD-1 and CTLA-4 blockade and improves survival in pancreatic carcinoma. *Cancer Immunol Res.* (2015) 3:399–411. doi: 10.1158/2326-6066.CIR-14-0215
30. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol.* (2007) 68:779–88. doi: 10.1016/j.humimm.2007.04.005

Conflict of Interest: The authors currently have a patent pending for methods of prioritizing tumor antigens, which is based on some of the work presented in this manuscript.

frontiers in Immunology

ORIGINAL RESEARCH
published: 12 March 2020
doi: 10.3389/fimmu.2020.00373

Tumor-Infiltrating T Cells From Clear Cell Renal Cell Carcinoma Patients Recognize Neoepitopes Derived From Point and Frameshift Mutations

Ulla Kring Hansen[1†], Sofie Ramskov[1†], Anne-Mette Bjerregaard[1], Annie Borch[1], Rikke Andersen[2], Arianna Draghi[2], Marco Donia[2], Amalie Kai Bentzen[1], Andrea Marion Marquard[1], Zoltan Szallasi[3], Aron Charles Eklund[1,4], Inge Marie Svane[2] and Sine Reker Hadrup[1]*

[1] Department of Health Technology, Technical University of Denmark, Lyngby, Denmark, [2] Center for Cancer Immune Therapy, Copenhagen University Hospital, Copenhagen, Denmark, [3] Danish Cancer Society Research Center, Copenhagen, Denmark, [4] Clinical Microbiomics A/S, Copenhagen, Denmark

OPEN ACCESS

Edited by:
Zlatko Trajanoski,
Innsbruck Medical University, Austria

Reviewed by:
Guilan Shi,
University of South Florida,
United States
John M. Maris,
University of Pennsylvania,
United States

***Correspondence:**
Sine Reker Hadrup
sirha@dtu.dk

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: 31 July 2019
Accepted: 17 February 2020
Published: 12 March 2020

Citation:
Hansen UK, Ramskov S, Bjerregaard A-M, Borch A, Andersen R, Draghi A, Donia M, Bentzen AK, Marquard AM, Szallasi Z, Eklund AC, Svane IM and Hadrup SR (2020) Tumor-Infiltrating T Cells From Clear Cell Renal Cell Carcinoma Patients Recognize Neoepitopes Derived From Point and Frameshift Mutations. Front. Immunol. 11:373. doi: 10.3389/fimmu.2020.00373

Mutation-derived neoantigens are important targets for T cell-mediated reactivity toward tumors and, due to their unique tumor expression, an attractive target for immunotherapy. Neoepitope-specific T cells have been detected across a number of solid cancers with high mutational burden tumors, but neoepitopes have been mostly selected from single nucleotide variations (SNVs), and little focus has been given to neoepitopes derived from in-frame and frameshift indels, which might be equally important and potentially highly immunogenic. Clear cell renal cell carcinomas (ccRCCs) are medium-range mutational burden tumors with a high pan-cancer proportion of frameshift mutations. In this study, the mutational landscape of tumors from six RCC patients was analyzed by whole-exome sequencing (WES) of DNA from tumor fragments (TFs), autologous tumor cell lines (TCLs), and tumor-infiltrating lymphocytes (TILs, germline reference). Neopeptides were predicted using MuPeXI, and patient-specific peptide–MHC (pMHC) libraries were created for all neopeptides with a rank score < 2 for binding to the patient's HLAs. T cell recognition toward neoepitopes in TILs was evaluated using the *high-throughput* technology of DNA barcode-labeled pMHC multimers. The patient-specific libraries consisted of, on average, 258 putative neopeptides (range, 103–397, $n = 6$). In four patients, WES was performed on two different sources (TF and TCL), whereas in two patients, WES was performed only on TF. Most of the peptides were predicted from both sources. However, a fraction was predicted from one source only. Among the total predicted neopeptides, 16% were derived from frameshift indels. T cell recognition of 52 neoepitopes was detected across all patients (range, 4–18, $n = 6$) and spanning two to five HLA restrictions per patient. On average, 21% of the recognized neoepitopes were derived from frameshift indels (range, 0–43%, $n = 6$). Thus, frameshift indels are equally

represented in the pool of immunogenic neoepitopes as SNV-derived neoepitopes. This suggests the importance of a broad neopeptide prediction strategy covering multiple sources of tumor material, and including different genetic alterations. This study, for the first time, describes the T cell recognition of frameshift-derived neoepitopes in RCC and determines their immunogenic profile.

Keywords: renal cell carcinoma, neoepitopes, neoantigens, frameshift mutations, T cell screening

INTRODUCTION

Tumor neoantigens are important targets for the immune system to mediate tumor control. Tumor-specific mutations give rise to altered proteins that are processed into short peptides. These are presented at the cell surface in the context of major histocompatibility complex (MHC) molecules, where they serve as targets for cytotoxic T cell killing of the tumor (1). Compared to shared tumor antigens, which can be expressed at low levels in healthy tissue, neoantigens have the advantage of being uniquely expressed in the tumor. Also, there is less T cell tolerance toward neoantigens since the T cell repertoire has not been negatively selected based on these sequences (2). Therefore, neoantigens are attractive targets for immunotherapy. Untargeted therapies, such as immune checkpoint inhibitors and adoptive T cell transfer with tumor infiltrating lymphocytes, have been shown to increase neoantigen reactive T cells, and the clinical response correlates with the mutational burden and predicted number of neoantigens (3–5). Neoantigens have also been directly targeted in personalized therapies by adoptive transfer of specifically expanded T cells (6, 7) and in personalized neoepitope vaccines (8, 9). The challenge for these strategies is, however, to determine which neoepitopes to preferentially target in each patient. Neoantigen reactive T cells have been detected across a number of solid cancers with high mutational burdens, such as melanoma and non-small cell lung cancer (10–12). The described neopeptides have, however, mainly been derived from single nucleotide variations (SNVs) with less focus on in-frame and frameshift indels, mutation types that are likely to be immunogenic based on their large sequence variance to the germline DNA. Even though the total number of frameshift indels are lower than SNVs, they have been shown to give rise to three times as many predicted high-affinity ($IC_{50} < 50$ nM) neoantigens per mutation compared to SNVs and are highly enriched for mutant-specific binding (i.e., neopeptides for which the wild-type peptide is not predicted to bind the HLA) (13). Hence, this mutation type is potentially highly relevant as a tumor neoantigen target (14, 15).

Clear cell renal cell carcinomas (ccRCCs) are medium-range mutational burden tumors that present with the highest pan-cancer proportion of frameshift indels (13, 16). ccRCCs appear to be immune sensitive, as suggested by high levels of T cells infiltrating the tumor site (17), and clinical benefit can be achieved using cytokine-based immunotherapies with interferon-α and high-dose interleukin 2 (18, 19) and checkpoint inhibitors (20, 21). Nevertheless, the tumor microenvironment of ccRCCs is characterized as highly immunosuppressive (22), which is reflected by the poor functional quality of T cell responses observed, with implications for adoptive cell therapy (23).

To our knowledge, as yet, no reports have described the neoantigens recognized by T cells in ccRCC and investigated the contribution of frameshift indels to T cell recognition of neoantigens. Such investigation is critical for using neoantigens as therapeutic targets and biomarkers of relevance to immunotherapy in this cancer type. For that reason, we evaluated the T cell recognition of neopeptides predicted from SNVs, in-frame, and frameshift indels in six ccRCC patients previously described in (23). The prediction was performed with WES from two sources of tumor material (TCL and TF) to include all potential neopeptides in our screenings.

MATERIALS AND METHODS

Patient and Healthy Donor Samples

Healthy donor samples were collected by approval of the local Scientific Ethics Committee, with donor written informed consent obtained according to the Declaration of Helsinki. Healthy donor blood samples were obtained from the blood bank at Rigshospitalet, Copenhagen, Denmark. All samples were obtained anonymously. Peripheral blood mononuclear cells (PBMCs) from healthy donors were obtained from whole blood by density centrifugation on Lymphoprep (Axis-Shield PoC, cat# 1114544) in Leucosep tubes (Greiner Bio-One, cat# 227288) and cryopreserved at −150°C in fetal calf serum (FCS, Gibco, cat#10500064) + 10% dimethyl sulfoxide (DMSO, Sigma-Aldrich, cat#C6164).

Tumor-infiltrating lymphocytes (TILs), tumor fragments (TFs), and tumor cell lines (TCLs) from ccRCC patients were obtained at the Department of Oncology and Center for Cancer Immune Therapy, Copenhagen University Hospital, Denmark, under approval by the Ethics Committee of the Capital Region of Denmark and the Danish Data Protection Agency. Young TIL cultures were obtained from resected tumor lesions from individuals with ccRCC with a Fuhrman grade between 1 and 3 (23). Tumor lesions were resected following surgical removal, and TFs were cultured individually in complete medium [RPMI1640 + GlutaMAX^{TM} (Gibco, cat#61870010) with 10% human serum (Sigma-Aldrich, cat#H3667), 100 U/ml penicillin (P/S, Sigma-Aldrich, cat#P0781), 100 μg/ml streptomycin (P/S, Sigma-Aldrich, cat#P0781), 1.25 μg/ml fungizone (Bristol-Myers Squibb), and 6,000 U/ml IL-2 (Proleukin, Novartis, cat#200-02)] at 37°C and 5% CO_2, allowing TILs to migrate into the

medium. TILs were expanded to reach $>50 \times 10^6$ total cells originating from ∼48 individual fragments, which had expanded to confluent growth in 2 ml wells and eliminated adherent tumor cells (average of $\sim 2 \times 10^6$ cells per well from each TF). TIL cultures were further expanded using a standard rapid expansion protocol (REP) as previously described (24). Briefly, TILs were stimulated with 30 ng/ml anti-CD3 antibody (OKT-3, Ortho Biotech) and 6,000 U/ml IL-2 in the presence of irradiated (40 Gy) allogeneic feeder cells (healthy donor PBMCs) at a feeder/TIL ratio of 200:1. Initially, TILs were rapidly expanded in a 1:1 mix of complete medium and REP medium [AIM-V (Invitrogen) + 10% human serum, 1.25 μg/ml fungizone, and 6,000 U/ml IL-2], but after 7 days, complete medium and serum were removed stepwise from the culture by adding REP medium without serum to maintain cell densities around $1–2 \times 10^6$ cells/ml. TIL cultures were cryopreserved at −150°C in human serum + 10% DMSO.

DNA and RNA Extraction and Sequencing Preparation

DNA and RNA were extracted and purified from TCLs, TFs, and TILs (germline DNA reference) using the AllPrep DNA/RNA Mini kit (Qiagen, cat#80204), with the addition of DNase during RNA purification (Qiagen, 79254). Next, DNA/RNA concentrations were analyzed by NanoDrop (Thermo Fischer Technologies), and RNA RIN values were analyzed by 2100 Bioanalyzer (Agilent Technologies). DNA whole-exome and RNA sequencing (RNAseq) were performed at the DTU Multi Assay Core (DMAC).

Next-Generation Sequencing Data Processing

Raw FASTQ files from whole WES and RNAseq were analyzed in the following manner. First, both data sets were pre-processed for quality using Trim Galore version 0.4.0 (25), which combines the functions of Cutadapt (26) and FastQC 0.11.2 (27): trimming the reads below an average Phred score of 20 (default value), cutting out standard adaptors such as those from Illumina, and running FastQC to evaluate data quality. Variant calling was performed following the Genome Analysis Toolkit (GATK) best practice guidelines for somatic variant detection (28). Reads were aligned to the human genome (GRCh38) using the Burrows-Wheeler Aligner (29) version 0.7.10 with default mem options and with a reading group provided for each sample for compatibility with the following steps. Duplicate reads were marked using Picard-tools version 2.6.0 MarkDuplicates. Indel realignment and base recalibration were performed with GATK version 3.3.0 to reduce false-positive variant calls. SNV and indel calls were made using GATKs build in a version of MuTect2 (30) designed to call variants, both SNVs and indels, from matched tumor and normal samples. Kallisto 0.42.1 (31) was used to determine the gene expression in transcript per million (TPM) from RNAseq data.

Neopeptide Prediction

The VCF output files from GATK's MuTect2 was given as input to the neopeptide predictor MuPeXI version 1.1 (32) together with RNAseq expression values obtained from Kallisto. HLA alleles of each patient were inferred from the WES data using OptiType version 1.0 (33) with default settings after filtering the reads aligning to the HLA region with RazerS version 3.4.0 (34). Identified mutations from TFs and TCLs were used to predict 9, 10, and 11 amino acid peptides, sorted according to the eluted ligand percentile rank (EL% Rank) score of the mutated neopeptides using netMHCpan 2.8 (35). All neopeptides with a rank score < 2 were selected for peptide synthesis, giving a total of 1,545 neopeptides across all six patients. Additionally, the tumor mutational burden of non-synonymous mutations was determined from the MuPeXI output logfile summarizing peptides originating from missense variant mutations, in-frame insertions, and deletions, together with frameshift mutations. Mutation types were determined by Ensembl's variant effect predictor as a dependency of MuPeXI. The neopeptide prediction has, prior to publication, been reanalyzed with MuPeXI 1.2.0 using netMHCpan 4.0 (36).

Peptides

All selected mutation derived and virus control peptides were purchased from Pepscan (Pepscan Presto BV, Lelystad, Netherlands) and dissolved to 10 mM in DMSO.

MHC Monomer Production and Generation of Specific Peptide–MHC Complexes

The production of MHC monomers was performed as previously described (37, 38). In brief, the heavy chains of the included HLA types and human β_2 microglobulin (β_2m) light chain were expressed in bacterial Bl21 (DE3) pLysS strain (Novagen, cat#69451) and purified as inclusion bodies. After solubilization, heavy-chain and β2m light-chain complexes were folded using a UV-sensitive ligand (39, 40), biotinylated with BirA biotin-protein ligase standard reaction kit (Avidity, 318 LLC-Aurora, Colorado), and purified using size-exclusion column (Waters, BioSuite125, 13 μm SEC 21.5 × 300 mm) HPLC (Waters 2489). Specific peptide–MHC (pMHC) complexes were generated by UV-induced peptide exchange (37, 39).

Detection of pMHC Specific T Cells by DNA Barcode-Labeled Multimers

Patient-specific libraries of predicted neopeptides and virus control peptides (size range 114–415 peptides) were generated as previously described (41). Briefly, the pMHC complexes generated above were coupled to a phycoerythrin (PE)- and DNA barcode-labeled dextran backbone. Hence, a specific peptide was given a unique DNA barcode together with a PE-fluorescent label. ccRCC patient TILs and healthy donor PBMCs were stained with an up-concentrated pool of all multimers in the presence of 50 nM dasatinib, followed by staining with a 5× antibody mix composed of CD8-BV510 (BD 563256, clone RPA-T8) or -BV480 (BD, cat#566121, clone RPA-T8), dump channel antibodies [CD4-FITC (BD, cat#345768), CD14-FITC (BD, cat#345784), CD19-FITC (BD, cat#345776), CD40-FITC (Serotech, cat#MCA1590F), and CD16-FITC (BD, cat#335035)], and a dead cell marker (LIVE/DEAD Fixable Near-IR; Invitrogen, cat#L10119). Multimer binding T cells were sorted as lymphocytes, single, live, $CD8^+$, $FITC^-$, and PE^+ and

pelleted by centrifugation. DNA barcodes were amplified from the isolated cells and from a stored aliquot of multimer pool (diluted 50,000× in the final PCR reaction, used as a baseline). PCR products were purified with a QIAquick PCR Purification kit (Qiagen, cat#28104) and sequenced at Sequetech (USA) using an Ion Torrent PGM 316 or 318 chip (Life Technologies). Sequencing data were processed by the software package Barracoda (available online at http://www.cbs.dtu.dk/services/barracoda). This tool identifies the DNA barcodes annotated for a given experiment, assigns a sample ID and pMHC specificity to each DNA barcode, and counts the total number of reads and clonally reduced reads for each pMHC-associated DNA barcode. Log_2 fold changes in read counts mapped to a given sample relative to the mean read counts mapped to triplicate baseline samples are estimated using normalization factors determined by the trimmed mean of M-values method. False discovery rates (FDRs) were estimated using the Benjamini–Hochberg method. At least 1/1,000 reads associated with a given DNA barcode relative to the total number of DNA barcode reads in that given sample was set as threshold to avoid false-positive detection of T cell responses due to low number of reads in the baseline samples. An estimated cell frequency was calculated for each DNA barcode from their read count fraction out of the percentage of $CD8^+$ $multimer^+$ T cells. DNA barcodes with a $p < 0.001$, which is equal with FDR $< 0.1\%$, and an estimated cell frequency $> 0.005\%$, were considered to be true T cell responses.

Detection of pMHC-Specific T Cells by Fluorescently Labeled pMHC Tetramers

pMHCs for which T cell responses were detected with the DNA-barcode labeled multimers were generated as fluorescently labeled pMHC tetramers in a combinatorial manner as previously described (42). Briefly, pMHC complexes were multimerized on two different streptavidin-conjugated fluorochromes to give a unique two-color combination. The following streptavidin-conjugated fluorochromes were used: PE (Biolegend, cat#405203), allophycocyanin (APC) (Biolegend, cat#405207), phycoerythrin-cyanin 7 (PE-Cy7) (Biolegend, cat#405206), PE-CF594 (BD, cat#562284), brilliant ultraviolet (BUV)737 (BD, cat#564293), brilliant violet (BV)605 (BD, cat#563260), BV650 (BD, cat#563855), BUV395 (BD, cat#564176), and BV421 (Biolegend, cat#405226). RCC patient TILs were stained with tetramers, followed by a 5× antibody mix composed of CD8-BV510 or -BV480, dump channel antibodies (CD4-FITC, CD14-FITC, CD19-FITC, CD40-FITC, and CD16-FITC), and a dead cell marker (LIVE/DEAD Fixable Near-IR). Multimer positive T cells were gated as single, live, $CD8^+$, $FITC^-$ (dump channel), multimer $color1^+$, multimer $color2^+$, and negative for the remaining colors, and defined by a minimum of 10 dual-color positive events.

Flow Cytometry

All flow cytometry experiments were carried out on LSRFortessa and FACSAria Fusion instruments (BD Biosciences). Data were analyzed in FACSDiva Software version 8.0.2 (BD Biosciences) and FlowJo version 10.4.2 (TreeStar, Inc.).

Determination of T Cell Diversity

T cell diversity was determined through the identification of CDR3 sequences from bulk RNAseq data with MiXCR version 2.1.1 (43) with the optimized setting for this specific purpose (44). The quality trimmed reads from RNAseq were used as input to MiXCR, which identify specific clones with reference to known CDR3 sequences from the ImMunoGeneTics (IMGT) database. The clone count of each clone detected refers to the reads aligning to this specific clone of the CDR3 reference library. Shannon entropy (45) was calculated as a T cell diversity measurement (46).

Self-Similarity Score

MuPeXI predicts the corresponding normal peptide for any predicted neopeptide. For a neopeptide derived from SNVs, the most similar normal peptide is identified from the unmutated amino acid sequence in the reference proteome. However, for a neopeptide derived from indels, the reference proteome is searched for the most similar peptide with up to four mismatches, referred to as the nearest normal peptide (32). The self-similarity score between a neopeptide and normal peptides was calculated using the kernel similarity measure (47). In short, this similarity is calculated from matching, at different length scales, all kmers (a substring of length k) in one peptide to the kmers in the other peptide using a Blosum similarity measure. The measure gives a value between 0 and 1 for the similarity of two peptides, where a value of 1 indicates a perfect match.

Statistical Analyses

The difference in the distribution of predicted peptides and detected responses (**Figure 1A**) was analyzed with Fisher's exact test with the Freeman–Halton extension. The data presented in **Figure 2** were assessed for normal distribution with a Shapiro–Wilk normality test with a significance level of 0.05. Data were analyzed with a non-parametric Mann–Whitney U-test or Kruskal–Wallis test with Dunn's correction for multiple comparisons. The correlations presented in **Figure 3** were analyzed using Spearman's non-parametric correlation. These statistical analyses were conducted using either GraphPad Prism 8.1.2 or R statistically software version 3.5.1.

RESULTS

Neopeptides Predicted From Two Sources

The mutational landscape of tumors from four of the six ccRCC patients was analyzed by WES and RNAseq from TFs, autologous TCLs, and TILs (germline reference). For two patients, TCLs were not established, and the analysis was done on TF solely. *In silico* extraction and prediction of neopeptides based on tumor sequencing data was performed with MuPeXI (32). The mutational burden ranged from 51 to 159 mutations in the six patients (**Table 1**). From these, neopeptides were predicted as 9-, 10-, and 11-mer peptides with predicted binding capacity to the patients' HLAs. Based on available MHC monomers produced in-house, we selected only the HLA types we could cover for the generation of the peptide (p)MHC libraries. Based on this, four to five HLA types per patient were included in the

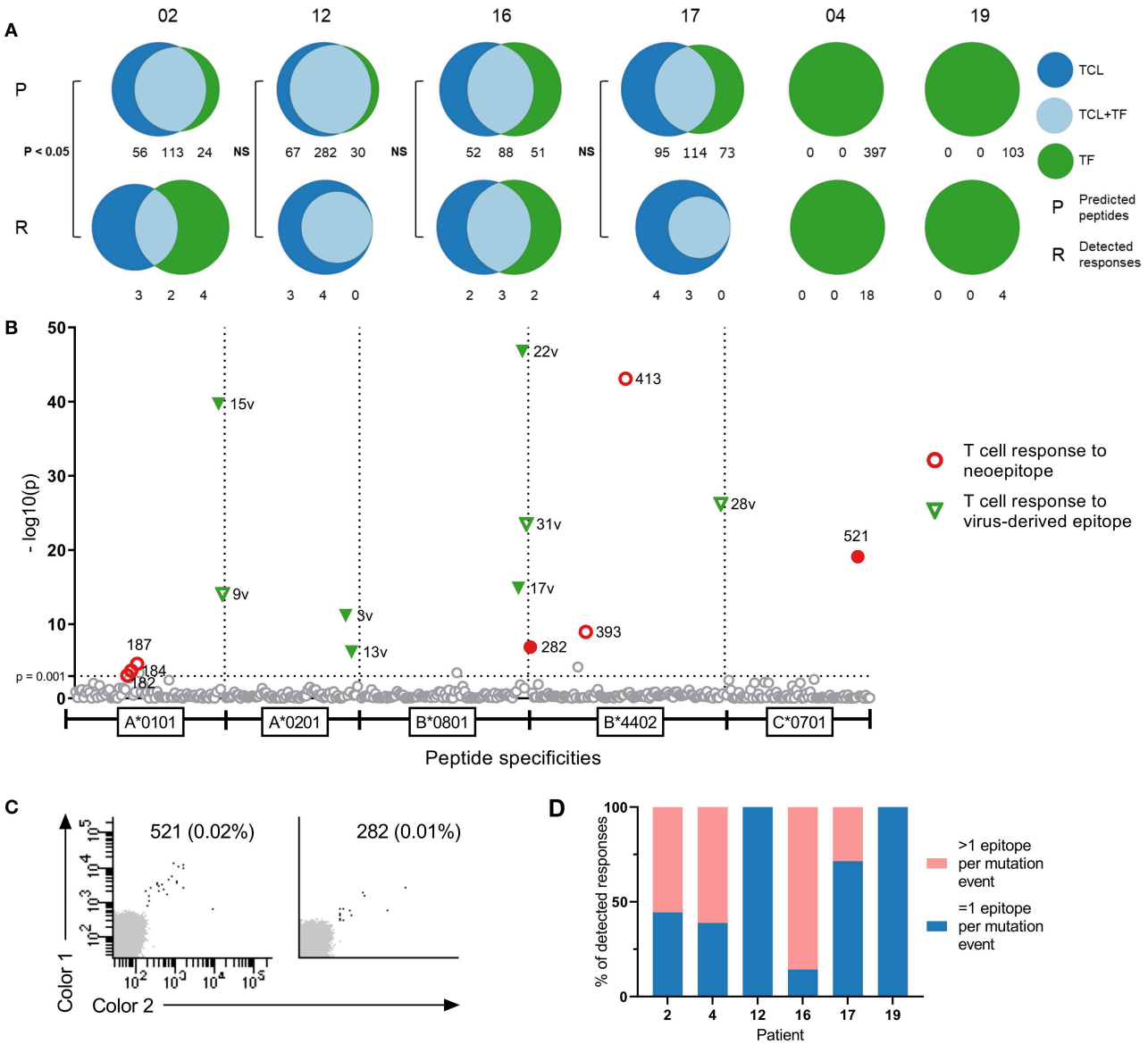


FIGURE 1 | Comparison of tumor cell lines and tumor fragments as sources for neopeptide prediction. **(A)** Distribution of predicted peptides (P) and detected responses (R) across tumor cell lines (TCL), tumor fragments (TF), and tumor cell line-tumor fragment overlap (TCL+TF) in each patient. Distribution of peptides analyzed with Fisher's exact test with Freeman-Halton extension. **(B)** T cell responses in patient RCC12 detected against neoepitopes and virus control epitopes with DNA barcode labeled multimers presented as −log10 of their significance level, distributed on HLA types. Dotted line at x = 3 [−log10(0.001)] represent the selected threshold of FDR < 0.1%. Filled labels indicate responses verified by tetramer staining. **(C)** Examples of tetramer verification plots for two of the responses detected in patient RCC12 against peptide 521 (C*0701) and peptide 282 (B*4402). **(D)** Distribution of responses, where the mutational event gave rise to more or less than 1 epitope in each patient.

neopeptide prediction. Binders were defined with a predicted rank score below 2 using NetMHCpan 2.8. On average, 258 putative neopeptides were predicted per patient, ranging from 103 to 397 (**Table 1**).

In the four patients with two tumor sources available for prediction, we conducted a comparison of the peptide origin. The mutational landscape overlapped substantially with average 50% of mutations detected in both tumor sources, and consequently, more than half of the neopeptides were predicted from both sources (57%, range 40–74%, $n = 4$). However, a proportion of the neopeptides were only predicted from one source: 17% from TF only (range, 8–27%, $n = 4$) and 26% from TCL only (range, 18–34%, $n = 4$). A similar trend is observed in the neoepitopes recognized by T cells (described in detail in the following section) with 40% of the neoepitopes being predicted from both sources (range, 22–57%, $n = 4$), whereas 20% were predicted only from TF (range, 0–44%, $n = 4$) and 40% only from TCL (range, 29–57%, $n = 4$) (**Figure 1A**). In three of

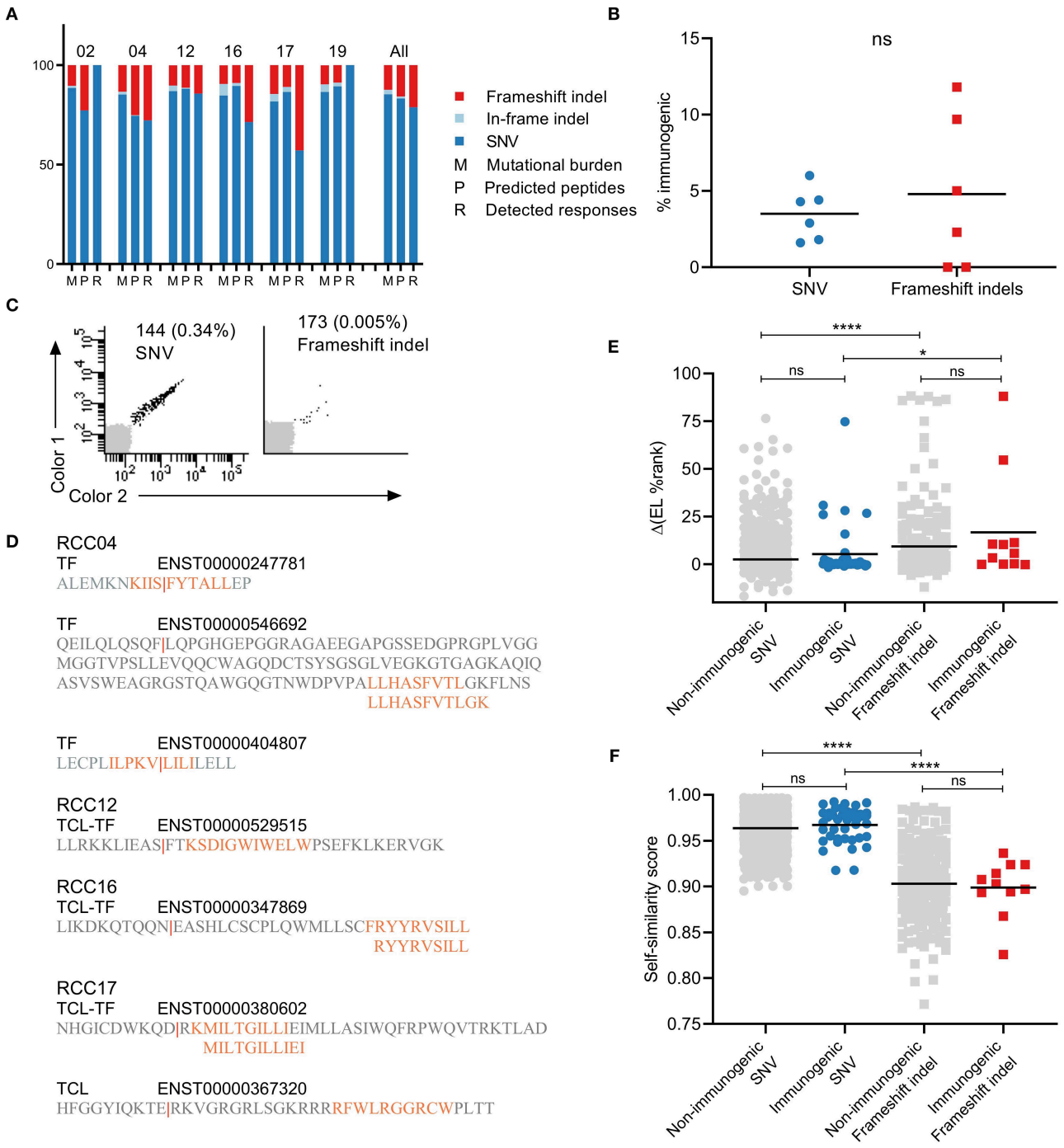


FIGURE 2 | Characterization of the contribution of different mutation types to immunogenicity. **(A)**, Distribution of frameshift indel (red), in-frame indel (light blue), and single nucleotide variation (SNV) (dark blue) mutations in each patient across tumor mutational burden (M), predicted peptides (P), and detected responses (R). **(B)** Percentages of immunogenic neoepitopes out of predicted peptides. NS difference found between mutation types (Mann-Withney U-test). **(C)** Examples of T cell responses detected against SNV mutation (left) and frameshift indel (right) derived neoepitopes. **(D)** Illustration of the frameshift mutational events giving rise to T cell responses in patients RCC04, 12, 16, and 17. **(E)** The difference in % eluted ligand (EL) rank scores between neoepitope and the corresponding wild-type. No difference between non-immunogenic and immunogenic neopeptide within the same mutation type. However, $^{****}p < 0.001$ and $^{*}p = 0.0315$ for comparison between mutation types within the same immunogenicity group (Kruskal-Wallis test with Dunn's correction). **(F)** Self-similarity score between neopeptide and the corresponding wild-type. No difference between non-immunogenic and immunogenic neopeptides within the same mutation type. However, $^{****}p < 0.001$ and $^{****}p < 0.001$ for comparison between mutation types within the same immunogenicity group (Kruskal-Wallis test with Dunn's correction).

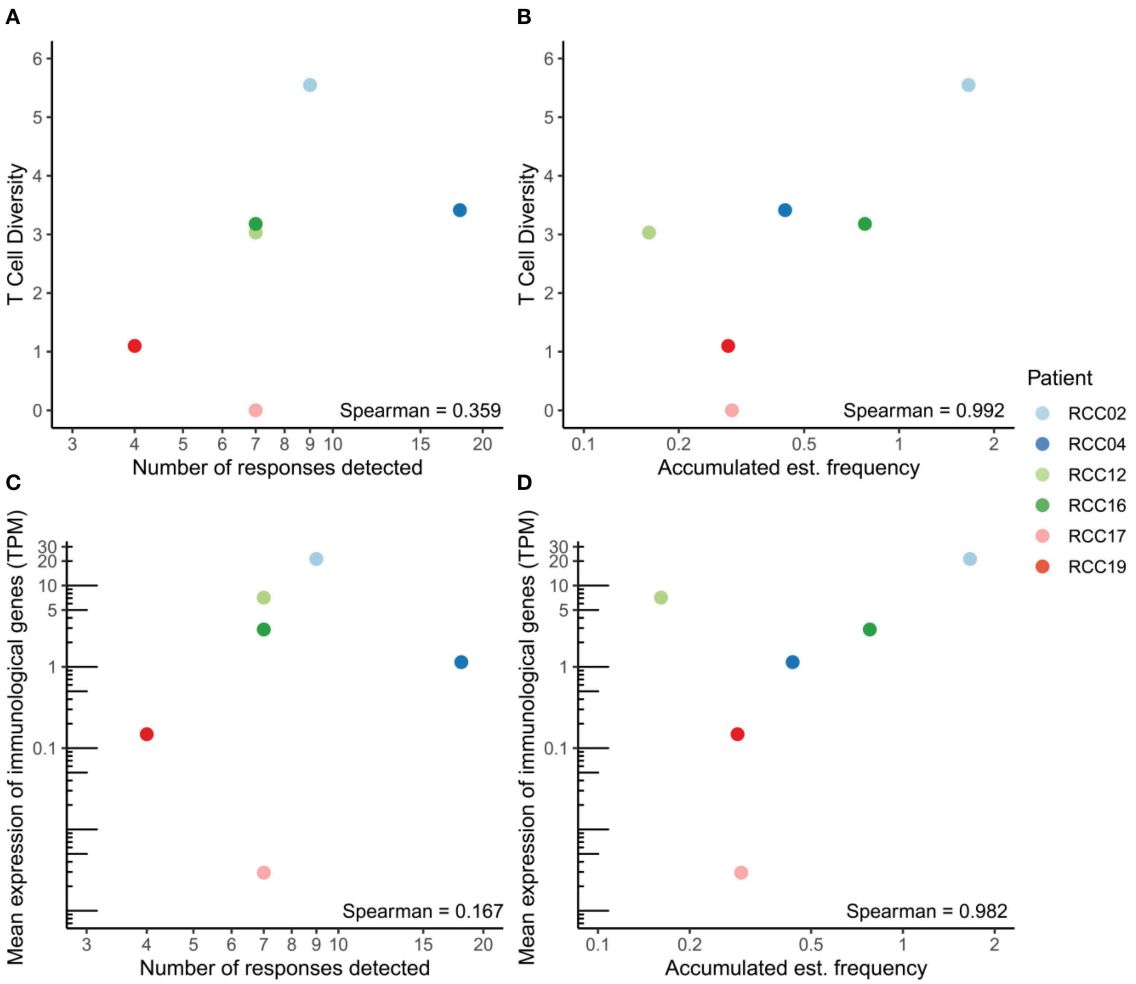


FIGURE 3 | Correlation between T cell diversity, functionality and immunogenicity. **(A,B)** Correlation between the number of detected responses **(A)** or the accumulated estimated frequency **(B)** and T cell diversity in each patient. **(C,D)** Correlation between the number of detected responses **(C)** or the accumulated estimated frequency **(D)** and the mean expression of immunological genes, measured as transcripts per million (TPM). The Spearman correlation coefficient is denoted in each plot.

the four patients, there was no difference in the distribution of peptides between predicted peptide and detected responses. However, in patient 02, the distribution was significantly different (Fisher's exact test with Freeman–Halton extension, $p < 0.05$). These results indicate the advantage of applying multiple sources of tumor material for neopeptide prediction to provide a comprehensive identification of T cell responses toward potential neopeptides.

Neoepitope-Specific CD8$^+$ T Cells Are Detected in ccRCC Patients

The 1,545 predicted neopeptides were synthesized and used to generate patient-specific libraries of DNA barcode-labeled pMHC multimers, as previously described (41). Included in each library were HLA matching epitopes derived from common viruses: influenza virus (FLU), Epstein–Barr virus (EBV), and cytomegalovirus (CMV). This resulted in patient-specific library sizes of 114 to 415 pMHC multimers that were used to stain cryopreserved TILs from the corresponding RCC patient and PBMCs from healthy donor controls. All CD8$^+$ T cells binding to a given pMHC multimer were selected and sorted based on their positive PE signal. The associated DNA barcodes were amplified and sequenced to reveal the neopeptide specificities recognized within the TIL samples. T cell responses were defined as any pMHC complex enriched in the sorted T cell fraction with a $p < 0.001$ and an estimated cell frequency above 0.005%.

T cell responses toward 54 neopeptides were detected across all patients, ranging from 4 to 18 responses per patient. **Figure 1B** shows a representation of patient 12. Results from the remaining patients are presented in **Figure S1** and with peptide information in **Table S1**. The recognized neopeptides spanned two to five HLA restrictions and covered, on average, 76% of the HLAs screened for (range, 50–100%). A number of the neopeptides were derived from the same mutational

TABLE 1 | Overview of number of mutations, predicted neopeptides, and detected T cell responses for each of the six patients.

	Mutations				Predicted neopeptides				Detected responses			
	TCL	TF	TCL/TF	Total	TCL	TF	TCL/TF	Total	TCL	TF	TCL/TF	Total
RCC02	30	13	43	86	56	24	113	193	3	4	2	9
RCC04	–	138	–	138	–	397	–	397	–	18	–	18
RCC12	28	19	97	144	67	30	282	379	3	0	4	7
RCC16	24	21	38	83	52	51	88	191	2	2	3	7
RCC17	55	39	65	159	95	73	114	282	4	0	3	7
RCC19	–	51	–	51	–	103	–	103	–	4	–	4

TCL, tumor cell line; TF, tumor fragment.

event, resulting in peptides with varying degrees of overlap in sequence. On average, 38% (range, 0–85.7%) of the T cell responses were directed toward mutations where >1 neoepitope was recognized by T cells (**Figure 1D**). Furthermore, in three of the six patients, T cell responses toward the common virus epitopes were detected (ranging from one to eight responses per patient) (**Figure 1B**, **Figure S1**). In the healthy donor cohort, we detected T cell responses toward several epitopes derived from common viruses, as expected. However, low-frequency responses toward neoepitopes were also detected.

The recognized neoepitopes were unique to each patient and none originated from known shared mutations. In a search of the COSMIC database, none of the mutations were previously described in renal cell carcinoma ($n = 6$), and in a broader search of kidney cancer [carcinoma ($n = 4512$), leiomyoblastoma ($n = 3$), renal cell carcinoma ($n = 6$), Wilms tumor ($n = 1354$), not specified ($n = 106$), and other ($n = 143$)], only two mutations were reported with a frequency above 1%: COL14A1 (1%, $n = 2168$) and PCDH11X (2.4%, $n = 2168$).

Fluorescently labeled combinatorial encoding pMHC tetramers were generated for the neoepitopes for which we observed responses with the barcode-labeling method, and these were used to validate the T cell reactivity for a number of the T cell responses observed (filled symbols, **Figure 1C** and **Figure S1**). Due to the combinatorial encoding of the tetramers, peptides with great sequence similarity (<2 amino acid difference) were not allowed in the same screen. This was, for instance, the case in patient 16 for neoepitope 144 and 173 with one amino acid difference. Tetramers were only generated for peptide 173, for which we detected T cell response toward, and we, therefore, consider peptide 144 as indirectly validated. In most cases, due to low cell numbers, the cells used for verification screens were from another TIL expanded cell product than the ones used in the original screen, whereby variation might occur—especially as many of the detected responses were of very low frequency. For patient 17 only, a CD107a sorted and expanded cell culture was available, and we screened it with the DNA barcode-labeled multimers. We detected T cell responses toward some of the same neopeptides, as in the original TIL sample (**Figure S1**).

Frameshift Indels Contribute to Immunogenicity

The tumor mutational burden of the patients included several non-synonymous mutation types: SNVs, frameshift indels, and in-frame indels (deletions and insertions) (**Figure 2A**). As expected, SNVs accounted for the largest fraction of mutations in the tumors of all six patients and resulted in a greater number of predicted neopeptides. The two other mutation types are less frequent; on average, across all patients, 12% of mutations and 16% of predicted neopeptides were frameshift indels (range, 9–15 and 9–25%), and 3% of mutations and 1% of predicted neopeptides were in-frame indels (range, 1–6 and 0–2.5%). Only neopeptides derived from SNVs (41/52, 79%) and frameshift indels (11/52, 21%) were recognized by T cells in our screen. There was no significant difference between the percentages of immunogenic neoepitopes out of predicted peptides between the two mutation groups (**Figure 2B**). However, a slightly increased average fraction was observed for frameshift indels (4.5%, range 0–12%) compared to SNVs (3.2%, range 1–6%). Validation plots of two responses toward each mutation type are presented in **Figure 2C**. The position of the original mutation that resulted in the frameshift varied between each event. In most cases, the mutation was upstream of the predicted neoepitope, and only a couple of neoepitopes were predicted at the mutation site (**Figure 2D**).

Frameshift Indels Have Increased Binding Capacity and Less Similarity to Self

The neoepitope predictor MuPeXI provides the corresponding wild-type peptide for any predicted neopeptide but through different means depending on the mutation type. For a neopeptide derived from SNVs, it is simply the unmutated amino acid sequence in the reference proteome. However, frameshift indels result in an entirely changed amino acid sequence. Instead, the reference proteome is searched for the most similar peptide with up to four mismatches, which will be defined as the nearest normal peptide to the neopeptide (32). In the following, both types will be referred to as wild-type peptides. We first investigated how both mutation types change the MHC binding capacity compared to wild-type. The prediction of neopeptides was performed with NetMHCpan 2.8. However, at the time of publication, a new version was

available (NetMHCpan 4.0) (36). Therefore, a second prediction of the current libraries was performed, and the % eluted ligand rank scores from the two versions were compared (**Figure S2A**). The outputs correlated well, with outliers representing a difference in prediction algorithms between the two versions. We continued with the prediction values from the newest version of NetMHCpan and used it to compare the binding capacity of neopeptides compared to wild-type peptide. The predicted rank scores for neopeptides were generally lower than the wild-type peptides (**Figures S2B**,**C** for individual patients). This difference was calculated as a delta(EL %Rank) value and divided into immunogenic and non-immunogenic peptides based on the T cell responses detected with the barcode-labeling method (**Figure 2E**). Within each mutation group, there was no significant difference between peptides based on their immunogenicity, even though, for both groups, slightly higher average delta values were detected for the immunogenic neoepitopes (SNVs: 2.6 and 5.4; frameshift indels: 9.4 and 16.7 for non-and immunogenic peptides, respectively). Furthermore, between the two mutation types, frameshift mutations had significantly enhanced MHC binding capacity compared to SNVs, relative to their wild-type sequence. Next, we determined the similarity between neopeptide and wild type using the kernel similarity measure giving a score between 0 and 1, where a value of 1 indicates a perfect match (47) (**Figure 2F**). This approach has previously been shown to focus on the central part of the peptide and could be an indication of similarity in T cell recognition of the presented peptide (48). As before, there is no significant difference within the same mutation group between non-immunogenic and immunogenic neopeptides. However, between the mutation types, neopeptides derived from frameshift indels are significantly less similar to wild type compared to SNV (SNVs: 0.96 and 0.97; frameshift indels: 0.9 and 0.89 for non- and immunogenic peptides, respectively).

T Cell Diversity and Functionality

We next investigated the T cell tumor infiltration and associated functional markers in the six ccRCC patients. The T cell receptor (TCR) CDR3 sequences were detected from bulk RNAseq data with MiXCR and T cell diversity was calculated using the Shannon Entropy, taking the number of reads per sample into account. Generally, few reads were detected, which is expected when extracting TCR CDR3s from RNAseq data. As a control measure, no TCRs were detected in the TCL samples, except one clone with a single read (data not shown). T cell diversity correlated with both the number of detected responses and accumulated estimated frequency from the DNA barcode screen (**Figures 3A,B**). The correlation was stronger for the accumulated estimated frequency than for the number of detected responses (Spearman correlation coefficient of 0.992 and 0.359, respectively), indicating T cell diversity as a potential surrogate marker for the number of (neo)antigen-specific T cells in the tumor. We further evaluated CD8 expression and expression of the perforin-granzyme pathway associated with CD8+ T cell activation. The mean expression of these genes correlated with both the number of detected responses and the accumulated estimated frequency from the DNA barcode screen (**Figures 3C,D**). Again, a strong positive correlation was observed for accumulated estimated frequency, whereas a weak correlation was observed for the number of detected responses (Spearman correlation coefficient of 0.982 and 0.167, respectively), demonstrating that the cell frequencies are better measurements relative to the number of recognized neopeptides.

DISCUSSION

This study details for the first time the identification and characterization of neoepitopes in renal cell carcinoma. By using a novel, high-throughput technology of DNA barcode-labeled pMHC multimers, we identified a total of 52 neoepitope-specific $CD8^+$ T cell responses in TILs from six patients with ccRCC. Renal cell carcinomas are known to harbor the highest number of insertions and deletion of all cancers (ccRCCs scoring highest of renal cell cancer subtypes), and in line with this, mutational analyses revealed the presence of frameshift and in-frame indel mutations in all six patients in the study cohort. Although we detected no responses toward in-frame indels, we observed a tendency of enrichment for T cell responses toward frameshift indel-derived neoepitopes compared to SNV-derived neoepitopes. This supports the notion that indels are a highly immunogenic subgroup of mutations, given their low self-similarity to the wild-type sequence and previous reports of enriched mutant-specific binding. We, therefore, advocate for the inclusion of indel-derived neopeptides in T cell investigational studies and neoepitope-based therapies, also in cancers with low numbers of indels. Although neoepitope prediction pipelines are undergoing intense development and optimization in these years, no consensus exists with respect to the material source for extraction of DNA and RNA for mutational mapping. Our comparison of TFs and TCLs revealed a substantial overlap in the mutational landscape identified based on the two sources (∼40% overlap), but none of the source materials performed better than the other in terms of identifying neoepitopes subjected to T cell recognition. Since a large number of epitopes were predicted from only one source or the other, it is advisable (when possible) to include both material sources as input for mutational analyses. A fraction of the variability that we observe between TFs and TCLs might be evenly present between two individual biopsies. Such tumor heterogeneity is well-documented, especially in renal cell carcinoma (49, 50). In the current study, the TFs and TCLs were generated from the same lumps of surgically removed tumors. However, they might still be influenced by tumor heterogeneity.

The neoepitopes detected in this study are all MHC class I restricted. Within recent years, growing interest have been on MHC class II neoantigens and the important role of $CD4^+$ T cells in tumor recognition and in generating a strong anti-tumor response (51, 52). Several cancer vaccines have shown to generate immune responses to class II neoepitopes either alone or in combination with class I neoepitopes (9, 53). $CD4^+$ T cells have also been suggested to be critical for tumor regression during checkpoint inhibitor therapy (54). Still, limitations in both *in silico* prediction algorithms and MHC-II multimer staining

reagents make identification of neoepitope-specific $CD4^+$ T cells challenging (55).

Although the number of RCC patients evaluated in this study is limited, the neoepitope screening presented here covers 1,545 predicted neoepitopes, derived from 572 SNV mutations and 99 frameshift/indel mutations, with ligands binding to 16 different HLA class I molecules. Thus, despite the limited number of patients analyzed, this represents a broad screening effort of class I neoepitopes from both SNVs and frameshift mutations, providing new insight into the neoepitope landscape in renal cell carcinoma patients. In line with previous studies of neoepitopes in other cancer types, all of the neoepitopes derived from mutations were unique to the given patient. Thus, therapeutic utilization in precision-targeted approaches will require patient-specific mutational mapping and prediction of neoepitopes, which can then be applied to tailor-made therapies such as personalized cancer vaccines or adoptive transfer of expanded neoepitope-specific patient TILs. The identification of virus-specific bystander T cells in the TIL products of half of the patients document the presence of therapeutically irrelevant T cells in current treatment products and further supports the rationale of developing precision-targeted therapies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Capital Region of Denmark. The Danish Data Protection Agency. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

UH and SR designed and performed experiments, data analysis, generated figures, and wrote the manuscript. A-MB designed the *in silico* prediction platform, performed data analysis, and generated figures. AB performed data analysis, generated figures, and revised the manuscript. RA provided donor material, diagnosed and characterized the patients, and generated tumor cell lines. AD, MD, and IS provided donor material and generated tumor cell lines. AKB and AM provided technical assistance, discussed data, and revised the manuscript. ZS and AE designed the *in silico* platforms. SH conceived the concept, supervised the study, discussed the data, and wrote the manuscript.

FUNDING

This research was funded through the European Research Council (ERC), StG 677268 NextDART, the Danish Research Council (DFF-4004-00422), the Danish Cancer Research Foundation, and the Neye Foundation.

ACKNOWLEDGMENTS

The authors would like to thank all cancer patients who contributed samples and all collaborators taking part in the study; Bente Rotbøl, Tripti Tamhane, Anna Gyllenberg Burkal, Anni Flarup Løye, and Julien Candrian for excellent technical assistance; and all current and former members of the SRH group for scientific discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.00373/full#supplementary-material

REFERENCES

1. Yarchoan M, Johnson BA, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment antitumour immunity. *Nat Rev Cancer.* (2017) 17:209–22. doi: 10.1038/nrc.2016.154
2. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science.* (2015) 348:69–74. doi: 10.1126/science.aaa4971
3. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med.* (2014) 371:2189–99. doi: 10.1056/NEJMoa1406498
4. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* (2015) 348:124–8. doi: 10.1126/science.aaa1348
5. Lauss M, Donia M, Harbst K, Andersen R, Mitra S, Rosengren F, et al. Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nat Commun.* (2017) 8:1738. doi: 10.1038/s41467-017-01460-0
6. Tran E, Turcotte S, Gros A, Robbins PF, Lu Y-C, Dudley ME, et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science.* (2014) 344:641–5. doi: 10.1126/science.1251102
7. Strønen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science.* (2016) 352:1337–41. doi: 10.1126/science.aaf2288
8. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature.* (2017) 547:222–6. doi: 10.1038/nature23003
9. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature.* (2017) 547:217–21. doi: 10.1038/nature22991
10. Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, et al. Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J Clin Invest.* (2015) 125:3981–91. doi: 10.1172/JCI82416
11. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science.* (2016) 351:1463–9. doi: 10.1126/science.aaf1490

12. Saini SK, Rekers N, Hadrup SR. Novel tools to assist neoepitope targeting in personalized cancer immunotherapy. *Ann Oncol.* (2017) 28:xii3–10. doi: 10.1093/annonc/mdx544
13. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* (2017) 18:1009–21. doi: 10.1016/S1470-2045(17)30516-8
14. Huang J, El-Gamil M, Dudley ME, Li YF, Rosenberg SA, Robbins PF. T cells associated with tumor regression recognize frameshifted products of the CDKN2A tumor suppressor gene locus and a mutated HLA class I gene product. *J Immunol.* (2004) 172:6057–64. doi: 10.4049/jimmunol.172.10.6057
15. Maby P, Galon J, Latouche J-B. Frameshift mutations, neoantigens and tumor-specific CD8(+) T cells in microsatellite unstable colorectal cancers. *Oncoimmunology.* (2016) 5:e1115943. doi: 10.1080/2162402X.2015.1115943
16. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature.* (2013) 500:415–21. doi: 10.1038/nature12477
17. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* (2016) 17:231. doi: 10.1186/s13059-016-1092-z
18. Gore ME, Griffin CL, Hancock B, Patel PM, Pyle L, Aitchison M, et al. Interferon alfa-2a versus combination therapy with interferon alfa-2a, interleukin-2, and fluorouracil in patients with untreated metastatic renal cell carcinoma (MRC RE04/EORTC GU 30012): an open-label randomised trial. *Lancet.* (2010) 375:641–8. doi: 10.1016/S0140-6736(09)61921-8
19. Yang JC, Childs R. Immunotherapy for Renal Cell Cancer. *J Clin Oncol.* (2006) 24:5576–83. doi: 10.1200/JCO.2006.08.3774
20. Wallin JJ, Bendell JC, Funke R, Sznol M, Korski K, Jones S, et al. Atezolizumab in combination with bevacizumab enhances antigen-specific T-cell migration in metastatic renal cell carcinoma. *Nat Commun.* (2016) 7:12624. doi: 10.1038/ncomms12624
21. Motzer RJ, Penkov K, Haanen J, Rini B, Albiges L, Campbell MT, et al. Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. *N Engl J Med.* (2019) 380:1103–15. doi: 10.1056/NEJMoa1816047
22. Matsushita H, Sato Y, Karasaki T, Nakagawa T, Kume H, Ogawa S, et al. Neoantigen load, antigen presentation machinery, and immune signatures determine prognosis in clear cell renal cell carcinoma. *Cancer Immunol Res.* (2016) 4:463–71. doi: 10.1158/2326-6066.CIR-15-0225
23. Andersen R, Westergaard MCW, Kjeldsen JW, Müller A, Pedersen NW, Hadrup SR, et al. T-cell responses in the microenvironment of primary renal cell carcinoma—implications for adoptive cell therapy. *Cancer Immunol Res.* (2018) 6:222–36. doi: 10.1158/2326-6066.CIR-17-0467
24. Donia M, Junker N, Ellebaek E, Andersen MH, Straten PT, Svane IM. Characterization and comparison of "Standard" and "Young" tumour-infiltrating lymphocytes for adoptive cell therapy at a danish translational research institution. *Scand J Immunol.* (2012) 75:157–67. doi: 10.1111/j.1365-3083.2011.02640.x
25. Babraham Bioinformatics. *Trim Galore*. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed March 15, 2016).
26. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* (2011) 17:10–2. doi: 10.14806/ej.17.1.200
27. Babraham Bioinformatics. *FastQC A Quality Control Tool for High Throughput Sequence Data*. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed March 15, 2016).
28. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics.* (2013) 43:11.10.1–33. doi: 10.1002/0471250953.bi1110s43
29. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
30. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* (2013) 31:213–9. doi: 10.1038/nbt.2514
31. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* (2016) 34:525–7. doi: 10.1038/nbt.3519
32. Bjerregaard A-M, Nielsen M, Hadrup SR, Szallasi Z, Eklund AC. MuPeXI: prediction of neo-epitopes from tumor sequencing data. *Cancer Immunol Immunother.* (2017) 66:1123–30. doi: 10.1007/s00262-017-2001-3
33. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics.* (2014) 30:3310–6. doi: 10.1093/bioinformatics/btu548
34. Weese D, Holtgrewe M, Reinert K. RazerS 3: faster, fully sensitive read mapping. *Bioinformatics.* (2012) 28:2592–9. doi: 10.1093/bioinformatics/bts505
35. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE.* (2007) 2:e796. doi: 10.1371/journal.pone.0000796
36. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol.* (2017) 199:3360–8. doi: 10.4049/jimmunol.1700893
37. Rodenko B, Toebes M, Hadrup SR, van Esch WJE, Molenaar AM, Schumacher TNM, et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. *Nat Protoc.* (2006) 1:1120–32. doi: 10.1038/nprot.2006.121
38. Hadrup SR, Toebes M, Rodenko B, Bakker AH, Egan D, Huib O, et al. High-throughput T-cell epitope discovery through MHC peptide exchange. *Methods Mol Biol.* (2009) 524:383–405. doi: 10.1007/978-1-59745-450-6_28
39. Toebes M, Coccoris M, Bins A, Rodenko B, Gomez R, Nieuwkoop NJ, et al. Design and use of conditional MHC class I ligands. *Nat Med.* (2006) 12:246–51. doi: 10.1038/nm1360
40. Bakker AH, Hoppes R, Linnemann C, Toebes M, Rodenko B, Berkers CR, et al. Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc Natl Acad Sci USA.* (2008) 105:3825–30. doi: 10.1073/pnas.0709717105
41. Bentzen AK, Marquard AM, Lyngaa R, Saini SK, Ramskov S, Donia M, et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol.* (2016) 34:1037–45. doi: 10.1038/nbt.3662
42. Andersen RS, Kvistborg P, Frøsig TM, Pedersen NW, Lyngaa R, Bakker AH, et al. Parallel detection of antigen-specific T cell responses by combinatorial encoding of MHC multimers. *Nat Protoc.* (2012) 7:891–902. doi: 10.1038/nprot.2012.037
43. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva E V, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* (2015) 12:380–1. doi: 10.1038/nmeth.3364
44. Brown SD, Hapgood G, Steidl C, Weng AP, Savage KJ, Holt RA. Defining the clonality of peripheral T cell lymphomas using RNA-seq. *Bioinformatics.* (2016) 33:1111–5. doi: 10.1093/bioinformatics/btw810
45. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* (1948) 27:379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x
46. Stewart JJ, Lee CY, Ibrahim S, Watts P, Shlomchik M, Weigert M, et al. A shannon entropy analysis of immunoglobulin and T cell receptor. *Mol Immunol.* (1997) 34:1067–82. doi: 10.1016/S0161-5890(97)00130-2
47. Shen W-J, Wong H-S, Xiao Q-W, Guo X, Smale S. Towards a mathematical foundation of immunology and amino acid chains. *arXiv [Preprint]. arXiv*:1205.6031 (2012). Available online at: http://arxiv.org/abs/1205.6031
48. Bjerregaard AM, Nielsen M, Jurtz V, Barra CM, Hadrup SR, Szallasi Z, et al. An analysis of natural T cell responses to predicted tumor neoepitopes. *Front Immunol.* (2017) 8:1566. doi: 10.3389/fimmu.2017.01566
49. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* (2012) 366:883–92. doi: 10.1056/NEJMoa1113205
50. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell.* (2017) 168:613–28. doi: 10.1016/j.cell.2017.01.018
51. Borst J, Ahrends T, Babała N, Melief CJM, Kastenmüller W. $CD4^+$ T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol.* (2018) 18:635–47. doi: 10.1038/s41577-018-0044-0

52. Linnemann C, Van Buuren MM, Bies L, Verdegaal EME, Schotte R, Calis JJA, et al. High-throughput epitope discovery reveals frequent recognition of neo-antigens by $CD4^+$ T cells in human melanoma. *Nat Med.* (2015) 21:81–5. doi: 10.1038/nm.3773
53. Kreiter S, Vormehr M, van de Roemer N, Diken M, Löwer M, Diekmann J, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature.* (2015) 520:692–6. doi: 10.1038/nature 14426
54. Alspach E, Lussier DM, Miceli AP, Kizhvatov I, DuPage M, Luoma AM, et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature.* (2019) 574:696–701. doi: 10.1038/s41586-019-1671-8
55. Schumacher TN, Scheper W, Kvistborg P. Cancer neoantigens. *Annu Rev Immunol.* (2019) 37:173–200. doi: 10.1146/annurev-immunol-042617-053402

Conflict of Interest: AE was employed by the company Clinical Microbiomics A/S.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 26 June 2020
doi: 10.3389/fimmu.2020.01215

Neoantigen-Specific Adoptive Cell Therapies for Cancer: Making T-Cell Products More Personal

Valentina Bianchi[1,2]*, *Alexandre Harari*[1,2†] *and George Coukos*[1†]

[1] *Department of Oncology, Lausanne University Hospital, Ludwig Institute for Cancer Research, University of Lausanne, Lausanne, Switzerland,* [2] *Center of Experimental Therapeutics, Department of Oncology, University Hospital of Lausanne (CHUV), Lausanne, Switzerland*

Mutation-derived neoantigens are taking central stage as a determinant in eliciting effective antitumor immune responses following adoptive T-cell therapies. These mutations are patient-specific, and their targeting calls for highly personalized pipelines. The promising clinical outcomes of tumor-infiltrating lymphocyte (TIL) therapy have spurred interest in generating T-cell infusion products that have been selectively enriched in neoantigen (or autologous tumor) reactivity. The implementation of an isolation step, prior to T-cell *in vitro* expansion and reinfusion, may provide a way to improve the overall response rates achieved to date by adoptive T-cell therapies in metastatic cancer patients. Here we provide an overview of the main technologies [i.e., peptide major histocompatibility complex (pMHC) multimers, cytokine capture, and activation markers] to enrich infiltrating or circulating T-cells in predefined neoantigen specificities (or tumor reactivity). The unique technical and regulatory challenges faced by such highly specialized and patient-specific manufacturing T-cell platforms are also discussed.

Keywords: cancer immunotherapy, adoptive cell therapy (ACT), tumor-infiltrating lymphocyte (TIL), neoantigens, enrichment

OPEN ACCESS

Edited by:
Nikolaos G. Sgourakis, University of California, Santa Cruz, United States

Reviewed by:
Maria Parkhurst, National Institutes of Health (NIH), United States
Gilles Marodon, INSERM U1135 Centre d'Immunologie et de Maladies Infectieuses, France

***Correspondence:**
Valentina Bianchi
valentina.bianchi@chuv.ch

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *31 March 2020*
Accepted: *15 May 2020*
Published: *26 June 2020*

Citation:
Bianchi V, Harari A and Coukos G (2020) Neoantigen-Specific Adoptive Cell Therapies for Cancer: Making T-Cell Products More Personal. Front. Immunol. 11:1215. doi: 10.3389/fimmu.2020.01215

INTRODUCTION

In the new age of personalized immune-oncology, tumor-infiltrating lymphocytes (TILs) generated from surgical resections, expanded *in vitro* and adoptively transferred, provide a unique opportunity to harness the specificity and diversity of the patient's endogenous T-cell repertoire. Building on the promising clinical outcomes achieved by TIL therapy in melanoma and cervical cancer (1, 2), efforts are now made to generate even more tailored T-cell products with predefined antigen specificities and, potentially, with enhanced *in vivo* tumor reactivity. The success of personalized adoptive cell therapies (ACTs) is therefore tightly linked to the identification of tumor-associated antigens, which are essential for tumor control.

Against this background, cancer neoantigens deriving from private mutations represent an ideal class of cancer antigens to target in that they are highly tumor-specific by nature, therefore reducing the potential induction of central and peripheral tolerance (3, 4). Most studies predominantly focus on single-nucleotide variants (SNVs) when referring to immunogenic tumor-specific mutant peptides; however, small insertions and deletions (indels), gene fusions, and posttranslational modifications (such as phosphorylation or glycosylation, which often alter the protein structure and function) have also been recognized as important neoantigen sources, therefore expanding the plethora of potential targets for cancer immunotherapy (5–9). Furthermore, non-canonical major histocompatibility complex (MHC) peptides derived from annotated noncoding regions are

emerging as critical immune regulators across cancer types and able to elicit tumor-specific T-cell responses (10, 11).

Neoantigen discovery is a multistep process performed on a patient-specific basis by cutting-edge preclinical pipelines integrating variant calling, *in silico* filtering, and immunogenicity evaluation, leading to private (and shared) neoantigen candidates (12–14). Briefly, mutations are called by whole-exome or whole-genome sequencing of tumor vs. germline DNA, are further filtered by *in silico* prediction algorithms and potentially tumor RNA sequencing immunopeptidomics, primarily taking into account peptide-MHC binding affinity and RNA expression as well as direct identification (15). Additional peptide features, such as stability, clonality, cleavage scores, variant allele frequency, dissimilarity to self, or mutation coverage, are now also taken into account as potential determinants of immunogenicity (16–18). The downstream number of short-listed neoepitopes varies among patients and tumor types and is further greatly reduced following cellular immunogenicity evaluation. Depending on the chosen experimental strategy, prioritized neoepitope candidates are synthesized in the form of short or long peptides, or mRNA encoding mutations, and screened for T-cell reactivities from patients' blood or tumor samples. In this context, functional assays [such as interferon (IFN)-γ ELISpot and CD137 assay] as well as peptide MHC (pMHC)-multimer direct stainings are typically used as sensitive readouts. Of note, cellular interrogation requires a significant number of patients' samples and often includes, prior to screening, a round of antigen-specific T-cell expansion with candidate neoepitope pools, which may alter the original clonotypes' composition.

Despite the variable mutational load across different human malignancies (19) and the technical challenges, tumor-infiltrating, as well as circulating, neoantigen-specific $CD8^+$ and $CD4^+$ T-cells have now been identified and characterized in several tumor types (20–25). Early clinical data also suggest that neoantigen load has a predictive role in patient response to checkpoint blockade and TIL ACT immunotherapy (26–29).

Bulk infiltrating T-cell populations can be very heterogeneous, and the frequency of private (and shared) tumor-associated antigen specificities is generally low (20, 21, 30). Dissection of melanoma and colorectal and lung cancers has highlighted that a significant fraction of TILs can contain antiviral $CD8^+$ T cells [such as Epstein-Barr virus (EBV)- and cytomegalovirus (CMV)-specific], extending observations that many tumor infiltrates may be in fact not tumor-specific (30–32). A study by Scheper et al. (33) has assessed the intrinsic tumor reactivity of TILs in melanoma and ovarian and colorectal cancer, demonstrating how indeed only a small fraction of the intratumoral $CD8^+$ T-cell receptor (TCR) repertoire is able to recognize autologous cancer cells. Yet, the frequency of $CD8^+$ in TILs correlates with favorable prognosis, and increasing evidence has shown how a relatively limited set of neoantigen-specific T-cells from melanoma TILs can mediate tumor recognition, despite the tumor cells harboring hundreds of somatic mutations (34–37). Collectively, these data suggest that enriching TIL infusion products for a few T-cell clonotypes specific for key immunogenic neoantigens could guide more effective antitumor responses *in vivo*.

One might argue that the need for available resected tumor specimens, from which infiltrating T-cells are isolated *ex vivo*, limits a broader application of standard TIL therapies to other tumor types. In this regard, Cohen et al. (22) first provided a simplified and noninvasive blood-based strategy as an alternative to current TIL production by demonstrating that neoantigen and self-antigen reactive T-cells can be reliably isolated from the peripheral blood of melanoma patients. Detection of neoantigen-specific $CD8^+$ and $CD4^+$ lymphocytes from peripheral blood has been subsequently described in patients with relatively low tumor mutation burden, such as ovarian and gastrointesinal cancers (20, 38–40). However, circulating neoantigen-specific T-cells share with their infiltrating counterpart very low detection frequencies (ranging from 0.5 to 0.002%) (20, 22, 37, 41, 42), hence the need for specific enrichment strategies. Of note, novel evidence has shown that the patient neoantigen-reactive $CD8^+$ TCR repertoire can be largely discordant (in terms of specificity and functional avidity) between circulating and infiltrating T-cells in ovarian cancer patients (20). In particular, neoantigen-specific TILs exhibited on average higher functional avidity than their peripheral blood lymphocyte (PBL) counterpart. Further studies are therefore required to assess whether PBL and TIL cultures can be an equally suitable source for successful personalized T-cell therapies. Of note, it has also been shown that non-tolerized $CD8^+$ T-cell repertoires of healthy donors were able to specifically recognize neoantigens which were ignored by tumor-infiltrating T cells in melanoma patients (43, 44).

Taken together, the selective enrichment of bulk TIL or PBL cultures for private (and shared) tumor-antigen specificities may improve the response rates achieved to date by adoptive T-cell therapies. One can envision highly personalized and specialized platforms, which integrate tumor-antigen identification and the generation of T-cell infusion products with a predefined reactivity composition (**Figure 1**). Here, we provide an overview of the current toolbox of technologies for the tailored enrichment of T-cell products in tumor-specific reactivities, addressing main advantages and disadvantages of individual approaches. A first general distinction can be made between techniques which require T-cells to be reactivated with cognate antigen (or autologous tumor) prior to downstream readout and separation (i.e., cytokine production or surface activation marker expression), and methods (such as pMHC multimer-based labeling) in which unstimulated antigen-specific T-cells can be directly selected. Different ways to restimulate antigen-specific T-cells are beyond the scope of this review and have been extensively discussed elsewhere (4, 45). A second distinction can be made between antigen-specific purification pipelines based on predefined specificities of interest and requiring a thorough antigen discovery phase and tumor reactivity-based pipelines which aim for a more "agnostic" enrichment in that they do not require *a priori* target prediction and identification

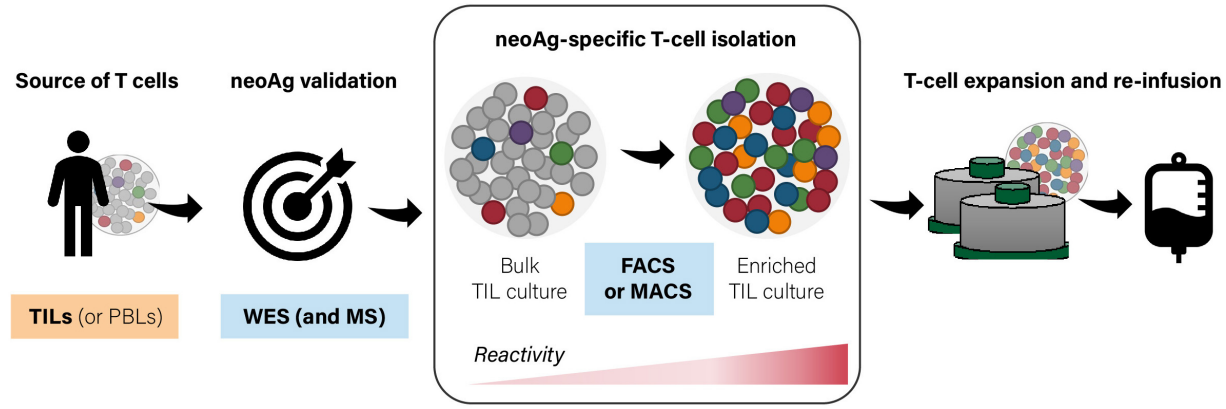


FIGURE 1 | General workflow for personalized enrichment of antigen-specific T-cells from bulk tumor-infiltrating lymphocyte (TIL) [or peripheral blood lymphocyte (PBL)] cultures. T-cells can be isolated from the patient's infiltrating or circulating lymphocyte populations. Following neoantigen discovery and validation, antigen-specific T-cells are enriched by bulk cultures and expanded *in vitro* to meet the numbers required for reinfusion. FACS, fluorescence-activated cell sorting; MACS, magnetic bead-activated cell sorting; WES, whole-exome sequencing; MS, mass spectrometry.

(**Figure 2**). The two main technologies for cell isolation are fluorescence-activated and magnetic bead-activated cell sorting (FACS and MACS, respectively), both of which are extensively employed in preclinical research environments. Finally, we will address some of the challenges and limitations that such individualized T-cell manufacturing platforms necessarily entail for clinical application from both a technical and regulatory point of view.

CURRENT TOOLSET FOR THE ENRICHMENT OF PREDEFINED NEOANTIGEN SPECIFICITIES

Peptide Major Histocompatibility Complex-Based Strategies

Labeling of a specific TCR by means of fluorochrome-conjugated pMHC multimers allows to directly identify $CD8^+$ T-cell reactivities without restriction to functional parameters. MHC-based reagents have rapidly evolved from single fluorescent-labeled pMHC tetramers to increasingly advanced and optimized staining protocols with higher detection sensitivity (46, 47). For example, the screening of multiple T-cell reactivities can be achieved by combinatorial multimer staining either assigning a unique binary color code to each antigen specificity (48, 49) or using a high number of possible fluorochrome combinations (50). Several groups have speculated a possible clinical implementation of MHC multimer-based approaches in order to screen samples and generate neoantigen-enriched therapeutic cellular products (22, 51–53).

Alternative pMHC multimeric reagents, such as Streptamers and NTAmers, are built on reversible complexes and can therefore rapidly dissociate in the presence of biotin or imidazole, respectively (54–56). Antigen-specific T-cell staining with reversible multimers not only improves conventional pMHC reagents by reducing activation-induced T-cell death but also allows pMHC monomer dissociation kinetic measurements which have been shown to correlate with T-cell functionality (55, 57, 58). These technologies could therefore further aid in the precise selection of the "fittest" T-cell clonotypes within a single antigen specificity.

A more recent addition to the pMHC multimer portfolio is represented by a different labeling whereby DNA bar code tags are attached to the multimer scaffold. T-cells are collectively sorted based on one fluorescence, and distinct pMHCs are retrospectively revealed by means of large-scale sequencing of cognate bar codes (59). As a result, up to thousands of unique specificities can be potentially identified simultaneously, paving the way for high-throughput platforms and downstream applications such as TCR redirected T-cell therapy (60, 61). Of note, state-of-the-art microfluidic devices can help with potential sample size limitation by allowing *on-chip* detection and manipulation of multimer-sorted neoantigen-specific T cells for downstream analysis and applications (38).

If on one hand, pMHC-based strategies have facilitated the characterization of complex T-cell repertoires, on the other, they present some limitations for clinical application. First of all, the MHC restriction of the antigenic peptide has to be well-characterized; they do not provide information on T-cell function and are limited for $CD4^+$ T-cell isolation (discussed below). Most importantly, given that pMHC multimer production is quite time-consuming and has to be manufactured under Good Manufacturing Practice (GMP) conditions, generation of a library of pMHC multimers on a patient-specific basis may be cost prohibitive for clinical implementation. In this regard, UV-based peptide exchange technologies (62) could aid the rapid engineering and manufacturing of multiple distinct pMHC reagents for individual patients.

Cytokine Capture

Cytokines secreted by previously activated T-cells can be retained on the cell surface via a capture matrix, allowing the molecules

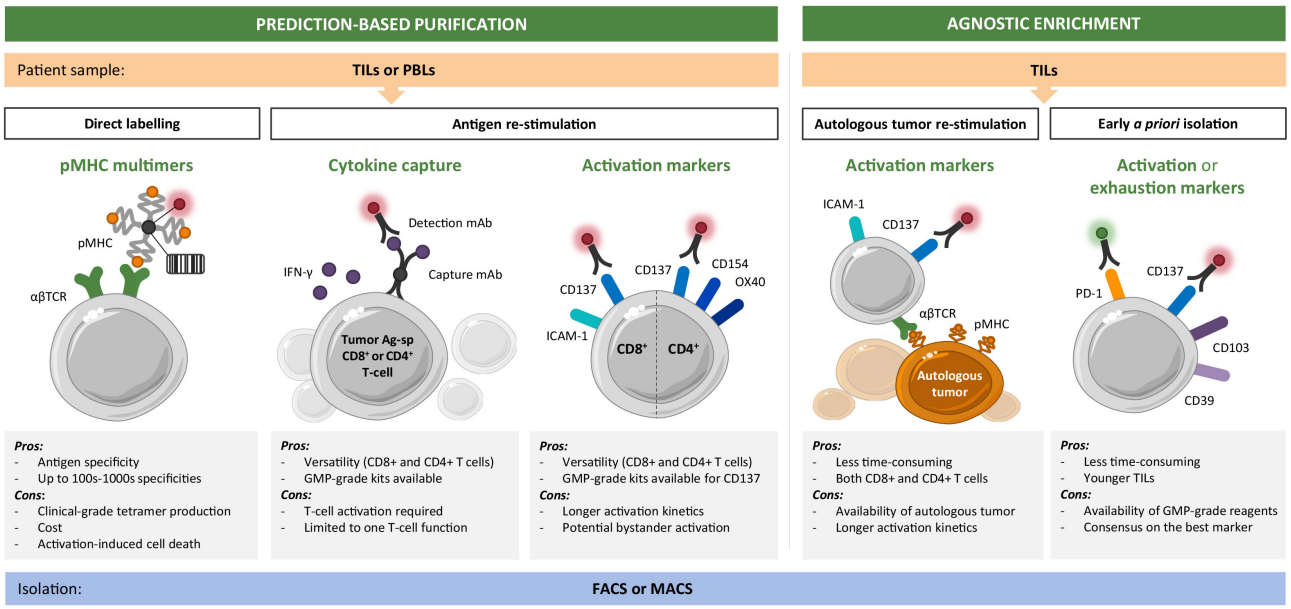


FIGURE 2 | Toolset for personalized enrichment of T-cell infusion products. Current technologies can be grouped into neoantigen-specific purification strategies, which rely on predictions and predefined epitope selection, and agnostic enrichment strategies based on coculture with autologous tumor or *a priori* identification of tumor-reactive T-cells. The main advantages and disadvantages of each approach are listed. $CD8^+$ (and $CD4^+$) T-cells of interest can be isolated starting from tumor infiltrating lymphocyte (TIL) or peripheral blood lymphocyte (PBL) cultures by fluorescence-activated (FACS) or magnetic bead-activated (MACS) cell sorting.

detection and the isolation of viable antigen-specific T-cells via MACS (63, 64). In particular, IFN-γ secretion by activated $CD8^+$ (and $CD4^+$) T-cells has long been associated with effective tumor recognition and used as a functional readout to detect tumor-reactive T-cells (65, 66). However, there are only a couple of examples of preclinical isolation of tumor-specific T-cells by means of IFN-γ capture (65). Jedema et al. (67) describe a strategy to isolate leukemia-reactive $CD8^+$ (and $CD4^+$) T-cells upon specific IFN-γ secretion to be used for adoptive transfer. Another group has reported a GMP-grade isolation of protocol of polyclonal and polyfunctional antigen-specific T-cells from healthy donor PBLs, by IFN-γ labeling followed by FACS, using NY-ESO-1 as a model system (68). Taken together, fully automated IFN-γ-based T-cell enrichment procedures are commercially available and could be more easily implemented in a clinical pipeline. However, cytokine production is known to be restricted to certain T-cell subsets; therefore, the enrichment of antigen-specific T-cell frequencies uniquely based on cytokine secretion profile might be incomplete.

Activation Markers

An alternative approach to direct labeling and cytokine detection is the use of activation-induced surface markers, which are upregulated upon antigen-specific TCR engagement. Expression of some of these markers is independent of cytokine production or T-cell phenotype, therefore potentially allowing the capture of the total pool of functional and reactive T-cells. Several surface markers have been suggested over time; however, only a limited number has been selected and extensively characterized because of reduced bystander activation, high specificity, and upregulation kinetics (69, 70).

The tumor necrosis factor receptor (TNFR) family member CD137 (or 4-1BB) has been initially characterized as a specific marker of TCR-induced activation of viral-specific $CD8^+$ T-cells (70–72). Following antigen-specific stimulation, CD137 is upregulated on $CD8^+$ T-cells, allowing the detection of viable antigen-specific T-cells. CD137 surface expression is now being extensively used as a marker to detect shared as well as neoantigen-reactive circulating and infiltrating T-cells, in combination with standard IFN-gamma ELISPOT screening (34, 42, 73–75). For example, Parkhurst et al. (73) isolated $CD137^+$ TILs by FACS following restimulation with dendritic cells (DCs) transfected with mutation-encoding RNA and showed that the expanded $CD137^+$ fraction had indeed been enriched in neoantigen-specific T-cells. At this point in time, anti-CD137 and anti-IFN-γ are among the few clinical-grade commercially available antibodies for the selection of antigen-specific T-cells; several companies though now provide custom monoclonal antibody development and conjugation under GMP guidelines.

On a different note, structural changes of activated integrins upon early TCR engagement can be exploited as the inside-out signal to detect functional T-cells (76). Dimitrov et al. (76) have successfully applied this strategy in order to monitor viral-specific T-cell responses within minutes, thereby stressing the advantageous very short incubation time compared to other activation markers. A parallel assessment of pMHC multimer and activation-based sorting would be highly informative in highlighting whether distinct markers are

able to isolate overlapping populations of heterogeneous antigen-specific T-cells.

CD4$^+$ NEOANTIGEN T-CELL RESPONSES

Screening of naturally occurring or induced neoantigen T-cell responses in patients with solid tumors has provided evidence that both CD8$^+$ and CD4$^+$ T-cells recognize private mutated epitopes (24, 77–80). Furthermore, a number of single-patient case reports seem to indicate that neoantigen-specific CD4$^+$ T-cells can mediate therapeutic immune responses to tumors (36, 81–83). A breakthrough paper by Tran et al. (36) provided the first demonstration of clinical activity of neoantigen-specific CD4$^+$ T-cell infusion in a metastatic cancer patient.

Screening for MHC class-II-restricted T-cell has been long under-appreciated because of the limited accuracy of neoantigen prediction algorithms (84, 85). However, rapidly improving prediction tools for MHC class-II ligands (86–90) and the notion that TIL cultures can include a substantial fraction of functional CD4$^+$ T-cells calls for flexible strategies to enable the enrichment of both CD4$^+$ and CD8$^+$ reactive compartments from bulk populations and downstream therapeutic infusion. In the framework of technologies validated for the CD8$^+$ counterpart, MHC class-II multimers have so far progressed at a lower rate because of technical issues with recombinant pMHC class-II heterodimer production, and the assumption that CD4$^+$ TCR binding affinity to cognate pMHC is significantly lower (91–93). On the other hand, activation marker upregulation following antigen restimulation offers the advantage of capturing cytokine-independent and heterogeneous CD4$^+$ T-cell responses (71, 94). Indeed, CD137 may allow the capture of both CD4$^+$ and CD8$^+$ functional T-cells with high specificity (69, 95). However, a few publications have described the use of alternative TCR-dependent surface markers such as CD154 and CD134 (or OX-40) to detect neoantigen-specific CD4$^+$ T-cells (34, 74, 75). A comprehensive comparison of activation-induced marker assays has yet not been investigated. In addition, care should be taken in discriminating regulatory T-cell from effector antigen-specific CD4$^+$ T-cells when exploiting activation markers. For instance, the inverse expression of CD137 and CD154 has been described to discriminate between activated regulatory and effector CD4$^+$ T-cells *ex vivo* (96).

Agnostic Enrichment of Tumor-Reactive Tumor-Infiltrating Lymphocytes

The identification and validation of patient-specific immunogenic neoantigen specificities require advanced technologies (such as high-throughput sequencing, mass spectrometry, and synthetic peptide production) and adds complexity and time (several weeks to months) to an already labor-intensive TIL production pipeline. Less time-consuming and unbiased methods are therefore being evaluated to generate patient-specific T-cell products, which are clinically feasible for adoptive transfer.

Coculture With Autologous Tumor

Using autologous tumor cells as targets circumvents the need for screening of immunogenic private or shared tumor antigens, while presenting to T-cells the complete range of naturally presented tumor antigens. In TIL production history, IFN-γ secretion has been exploited to prescreen which tumor fragments to expand: only TILs showing tumor reactivity above a predefined cutoff value were selected for downstream expansion and infused (97).

The role of activation marker CD137 was initially investigated by Ye et al. (98) for the quick and sensitive enrichment of tumor-reactive TILs from ovarian cancer and melanoma patients. The authors showed that CD137$^+$-sorted TILs demonstrated increased reactivity against shared antigens following overnight incubation in the presence of MHC-matched tumor cell lines. Importantly, the CD137-enriched fraction resulted in enhanced *in vitro* and *in vivo* antitumor reactivity (98). Seliktar-Ofir et al. (99) presented proof of concept of a GMP-compatible CD137-based separation method for personalized adoptive cell therapy. Melanoma TILs were sorted by MACS based on CD137 surface upregulation following overnight coincubation with autologous tumor cultures (99). CD137$^+$ TIL populations showed increased *in vitro* antitumor reactivity and contained a higher fraction of neoantigen and shared tumor antigen-specific T-cells when compared to the starting unseparated cultures. This approach might be limited by the establishment of autologous primary tumor cell lines, for which the success rate can be very low in tumors other than melanoma. In this regard, Dijkstra et al. (100) have presented a proof-of-concept study in which tumor-reactive T-cells from non-small-cell lung cancer and colon rectal cancer patients can be obtained by coculturing autologous PBLs with matched tumor epithelial organoids. Organoids are 3D cultures of primary solid tumors and can be established with a higher success rate from very limited amounts of tumor biopsies or surgical resections.

Considering the importance of costimulation in the context of successful tumor-specific T-cell activation, antigen presenting cells (APCs) should also be taken into account when establishing *ex vivo* cocultures of T cells and autologous tumor. The combination of natural or artificial APC with tumor lysate preparations can provide a wide array of tumor antigens in a more physiological costimulation context, therefore boosting the downstream antitumor activity of adoptively transferred T cells. Initial protocols used *ex vivo*-derived autologous DCs, stimulated with a defined maturation cocktail and pulsed with whole tumor lysate, to preferentially expand TILs to treat patients with melanoma (101–103). However, such strategies introduced additional time and numerous cytokines required for DC cell generation and maturation and prompted the quest for easily tailored and artificial APC (aAPC) platforms. Clinical-grade aAPCs have now limitless application potential: they can be coated with any number of costimulatory molecules (such as CD80, CD86, and CD137L) and membrane-bound cytokines to elicit rapid and improved TIL activation (104–106).

A Priori Enrichment of Tumor-Reactive Tumor-Infiltrating Lymphocytes

Efforts from several groups are focusing on improving the *a priori* identification of tumor-reactive TILs solely based on phenotypic profiling. The rationale behind this strategy lies in the fact that naturally occurring tumor-reactive TILs are chronically exposed to their cognate antigen in the tumor site, therefore expressing a defined set of surface activation- and/or exhaustion-associated markers, providing the opportunity for their direct isolation.

Initial evidence suggested that preselection of melanoma-infiltrating or peripheral blood T-cells by PD-1 expression prior to expansion could directly enrich tumor-reactive T-cells (39, 107). In another study, CD137 was identified as a better marker than PD-1 for the prospective selection of naturally occurring tumor-reactive fresh TILs in ovarian cancer (98). Building upon these previous works, a defined set of tissue residency markers (such as CD103, or integrin αE), necessary for recruitment and retention of TILs in the tumor site, has been suggested as a prospective marker of TIL tumor reactivity (108–110). Duhen et al. (111) have shown that co-expression of CD103 and CD39 further enriches the TIL population for tumor-reactive $CD8^+$ T-cells. $CD103^+$ $CD39^+$ TILs were sorted from tumor digests and expanded *in vitro*, resulting in increased cytotoxicity toward autologous tumor cells when compared to the respective single positive populations (111).

CHALLENGES AND LIMITATIONS TO CONSIDER FOR CLINICAL IMPLEMENTATION

Compared to traditional biological molecules, personalized adoptive T-cell platforms are developed on a patient-specific basis, therefore presenting unique challenges not only for preclinical developers and manufacturers but also for regulatory authorities and healthcare providers. Starting from the initial step of private tumor antigen discovery and validation, throughout TIL (or PBL) enrichment and *in vitro* scale-up expansion, individual processes, facilities, and technologies must be carefully reviewed and adjusted according to clinical requirements. Indeed, specific regulations may differ slightly among countries and regions, but most challenges and limitations linked to clinical implementation are shared.

Isolation Phase

Starting from the isolation step itself, one has to consider not only the technical aspects of the sorting strategy but more importantly its compatibility with regulatory requirements. The choice of a FACS or MACS-based enrichment depends on several factors, including the number of cells in the source material (TILs or PBLs), the relative frequency of antigen-specific cells within, the level of purity needed for the final product. On one hand, magnetic beads are of lower technical complexity and clinical-grade isolation kits are already commercially available, on the other, FACS separation performs multiparameter analysis of single cells, achieving resolution and purity levels, which are not always possible by MACS. In addition, FACS analysis can characterize in real time the sorted bulk T-cell population (in terms of identity and purity), as a first in-process quality control. However, FACS is still not routinely applicable under GMP conditions, which require single-use and a closed fluidic system for clinical implementation.

Expansion Phase

As the absolute cell counts of neoantigen-specific T-cells after isolation are extremely low for direct reinfusion, a rapid expansion procedure (namely, REP) of sorted cells is typically performed with allogeneic irradiated feeder cells in the presence of high-dose interleukin (IL)-2 and anti-CD3 (97). Depending on the yield, the best scale-up closed-system expansion devices and culture conditions can be optimized to meet the numbers required for the adoptive transfer (typically in the order of 10^9 cells per patient). Several distributors supply culture bags or gas-permeable flasks, sterile tubing accessories, and welding to facilitate the conversion of research protocols to GMP closed manufacturing processes, where the risk of cross contamination has to be minimized.

Absolute numbers aside, critical parameters for a successful ACT is ensuring that the final TIL product has maintained purity, TCR clonal diversity, and tumor reactivity following *in vitro* expansion. Indeed, during the REP phase, there can be interclonal competition resulting in an increased or decreased frequency of given specificities of interest compared to the starting culture. In this sense, enriching for tumor-specific T-cells at appreciable frequencies prior to up-scale would higher the chances of obtaining a final TIL product with adequate tumor reactivity upon infusion. The extensive expansion can also drive progressive T-cell differentiation and phenotype changes which may affect TIL *in vivo* persistence, homing, and proliferative capacity shortly after transfer, as these cells reencounter cognate antigens within the tumor microenvironment (112–115). Increased TIL proliferation and reactivity toward autologous tumor have been recently reported in a study introducing CTLA-4 blockade *in vitro* during the initial TIL pre-REP from ovarian tumor fragments (26). As mentioned previously, cytokines used during the *in vitro* manufacturing of the product can also significantly affect TIL immune profiles. Alternative cytokines to standard IL-2 have been tested during TIL REP (114) or during initial priming period (e.g., IL-21) (116, 117).

An additional aspect to consider is that TILs may fail to perform their expected therapeutic effector functions upon infusion due to activation-induced cell death (AICD) and exhaustion; both are peripheral tolerance mechanisms restricting an escalating, therefore potentially damaging, immune response. Melanoma TILs undergoing intense polyclonal TCR stimulation during REP have been shown to be more sensitive to AICD when cocultured *in vitro* with autologous tumor, whereas "younger" and less differentiated TILs are less susceptible and have a better *in vivo* tumor control (118–121). In a similar fashion, alternative costimulatory pathways, such as through CD137 via the addition of agonist antibodies during or following REP, can increase the polyclonal expansion of infiltrating or circulating $CD8^+$ TILs while preserving their responsiveness (122–124).

Overall, how infused antigen-specific TIL clonotypes persist *in vivo* and respond to tumor antigen restimulation upon transfer and how their gene signature correlates to clinical benefit should be studied systematically on a larger number of patients receiving the same infusion regimen.

General Improvements

Further pre-sensitization approaches could help increase neoantigen-specific T-cell frequency in starting TIL (or PBL) cultures and facilitate downstream sorting of the population of interest. Our group has reported that the addition of synthetic peptide pools of all predicted class-I neoantigens can improve conventional TIL generation in ovarian cancer (20). Primed TIL cultures were significantly enriched in neoepitope-specific $CD8^+$ T-cells as compared with standard TILs generated from the same patients.

Finally, while it is tempting to focus on private neoantigens deriving from single point mutations, driver genes (such as RAS and BRAF) recurrently affected by mutation or fusion events across individuals and cancer types would be expected to yield semiprivate (or even "shared") neoantigens. This seems to be especially the case for hematological malignancies, where immunogenic neoantigens have been reported to be mutated in up to 30% of patients (125, 126), along with case reports of mutation hot spots in solid tumors (127–130). In addition, potential semiprivate neoantigens derived from aberrant phosphorylation, resulting from dysregulated protein kinase activity during transformation, can be detected with mass spectrometry using relatively small amounts of patient samples (9, 131, 132). The targeting of shared or semiprivate neoantigens in solid tumors is a particularly desirable possibility which has to be further investigated, considering it would contribute to greatly reducing costs and production time of highly enriched T-cell infusion products. Given the multistep and laborious nature of these enriched T-cell therapies, which require coordination between highly specialized healthcare centers and manufacturing cell facilities, one has to ultimately consider if the added time frame is clinically reasonable. An additional month to the pipeline can result in significant patient dropout because of rapid disease progression (133).

CONCLUDING REMARKS

The identification of neoantigens as drivers of successful antitumor immunity is offering exciting new opportunities for cancer immunotherapies, including making T-cell infusion products highly individualized for more effective treatment. Moving forward, patient-specific T-cell enrichment technologies will need to be integrated into clinically compliant pipelines. In this respect, the first Food and Drug Administration (FDA) approval of the first adoptive cell therapy [i.e., chimeric antigen receptor (CAR) T-cell therapy] represents a huge achievement in the immune-oncology field and will hopefully pave the way for further approvals of personalized immunotherapies.

AUTHOR CONTRIBUTIONS

VB wrote the manuscript and made the figures. VB, AH, and GC contributed to the concept, discussion of content, and editing of the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Rosenberg SA, Restifo NP. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science.* (2015) 348:62–8. doi: 10.1126/science.aaa4967
2. Stevanović S, Draper LM, Langhan MM, Campbell TE, Kwong ML, Wunderlich JR, et al. Complete regression of metastatic cervical cancer after treatment with human papillomavirus-targeted tumor-infiltrating T cells. *J Clin Oncol.* (2015) 33:1543–50. doi: 10.1200/JCO.2014.58.9093
3. Linette GP, Carreno BM. Neoantigen vaccines pass the immunogenicity test. *Trends Mol Med.* (2017) 23:869–71. doi: 10.1016/j.molmed.2017.08.007
4. Bobisse S, Foukas PG, Coukos G, Harari A. Neoantigen-based cancer immunotherapy. *Ann Transl Med.* (2016) 4:262. doi: 10.21037/atm.2016.06.17
5. Wei Z, Zhou C, Zhang Z, Guan M, Zhang C, Liu Z, et al. The landscape of tumor fusion neoantigens: a pan-cancer analysis. *iScience.* (2019) 21:249. doi: 10.1016/j.isci.2019.10.028
6. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* (2017) 18:1009–21. doi: 10.1016/S1470-2045(17)30516-8
7. Raposo B, Merky P, Lundqvist C, Yamada H, Urbonaviciute V, Niaudet C, et al. T cells specific for post-translational modifications escape intrathymic tolerance induction. *Scientific Reports.* (2018) 9:353. doi: 10.1038/s41467-017-02763-y
8. Gao Q, Liang W-W, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, et al. Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* (2018) 23:227–38.e3. doi: 10.1016/j.celrep.2018.03.050
9. Solleder M, Guillaume P, Racle J, Michaux J, Pak H-S, Müller M, et al. Mass spectrometry based immunopeptidomics leads to robust predictions of phosphorylated HLA class I ligands. *Mol Cell Proteomics.* (2019) 19:390–404. doi: 10.1074/mcp.TIR119.001641
10. Müller M, Pak H, Harnett D, Huber F, Grun D, Leleu M, et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun.* (2020) 11:14968. doi: 10.1038/s41467-020-14968-9
11. Li Y, Jiang T, Zhou W, Li J, Li X, Wang Q, et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun.* (2020) 11:1000. doi: 10.1038/s41467-020-14802-2
12. Purcell AW, Ramarathinam SH, Ternette N. Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nat Protoc.* (2019) 14:1687–707. doi: 10.1038/s41596-019-0133-y
13. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature.* (2014) 515:572–6. doi: 10.1038/nature14001
14. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun.* (2016) 7:13404–16. doi: 10.1038/ncomms13404
15. Garcia-Garijo A, Fajardo CA, Gros A. Determinants for neoantigen identification. *Front Immunol.* (2019) 10:1392. doi: 10.3389/fimmu.2019.01392

16. McGranahan N, Swanton C. Neoantigen quality, not quantity. *Gene Ther.* (2019) 11:eaax7918. doi: 10.1126/scitranslmed.aax7918
17. Hundal J, Kiwala S, Feng Y-Y, Liu CJ, Govindan R, Chapman WC, et al. Accounting for proximal variants improves neoantigen prediction. *Nat Genet.* (2018) 51:175–79. doi: 10.1038/s41588-018-0283-9
18. Balachandran VP, Łuksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nat Cancer Rev.* (2017) 551:512–6. doi: 10.1038/nature24462
19. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science.* (2013) 339:1546–58. doi: 10.1126/science.1235122
20. Bobisse S, Genolet R, Roberti A, Tanyi JL, Racle J, Stevenson BJ, et al. Sensitive and frequent identification of high avidity neo-epitope specific $CD8^+$ T cells in immunotherapy-naive ovarian cancer. *Nat Commun.* (2018) 9:1092. doi: 10.1038/s41467-018-03301-0
21. Kvistborg P, Shu CJ, Heemskerk B, Fankhauser M, Thrue CA, Toebes M, et al. TIL therapy broadens the tumor-reactive $CD8^+$ T cell compartment in melanoma patients. *OncoImmunology.* (2012) 1:409–18. doi: 10.4161/onci.18851
22. Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, et al. Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J Clin Invest.* (2015) 125:3981–91. doi: 10.1172/JCI82416
23. Zhang X, Kim S, Hundal J, Herndon JM, Li S, Petti AA, et al. Breast cancer neoantigens can induce CD8 T cell responses and antitumor immunity. *Cancer Immunol Res.* (2017) 5:516–23. doi: 10.1158/2326-6066.CIR-16-0264
24. Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Cancer Rev.* (2013) 19:747–52. doi: 10.1038/nm.3161
25. Tran E, Ahmadzadeh M, Lu Y-C, Gros A, Turcotte S, Robbins PF, et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science.* (2015) 350:1387–90. doi: 10.1126/science.aad1253
26. Friese C, Harbst K, Borch TH, Westergaard MCW, Pedersen M, Kverneland A, et al. CTLA-4 blockade boosts the expansion of tumor-reactive $CD8^+$ tumor-infiltrating lymphocytes in ovarian cancer. *Sci Rep.* (2020) 10:3914. doi: 10.1038/s41598-020-60738-4
27. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol Cancer Ther.* (2017) 16:2598–608. doi: 10.1158/1535-7163.MCT-17-0386
28. Hellmann MD, Callahan MK, Awad MM, Calvo E, Ascierto PA, Atmaca A, et al. Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer. *Cancer Cell.* (2018) 33:853–61.e4. doi: 10.1016/j.ccell.2018.04.001
29. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* (2015) 348:124–8. doi: 10.1126/science.aaa1348
30. Andersen RS, Thrue CA, Junker N, Lyngaa R, Donia M, Ellebæk E, et al. Dissection of T-cell antigen specificity in human melanoma. *Cancer Res.* (2012) 72:1642–50. doi: 10.1158/0008-5472.CAN-11-2614
31. Rosato PC, Wijeyesinghe S, Stolley JM, Nelson CE, Davis RL, Manlove LS, et al. Virus-specific memory T cells populate tumors and can be repurposed for tumor immunotherapy. *Nat Commun.* (2019) 10:567. doi: 10.1038/s41467-019-08534-1
32. Simoni Y, Becht E, Fehlings M, Loh CY, Koo S-L, Teng KWW, et al. Bystander $CD8^+$ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature.* (2018) 557:575–9. doi: 10.1038/s41586-018-0130-2
33. Scheper W, Kelderman S, Fanchi LF, Linnemann C, Bendle G, de Rooij MAJ, et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat Med.* (2018) 25:89–94. doi: 10.1038/s41591-018-0266-5
34. Kalaora S, Wolf Y, Feferman T, Barnea E, Greenstein E, Reshef D, et al. Combined analysis of antigen presentation and T cell recognition reveals restricted immune responses in melanoma. *Cancer Discov.* (2018) 8:1366–75. doi: 10.1158/2159-8290.CD-17-1418
35. Lu YC, Yao X, Li YF, El-Gamil M, Dudley ME, Yang JC, et al. Mutated PPP1R3B is recognized by T cells used to treat a melanoma patient who experienced a durable complete tumor regression. *J Immunol.* (2013) 190:6034–42. doi: 10.4049/jimmunol.1202830
36. Tran E, Turcotte S, Gros A, Robbins PF, Lu Y-C, Dudley ME, et al. Cancer immunotherapy based on mutation-specific $CD4^+$ T cells in a patient with epithelial cancer. *Science.* (2014) 344:641–5. doi: 10.1126/science.1251102
37. Pritchard AL, Burel JG, Neller MA, Hayward NK, Lopez JA, Fatho M, et al. Exome sequencing to predict neoantigens in melanoma. *Cancer Immunol Res.* (2015) 3:992–8. doi: 10.1158/2326-6066.CIR-15-0088
38. Peng S, Zaretsky JM, Chour W, Bethune MT, Choi J, Hsu A, et al. Sensitive detection and analysis of neoantigen-specific T cell populations from tumors and blood. *Cell Rep.* (2019) 28:2728. doi: 10.1016/j.celrep.2019.07.106
39. Gros A, Tran E, Parkhurst MR, Ilyas S, Pasetto A, Groh EM, et al. Recognition of human gastrointestinal cancer neoantigens by circulating $PD\text{-}1^+$ lymphocytes. *Nat Med.* (2019) 129:4992. doi: 10.1172/JCI127967
40. Cafri G, Yossef R, Pasetto A, Deniger DC, Lu Y-C, Parkhurst M, et al. Memory T cells targeting oncogenic mutations detected in peripheral blood of epithelial cancer patients. *Nat Commun.* (2019) 10:449. doi: 10.1038/s41467-019-08304-z
41. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat Cancer Rev.* (2016) 22:433–8. doi: 10.1038/nm.4051
42. Martin SD, Wick DA, Nielsen JS, Little N, Holt RA, Nelson BH. A library-based screening method identifies neoantigen-reactive T cells in peripheral blood prior to relapse of ovarian cancer. *Oncoimmunology.* (2017) 7:e1371895. doi: 10.1080/2162402X.2017.1371895
43. Strønen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science.* (2016) 352:1337–41. doi: 10.1126/science.aaf2288
44. Ali M, Foldvari Z, Giannakopoulou E, Böschen M-L, Strønen E, Yang W, et al. Induction of neoantigen-reactive T cells from healthy donors. *Nat Protoc.* (2019) 14:1926–43. doi: 10.1038/s41596-019-0170-6
45. Arnaud M, Duchamp M, Bobisse S, Renaud P, Coukos G, Harari A. Biotechnologies to tackle the challenge of neoantigen identification. *Curr Opin Biotechnol.* (2020) 65:52. doi: 10.1016/j.copbio.2019.12.014
46. Dolton G, Tungatt K, Lloyd A, Bianchi V, Theaker SM, Trimby A, et al. More tricks with tetramers: a practical guide to staining T cells with peptide-MHC multimers. *Immunology.* (2015) 146:11–22. doi: 10.1111/imm.12499
47. Bentzen AK, Hadrup SR. Evolution of MHC-based technologies used for detection of antigen-responsive T cells. *Cancer Immunol Immunother.* (2017) 66:657–66. doi: 10.1007/s00262-017-1971-5
48. Hadrup SR, Bakker AH, Shu CJ, Andersen RS, Veluw J, Hombrink P, et al. Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat Methods.* (2009) 6:520–6. doi: 10.1038/nmeth.1345
49. Hadrup SR, Schumacher TN. MHC-based detection of antigen-specific $CD8^+$ T cell responses. *Cancer Immunol Immunother.* (2010) 59:1425–33. doi: 10.1007/s00262-010-0824-2
50. Newell E, Klein L, Yu W, Davis M. Simultaneous detection of many T-cell specificities using combinatorial tetramer staining. *Nat methods.* (2009) 6:497–9. doi: 10.1038/nmeth.1344
51. Kelderman S, Heemskerk B, Fanchi L, Philips D, Toebes M, Kvistborg P, et al. Antigen-specific TIL therapy for melanoma: a flexible platform for personalized cancer immunotherapy. *Eur J Immunol.* (2016) 46:1351–60. doi: 10.1002/eji.201545849
52. Tubb VM, Schrikkema DS, Croft NP, Purcell AW, Linnemann C, Freriks MR, et al. Isolation of T cell receptors targeting recurrent neoantigens in hematological malignancies. *J Immuno Therapy Cancer.* (2018) 6:70. doi: 10.1186/s40425-018-0386-y
53. Gubin MM, Artyomov MN, Mardis ER, Schreiber RD. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J Clin Invest.* (2015) 125:3413–21. doi: 10.1172/JCI80008
54. Neudorfer J, Schmidt B, Huster KM, Anderl F, Schiemann M, Holzapfel G, et al. Reversible HLA multimers (Streptamers) for the isolation of human cytotoxic T lymphocytes functionally active

against tumor- and virus-derived antigens. *J Immunol Methods.* (2007) 320:119–31. doi: 10.1016/j.jim.2007.01.001
55. Schmidt J, Guillaume P, Irving M, Baumgaertner P, Speiser D, Luescher IF. Reversible major histocompatibility complex I-peptide multimers containing $Ni2^{+}$-nitrilotriacetic acid peptides and histidine tags improve analysis and sorting of $CD8^{+}$ T cells. *J Biol Chem.* (2011) 286:41723–35. doi: 10.1074/jbc.M111.283127
56. Hebeisen M, Schmidt J, Guillaume P, Baumgaertner P, Speiser DE, Luescher I, et al. Identification of rare high-avidity, tumor-reactive $CD8^{+}$ T cells by monomeric TCR-ligand off-rates measurements on living cells. *Cancer Res.* (2015) 75:1983–91. doi: 10.1158/0008-5472.CAN-14-3516
57. Allard M, Couturaud B, Carretero-Iglesia L, Duong MN, Schmidt J, Monnot GC, et al. TCR-ligand dissociation rate is a robust and stable biomarker of $CD8^{+}$ T cell potency. *JCI Insight.* (2017) 2:e92570. doi: 10.1172/jci.insight.92570
58. Gannon PO, Wieckowski S, Baumgaertner P, Hebeisen M, Allard M, Speiser DE, et al. Quantitative TCR:pMHC dissociation rate assessment by NTAmers reveals antimelanoma T cell repertoires enriched for high functional competence. *J Immunol.* (2015) 195:356–66. doi: 10.4049/jimmunol.1403145
59. Bentzen AK, Marquard AM, Lyngaa R, Saini SK, Ramskov S, Donia M, et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol.* (2016) 34:1037–45. doi: 10.1038/nbt.3662
60. Bentzen AK, Such L, Jensen KK, Marquard AM, Jessen LE, Miller NJ, et al. T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide-MHC complexes. *Nat Biotechnol.* (2018) 36:1191–6. doi: 10.1038/nbt.4303
61. Zhang S-Q, Ma K-Y, Schonnesen AA, Zhang M, He C, Sun E, et al. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat Biotechnol.* (2018) 36:1156–9. doi: 10.1101/457069
62. Toebes M, Coccoris M, Bins A, Rodenko B, Gomez R, Nieuwkoop NJ, et al. Design and use of conditional MHC class I ligands. *Nat Med.* (2006) 12:246–51. doi: 10.1038/nm1360
63. Brosterhus H, Brings S, Leyendeckers H, Manz RA, Miltenyi S, Radbruch A, et al. Enrichment and detection of live antigen-specific $CD4^{+}$ and $CD8^{+}$ T cells based on cytokine secretion. *Eur J Immunol.* (1999) 29:4053–9. doi: 10.1002/(SICI)1521-4141(199912)29:12<4053::AID-IMMU4053>3.0.CO;2-C
64. Manz R, Assenmacher M, Pflüger E, Miltenyi S, Radbruch A. Analysis and sorting of live cells according to secreted molecules, relocated to a cell-surface affinity matrix. *Proc Natl Acad Sci U S A.* (1995) 92:1921–5. doi: 10.1073/pnas.92.6.1921
65. Becker C, Pohla H, Frankenberger B, Schüler T, Assenmacher M, Schendel DJ, et al. Adoptive tumor therapy with T lymphocytes enriched through an IFN-γ capture assay. *Nat Med.* (2001) 7:1159–62. doi: 10.1038/nm1001-1159
66. Kaplan DH, Shankaran V, Dighe AS, Stockert E, Aguet M, Old LJ, et al. Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. *Proc Natl Acad Sci U S A.* (1998) 95:7556–61. doi: 10.1073/pnas.95.13.7556
67. Jedema I, Meij P, Steeneveld E, Hoogendoorn M, Nijmeijer BA, van de Meent M, et al. Early detection and rapid isolation of leukemia-reactive donor T cells for adoptive transfer using the IFN-γ secretion assay. *Clin Cancer Res.* (2007) 13:636–43. doi: 10.1158/1078-0432.CCR-06-2093
68. Kayser S, Boβ C, Feucht J, Witte K-E, Scheu A, Bülow H-J, et al. Rapid generation of NY-ESO-1-specific $CD4^{+}$ T HELPER1 cells for adoptive T-cell therapy. *Oncoimmunology.* (2015) 4:e1002723. doi: 10.1080/2162402X.2014.1002723
69. Wehler TC, Karg M, Distler E, Konur A, Nonn M, Meyer RG, et al. Rapid identification and sorting of viable virus-reactive $CD4^{+}$ and $CD8^{+}$ T cells based on antigen-triggered CD137 expression. *J Immunol Methods.* (2008) 339:23–37. doi: 10.1016/j.jim.2008.07.017
70. Wölfl M, Kuball J, Ho WY, Nguyen H, Manley TJ, Bleakley M, et al. Activation-induced expression of CD137 permits detection, isolation, and expansion of the full repertoire of $CD8^{+}$ T cells responding to antigen without requiring knowledge of epitope specificities. *Blood.* (2007) 110:201–10. doi: 10.1182/blood-2006-11-056168
71. Wölfl M, Kuball J, Eyrich M, Schlegel PG, Greenberg PD. Use of CD137 to study the full repertoire of $CD8^{+}$ T cells without the need to know epitope specificities. *Cytometry.* (2008) 73A:1043–9. doi: 10.1002/cyto.a.20594
72. Zandvliet ML, van Liempt E, Jedema I, Kruithof S, Kester MGD, Guchelaar H-J, et al. Simultaneous isolation of $CD8^{+}$ and $CD4^{+}$ T cells specific for multiple viruses for broad antiviral immune reconstitution after allogeneic stem cell transplantation. *J Immunother.* (2011) 34:307–19. doi: 10.1097/CJI.0b013e318213cb90
73. Parkhurst MR, Gros A, Pasetto A, Prickett TD, Crystal JS, Robbins PF, et al. Isolation of T cell receptors specifically reactive with mutated tumor associated antigens from tumor infiltrating lymphocytes based on CD137 expression. *Clin Cancer Res.* (2016) 23:2491–505. doi: 10.1186/2051-1426-3-S2-P40
74. Yossef R, Tran E, Deniger DC, Gros A, Pasetto A, Parkhurst MR, et al. Enhanced detection of neoantigen-reactive T cells targeting unique and shared oncogenes for personalized cancer immunotherapy. *JCI Insight.* (2018) 3:4579. doi: 10.1172/jci.insight.122467
75. Deniger DC, Pasetto A, Robbins PF, Gartner JJ, Prickett TD, Paria BC, et al. T-cell responses to TP53 "Hotspot" mutations and unique neoantigens expressed by human ovarian cancers. *Clin Cancer Res.* (2018) 24:5562–73. doi: 10.1158/1078-0432.CCR-18-0573
76. Dimitrov S, Gouttefangeas C, Besedovsky L, Jensen ATR, Chandran PA, Rusch E, et al. Activated integrins identify functional antigen-specific $CD8^{+}$ T cells within minutes after antigen stimulation. *Proc Natl Acad Sci U S A.* (2018) 115:E5536–E45. doi: 10.1073/pnas.1720714115
77. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nat Cancer Rev.* (2017) 547:222–6. doi: 10.1038/nature23003
78. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature.* (2017) 547:217–21. doi: 10.1038/nature22991
79. van Rooij N, van Buuren MM, Philips D, Velds A, Toebes M, Heemskerk B, et al. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J Clin Oncol.* (2013) 31:e439–e42. doi: 10.1200/JCO.2012.47.7521
80. Linnemann C, van Buuren MM, Bies L, Verdegaal EME, Schotte R, Calis JJA, et al. High-throughput epitope discovery reveals frequent recognition of neo-antigens by $CD4^{+}$ T cells in human melanoma. *Nat Med.* (2015) 21:81–5. doi: 10.1038/nm.3773
81. Friedman KM, Prieto PA, Devillier LE, Gross CA, Yang JC, Wunderlich JR, et al. Tumor-specific $CD4^{+}$ melanoma tumor-infiltrating lymphocytes. *J Immunother.* (2012) 35:400–8. doi: 10.1097/CJI.0b013e31825898c5
82. Robbins PF, Li YF, El-Gamil M, Zhao Y, Wargo JA, Zheng Z, et al. Single and dual amino acid substitutions in TCR CDRs can enhance antigen-specific T cell functions. *J Immunol.* (2008) 180:6116. doi: 10.4049/jimmunol.180.9.6116
83. Kreiter S, Vormehr M, van de Roemer N, Diken M, Löwer M, Diekmann J, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature.* (2015) 520:692–96. doi: 10.1038/nature14426
84. Lundegaard C, Lund O, Nielsen M. Prediction of epitopes using neural network based methods. *J Immunol Methods.* (2011) 374:26–34. doi: 10.1016/j.jim.2010.10.011
85. Nielsen M, Lund O, Buus S, Lundegaard C. MHC class II epitope predictive algorithms. *Immunology.* (2010) 130:319–28. doi: 10.1111/j.1365-2567.2010.03268.x
86. Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat Biotechnol.* (2019) 37:1332. doi: 10.1038/s41587-019-0280-2
87. Racle J, Michaux J, Rockinger GA, Arnaud M, Bobisse S, Chong C, et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat Biotechnol.* (2019) 37:1283–6. doi: 10.1038/s41587-019-0289-6
88. Gfeller D, Bassani-Sternberg M. Predicting antigen presentation-what could we learn from a million peptides? *Front Immunol.* (2018) 9:1716. doi: 10.3389/fimmu.2018.01716
89. Schneidman-Duhovny D, Khuri N, Dong GQ, Winter MB, Shifrut E, Friedman N, et al. Predicting CD4 T-cell epitopes based on antigen

cleavage, MHCII presentation, and TCR recognition. *PLoS One.* (2018) 13:e0206654. doi: 10.1371/journal.pone.0206654
90. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics.* (2015) 67:641–50. doi: 10.1007/s00251-015-0873-y
91. Holland CJ, Dolton G, Scurr M, Ladell K, Schauenburg AJ, Miners K, et al. Enhanced detection of antigen-specific $CD4^+$ T cells using altered peptide flanking residue peptide-MHC class II multimers. *J Immunol.* (2015) 195:5827–36. doi: 10.4049/jimmunol.1402787
92. Ayyoub M, Dojcinovic D, Pignon P, Raimbaud I, Schmidt J, Luescher I, et al. Monitoring of NY-ESO-1 specific $CD4^+$ T cells using molecularly defined MHC class II/His-tag-peptide tetramers. *Proc Natl Acad Sci U S A.* (2010) 107:7437–42. doi: 10.1073/pnas.1001322107
93. Cole DK, Pumphrey NJ, Boulter JM, Sami M, Bell JI, Gostick E, et al. Human TCR-binding affinity is governed by MHC class restriction. *J Immunol.* (2007) 178:5727–34. doi: 10.4049/jimmunol.178.9.5727
94. Chattopadhyay PK, Yu J, Roederer M. A live-cell assay to detect antigen-specific $CD4^+$ T cells with diverse cytokine profiles. *Nat Med.* (2005) 11:1113–7. doi: 10.1038/nm1293
95. Danilova L, Anagnostou V, Caushi JX, Sidhom J-W, Guo H, Chan HY, et al. The mutation-associated neoantigen functional expansion of specific T cells (MANAFEST) assay: a sensitive platform for monitoring antitumor immunity. *Cancer Immunol Res.* (2018) 6:888–99. doi: 10.1158/2326-6066.CIR-18-0129
96. Nowak A, Lock D, Bacher P, Hohnstein T, Vogt K, Gottfreund J, et al. $CD137^+CD154$– expression as a regulatory T cell (Treg)-specific activation signature for identification and sorting of stable human Tregs from *in vitro* expansion cultures. *Front Immunol.* (2018) 9:3921–15. doi: 10.3389/fimmu.2018.00199
97. Dudley ME, Rosenberg SA. Adoptive-cell-transfer therapy for the treatment of patients with cancer. *Nat Rev Cancer.* (2003) 3:666–75. doi: 10.1038/nrc1167
98. Ye Q, Song D-G, Poussin M, Yamamoto T, Best A, Li C, et al. CD137 accurately identifies and enriches for naturally occurring tumor-reactive T cells in tumor. *Clin Cancer Res.* (2014) 20:44–55. doi: 10.1158/1078-0432.CCR-13-0945
99. Seliktar-Ofir S, Merhavi-Shoham E, Itzhaki O, Yunger S, Markel G, Schachter J, et al. Selection of shared and neoantigen-reactive T cells for adoptive cell therapy based on CD137 separation. *Front Immunol.* (2017) 8:4550–14. doi: 10.3389/fimmu.2017.01211
100. Dijkstra KK, Cattaneo CM, Weeber F, Chalabi M, van de Haar J, Fanchi LF, et al. Generation of tumor-reactive T cells by co-culture of peripheral blood lymphocytes and tumor organoids. *Cell.* (2018) 174:1586–98.e12. doi: 10.1016/j.cell.2018.07.009
101. Lou Y, Wang G, Lizée G, Kim GJ, Finkelstein SE, Feng C, et al. Dendritic cells strongly boost the antitumor activity of adoptively transferred T cells *in vivo*. *Cancer Res.* (2004) 64:6783–90. doi: 10.1158/0008-5472.CAN-04-1621
102. Chodon T, Comin-Anduix B, Chmielowski B, Koya RC, Wu Z, Auerbach M, et al. Adoptive transfer of MART-1 T-cell receptor transgenic lymphocytes and dendritic cell vaccination in patients with metastatic melanoma. *Clin Cancer Res.* (2014) 20:2457–65. doi: 10.1158/1078-0432.CCR-13-3017
103. Mackensen A, Meidenbauer N, S, Vogl S, Vogl R, Laumer M, et al. Phase I study of adoptive T-cell therapy using antigen-specific $CD8^+$ T cells for the treatment of patients with metastatic melanoma. *J Clin Oncol.* (2006) 24:5060–9. doi: 10.1200/JCO.2006.07.1100
104. Forget M-A, Malu S, Liu H, Toth C, Maiti S, Kale C, et al. Activation and propagation of tumor-infiltrating lymphocytes on clinical-grade designer artificial antigen-presenting cells for adoptive immunotherapy of melanoma. *J Immunother.* (2014) 37:448–60. doi: 10.1097/CJI.0000000000000056
105. Eggermont LJ, Paulis LE, Tel J, Figdor CG. Towards efficient cancer immunotherapy: advances in developing artificial antigen-presenting cells. *Trends Biotechnol.* (2014) 32:456–65. doi: 10.1016/j.tibtech.2014.06.007
106. Su Q, Igyártó BZ. One-step artificial antigen presenting cell-based vaccines induce potent effector CD8 T cell responses. *Sci Rep.* (2019) 9:18949. doi: 10.1038/s41598-019-55286-5
107. Gros A, Robbins PF, Yao X, Li YF, Turcotte S, Tran E, et al. PD-1 identifies the patient-specific $CD8^+$ tumor-reactive repertoire infiltrating human tumors. *J Clin Invest.* (2014) 124:2246–59. doi: 10.1172/JCI73639
108. Djenidi F, Adam J, Goubar A, Durgeau A, Meurice G, de Montpréville V, et al. $CD8^+CD103^+$ tumor-infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. *J Immunol.* (2015) 194:3475–86. doi: 10.4049/jimmunol.1402711
109. Ganesan A-P, Clarke J, Wood O, Garrido-Martin EM, Chee SJ, Mellows T, et al. Tissue-resident memory features are linked to the magnitude of cytotoxic T cell responses in human lung cancer. *Nat Immunol.* (2017) 18:940–50. doi: 10.1038/ni.3775
110. Webb JR, Milne K, Watson P, Deleeuw RJ, Nelson BH. Tumor-infiltrating lymphocytes expressing the tissue resident memory marker CD103 are associated with increased survival in high-grade serous ovarian cancer. *Clin Cancer Res.* (2014) 20:434–44. doi: 10.1158/1078-0432.CCR-13-1877
111. Duhen T, Duhen R, Montler R, Moses J, Moudgil T, de Miranda NF, et al. Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat Commun.* (2018) 9:56–13. doi: 10.1038/s41467-018-05072-0
112. Lu Y-C, Jia L, Zheng Z, Tran E, Robbins PF. Single-cell transcriptome analysis reveals gene signatures associated with T-cell persistence following adoptive cell therapy. *Cancer Immunol Res.* (2019) 7:1824–36. doi: 10.1158/2326-6066.CIR-19-0299
113. Radvanyi LG, Bernatchez C, Zhang M, Fox PS, Miller P, Chacon J, et al. Specific lymphocyte subsets predict response to adoptive cell therapy using expanded autologous tumor-infiltrating lymphocytes in metastatic melanoma patients. *Clin Cancer Res.* (2012) 18:6758–70. doi: 10.1158/1078-0432.CCR-12-1177
114. Li Y, Liu S, Hernandez J, Vence L, Hwu P, Radvanyi L. MART-1-specific melanoma tumor-infiltrating lymphocytes maintaining CD28 expression have improved survival and expansion capability following antigenic restimulation *in vitro*. *J Immunol.* (2010) 184:452–65. doi: 10.4049/jimmunol.0901101
115. Robbins PF, Dudley ME, Wunderlich J, El-Gamil M, Li YF, Zhou J, et al. Cutting edge: persistence of transferred lymphocyte clonotypes correlates with cancer regression in patients receiving cell transfer therapy. *J Immunol.* (2004) 173:7125–30. doi: 10.4049/jimmunol.173.12.7125
116. Chapuis AG, Lee SM, Thompson JA, Roberts IM, Margolin KA, Bhatia S, et al. Combined IL-21-primed polyclonal CTL plus CTLA4 blockade controls refractory metastatic melanoma in a patient. *J Exp Med.* (2016) 213:1133–9. doi: 10.1084/jem.20152021
117. Chapuis AG, Desmarais C, Emerson R, Schmitt TM, Shibuya KC, Lai IP, et al. Tracking the fate and origin of clinically relevant adoptively transferred CD8 $^+$T cells *in vivo*. *Sci Immunol.* (2017) 2:eaal2568. doi: 10.1126/sciimmunol.aal2568
118. Scheffel MJ, Scurti G, Simms P, Garrett-Mayer E, Mehrotra S, Nishimura MI, et al. Efficacy of adoptive T-cell therapy is improved by treatment with the antioxidant N-acetyl cysteine, which limits activation-induced T-cell death. *Cancer Res.* (2016) 76:6006–16. doi: 10.1158/0008-5472.CAN-16-0587
119. Gattinoni L. Acquisition of full effector function in vitro paradoxically impairs the in vivo antitumor efficacy of adoptively transferred $CD8^+$ T cells. *J Clin Invest.* (2005) 115:1616–26. doi: 10.1172/JCI24480
120. Tran KQ, Zhou J, Durflinger KH, Langhan MM, Shelton TE, Wunderlich JR, et al. Minimally cultured tumor-infiltrating lymphocytes display optimal characteristics for adoptive cell therapy. *J Immunother.* (2008) 31:742–51. doi: 10.1097/CJI.0b013e31818403d5
121. Hernandez-Chacon JA, Li Y, Wu RC, Bernatchez C, Wang Y, Weber JS, et al. Costimulation through the CD137/4-1BB pathway protects human melanoma tumor-infiltrating lymphocytes from activation-induced cell death and enhances antitumor effector function. *J Immunother.* (2011) 34:236–50. doi: 10.1097/CJI.0b013e318209e7ec
122. Chacon JA, Wu RC, Sukhumalchandra P, Molldrem JJ, Sarnaik A, Pilon-Thomas S, et al. Co-stimulation through 4-1BB/CD137 improves the expansion and function of $CD8^+$ melanoma tumor-infiltrating lymphocytes for adoptive T-cell therapy. *PLoS One.* (2013) 8:e60031. doi: 10.1371/journal.pone.0060031
123. Choi BK, Lee SC, Lee MJ, Kim YH, Kim Y-W, Ryu K-W, et al. 4-1BB-based isolation and expansion of $CD8^+$ T cells specific for self-tumor and

non-self-tumor antigens for adoptive T-cell therapy. *J Immunother*. (2014) 37:225–36. doi: 10.1097/CJI.0000000000000027

124. Sakellariou-Thompson D, Forget M-A, Creasy C, Bernard V, Zhao L, Kim YU, et al. 4-1BB agonist focuses $CD8^+$ tumor-infiltrating T-cell growth into a distinct repertoire capable of tumor recognition in pancreatic cancer. *Clin Cancer Res.* (2017) 23:7263–75. doi: 10.1158/1078-0432.CCR-17-0831
125. Yang W, Lee K-W, Srivastava RM, Kuo F, Krishna C, Chowell D, et al. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat Med.* (2019) 25:767–75. doi: 10.1038/s41591-019-0434-2
126. van der Lee DI, Reijmers RM, Honders MW, Hagedoorn RS, de Jong RC, Kester MG, et al. Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia. *Nat Med.* (2019) 129:774–85. doi: 10.1172/JCI97482
127. Inderberg EM, Wälchli S, Myhre MR, Trachsel S, Almåsbak H, Kvalheim G, et al. T cell therapy targeting a public neoantigen in microsatellite instable colon cancer reduces *in vivo* tumor growth. *Oncoimmunology*. (2017) 6:e1302631. doi: 10.1080/2162402X.2017.1302631
128. Chheda ZS, Kohanbash G, Okada K, Jahan N, Sidney J, Pecoraro M, et al. Novel and shared neoantigen derived from histone 3 variant H3.3K27M mutation for glioma T cell therapy. *J Exp Med.* (2018) 215:141–57. doi: 10.1084/jem.20171046
129. Wang QJ, Yu Z, Griffith K, Hanada K-I, Restifo NP, Yang JC. Identification of T-cell receptors targeting KRAS-mutated human tumors. *Cancer Immunol Res.* (2016) 4:204–14. doi: 10.1158/2326-6066.CIR-15-0188
130. Tran E, Robbins PF, Lu Y-C, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med.* (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
131. Cobbold M, La Peña De H, Norris A, Polefrone JM, Qian J, English AM, et al. MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. *Sci Transl Med.* (2013) 5:203ra125. doi: 10.1126/scitranslmed.3006061
132. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, et al. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat Protoc.* (2015) 10:1308–18. doi: 10.1038/nprot.2015.086
133. Svane IM, Verdegaal EM. Achievements and challenges of adoptive T cell therapy with tumor-infiltrating or blood-derived lymphocytes for metastatic melanoma: what is needed to achieve standard of care? *Cancer Immunol Immunother.* (2014) 63:1081–91. doi: 10.1007/s00262-014-1580-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

TECHNOLOGY AND CODE
published: 10 November 2020
doi: 10.3389/fimmu.2020.483296

ProTECT—Prediction of T-Cell Epitopes for Cancer Therapy

Arjun A. Rao[1,2,3], *Ada A. Madejska*[2,4], *Jacob Pfeil*[1,2,3], *Benedict Paten*[1,2,3], *Sofie R. Salama*[1,3,5*†] *and David Haussler*[1,2,3,5†]

[1] *Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, United States,* [2] *Computational Genomics Lab, University of California, Santa Cruz, Santa Cruz, CA, United States,* [3] *UC Santa Cruz Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA, United States,* [4] *Department of Molecular, Cell, and Developmental Biology, University of California, Santa Cruz, Santa Cruz, Santa Cruz, CA, United States,* [5] *Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, CA, United States*

Somatic mutations in cancers affecting protein coding genes can give rise to potentially therapeutic neoepitopes. These neoepitopes can guide Adoptive Cell Therapies and Peptide- and RNA-based Neoepitope Vaccines to selectively target tumor cells using autologous patient cytotoxic T-cells. Currently, researchers have to independently align their data, call somatic mutations and haplotype the patient's HLA to use existing neoepitope prediction tools. We present ProTECT, a fully automated, reproducible, scalable, and efficient end-to-end analysis pipeline to identify and rank therapeutically relevant tumor neoepitopes in terms of potential immunogenicity starting directly from raw patient sequencing data, or from pre-processed data. The ProTECT pipeline encompasses alignment, HLA haplotyping, mutation calling (single nucleotide variants, short insertions and deletions, and gene fusions), peptide:MHC binding prediction, and ranking of final candidates. We demonstrate the scalability, efficiency, and utility of ProTECT on 326 samples from the TCGA Prostate Adenocarcinoma cohort, identifying recurrent potential neoepitopes from TMPRSS2-ERG fusions, and from SNVs in SPOP. We also compare ProTECT with results from published tools. ProTECT can be run on a standalone computer, a local cluster, or on a compute cloud using a Mesos backend. ProTECT is highly scalable and can process TCGA data in under 30 min per sample (on average) when run in large batches. ProTECT is freely available at https://www.github.com/BD2KGenomics/protect.

Keywords: cancer, neoepitope, neoantigen, automated prediction, vaccine, cancer immunotherapy, adoptive cell therapy

OPEN ACCESS

Edited by:
Mustafa Diken, Johannes Gutenberg-Universität Mainz, Germany

Reviewed by:
Jonas Ibn-Salem, Johannes Gutenberg-Universität Mainz, Germany
Zoltan Vereb, University of Szeged, Hungary

***Correspondence:**
Sofie R. Salama
ssalama@ucsc.edu

†These authors share senior authorship

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: *06 July 2019*
Accepted: *13 October 2020*
Published: *10 November 2020*

Citation:
Rao AA, Madejska AA, Pfeil J, Paten B, Salama SR and Haussler D (2020) ProTECT—Prediction of T-Cell Epitopes for Cancer Therapy. Front. Immunol. 11:483296. doi: 10.3389/fimmu.2020.483296

INTRODUCTION

Tumor recognition by the adaptive immune system has been described in the literature as early as the 1980s. In 1987, Muul et al. described tumor infiltrating lymphocytes in a cohort of six melanoma samples that showed high cytotoxicity towards fresh, autologous melanoma tumor cells (1). However, at the time, T-cell responses were observed to be short lived, often lasting only a few days. Later studies showed that tumors were capable of suppressing immune responses *via* different mechanisms (2–5).

Checkpoint blockade therapy has seen a great increase in interest in the past few years with numerous drugs being approved by the FDA for clinical treatment (6–8). Prevention of PD-1:PD-L1 (9) and CTLA-4:B7.1/2 (10) binding *via* monoclonal antibodies re-enables the immune attack against the tumor; however, it can leave the patient open to development of autoimmunity or other toxicities associated with unchecked immune action (11, 12). The mutational load of a tumor (or Tumor Mutational Burden) is a good predictor of response to checkpoint therapy (13, 14). The observation that aberrations in DNA Mismatch repair genes impair tumor growth (15) suggests this effect is due to tumor "neoantigens" that act as markers for immune targeting.

Tumor infiltrating lymphocytes (TILs) from patient tumors can be activated and expanded *in-vitro* using minced autologous tumor (16). TILs can also be activated using autologous dendritic cells that are experimentally primed *in-vitro* with synthetically generated, neoepitope-bearing peptides or with RNA vaccines that contain coding transcripts for neoepitope-bearing peptides (17, 18). These cells selectively target cell-surface MHC-presented antigen produced by the tumor. Peptide vaccines attempt to produce the same result by stimulating dendritic cells *in-vivo via* synthetically produced peptides delivered subcutaneously to the patient. Experimentally primed dendritic cells and peptide vaccine therapies require prior knowledge of the mutations in the tumor in order to identify the potentially targetable sequence.

Bioinformatic analysis of tumor sequencing data can aid in the selection of neoepitopes to target in vaccine and adaptive immune system-based cancer therapies. pVAC-Seq (19) is an automated pipeline that identifies neoepitopes generated from a pre-computed, VEP-annotated (20) VCF file run with specialized plug-ins that incorporate wildtype and mutant protein sequence. Vaxrank (21) provides a ranked list of epitopes given an input mutation VCF, RNA-Seq BAMs and the patient MHC haplotype. Epidisco (22), the predecessor of Vaxrank, was capable of starting from input FASTQs. INTEGRATE-Neo (23) identifies neoepitopes from fusion genes provided in a pre-computed BEDPE file. NeoepitopePred (24) provided a workflow for epitope prediction from fusion genes and can be run through the applets on the DNAnexus cloud platform. These tools all require a user to previously align the sequencing data to a reference of choice and call variants before following the same logical paradigm of identifying mutant peptides and predicting peptide:MHC (pMHC) affinity binding [often *via* netMHC (25)]. The pipelines differ in their degree of automation, input mutation type and annotation, and presence or absence of a ranking schema. There is a clear need for a fully automated pipeline from end-to-end, beginning at the raw FASTQ files emitted by the sequencer from DNA and RNA sequencing. Recently, NeoFuse (26) was published, which automates fusion-gene-based neoepitope prediction from paired RNA-seq, but this tool does not include neoepitopes derived from single nucleotide variants (SNVs) or short insertions and deletions (INDELs).

We developed ProTECT, a fully automated tool for the Prediction of T-cell Epitopes for Cancer Therapy. We previously demonstrated the utility of ProTECT using an early version to analyze externally called SNVs in a neuroblastoma cohort (27). There we identified a potentially therapeutic neoepitope from the ALK:R1275Q hotspot mutation and proved that $CD8^+$ cytotoxic T-cells could recognize it using *in-vitro* MHC tetramer staining of peripheral blood mononuclear cells from two HLA-matched donors. The full ProTECT codebase, reported here, is completely self-contained. It accepts an input trio of sequencing data from a patient consisting of the paired tumor and normal DNA, and the tumor RNA reads in the FASTQ format and processes the data from end-to-end including alignment, *in-silico* HLA haplotyping, expression profiling, mutation calling, and neoepitope prediction.

Here we evaluate the scalability, utility, and performance of ProTECT using publicly available data. We use the 326 samples from The Cancer Genome Atlas (TCGA) Prostate Adenocarcinoma (PRAD) cohort (28) with trios of genomic data (tumor DNA, normal DNA, and tumor RNA), augmenting these data with eight previously published clinical melanoma samples (29). The TCGA PRAD cohort has an average of 21.5 exonic mutations per sample (30) and 31% of all samples are predicted to contain a fusion transcript (31), making it a good choice for detecting both SNV and fusion neoepitopes. Further, it was previously evaluated for fusion-gene-derived epitopes using INTEGRATE-Neo (23). The melanoma dataset was reported to have between 219 and 598 missense exonic mutations per sample and was previously analyzed by pVAC-Seq (19) as part of a clinical trial. We compared ProTECT's performance to the performance of these other tools.

MATERIALS AND METHODS

Procurement of Input Data

Genomic Trio (tumor DNA, normal DNA, and tumor RNA) BAM files containing sequences from 326 samples in the TCGA Prostate Adenocarcinoma (PRAD) cohort were downloaded from the Genomics Data Commons (GDC) at the National Cancer Institute using the GDC data transfer tool. The downloaded BAM files were converted back to FASTQ format as would be produced by direct sequencing using the SamToFastq module from Picard version 1.125[1]. MHCI haplotype calls using POLYSOLVER (32) for all samples were obtained externally and used for MHC haplotype prediction comparisons.

Genomic trios from three additional samples (Mel-21, Mel-38, Mel-218) were downloaded from the NCBI short read archive (SRA) (33) *via* Bioproject PRJNA278450/dbGaP accession phs001005. These patients were diagnosed with stage III resected cutaneous melanoma and had all previously received ipilimumab. Data from seven A*02:01 restricted vaccines tested for each patient were obtained from the supplementary information of the original manuscript (29).

[1] Obtained from http://broadinstitute.github.io/picard/

The input data for the INTEGRATE-Neo comparison included haplotype and fusion calls from 240 samples in the supplementary data of the INTEGRATE-Neo paper. The fusions from supplementary Excel sheet 1 were parsed into individual BEDPE format files and the epitopes from sheet 3 were extracted into individual haplotype list files with one MHC allele per line.

Indexes for the various tools were generated using the hg38 (GRCh38) reference sequence obtained from the UCSC genome browser (34). GENCODE (35) v25 was chosen as the reference annotation and was used in all relevant parts of the pipeline. Every generic hg38 index used in the analysis is available in our AWS S3 bucket 'protect-data' under the folders 'hg38_references'. These indexes can be pulled by any user to run ProTECT locally. A detailed list of commands used to create the various indexes is available in the same bucket in the 'README'.

Compute Resources Utilized

All TCGA-related analyses were conducted on a Mesos (36) cluster with one leader (12 cpus, 62 GB RAM, 500 GB Local disk) and eight identical agents (56 cpus, 250 GB RAM, 1.8 TB local disk).

The Melanoma data was analyzed on the Amazon Web Services EC2, and the data was stored securely using SSE-C encryption on S3.

326-Sample PRAD Compute

The 326 samples were run in batches of 1, 2, 5, 10, 20, or 50 samples in order to gauge the efficiency and scalability of the pipeline engine, Toil (37). Each batch size was run five times with unique samples to normalize the runtime information. The configuration file for each run was generated from a template containing all the required tool options and paths to the input reference files on the Network File System (NFS) storage server. Each batch was run once on the Mesos cluster using all nodes and an NFS-based Toil file job store to save the state of the pipeline. The five single-sample batches were also run separately without Mesos on individual nodes of the cluster using an NFS-based Toil file job store to document the time taken per sample on a single machine.

Comparison With pVAC-Seq

To compare our results with pVAC-Seq, we ran ProTECT on the input samples on AWS EC2 using an S3-based cloud job store. The input configuration for the run included paths to hg38-mapped reference files from our S3 bucket 'protect-data' and paths to the input FASTQ files in another secure bucket. The results were stored on S3 in the same bucket as the input. This analysis was conducted consistent with the mandatory cloud data use limitations on the input dataset.

Comparison With INTEGRATE-Neo

To compare our results with INTEGRATE-Neo, we parsed the data from the manuscript supplement into files acceptable by ProTECT *via* a python script. The initial input configuration file consisted of links to the fusion BEDPE format file for each of 240 samples, along with the haplotype and expression data called from the ProTECT 326 sample run. The final analysis included fusion and inferred haplotype calls for 83 samples from INTEGRATE-Neo along with ProTECT expression estimates. All ProTECT runs were conducted on the Mesos cluster.

PIPELINE SPECIFICS

ProTECT consists of eight major sections: sequence alignment, haplotyping, expression profiling, mutation calling, mutation translation, MHC:peptide binding prediction, neoepitope ranking, and reporting. **Figure 1** shows the schema for the run. Every tool used in the pipeline was hand-picked from industry-standard choices and literature reviews. Some aspects of the pipeline, notably TransGene and Rankboost, were developed in-house due to a lack of publicly available alternatives. Both tools are available as open-source repositories on github.

The entire analysis from end-to-end is built to process data against the same reference sequence and annotation. The user provides links to the properly generated indexes for each tool in the pipeline. We provide Gencode (35) version 19 annotated references for hg19 and Gencode version 25 annotated references for hg38 on our public AWS S3 bucket "protect-data"[2]. The input for a ProTECT run is a single configuration file that lists input files for each patient that will be processed and all the options and links to indexes that will be used during the run.

While ProTECT is built for end-to-end processing of sequencing trios per patient using our choice of software at each step, we understand that researchers have personal preferences for some software over others for mutation calling, gene expression estimation, *etc*. We have engineered ProTECT such that a user may run it with pre-computed SNVs, fusion calls, gene expression, and HLA haplotypes, provided they are formatted appropriately.

Sequence Alignment

DNA sequence alignment is carried out using the Burrows–Wheeler aligner (BWA) (38). The reads are aligned with BWA-mem to the provided BWA reference using default parameters. The SAM file produced upon alignment is processed to properly format the SAM header and is then converted to a coordinate-sorted BAM file with a corresponding index.

RNA sequence alignment is carried out using the ultra-fast aligner, STAR (39). The parameters for the run are optimized for fusion detection *via* STAR-fusion (40).

Alternatively, ProTECT accepts pre-aligned BAM files as an input if the MHC haplotype is provided as well. ProTECT assumes that the user has aligned the DNA and RNA using the same reference genome with the same genomic annotation.

Haplotyping

The HLA Haplotype of the patient is predicted using PHLAT (41). The haplotype is predicted using each input source of

[2]https://s3.console.aws.amazon.com/s3/buckets/protect-data/?region=us-west-2&tab=overview (Requires AWS account to access).

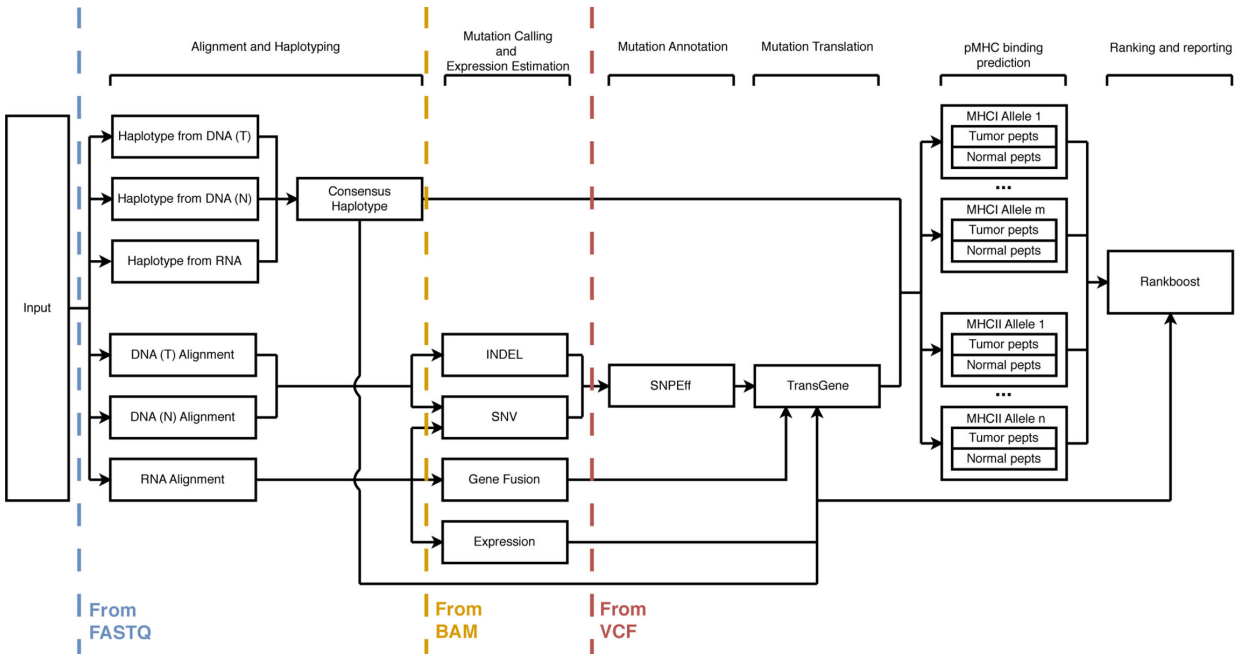


FIGURE 1 | A schematic description of the ProTECT workflow. ProTECT can process FASTQs all the way through the prediction of ImmunoActive Regions, including alignment, HLA haplotyping, variant calling, expression estimation, mutation translation, and pMHC binding affinity prediction. ProTECT also allows users to provide pre-computed inputs for various steps instead.

information (normal and tumor DNA, tumor RNA), and the consensus haplotype is generated based on agreement between any two of the three haplotype predictions. Due to limitations in the tool, we only proceed with HLA-A, HLA-B, and HLA-C for MHCI, and HLA-DPA/B and HLA-DRB for MHCII.

Expression Profiling

The gene-level and isoform-level expression is estimated using RSEM (42) with default parameters.

Mutation Calling

SNVs are predicted on a per-chromosome basis using five separate mutation prediction algorithms: MuTECT (43), MuSE (44), RADIA (45), Somatic Sniper (46), and Strelka (47). The choice of mutation callers was guided by the results from the ICGC DREAM mutation calling challenge (48). All called mutations are merged into a common file, and only events supported by two or more predictors advance to the translation step. Strelka additionally produces a callout for short insertions and deletions (INDELs). These are also used to identify neoepitopes.

Fusion calling occurs using STAR-Fusion (40) with default parameters. Candidate fusions are annotated using Fusion-Inspector[3] along with an optional assembly step using Trinity (49).

Mutation Translation

SNVs and INDELs are annotated using SNPEff (50). Mutations identified in coding regions of the genome are processed using an in-house translation tool, TransGene[4]. TransGene filters the input SNPEff-produced VCF file to exclude non-expressed calls based on the gene expression data obtained in the previous step. SNVs and in-frame INDELs are directly injected into the amino acid chain to produce the mutant sequence. Frameshift INDELs are translated downstream of the mutation event till a stop codon is found (or a user-defined threshold is reached). Events lying within 27, 30, and 45 bp of each other (for 9-mer-, 10-mer-, and 15-mer-containing peptides respectively) are chained together into an "immunoactive region" (IAR), or a region that will potentially produce an immunogenic peptide. Separate mutation events that are combined into a single immunoactive region are phased using the RNA-Seq data to ensure that they truly are co-expressed on the same haplotype.

Fusion IARs are generated using the breakpoints provided in an input BEDPE file. TransGene uses provided junction sequences or infers them from the input annotation file. The predicted IAR contains (n − 1)*3 bp on either side of the fusion junction from each donor for each n in 9-, 10-, and 15-mer. Fusion calls are optionally filtered at this stage to remove events arising from two mitochondrial genes or two immunoglobulin genes since these are usually false positive events arising from sequence similarity. Fusions can also be filtered for being

[3]Obtained from https://github.com/FusionInspector

[4]Hosted at https://github.com/arkal/transgene

potential transcriptional readthroughs (by default, two genes on the same chromosome within 500 kb of each other are rejected) or for having a 5′ lincRNA (under the assumption that these events are unlikely to be translated).

MHC:Peptide Binding Prediction

The predicted neoepitopes are assayed against each of the MHCI (9- and 10-mers) and MHCII (15-mers) predicted to be in the patient's HLA haplotype using the IEDB MHCI and MHCII binding predictions tools.

The IEDB tools run a panel of methods (51–57) on each input query (input peptide FASTQ + MHC allele) and provide a consensus "percentile rank" that describes on average, how well each peptide is predicted to bind against a background set of 100,000 UniPROT derived peptides. Calls predicted to bind within the top 5% of all binders are selected for further study. The normal, unmutated ("wildtype") counterpart peptide for each selected neoepitope is then also assayed against the MHC(s) identified to determine how well it binds, so that this can be compared to the binding affinity of the mutant version.

Neo-Epitope Ranking

Neoepitope: MHC calls are consolidated by the candidate IAR of origin. An in-house method, Rankboost[5], first arranges the IARs in descending order based on the best binding score of a contained neoepitope and then uses the boosting strategy described in Algorithm 1 to produce a final list of ranked IARs. Candidates satisfying certain biologically relevant criteria are boosted in rank based on user-specified weights. The features considered are the total number of calls originating from the IAR (npa) and ones with high predicted binding score (nph, percentile rank <=1.0), the promiscuity of the region (nmhc, *i.e.* the number of MHCs stimulated by peptides from the region), the combined expression of the isoforms displaying a neoepitope-generating mutation (TPM), the number of neoepitopes in the region predicted to bind to an MHC better than their wildtype counterpart, and the number of events where a 10-mer and 9-mer subsequence of it both bind well to an MHC (ovlp, this is only done for MHCI). Each candidate is assigned a score from 0-1 for each feature that is multiplied by a user-specified weight. The sum of the weighted score provides the boost received by the candidate. Feature score functions were generated based on empirical distributions of the features seen in IARS predicted in other TCGA and internal datasets. The algorithm iterates over the table of candidates three times and performs per-candidate boosting, resulting in a ranked list of epitopes in the sample. We ran our samples prioritizing overlap and promiscuity (0.68 and 0.32 respectively) for MHCI calls and set each covariate to 0.2 (equally important) for MHCII calls.

Algorithm 1. Pseudocode for the rank boosting strategy. W_x describes the weight for covariate x, boost_x describes the score for the candidate x from 0 to 1, npa = number of peptides constituting an IAR, nph = number of strongly binding peptides constituting the IAR, nMHC = number of MHCs predicted to recognize a neoepitope from this IAR, TPM = expression of the transcript harboring the IAR, and ovlp = number of events where a 9-mer and 10-mer overlap and are predicted to bind to the same MHC (only valid for MHCI).

```
For i in 1, 2, 3
  ∀ candidate in candidates
    boost = W_npa * boost_npa + \
            W_nph * boost_nph + \
            W_nMHC * boost_nMHC + \
            W_TPM * boost_TPM + \
            W_ovlp * boost_ovlp
    new_rank = old_rank * (1-boost)
```

[5]Hosted at https://github.com/arkal/rankboost

RESULTS AND DISCUSSION

We ran three experiments to demonstrate our pipeline. The first experiment was run on 326 samples from the TCGA PRAD cohort and highlights the scalability, efficiency, and utility of ProTECT. We also identify recurrent IARs in the cohort (containing mutations that occurred in more than one case) suggesting possible public neoepitopes for PRAD. The second experiment compares ProTECT to the published SNV- and INDEL-based neoepitope prediction pipeline, pVAC-Seq. The third experiment compares ProTECT to the published fusion-based neoepitope predictor, INTEGRATE-Neo. In all experiments, ProTECT was run using a consensus of two out of five mutation callers (as described above) and using all TransGene fusion filters to remove inter-mitochondrial, inter-immunoglobulin, 5′ lincRNA, and transcriptomic readthrough events. Results were tabulated using a mix of python scripts and manual curation on a local machine.

326 Sample Run

To describe the scalability, utility, and efficiency of ProTECT, we ran ProTECT on a total of 326 genomic trios from the TCGA PRAD cohort. We called a median of 79.5 SNVs and INDELs, and seven fusion genes per sample, and accepted 20 and three respectively for the production of IARs. We identified a median of 11 IARs per sample. Of the 326 samples, only three samples were predicted to have no IARs. These samples were observed to have no expressed non-synonymous mutations or filter-passing fusions. The entire metrics table is presented in **Supplementary Table 1**, and the results are submitted as **Supplementary File 1**.

Figure 2 shows the results from running ProTECT with different batch sizes on our local cluster (see section *Compute Resources Utilized*). As the number of samples increases, we see an expected increase in overall time, but the average time per sample decreases drastically because our pipeline engine maximizes resource utilization. We processed samples from end-to-end at a rate of 24.6 min per sample (calculated as total time divided by total number of samples) when run in a batch of 50 samples.

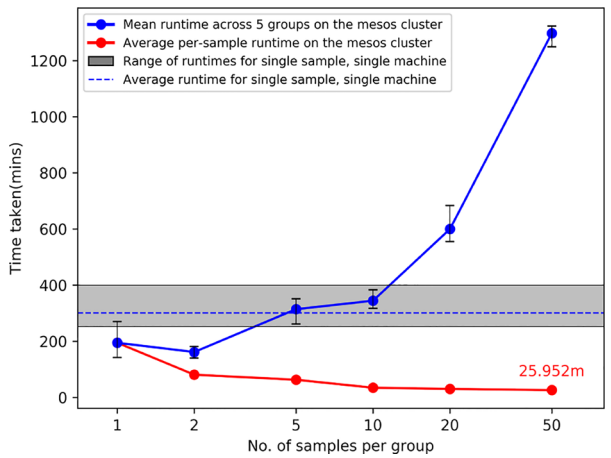


FIGURE 2 | Average runtimes on our cluster when ProTECT is run in a batch of 'n' samples. Each batch of size 'n' is run with five unique sample sets, and the range of runtimes is described by the whiskers at each datapoint. The gray bar describes the result of running ProTECT on a single sample on one machine. ProTECT takes considerably less time on average when run in a large group.

Recurrent Fusion-Gene-Derived IARs in PRAD

We detected the well-documented TMPRSS2-ERG fusion gene (58–60) in 131 samples. We predicted at least one IAR each arising from five of the 10 unique breakpoints called (**Table 1**). Of the five breakpoints that do not result in an IAR, four of these breakpoints are located in the 5′ UTR of TMPRSS2 and will not result in a neoepitope. The 5th breakpoint has a 5′ intronic breakpoint and a 3′ exonic one, and the resulting neoepitope should contain the translated product from the last few bases of TMPRRS2 Exon 1 and the first bases after the *de novo* splice acceptor is reached in ERG. This case is not handled by TransGene at this time, and so no neoepitope call was made. One IAR of particular interest is DNSKMALNSEALSVVSED from the junction chr21:41498119–chr21:38445621, which is found in 37 of the 48 unique samples harboring that junction (11% of the entire cohort). Peptides from this IAR are predicted to bind well to HLA-A*02:01 (Allele Freq: 0.26) and HLA-C*07:01 (allele Freq: 0.17), alleles frequently seen in Caucasian populations, which are highly represented in the TCGA cohort. Similarly, we predict SGCEERGAAGSLISCE from 22/35 samples with chr21:41507950–chr21:38445621, binding to C*07:01, C*04:01, B*44:02 (allele frequencies 0.14, 0.12, 0.08 respectively). The distributions of MHC alleles detected in patients harboring these events are shown in **Supplementary Figures 1** and **2**, respectively. These events are potentially viable candidates for public epitopes for patients with TMPRSS2-ERG and could be pursued as vaccines for these cancers.

Recurrent SNV-Derived IARs in PRAD

We detected a number of recurrent mutations in the SPOP gene concordant with previous reports (28, 61, 62). We detected seven unique recurrent variants across 19 samples that map to three different amino acid positions in the SPOP protein, p.F133C/V/I/L, p.F102C/V, and p.W131G (**Table 2**). The mutation at position 133 might be of immunological interest since Leucine, Isoleucine, and Valine have small hydrophobic side-chains and may stimulate the same TCR depending on pMHC binding. This hypothesis however, would require biological validation. Samples with SPOP mutations lack ETV family fusions, suggesting that vaccine therapies against SPOP and the TMPRSS2-ERG fusion would target different populations of PRAD patients.

Comparison of HLA Haplotypes Between PHLAT and POLYSOLVER

An important topic to highlight is HLA haplotypes called by PHLAT (41). We compared our results to the POLYSOLVER (32) calls, and consistent with prior work (63), we see that PHLAT miscalled HLA-A*02:01 as HLA-A*01:81 in 33 samples. However, 29 of these samples are predicted to be homozygous HLA-A*02:01 by POLYSOLVER so the effect of this miscall will be to add information to the final ranked IARS from one additional allele. Since most IARs contain peptides predicted to bind to more than one allele, the noise produced by this artifact should not adversely affect the scores generated *via* the signal from calls against the correct partners. The remaining four samples were predicted to be heterozygous HLA-A*02:01/HLA-A*01:01 *via* POLYSOLVER, and ProTECT identified these samples as HLA-*02:01/HLA-A*01:81. This is slightly worse

TABLE 1 | Recurrent TMPRSS2-ERG breakpoints in the cohort.

Breakpoint	Count	5' breakpoint	3' breakpoint	Neoepitope Expected?	IAR	Count
21:41508081–21:38445621	122	5′ UTR	Exon 2	No	NA	
21:41498119–21:38445621	48	Exon 2	Exon 2	Yes	DNSKMALNS EALSVVSED	37
21:41507950–21:38445621	35	Exon 1	Exon 2	Yes***	SGCEERGAA GSLISCE	22
21:41508081–21:38474121	18	5′ UTR	Intron 1	No	NA	
21:41506445–21:38445621	18	Intron 1	Exon 2	Yes*	NA	
21:41508081–21:38584945	11	5′ UTR	5'UTR	No	NA	
21:41498119–21:38474121	7	Exon 2	Intron 1	Yes**	DNSKMALNS LNSIDDAQL	7
21:41508081–21:38423561	7	5′ UTR	Exon 3	No	NA	
21:41498119–21:38423561	4	Exon 2	Exon 3	Yes***	DNSKMALNS ELS	1
21:41494356–21:38445621	3	Exon 3	Exon 2	Yes***	SPSGTVCTS RSLISCE	3

*IARs from 21:41498119 to 21:38445621 and 21:41507950 to 21:38445621 are recurrent suggesting the viability of universal peptide vaccine candidates. We do not expect to see an IAR from fusions with 5′UTR breakpoints. *TransGene cannot handle de novo splice acceptors. **An epitope will exist where the TMPRSS2 reads into the intron of ERG. ***A frameshift is seen on the ERG side of the fusion.*

TABLE 2 | Recurrent mutants in the SPOP gene target three codons.

Variant	Count	Gene	Mutant	IAR	Frequency
chr17:49619064A>C	5	SPOP	p.F133V	RFVQGKDWG **V** KKFIRRDFL	4
chr17:49619063A>C	3	SPOP	p.F133C	RFVQGKDWG **C** KKFIRRDFL	2
chr17:49619064A>T	2	SPOP	p.F133I	RFVQGKDWG **I** KKFIRRDFL	1
chr17:49619062G>T	2	SPOP	p.F133L	RFVQGKDWG **L** KKFIRRDFL	2
chr17:49619281A>C	2	SPOP	p.F102C	CPKSEVRAK **C** KFSILNAKG	2
chr17:49619282A>C	3	SPOP	p.F102V	CPKSEVRAK **V** KFSILNAKG	3
chr17:49619070A>C	2	SPOP	p.W131G	AYRFVQGKD **G** GFKKFIRRD	2

The F133V/C/I/L mutant may be of interest as a universal neoepitope due to the similar chemical properties of Leucine, Isoleucine and Valine.

than the first case since we're completely lacking HLA-A*01:01 peptide binding affinity predictions for all these samples. Overall, 67.5% of all samples had perfectly concordant haplotypes with POLYSOLVER, 28.8% differed by one allele and 3.7% differed by two (**Figure 3**). A large chunk of the second group consists of the miscalls mentioned above. ProTECT allows users to provide pre-computed MHC haplotype calls if they trust another external caller more than PHLAT, or if they have haplotype information from another source.

Comparison With Published Callers

Comparison With an SNV-Based Neoepitope Predictor

We ran ProTECT on the eight melanoma samples from three patients (one primary lymph node tumor each and multiple metachronous tumors in two samples) (29) that were used to benchmark pVAC-Seq (19). Carreno et al. predicted 11–28 expressed, HLA-A*02:01 binding candidate peptides per sample and synthesized seven unique peptide vaccines per patient based on presence of the mutants in the metachronous tumors and assessed binding of the predicted peptide to HLA-A*02:01 in T2 assays. Three peptides per patient were found to induce an immune reaction. ProTECT correctly identified the expected immunogenic mutations in every reported mutation: sample pair. In some cases, ProTECT even predicted the expected variant in a metachronous tumor where the original paper missed it (E.g. CDKN2A:E153K in the Lymph Node of Mel-21) (**Table 3**). Overall, ProTECT ranked IARs containing the validated variants relatively highly (in the top 15–20%, median absolute rank of 11) except in Mel218. We cannot definitively comment on the ranking in Mel218 since ProTECT considers every mutant and MHC allele in the MHC haplotype, while Carreno et al. only considered a curated list of peptides against HLA-A*02:01. In addition to the validated variants, we also provided a larger ranked set of possible candidates that broaden the spectrum of testable epitopes. The

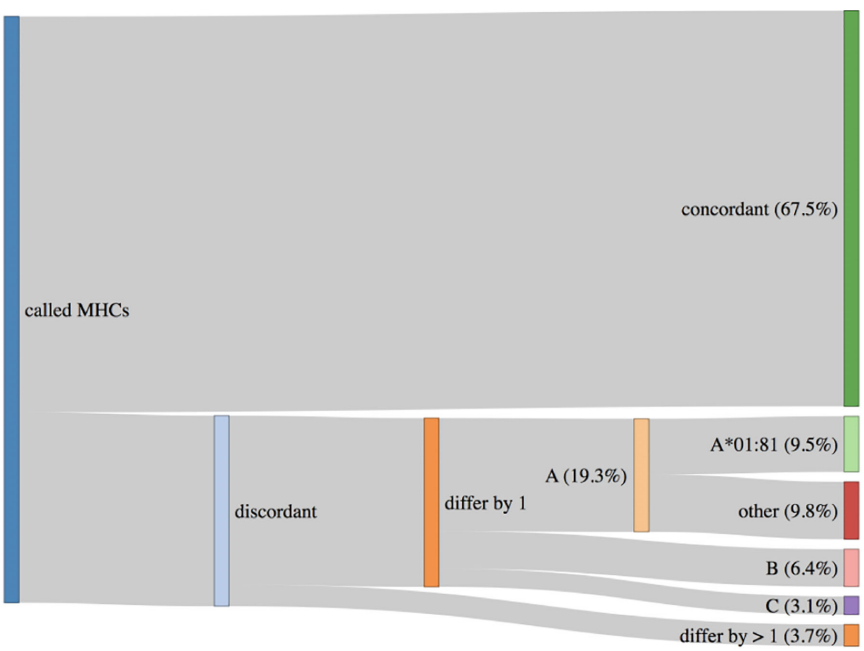


FIGURE 3 | HLA haplotypes called by ProTECT (using PHLAT) are fully concordant with POLYSOLVER haplotypes in only 67.5% of samples. 28.8% differ by one call and 3.7% by 2 calls. A majority of the miscalled HLA-A alleles are a documented PHLAT artifact.

TABLE 3 | ProTECT ranks on eight metachronous tumors across three MELANOMA patients.

Sample and Source	Mel 21				Mel 38			Mel 218
	LN	Skin (2012)	Skin (2013)		Abdominal Wall	Axilla LN	Breast	LN
			RNA 1	RNA 2				
Collection date	1/30/2011	5/10/2012	6/6/2013	6/6/2013	4/16/2013	4/19/2012	2/14/2013	4/4/2005
Total Variants	1,532	2,140	1,681	1,679	1,213	1,121	1,259	2,176
Actionable Variants	332	400	393	391	219	216	224	449
Total IARs	105	137	114	116	73	80	86	155
Vaccine Candidates with ProTECT ranks	NDC1:F169L (Reported as TMEM48 F169L)				SEC24A:P469L			EXOC8:Q656P
	4	5	10	3	8	9	2	152
	TKT:R438W				AKAP13:Q285K			PABPC1:R520Q
	6	6	4	4	14	80	17	140
	CDKN2A:E153K				OR8B3:T190I			MRPS5:P59L
	21	–	18	17	11	13	14	26

Highlighted ranks describe instances where pVAC-Seq and ProTECT both call a neoepitope. Green: Dominant epitope (existing immunity, neoantigen processed from endogenous protein), Orange: Subdominant epitope (immunity after vaccination, neoantigen processed from endogenous protein), Red: Cryptic epitope (immunity after vaccination, neoantigen not processed from endogenous protein).

data for all seven tested peptides is provided in **Supplementary Table 2**, and all neoepitopes predicted by ProTECT in **Supplementary Table 3** and **Supplementary File 2**.

Comparison With a Fusion-Gene-Based Neoepitope Predictor

We compared our fusion prediction accuracy with INTEGRATE-Neo (23). INTEGRATE-Neo was demonstrated on 321 samples from the TCGA PRAD cohort, and at least one neoepitope was predicted from 161 samples. 240 of the 321 samples overlap with our 326 sample dataset, and this subset was used for this experiment. None of the predicted neoepitopes in this study have been validated using any biological experiments. We first attempted to compare our fusions (called using STAR-Fusion) with the fusion calls generated from INTEGRATE (64) as fusion callers are known for having varied performance across different datasets (65). As expected, the overlap between the ProTECT and INTEGRATE calls was relatively low (595/1519, with 120 unique calls in ProTECT), but a large chunk of the non-overlapping calls were from events with one spanning read support in INTEGRATE (**Supplementary Figure 3**). We see a better overlap when we increase the minimum support to two (an internal metric within ProTECT) and also find that 44 events rejected for having one read support in INTEGRATE were detected by STAR-Fusion (40) with >1 read support. Some of the INTEGRATE-specific calls were picked up by ProTECT but filtered out as low-read-evidence events. We further noticed that the concordance between MHC haplotypes called by HLA Miner (66) (used by INTEGRATE-Neo), PHLAT (41) (Used by ProTECT), and POLYSOLVER (32) was very low (**Supplementary Figure 4**). 61 of the unique HLAMiner predictions across the cohort did not match any of the other two callers, and 41 matched both. (Homozygous calls in a patient were treated as one call.) Two alleles were shared exclusively between ProTECT and INTEGRATE and only one between INTEGRATE and POLYSOLVER. In order to conduct a more comparable analysis, we reran ProTECT with the INTEGRATE fusion calls and the MHC haplotypes from the INTEGRATE-Neo manuscript (182 neoepitopes from 720 fusions over 83 samples, **Supplementary Figure 5**). ProTECT rejected 100 of the 720 provided fusion events as transcriptional readthroughs (92 events) or for having a 5′ non-coding RNA partner (eight events). ProTECT correctly identified 139/182 neoepitopes as IARs and rejected the remaining for being in a rejected fusion (23 neoepitopes), scoring below the 5% predicted binding score threshold (16 neoepitopes), having a 5′ breakpoint in the UTR (three neoepitopes), or for having a 5′ non-coding partner (one neoepitope) (**Supplementary Table 4**). On further inspection, we noticed that the three neoepitopes arising from the 5′ UTR breakpoints (TCGA-HC-7080, PRH1>>RP11-259O18.4 and PRH1>>M6PR) could have been detected if the 5′ partner had been annotated with a different gene (PRR4) at the same locus (**Supplementary Figure 6**), an issue arising due to the differing gene annotation GTFs used between the methods (Gencode v25 for ProTECT and Ensembl v85 for INTEGRATE). Interestingly, this type of event occurred in one other sample (TCGA-EJ-8474, C1QTNF3-AMACR>>NDUFAF2); however, INTEGRATE called the overlapping call as well (AMACR>>NDUFAF2), and since the epitopes were identical from both, ProTECT picked them up under the correct call (**Supplementary Figure 7**). The full set of results from running ProTECT on 83 INTEGRATE-Neo inputs is provided in **Supplementary File 3**. Easing ProTECT's 5% filter would increase the number of false positives called by too large a margin, so we stand by our decision to reject the 16 neoepitopes missed due to this filter. This experiment also highlights the modularity of ProTECT and its flexibility in accepting pre-computed inputs to run only the necessary steps to produce a ranked list of IARs.

Reproducibility

Every tool used the pipeline, from established aligners to the in-house script used to translate mutations, is wrapped in a Docker

image (67) tagged with the appropriate tool version. Docker allows a developer to wrap a piece of code and any requirements into an image that can be instantiated into a container on any other machine. The tool within the container will run in the same manner on any machine, under the same environmental constraints, barring minor differences that may arise from asynchronous multiprocessing/multithreading. This way, results from ProTECT run on different machines with the same inputs will always be near-identical. The default versions of each tool used by ProTECT are mentioned in the repository, and users can containerize other versions of the same tools and specify the new version to ProTECT at runtime.

Automation, Scalability, and Efficiency

ProTECT is built to be run end-to-end without any user intervention. ProTECT is written in the Toil framework and will attempt to run the pipeline on the given input samples in a resource-efficient manner. The pluggable backend Toil APIs allow ProTECT to run on a single machine, a grid engine cluster, or a Mesos cluster setup on a local network or on AWS. Toil allows users to deploy scripts on Azure and Google cloud as well; however, ProTECT does not yet support these environments.

Users provide ProTECT a config file that details the input files and the various indexes and versions of tools to use during the run. ProTECT copies (or downloads) the files to a "file store" and then queues a graph of jobs for each input sample culminating in a ranked list of epitopes. The nodes in the graph are tuned to request an appropriate number of CPUs (for multithreaded jobs), memory, and disk space. Toil ensures that these queued jobs are spawned in a way that all available resources are utilized to the maximum extent. In practice, this means that smaller, low-compute, low-memory, short-duration jobs like variant calling, mutation translation, *etc.* from one sample can run parallel to higher-compute, high-memory, long-running jobs like alignment and haplotyping in another. The processing time of any single sample is strongly influenced by the long-running jobs but utilization of free compute to run queued short-jobs reduced the overall per-sample runtime.

CONCLUSION

We have described an efficient, automated, and portable workflow for the prediction of neoepitope candidates that can guide vaccine-based or adoptive T-cell therapies. We have shown that ProTECT scales well on a parallel processing environment and is highly efficient processing samples in large batches. On average, we processed a sample from end-to-end in 26.4 min when we ran 50 samples in a single batch on an eight-node cluster. We have shown that ProTECT is comparable to existing callers and improves on them by providing a ranked list of neoepitopes arising from SNVs, INDELs, and fusion genes. None of the currently published pipelines give results for all three types of mutations. Positive results from a clinical trial were ranked highly in our results, and we retrospectively identified additional events that could have been used in the trial. We identified recurrent epitopes arising from the well-documented TMPRSS2-ERG fusion, and these results suggest a peptide or RNA vaccine could be developed for one of the common breakpoints. While designed for use in the rapidly growing fields of cancer vaccines and Autologous T-cell therapies, ProTECT can also be used to understand the link between tumor mutational burden and response to checkpoint blockade therapies. It is our fervent hope that improvements in these fields will quickly establish neoepitope-targeted immunotherapies as standard-of-care for cancer treatment.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

AR, SS, and DH contributed to the conception and design of the study. AR developed the entire codebase available on github with significant contributions from JP. AM was involved with analyses on the 326 sample run. AR wrote the first draft of the manuscript; SS, DH, and AR contributed to manuscript revision. DH, SS, and BP acquired funding for this work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institutes of Health/National Cancer Institute [5U24CA143858], an Alex's Lemonade Stand Innovation Award, and the Howard Hughes Medical Institute.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of Hannes Schmidt, Christopher Ketchum, Aisling O'Farrell, Swapnil Patil, Akul Goyal and Walter Shands towards the ProTECT source code, and Bala Ganeshan, Giridhar Basava, Hemant Trivedi of Stellus Technologies for the compute support that enabled the 326 TCGA sample run. We thank Dr. Catherine Wu for her permission to use the POLYSOLVER MHC calls for the TCGA PRAD study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.483296/full#supplementary-material

REFERENCES

1. Muul LM, Spiess PJ, Director EP, Rosenberg SA. Identification of specific cytolytic immune responses against autologous tumor in humans bearing malignant melanoma. *J Immunol* (1987) 138:989–95.
2. Facciabene A, Motz GT, Coukos G. T-Regulatory Cells: Key Players in Tumor Immune Escape and Angiogenesis. *Cancer Res* (2012) 72:2162–71. doi: 10.1158/0008-5472.CAN-11-3687
3. Ferris RL, Hunt JL, Ferrone S. Human leukocyte antigen (HLA) class I defects in head and neck cancer: molecular mechanisms and clinical significance. *Immunol Res* (2005) 33:113–33. doi: 10.1385/IR:33:2:113
4. Ghiringhelli F, Puig PE, Roux S, Parcellier A, Schmitt E, Solary E, et al. Tumor cells convert immature myeloid dendritic cells into TGF-β–secreting cells inducing CD4+CD25+ regulatory T cell proliferation. *J Exp Med* (2005) 202:919–29. doi: 10.1084/jem.20050463
5. Shevach EM. Fatal attraction: tumors beckon regulatory T cells. *Nat Med* (2004) 10:900–1. doi: 10.1038/nm0904-900
6. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N Engl J Med* (2016) 375:1823–33. doi: 10.1056/NEJMoa1606774
7. Carbone DP, Reck M, Paz-Ares L, Creelan B, Horn L, Steins M, et al. First-Line Nivolumab in Stage IV or Recurrent Non–Small-Cell Lung Cancer. *N Engl J Med* (2017) 376:2415–26. doi: 10.1056/NEJMoa1613493
8. Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N Engl J Med* (2010) 363:711–23. doi: 10.1056/NEJMoa1003466
9. Alsaab HO, Sau S, Alzhrani R, Tatiparti K, Bhise K, Kashaw SK, et al. PD-1 and PD-L1 Checkpoint Signaling Inhibition for Cancer Immunotherapy: Mechanism, Combinations, and Clinical Outcome. *Front Pharmacol* (2017) 8 (561). doi: 10.3389/fphar.2017.00561
10. Chambers CA, Kuhns MS, Egen JG, Allison JP. CTLA-4-Mediated Inhibition in Regulation of T Cell Responses: Mechanisms and Manipulation in Tumor Immunotherapy. *Annu Rev Immunol* (2001) 19:565–94. doi: 10.1146/annurev.immunol.19.1.565
11. Ryder M, Callahan M, Postow MA, Wolchok J, Fagin JA. Endocrine-related adverse events following ipilimumab in patients with advanced melanoma: a comprehensive retrospective review from a single institution. *Endocr Relat Cancer* (2014) 21:371–81. doi: 10.1530/ERC-13-0499
12. van Elsas A, Sutmuller RPM, Hurwitz AA, Ziskin J, Villasenor J, Medema J-P, et al. Elucidating the Autoimmune and Antitumor Effector Mechanisms of a Treatment Based on Cytotoxic T Lymphocyte Antigen-4 Blockade in Combination with a B16 Melanoma Vaccine: Comparison of Prophylaxis and Therapy. *J Exp Med* (2001) 194:481–90. doi: 10.1084/jem.194.4.481
13. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther* (2017) 16:2598–608. doi: 10.1158/1535-7163.MCT-17-0386
14. Samstein RM, Lee C-H, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* (2019) 51:202. doi: 10.1038/s41588-018-0312-8
15. Germano G, Lamba S, Rospo G, Barault L, Magrì A, Maione F, et al. Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. *Nature* (2017) 552:116–20. doi: 10.1038/nature24673
16. Rosenberg SA, Yang JC, Sherry RM, Kammula US, Hughes MS, Phan GQ, et al. Durable Complete Responses in Heavily Pretreated Patients with Metastatic Melanoma Using T-Cell Transfer Immunotherapy. *Clin Cancer Res* (2011) 17:4550–7. doi: 10.1158/1078-0432.CCR-11-0116
17. Zacharakis N, Chinnasamy H, Black M, Xu H, Lu Y-C, Zheng Z, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat Med* (2018) 24:724–30. doi: 10.1038/s41591-018-0040-8
18. McNamara MA, Nair SK, Holl EK. RNA-Based Vaccines in Cancer Immunotherapy. *J Immunol Res* (2015) 2015:794528. doi: 10.1155/2015/794528
19. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, et al. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med* (2016) 8:11. doi: 10.1186/s13073-016-0264-5
20. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol* (2016) 17:122. doi: 10.1186/s13059-016-0974-4
21. Rubinsteyn A, Hodes I, Kodysh J, Hammerbacher J. Vaxrank: A computational tool for designing personalized cancer vaccines. *bioRxiv* (2018) 142919. doi: 10.1101/142919
22. Rubinsteyn A, Kodysh J, Hodes I, Mondet S, Aksoy BA, Finnigan JP, et al. Computational Pipeline for the PGV-001 Neoantigen Vaccine Trial. *Front Immunol* (2018) 8:1807. doi: 10.3389/fimmu.2017.01807
23. Zhang J, Mardis ER, Maher CA. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics* (2017) 33:555–7. doi: 10.1093/bioinformatics/btw674
24. Chang T, Carter RA, Li Y, Li Y, Wang H, Edmonson MN, et al. The neoepitope landscape in pediatric cancers. *Genome Med* (2017) 9:78. doi: 10.1186/s13073-017-0468-3
25. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res* (2008) 36:W509–12. doi: 10.1093/nar/gkn202
26. Fotakis G, Rieder D, Haider M, Trajanoski Z, Finotello F. NeoFuse: predicting fusion neoantigens from RNA sequencing data. *Bioinformatics* 36(7):2260–1. doi: 10.1093/bioinformatics/btz879
27. Toor JS, Rao AA, McShan AC, Yarmarkovich M, Nerli S, Yamaguchi K, et al. A Recurrent Mutation in Anaplastic Lymphoma Kinase with Distinct Neoepitope Conformations. *Front Immunol* (2018) 9:99. doi: 10.3389/fimmu.2018.00099
28. The Cancer Genome Atlas Research network. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* (2015) 163:1011–25. doi: 10.1016/j.cell.2015.10.025
29. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* (2015) 348(6236):803–8. doi: 10.1126/science.aaa3828
30. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* (2013) 499:214–8. doi: 10.1038/nature12213
31. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* (2015) 34:4845–54. doi: 10.1038/onc.2014.406
32. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* (2015) 33:1152–8. doi: 10.1038/nbt.3344
33. Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. *Nucleic Acids Res* (2011) 39:D19–21. doi: 10.1093/nar/gkq1019
34. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res* (2002) 12:996–1006. doi: 10.1101/gr.229102
35. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* (2012) 22:1760–74. doi: 10.1101/gr.135350.111
36. Hindman B, Konwinski A, Zaharia M, Ghodsi A, Joseph AD, Katz R, et al. *Mesos: A Platform for Fine-grained Resource Sharing in the Data Center. in Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation NSDI"11*. Berkeley, CA, USA: USENIX Association. Available at: http://dl.acm.org/citation.cfm?id=1972457.1972488 (Accessed June 7, 2019).
37. Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotech* (2017) 35:314–6. doi: 10.1038/nbt.3772
38. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
39. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635
40. Haas B, Dobin A, Stransky N, Li B, Yang X, Tickle T, et al. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv* (2017). doi: 10.1101/120295

41. Bai Y, Ni M, Cooper B, Wei Y, Fury W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics* (2014) 15:325. doi: 10.1186/1471-2164-15-325
42. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf* (2011) 12:323. doi: 10.1186/1471-2105-12-323
43. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotech* (2013) 31:213–9. doi: 10.1038/nbt.2514
44. Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal PA, et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol* (2016) 17:178. doi: 10.1186/s13059-016-1029-6
45. Radenbaugh AJ, Ma S, Ewing A, Stuart JM, Collisson EA, Zhu J, et al. RADIA: RNA and DNA Integrated Analysis for Somatic Mutation Detection. *PloS One* (2014) 9:e111516. doi: 10.1371/journal.pone.0111516
46. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* (2012) 28:311–7. doi: 10.1093/bioinformatics/btr665
47. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* (2012) 28:1811–7. doi: 10.1093/bioinformatics/bts271
48. Ewing AD, Houlahan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, et al. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat Methods* (2015) 12:623–30. doi: 10.1038/nmeth.3407
49. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* (2013) 8:1494–512. doi: 10.1038/nprot.2013.084
50. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* (2012) 6:80–92. doi: 10.4161/fly.19695
51. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* (2003) 12:1007–17. doi: 10.1110/ps.0239403
52. Bui H-H, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton K-A, et al. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics* (2005) 57:304–14. doi: 10.1007/s00251-005-0798-y
53. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinf* (2005) 6:132. doi: 10.1186/1471-2105-6-132
54. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* (2012) 64:177–86. doi: 10.1007/s00251-011-0579-8
55. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics* (2009) 25:1293–9. doi: 10.1093/bioinformatics/btp137
56. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinf* (2007) 8:238. doi: 10.1186/1471-2105-8-238
57. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* (1999) 17:555–61. doi: 10.1038/9858
58. Tomlins SA, Laxman B, Varambally S, Cao X, Yu J, Helgeson BE, et al. Role of the TMPRSS2-ERG Gene Fusion in Prostate Cancer. *Neoplasia* (2008) 10:177–88. doi: 10.1593/neo.07822
59. Wang Z, Wang Y, Zhang J, Hu Q, Zhi F, Zhang S, et al. Significance of the TMPRSS2:ERG gene fusion in prostate cancer. *Mol Med Rep* (2017) 16:5450–8. doi: 10.3892/mmr.2017.7281
60. Kron KJ, Murison A, Zhou S, Huang V, Yamaguchi TN, Shiah Y-J, et al. TMPRSS2–ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat Genet* (2017) 49:1336–45. doi: 10.1038/ng.3930
61. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat J-P, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* (2012) 44:685–9. doi: 10.1038/ng.2279
62. Blattner M, Lee DJ, O'Reilly C, Park K, MacDonald TY, Khani F, et al. SPOP Mutations in Prostate Cancer across Demographically Diverse Patient Cohorts. *Neoplasia* (2014) 16:14–W10. doi: 10.1593/neo.131704
63. Kiyotani K, Mai TH, Nakamura Y. Comparison of exome-based HLA class I genotyping tools: identification of platform-specific genotyping errors. *J Hum Genet* (2017) 62:397. doi: 10.1038/jhg.2016.141
64. Zhang J, White NM, Schmidt HK, Fulton RS, Tomlinson C, Warren WC, et al. INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Genome Res* (2016) 26:108–18. doi: 10.1101/gr.186114.114
65. Liu S, Tsai W-H, Ding Y, Chen R, Fang Z, Huo Z, et al. Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Res* (2016) 44:e47–7. doi: 10.1093/nar/gkv1234
66. Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, et al. Derivation of HLA types from shotgun sequence datasets. *Genome Med* (2012) 4:95. doi: 10.1186/gm396
67. Merkel D. *Docker: Lightweight Linux Containers for Consistent Development and Deployment. Linux J* (2014). Available at: http://dl.acm.org/citation.cfm?id=2600239.2600241 (Accessed June 7, 2019).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JI-S and handling editor declared their shared affiliation at the time of the review.

frontiers in Immunology

REVIEW
published: 15 July 2021
doi: 10.3389/fimmu.2021.592031

Targeting Neoepitopes to Treat Solid Malignancies: Immunosurgery

*Eric de Sousa[1], Joana R. Lérias[1], Antonio Beltran[2], Georgia Paraschoudi[1], Carolina Condeço[1], Jéssica Kamiki[1], Patrícia Alexandra António[1], Nuno Figueiredo[3], Carlos Carvalho[3], Mireia Castillo-Martin[2], Zhe Wang[4], Dário Ligeiro[5], Martin Rao[1] and Markus Maeurer[1,6]**

[1] ImmunoSurgery Unit, Champalimaud Centre for the Unknown, Lisbon, Portugal, [2] Department of Pathology, Champalimaud Clinical Centre, Lisbon, Portugal, [3] Digestive Unit, Champalimaud Clinical Centre, Lisbon, Portugal, [4] Jiangsu Industrial Technology Research Institute (JITRI), Applied Adaptome Immunology Institute, Nanjing, China, [5] Lisbon Centre for Blood and Transplantation, Instituto Português do Sangue e Transplantação (IPST), Lisbon, Portugal, [6] I Medical Clinic, Johannes Gutenberg University of Mainz, Mainz, Germany

OPEN ACCESS

Edited by:
Nikolaos G. Sgourakis, University of Pennsylvania, United States

Reviewed by:
Juliane Liepe, Max Planck Institute for Biophysical Chemistry, Germany
Karen Taraszka Hastings, University of Arizona, United States

***Correspondence:**
Markus Maeurer
markus.maeurer@fundacaochampalimaud.pt

Specialty section:
This article was submitted to Cancer Immunity and Immunotherapy, a section of the journal Frontiers in Immunology

Received: 06 August 2020
Accepted: 07 May 2021
Published: 15 July 2021

Citation:
de Sousa E, Lérias JR, Beltran A, Paraschoudi G, Condeço C, Kamiki J, António PA, Figueiredo N, Carvalho C, Castillo-Martin M, Wang Z, Ligeiro D, Rao M and Maeurer M (2021) Targeting Neoepitopes to Treat Solid Malignancies: Immunosurgery. Front. Immunol. 12:592031. doi: 10.3389/fimmu.2021.592031

Successful outcome of immune checkpoint blockade in patients with solid cancers is in part associated with a high tumor mutational burden (TMB) and the recognition of private neoantigens by T-cells. The quality and quantity of target recognition is determined by the repertoire of 'neoepitope'-specific T-cell receptors (TCRs) in tumor-infiltrating lymphocytes (TIL), or peripheral T-cells. Interferon gamma (IFN-γ), produced by T-cells and other immune cells, is essential for controlling proliferation of transformed cells, induction of apoptosis and enhancing human leukocyte antigen (HLA) expression, thereby increasing immunogenicity of cancer cells. TCR αβ-dependent therapies should account for tumor heterogeneity and availability of the TCR repertoire capable of reacting to neoepitopes and functional HLA pathways. Immunogenic epitopes in the tumor-stroma may also be targeted to achieve tumor-containment by changing the immune-contexture in the tumor microenvironment (TME). Non protein-coding regions of the tumor-cell genome may also contain many aberrantly expressed, non-mutated tumor-associated antigens (TAAs) capable of eliciting productive anti-tumor immune responses. Whole-exome sequencing (WES) and/or RNA sequencing (RNA-Seq) of cancer tissue, combined with several layers of bioinformatic analysis is commonly used to predict possible neoepitopes present in clinical samples. At the ImmunoSurgery Unit of the Champalimaud Centre for the Unknown (CCU), a pipeline combining several tools is used for predicting private mutations from WES and RNA-Seq data followed by the construction of synthetic peptides tailored for immunological response assessment reflecting the patient's tumor mutations, guided by MHC typing. Subsequent immunoassays allow the detection of differential IFN-γ production patterns associated with (intra-tumoral) spatiotemporal differences in TIL or peripheral T-cells versus TIL. These bioinformatics tools, in addition to histopathological assessment, immunological readouts from functional bioassays and deep T-cell 'adaptome' analyses, are expected to advance discovery and development of next-generation personalized precision medicine strategies to improve clinical outcomes in cancer in the context of i) anti-tumor vaccination

strategies, ii) gauging mutation-reactive T-cell responses in biological therapies and iii) expansion of tumor-reactive T-cells for the cellular treatment of patients with cancer.

Keywords: T-cells, antigens, TIL, neoepitopes, precision medicine, vaccination, T-cell receptor, immunotherapy

INTRODUCTION

'*Personalized immunotherapy is all the rage, but neoantigen discovery and validation remains a daunting problem*' echoed an Editorial in Nature Biotechnology 2017 (1). Advances in the last three years in whole exome sequencing (WES), RNA sequencing (RNA-Seq) and combinational peptide vaccination trials combined with checkpoint inhibitors addressed some of the unanswered questions and challenges in therapeutic vaccinations using neoepitopes. Biologically and clinically relevant immune responses happen in distinct immunological contexts, they are dependent on antigen processing, presentation and on the available T-cell receptor (TCR) repertoire that is shaped by previous encounters with antigens. The immune synapse between the major histocompatibility complex (MHC)-peptide and TCR interaction is the center of T-cell activation, which is orchestrated by cells of the innate and adaptive immune response that guides and edit neoepitope-specific T-cell responses. We will therefore review various immune cell types that contribute to successful cellular immune responses and expansion of neoepitope-directed T-cells. Finally, we address in practical terms how neoepitopes are identified in cancer tissue specimens starting with immunohistology, WES, RNA-Seq and epitope prediction algorithms using standard prediction programs.

Tumor mutational burden (TMB) is a key factor in determining the response of patients with cancer to immunotherapy with immune checkpoint inhibitors (anti-programmed cell death 1 [PD-1] or anti-cytotoxic T lymphocyte-associated antigen 4 [CTLA-4]) (2–7). The 'mutanome', the summary of mutations developing over the course of disease is unique from one patient to another, thus making the TMB a unique biological signature comprising of druggable targets and epitopes to elicit anti-cancer immune responses. Alexandrov and colleagues elegantly showed that varying degrees of TMB are associated with different cancer types, and that disease-specific mutational signatures may either be widespread (e.g. melanoma and lung cancer) or restricted (e.g. pancreatic cancer) to certain parts of the genome – thus influencing the number of mutant genes and inevitably the availability and immunogenicity of neoantigens (8). A large proportion of favorable clinical responses rely on a rich reservoir of tumor-infiltrating lymphocytes (TIL) as well as circulating tumor-directed T-cells and, therefore, TCRs which recognize neoepitopes presented by human leukocyte antigen (HLA) molecules on tumor cells (9–19). The number of mutations which are identified through bioinformatics directly influence the repertoire size of immunogenic targets that may induce T-cell responses and potentially anti-tumor directed T-cell responses (**Figure 1**). Although companion diagnostics for PD-1, programmed death-ligand 1 (PD-L1) and CTLA-4 are actively used prior to initiating immunotherapy to confirm expression in tumor tissue samples, mutations in the HLA pathways may often be overlooked – which will impair or abolish productive anti-cancer directed cellular immune responses. In addition, other immunologically relevant mutations or natural variations which may inherently affect immune function and T-cell responses deserve equal attention if these factors influence the quality and quantity of anti-cancer directed immune responses. The TMB is still considered a key factor in predicting clinical responsiveness or to gauge the possibility of the immune system to productively react against cancer cells. Yet the TMB represents only the substrate of potential immune reactivity and the immune system is not objectively considered and analyzed. The TMB is therefore increasingly viewed as an important yet 'imperfect' surrogate marker for clinical responsiveness and the corresponding elements in orchestrating a cellular immune response, namely the MHC genetic background as well as the T-cell receptor repertoire capable of reacting to potential cancer neopitopes, are now considered to be analyzed as well to gauge for immune response analysis (20). The nature and the histological location of T-cells that serve to functionally test for immune recognition of neoepitopes are therefore also considered in this review. We will also highlight in this review relevant findings from clinical and translational studies pertaining to personalized cancer immunotherapy. We discuss HLA mutations in tumor lesions from patients with cancer and discuss how this information is necessary for designing personalized immunotherapy clinical trials. Finally, we propose the combined use of well-established techniques such as immunohistochemistry (IHC) and flow cytometry in conjunction with next-generation sequencing methods to assist in making better informed clinical decisions for treatment regimens, a concept that has been implemented at the ImmunoSurgery Unit and Anatomic Pathology Clinical Service at the Champalimaud Centre for the Unknown (CCU) and the Champalimaud Clinical Centre (CCC) in Lisbon, Portugal (**Figure 2**).

NEXT-GENERATION SEQUENCING: THE FUEL OF PRECISION ONCOLOGY

Advances in next-generation sequencing (NGS) techniques such as WES and RNA-seq form the bedrock of personalized precision medicine in neoantigen-directed immuno-oncology (22, 23). Immunoediting leading to neoantigen generation and turnover in the tumor microenvironment (TME) influencing T-cell infiltration and survival in patients with advanced cancer (24–26). This also goes hand-in-hand with the MHC background of the patient as well as the capacity to present the 'best' neoepitope candidates to evoke meaningful and clinically

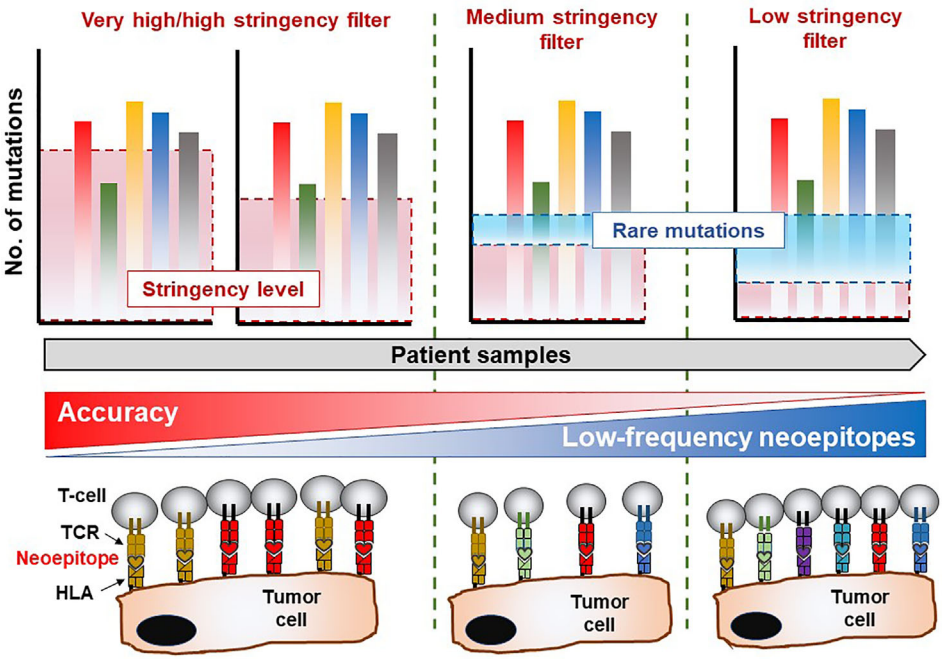


FIGURE 1 | Mutation analysis reveals immune-recognition profile in the TME. Whole-exome sequencing data allows for mining of private somatic mutations in tumor samples compared to healthy (non-transformed) tissue or cells, which is unique to each patient. The stringency of the filtering parameters applied in bioinformatics and statistical analysis of the sequencing data will greatly influence the number of mutations recovered, which are essential for downstream characterization of immune responses of T-cell products. Highly stringent parameters may yield a lower number of mutations albeit with an exceptional level of accuracy. Nevertheless, this approach suffers the risk of overlooking several infrequent mutations which also give rise to immunogenic (T-cell reactive) neoepitopes in the patient. On the contrary, reducing the stringency levels of analysis may reveal rare mutations which facilitate the identification of potentially immunogenic molecular targets recognized by certain TCRs capable of eliciting a biologically relevant anti-tumor immune response. The drawback in the latter scenario is that a high degree of false positive hits may be obtained and included in the final list of legitimate cancer-associated somatic mutations. Thus, a balanced yet wholistic approach is required to identify all immunogenic mutations in tumor tissue which will be instrumental in developing personalized cancer therapies.

beneficial T-cell responses (27–29). Importantly, juxtaposition of tumor-specific T-cells to the tumor cells themselves provides promising prognosis, suggesting that the local 'cell-cell' interaction between neoepitope-specific lymphocytes and tumor cells is clinically beneficial and desirable (30). Treatments affecting the activity of cancer-associated fibroblasts (CAF) or tumor-associated macrophages (TAM) (e.g. monoclonal antibodies such as anti-CD47 or anti-CD40) redirect T-cells to these nominal target cells which appear to be associated with improved anti-tumor responses in a clinical setting (31–36). Furthermore, evolution of the neoantigen landscape under treatment pressure, such as standard chemotherapy, immune checkpoint blockade or active cellular therapy/adoptive cell transfer (ACT) is an essential determinant of how patient immune response patterns are modulated and change over time (37, 38). In line with this, neoantigen fitness – the propensity of mutated host targets which differ significantly enough from the wildtype form to be able to produce a biologically meaningful anti-tumor response, further to their HLA-binding strength – can be mathematically modeled to predict survival dynamics of patients and aid in the selection of promising neoepitope candidates for immunotherapy protocols (39). The fitness of a (cancer) cell clone is defined by several factors, e.g. the recognition potential of the (immunodominant) neoepitopes by nominal anti-cancer directed TCRs that will aid to estimate the future size of the cancer cell population. 'Immunodominance' can be gauged by comparing the affinity of the wildtype and the corresponding mutant candidate target epitope that would impose selective pressure on the clonal pool of available TCRs that recognize the MHC-peptide complex.

Clinically, TIL therapy targeting individual neoepitopes has been proven to be successful, with the capacity to promote durable anti-tumor responses in patients with solid tumors (17, 40). Clinical responses appear to be associated with the frequency of neoepitope-specific T-cells in the T-cell product (40). Mutant KRAS-directed TIL and TCR transfer therapy has also shown great clinical promise, albeit in an HLA allele-dependent manner (41). In addition to neoantigens, non-coding regions of the cancer genome giving rise to previously undefined, non-mutated peptides with immunogenic properties can also be mined for, using NGS strategies (42), as well as peptides resulting from novel gene fusions (43). This pattern may differ from patient to patient, necessitating the use of *in-silico* analyses to select matching HLA-epitope sets for a personalized therapy protocol. Thus, private and shared neoantigens as well as hitherto unknown immunogenic peptides can trigger beneficial clinical responses in patients with advanced cancer (16, 44, 45).

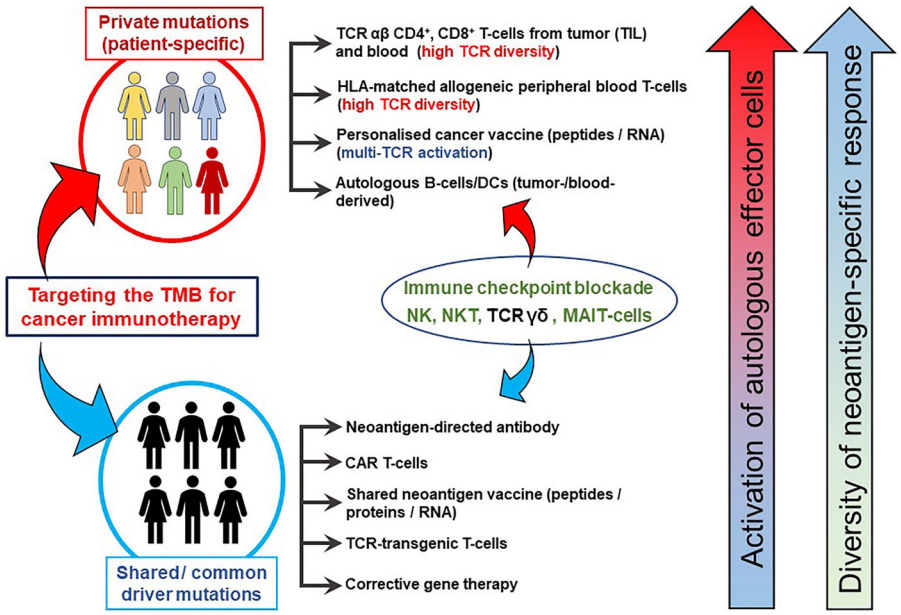


FIGURE 2 | TMB-directed immunotherapy approaches at the Champalimaud Centre for the Unknown. The schematic shows strategies aimed at therapeutic targeting of private (personalized and patient-specific) and shared (often driver) mutations. For personalized therapy, $CD4^+$ and $CD8^+$ T-cells from TIL or peripheral blood expressing a highly diverse TCR αβ repertoire recognizing a private neoepitopes can be procured. HLA-matched, healthy donor-derived TCRs have also been shown to recognize patient-specific neoepitopes (21). Personalized cancer vaccines, comprising private neoepitopes as a peptide formulation or as RNA constructs, promote durable immune responses in patients with advanced cancer. Autologous B-cells can be used as a source of APCs as well as cytokine producers, in addition to their differentiation into plasma cells to secrete tumor antigen-specific antibodies *in vivo*. Approaches targeting shared mutations serve as excellent 'off-the-shelf' options which can be used for larger groups of patients simultaneously. Cancer vaccines based on shared mutations are also clinically important, provided the patients' HLA profiles are matched to the epitope binding characteristics. Antibodies derived from tumor-infiltrating B-cells or from peripheral blood B-cells targeting surface-bound shared neoantigens may mediating cellular cytotoxicity and aid in the development of CAR T-cells. Gene therapy to correct shared driver mutations may promote tumor susceptibility to immune attack. Immune checkpoint blockade has been placed between the two domains as its clinical activity targets both private and shared mutated targets. Similarly, NK, TCR γδ T-cells and possibly NKT T-cells or MAIT-cells may be instrumental in patients presenting with private and/or shared HLA pathway mutations and can be derived from allogeneic sources for treatment.

NGS readouts combined with recent advancements in immune-based analysis of patient-derived tumor and blood samples are able to provide a wealth of information concerning the presence of dynamics of cancer-specific T-cells suitable for immunotherapy development or for immuno-monitoring following treatment, including neoepitope specificity and TCR tracking (14, 18, 46–49). Neoepitope screening has enabled the identification of private mutation-directed TIL from pancreatic cancer (18, 45, 50) and glioblastoma (14) demonstrating that tumor histologies previously considered poorly immunogenic may also contain a broad repertoire of neoantigen-reactive immune cells (11, 16, 51–53). Specific neoepitopes involved in eliciting productive immune responses that promote tumor regression either by engaging cellular cytotoxicity or by cytokine production (e.g. IFN-γ) are therefore particularly attractive for developing personalized therapies within the framework of precision cancer medicine (54, 55). However, there is also a need to identify target neoepitopes which are most likely to induce regulatory T-cell responses among TIL to the effect of dampening productive anti-tumor reactivity in patients (56).

Druggable mutations (e.g. those associated with ROS1, ALK, tropomycin receptor kinase [TRK] and NTRK1/2/3 chromosomal fusions) which have been implicated to be responsible for clinical responses in pediatric central nervous system (CNS) malignancies provide a roadmap for how NGS is able to support precision oncology based on selected small molecules (i.e. Entrectinib and Larotrectinib) (57). Hitherto unknown mutational events can be captured *via* NGS, possibly expanding the use of existing targeted cancer drugs and, in addition, newly devised immunotherapeutic strategies. These probable candidates can be tested for T-cell and antibody reactivity *in vitro*, and positive results can then be followed up with more detailed analysis to enable the formulation of personalized cancer vaccines (PCVs) or cellular immunotherapy development (chimeric antigen receptor T-cells [CAR-T], TCR transfer, ACT of TIL and/or memory B-cells), paving the way for combination therapies, e.g. with tyrosine kinase inhibitors (TKIs) and immune-based interventions.

PERSONALIZED CANCER VACCINES

Building on the therapeutic value of targeting cancer-associated mutations, mutation-directed cancer immunotherapy based on PCVs represent a highly specialized approach to induce clinically

relevant and specifically tailored anti-tumor immune responses in patients with advanced malignancies (54, 58, 59). A central point is whether epitopes can be presented by HLA class I or II molecules based on their fitting into the epitope-binding groove and be tailored *in silico*, or whether natural processing by antigen-presenting cells (APCs) in the host or dendritic cell (DC)-based vaccines would be more advantageous (e.g. if antigens were delivered as "long peptides") (60–62), or in a vectored format (e.g. genetically-reengineered viruses and bacteria) (63–68). A carefully selected panel of private and shared cancer-related mutations (e.g. common driver mutations in genes such as *KRAS*, *SMAD4*, *TP53*) identified by WES that bind to the HLA class I and II restriction elements of the patient constitute the formulation of some PCVs (27, 69–72). New research based on high-throughput NGS data shows that the hydrophobicity of predicted neoepitopes could, in part, determine better HLA-binding capacity (28). Longer peptide sequences are likely to contain both HLA class I and class II peptides and would, therefore, activate tumor-directed $CD8^+$ and $CD4^+$ T-cells facilitated by cross presentation of antigens in antigen-presenting cells (i.e. DC, macrophages, B-cells as well as tumor cells) (58, 73). PCV constituent peptides may also be used as lead molecules to construct HLA tetramers or as T-cell stimulants to screen for the presence of neoantigen-specific TCRs in blood samples of patients with cancer (16, 44, 74).

A number of trials to test neoantigen-based PCVs in patients with advanced cancer – including pancreatic cancer – have been registered (58, 59, 75). PCV strategies which have been clinically evaluated are based on direct delivery of messenger RNA (mRNA) sequences of private neoepitopes to the lymph nodes (76), dendritic cells loaded with the patient's tumor lysate, private mutated peptides (neoantigens) (12, 71, 77–80) or clinical-grade neoepitope peptide sequences injected alongside a strong adjuvant or immunostimulant (i.e. poly-ICLC) (69, 72). Montanide®, which is based on antigens from *Mycobacterium tuberculosis* (81–84), and QS-21, extract derived from the soap bark tree *Quillaja saponaria* (85, 86), are also candidates for use as adjuvants in PCVs based on previous clinical experience. Hu and colleagues have comprehensively summarized and elaborated on the current landscape in PCV development (23).

Pertaining to the clinical testing of cell-free, neoepitope-based peptide vaccines, Keskin and colleagues recently reported a phase 1b PCV clinical trial in eight patients with glioblastoma, where specific $CD4^+$ T-cell responses to a mutation-bearing sequence from Rho GTPase Activating Protein 35 (ARHGAP35), which is naturally processed and presented to the immune system, were demonstrated in one patient (72). Furthermore, the authors also noted increased T-cell infiltration into the tumor – concomitant with neoantigen-specific T-cell in peripheral blood – following PCV administration. In a previous study, the authors had treated six patients with advanced melanoma and showed, that despite including HLA class I-binding peptides (for $CD8^+$ T-cell recognition) in the vaccine design, superior $CD4^+$ T-cell responses directed against the patients' neoantigens was observed (69). There is until this point not a convincing biological model for the observation that presumed MHC class I binding peptides (87), delivered as 9mers, induce rather $CD4^+$ T-cell responses as compared to $CD8^+$ T-cells; an observation that has been found to be true in several vaccination studies using tumor-associated mutant targets (69). In a different study, in patients with glioblastoma, $CD4^+$ T-cell responses were dominant in the case if mutant (nested) MHC class I-restricted epitopes were used for vaccination. None of the mutant epitopes elicited solely a $CD8^+$ T-cell response (although MHC class I epitope clusters were used), yet rather immune responses restricted by $CD4^+$ T-cells alone or by $CD4^+$ and $CD8^+$ T-cells were detected (88).

PCV-induced immune responses can also be enhanced with anti-PD-1 therapy (62, 69, 76). The TCR repertoire identified in clinically relevant and successful immune checkpoint therapy responses is associated with different anatomical compartments (89) as well as distinct T-cell markers in antigen-specific T-cells including $CD103^+$ T-cells (90) or – more recently with an stem-cell like $CD8^+CD69^-CD39^-$ phenotype in TIL that is strongly associated with clinical responses (90).

The rationale for increasing clinical responsiveness to PCV with checkpoint inhibitors would be to mobilize mutation-specific T-cells and $PD\text{-}1^+$ B-cell populations specific for cancer mutations (91–93). A different approach which may improve fine-tuning of anti-tumor responses following PCV treatment is the removal of non-productive inflammation caused by interleukin 6 (IL-6). Generally important for priming T-cell responses (94), IL-6 is a pleiotropic cytokine implicated in the pathogenesis of several cancer histologies, particularly gastrointestinal malignancies including pancreatic cancer and colorectal cancer (95, 96), one of the key factors being the suppression of productive immune responses in the TME (97–101). Treatments targeting IL-6 are approved for clinical use in patients with rheumatoid arthritis and Castleman's disease (102), although their use in patients with advanced cancer has yet to be fully realized despite promising results from preclinical models of solid tumors (94). Furthermore, IL-6 is among the cytokines released in large amounts following T-cell therapies for cancer (i.e., adoptive cell transfer ACT or CAR-T cell therapy) (103). IL-6 can lead to upregulation of PD-1 and immune exhaustion (104), while promoting interleukin 17 (IL-17) production – which can be a disadvantage in patients with cancer to subduing T-cell activity and augmenting tumor cell proliferation (105–107). Combined targeting of IL-6 and the PD-1/PD-L1 axis has shown reversal of immunosuppression in the TME leading to immune activation and tumor rejection in murine models of human cancer (108, 109). Other approaches in increasing the amenability of the TME to therapeutic intervention are to target the extracellular matrix and tumor stroma which provide scaffold support for the cancer cells (110, 111) and TAMs which have a pro-tumor effect in the TME (112).

Recent peptide vaccine trials show the complex neoepitope selection process and validation process – and underline also the need for a more harmonized approach that will enable to compare results across different studies to gauge T-cell responses against the immunizing peptides. In a clinical study for patients with melanoma (69), WES of the tumor was conducted, validated by RNA-Seq and mutant (tumor)

peptides were selected based on the patient's HLA-A and -B alleles followed by production of long-peptides representing up to 20 neoepitopes per patient. The 'private vaccine' was administered with an TLR3 adjuvant (poly-ICLC). MHC class I binding was predicted *via* NetMHCpan v2.4 and neoepitopes were selected with a hierarchy of criteria: i) frameshift mutations where the algorithm predicted binding, ii) single nucleotide variants (SNVs) where the algorithm predicted binding due to the mutation being in an anchor residue, iii) SNVs where the algorithm predicted binding due to the mutation being in residues other than anchor residues, iv) frameshift mutations where the algorithm did not predict binding and v) SNVs where the algorithm predicted low binding. In addition to the criteria listed above, oncogenes were prioritized and biochemical constraints concerning peptide synthesis were considered. 'Long peptides' allow for antigen processing and presentation for both $CD4^+$ and $CD8^+$ T-cells. The immunological readout addressed biological and clinically relevant questions like the frequency of peptide-specific T-cells (*in vitro*) upon re-stimulation assays using peripheral blood mononuclear cell (PBMCs) as immune effector cells showing that T-cell responses could indeed be induced against each individual vaccine target antigen. This point has practical implications: target antigen peptides were screened for T-cell reactivity and the biological readout is usually IFN-γ production. Non detectable IFN-γ production could imply that – assuming that the candidate peptide shows MHC binding – the frequency of T-cells directed against the candidate epitope is either low or that there are no peptide specific TCRs available in an individual patient. Low T-cell frequencies specific for a candidate peptide implies that these T-cells have not yet been expanded *in vivo*. Such a candidate peptide may represent a viable neoepitope for vaccination or T-cell expansion particularly if it is able to recruit T-cells from a stem cell pool with promising expansion potential and anti-cancer directed immune effector functions (90). Different peptide testing formats were used to gauge for the immune T-cell reactivity readout: a) peptides, b) minigenes (that allow the use of surrogate antigen presenting cells in order to test whether the peptides are naturally processed and presented, assuming that these minigenes are similarly processed as compared to the wildtype target tumor antigen), c) autologous tumor cells in order to test whether tumor cells are recognized by peptide-expanded T-cells since antigen processing and presentation may be different in tumor cells as compared to non-transformed cells as reviewed in Vigneron (113). This example shows the critical steps in the workflow and decision making process for which antigens should be selected (oncogenes, frameshift mutations), the format for vaccination (long peptides), the nature of the adjuvant and the question whether candidate peptide-reactive T-cells recognize naturally processed epitopes, the nature of the immune response, i.e., cytokine production (using intracellular cytokine staining to gauge for polyfunctional T-cells), a CD107a induction assay (to gauge for cytotoxicity), as well as a direct enumeration of MHC-specific T-cells using soluble MHC-peptide complexes. The list of different assays above reflects that peptide recognition may be tested positive in a specific biological readout (e.g. cytokine production), but not in another (e.g. cytotoxicity). Of practical interest is that MHC-class II peptide-tetramer guided enumeration often underestimates antigen-reactive $CD4^+$ T-cell numbers since the MHC-peptide interaction is fixed. In contrast, the incubation time of candidate test peptides with PBMCs is usually a few hours (and takes place at a different temperature) – allowing to accommodate a more 'promiscuous' binding of peptide species to several MHC class II alleles. Analysis of peptide-reactive T-cells with soluble MHC-peptide complexes provides unbiased enumeration of MHC-peptide reactive T-cells since it enables *ex vivo* analysis without the need of *in vivo* T-cell expansion, it enables – *via* the co-staining of T-cell differentiation and activation markers (that define in which T-cell compartment the antigen-specific T-cells reside) – to link the *ex vivo* analysis of antigen-specific T-cells with T-cell homing, differentiation, maturation or functional markers associated with cytokine production. This is clinically relevant since tumor-reactive T-cells that – upon adoptive transfer – lead to clinically relevant response reside preferentially in the central memory T-cell subset and/or exhibit distinct activation ($CD69^-$ $CD39^-$) phenotypes (90). Examples of two vaccination trials with peptides identified from glioblastomas addressed different, clinically relevant points, namely whether the presence of (candidate) antigen specific T-cells prior to vaccination would predict successful vaccination outcomes and whether the nature (mutant versus non-mutant targets) would make a difference in regard to T-cells expansion (88). This first, rather complex, clinical trial, was conducted using an 'off-the-shelf' cocktail of non-mutant peptides of glioblastoma-associated antigens targeting HLA-A2 and HLA-A24-positive patients, plus candidate 'private peptides', either from mutant or non-mutant 'private' glioblastoma targets. Key observations were that i) binding of some candidate peptides to MHC molecules was confirmed by mass spectrometry, i.e., these peptides were found to be naturally processed and presented, ii) mass spectrometry allowed an unbiased analysis of the peptide repertoire displayed by cancer cells, within the detection limits, yet requires approximately $10e^7$ tumor cells for analysis (114), iii) MHC-class I restricted $CD8^+$ T-cell responses, usually residing in precursor T-cells, to non-mutant epitopes prior to vaccination predicted successful T-cell responses after vaccination, iv) some peptide vaccine-induced T-cells recognized naturally processed and presented epitopes on the patients' autologous tumor cells, v) vaccination with $CD8^+$ T-cell epitopes induced $CD4^+$ T-cell responses, iv) some mutant vaccine epitope resulted in T-cells recognizing wildtype and mutant target antigens, v) none of the mutant epitopes evoked an exclusive $CD8^+$ T-cell response, but rather $CD4^+$, or T-cell responses in $CD4^+$ and in $CD8^+$ T-cells, vi) T-cells expanded from glioblastoma tissue harvested prior to vaccination did not contain T-cells responding to the selected candidate target epitopes, vii) no preferential expansion of T-cells using mutant epitopes as compared to non-mutant epitopes. A different clinical trial, also in patients with glioblastoma, showed that T-cells induced by vaccination infiltrated into tumor lesions after peptide vaccination (115).

This demonstration is clinically very relevant since only a few studies were able to demonstrate that T-cell clones elicited by peptide vaccination – reacting against the immunizing peptide – would then hone to the patient's tumor and aid to mediate tumor regression. This argues that it is necessary to obtain biopsies in progressing and regressing tumor lesions from patients with cancer who undergo peptide-based vaccination. Of particularly clinical relevance is i) that patients who received corticosteroids (which most patients with glioblastoma receive) during vaccine-priming failed to generate IFN-γ production to vaccine peptides, ii) some vaccine-peptides induced T-cells reacting against minigenes (coding for these targets, expressed as transgenes into surrogate recipient targets cells), but not to tumor cells, iii) that a round of *in vitro* stimulation was needed in order to detect antigen-specific T-cells, reflecting most likely low antigen-specific T-cell frequency, iv) peptide antigen-driven expansion *in vitro* and subsequent single cell PCR sequencing allowed to link TCR usage to peptide specificity. The identification of the unique peptide-specific TCR CDR3 motif allowed to 'trace back' the antigen specific T-cells to time points prior to vaccination (and post-vaccination samples) to the tumor sample used to identify the private mutations, as well as to post-vaccination tumor samples in case of tumor recurrence. Some mutant peptide specific TCRs were not detected prior to vaccination in PBMCs (which represent only 2% of the entire lymphocyte pool), nor have they been found in the tumor specimen used for mutational analysis, yet they were detectable in the tumor recurrence, an observation that was also observed in a rather more anatomically accessible basal cell cancer study (116). Such antigen-specific T-cells are mediating anti-tumor responses and their detection allow therefore a biologically relevant clue how neoepitope-specific T-cell therapies could be improved: Anti-cancer-directed T-cells after checkpoint inhibitor therapy were not 'rescued' or epigenetically rewired in response to checkpoint inhibitor therapies (117), yet rather 'new' T-cells were able to access the tumor site upon checkpoint inhibitor treatment. This phenomenon was dubbed 'clonal replacement' and would also support the notion that peptide-induced T-cells are able to access cancer lesions after vaccination (116). These observations are reminiscent of anti-cancer directed vaccine trials almost 2 decades ago. Although vaccination with (non-mutant) tumor associated antigens resulted in clinically relevant responses, the regressing tumor lesions showed 'spontaneous' anti-cancer immune reactivity, yet anti-cancer vaccine responses were not detected in the regressing cancer lesions suggesting that tumor vaccination aids to reinvigorate immune responses to private cancer antigens – and that a competent TCR is a prerequisite to achieve clinically meaningful T-cell responses (118). The practical consequence of these observations is to perform 2 or 4 mm needle biopsies in accessible tumor lesions that would allow to gauge for deep TCR sequence analysis and to trace mutant-epitope specific T-cell clones. In addition, both studies targeting glioblastoma epitopes showed that mutant peptide epitopes favored expansion of cytotoxic $CD4^+$ T-cells (115) and even if peptides were used to target $CD8^+$ T-cells, peptide antigen-specific $CD4^+$ T-cell expansion was observed in both studies (88). These and other clinical trial data were recently excellently reviewed addressing the clinical utility of neoantigen identification, peptide processing and MHC presentation of candidate epitope targets for rational vaccine design (119).

IMMUNE FUNCTION AND PERSONALIZED IMMUNOTHERAPY

Personalized immunotherapy is based on the capacity of the immune system to recognize, to be activated, to clonally expand and ultimately to kill off or to contain cancer cells. This involves several biological pathways, including the recognition and response to danger-associated molecular patterns (DAMPs) by cognate receptors present on the surface of APCs, T-cells as well as parenchymal cells (120). DAMPs are either released into the environment [e.g. high-mobility group box 1 (HMGB1), adenosine triphosphate, calreticulin; reviewed in (121)] or they can be cell-bound (Fas ligand [FasL] (122), heat shock proteins (123), MHC class I polypeptide-related proteins A/B [MICA/B]) (124, 125), and upon encountering the suitable receptor, elicit a signaling program resulting in an pro-inflammatory response. Although DAMPs can lead to non-productive inflammation resulting in organ damage, they play nevertheless an essential role in promoting cancer-directed immune responses and form an integral component of personalized immunotherapy strategies (126). Equally important is the epigenetic regulation of DAMPs which aid to the successful orchestration of innate and adaptive immune responses in PCV trials and clinically relevant immune responses (127). The use of 'in-built', molecularly defined adjuvants such as non-coding RNA may also augment the capacity of immune cells to orchestrate a potent anti-tumor immune responses *in vivo*, e.g. by activating the RNA sensing molecule retinoic acid-inducible gene I (RIG-1) (128). Indeed, RIG-I and a related intracellular RNA-sensing molecular melanoma-differentiation factor 5 (MDA-5) have been discussed as potential players in amplifying anti-tumor immune responses following recognition of cancer-associated RNA structures (128, 129). Other players in immunosurveillance are the toll-like receptors (TLR) 3 and 7, that are also involved in recognizing RNA derived from pathogens, with TLR3-mediated immune activation playing an essential role in the clinically relevant immunogenicity of poly-ICLC, a synthetic, double-stranded RNA-based polymer used as an adjuvant in the formulation of personalized cancer vaccines (69, 72, 76).

In a similar manner, different microRNA species are likely to be involved in immunomodulation and enhancement of local immune surveillance in cancer lesions (13, 130–132). Although the immunosuppressive TME may result in the downregulation of microRNA species, an unbiased identification of promising microRNAs and non-coding RNA sequences is possible *via* NGS and would allow to test synthetically produced RNA sequences as components in immuno-stimulation in PCV studies. MicroRNA species may also subdue expression of neoantigen-encoding genes, identifying microRNA using RNA-seq will therefore reveal additional layers of genetic regulation interfering with

optimal anti-tumor immune responses, yet it also opens new molecularly defined ways to optimize anti-cancer directed therapies in a more evidence-based fashion.

The stimulator of interferon genes (STING) pathway augments as well anti-tumor cellular immune responses (previously reviewed (133–135)), including potent B-cell activation and antibody production (136). STING is encoded by the TMEM173 gene in humans and acts as an intracellular DNA-sensing molecule (thus a pattern recognition receptor [PRR]) requiring cyclic guanosine monophosphate–adenosine monophosphate (GMP-AMP) synthase recognition of cytosolic DNA involved in expression of type I IFN-regulated pro-inflammatory genes. Activation of STING leads to transcription of IFN-stimulated response elements (ISREs) *via* TANK Binding Kinase 1 (TBK1) activation and interferon regulatory factor 3 (IRF3) localization into the nucleus to initiate gene transcription. Several clinical trials are underway, testing STING pathway agonists to induce anti-tumor immune responses in patients with cancer. A three-prime repair exonuclease 1 (TREX1) expression is involved in dampening STING-mediated immune activation by eliminating damaged DNA from the cytosol (135). However, it is possible that TREX1 mutations in patients with cancer may instead increase immune activation in cancer cell along with STING stimulation. Mutations in the STING pathway have been reported in patients with colorectal cancer, where STING-deficient cancer cells were unable to produce interleukin 1β (IL-1β) in response to DNA damage (137). Preclinical research showed that STING-dependent immune activation was able to enhance neoantigen vaccine responses and changed favorably the TME immune profile (138–140). Thus, primary and secondary immunodeficiencies defined by NGS can be identified molecularly and should supplement the NGS information obtained from cancer cells. STING variants may be naturally occurring genetic aberrations (e.g. silencing mutations in TLRs or receptors recognizing DAMPs), be associated with infections (e.g. human immunodeficiency virus [HIV]) or with immunosuppressive therapies (e.g. solid organ/stem cell transplantation and therapies used for patients with autoimmune diseases (120, 141–144)). NGS readouts allow also to visualize the mutational status of the STING pathway – among other immune-activating genes – in validating neoantigen-directed immune responses when designing PCVs and transgenic TCR-based cancer treatments. A comprehensive panel of mutations and natural variants in key molecules orchestrating the quality and quantity of anti-cancer directed immune responses is screened within the cancer NGS analysis in the ImmunoSurgery unit at the Champalimaud Foundation (see below) in order to better define immunological landscape of local and systemic immune responses that may influence immunotherapeutic strategies.

ANTIGEN PROCESSING AND PRESENTATION MACHINERY MUTATIONS IN THE CONTEXT OF IMMUNOTHERAPY

The antigen presentation machinery (APM), mainly constituted by the HLA class I and class II antigen processing and presentation pathway, are central to immune recognition and immune surveillance. While the HLA class I pathway generally processes and presents endogenous antigens derived from intracellular pathogens or autoantigens (such as neoantigens in cancer), the HLA class II pathway processes and presents exogenous antigens, which could be host- or pathogen-derived. $CD8^+$ T-cells are HLA class I restricted while $CD4^+$ T-cells recognize HLA class II epitopes (145). While all cells of the body (except erythrocytes) constitutively express the HLA class I pathway (except in the CNS, where MHC class I is downregulated); the HLA class II pathway can be activated in the presence of IFN-γ *via* transcription of the class II transactivator, thus underlining the need for IFN-γ in the TME (146). The standard APM in human cells is shown in **Figure 3**.

The antigen processing and presentation machinery is of major importance in immunosurveillance, as mutations occurring in the HLA class I and class II pathways bear great significance to cancer immunotherapy. Loss of MHC molecules may lead to immune-escape which may entail the failure of clinically relevant immune surveillance, loss of individual MHC class I loci in cancer lesions prohibits targeted therapy using PCV, since the identification of allelic losses limits naturally the choice of peptides to be used in a PCV. Thus, detailed analysis of the HLA haplotypes (HLA-A/B/C), is a prerequisite in selecting the 'best-fitting' neoepitopes in the design of personalized cancer vaccines as well as TCR-dependent T-cell therapies (28, 71, 147, 148). Downregulation of and loss-of-function mutations in the HLA class I and II pathways abrogate or dampen immune recognition of tumor cells *in vivo* (149–155), which goes along with TME evolution in response to immune activation (26, 156). Components of the HLA class I and class II pathways, if affected by genetic aberrations, may lead to 'tumor antigen loss variants' (thus the inability to process and present immunologically viable neoantigens) (157, 158). The 'hyper progression' effect described in patients treated with checkpoint inhibitors may, in fact, reflect an HLA loss *in vivo* while subtle MHC mutations may also have a similar deleterious effect on immune recognition by TIL if such mutations affect the nature and diversity of the peptide repertoire loaded onto the nominal MHC molecule. In accordance, mutations in the canonical HLA class I pathway (HLA-A/B/C) and its associated components in patients with cancer have been previously described, e.g. transporter associated with Antigen Processing 1/2 (TAP1/2), latent membrane protein 2/7 (LMP2/7) and β2-microglobulin (151, 159–164). Some tumor cells may also contain alternate splice forms of tapasin that can alter the repertoire of peptides loaded into the MHC class I antigen presentation pathway (165), mutations in the MHC class II antigen processing and presentation pathway have also been described (166, 167), yet they are not reported as frequently as molecular defects in the HLA class I pathway. There has been however much attention given to MHC class II expression in various cancer types such as colorectal (168), cervical (169), lung (170), breast (171), melanoma (172) and pancreatic cancer (173), pointing also to the significance of $CD4^+$-mediated anti-tumor responses (174, 175). Components associated with HLA class II peptide loading (the invariant chain [CLIP]), as well as the "peptide editors", HLA-DMA/DMB/DOA/DOB also play a

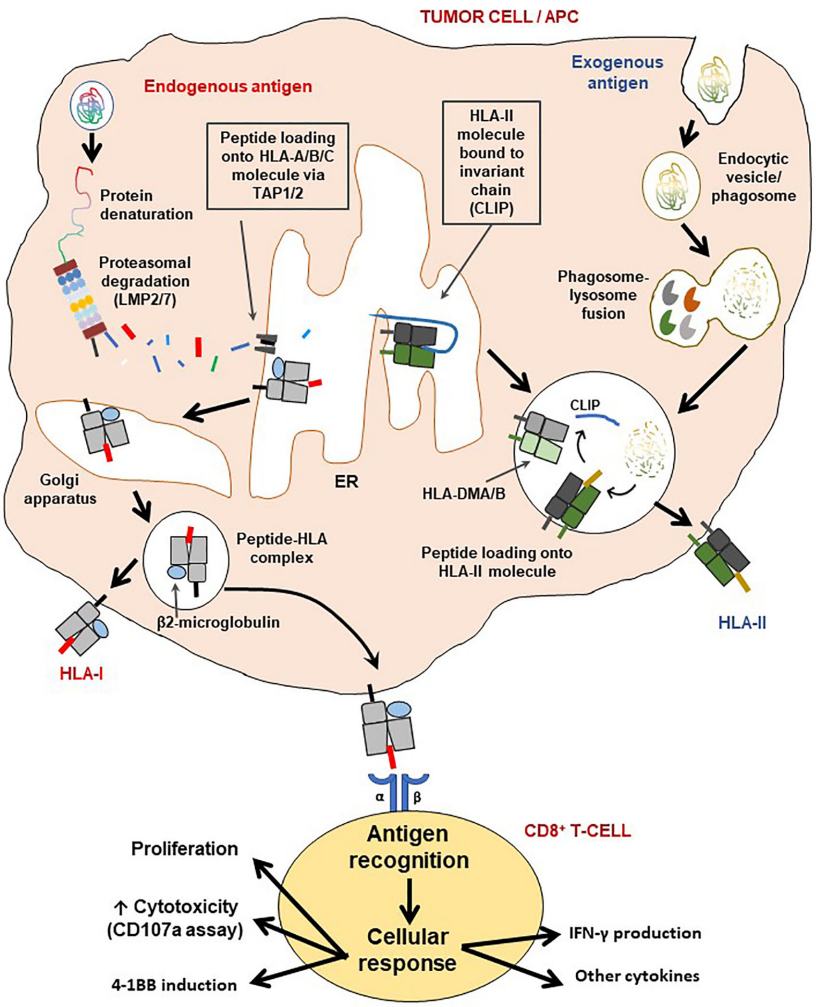


FIGURE 3 | Schematic representation of the HLA class I and II pathways and T-cell activation. The HLA class I pathway is also known as the intrinsic pathway as it processes and presents endogenous antigens while antigens derived from the extracellular environment are processed and presented *via* the HLA class II (extrinsic) pathway. LMP2/7 are immunoproteasome subunits necessary for generating short epitopes (7-11 amino acids along), which are then loaded on the HLA class I molecule for presentation to $CD8^+$ T-cells. The β2-microglobulin (β2M) is critical for the assembly and stable expression of HLA class I-peptide complexes on the cell surface. On the other hand, HLA class II molecules first exist with the class II-associated invariant chain (CLIP) for stability, which is then removed with assistance from the HLA-DMA/B complex, for loading of $CD4^+$ T-cell epitopes generated *via* lysosomal degradation. Processed antigens are then presented by either HLA-II (extrinsic pathway) or HLA-I (intrinsic pathway), to T-cells to initiate an immune synapse followed by activation of the latter. Indeed, as a result of cognate antigen recognition, T-cells may produce one or a combination of effects: i) cellular proliferation (also involves IL-2), ii) increase in cytotoxicity (may be measured by surface CD107a induction assay), iii) induction of 4-1BB expression and/or iv) production of cytokines, such as IFN-γ, TNF-α, IL-2, IL-17c.

significant role in producing meaningful $CD4^+$ T-cell responses (153, 168, 169) and are needed to present tumor antigen derived epitopes to $CD4^+$ T-cells either by cancer cells directly or APCs in the TME cross-presenting tumor-associated antigens. This is of clinical relevance since PCV enriched for $CD8^+$ T-cell epitopes tend to induce target-specific $CD4^+$ T-cell responses as discussed above. The loss of HLA class I expression and therefore subsequent $CD8^+$ T-cell responses in a patient with pancreatic cancer has been observed to be compensated with HLA class II-restricted $CD4^+$ T-cells (with cytokine production and cytolytic activity) (18) arguing that a molecular and immuno-histological examination of cancer lesions should include the MHC class I as well as the MHC class II antigen processing and presentation pathways that can be subject to therapeutic modulations, e.g. using HDAC inhibitors.

A ROLE FOR UNCONVENTIONAL NEOANTIGEN PRESENTATION IN CANCER?

Reduced expression of HLA-E (non-canonical HLA class I) has been linked to increased survival of patients with ovarian cancer (176), while it may also inhibit $CD8^+$ TIL activity. MICA and MICB

are expressed in the gastrointestinal epithelium, thus with relevance for metastatic GI cancers, e.g. colorectal or pancreatic cancer. Both MICA and MICB molecules are stress induced and bind to NKG2D in T-cell engagement, which could lead to activation of NK cells (or – not mutually exclusive – activation of TCR $\gamma\delta^+$ T-cells) instead of the TCR $\alpha\beta^+$ T-cells. HLA-G is yet another non-classical HLA class I member, whose expression is associated with a poor prognosis for patients with cancer, including patients with glioblastoma, colorectal and pancreatic cancer (177). Furthermore, HLA-G can be found in soluble form in blood while also secreted in exosomes; it can also be readily detected in IHC – thus making it feasible also for immunodiagnostics. Non-classical HLA molecules should also be considered in the development of anti-tumor directed vaccination – and future preclinical developments will target the identification of tumor-associated antigens presented by non-classical MHC-molecules to anti-cancer directed immune cells including the non-classical Major histocompatibility complex class I-related (MR1) molecule (178).

The presentation of non-peptide antigens by the cluster of differentiation 1 (CD1) family of molecules (a, b, c, d) – which is related to HLA class I – leads to activation of unconventional T-cell subsets, such as natural killer T-cells (NKT) (e.g. lipid antigens) and TCR γδ T-cells (e.g. phosphoantigens), with CD1d being the most prominent member, expressed on epithelial cells – and epithelial cancer cells. Expression of CD1 molecules in cancer are associated with different clinical outcomes while associated with poor prognosis in renal cell carcinoma (179), CD1d promotes NKT-mediated cytolysis of cancer cells in lung adenocarcinoma (180). A similar scenario exists for multiple myeloma and B-cell lymphoma, since CD1d is downregulated and associated with poor outcomes in contrast to higher CD1d expression levels in PBMCs from healthy individuals (181).

Nevertheless, neoantigen classes (and neoepitopes thereof) are limited to protein-based structures at this juncture due to their recognition by conventional HLA/TCR interactions involving $CD4^+$ and $CD8^+$ T-cells (13, 38, 182). More research is necessary to project a more concise picture of the role of non-peptide entities, e.g. lipids and carbohydrates bearing cancer-associated molecular abnormalities in augmenting productive immune responses in patients. For example, overexpressed or aberrantly glycosylated carbohydrates (e.g. gangliosides) is now hailed as a CAR-T target to treat pediatric patients with solid tumors (183). Also, the recognition of several pathogen-derived carbohydrate structures by conventional T-cells has been previously reviewed (184). The clinical studies associated with such therapeutic approaches could provide a template for precision oncology methods e.g. investigating which sugar moieties harboring abnormalities would be recognized by specific T-cell subsets using NGS and immunological assays.

ACCOUNTING FOR LYMPHOCYTE CLASSES IN PRECISION IMMUNOTHERAPY DESIGN

Multimodal studies have shown that the local immune landscape as well as neoantigen expression are quintessential parameters in determining and driving clinical responses in patients with cancer (13, 37, 185–188). Data from translational and clinical cancer immunotherapy studies collectively advocate for the development of 'composite lymphocyte grafts' comprising several immune-cell types interacting with a broad array of neoantigen profiles and subsequently a diverse set of effector functions with the unified aim to minimize disease progression, while eliminating existing cancer cell reservoirs in the patient (189, 190). Tumor infiltrating immune cells as well as tissue resident cells contribute to shape the immunological milieu, which is worthwhile to consider in precision immunotherapy protocols.

T-cells can be harvested and expanded for immunotherapy mainly from cancer tissue (TILs, inflamed tissue-derived cells), and/or PBMCs (11, 19, 44, 47, 191–193), cells from pleural effusions may also serve to isolate tumor-reactive T-cells (194) as a possible T-cell source, as well as immune cells from bronchoalveolar lavage (195), cerebrospinal fluid (196) or bone marrow aspirates (197). This biological material is a yet rather untapped source for future assessments in T-cell immunotherapy trials.

Not only the nature of the tissue specimen, yet also the different anatomical location of the T-cell harvest is critical if T-cells are tested for recognition of neoepitopes, exemplified in **Figure 4**. Not only the frequency of $CD4/CD8^+$ T-cells changes in relation from the tumor center to the tumor periphery, also the epitope target specificity changes, mutant KRAS reactive T-cells were detected in the tumor center, anti-mesothelin reactive T-cells were found in the tumor periphery. For clinical usage, it is important to emphasize that the location of the T-cell harvest has to be documented along with caution that different cancer tissue regions harbor different immune cell populations with different neoepitope specificities. While this is not surprising due to the tumor mutanome heterogeneity and consequent TCR diversity, it has to be taken into practical considerations as different areas of cancer specimens are harvested to expand TIL for cellular therapy. While TIL isolation and propagation for immunotherapy is feasible for some cancer types, patients with certain malignancies may present with cancer lesions that are – even with minimal invasive procedures or biopsies – very difficult to access (198). For those cases, PBMCs may be a viable and less invasive option, since peripheral blood T-cells are easily accessible and can be later used as a cell source for T-cell engineering to express a specific TCR or CAR (199, 200). HLA-matched donor-derived T-cells from donor PBMCs reactive to patient-derived neoantigens also present a viable option for neoepitope directed cellular immunotherapy (21).

There is also a different source of T-cells that can be considered for anti-cancer directed cellular therapy and for screening of neoepitope-reactive T-cell population, namely tissue-resident memory T-cells (TRM), a population of non-recirculating $CD8^+$ T-cells, residing long-term in peripheral tissues. TRMs contribute to tumor surveillance and to protection against viral and bacterial infections (201, 202), TRMs express a variety of homing markers that allow them to recirculate in peripheral tissues, such as CD103 (αe integrin) and CD49a (collagen-binding molecule antigen-1 (202–204), they

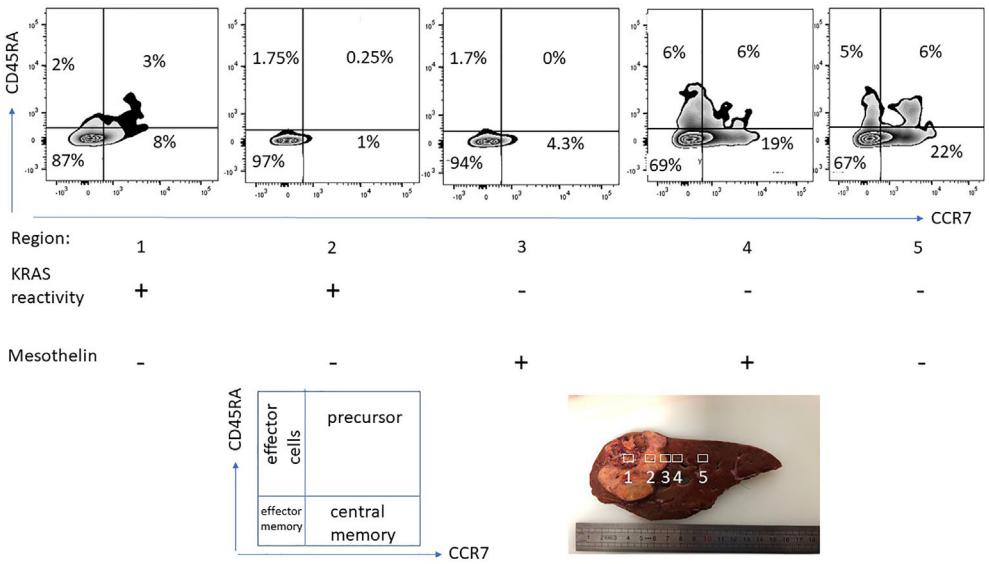


FIGURE 4 | T-cell phenotype and functional-spatial differences. TIL were expanded from different regions from a pancreas cancer lesion metastatic to the liver, 5 regions were harvested in different proximity to the tumor center. Note the different homing/maturation phenotype based on CD45RA/CCR7 expression, central memory T-cells in the tumor periphery. Thus, the quality of the T-cell response (to neoepitopes) is also associated with the immune cell maturation status. Reactivity to (mutant) KRAS or mesothelin was tested by pre-incubation of TIL for 5 days followed by IFN-gamma production analysis. Exclusive KRAS recognition in the tumor center versus mesothelin recognition in the tumor periphery and in macroscopically cancer-negative tissue demonstrating that the selection of neoepitope specific T-cells depends on the anatomical location.

produce Th1-type cytokines, namely IFN-γ, tumor necrosis factor alpha (TNF-α) and interleukin 2 (IL-2) upon stimulation (203, 205), yet may also elaborate Th17- or Th2-type cytokines (206–208). The impact of TRM cells in tumor surveillance is also related to the fact that TIL that express TRM cellular markers have been identified in several human solid cancers (209–215) often in correlation with improved clinical outcome (216–220). Therefore, the presence of homing markers and TRM cells among TIL, in addition to the tumor localization from where TIL are being expanded merit more attention in clinical studies pertaining to their role in neoantigen recognition, tumor surveillance and the selection of TIL for improved cellular immunotherapy.

CONVENTIONAL CD4$^+$ AND CD8$^+$ T-CELLS

T-cells bearing the conventional TCR αβ have been associated with augmenting clinical responses in patients receiving immunotherapy – both cell-based (191) and immune checkpoint inhibitors (221) and, more recently, personalized vaccines (78, 222). Both CD8$^+$ and CD4$^+$ neoantigen-specific TCR αβ responses in hard-to-treat cancers such as glioblastoma (14, 72), pancreatic malignancies (18), non-small-cell lung cancer (2, 3), melanoma (69, 76), bile duct (40) and colorectal cancers (41) are now regarded as vital to promote durable clinical responses in patients, further to the presence of suitable neoepitope restricting HLA elements (27, 223). Much focus has been placed on CD8$^+$ TILs in mediating anti-tumor activity due to their cytotoxic capacity and responsiveness to immune checkpoint blockade in view of their neoantigen-directed immune reactivity (224). In contrast, CD4$^+$ T-cells are largely attributed with helper functions, i.e., production of effectors cytokines, such as IFN-γ and TNF-α while responding to stimuli provided by IL-2, IL-6, interleukin 18 (IL-18), and IL-1β to list a few. Nevertheless, the cytotoxic activity of some tumor-directed CD4$^+$ T-cell subsets is now considered an important arm of the MHC class II restricted immune defense (174), particularly in patients with a defective HLA-I pathway. Further to autologous TCRs, allogeneic T-cells from HLA-matched healthy donors can either naturally react to – as part of the naturally occurring TCR repertoire – neoepitopes or they can be specifically selected and re-programmed to specifically react to neoepitopes and kill cancer cells without overt off-target toxicity (21, 225). Past and emerging studies consolidate the utility of conventional T-cells as sources of TCRs strongly reactive to peptide-HLA complexes based on the best-fitting epitopes to transduce PBMCs for developing possible 'off-the-shelf' TCRs options for patients with cancer expressing distinct tumor-associated antigens and sharing the respective restricting MHC allele (41, 226–230).

NON-CONVENTIONAL T-CELLS: TCR γδ AND INVARIANT NKT-CELLS

The relevance of TCR γδ T-cell subsets has received a substantial deal of attention in the last decade owing to clinically meaningful observations of anti-cancer reactivity in several cancer types (231–236). In patients with malignancies showing a defective

HLA system, TCR γδ T-cells may have the upper hand in immune recognition as they do not need the classical antigen presentation machinery for antigen recognition and activation (237). TCR γδ T-cells participate in a wide array of immunological processes which can activate or dampen the ensuing T-cell response, including the production of IL-17 which has been implicated in the pathogenesis of inflammatory disease as well as cancer (238–241). While several different gamma chains are known, two main delta chains have been described in humans TCR γδ T-cells, namely Vδ1 and Vδ2, although $Vδ3^+$ T-cells have been isolated from the liver (238, 242). As mentioned earlier, an important feature of TCR γδ cells is their expression of natural cytotoxicity receptors (NCRs), which are also present on NK-cells, namely NKG2D, NKp30 and DNAM-1 (243, 244). The most commonly occurring subclass of these cells are those expressing the Vγ9Vδ2 TCR which, *via* NKG2D, can bind to the HLA class I-like molecules MICA/B, akin to NK cells (244). Daley and colleagues recently showed that γδ T-cells outnumber $CD8^+$ T-cells in human pancreatic adenocarcinoma tissues, and potentially dampen the anti-tumor activity of conventional TCR αβ T-cells (245). Although anti-tumor γδ T-cell subsets expressing TCR Vγ9Vδ2 comprise a very small percentage in the tumor microenvironment, approximately 30% of circulating γδ T-cells expressed Vγ9. Apart from these, $Vδ1^+$ T-cell subsets have been shown to mediate productive immune responses against gastrointestinal tumors (246, 247), and are likely to be important players – in addition to the much studied Vγ9Vδ2 subset – in developing cellular immunotherapies for cancer (238).

In addition to TCR γδ T-cells, invariant natural killer (iNKT)–cells – bearing the invariant TCR V24α chain may also be relevant in recognizing neoepitopes in cancer (248). Alpha-galactosylceramide-driven activation of iNKT-cells (afore-mentioned CD1d-mediated antigen presentation) in patients with solid tumors has resulted in stable disease and detectable immune responses, including in protocols involving DCs pre-activated with alpha-galactosylceramide prior to infusion into patients (248–250). iNKT-cells can also exhibit cytolytic activity akin to NK-cells and cytotoxic $CD8^+$ T-cells, requiring the NCRs NKG2D and NKp44. Considering the characteristics of 'non-conventional' T-cells and clinical studies that support their individual use in immunotherapy (251), the combination of these immune cells with αβ T-cells should be considered in order to augment anti-cancer directed T-cell responses.

While isolation and cultivation of autologous T-cells from patients with cancer is a tailor-made drug development strategy, it is also time-consuming and can only cater for a limited number of patients at a time. Importantly, not all patients qualify for surgery and tumor biopsies are not always sufficient for TIL propagation after allocation for histopathological analysis. Therefore, TCRs from peripheral blood T-cells – recognizing shared or common cancer mutations – can be used to generate a cellular product generation with heterologous expression in patient T-cells. This approach has been shown to be successful in a patient with metastatic colorectal cancer who received an HLA-Cw08*02-restricted TCR targeting the $KRAS_{G12D}$ driver mutation (41). An integrated approach using NGS and immunology may be able to identify new neoantigens which are shared among certain patient groups to be adapted for developing transgenic TCR-based cellular drugs. The use of mucosal associated invariant T-cells (MAIT) and their respective targets for the potential use in personalized therapies is not discussed here.

B-CELLS

Unlike T-cells, B-cells have received the least attention although emerging evidence suggests that they should be accounted for in future treatment regimens due to their association with beneficial anti-tumor responses, including the production of cancer antigen-specific antibodies (252–254). TIB (tumor infiltrating B-cell) mediated responses, visualized by antibodies recognizing KRAS mutations, have been described in patients with pancreas adenocarcinoma (PDAC) (255), highlighting their clinical utility in anti-tumor immune responses. NGS platforms can supplement innovations in surgical oncology by the use of fluorescently-labelled antibodies and imaging to precisely mark the specific location of cancer disease for resection in the patient (256) coupled with *in vitro* laser microdissection of specific intra-tumoral regions of interest by identifying areas which are likely to represent varying mutational cancer profiles and matching T-cell reactivities (257). New data has revealed that an intact B-cell compartment in patients with advanced melanoma undergoing immune checkpoint blockade therapy (anti-PD-1, anti-CTLA-4 or both) are predictive of improved patient survival, given that no immune-related adverse events (irAEs) occur (258). This observation was associated with an increased proportion of circulating $CD21^+$ B-cells and plasmablasts after therapy. Also, the role of antibodies in the recognition of cancer-specific mutated proteins such as KRAS (255) as well as CMV- and EBV-derived epitopes in the TME (259) and carbohydrates (183) cannot be dismissed and warrant deeper insights in clinical studies examining the role of TIBs in immune-recognition and immunomodulation in the tumor. Antibodies binding to neoantigens of interest can be used for designing CARs, provided these neoantigens are expressed on the tumor-cell surface (e.g. MUC4), for which TCRs in blood were recently described (44). Waltari and colleagues recently showed how combining NGS and immunoassays platforms, while incorporating RNA-Seq and downstream bioinformatics analysis followed by *in vitro* stimulation with CpG and clonal expansion, can help identify and isolate memory B-cells from blood with B-cell receptors (BCRs) for a specific antigen in association with protection from disease which, in this case, was influenza (260). A more recent study reported the use of RNA-based Repertoire and Gene Expression by Sequencing (RAGE-Seq) that was able to identify BCR and TCR species circulating in blood of a patient with breast cancer that facilitated the tracking of lymphocyte populations in different tissue compartments (261). Thus, novel innovations in NGS techniques may also aid the discovery of distinct neoantigen- and tumor-reactive lymphocytes with clinical applicability. Neoepitope vaccination strategies may also induce mutation-specific antibodies in

antigen-driven expansion of B-cells that may also produce anti-cancer directed cytokines (262), functional TIB (263) that are associated with increased survival (264, 265) have been shown to produce antibodies that target TAAs, including mutant KRAS molecules (255). Thus, neoepitope-based vaccination immuno-monitoring should also include screening of vaccine peptide-specific humoral immune responses in the peripheral circulation as well as in TIL, even if the vaccine peptides are designed for MHC class I or -class II binding.

Laboratory-Based Platforms to Complement NGS and Facilitate Personalized Immuno-Oncology

A close collaboration with the pathology unit at healthcare facilities and allowing their active involvement in all phases of the clinical trial is crucial. Routine as well as specialized immunohistology panels can be designed to aim at analyzing HLA profiles in patients with cancer before, during and after immunotherapy. Furthermore, antibodies that can differentiate between misfolded and native HLA class I molecules on paraffinized tissue samples would be an immense advantage, since aberrant HLA class I molecules on the surface of tumor cells are likely to be missed by $CD8^+$ T-cells. Reagents that recognize all components of the HLA class I pathway such as TAP (151, 266), tapasin, β2-microglobulin-free HLA-A variants (151, 267, 268), LMP2/7 (151) and β2-microglobulin (151) have been described before, while those that recognize HLA class II components are also commercially available. In addition, immunostaining panels for histology, encompassing mutant epitopes of cellular proteins which can identify cancer cells and indicate whether druggable mutations are present in cancer tissue would be of great clinical value. Expression of the TNF-related apoptosis inducing ligand (TRAIL) molecules on the surface of cancer cell may also be a good indicator of their sensitivity to treatment-induced apoptosis (269). This approach may, in fact, serve as a means of 'companion diagnostics' to facilitate mutation-directed T-cell therapies. Circulating tumor cells (CTCs) that may be present in liquid biopsies can also be purified for immunocytochemistry (270–272). Two newly published reports describe how stable HLA molecules with an empty epitope-binding groove can be customized to bind peptides of interest and leveraged to screen for the best-fitting epitopes which induce an immune response (273, 274). Indeed, all of these methods could be exploited to screen for best-fitting neoepitopes using information arising from immunohistology and NGS data and obtain a better personalized anti-cancer vaccine and/or another treatment type, that may include pre-incubation of TILs with neoepitopes (to increase the frequency of TIL against mutant peptides). Multiparametric flow cytometry constitutes an integral part of screening for cellular immune responses and their physiological status is an addition to qualifying them for release as cellular products for personalized immunotherapy (55, 275). A wider panel of flow cytometry-compliant antibodies which can assess lymphocyte subsets present in cancer tissue based on phenotype and physiological status (e.g. exhausted vs. active, cytotoxic potential, type of recognition including MHC class I/II, CD1, MIC1A/B, MR1 restricted T-cells) prior to processing for cell culture would be an immense addition to clinical immunotherapy protocols (**Table 1**), augmenting findings from immunohistology analysis of tumor tissue. A dynamic set of flow cytometric analysis panels for further characterization of cellular products over the course of immune cell expansion for adoptive therapies will be of practical help, linked with immunohistology data from the resected tumor specimen. Ideally, functional T-cell data either from *ex vivo* expanded T-cells for adoptive therapy, or T-cell data obtained during immuno-monitoring in the context of peptide-based vaccination will yield extended information which can be amalgamated to the IHC data. Recent studies show that TP53 and KRAS mutations increase the expression of PD-L1 on tumor cells (276, 277), indicating that the presence of shared mutations can also be used as a companion diagnostic readout in personalized immunotherapy protocols.

PEPTIDE-HLA STABILITY ASSESSMENT

In addition to predicted HLA-matched neoepitopes, the use of an effective *in vitro* assay to measure the strength of peptide binding and stability may be able to improve the decision-making in selecting the most suitable neoantigens for personalized vaccine design. The measurement of the stability and half-life dynamics (also referred to as the 'off-rate'), which informs of how long a given peptide sequence can bind to the groove of the HLA molecule, was previously shown in the context of the TAA survivin (278), *Mycobacterium tuberculosis* protein TB10.4 (279, 280), HIV-1 epitopes (281), HA-1His autoantigen (282) and recently using HLA-B07*02-restricted myeloperoxidase (MPO) epitopes (226), allowing for the selection of strong binders capable of inducing optimal T-cell recognition and cytokine responses and/or cytotoxicity (283, 284). The half-life of the peptide-HLA complex class I/II constitutes an important parameter in dictating immunogenicity – the duration of time for which the peptide-HLA complex can be stably be expressed on the surface of the APC (including transformed cells) and evoke a strong $CD4^+$ or $CD8^+$ T-cell response (281, 283, 285–288). A recent study demonstrated that the stability of the peptide appears to be a better correlate of immunogenicity than affinity (287), and this may suggest that highly stable epitopes (those with a long half-life) can have a very strong affinity for their cognate HLA allele (283, 284, 289). The evaluation of peptide recognition defined by IFN-γ production – as a result from tumor mutanome analysis either in the format of synthetic peptides or as minigenes – can be used to gauge for biological activity in TIL as proficiency assay to gauge for T-cell reactivity against commonly strongly expressed mutant or non-mutant tumor associated antigens (290).

The absence of IFN-γ in the TIL culture supernatants from assays probing antigen reactivity does not necessarily imply that the predicted peptides do not induce any type of effector response by the candidate testing T-cells. For instance, matching TCR-HLA immune synapses may also lead to cytotoxicity instead of cytokine production, which can be

TABLE 1 | Lymphocyte markers for use in IHC and flow cytometry studies to support clinical decision making in personalized cancer immunotherapy.

Lymphocytes	Standard Analysis	Additional	Remarks
T-cells (TCR αβ, TCR γδ, NKT, MAIT cells)	CD3, CD4, CD8, CD25, TCR Vα/Vβ, TCR Vγ/Vδ, CD56, classical MAIT TCR Vα 7.2	NKG2D	Cytotoxic effector molecule (also applies to NK-cells)
		PD-1	Immune checkpoint molecules
		CTLA-4	
		LAG-3	
		TIM-3	
		IL-7R	IL-7 receptor/CD127; for Treg identification
		4-1BB	CD137; activation marker
		CD45RA	To assess the memory phenotype of T-cells
		CCR7	
		CXCR3	To assess the T-helper phenotype and tissue-penetration capacity of T-cells
		CCR4	
		CCR6	
		FoxP3	Transcription factor upregulated in activated T-cells and Tregs
		Helios	Aids in Treg identification
		Perforin	Cytolytic effector molecule
		Granzyme, Granylysin	Apoptosis-inducing effector molecule
		CD8+CD69-CD39-	CD8+ TIL with stem cell like properties and a CD69/CD39- phenotype are associated with response to therapy
		Cytokine receptors i.e., IL-6R, IL-1βR, IL-18R, IL-21R	For T-cell activation by APCs, and may help identify high-affinity antigen-specific cells
		BTN3A1/CD277	Antigen presentation to γδ T-cells
		IL-17	Can be useful as a marker for potentially pathogenic γδ T-cells
		Fas	Involved in apoptosis induction
		FasL	
B-cells (also act as APCs)	CD19, CD20	CD21	May have positive prognosis for patients with cancer
		FasL	Involved in apoptosis induction
		Fas	
		HLA class I pathway components	HLA alleles, TAP, tapasin, LMP2/7, β2M; to predict response to immunotherapy
		HLA class II pathway components	HLA-DR/DMA/DMB/DOA/DOB; to predict response to immunotherapy
		BTK	Bruton tyrosine kinase; may impede anti-tumor responses

determined by measuring surface CD107a induction or Fas expression. Production of granulocyte-macrophage colony-stimulating factor (GM-CSF) in lieu of cytotoxicity by $CD8^+$ T-cells has been shown in the context of HLA-A1-restricted melanoma epitopes (291). The absence of cytokine production may also stem from the fact that the predicted neoepitope is not naturally processed and presented to the immune system. Alternatively, some of the TCRs which may recognize predicted peptides are present in the general TCR repertoire of the patient but they may not be present in the tissue sample harvested to test for T-cell recognition of the predicted epitopes. For instance, TIL may represent a rather focused and antigen-specific enriched TCR repertoire (29, 292) and PBMCs from a standard blood draw represent usually 2% of the entire T-cell pool. One of the 'gold' standards in gauging anti-neoepitope-specific T-cell responses is certainly whether peptide-reactive T-cells – after sorting by tetramers, IFN-γ–capture or by using activation markers (e.g. CD137)-are able to recognize the patient's own tumor cells. This would strongly argue that epitope-specific T-cells recognize as well naturally processed and presented peptides – and that the selected candidate epitopes are biologically relevant. A different 'reversed' procedure is the repetitive stimulation with autologous tumor cells and the subsequently enriched T-cells are then tested for epitope specific reactivity (18).

'REAL WORLD' DATA: APPROACH AT THE CHAMPALIMAUD CLINICAL CENTRE

The points discussed above guide the immunotherapy program at the Champalimaud Foundation termed 'ImmunoSurgery' to underline i) the seamless connection with the surgical team and the subsequent examination of the resected tissue specimens from a clinical pathologist. The tissue used to produce TIL – and to perform tumor exome sequencing – requires documentation and histopathological information about the functional loco-regional diversity of the T-cell infiltrate into the tumor tissue which can be further assessed by deep TCR sequencing, ii) that T-cells expanded from surgical specimens are tested for neoepitope specificities as defined by tumor mutanome analysis and represent a 'biological knife'.

The workflow used for tissue procurement, neoepitope identification and T-cell screening is provided in **Figure 5**, factors that may impact on the nature of neoepitopes, neoepitope generation and factors shaping the responding T-cell repertoire are compiled in **Figure 6**. Tumor-epitope identification by WES and RNA-Seq is guided by careful selection of the tumor area for genetic analysis. A more recent excellent review addressed the clinical utility of neoantigen identification, peptide processing and MHC presentation of candidate epitope targets for rational vaccine design (119). Ideally, a tumor area that shows more than

80% of transformed cells is selected (**Figure 7**). In order to better define the tumor specimen, a standard analysis for the immune-contexture is carried out at our institution. A general (HLA-A, B and C) MHC class I loss would exclude patients from entering into peptide vaccination trials. $CD3^+$, $CD4^+$ and $CD8^+$ T-cells are being described along with the presence of $CD68^+$ macrophages, the presence of MHC class I antigens (HLA-A, B and C), the presence of HLA-DR, the expression of tumor-associated antigens (e.g. NY-ESO-1, mesothelin or survivin) to test for T-cell responses directed against non-mutant target antigens, as well as molecules associated with immune-suppression/evasion (e.g. PD-1, PDL-1 and CD47) along with the description if immune cells reside within the tumor or rather around, as single cells, or in clusters (**Figure 8**). The immune microenvironment imposes a strong pressure in untreated non-small-cell-lung cancers that subsequently show different routes of immune evasion. Different qualities of immune cell infiltration are associated with immune-editing (and therefore the diversity of neoepitopes available for T-cell expansion), MHC loss and defects in the antigen processing and presentation pathways (293). This may be differently associated with distinct tumor locations, ideally, parallel immunohistological sections are selected for WES and RNA-Seq. We combine different platforms to identify mutations in tumor exome data, combining the results of four different tools [Mutect2 (294), Varscan2 (295), Strelka2 (296) and Lancet (297)] and keep only mutations that are identified by at least two of these platforms. pVACtools takes results from the exome sequencing, complemented with mutations and fusions found in the transcriptomics data set which is then combined with several prediction algorithms resulting in a consensus ranking of neoantigens based on four criteria: rank of peptide binding affinity to the nominal MHC allele, rank of fold change between mutant and wild-type alleles, rank of mutant allele expression and the rank of DNA variant allele fraction (298). We test routinely two peptide formats to screen for cancer associated antigens in TIL or in PBMCs, namely i) peptides with 15 residues where the mutation is centered (and 7 amino flanking the mutation) or ii) the full downstream protein sequence in case of a frameshift. These different formats are used for immunoassays to gauge for INF-γ production in a 96-well format and supernatants are harvested at days 3 and 7 (see **Table 2** for references). Peptides

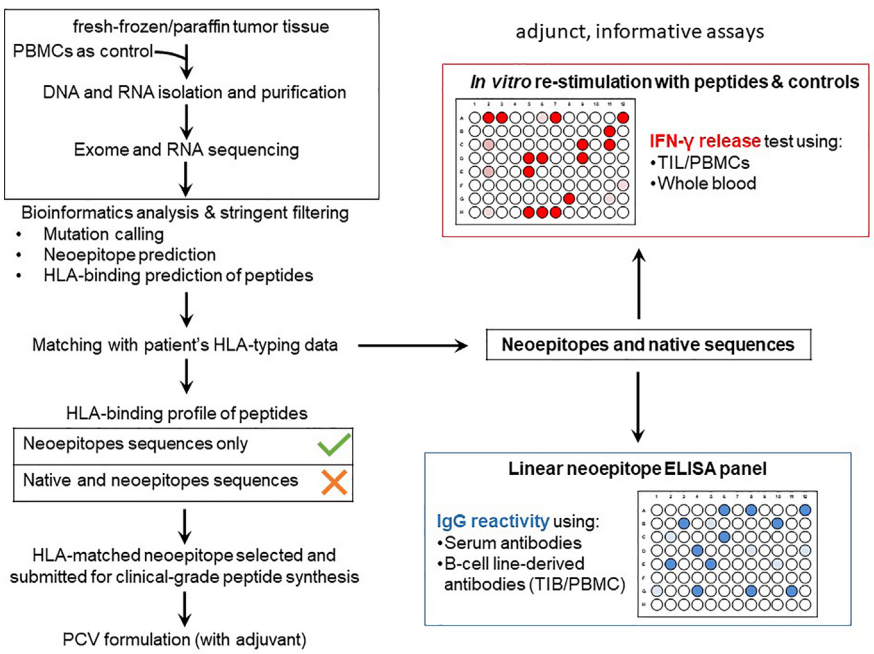


FIGURE 5 | PCV development and immuno-analyses workflow at the ImmunoSurgery Unit. Formalin-fixed paraffin-embedded (FFPE) or fresh-frozen tissue samples prepared by the Pathology Unit at the CCC is submitted for WES of tumor DNA with the patient's PBMCs as an internal control for downstream analysis, RNA-Seq is also sometimes performed to tumor RNA. The WES and RNA-Seq raw data is then analyzed at the ImmunoSurgery Unit at the CCU to predict private mutations followed by HLA class I and II binding prediction matched to the patients' HLA restriction profile to select candidates for inclusion in the PCV formulation. Only HLA-binding, neoepitope-containing peptides but not the wildtype counterparts are considered. The same and also 15-mer equivalent but non-clinical grade peptides, alongside the corresponding native sequences, are used for gauging TIL and/or PBMCs reactivities based on IFN-γ production (the peptide is at a concentration 1ug/mL tested with $10e^4$ responder T-cells; the fixed T-cell number allows to compare results obtained at different timepoints or from tissues harvested from different tumor areas). This part of the immunological evaluation of the neoepitopes is used in the follow-up phase of the trial which aims to assess T-cell responses of patients to the PCV and TIL therapies (possibly also for patients receiving immune checkpoint inhibitors). A different platform is an ELISA panel comprising the patient's neoepitopes in linear format to assess IgG reactivity using antibodies from serum as well as those secreted by TIB and PBMC-derived B-cell lines. Neoepitopes are also screened for TIL recognition since TIL are routinely generated to gauge for differences in TIL versus PBMC recognition. This will allow to describe whether selected neoepitopes are recognized in the tumor lesion that was used to identify the tumor neoepitopes (by NGS), it also allows to screen for differences in TIL recognition from tumor lesions harvested at different anatomical sites or at different timepoints in the course of the disease.

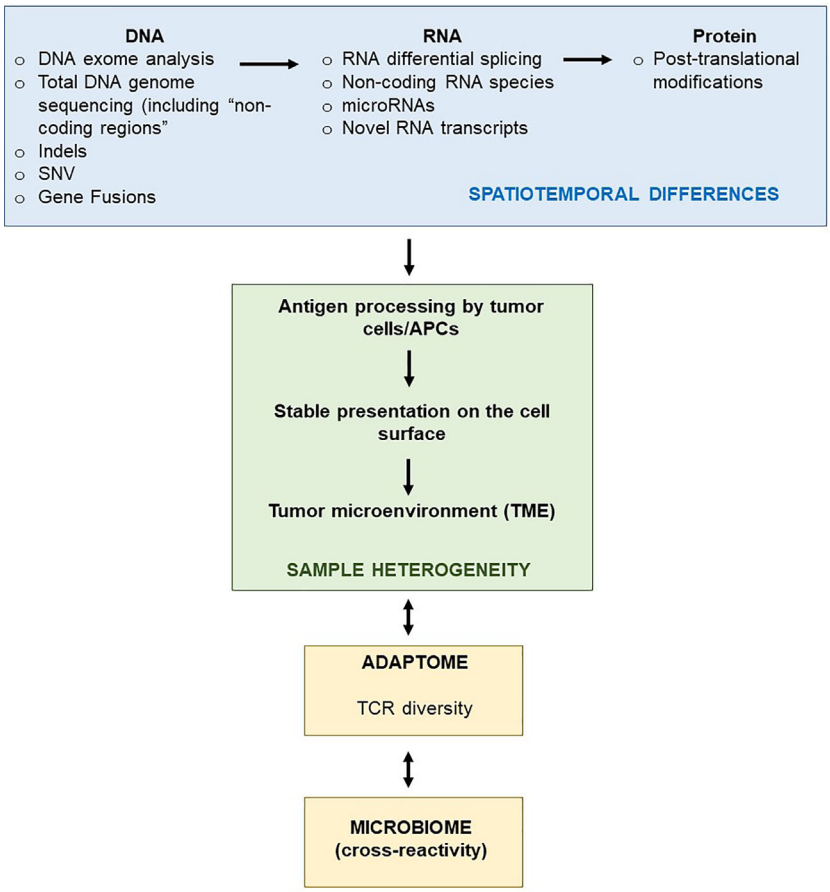


FIGURE 6 | Schematic representation of the general molecular paradigm of neoantigen recognition in the TME. The process of transcription of DNA to RNA and then to protein (antigens) is prone to generate heterogeneity in the context of cancer, i.e., the same DNA molecule may be differentially transcribed (due to RNA alternative splicing or mutations) and then translated to different proteins isotype (also as a result of post-translation modifications) or there might be gene fusions that result in novel RNA transcripts. The heterogeneous expression of tumor antigens, as a result of spatial-temporal differences in DNA to antigen production, results in different antigens being presented to the immune system by HLA complexes (as well as whole antigens) at the cell surface of a tumor cell and, therefore, contributing to different sub-regional TMEs within the same tumor tissue sample. These are likely to be neoantigens, as they are not present in healthy (non-transformed) tissue. The TCR diversity ("adaptome") will also change depending on the specific TME, i.e., different TCRs will be encountered depending on intratumoral spatial differences. Along the same lines, molecular structures associated with the microbiome present in the tumor tissue may cross-react with some T-cells, depending on the presence of absence of TCRs that recognize such microorganisms. The possible cross-reactivity, if present, may favor the expansion of the relevant immune-cell populations and, therefore, change the TCR repertoire.

of different lengths tailored according to the MHC typing of the patient are selected for candidates for PCVs based on i) if they are driver mutations, ii) strong expression in RNAseq, iii) superior binding of the mutant epitope as compared to the wildtype sequence, iv) frameshift mutations and v) practical considerations concerning peptide synthesis. If there are obvious different areas in the tumor specimen (**Figure 7**), micro-dissection of such areas is considered to estimate differences in tumor-heterogeneity. We are currently testing whether neoepitope directed T-cell responses are different in the primary cancer lesion as compared to metastatic lesions – that are usually harvested at a later time point during the cancer disease. Recognition analysis of MHC class I or -class II restricted epitopes defined by IFN-γ production in PBMCs versus TIL as a predictor of which neoepitopes are most likely immunogenic and also lead to clinically relevant responses in the course of a peptide-based vaccination strategy can only be tested in a phase I clinical trial that is currently being prepared.

In general, the resected cancer specimen is the result of already immuno-edited cancer cell clones, areas of potential neoantigen depletion and clinical tumor progress, despite the presence of immune infiltrates (24). The timing of cancer lesion harvest is also clinically relevant in the context for vaccination, if we presume that the landscape of tumor mutations within the same tumor lesion, and also at different spatiotemporal lesions, represents an active process between cancer evolution and the immune system. Not only tumor cells may be edited, also the available T-cell repertoire undergoes selective pressure. Timing of the tumor lesions for selection of vaccine epitopes determines the mutational burden, yet also the TCR repertoire that changes

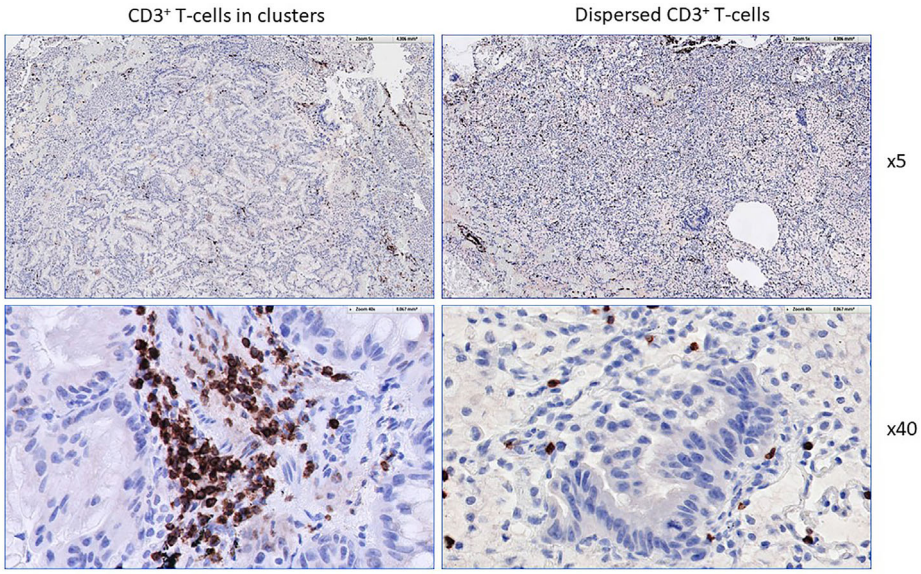


FIGURE 7 | Different immune-textures in cancer lesions. Starting point for WES and RNA-Seq. Definition and documentation of the immune cell infiltrate. Parallel slices of the paraffin-embedded tissues are procured and subjected to DNA and RNA analysis. Note the different patterns of $CD3^+$ T-cell clusters (left) versus individual $CD3^+$ T-cells in close proximity to tumor cells. RNA isolated from this tumor section would also allow for deep TCR-sequencing and allow to trace back individual TCR CDR3 motifs in case if neoepitope specific TCRs are identified.

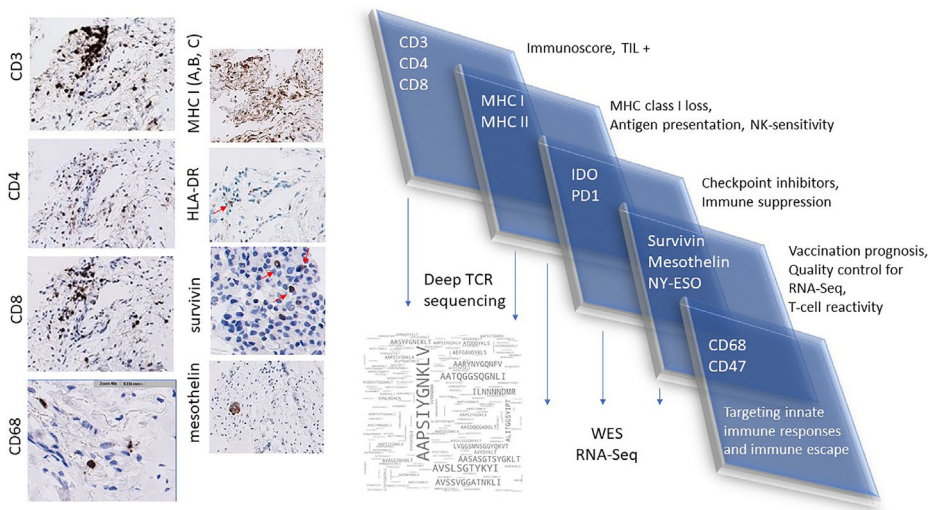


FIGURE 8 | Example of a standard immuno-histological analysis of a tumor sample at the Clinical Pathology Unit. Analysis of $CD3^+$, $CD4^+$ and $CD8^+$ T-cell infiltrates along with tumor-associated $CD68^+$ macrophages. Testing for MHC class I (HLA-A, B and C) expression to screen whether transformed calls can be recognized by $CD8^+$ T-cells, general MHC class I loss would not support vaccination strategies of adoptive T-cell therapy targeting TCR alpha/beta T-cells as the immune effector population. CD47, PD-1 and PDL-1 expression to gauge immune escape. Examination of commonly shared, non-mutant TAAs (NY-ESO-1, survivin, mesothelin) to identify T-cell responses in TIL and in corresponding PBMCs. Expression analysis of TAAs aids in quality control concerning RNA-Seq (of the corresponding gene coding for the TAA) and deep TCR analysis of T-cells reacting to TAAs.

over time (314). These very basic considerations bear very concrete consequences, i.e., usually tumor specimens are collected to choose mutant target epitopes in vaccine trials should be harvested for analysis after the most recent (standard) chemotherapy or immuno-therapy to reflect potential changes in the neoantigen landscape. Some of these practical considerations that are already currently discussed in clinical decision making or considered in future clinical trials are

TABLE 2 | Examples of molecular analysis guiding future therapeutic decision making.

Analysis	Examples of target genes	Potential biological and clinical effects	Potential practical consequences	Reference
WES or RNAseq Mutations in immunological response genes, i) e.g. induced by radiation, ii) germline mutations, iii) or natural variants that impact on immune function.	Immune responses genes in innate or adaptive immune responses including immune cell signaling, e.g. C2, CD163L1, FCγR2A	Gene variants or mutated genes edit immune infiltration, quality and quantity of the tumor-microenvironment	Despite identification of neoepitopes for neoepitope vaccination therapy plus checkpoint inhibitors, the innate or adaptive immune response may be blunted. The anti-cancer vaccination effect may not be achieved due to the incapacity to mount strong and cancer antigen specific immune response. Other therapeutic strategies are to be considered	(299, 300)
CHIP analysis, WES and RNAseq	Not only mutations in *bona fide* immune response genes, yet in cancer-associated genes, i.e. ARIDA, shape the quality of immune responses and T-cell infiltration	ARID1A aberrations may lead to differential chromatin accessibility and therefore to blunted anti-cancer directed immune responses, e.g. by reduction of overall IFN-gamma production, diminished immune cell infiltration and insufficient long- term immune memory responses.	Awareness that immunological treatment strategies may be challenging due to reduced IFN-gamma production. Detailed molecular analysis may aid to decipher how an effective anti-cancer directed milieu could be achieved without ARIDA1A interference	(301)
Deep TCR sequencing, TCRalpha, beta, gamma and delta chain. Bulk sequencing may be sufficient for most clinical questions. Single cell analysis possible.	Detailed molecular description of TCR infiltrate to objectively describe the situation prior to therapy. Different TCR repertoires in spatiotemporal cancer lesions.	A focused TCR repertoire can represent a relevant clonal immune response. Clonal immuno-editing may occur and lead to antigen – loss variants. 'Clonal replacement' appears to be associated with response to checkpoint inhibitors.	TCR convergence in PBMCs or tumor lesion (biopsies) and/or clonal convergence as companion diagnostics for immunological treatments. Knowledge of neoepitope specific TCR allows to follow antigen-specific reactivities. Broader TCR repertoire may provide more possibilities to react to neoepitopes imposed by the structural constraints of the MHC – peptide complexes. Long term neoepitope specific responses have been identified in patients with melanoma after peptide vaccination with different TCR clonotypes (directed against the identical epitope, this allows to link epitope-specific recognition with TCR diversity and functional avidity.	(116, 302, 303)
Epitope specific recognition defined by IFN or other cytokines in TIL from surgically removed tumor specimens and PBMCs	Either 'private mutations' or commonly shared tumor – associated antigens, i.e. NY-ESO-1, mesothelin, or common infectious pathogen antigens, e.g. EBV or CMV, provide a 'recognition fingerprint' to follow the immune response pattern in immunological therapies	Standard chemotherapy or immunological therapies shape the immune-competence to indicator targets (private antigens, TAAs or infectious disease targets).	Loss of anti-EBV or CMV recognition in peripheral blood, or anti-tumor antigen directed T-cell responses may represent one factor in the complex decision making choosing second or third line treatment therapies.	(259, 304–306)
Immuno-histology, RNA expression of commonly shared tumor associated antigens	NY-ESO-1, survivin or mesothelin expression	Commonly shared TAA-vaccines, e.g. anti-survivin, mesothelin or NY-ESO-1 are available. Anti-Mesothelin CARs or transgenic TCRs. MHC class I or class II-restricted NY-ESO-1 restricted TCRs are in clinical trials.	Strong antigenic heterogeneity in solid tumors defined by neoepitopes may still allow to use the immunogenic cancer – testis antigen NY-ESO-1 if sufficiently expressed. Mesothelin CARs have shown to be associated with epitope spreading and induce T-cell responses against private antigens. Commonly shared TAAs may represent a cellular 'first line' treatment, enhancement possible with checkpoint inhibitors.	(307–310)
WES and RNAseq bulk or single sequencing	Clonal spatiotemporal evolution in metastatic cancer lesions	'Immuno-edited' tumor clones may be eliminated during the course of the tumor disease while progressing tumor clones are 'Immune-privileged' despite the presence of tumor-infiltrating lymphocytes. Neoantigen depletion was observed in tumors with high Immunoscore and spatial proximity between tumor cells and T-cells.	'Immuno-edited' tumor lesions may still be accessible to commonly shared TAAs. If neoepitope-directed therapies are contemplated, a 'fresh' tumor biopsy after chemotherapy or immunotherapy is advisable due to the tumor evolution in order to obtain the most 'updated' antigenic profile.	(311)
WES in tumor versus metastasis	Mutanome in association with spatiotemporal differences.	Standard chemotherapy or immunotherapy may drive private mutations and clonal evolution: Treated metastases exhibit private 'driver'	Private mutations bear the risk of chemoresistance. Obtain clinical material from the most recent cancer lesions to assess spatiotemporal differences of	(312)

(Continued)

TABLE 2 | Continued

Analysis	Examples of target genes	Potential biological and clinical effects	Potential practical consequences	Reference
		mutations more frequently as compared to untreated metastases.	mutations in case if 'druggable' targets are considered or neoepitope-directed therapies.	
Immunological landscape analysis, RNAseq and/or immuno-histology	Cytokines, such as TGFbeta or IL-17.	TGFbeta may be strongly immuno-suppressive, promote desmoplastic changes in the TME that further inhibit anti-cancer immune responses, IL-17 may drive tumorigenesis.	A strong immuno-suppressive TME may counteract anti-cancer directed immunotherapies, e.g. neo-epitope-directed vaccination. Anti-TGFbeta directed therapies could be considered, either in the format of monoclonal antibodies or – in the case of active cellular therapies, gene-edited (TGFbeta-receptor) negative T-cells.	(313)

listed in **Table 2** where we also list the impact of tumor mutations or mutations in immune response genes on anti-cancer directed immune responses. The list of immune response genes that are particularly scrutinized and reported in the course of a standard WES is listed in the **Supplementary Data Set 1**. Although a high mutational burden is generally viewed as beneficial to elicit clinically relevant tumor responses, this may be less accentuated if the tumor lesion is very heterogenous (3), also reflected in the 'hot' and 'cold' areas in the same tumor specimen (293). Also, a 'low' mutational burden does not necessarily imply that neoepitopes within a tumor lesions are not able to elicit biologically and clinically relevant T-cell responses, as evidenced by glioblastoma-specific T-cell expansion, discussed above, and that clonally expanded T-cells, even in 'low-mutational burden' tumors (e.g. rhabdoid tumors) show tumor-specific T-cell responses (315). The concordant analysis of the tumor-associated neoepitope landscape does not only allow to link immuno-histological detection of T-cell responses with mutational events, it also enables the spatiotemporal analysis of the molecular composition of the T-cell repertoire with tumor mutations. The TCR landscape, defined by deep TCR sequencing allows the identification of motifs in TCRs infiltrating into tumor tissue as compared to non-transformed tissue (292). A more detailed discussion of this topic is not subject of this review. Yet it may allow to validate – although most likely not at this point in a routine fashion – whether MHC-peptide specific clonal TCRs are present within cancer or tissue lesions by modeling T-cell MHC epitope specificity (316) using yeast-display libraries of MHC-peptide complexes tested for TCR recognition as shown for TIL in colorectal adenocarcinoma (317). This has also been confirmed for the identification of pathogen-specific epitopes – starting with T-cell receptor sequences (318). Thus, *ex vivo* identification of mutant epitope targets may be validated by the identification of the nominal TCR ligands modeling the TCR and MHC-peptide interaction which is beyond the scope of this review.

CONCLUSIONS

Prediction of the best neoepitope candidates for immunotherapy is a multistep process combining several technology platforms ranging from NGS to histopathology and cellular assays. The specifics of the predicted neoepitopes, e.g. defined by the interaction with the nominal classical or non-classical MHC molecules, play a central role in developing clinical products in designing PCVs or in gauging TIL reactivities in association with the spatiotemporal diversity. New findings from translational and clinical research efforts would need to account for different genetic backgrounds and TCR diversities in order to objectively describe differences in immune cells capable of reacting to tumor-associated antigens with the goal to advance personalized cancer immunotherapy to expand potential treatment modalities for patients with cancer.

AUTHOR CONTRIBUTIONS

EdS co-wrote the first draft and created figures. JL performed experiments and edited text. AB performed immunohistology. GP performed experiments. CCon performed experiments and wrote passages concerning T-cell subsets. JK and PA contributed to peptide vaccine overviews and edited. NF provided surgical specimens. CCar edited and conceptualized. MC-M edited and performed immunohistology. ZW edited and commented on TCR analysis. DL edited and commented on the HLA and NK part. MR wrote the first draft and combined the overall efforts for the review. MM initiated the work, wrote the final draft, and conceptualized the entire program concerning epitope-based vaccination. All authors contributed to the article and approved the submitted version.

FUNDING

This work has been funded by the Champalimaud Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.592031/full#supplementary-material

REFERENCES

1. The Problem With Neoantigen Prediction. *Nat Biotechnol* (2017) 35:97. doi: 10.1038/nbt.3800
2. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer Immunology. Mutational Landscape Determines Sensitivity to PD-1 Blockade in Non-Small Cell Lung Cancer. *Science* (2015) 348:124–8. doi: 10.1126/science.aaa1348
3. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal Neoantigens Elicit T Cell Immunoreactivity and Sensitivity to Immune Checkpoint Blockade. *Science* (2016) 351:1463–9. doi: 10.1126/science.aaf1490
4. Maleki Vareki S. High and Low Mutational Burden Tumors Versus Immunologically Hot and Cold Tumors and Response to Immune Checkpoint Inhibitors. *J Immunother Cancer* (2018) 6:157. doi: 10.1186/s40425-018-0479-7
5. Smith KN, Llosa NJ, Cottrell TR, Siegel N, Fan H, Suri P, et al. Persistent Mutant Oncogene Specific T Cells in Two Patients Benefitting From Anti-PD-1. *J Immunother Cancer* (2019) 7:40. doi: 10.1186/s40425-018-0492-x
6. Gandara DR, Paul SM, Kowanetz M, Schleifman E, Zou W, Li Y, et al. Blood-Based Tumor Mutational Burden as a Predictor of Clinical Benefit in Non-Small-Cell Lung Cancer Patients Treated With Atezolizumab. *Nat Med* (2018) 24:1441–8. doi: 10.1038/s41591-018-0134-3
7. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, et al. Genomic Features of Response to Combination Immunotherapy in Patients With Advanced Non-Small-Cell Lung Cancer. *Cancer Cell* (2018) 33:843–852 e4. doi: 10.1016/j.ccell.2018.03.018
8. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of Mutational Processes in Human Cancer. *Nature* (2013) 500:415–21. doi: 10.1038/nature12477
9. Lu YC, Yao X, Crystal JS, Li YF, El-Gamil M, Gross C, et al. Efficient Identification of Mutated Cancer Antigens Recognized by T Cells Associated With Durable Tumor Regressions. *Clin Cancer Res* (2014) 20:3401–10. doi: 10.1158/1078-0432.CCR-14-0433
10. Andersen R, Donia M, Westergaard MC, Pedersen M, Hansen M, Svane IM. Tumor Infiltrating Lymphocyte Therapy for Ovarian Cancer and Renal Cell Carcinoma. *Hum Vaccin Immunother* (2015) 11:2790–5. doi: 10.1080/21645515.2015.1075106
11. Cohen CJ, Gartner JJ, Horovitz-Fried M, Shamalov K, Trebska-McGowan K, Bliskovsky VV, et al. Isolation of Neoantigen-Specific T Cells From Tumor and Peripheral Lymphocytes. *J Clin Invest* (2015) 125:3981–91. doi: 10.1172/JCI82416
12. Tanyi JL, Bobisse S, Ophir E, Tuyaerts S, Roberti A, Genolet R, et al. Personalized Cancer Vaccine Effectively Mobilizes Antitumor T Cell Immunity in Ovarian Cancer. *Sci Trans Med* (2018) 10:eaao5931. doi: 10.1126/scitranslmed.aao5931
13. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023
14. Valentini D, Rao M, Meng Q, von Landenberg A, Bartek J Jr., Sinclair G, et al. Identification of Neoepitopes Recognized by Tumor-Infiltrating Lymphocytes (Tils) From Patients With Glioma. *Oncotarget* (2018) 9:19469–80. doi: 10.18632/oncotarget.24955
15. Velcheti V, Kim ES, Mekhail T, Dakhil C, Stella PJ, Shen X, et al. Prospective Clinical Evaluation of Blood-Based Tumor Mutational Burden (Btmb) as a Predictive Biomarker for Atezolizumab (Atezo) in 1L Non-Small Cell Lung Cancer (NSCLC): Interim B-F1RST Results. *J Clin Oncol* (2018) 36:12001–1. doi: 10.1200/JCO.2018.36.15_suppl.12001
16. Yossef R, Tran E, Deniger DC, Gros A, Pasetto A, Parkhurst MR, et al. Enhanced Detection of Neoantigen-Reactive T Cells Targeting Unique and Shared Oncogenes for Personalized Cancer Immunotherapy. *JCI Insight* (2018) 3:122467. doi: 10.1172/jci.insight.122467
17. Zacharakis N, Chinnasamy H, Black M, Xu H, Lu YC, Zheng Z, et al. Immune Recognition of Somatic Mutations Leading to Complete Durable Regression in Metastatic Breast Cancer. *Nat Med* (2018) 24:724–30. doi: 10.1038/s41591-018-0040-8
18. Meng Q, Valentini D, Rao M, Moro CF, Paraschoudi G, Jager E, et al. Neoepitope Targets of Tumour-Infiltrating Lymphocytes From Patients With Pancreatic Cancer. *Br J Cancer* (2019) 120:97–108. doi: 10.1038/s41416-018-0262-z
19. Liu S, Matsuzaki J, Wei L, Tsuji T, Battaglia S, Hu Q, et al. Efficient Identification of Neoantigen-Specific T-Cell Responses in Advanced Human Ovarian Cancer. *J Immunother Cancer* (2019) 7:156. doi: 10.1186/s40425-019-0629-6
20. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R. The Challenges of Tumor Mutational Burden as An Immunotherapy Biomarker. *Cancer Cell* (2021) 39:154–73. doi: 10.1016/j.ccell.2020.10.001
21. Stronen E, Toebes M, Kelderman S, van Buuren MM, Yang W, van Rooij N, et al. Targeting of Cancer Neoantigens With Donor-Derived T Cell Receptor Repertoires. *Science* (2016) 352:1337–41. doi: 10.1126/science.aaf2288
22. Lee C-H, Yelensky R, Jooss K, Chan TA. Update on Tumor Neoantigens and Their Utility: Why It Is Good to Be Different. *Trends Immunol* (2018) 39:536–48. doi: 10.1016/j.it.2018.04.005
23. Hu Z, Ott PA, Wu CJ. Towards Personalized, Tumour-Specific, Therapeutic Vaccines for Cancer. *Nat Rev Immunol* (2018) 18:168–82. doi: 10.1038/nri.2017.131
24. Angelova M, Mlecnik B, Vasaturo A, Bindea G, Fredriksen T, Lafontaine L, et al. Evolution of Metastases in Space and Time Under Immune Selection. *Cell* (2018) 175:751–65.e16. doi: 10.1016/j.cell.2018.09.018
25. Kalaora S, Wolf Y, Feferman T, Barnea E, Greenstein E, Reshef D, et al. Combined Analysis of Antigen Presentation and T-Cell Recognition Reveals Restricted Immune Responses in Melanoma. *Cancer Discov* (2018) 8:1366–75. doi: 10.1158/2159-8290.CD-17-1418
26. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* (2016) 17:1206. doi: 10.1016/j.celrep.2016.10.009
27. Cai W, Zhou D, Wu W, Tan WL, Wang J, Zhou C, et al. MHC Class II Restricted Neoantigen Peptides Predicted by Clonal Mutation Analysis in Lung Adenocarcinoma Patients: Implications on Prognostic Immunological Biomarker and Vaccine Design. *BMC Genomics* (2018) 19:582–2. doi: 10.1186/s12864-018-4958-5
28. Teku GN, Vihinen M. Pan-Cancer Analysis of Neoepitopes. *Sci Rep* (2018) 8:12735. doi: 10.1038/s41598-018-30724-y
29. Li B, Li T, Pignon JC, Wang B, Wang J, Shukla SA, et al. Landscape of Tumor-Infiltrating T Cell Repertoire of Human Cancers. *Nat Genet* (2016) 48:725–32. doi: 10.1038/ng.3581
30. Enfield KSS, Martin SD, Marshall EA, Kung SHY, Gallagher P, Milne K, et al. Hyperspectral Cell Sociology Reveals Spatial Tumor-Immune Cell Interactions Associated With Lung Cancer Recurrence. *J ImmunoTher Cancer* (2019) 7:13. doi: 10.1186/s40425-018-0488-6
31. Cioffi M, Trabulo S, Hidalgo M, Costello E, Greenhalf W, Erkan M, et al. Inhibition of CD47 Effectively Targets Pancreatic Cancer Stem Cells *Via* Dual Mechanisms. *Clin Cancer Res* (2015) 21:2325–37. doi: 10.1158/1078-0432.CCR-14-1399
32. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-Associated Macrophages as Treatment Targets in Oncology. *Nat Rev Clin Oncol* (2017) 14:399–416. doi: 10.1038/nrclinonc.2016.217
33. Ring NG, Herndler-Brandstetter D, Weiskopf K, Shan L, Volkmer JP, George BM, et al. Anti-Sirpalpha Antibody Immunotherapy Enhances Neutrophil and Macrophage Antitumor Activity. *Proc Natl Acad Sci USA* (2017) 114:E10578–85. doi: 10.1073/pnas.1710877114
34. Schurch CM, Roelli MA, Forster S, Wasmer MH, Bruhl F, Maire RS, et al. Targeting CD47 in Anaplastic Thyroid Carcinoma Enhances Tumor Phagocytosis by Macrophages and is a Promising Therapeutic Strategy. *Thyroid* (2019) 29:979–92. doi: 10.1089/thy.2018.0555
35. Beatty GL, Chiorean EG, Fishman MP, Saboury B, Teitelbaum UR, Sun W, et al. CD40 Agonists Alter Tumor Stroma and Show Efficacy Against Pancreatic Carcinoma in Mice and Humans. *Science* (2011) 331:1612–6. doi: 10.1126/science.1198443
36. Beatty GL, Torigian DA, Chiorean EG, Saboury B, Brothers A, Alavi A, et al. A Phase I Study of an Agonist CD40 Monoclonal Antibody (CP-870,893) in Combination With Gemcitabine in Patients With Advanced Pancreatic Ductal Adenocarcinoma. *Clin Cancer Res* (2013) 19:6286–95. doi: 10.1158/1078-0432.CCR-13-1320
37. Anagnostou V, Smith KN, Forde PM, Niknafs N, Bhattacharya R, White J, et al. Evolution of Neoantigen Landscape During Immune Checkpoint

Blockade in Non-Small Cell Lung Cancer. *Cancer Discov* (2017) 7:264–76. doi: 10.1158/1538-7445.AM2017-NG01
38. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-Cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* (2017) 18:248–62. doi: 10.1016/j.celrep.2016.12.019
39. Luksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A Neoantigen Fitness Model Predicts Tumour Response to Checkpoint Blockade Immunotherapy. *Nature* (2017) 551:517–20. doi: 10.1038/nature24473
40. Tran E, Turcotte S, Gros A, Robbins PF, Lu YC, Dudley ME, et al. Cancer Immunotherapy Based on Mutation-Specific CD4+ T Cells in a Patient With Epithelial Cancer. *Science* (2014) 344:641–5. doi: 10.1126/science.1251102
41. Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N Engl J Med* (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
42. Laumont CM, Vincent K, Hesnard L, Audemard É., Bonneil É., Laverdure J-P, et al. Noncoding Regions Are the Main Source of Targetable Tumor-Specific Antigens. *Sci Trans Med* (2018) 10:eaau5516. doi: 10.1126/scitranslmed.aau5516
43. Yang W, Lee K-W, Srivastava RM, Kuo F, Krishna C, Chowell D, et al. Immunogenic Neoantigens Derived From Gene Fusions Stimulate T Cell Responses. *Nat Med* (2019) 25:767–75. doi: 10.1038/s41591-019-0434-2
44. Cafri G, Yossef R, Pasetto A, Deniger DC, Lu Y-C, Parkhurst M, et al. Memory T Cells Targeting Oncogenic Mutations Detected in Peripheral Blood of Epithelial Cancer Patients. *Nat Commun* (2019) 10:449. doi: 10.1038/s41467-019-08304-z
45. Chen F, Zou Z, Du J, Su S, Shao J, Meng F, et al. Neoantigen Identification Strategies Enable Personalized Immunotherapy in Refractory Solid Tumors. *J Clin Invest* (2019) 130(5):2056–2070. doi: 10.1172/JCI99538
46. Pasetto A, Gros A, Robbins PF, Deniger DC, Prickett TD, Matus-Nicodemos R, et al. Tumor- and Neoantigen-Reactive T-Cell Receptors Can Be Identified Based on Their Frequency in Fresh Tumor. *Cancer Immunol Res* (2016) 4:734–43. doi: 10.1158/2326-6066.CIR-16-0001
47. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective Identification of Neoantigen-Specific Lymphocytes in the Peripheral Blood of Melanoma Patients. *Nat Med* (2016) 22:433–8. doi: 10.1038/nm.4051
48. Lee M, Park C, Woo J, Kim J, Kho I, Nam D-H, et al. Preferential Infiltration of Unique Vγ9Jγ2-Vδ2 T Cells Into Glioblastoma Multiforme. *Front Immunol* (2019) 10:555. doi: 10.3389/fimmu.2019.00555
49. Aoki H, Ueha S, Shichino S, Ogiwara H, Hashimoto S-I, Kakimi K, et al. TCR Repertoire Analysis Reveals Mobilization of Novel CD8+ T Cell Clones Into the Cancer-Immunity Cycle Following Anti-CD4 Antibody Administration. *Front Immunol* (2019) 9:3185. doi: 10.3389/fimmu.2018.03185
50. Balachandran VP, Luksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of Unique Neoantigen Qualities in Long-Term Survivors of Pancreatic Cancer. *Nature* (2017) 551:512–6. doi: 10.1038/nature24462
51. Wang QJ, Yu Z, Griffith K, Hanada K, Restifo NP, Yang JC. Identification of T-Cell Receptors Targeting KRAS-Mutated Human Tumors. *Cancer Immunol Res* (2016) 4:204–14. doi: 10.1158/2326-6066.CIR-15-0188
52. Matsuda T, Leisegang M, Park J-H, Ren L, Kato T, Ikeda Y, et al. Induction of Neoantigen-Specific Cytotoxic T Cells and Construction of T-Cell Receptor–Engineered T Cells for Ovarian Cancer. *Clin Cancer Res* (2018) 24:5357–67. doi: 10.1158/1078-0432.CCR-18-0142
53. Peng S, Peng Z, Xu L, Mei J, Tang L, Han Y, et al. Abstract LB-076: Personal Neoantigen Immunotherapy in Hepatocellular Carcinoma. *Cancer Res* (2019) 79:LB–076-LB-076. doi: 10.1158/1538-7445.SABCS18-LB-076
54. Tran E, Robbins PF, Rosenberg SA. 'Final Common Pathway' of Human Cancer Immunotherapy: Targeting Random Somatic Mutations. *Nat Immunol* (2017) 18:255–62. doi: 10.1038/ni.3682
55. Rao M, Zhenjiang L, Meng Q, Sinclair G, Dodoo E, Maeurer M. Mutant Epitopes in Cancer. In: L Zitvogel and G Kroemer, editors. *Oncoimmunology: A Practical Guide for Cancer Immunotherapy*. Cham: Springer International Publishing (2018). p. 41–67.
56. Ahmadzadeh M, Pasetto A, Jia L, Deniger DC, Stevanović S, Robbins PF, et al. Tumor-Infiltrating Human CD4+ Regulatory T Cells Display a Distinct TCR Repertoire and Exhibit Tumor and Neoantigen Reactivity. *Sci Immunol* (2019) 4:eaao4310. doi: 10.1126/sciimmunol.aao4310
57. Robinson G, Gajjar A, Gauvain K, Basu E, Macy M, Maese L, et al. Phase 1/1B Trial to Assess the Activity of Entrectinib in Children and Adolescents With Recurrent or Refractory Solid Tumors Including Central Nervous System (CNS) Tumors. *Am Soc Clin Oncol Annu Meeting Chicago Illinois* (2019), 37:Abstract 10009. doi: 10.1200/JCO.2019.37.15_suppl.10009
58. Aldous AR, Dong JZ. Personalized Neoantigen Vaccines: A New Approach to Cancer Immunotherapy. *Bioorg Med Chem* (2018) 26:2842–9. doi: 10.1016/j.bmc.2017.10.021
59. Aurisicchio L, Pallocca M, Ciliberto G, Palombo F. The Perfect Personalized Cancer Therapy: Cancer Vaccines Against Neoantigens. *J Exp Clin Cancer Res* (2018) 37:86. doi: 10.1186/s13046-018-0751-1
60. Kenter GG, Welters MJP, Valentijn ARPM, Lowik MJG, Berends-van der Meer DMA, Vloon APG, et al. Vaccination Against HPV-16 Oncoproteins for Vulvar Intraepithelial Neoplasia. *N Engl J Med* (2009) 361:1838–47. doi: 10.1056/NEJMoa0810097
61. Welters MJ, van der Sluis TC, van Meir H, Loof NM, van Ham VJ, van Duikeren S, et al. Vaccination During Myeloid Cell Depletion by Cancer Chemotherapy Fosters Robust T Cell Responses. *Sci Transl Med* (2016) 8:334ra52. doi: 10.1126/scitranslmed.aad8307
62. Massarelli E, William W, Johnson F, Kies M, Ferrarotto R, Guo M, et al. Combining Immune Checkpoint Blockade and Tumor-Specific Vaccine for Patients With Incurable Human Papillomavirus 16–Related Cancer: A Phase 2 Clinical Trialimmune Checkpoint Blockade and Tumor-Specific Vaccine for Patients With Incurable HPV-16 Cancerimmune Checkpoint Blockade and Tumor-Specific Vaccine for Patients With Incurable HPV-16 Cancer. *JAMA Oncol* (2019) 5:67–73. doi: 10.1001/jamaoncol.2018.4051
63. Le DT, Brockstedt DG, Nir-Paz R, Hampl J, Mathur S, Nemunaitis J, et al. A Live-Attenuated Listeria Vaccine (ANZ-100) and a Live-Attenuated Listeria Vaccine Expressing Mesothelin (CRS-207) for Advanced Cancers: Phase I Studies of Safety and Immune Induction. *Clin Cancer Res* (2012) 18:858–68. doi: 10.1158/1078-0432.CCR-11-2121
64. Le DT, Wang-Gillam A, Picozzi V, Greten TF, Crocenzi T, Springett G, et al. Safety and Survival With GVAX Pancreas Prime and Listeria Monocytogenes-Expressing Mesothelin (CRS-207) Boost Vaccines for Metastatic Pancreatic Cancer. *J Clin Oncol* (2015) 33:1325–33. doi: 10.1200/JCO.2014.57.4244
65. Le DT, Ko AH, Wainberg ZA, Picozzi VJ, Kindler HL, Wang-Gillam A, et al. Results From a Phase 2b, Randomized, Multicenter Study of GVAX Pancreas and CRS-207 Compared to Chemotherapy in Adults With Previously-Treated Metastatic Pancreatic Adenocarcinoma (ECLIPSE Study). *J Clin Oncol* (2017) 35:345–5. doi: 10.1200/JCO.2017.35.4_suppl.345
66. Le DT, Picozzi VJ, Ko AH, Wainberg ZA, Kindler H, Wang-Gillam A, et al. A Randomized Phase 2b Study of GVAX Pancreas and CRS-207 Compared to Chemotherapy in Previously-Treated Metastatic Pancreatic Adenocarcinoma Patients (ECLIPSE Study). *Clin Cancer Res* (2019) 25 (18):5493–502. doi: 10.1158/1078-0432.CCR-18-2992
67. Ramlau R, Quoix E, Rolski J, Pless M, Lena H, Levy E, et al. A Phase II Study of Tg4010 (Mva-Muc1-Il2) in Association With Chemotherapy in Patients With Stage III/IV Non-Small Cell Lung Cancer. *J Thorac Oncol* (2008) 3:735–44. doi: 10.1097/JTO.0b013e31817c6b4f
68. Hui EP, Taylor GS, Jia H, Ma BB, Chan SL, Ho R, et al. Phase I Trial of Recombinant Modified Vaccinia Ankara Encoding Epstein-Barr Viral Tumor Antigens in Nasopharyngeal Carcinoma Patients. *Cancer Res* (2013) 73:1676–88. doi: 10.1158/0008-5472.CAN-12-2448
69. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An Immunogenic Personal Neoantigen Vaccine for Patients With Melanoma. *Nature* (2017) 547:217–21. doi: 10.1038/nature22991
70. Shindo Y, Hazama S, Suzuki N, Iguchi H, Uesugi K, Tanaka H, et al. Predictive Biomarkers for the Efficacy of Peptide Vaccine Treatment: Based on the Results of a Phase II Study on Advanced Pancreatic Cancer. *J Exp Clin Cancer Res* (2017) 36:36. doi: 10.1186/s13046-017-0509-1
71. Chen F, Zou Z, Du J, Su S, Shao J, Meng F, et al. Neoantigen Identification Strategies Enable Personalized Immunotherapy in Refractory Solid Tumors. *J Clin Invest* (2019) 129:2056–70. doi: 10.1172/JCI99538
72. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, et al. Neoantigen Vaccine Generates Intratumoral T Cell Responses in Phase Ib Glioblastoma Trial. *Nature* (2019) 565:234–9. doi: 10.1038/s41586-018-0792-9

73. Müller M, Gfeller D, Coukos G, Bassani-Sternberg M. 'Hotspots' of Antigen Presentation Revealed by Human Leukocyte Antigen Ligandomics for Neoantigen Prioritization. *Front Immunol* (2017) 8:1367–7. doi: 10.3389/fimmu.2017.01367
74. Bharathan M, Trebska-McGowan K, Anna P, Deniger DC, Hanada K-I, Gartner JJ, et al. Tetramer Based Approach for Efficient Identification and Isolation of Neo-Antigen Specific CD8 T Cells From Peripheral Blood (PBL) of Patients With Metastatic Cancers. *J Immunother Cancer* (2015) 3:P47–7. doi: 10.1186/2051-1426-3-S2-P47
75. Bezu L, Kepp O, Cerrato G, Pol J, Fucikova J, Spisek R, et al. Trial Watch: Peptide-Based Vaccines in Anticancer Therapy. *Oncoimmunology* (2018) 7: e1511506–e1511506. doi: 10.1080/2162402X.2018.1511506
76. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, et al. Personalized RNA Mutanome Vaccines Mobilize Poly-Specific Therapeutic Immunity Against Cancer. *Nature* (2017) 547:222–6. doi: 10.1038/nature23003
77. Wheeler CJ, Black KL. Dcvax ®-Brain and DC Vaccines in the Treatment of GBM. *Expert Opin Invest Drugs* (2009) 18:509–19. doi: 10.1517/13543780902841951
78. Guo Y, Lei K, Tang L. Neoantigen Vaccine Delivery for Personalized Anticancer Immunotherapy. *Front Immunol* (2018) 9:1499–9. doi: 10.3389/fimmu.2018.01499
79. Mastelic-Gavillet B, Balint K, Boudousquie C, Gannon PO, Kandalaft LE. Personalized Dendritic Cell Vaccines-Recent Breakthroughs and Encouraging Clinical Results. *Front Immunol* (2019) 10:766–6. doi: 10.3389/fimmu.2019.00766
80. Carreno BM, Magrini V, Becker-Hapak M, Kaabinejadian S, Hundal J, Petti AA, et al. A Dendritic Cell Vaccine Increases the Breadth and Diversity of Melanoma Neoantigen-Specific T Cells. *Science* (2015) 348:803–8. doi: 10.1126/science.aaa3828
81. Cormier JN, Salgaller ML, Prevette T, Barracchini KC, Rivoltini L, Restifo NP, et al. Enhancement of Cellular Immunity in Melanoma Patients Immunized With a Peptide From MART-1/Melan a. *Cancer J Sci Am* (1997) 3:37–44.
82. Wang F, Bade E, Kuniyoshi C, Spears L, Jeffery G, Marty V, et al. Phase I Trial of a MART-1 Peptide Vaccine With Incomplete Freund's Adjuvant for Resected High-Risk Melanoma. *Clin Cancer Res* (1999) 5:2756–65.
83. Yamshchikov GV, Barnd DL, Eastham S, Galavotti H, Patterson JW, Deacon DH, et al. Evaluation of Peptide Vaccine Immunogenicity in Draining Lymph Nodes and Peripheral Blood of Melanoma Patients. *Int J Cancer* (2001) 92:703–11. doi: 10.1002/1097-0215(20010601)92:5<703::AID-IJC1250>3.0.CO;2-5
84. Karbach J, Gnjatic S, Bender A, Neumann A, Weidmann E, Yuan J, et al. Tumor-Reactive CD8+ T-Cell Responses After Vaccination With NY-ESO-1 Peptide, Cpg 7909 and Montanide ISA-51: Association With Survival. *Int J Cancer* (2010) 126:909–18. doi: 10.1002/ijc.24850
85. Bigaeva E, Doorn EV, Liu H, Hak E. Meta-Analysis on Randomized Controlled Trials of Vaccines With QS-21 or ISCOMATRIX Adjuvant: Safety and Tolerability. *PLoS One* (2016) 11:e0154757–e0154757. doi: 10.1371/journal.pone.0154757
86. Zhu D, Tuo W. Qs-21: A Potent Vaccine Adjuvant. *Natural Products Chem Res* (2016) 3:e113. doi: 10.4172/2329-6836.1000e113
87. Wells DK, van Buuren MM, Dang KK, Hubbard-Lucey VM, Sheehan KCF, Campbell KM, et al. Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* (2020) 183:818–34.e13. doi: 10.1016/j.cell.2020.09.015
88. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanovic S, Gouttefangeas C, et al. Actively Personalized Vaccination Trial for Newly Diagnosed Glioblastoma. *Nature* (2019) 565:240–5. doi: 10.1038/s41586-018-0810-y
89. Kidman J, Principe N, Watson M, Lassmann T, Holt RA, Nowak AK, et al. Characteristics of TCR Repertoire Associated With Successful Immune Checkpoint Therapy Responses. *Front Immunol* (2020) 11:587014. doi: 10.3389/fimmu.2020.587014
90. Krishna S, Lowery FJ, Copeland AR, Bahadiroglu E, Mukherjee R, Jia L, et al. Stem-Like CD8 T Cells Mediate Response of Adoptive Cell Immunotherapy Against Human Cancer. *Science* (2020) 370:1328–34. doi: 10.1126/science.abb9847
91. Ren Z, Peng H, Fu Y-X. PD-1 Shapes B Cells as Evildoers in the Tumor Microenvironment. *Cancer Discov* (2016) 6:477–8. doi: 10.1158/2159-8290.CD-16-0307
92. Xiao X, Lao XM, Chen MM, Liu RX, Wei Y, Ouyang FZ, et al. PD-1hi Identifies a Novel Regulatory B-Cell Population in Human Hepatoma That Promotes Disease Progression. *Cancer Discov* (2016) 6:546–59. doi: 10.1158/2159-8290.CD-15-1408
93. Zhao Y, Shen M, Feng Y, He R, Xu X, Xie Y, et al. Regulatory B Cells Induced by Pancreatic Cancer Cell-Derived Interleukin-18 Promote Immune Tolerance Via the PD-1/PD-L1 Pathway. *Oncotarget* (2017) 9:14803–14. doi: 10.18632/oncotarget.22976
94. Kumari N, Dwarakanath BS, Das A, Bhatt AN. Role of Interleukin-6 in Cancer Progression and Therapeutic Resistance. *Tumour Biol* (2016) 37:11553–72. doi: 10.1007/s13277-016-5098-7
95. Vainer N, Dehlendorff C, Johansen JS. Systematic Literature Review of IL-6 as a Biomarker or Treatment Target in Patients With Gastric, Bile Duct, Pancreatic and Colorectal Cancer. *Oncotarget* (2018) 9:29820–41. doi: 10.18632/oncotarget.25661
96. Lippitz BE, Harris RA. Cytokine Patterns in Cancer Patients: A Review of the Correlation Between Interleukin 6 and Prognosis. *Oncoimmunology* (2016) 5:e1093722–e1093722. doi: 10.1080/2162402X.2015.1093722
97. Kristiansen OP, Mandrup-Poulsen T. Interleukin-6 and Diabetes: The Good, the Bad, or the Indifferent? *Diabetes* (2005) 54(Suppl 2):S114–24. doi: 10.2337/diabetes.54.suppl_2.S114
98. Ohmoto K, Yamamoto S. Serum Interleukin-6 and Interleukin-10 in Patients With Acute Pancreatitis: Clinical Implications. *Hepatogastroenterology* (2005) 52:990–4.
99. Zhang Y, Yan W, Collins MA, Bednar F, Rakshit S, Zetter BR, et al. Interleukin-6 is Required for Pancreatic Cancer Progression by Promoting MAPK Signaling Activation and Oxidative Stress Resistance. *Cancer Res* (2013) 73:6359–74. doi: 10.1158/0008-5472.CAN-13-1558-T
100. Holmer R, Goumas FA, Waetzig GH, Rose-John S, Kalthoff H. Interleukin-6: A Villain in the Drama of Pancreatic Cancer Development and Progression. *Hepatobiliary Pancreat Dis Int* (2014) 13:371–80. doi: 10.1016/S1499-3872(14)60259-9
101. Flint TR, Janowitz T, Connell CM, Roberts EW, Denton AE, Coll AP, et al. Tumor-Induced IL-6 Reprograms Host Metabolism to Suppress Anti-Tumor Immunity. *Cell Metab* (2016) 24:672–84. doi: 10.1016/j.cmet.2016.10.010
102. Rossi J-F, Lu Z-Y, Jourdan M, Klein B. Interleukin-6 as a Therapeutic Target. *Clin Cancer Res* (2015) 21:1248–57. doi: 10.1158/1078-0432.CCR-14-2291
103. Lee DW, Gardner R, Porter DL, Louis CU, Ahmed N, Jensen M, et al. Current Concepts in the Diagnosis and Management of Cytokine Release Syndrome. *Blood* (2014) 124:188–95. doi: 10.1182/blood-2014-05-552729
104. Kato T, Nishida T, Ito Y, Murase M, Murata M, Naoe T. Correlations of Programmed Death 1 Expression and Serum IL-6 Level With Exhaustion of Cytomegalovirus-Specific T Cells After Allogeneic Hematopoietic Stem Cell Transplantation. *Cell Immunol* (2014) 288:53–9. doi: 10.1016/j.cellimm.2014.02.007
105. Hou W, Jin YH, Kang HS, Kim BS. Interleukin-6 (IL-6) and IL-17 Synergistically Promote Viral Persistence by Inhibiting Cellular Apoptosis and Cytotoxic T Cell Function. *J Virol* (2014) 88:8479–89. doi: 10.1128/JVI.00724-14
106. Chen XW, Zhou SF. Inflammation, Cytokines, the IL-17/IL-6/STAT3/NF-Kappab Axis, and Tumorigenesis. *Drug Des Devel Ther* (2015) 9:2941–6. doi: 10.2147/DDDT.S86396
107. Chang SH. T Helper 17 (Th17) Cells and Interleukin-17 (IL-17) in Cancer. *Arch Pharm Res* (2019) 42(7):549–59. doi: 10.1007/s12272-019-01146-9
108. Tsukamoto H, Fujieda K, Miyashita A, Fukushima S, Ikeda T, Kubo Y, et al. Combined Blockade of IL6 and PD-1/PD-L1 Signaling Abrogates Mutual Regulation of Their Immunosuppressive Effects in the Tumor Microenvironment. *Cancer Res* (2018) 78:5011–22. doi: 10.1158/0008-5472.CAN-18-0118
109. Bialkowski L, Van der Jeught K, Bevers S, Tjok Joe P, Renmans D, Heirman C, et al. Immune Checkpoint Blockade Combined With IL-6 and TGF-Beta Inhibition Improves the Therapeutic Outcome of Mrna-Based Immunotherapy. *Int J Cancer* (2018) 143:686–98. doi: 10.1002/ijc.31331
110. Vennin C, Murphy KJ, Morton JP, Cox TR, Pajic M, Timpson P. Reshaping the Tumor Stroma for Treatment of Pancreatic Cancer. *Gastroenterology* (2018) 154:820–38. doi: 10.1053/j.gastro.2017.11.280
111. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR. Targeting Tumor Microenvironment for Cancer Therapy. *Int J Mol Sci* (2019) 20 (4):840. doi: 10.3390/ijms20040840

112. Poh AR, Ernst M. Targeting Macrophages in Cancer: From Bench to Bedside. *Front Oncol* (2018) 8:49. doi: 10.3389/fonc.2018.00049
113. Vigneron N. Human Tumor Antigens and Cancer Immunotherapy. *BioMed Res Int* (2015) 2015:948501. doi: 10.1155/2015/948501
114. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-Allelic Cells Enables More Accurate Epitope Prediction. *Immunity* (2017) 46:315–26. doi: 10.1016/j.immuni.2017.02.007
115. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, et al. Neoantigen Vaccine Generates Intratumoral T Cell Responses in Phase Ib Glioblastoma Trial. *Nature* (2019) 565:234–9. doi: 10.1038/s41586-018-0792-9
116. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal Replacement of Tumor-Specific T Cells Following PD-1 Blockade. *Nat Med* (2019) 25:1251–9. doi: 10.1038/s41591-019-0522-3
117. Philip M, Fairchild L, Sun L, Horste EL, Camara S, Shakiba M, et al. Chromatin States Define Tumour-Specific T Cell Dysfunction and Reprogramming. *Nature* (2017) 545:452–6. doi: 10.1038/nature22367
118. Germeau C, Ma W, Schiavetti F, Lurquin C, Henry E, Vigneron N, et al. High Frequency of Antitumor T Cells in the Blood of Melanoma Patients Before and After Vaccination With Tumor Antigens. *J Exp Med* (2005) 201:241–8. doi: 10.1084/jem.20041379
119. Richters MM, Xia H, Campbell KM, Gillanders WE, Griffith OL, Griffith M. Best Practices for Bioinformatic Characterization of Neoantigens for Clinical Utility. *Genome Med* (2019) 11:56. doi: 10.1186/s13073-019-0666-2
120. Hernandez C, Huebener P, Schwabe RF. Damage-Associated Molecular Patterns in Cancer: A Double-Edged Sword. *Oncogene* (2016) 35:5931–41. doi: 10.1038/onc.2016.104
121. Krysko O, Løve Aaes T, Bachert C, Vandenabeele P, Krysko DV. Many Faces of Damps in Cancer Therapy. *Cell Death Dis* (2013) 4:e631. doi: 10.1038/cddis.2013.156
122. Peter ME, Hadji A, Murmann AE, Brockway S, Putzbach W, Pattanayak A, et al. The Role of CD95 and CD95 Ligand in Cancer. *Cell Death Differ* (2015) 22:549. doi: 10.1038/cdd.2015.3
123. Chatterjee S, Burns TF. Targeting Heat Shock Proteins in Cancer: A Promising Therapeutic Approach. *Int J Mol Sci* (2017) 18:1978. doi: 10.3390/ijms18091978
124. Ghadially H, Brown L, Lloyd C, Lewis L, Lewis A, Dillon J, et al. MHC Class I Chain-Related Protein a and B (MICA and MICB) Are Predominantly Expressed Intracellularly in Tumour and Normal Tissue. *Br J Cancer* (2017) 116:1208–17. doi: 10.1038/bjc.2017.79
125. Zhao Y, Chen N, Yu Y, Zhou L, Niu C, Liu Y, et al. Prognostic Value of MICA/B in Cancers: A Systematic Review and Meta-Analysis. *Oncotarget* (2017) 8:96384–95. doi: 10.18632/oncotarget.21466
126. Li X. The Inducers of Immunogenic Cell Death for Tumor Immunotherapy. *Tumori* (2018) 104:1–8. doi: 10.5301/tj.5000675
127. Cruickshank B, Giacomantonio M, Marcato P, McFarland S, Pol J, Gujar S. Dying to Be Noticed: Epigenetic Regulation of Immunogenic Cell Death for Cancer Immunotherapy. *Front Immunol* (2018) 9:654. doi: 10.3389/fimmu.2018.00654
128. Ranoa DR, Parekh AD, Pitroda SP, Huang X, Darga T, Wong AC, et al. Cancer Therapies Activate RIG-I-Like Receptor Pathway Through Endogenous Non-Coding Rnas. *Oncotarget* (2016) 7:26496–515. doi: 10.18632/oncotarget.8420
129. Wu Y, Wu X, Wu L, Wang X, Liu Z. The Anticancer Functions of RIG-I-Like Receptors, RIG-I and MDA5, and Their Applications in Cancer Therapy. *Transl Res* (2017) 190:51–60. doi: 10.1016/j.trsl.2017.08.004
130. Schwickert A, Weghake E, Brüggemann K, Engbers A, Brinkmann BF, Kemper B, et al. MicroRNA miR-142-3p Inhibits Breast Cancer Cell Invasiveness by Synchronous Targeting of WASL, Integrin Alpha V, and Additional Cytoskeletal Elements. *PLoS One* (2015) 10:e0143993. doi: 10.1371/journal.pone.0143993
131. Zhang J, Zhang J, Zhang J, Qiu W, Xu S, Yu Q, et al. Microrna-625 Inhibits the Proliferation and Increases the Chemosensitivity of Glioma by Directly Targeting AKT2. *Am J Cancer Res* (2017) 7:1835–49.
132. Anandagoda N, Willis JCD, Hertweck A, Roberts LB, Jackson I, Gökmen MR, et al. MicroRNA-142–Mediated Repression of Phosphodiesterase 3B Critically Regulates Peripheral Immune Tolerance. *J Clin Invest* (2019) 129 (3):1257–71. doi: 10.1172/JCI124725
133. Corrales L, McWhirter SM, Dubensky TW Jr, Gajewski TF. The Host STING Pathway At the Interface of Cancer and Immunity. *J Clin Invest* (2016) 126:2404–11. doi: 10.1172/JCI86892
134. Zhou X, Jiang Z. STING-Mediated DNA Sensing in Cancer Immunotherapy. *Sci China Life Sci* (2017) 60:563–74. doi: 10.1007/s11427-016-9066-0
135. Barber GN. STING: Infection, Inflammation and Cancer. *Nat Rev Immunol* (2015) 15:760–70. doi: 10.1038/nri3921
136. Walker MM, Crute BW, Cambier JC, Getahun A. B Cell–Intrinsic STING Signaling Triggers Cell Activation, Synergizes With B Cell Receptor Signals, and Promotes Antibody Responses. *J Immunol* (2018) 201(9):2641–53. doi: 10.4049/jimmunol.1701405
137. Xia T, Konno H, Ahn J, Barber GN. Deregulation of STING Signaling in Colorectal Carcinoma Constrains DNA Damage Responses and Correlates With Tumorigenesis. *Cell Rep* (2016) 14:282–97. doi: 10.1016/j.celrep.2015.12.029
138. Luo M, Wang H, Wang Z, Cai H, Lu Z, Li Y, et al. A STING-Activating Nanovaccine for Cancer Immunotherapy. *Nat Nanotechnol* (2017) 12:648–54. doi: 10.1038/nnano.2017.52
139. Kinkead HL, Hopkins A, Lutz E, Wu AA, Yarchoan M, Cruz K, et al. Combining STING-Based Neoantigen-Targeted Vaccine With Checkpoint Modulators Enhances Antitumor Immunity in Murine Pancreatic Cancer. *JCI Insight* (2018) 3:e122857. doi: 10.1172/jci.insight.122857
140. Sivick KE, Desbien AL, Glickman LH, Reiner GL, Corrales L, Surh NH, et al. Magnitude of Therapeutic STING Activation Determines CD8+ T Cell-Mediated Anti-Tumor Immunity. *Cell Rep* (2018) 25:3074–85.e5. doi: 10.1016/j.celrep.2018.11.047
141. Satgé D. A Tumor Profile in Primary Immune Deficiencies Challenges the Cancer Immune Surveillance Concept. *Front Immunol* (2018) 9:1149–9. doi: 10.3389/fimmu.2018.01149
142. Haas OA. Primary Immunodeficiency and Cancer Predisposition Revisited: Embedding Two Closely Related Concepts Into an Integrative Conceptual Framework. *Front Immunol* (2019) 9:3136–6. doi: 10.3389/fimmu.2018.03136
143. de Jong D, Roemer MGM, Chan JKC, Goodlad J, Gratzinger D, Chadburn A, et al. B-Cell and Classical Hodgkin Lymphomas Associated With Immunodeficiency: 2015 SH/EAHP Workshop Report-Part 2. *Am J Clin Pathol* (2017) 147:153–70. doi: 10.1093/ajcp/aqw216
144. Sánchez-Ramón S, Bermúdez A, González-Granado LI, Rodríguez-Gallego C, Sastre A, Soler-Palacín P, et al. Primary and Secondary Immunodeficiency Diseases in Oncohaematology: Warning Signs, Diagnosis, and Management. *Front Immunol* (2019) 10:586. doi: 10.3389/fimmu.2019.00586
145. Abbas A, Lichtman A, Pilai S. *Basic Immunology: Functions and Disorders of the Immune System*. Philadelphia: Elsevier Saunders (2014).
146. Piskurich JF, Wang Y, Linhoff MW, White LC, Ting JP. Identification of Distinct Regions of 5' Flanking DNA That Mediate Constitutive, IFN-Gamma, STAT1, and TGF-Beta-Regulated Expression of the Class II Transactivator Gene. *J Immunol* (1998) 160:233–40.
147. Zhou J, Zhao W, Wu J, Lu J, Ding Y, Wu S, et al. Neoantigens Derived From Recurrently Mutated Genes as Potential Immunotherapy Targets for Gastric Cancer. *BioMed Res Int* (2019) 2019:11. doi: 10.1155/2019/8103142
148. Bräunlein E, Krackhardt AM. Identification and Characterization of Neoantigens as Well as Respective Immune Responses in Cancer Patients. *Front Immunol* (2017) 8:1702–2. doi: 10.3389/fimmu.2017.01702
149. Seliger B, Maeurer MJ, Ferrone S. TAP Off–Tumors on. *Immunol Today* (1997) 18:292–9. doi: 10.1016/S0167-5699(97)80026-6
150. Seliger B, Maeurer MJ, Ferrone S. Antigen-Processing Machinery Breakdown and Tumor Growth. *Immunol Today* (2000) 21:455–64. doi: 10.1016/S0167-5699(00)01692-3
151. Atkins D, Breuckmann A, Schmahl GE, Binner P, Ferrone S, Krummenauer F, et al. MHC Class I Antigen Processing Pathway Defects, Ras Mutations and Disease Stage in Colorectal Carcinoma. *Int J Cancer* (2004) 109:265–73. doi: 10.1002/ijc.11681
152. Maccalli C, Parmiani G, Ferrone S. Immunomodulating and Immunoresistance Properties of Cancer-Initiating Cells: Implications for the Clinical Success of Immunotherapy. *Immunol Invest* (2017) 46:221–38. doi: 10.1080/08820139.2017.1280051
153. Seliger B, Kloor M, Ferrone S. HLA Class II Antigen-Processing Pathway in Tumors: Molecular Defects and Clinical Relevance. *Oncoimmunology* (2017) 6:e1171447. doi: 10.1080/2162402X.2016.1171447

154. Zaretsky JM, Garcia-Diaz A, Shin DS, Escuin-Ordinas H, Hugo W, Hu-Lieskovan S, et al. Mutations Associated With Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* (2016) 375:819–29. doi: 10.1056/NEJMoa1604958
155. Haworth KB, Leddon JL, Chen CY, Horwitz EM, Mackall CL, Cripe TP. Going Back to Class I: MHC and Immunotherapies for Childhood Cancer. *Pediatr Blood Cancer* (2015) 62:571–6. doi: 10.1002/pbc.25359
156. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* (2017) 171:1259–71.e11. doi: 10.1016/j.cell.2017.10.001
157. Chowell D, Morris LGT, Grigg CM, Weber JK, Samstein RM, Makarov V, et al. Patient HLA Class I Genotype Influences Cancer Response to Checkpoint Blockade Immunotherapy. *Science* (2018) 359:582–7. doi: 10.1126/science.aao4572
158. Arakawa A, Vollmer S, Tietze J, Galinski A, Heppt MV, Burdek M, et al. Clonality of CD4(+) Blood T Cells Predicts Longer Survival With CTLA4 or PD-1 Checkpoint Inhibition in Advanced Melanoma. *Front Immunol* (2019) 10:1336. doi: 10.3389/fimmu.2019.01336
159. Koopman LA, van Der Slik AR, Giphart MJ, Fleuren GJ. Human Leukocyte Antigen Class I Gene Mutations in Cervical Cancer. *J Natl Cancer Inst* (1999) 91:1669–77. doi: 10.1093/jnci/91.19.1669
160. Leone P, Shin EC, Perosa F, Vacca A, Dammacco F, Racanelli V. MHC Class I Antigen Processing and Presenting Machinery: Organization, Function, and Defects in Tumor Cells. *J Natl Cancer Inst* (2013) 105:1172–87. doi: 10.1093/jnci/djt184
161. Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, et al. Somatic CALR Mutations in Myeloproliferative Neoplasms With Nonmutated JAK2. *N Engl J Med* (2013) 369:2391–405. doi: 10.1056/NEJMoa1312542
162. Li L, Dong M, Wang XG. The Implication and Significance of Beta 2 Microglobulin: A Conservative Multifunctional Regulator. *Chin Med J (Engl)* (2016) 129:448–55. doi: 10.4103/0366-6999.176084
163. Leney AC, Pashley CL, Scarff CA, Radford SE, Ashcroft AE. Insights Into the Role of the Beta-2 Microglobulin D-Strand in Amyloid Propensity Revealed by Mass Spectrometry. *Mol Biosyst* (2014) 10:412–20. doi: 10.1039/C3MB70420C
164. Valleix S, Gillmore JD, Bridoux F, Mangione PP, Dogan A, Nedelec B, et al. Hereditary Systemic Amyloidosis Due to Asp76Asn Variant Beta2-Microglobulin. *N Engl J Med* (2012) 366:2276–83. doi: 10.1056/NEJMoa1201356
165. Belicha-Villanueva A, Golding M, McEvoy S, Sarvaiya N, Cresswell P, Gollnick SO, et al. Identification of an Alternate Splice Form of Tapasin in Human Melanoma. *Hum Immunol* (2010) 71:1018–26. doi: 10.1016/j.humimm.2010.05.019
166. Michel S, Linnebacher M, Alcaniz J, Voss M, Wagner R, Dippold W, et al. Lack of HLA Class II Antigen Expression in Microsatellite Unstable Colorectal Carcinomas Is Caused by Mutations in HLA Class II Regulatory Genes. *Int J Cancer* (2010) 127:889–98. doi: 10.1002/ijc.25106
167. Surmann E-M, Voigt AY, Michel S, Bauer K, Reuschenbach M, Ferrone S, et al. Association of High CD4-Positive T Cell Infiltration With Mutations in HLA Class II-Regulatory Genes in Microsatellite-Unstable Colorectal Cancer. *Cancer Immunol Immunother* (2015) 64:357–66. doi: 10.1007/s00262-014-1638-4
168. Sconocchia G, Eppenberger-Castori S, Zlobec I, Karamitopoulou E, Arriga R, Coppola A, et al. HLA Class II Antigen Expression in Colorectal Carcinoma Tumors as a Favorable Prognostic Marker. *Neoplasia* (2014) 16:31–42. doi: 10.1593/neo.131568
169. Zehbe I, Hohn H, Pilch H, Neukirch C, Freitag K, Maeurer MJ. Differential MHC Class II Component Expression in HPV-Positive Cervical Cancer Cells: Implication for Immune Surveillance. *Int J Cancer* (2005) 117:807–15. doi: 10.1002/ijc.21226
170. Kamma H, Yazawa T, Ogata T, Horiguchi H, Iijima T. Expression of MHC Class II Antigens in Human Lung Cancer Cells. *Virchows Arch B Cell Pathol Incl Mol Pathol* (1991) 60:407–12. doi: 10.1007/BF02899573
171. Park IA, Hwang SH, Song IH, Heo SH, Kim YA, Bang WS, et al. Expression of the MHC Class II in Triple-Negative Breast Cancer is Associated With Tumor-Infiltrating Lymphocytes and Interferon Signaling. *PLoS One* (2017) 12:e0182786. doi: 10.1371/journal.pone.0182786
172. Johnson DB, Estrada MV, Salgado R, Sanchez V, Doxie DB, Opalenik SR, et al. Melanoma-Specific MHC-II Expression Represents a Tumour-Autonomous Phenotype and Predicts Response to Anti-PD-1/PD-L1 Therapy. *Nat Commun* (2016) 7:10582. doi: 10.1038/ncomms10582
173. Scupoli MT, Sartoris S, Tosi G, Ennas MG, Nicolis M, Cestari T, et al. and Class II Antigens in Pancreatic Adenocarcinomas. *Tissue Antigens* (1996) 48:301–11. doi: 10.1111/j.1399-0039.1996.tb02649.x
174. Kreiter S, Vormehr M, van de Roemer N, Diken M, Lower M, Diekmann J, et al. Mutant MHC Class II Epitopes Drive Therapeutic Immune Responses to Cancer. *Nature* (2015) 520:692–6. doi: 10.1038/nature14426
175. Sun Z, Chen F, Meng F, Wei J, Liu B. MHC Class II Restricted Neoantigen: A Promising Target in Tumor Immunotherapy. *Cancer Lett* (2017) 392:17–25. doi: 10.1016/j.canlet.2016.12.039
176. Gooden M, Lampen M, Jordanova ES, Leffers N, Trimbos JB, van der Burg SH, et al. HLA-E Expression by Gynecological Cancers Restrains Tumor-Infiltrating CD8(+) T Lymphocytes. *Proc Natl Acad Sci USA* (2011) 108:10656–61. doi: 10.1073/pnas.1100354108
177. Lin A, Yan WH. HLA-G Expression in Cancers: Roles in Immune Evasion, Metastasis and Target for Therapy. *Mol Med* (2015) 21(1):782–91. doi: 10.2119/molmed.2015.00083
178. Vacchini A, Chancellor A, Spagnuolo J, Mori L, De Libero G. Mr1-Restricted T Cells are Unprecedented Cancer Fighters. *Front Immunol* (2020) 11:751. doi: 10.3389/fimmu.2020.00751
179. Chong TW, Goh FY, Sim MY, Huang HH, Thike AA, Lim WK, et al. CD1d Expression in Renal Cell Carcinoma is Associated With Higher Relapse Rates, Poorer Cancer-Specific and Overall Survival. *J Clin Pathol* (2015) 68:200–5. doi: 10.1136/jclinpath-2014-202735
180. Yang PM, Lin PJ, Chen CC. CD1d Induction in Solid Tumor Cells by Histone Deacetylase Inhibitors Through Inhibition of HDAC1/2 and Activation of Sp1. *Epigenetics* (2012) 7:390–9. doi: 10.4161/epi.19373
181. Chaudhry MS, Karadimitris A. Role and Regulation of CD1d in Normal and Pathological B Cells. *J Immunol* (2014) 193:4761–8. doi: 10.4049/jimmunol.1401805
182. Yarchoan M, Johnson BA, Lutz 3ER, Laheru DA, Jaffee EM. Targeting Neoantigens to Augment Antitumour Immunity. *Nat Rev Cancer* (2017) 17:209–22. doi: 10.1038/nrc.2016.154
183. Rossig C, Kailayangiri S, Jamitzky S, Altvater B. Carbohydrate Targets for CAR T Cells in Solid Childhood Cancers. *Front Oncol* (2018) 8:513. doi: 10.3389/fonc.2018.00513
184. Sun L, Middleton DR, Wantuch PL, Ozdilek A, Avci FY. Carbohydrates as T-Cell Antigens With Implications in Health and Disease. *Glycobiology* (2016) 26:1029–40. doi: 10.1093/glycob/cww062
185. Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, et al. Spatial Maps of Prostate Cancer Transcriptomes Reveal an Unexplored Landscape of Heterogeneity. *Nat Commun* (2018) 9:2419. doi: 10.1038/s41467-018-04724-5
186. Balli D, Rech AJ, Stanger BZ, Vonderheide RH. Immune Cytolytic Activity Stratifies Molecular Subsets of Human Pancreatic Cancer. *Clin Cancer Res* (2017) 23:3129–38. doi: 10.1158/1078-0432.CCR-16-2128
187. Neelamraju Y, Gonzalez-Perez A, Bhat-Nakshatri P, Nakshatri H, Janga SC. Mutational Landscape of RNA-Binding Proteins in Human Cancers. *RNA Biol* (2017) 15(1):115–29. doi: 10.1080/15476286.2017.1391436
188. Trujillo JA, Sweis RF, Bao R, Luke JJ. T Cell–Inflamed Versus Non-T Cell–Inflamed Tumors: A Conceptual Framework for Cancer Immunotherapy Drug Development and Combination Therapy Selection. *Cancer Immunol Res* (2018) 6:990–1000. doi: 10.1158/2326-6066.CIR-18-0277
189. Blank CU, Haanen JB, Ribas A, Schumacher TN. The "Cancer Immunogram". *Science* (2016) 352:658–60. doi: 10.1126/science.aaf2834
190. Karasaki T, Nagayama K, Kuwano H, Nitadori JI, Sato M, Anraku M, et al. An Immunogram for the Cancer-Immunity Cycle: Towards Personalized Immunotherapy of Lung Cancer. *J Thorac Oncol* (2017) 12(5):791–803. doi: 10.1016/j.jtho.2017.01.005
191. Rosenberg SA, Restifo NP. Adoptive Cell Transfer as Personalized Immunotherapy for Human Cancer. *Science* (2015) 348:62–8. doi: 10.1126/science.aaa4967
192. Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, et al. Cancer Regression in Patients After Transfer of Genetically Engineered Lymphocytes. *Science* (2006) 314:126–9. doi: 10.1126/science.1129003
193. Dudley ME, Yang JC, Sherry R, Hughes MS, Royal R, Kammula U, et al. Adoptive Cell Therapy for Patients With Metastatic Melanoma: Evaluation

of Intensive Myeloablative Chemoradiation Preparative Regimens. *J Clin Oncol* (2008) 26:5233–9. doi: 10.1200/JCO.2008.16.5449
194. Chu H, Du F, Gong Z, Lian P, Wang Z, Li P, et al. Better Clinical Efficiency of Tils for Malignant Pleural Effusion and Ascites Than Cisplatin Through Intrapleural and Intraperitoneal Infusion. *Anticancer Res* (2017) 37:4587–91. doi: 10.21873/anticanres.11857
195. Domagala-Kulawik J, Guzman J, Costabel U. Immune Cells in Bronchoalveolar Lavage in Peripheral Lung Cancer–Analysis of 140 Cases. *Respiration* (2003) 70:43–8. doi: 10.1159/000068414
196. Subira D, Castanon S, Aceituno E, Hernandez J, Jimenez-Garofano C, Jimenez A, et al. Flow Cytometric Analysis of Cerebrospinal Fluid Samples and Its Usefulness in Routine Clinical Practice. *Am J Clin Pathol* (2002) 117:952–8. doi: 10.1309/123P-CE6V-WYAK-BB1F
197. Khan AB, Carpenter B, Santos ESP, Pospori C, Khorshed R, Griffin J, et al. Redirection to the Bone Marrow Improves T Cell Persistence and Antitumor Functions. *J Clin Invest* (2018) 128:2010–24. doi: 10.1172/JCI97454
198. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer Treatment and Survivorship Statistics, 2016. *CA: A Cancer J Clin* (2016) 66:271–89. doi: 10.3322/caac.21349
199. Themeli M, Riviere I, Sadelain M. New Cell Sources for T Cell Engineering and Adoptive Immunotherapy. *Cell Stem Cell* (2015) 16:357–66. doi: 10.1016/j.stem.2015.03.011
200. Met O, Jensen KM, Chamberlain CA, Donia M, Svane IM. Principles of Adoptive T Cell Therapy in Cancer. *Semin Immunopathol* (2019) 41:49–58. doi: 10.1007/s00281-018-0703-z
201. Mueller SN, Mackay LK. Tissue-Resident Memory T Cells: Local Specialists in Immune Defence. *Nat Rev Immunol* (2016) 16:79–89. doi: 10.1038/nri.2015.3
202. Park SL, Gebhardt T, Mackay LK. Tissue-Resident Memory T Cells in Cancer Immunosurveillance. *Trends Immunol* (2019) 40:735–47. doi: 10.1016/j.it.2019.06.002
203. Schenkel JM, Fraser KA, Beura LK, Pauken KE, Vezys V, Masopust D. T Cell Memory. Resident Memory CD8 T Cells Trigger Protective Innate and Adaptive Immune Responses. *Science* (2014) 346:98–101. doi: 10.1126/science.1254536
204. Mackay LK, Rahimpour A, Ma JZ, Collins N, Stock AT, Hafon ML, et al. The Developmental Pathway for CD103(+)CD8+ Tissue-Resident Memory T Cells of Skin. *Nat Immunol* (2013) 14:1294–301. doi: 10.1038/ni.2744
205. Gebhardt T, Wakim LM, Eidsmo L, Reading PC, Heath WR, Carbone FR. Memory T Cells in Nonlymphoid Tissue That Provide Enhanced Local Immunity During Infection With Herpes Simplex Virus. *Nat Immunol* (2009) 10:524–30. doi: 10.1038/ni.1718
206. Cheuk S, Schlums H, Gallais Serezal I, Martini E, Chiang SC, Marquardt N, et al. CD49a Expression Defines Tissue-Resident CD8(+) T Cells Poised for Cytotoxic Function in Human Skin. *Immunity* (2017) 46:287–300. doi: 10.1016/j.immuni.2017.01.009
207. Naik S, Bouladoux N, Linehan JL, Han SJ, Harrison OJ, Wilhelm C, et al. Commensal-Dendritic-Cell Interaction Specifies a Unique Protective Skin Immune Signature. *Nature* (2015) 520:104–8. doi: 10.1038/nature14052
208. Harrison OJ, Linehan JL, Shih HY, Bouladoux N, Han SJ, Smelkinson M, et al. Commensal-Specific T Cell Plasticity Promotes Rapid Tissue Adaptation to Injury. *Science* (2019) 363(6422):eaat6280. doi: 10.1126/science.aat6280
209. Boddupalli CS, Bar N, Kadaveru K, Krauthammer M, Pornputtapong N, Mai Z, et al. Interlesional Diversity of T Cell Receptors in Melanoma With Immune Checkpoints Enriched in Tissue-Resident Memory T Cells. *JCI Insight* (2016) 1:e88955. doi: 10.1172/jci.insight.88955
210. Djenidi F, Adam J, Goubar A, Durgeau A, Meurice G, de Montpreville V, et al. CD8+CD103+ Tumor-Infiltrating Lymphocytes Are Tumor-Specific Tissue-Resident Memory T Cells and a Prognostic Factor for Survival in Lung Cancer Patients. *J Immunol* (2015) 194:3475–86. doi: 10.4049/jimmunol.1402711
211. Wang B, Wu S, Zeng H, Liu Z, Dong W, He W, et al. CD103+ Tumor Infiltrating Lymphocytes Predict a Favorable Prognosis in Urothelial Cell Carcinoma of the Bladder. *J Urol* (2015) 194:556–62. doi: 10.1016/j.juro.2015.02.2941
212. Webb JR, Wick DA, Nielsen JS, Tran E, Milne K, McMurtrie E, et al. Profound Elevation of CD8+ T Cells Expressing the Intraepithelial Lymphocyte Marker CD103 (Alphae/Beta7 Integrin) in High-Grade Serous Ovarian Cancer. *Gynecol Oncol* (2010) 118:228–36. doi: 10.1016/j.ygyno.2010.05.016
213. Komdeur FL, Prins TM, van de Wall S, Plat A, Wisman GBA, Hollema H, et al. CD103+ Tumor-Infiltrating Lymphocytes Are Tumor-Reactive Intraepithelial CD8+ T Cells Associated With Prognostic Benefit and Therapy Response in Cervical Cancer. *Oncoimmunology* (2017) 6:e1338230. doi: 10.1080/2162402X.2017.1338230
214. Wang ZQ, Milne K, Derocher H, Webb JR, Nelson BH, Watson PH. CD103 and Intratumoral Immune Response in Breast Cancer. *Clin Cancer Res* (2016) 22:6290–7. doi: 10.1158/1078-0432.CCR-16-0732
215. Quinn E, Hawkins N, Yip YL, Suter C, Ward R. CD103+ Intraepithelial Lymphocytes–A Unique Population in Microsatellite Unstable Sporadic Colorectal Cancer. *Eur J Cancer* (2003) 39:469–75. doi: 10.1016/S0959-8049(02)00633-0
216. Edwards J, Wilmott JS, Madore J, Gide TN, Quek C, Tasker A, et al. CD103 (+) Tumor-Resident CD8(+) T Cells Are Associated With Improved Survival in Immunotherapy-Naive Melanoma Patients and Expand Significantly During Anti-PD-1 Treatment. *Clin Cancer Res* (2018) 24:3036–45. doi: 10.1158/1078-0432.CCR-17-2257
217. Bosmuller HC, Wagner P, Peper JK, Schuster H, Pham DL, Greif K, et al. Combined Immunoscore of CD103 and CD3 Identifies Long-Term Survivors in High-Grade Serous Ovarian Cancer. *Int J Gynecol Cancer* (2016) 26:671–9. doi: 10.1097/IGC.0000000000000672
218. Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, et al. Single-Cell Profiling of Breast Cancer T Cells Reveals a Tissue-Resident Memory Subset Associated With Improved Prognosis. *Nat Med* (2018) 24:986–93. doi: 10.1038/s41591-018-0078-7
219. Workel HH, Komdeur FL, Wouters MC, Plat A, Klip HG, Eggink FA, et al. CD103 Defines Intraepithelial CD8+ PD1+ Tumour-Infiltrating Lymphocytes of Prognostic Significance in Endometrial Adenocarcinoma. *Eur J Cancer* (2016) 60:1–11. doi: 10.1016/j.ejca.2016.02.026
220. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global Characterization of T Cells in Non-Small-Cell Lung Cancer by Single-Cell Sequencing. *Nat Med* (2018) 24:978–85. doi: 10.1038/s41591-018-0045-3
221. Darvin P, Toor SM, Sasidharan Nair V, Elkord E. Immune Checkpoint Inhibitors: Recent Progress and Potential Biomarkers. *Exp Mol Med* (2018) 50:165–5. doi: 10.1038/s12276-018-0191-1
222. Sahin U, Türeci Ö. Personalized Vaccines for Cancer Immunotherapy. *Science* (2018) 359:1355–60. doi: 10.1126/science.aar7112
223. Chowell D, Morris LGT, Grigg CM, Weber JK, Samstein RM, Makarov V, et al. Patient HLA Class I Genotype Influences Cancer Response to Checkpoint Blockade Immunotherapy. *Science (New York, NY)* (2018) 359:582–7. doi: 10.1126/science.aao4572
224. Durgeau A, Virk Y, Corgnac S, Mami-Chouaib F. Recent Advances in Targeting CD8 T-Cell Immunity for More Effective Cancer Immunotherapy. *Front Immunol* (2018) 9:14–4. doi: 10.3389/fimmu.2018.00014
225. Ali M, Foldvari Z, Giannakopoulou E, Boschen ML, Stronen E, Yang W, et al. Induction of Neoantigen-Reactive T Cells From Healthy Donors. *Nat Protoc* (2019) 14:1926–43. doi: 10.1038/s41596-019-0170-6
226. Audehm S, Glaser M, Pecoraro M, Bräunlein E, Mall S, Klar R, et al. Key Features Relevant to Select Antigens and TCR From the MHC-Mismatched Repertoire to Treat Cancer. *Front Immunol* (2019) 10:1485. doi: 10.3389/fimmu.2019.01485
227. Lu YC, Parker LL, Lu T, Zheng Z, Toomey MA, White DE, et al. Treatment of Patients With Metastatic Cancer Using a Major Histocompatibility Complex Class II-Restricted T-Cell Receptor Targeting the Cancer Germline Antigen MAGE-A3. *J Clin Oncol* (2017) 35:3322–9. doi: 10.1200/JCO.2017.74.5463
228. Rapoport AP, Stadtmauer EA, Binder-Scholl GK, Goloubeva O, Vogl DT, Lacey SF, et al. NY-ESO-1-Specific TCR-Engineered T Cells Mediate Sustained Antigen-Specific Antitumor Effects in Myeloma. *Nat Med* (2015) 21:914–21. doi: 10.1038/nm.3910
229. Robbins PF, Kassim SH, Tran TL, Crystal JS, Morgan RA, Feldman SA, et al. A Pilot Trial Using Lymphocytes Genetically Engineered With An NY-ESO-1-Reactive T-Cell Receptor: Long-Term Follow-Up and Correlates With Response. *Clin Cancer Res* (2015) 21:1019–27. doi: 10.1158/1078-0432.CCR-14-2708

230. D'Angelo SP, Melchiori L, Merchant MS, Bernstein D, Glod J, Kaplan R, et al. Antitumor Activity Associated With Prolonged Persistence of Adoptively Transferred NY-ESO-1 (C259)T Cells in Synovial Sarcoma. *Cancer Discov* (2018) 8:944–57. doi: 10.1158/2159-8290.CD-17-1417
231. Legut M, Cole DK, Sewell AK. The Promise of Gammadelta T Cells and the Gammadelta T Cell Receptor for Cancer Immunotherapy. *Cell Mol Immunol* (2015) 12:656–68. doi: 10.1038/cmi.2015.28
232. Kobayashi H, Tanaka Y, Yagi J, Osaka Y, Nakazawa H, Uchiyama T, et al. Safety Profile and Anti-Tumor Effects of Adoptive Immunotherapy Using Gamma-Delta T Cells Against Advanced Renal Cell Carcinoma: A Pilot Study. *Cancer Immunol Immunother* (2007) 56:469–76. doi: 10.1007/s00262-006-0199-6
233. Kobayashi H, Tanaka Y, Shimmura H, Minato N, Tanabe K. Complete Remission of Lung Metastasis Following Adoptive Immunotherapy Using Activated Autologous Gammadelta T-Cells in a Patient With Renal Cell Carcinoma. *Anticancer Res* (2010) 30:575–9.
234. Kobayashi H, Tanaka Y, Yagi J, Minato N, Tanabe K. Phase I/II Study of Adoptive Transfer of Gammadelta T Cells in Combination With Zoledronic Acid and IL-2 to Patients With Advanced Renal Cell Carcinoma. *Cancer Immunol Immunother* (2011) 60:1075–84. doi: 10.1007/s00262-011-1021-7
235. Nakajima J, Murakawa T, Fukami T, Goto S, Kaneko T, Yoshida Y, et al. A Phase I Study of Adoptive Immunotherapy for Recurrent Non-Small-Cell Lung Cancer Patients With Autologous Gammadelta T Cells. *Eur J Cardiothorac Surg* (2010) 37:1191–7. doi: 10.1016/j.ejcts.2009.11.051
236. Sakamoto M, Nakajima J, Murakawa T, Fukami T, Yoshida Y, Murayama T, et al. Adoptive Immunotherapy for Advanced Non-Small Cell Lung Cancer Using Zoledronate-Expanded Gammadeltatcells: A Phase I Clinical Study. *J Immunother* (2011) 34:202–11. doi: 10.1097/CJI.0b013e318207ecfb
237. Ferreira LM. Gammadelta T Cells: Innately Adaptive Immune Cells? *Int Rev Immunol* (2013) 32:223–48. doi: 10.3109/08830185.2013.783831
238. Silva-Santos B, Mensurado S, Coffelt SB. Gammadelta T Cells: Pleiotropic Immune Effectors With Therapeutic Potential in Cancer. *Nat Rev Cancer* (2019) 19(7):392–404. doi: 10.1038/s41568-019-0153-5
239. Coffelt SB, Kersten K, Doornebal CW, Weiden J, Vrijland K, Hau CS, et al. IL-17-Producing Gammadelta T Cells and Neutrophils Conspire to Promote Breast Cancer Metastasis. *Nature* (2015) 522:345–8. doi: 10.1038/nature14282
240. Kimura Y, Nagai N, Tsunekawa N, Sato-Matsushita M, Yoshimoto T, Cua DJ, et al. IL-17A-Producing CD30(+) Vdelta1 T Cells Drive Inflammation-Induced Cancer Progression. *Cancer Sci* (2016) 107:1206–14. doi: 10.1111/cas.13005
241. Akitsu A, Iwakura Y. Interleukin-17-Producing Gammadelta T (Gammadelta17) Cells in Inflammatory Diseases. *Immunology* (2018) 155:418–26. doi: 10.1111/imm.12993
242. Kenna T, Golden-Mason L, Norris S, Hegarty JE, O'Farrelly C, Doherty DG. Distinct Subpopulations of Gamma Delta T Cells Are Present in Normal and Tumor-Bearing Human Liver. . *Clin Immunol (Orlando Fla.)* (2004) 113:56–63. doi: 10.1016/j.clim.2004.05.003
243. Correia DV, Fogli M, Hudspeth K, da Silva MG, Mavilio D, Silva-Santos B. Differentiation of Human Peripheral Blood Vdelta1+ T Cells Expressing the Natural Cytotoxicity Receptor Nkp30 for Recognition of Lymphoid Leukemia Cells. *Blood* (2011) 118:992–1001. doi: 10.1182/blood-2011-02-339135
244. Wu D, Wu P, Qiu F, Wei Q, Huang J. Human Gammadeltat-Cell Subsets and Their Involvement in Tumor Immunity. *Cell Mol Immunol* (2017) 14:245–53. doi: 10.1038/cmi.2016.55
245. Daley D, Zambirinis CP, Seifert L, Akkad N, Mohan N, Werba G, et al. Gammadelta T Cells Support Pancreatic Oncogenesis by Restraining Alphabeta T Cell Activation. *Cell* (2016) 166:1485–99.e15. doi: 10.1016/j.cell.2016.07.046
246. Maeurer M, Zitvogel L, Elder E, Storkus WJ, Lotze MT. Human Intestinal V Delta 1+ T Cells Obtained From Patients With Colon Cancer Respond Exclusively to SEB But Not to SEA. *Nat Immun* (1995) 14:188–97.
247. Maeurer MJ, Martin D, Walter W, Liu K, Zitvogel L, Halusczcak K, et al. Human Intestinal Vdelta1+ Lymphocytes Recognize Tumor Cells of Epithelial Origin. *J Exp Med* (1996) 183:1681–96. doi: 10.1084/jem.183.4.1681
248. Wolf BJ, Choi JE, Exley MA. Novel Approaches to Exploiting Invariant NKT Cells in Cancer Immunotherapy. *Front Immunol* (2018) 9:384. doi: 10.3389/fimmu.2018.00384
249. Giaccone G, Punt CJ, Ando Y, Ruijter R, Nishi N, Peters M, et al. A Phase I Study of the Natural Killer T-Cell Ligand Alpha-Galactosylceramide (KRN7000) in Patients With Solid Tumors. *Clin Cancer Res* (2002) 8:3702–9.
250. Uchida T, Horiguchi S, Tanaka Y, Yamamoto H, Kunii N, Motohashi S, et al. Phase I Study of Alpha-Galactosylceramide-Pulsed Antigen Presenting Cells Administration to the Nasal Submucosa in Unresectable or Recurrent Head and Neck Cancer. *Cancer Immunol Immunother* (2008) 57:337–45. doi: 10.1007/s00262-007-0373-5
251. Kabelitz D, Serrano R, Kouakanou L, Peters C, Kalyan S. Cancer Immunotherapy With Gammadelta T Cells: Many Paths Ahead of Us. *Cell Mol Immunol* (2020) 17:925–39. doi: 10.1038/s41423-020-0504-x
252. Garg K, Maurer M, Griss J, Bruggen MC, Wolf IH, Wagner C, et al. Tumor-Associated B Cells in Cutaneous Primary Melanoma and Improved Clinical Outcome. *Hum Pathol* (2016) 54:157–64. doi: 10.1016/j.humpath.2016.03.022
253. Fremd C, Stefanovic S, Beckhove P, Pritsch M, Lim H, Wallwiener M, et al. Mucin 1-Specific B Cell Immune Responses and Their Impact on Overall Survival in Breast Cancer Patients. *Oncoimmunology* (2016) 5:e1057387. doi: 10.1080/2162402X.2015.1057387
254. Chen JL, Dawoodji A, Tarlton A, Gnjatic S, Tajar A, Karydis I, et al. NY-ESO-1 Specific Antibody and Cellular Responses in Melanoma Patients Primed With NY-ESO-1 Protein in ISCOMATRIX and Boosted With Recombinant NY-ESO-1 Fowlpox Virus. *Int J Cancer* (2015) 136:E590–601. doi: 10.1002/ijc.29118
255. Meng Q, Valentini D, Rao M, Maeurer M. KRAS RENAISSANCE(S) in Tumor Infiltrating B Cells in Pancreatic Cancer. *Front Oncol* (2018) 8:384. doi: 10.3389/fonc.2018.00384
256. van Dam GM, Themelis G, Crane LMA, Harlaar NJ, Pleijhuis RG, Kelder W, et al. Intraoperative Tumor-Specific Fluorescence Imaging in Ovarian Cancer by Folate Receptor-A Targeting: First in-Human Results. *Nat Med* (2011) 17:1315. doi: 10.1038/nm.2472
257. Zhang Y, Chang L, Yang Y, Fang W, Guan Y, Wu A, et al. The Correlations of Tumor Mutational Burden Among Single-Region Tissue, Multi-Region Tissues and Blood in Non-Small Cell Lung Cancer. *J ImmunoTher Cancer* (2019) 7:98. doi: 10.1186/s40425-019-0581-5
258. Das R, Bar N, Ferreira M, Newman AM, Zhang L, Bailur JK, et al. Early B Cell Changes Predict Autoimmunity Following Combination Immune Checkpoint Blockade. *J Clin Invest* (2018) 128(2):715–20. doi: 10.1172/JCI96798
259. Meng Q, Valentini D, Rao M, Dodoo E, Maeurer M. CMV and EBV Targets Recognized by Tumor-Infiltrating B Lymphocytes in Pancreatic Cancer and Brain Tumors. *Sci Rep* (2018) 8:17079–89. doi: 10.1038/s41598-018-34710-2
260. Waltari E, McGeever A, Friedland N, Kim PS, McCutcheon KM. Functional Enrichment and Analysis of Antigen-Specific Memory B Cell Antibody Repertoires in Pbmcs. *Front Immunol* (2019) 10:1452. doi: 10.3389/fimmu.2019.01452
261. Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-Throughput Targeted Long-Read Single Cell Sequencing Reveals the Clonal and Transcriptional Landscape of Lymphocytes. *Nat Commun* (2019) 10:3120. doi: 10.1038/s41467-019-11049-4
262. Nicodemus CF. Antibody-Based Immunotherapy of Solid Cancers: Progress and Possibilities. *Immunotherapy* (2015) 7(8):923–39. doi: 10.2217/imt.15.57
263. Garaud S, Buisseret L, Solinas C, Gu-Trantien C, de Wind A, Van den Eynden G, et al. Tumor Infiltrating B-Cells Signal Functional Humoral Immune Responses in Breast Cancer. *JCI Insight* (2019) 5(18):e129641. doi: 10.1172/jci.insight.129641
264. Petitprez F, de Reynies A, Keung EZ, Chen TW, Sun CM, Calderaro J, et al. B Cells are Associated With Survival and Immunotherapy Response in Sarcoma. *Nature* (2020) 577:556–60. doi: 10.1038/s41586-019-1906-8
265. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, et al. B Cells and Tertiary Lymphoid Structures Promote Immunotherapy Response. *Nature* (2020) 577:549–55. doi: 10.1038/s41586-019-1922-8
266. Ogino T, Wang X, Kato S, Miyokawa N, Harabuchi Y, Ferrone S. Endoplasmic Reticulum Chaperone-Specific Monoclonal Antibodies for Flow Cytometry and Immunohistochemical Staining. *Tissue Antigens* (2003) 62:385–93. doi: 10.1034/j.1399-0039.2003.00114.x
267. Kikuchi E, Yamazaki K, Torigoe T, Cho Y, Miyamoto M, Oizumi S, et al. HLA Class I Antigen Expression is Associated With a Favorable Prognosis in Early Stage Non-Small Cell Lung Cancer. *Cancer Sci* (2007) 98:1424–30. doi: 10.1111/j.1349-7006.2007.00558.x
268. Torigoe T, Asanuma H, Nakazawa E, Tamura Y, Hirohashi Y, Yamamoto E, et al. Establishment of a Monoclonal Anti-Pan HLA Class I Antibody Suitable for Immunostaining of Formalin-Fixed Tissue: Unusually High

Frequency of Down-Regulation in Breast Cancer Tissues. *Pathol Int* (2012) 62:303–8. doi: 10.1111/j.1440-1827.2012.02789.x
269. Twomey JD, Kim SR, Zhao L, Bozza WP, Zhang B. Spatial Dynamics of TRAIL Death Receptors in Cancer Cells. *Drug Resist Update* (2015) 19:13–21. doi: 10.1016/j.drup.2015.02.001
270. Z'Graggen K, Centeno BA, Fernandez-del Castillo C, Jimenez RE, Werner J, Warshaw AL. Biological Implications of Tumor Cells in Blood and Bone Marrow of Pancreatic Cancer Patients. *Surgery* (2001) 129:537–46. doi: 10.1067/msy.2001.113819
271. Sho S, Court CM, Winograd P, Lee S, Hou S, Graeber TG, et al. Precision Oncology Using a Limited Number of Cells: Optimization of Whole Genome Amplification Products for Sequencing Applications. *BMC Cancer* (2017) 17:457. doi: 10.1186/s12885-017-3447-6
272. Lin E, Cao T, Nagrath S, King MR. Circulating Tumor Cells: Diagnostic and Therapeutic Applications. *Annu Rev BioMed Eng* (2018) 20:329–52. doi: 10.1146/annurev-bioeng-062117-120947
273. Moritz A, Anjanappa R, Wagner C, Bunk S, Hofmann M, Pszolla G, et al. High-Throughput Peptide-MHC Complex Generation and Kinetic Screenings of Tcrs With Peptide-Receptive HLA-a*02:01 Molecules. *Sci Immunol* (2019) 4:eaav0860. doi: 10.1126/sciimmunol.aav0860
274. Saini SK, Tamhane T, Anjanappa R, Saikia A, Ramskov S, Donia M, et al. Empty Peptide-Receptive MHC Class I Molecules for Efficient Detection of Antigen-Specific T Cells. *Sci Immunol* (2019) 4:eaau9039. doi: 10.1126/sciimmunol.aau9039
275. Danova M, Torchio M, Comolli G, Sbrana A, Antonuzzo A, Mazzini G. The Role of Automated Cytometry in the New Era of Cancer Immunotherapy. *Mol Clin Oncol* (2018) 9:355–61. doi: 10.3892/mco.2018.1701
276. Coelho MA, de Carne Trecesson S, Rana S, Zecchin D, Moore C, Molina-Arcas M, et al. Oncogenic RAS Signaling Promotes Tumor Immunoresistance by Stabilizing PD-L1 Mrna. *Immunity* (2017) 47 (6):1083–99.e6. doi: 10.1016/j.immuni.2017.11.016
277. Dong ZY, Zhong WZ, Zhang XC, Su J, Xie Z, Liu SY, et al. Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma. *Clin Cancer Res* (2017) 23:3012–24. doi: 10.1016/j.jtho.2016.11.504
278. Bachinsky MM, Guillen DE, Patel SR, Singleton J, Chen C, Soltis DA, et al. Mapping and Binding Analysis of Peptides Derived From the Tumor-Associated Antigen Survivin for Eight HLA Alleles. *Cancer Immun* (2005) 5:6.
279. Axelsson-Robertson R, Ahmed RK, Weichold FF, Ehlers MM, Kock MM, Sizemore D, et al. Human Leukocyte Antigens a*3001 and a*3002 Show Distinct Peptide-Binding Patterns of the Mycobacterium Tuberculosis Protein TB10.4: Consequences for Immune Recognition. *Clin Vaccine Immunol* (2011) 18:125–34. doi: 10.1128/CVI.00302-10
280. Axelsson-Robertson R, Weichold F, Sizemore D, Wulf M, Skeiky YA, Sadoff J, et al. Extensive Major Histocompatibility Complex Class I Binding Promiscuity for Mycobacterium Tuberculosis TB10.4 Peptides and Immune Dominance of Human Leucocyte Antigen (HLA)-B*0702 and HLA-B*0801 Alleles in TB10.4 CD8 T-Cell Responses. *Immunology* (2010) 129:496–505. doi: 10.1111/j.1365-2567.2009.03201.x
281. van der Burg SH, Visseren MJ, Brandt RM, Kast WM, Melief CJ. Immunogenicity of Peptides Bound to MHC Class I Molecules Depends on the MHC-Peptide Complex Stability. *J Immunol* (1996) 156:3308–14.
282. Spierings E, Gras S, Reiser J-B, Mommaas B, Almekinders M, Kester MGD, et al. Steric Hindrance and Fast Dissociation Explain the Lack of Immunogenicity of the Minor Histocompatibility HA-1Arg Null Allele. *J Immunol* (2009) 182:4809–16. doi: 10.4049/jimmunol.0803911
283. Micheletti F, Guerrini R, Formentin A, Canella A, Marastoni M, Bazzaro M, et al. Selective Amino Acid Substitutions of a Subdominant Epstein-Barr Virus LMP2-Derived Epitope Increase HLA/Peptide Complex Stability and Immunogenicity: Implications for Immunotherapy of Epstein-Barr Virus-Associated Malignancies. *Eur J Immunol* (1999) 29:2579–89. doi: 10.1002/(SICI)1521-4141(199908)29:08<2579::AID-IMMU2579>3.0.CO;2-E
284. van Stipdonk MJB, Badia-Martinez D, Sluijter M, Offringa R, van Hall T, Achour A. Design of Agonistic Altered Peptides for the Robust Induction of CTL Directed Towards H-2Db in Complex With the Melanoma-Associated Epitope Gp100. *Cancer Res* (2009) 69:7784–92. doi: 10.1158/0008-5472.CAN-09-1724
285. Lazarski CA, Chaves FA, Jenks SA, Wu S, Richards KA, Weaver JM, et al. The Kinetic Stability of MHC Class II: Peptide Complexes Is a Key Parameter That Dictates Immunodominance. *Immunity* (2005) 23:29–40. doi: 10.1016/j.immuni.2005.05.009
286. Burrows JM, Wynn KK, Tynan FE, Archbold J, Miles JJ, Bell MJ, et al. The Impact of HLA-B Micropolymorphism Outside Primary Peptide Anchor Pockets on the CTL Response to CMV. *Eur J Immunol* (2007) 37:946–53. doi: 10.1002/eji.200636588
287. Harndahl M, Rasmussen M, Roder G, Dalgaard Pedersen I, Sorensen M, Nielsen M, et al. Peptide-MHC Class I Stability Is a Better Predictor Than Peptide Affinity of CTL Immunogenicity. *Eur J Immunol* (2012) 42:1405–16. doi: 10.1002/eji.201141774
288. Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-Specific Prediction of Peptide–MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. *J Immunol* (2016) 197:1517–24. doi: 10.4049/jimmunol.1600582
289. Nicholls S, Piper KP, Mohammed F, Dafforn TR, Tenzer S, Salim M, et al. Secondary Anchor Polymorphism in the HA-1 Minor Histocompatibility Antigen Critically Affects MHC Stability and TCR Recognition. *Proc Natl Acad Sci* (2009) 106:3889–94. doi: 10.1073/pnas.0900411106
290. Dudley ME, Rosenberg SA. Adoptive-Cell-Transfer Therapy for the Treatment of Patients With Cancer. *Nat Rev Cancer* (2003) 3:666–75. doi: 10.1038/nrc1167
291. Maeurer MJ, Martin D, Elder E, Storkus WJ, Lotze MT. Detection of Naturally Processed and HLA-A1-Presented Melanoma T-Cell Epitopes Defined by CD8(+) T-Cells' Release of Granulocyte-Macrophage Colony-Stimulating Factor But Not by Cytolysis. *Clin Cancer Res* (1996) 2:87–95.
292. Ostmeyer J, Christley S, Toby IT, Cowell LG. Biophysicochemical Motifs in T-Cell Receptor Sequences Distinguish Repertoires From Tumor-Infiltrating Lymphocyte and Adjacent Healthy Tissue. *Cancer Res* (2019) 79:1671–80. doi: 10.1158/0008-5472.CAN-18-2292
293. Rosenthal R, Cadieux EL, Salgado R, Bakir MA, Moore DA, Hiley CT, et al. And T.R. Consortium. *Neoantigen-Directed Immune Escape Lung Cancer Evol Nat* (2019) 567:479–85. doi: 10.1038/s41586-019-1032-7
294. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic Snvs and Indels With Mutect2. *bioRxiv* (2019) 861054. doi: 10.1101/861054
295. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. Varscan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing. *Genome Res* (2012) 22:568–76. doi: 10.1101/gr.129684.111
296. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: Fast and Accurate Calling of Germline and Somatic Variants. *Nat Methods* (2018) 15:591–4. doi: 10.1038/s41592-018-0051-x
297. Narzisi G, Corvelo A, Arora K, Bergmann EA, Shah M, Musunuri R, et al. Genome-Wide Somatic Variant Calling Using Localized Colored De Bruijn Graphs. *Commun Biol* (2018) 1:20–0. doi: 10.1038/s42003-018-0023-9
298. Hundal J, Kiwala S, McMichael J, Miller CA, Xia H, Wollam AT, et al. Pvactools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. *Cancer Immunol Res* (2020) 8:409–20. doi: 10.1158/2326-6066.CIR-19-0401
299. Olcina MM, Balanis NG, Kim RK, Aksoy BA, Kodysh J, Thompson MJ, et al. Mutations in an Innate Immunity Pathway Are Associated With Poor Overall Survival Outcomes and Hypoxic Signaling in Cancer. *Cell Rep* (2018) 25:3721–32.e6. doi: 10.1016/j.celrep.2018.11.093
300. Porta-Pardo E, Godzik A. Mutation Drivers of Immunological Responses to Cancer. *Cancer Immunol Res* (2016) 4:789–98. doi: 10.1158/2326-6066.CIR-15-0233
301. Li J, Wang W, Zhang Y, Cieslik M, Guo J, Tan M, et al. Epigenetic Driver Mutations in ARID1A Shape Cancer Immune Phenotype and Immunotherapy. *J Clin Invest* (2020) 130:2712–26. doi: 10.1172/JCI134402
302. Looney TJ, Topacio-Hall D, Lowman G, Conroy J, Morrison C, Oh D, et al. TCR Convergence in Individuals Treated With Immune Checkpoint Inhibition for Cancer. *Front Immunol* (2019) 10:2985. doi: 10.3389/fimmu.2019.02985
303. Hu Z, Leet DE, Allesoe RL, Oliveira G, Li S, Luoma AM, et al. Personal Neoantigen Vaccines Induce Persistent Memory T Cell Responses and Epitope Spreading in Patients With Melanoma. *Nat Med* (2021) 27:515–25. doi: 10.1038/s41591-020-01206-4
304. Meng Q, Valentini D, Rao M, Liu Z, Xie S, Morgell A, et al. Prediction of Improved Survival in Patients With Pancreatic Cancer *Via* IL-21 Enhanced Detection of Mesothelin Epitope-Reactive T-Cell Responses. *Oncotarget* (2018) 9:22451–9. doi: 10.18632/oncotarget.25121

305. Zhenjiang L, Rao M, Luo X, Sandberg E, Bartek J Jr, Schoutrop E, et al. Mesothelin-Specific Immune Responses Predict Survival of Patients With Brain Metastasis. *EBioMedicine* (2017) 23:20–4. doi: 10.1016/j.ebiom.2017.08.024
306. Zhenjiang L, Rao M, Luo X, Valentini D, von Landenberg A, Meng Q, et al. Cytokine Networks and Survivin Peptide-Specific Cellular Immune Responses Predict Improved Survival in Patients With Glioblastoma Multiforme. *EBioMedicine* (2018) 33:49–56. doi: 10.1016/j.ebiom.2018.06.014
307. Alcantara M, Du Rusquec P, Romano E. Current Clinical Evidence and Potential Solutions to Increase Benefit of CAR T-Cell Therapy for Patients With Solid Tumors. *OncoImmunology* (2020) 9:1777064. doi: 10.1080/2162402X.2020.1777064
308. Smith SM, Iwenofu OH. NY-ESO-1: A Promising Cancer Testis Antigen for Sarcoma Immunotherapy and Diagnosis. *Chin Clin Oncol* (2018) 7:44. doi: 10.21037/cco.2018.08.11
309. Thomas R, Al-Khadairi G, Roelands J, Hendrickx W, Dermime S, Bedognetti D, et al. NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives. *Front Immunol* (2018) 9:947. doi: 10.3389/fimmu.2018.00947
310. Klampatsa A, Dimou V, Albelda SM. Mesothelin-Targeted CAR-T Cell Therapy for Solid Tumors. *Expert Opin Biol Ther* (2020) 21(4):473–86. doi: 10.1080/14712598.2021.1843628
311. Angelova M, Mlecnik B, Vasaturo A, Bindea G, Fredriksen T, Lafontaine L, et al. Evolution of Metastases in Space and Time Under Immune Selection. *Cell* (2018) 175:751–65.e16. doi: 10.1016/j.cell.2018.09.018
312. Hu Z, Li Z, Ma Z, Curtis C. Multi-Cancer Analysis of Clonality and the Timing of Systemic Spread in Paired Primary Tumors and Metastases. *Nat Genet* (2020) 52:701–8. doi: 10.1038/s41588-020-0628-z
313. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. *Immunity* (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023
314. Trinh A, Polyak K. Tumor Neoantigens: When Too Much of a Good Thing is Bad. *Cancer Cell* (2019) 36:466–7. doi: 10.1016/j.ccell.2019.10.009
315. Leruste A, Tosello J, Ramos RN, Tauziede-Espariat A, Brohard S, Han ZY, et al. Clonally Expanded T Cells Reveal Immunogenicity of Rhabdoid Tumors. *Cancer Cell* (2019) 36:597–612.e8. doi: 10.1016/j.ccell.2019.10.008
316. Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable Predictive Features Define Epitope-Specific T Cell Receptor Repertoires. *Nature* (2017) 547:89–93. doi: 10.1038/nature22383
317. Gee MH, Han A, Lofgren SM, Beausang JF, Mendoza JL, Birnbaum ME, et al. Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. *Cell* (2018) 172:549–63.e16. doi: 10.1016/j.cell.2017.11.043
318. Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM. Analyzing the Mycobacterium Tuberculosis Immune Response by T-Cell Receptor Clustering With GLIPH2 and Genome-Wide Antigen Screening. *Nat Biotechnol* (2020) 38(10):1194–202. doi: 10.1038/s41587-020-0505-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership