HIGH THROUGHPUT COMPUTATIONAL METHODS FOR IMMUNO-ONCOLOGY: PRECISE PATIENT STRATIFICATION BASED ON NEOANTIGEN PROFILE ANALYSES

by

Xiaoshan (Melody) Shao

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

December 2022

Abstract

Computational prediction of binding between neoantigen peptides and major histocompatibility complex (MHC) proteins can be used to predict patient response to cancer immunotherapy for immune-checkpoint blockade (ICB) therapies. Current neoantigen predictors focus on *in silico* estimation of MHC binding affinity and are limited by low predictive value (PPV) for actual peptide presentation, inadequate support for rare MHC alleles, and poor scalability to high-throughput data sets. To address these limitations, I developed MHCnuggets, a deep neural network method that predicts peptide-MHC binding. MHCnuggets can predict binding for common or rare alleles of MHC class I or II with a single network. Using a long short-term memory (LSTM) network, MHCnuggets accepts peptides of variable length and is faster than other methods. MHCnuggets yields a 4-fold PPV improvement when tested with MHC-bound peptide (HLAp) data. MHCnuggets' was next used to explore ICB-treated patients' mutation-associated neoantigen (MANA) and immunogenic mutation (IMM) profiles. I found that responding tumors harbored a significantly higher HLA class II IMM burden for both melanoma and Non-Small Cell Lung Cancer (NSCLC) ($P \leq 9.6e\text{-}3$). HLA class II IMM burden was correlated with longer survival. HLA class I and II IMMs were largely distinct, suggesting a complementary role for class II IMMs in tumor rejection. A higher HLA class II IMM burden was associated with $CD4^+$ T-cell infiltration. In addition, I found distinct sex-based immunogenomic landscapes in NSCLC: TMB, class I IMM load, class II IMM load, and mutational smoking signature were significantly higher in female responding tumors as

compared to female non-responding tumors ($P \leq 0.005$). In contrast, among males, there was no significant association between durable-clinical benefit and any of these features. When IMM was considered in the context of HLA zygosity, high MHC-II restricted IMM load and high HLA class II diversity was significantly associated with overall survival in males ($p=0.017$). Such results indicated that inherent sex-driven differences in immune surveillance affect the immunogenomic determinants of response to ICB and likely mediate the dimorphic outcomes with ICB therapy. Taken together, my thesis work used computational methods to provide more precise patient stratification in immune oncology based on patients' genomic and neoantigen profiles.

Primary Reader and Advisor: Dr. Rachel Karchin

Secondary Reader: Dr. Valsamo Anagnostou

Acknowledgements

I am forever grateful to all the people who have supported me both academically and nonacademically to make this work possible. First, I would like to thank my thesis advisor, Rachel Karchin. Her guidance and encouragement were invaluable in shaping the direction and the final presentation of this dissertation. I would also like to thank my thesis committee members, Valsamo (Elsa) Anagnostou and Rob Scharpf, for their great support and insightful suggestions. In particular, many thanks to Elsa's constant support in helping me connect my computational work to biologically relevant and clinically actionable results. I am also incredibly thankful for the scientific and emotional support from past and current members of the Karchin lab: Lily Zheng, Justin Huang, Noushin Niknafs, Rohit Bhattachary, Ashok Sivakumar, Violeta Beleva Gutherie, and Collin Tokheim. Thank you all for making lab a fun place to go to everyday. I would also like to thank former Biomedical Engineering Department admin, Hong Lan, for the great supports she has given since day one of this journey. To all my fellow PhD students in the program, I am glad that it was you that I shared this journey with. PhD was hard, but all the ups and downs were memorable because you were there by my side. During COVID, I was stuck in China due to travel restrictions, and I had to work remotely overseas. In those times, all these aforementioned people believed in me, worked with me, and provided me support that I am forever in debt to. Lastly, Mom, Dad, Vinson, all the members of the Mickle family and all my family in China, I love you all.

Dedication

To my friends and family

Table of Contents

List of Figures

Chapter 1 Introduction

1.1 The need of mutation associated "neoantigens" (MANA) presentation prediction in immune checkpoint blockade (ICB) treatments.

The presentation of peptides bound to major histocompatibility complex (MHC) proteins on the surface of antigen-presenting cells and subsequent recognition by T-cell receptors is fundamental to the mammalian adaptive immune system. Neoantigens derived from somatic mutations are targets of immunoediting and drive therapeutic responses in cancer patients treated with immunotherapy [1, 2]. Because experimental characterization of neoantigens is both costly and time-consuming, computational methods have been developed to predict peptide–MHC binding and subsequent immune response [3, 4]. Supervised neural network machine learning approaches have performed the best [5-7] and are the most widely used in silico methods. Despite advances in computational approaches, improvements in predictive performance have been minimal, due in part to a lack of sufficiently large sets of experimentally characterized peptide binding affinities for most MHC alleles.

Although neoantigen prediction for common MHC class I alleles is well studied [8], predictive accuracy on rare and less-characterized MHC alleles remains poor[9, 10]. Class II predictors are scarce [11]. Current estimates suggest that class II antigen lengths primarily range from 13 to 25 amino acids [12], and this diversity has been an obstacle to developing in silico neoantigen predictors [11, 13]. As most neural network architectures are designed for fixed-length inputs, methods such as NetMHC [14-17]and MHCflurry [18] require preprocessing of peptide sequences or training of separate classifiers for each peptide length.

Clinical application of MHC–peptide binding predictors, to identify biomarkers for cancer immunotherapy, requires scalability to large patient cohorts and low false-positive rates [19]. A cancer may contain hundreds of candidate somatically altered peptides, but few will actually bind to MHC proteins and elicit an immune response [20]. For many years, most neoantigen predictors were trained primarily on quantitative peptide–MHC binding affinity data from in vitro experiments [21]. Advances in immunopeptidomics technologies have enabled the identification of thousands of naturally presented MHC-bound peptides (HLAp) from cancer patient samples and cell lines [19, 22]. Several neoantigen predictors are trained only on HLAp data for class I, and only for a limited number of peptide lengths [21, 23]. The EDGE neural network is trained primarily on multiallelic HLAp and RNA sequencing (RNA-seq) data from 74 cancer patients; ForestMHC is a random forest trained on HLAp from publicly available monoallelic and deconvoluted multiallelic cell lines. The potential to improve neoantigen predictors by integrating binding affinity and HLAp data [19] has motivated hybrid approaches [14, 18]. However, most methods predict more candidate neoantigens than are immunogenic in patients [11, 19].

1.2 MHCnuggets, a deep learning-based solution for computational neoantigen predictions

To address above mentioned limitations, I developed MHCnuggets, an open-source deep neural network method that predicts peptide-MHC binding. MHCnuggets can predict binding for common or rare alleles of MHC class I or II with a single neural network architecture. Using a long short-term memory network (LSTM) [24], MHCnuggets accepts peptides of variable length and does not require lengthening or shortening heuristics of peptides. By leveraging transfer

learning and allele clustering, rare allele predictions can be handled automatically. MHCnuggets was built based on parallel computing frameworks and is fast in both training and prediction processes. When compared to methods that integrate binding affinity and MHC-bound peptide (HLAp) data from mass spectrometry, MHCnuggets yields a fourfold increase in positive predictive value on independent HLAp data. To demonstrate MHCnuggets' utility in clinical settings, I applied MHCnuggets to large patient cohorts in The Cancer Genome Atlas (TCGA). Although these patients were not ICB-treated, HLA class I MANA and immunogenic mutation (IMM) landscape identified through MHCnuggets prediction showed positive correlation with lymphocyte infiltration in 6,613 patients from 26 tumor types. With these findings, I, next, applied MHCnuggets to ICB-treated cohorts to help shed light on mechanisms of patient response and aim to achieve more precise immunotherapy patient stratification based on their MANA profile.

1.3 Explore HLA class II immunogenic mutation burden as a predictor of ICB treatment response through a rank system extension of MHCnuggets.

Tumor mutational burden (TMB) has emerged as a biomarker for cancer treatments, particularly immune checkpoint blockade (ICB) treatments in solid tumors, but it remains an imperfect biomarker of response [25]. Nuanced approaches that focus on the quality rather than the quantity of mutations harbored in coding regions of the tumor genome are therefore needed to enhance the predictive value of TMB [26-31]. In the immunotherapy setting, previous efforts have focused on investigating the value of HLA class I neoepitope burden in predicting clinical outcomes with ICB [32-34]. Preclinical studies have shown that CD4+ T-cell neoepitope vaccination induced complete tumor rejection, tumor microenvironment (TME) re-shaping and

antigen spread [35]. Furthermore, tumor rejection in the context of immunotherapy has been shown to require the activity of both antigen specific CD8+ and CD4+ T-cells, suggesting the non-overlapping but complementary role of HLA class I and class II restricted neoantigens [36]. Here, I used computational approaches to assess HLA class II immunogenic mutation (IMM) burden in patients with melanoma and lung cancer treated with ICB.

First, I established immunogenicity criteria for predicted neoantigens. Rank systems of neoantigen binding affinities have previously been shown to overcome affinity distribution variabilities across HLA alleles and have achieved better binding prediction performance [37, 38]. Here, I extended the MHCnuggets framework by implementing an HLA affinity ranking system. Next, I applied this approach to define and to evaluate the role of HLA class II immunogenic mutation (IMM) burden in four independent cohorts of patients with non-small cell lung cancer (NSCLC) and melanoma treated with immunotherapy. I further integrated HLA class II IMMs with HLA germline variation, tumor clonal architecture and investigated the tumor microenvironment composition of tumors harboring a high class II IMM burden.

1.4 Sex-dimorphism assessed through ICB-treated patients' immunogenomic landscape powered by MHCnuggets' neoantigen analyses.

Several recent meta-analyses of immunotherapy clinical trials have suggested a decreased magnitude of benefit for ICB monotherapy in female patients as compared to male patients across multiple cancer types, including non-small cell lung cancer (NSCLC) [39-44]. A meta-analysis by Conforti et al. included six randomized trials with single- or dual-agent PD-1/PD-L1 inhibitors in NSCLC, including 3482 patients, and reported a greater magnitude of benefit in males with a pooled hazard ratio (HR) of 0.72 (95% CI 0.61-0.86) compared to a lack of

significant benefit in females with a pooled HR of 0.89 (95% CI (0.71-1.11) [39]. In all trials the HR for death was lower for males than females, and in three NSCLC trials [45-47] the HR demonstrated a significant benefit only for male and not for female patients. Interestingly, subsequent analysis has demonstrated that females have significantly increased magnitude of benefit from combinations of chemo-immunotherapy [48], and females experience higher rates of immune-related adverse events after immunotherapy for NSCLC and melanoma, as compared to males [49, 50]. Together, the data highlight the significant yet incompletely understood impact of sex on outcomes following ICB therapy in multiple tumor types and treatment settings. These findings have spurred interest in investigating the underlying mechanisms potentially mediating this sex dimorphism in therapeutic response in the context of immunotherapy.

Inherent differences in the male and female immune response are likely central to the observed sex dimorphism in ICB response. Females demonstrate increased humoral and cell-mediated responses to antigenic stimulation, vaccination, and infection, and there exists a strong female predominance of autoimmune diseases[51, 52]. Females demonstrate increased CD8+ T lymphocyte activity and IFNγ production as compared to males, and tumors in females demonstrate enrichment of nearly all T cell subpopulations in the tumor microenvironment [53, 54]. Furthermore, estrogen signaling has been demonstrated to modulate the tumor immune microenvironment, tumor antigen presentation, immune checkpoint expression, and intratumoral lymphocyte infiltration [55, 56]. Given these well-established differences in immune responses of males and females, I hypothesized that there are sex-based differences in tumor-intrinsic features that have been shaped by the immune system during tumor evolution and immunoediting, which are reflected in differential clinical outcomes following ICB therapy.

The therapeutic efficacy of ICBs relies on augmentation of the anti-tumor immune response, mediated by presentation of tumor-specific neoantigens primarily by major histocompatibility complex (MHC) class I proteins, along with multiple co-stimulatory signals, and resultant CD8+ effector cell activation. Recent studies have also shown the importance of accounting for HLA class II restricted neoantigens and IMMs when understanding efficacy of immune checkpoint blockade [36]. Additionally, the ability to present a mutation-associated neoantigen is dependent on HLA genetic variation, decreased HLA genetic variation may lead to immune escape and suboptimal responses to ICB therapy [25, 57].Taken together, I considered all the above factors in two non-small cell lung cancer (NSCLC) cohorts to assess sex-dimorphism in ICB response through intrinsic immunogenomic landscapes

Chapter 2 MHCnuggets

Here, I present a neoantigen predictor designed for MHC class I and II alleles in a single framework. The method leverages transfer learning and allele clustering to accommodate both common, well-characterized MHC alleles and rare, less studied alleles. Although existing computational neoantigen predictors generate a ranked list of candidate peptides, maximizing the number of predictions that identify immunogenic peptides would be preferred in many applications [18]. I demonstrated that MHCnuggets' predictive performance is competitive with widely used methods on binding affinity benchmark data sets. In comparison with hybrid methods that integrate binding affinity and HLAp data, MHCnuggets shows fewer false positives and increased positive predictive value (PPV) in a held-out cell line data set of ligands identified by mass spectrometry [22, 58]. To demonstrate the clinical utility and applicability of MHCnuggets to large patient cohorts, I investigated candidate immunogenic mutations from 26 tumor types in The Cancer Genome Atlas (TCGA). MHCnuggets yielded 101,326 predicted immunogenic missense mutations (IMM), observed in at least 1 individual (out of 1,124,266) in less than 2.3 hours. These mutations were correlated with increased lymphocyte infiltration; however, only 0.16% were observed in more than 2 patients. This chapter is based on published work in Shao *et al.* 2020 [31], further information and data are available at https://doi.org/10.1158/2326-6066.CIR-19-0464.

2.1 Computation implementation

2.1.1 Deep learning neural network architecture

MHCnuggets uses an LSTM neural network architecture [24](Fig. 2.1A). LSTM architectures excel at handling variable length sequence inputs and can learn long-term dependencies between noncontiguous elements, enabling an input encoding that does not require peptide shortening or splitting (Figure 2.1B). LSTMs are capable of handling peptides of any length. In practice, a maximum peptide length should be selected for network training. I set maximum peptide input length of 15 for class I and 30 for class II, for computational efficiency purposes. These values cover the majority of lengths observed in naturally presented MHC-bound peptides [12]. Peptides that are less than the maximum length are padded at the end with a character ("Z" which is not in the amino acid alphabet) until they reach the maximum length (Figure 2.1B). Peptides were represented to the network as a series of amino acids; each amino acid was represented as a 21-dimensional smoothed, one-hot encoded vector (0.9 and 0.005 replace 1 and 0, respectively).

In addition, MHCnuggets architectures can be trained using either continuous binding affinity measurements from in vitro experiments (half maximal affinity or IC50) and/or immunopeptidomic (HLAp) binary labels. For each MHC allele, I trained a neural network model consisting of an LSTM layer of 64 hidden units, a fully connected layer of 64 hidden units, and a final output layer of a single sigmoid unit (Figure 2.1A).

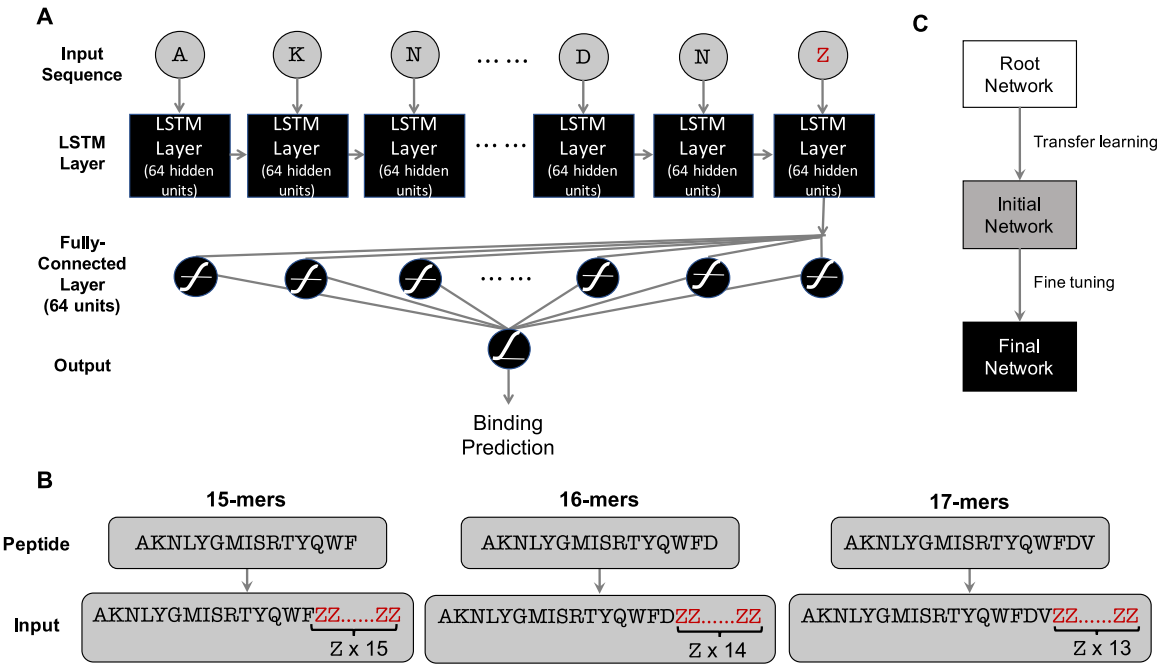

Figure 2.1 MHCnuggets architecture (A) A network is trained for each MHC allele. Each network has a LSTM layer with 64 hidden units, a Fully Connected (FC) layer with 64 hidden units and a final output layer of a single sigmoid unit. (B) Input scheme for peptides with variable lengths. MHCnuggets architecture is capable of handling peptides of any length, but in practice a maximum length should be selected. Peptides are extended with padding until they reach the maximum length, prior to input into the neural network. The example shows padding for class II peptides with maximum length set to 30 amino acids. (C) Transfer learning protocol for parameter sharing among alleles. A base allele-specific network is trained for each MHC class, with an allele selected by largest number of training examples. Transfer learning is applied to train networks for the remaining alleles with initial network weights set to final base network weights. A fine-tuning step identifies alleles that can be leveraged for a second round of transfer learning to produce a final network

2.1.2 Deep learning neural network training process

MSE loss L_{MSE} was used to train networks with continuous-valued binding affinity data and BCE loss L_{BCE} for binary HLAp data. For a data set with n samples,

$$L_{MSE}(\hat{y}, y) = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2$$

$$L_{BCE}(\hat{y}, y) = -\frac{1}{n}\sum_{i=1}^{n} y^{(i)} \log\left(\hat{y}^{(i)}\right) + (1 - y^{(i)}) \log\left(1 - \hat{y}^{(i)}\right)$$

All training used backpropagation with the Adam optimizer [59] and learning rate of 0.001. Regularization was performed with dropout and recurrent dropout [60] probabilities of 0.2. The number of hidden units, dropout rate, and number of training epochs were estimated by 3-fold

cross-validation on MHC class I HLA-A*02:01, a common allele with a large number of experimentally characterized binding peptides.

For the 16 alleles where allele-specific HLAp training data were available [61], I trained networks on both binding affinity and HLAp data (MHCnuggets). Next, I trained networks only with binding affinity measurements (MHCnuggets without mass spectrometry data or noMS) for all MHC class I alleles. Due to the lack of allele-specific HLAp training data for class II, all MHC class II networks were trained only on binding affinity measurements. In total, I trained 148 class I and 136 class II allele–specific networks. Common alleles comprise a small fraction (<1%) of all known MHC alleles [62]. To handle binding predictions for rare alleles, MHCnuggets selects a network by searching for the closest allele, based on previously published supertype clustering approaches. I prioritized approaches based on binding pocket biochemical similarity when available. Briefly, HLA-A and HLA-B alleles were clustered by MHC binding pocket amino acid residue composition [63], and HLA-C and all MHC II alleles were hierarchically clustered based upon experimental mass spectrometry and binding assay results [64, 65]. For alleles with no supertype classification, the closest allele was from the same HLA gene, and allele group if available, with preference for alleles with the largest number of characterized binding peptides. All networks were implemented with the Keras Python package (TensorFlow back-end) [66, 67]. Open-source software is available at https://github.com/KarchinLab/mhcnuggets, installable via pip or Docker, and has been integrated into the PepVacSeq [68], pvactools [69], and Neoepiscope [70] pipelines.

2.1.3 Transfer learning protocols

The networks were trained with transfer learning [71], which allows networks for less well-characterized alleles to leverage information from extensively studied alleles (Figure 2.1C). Transfer learning was also used to train networks combining binding affinity and HLAp data sets.

2.1.3.1 Transfer learning protocol for binding affinity data only

I used transfer learning to improve network learning for MHC alleles with limited characterized peptides available for training. I first trained base allele-specific networks for class I and class II, using alleles with the most training examples in IEDB (HLA-A*02:01 for class I and HLA-DRB1*01:01 for class II). For all other alleles, the final weights of the base network for its respective class were used to initialize network training, and then an allele-specific network was trained for each allele. Next, I assessed prediction performance of each allele-specific network on the training examples for each of the alleles. For each allele, if the network that performed best was not the HLA-A*02:01 network (for class I alleles) or HLA-DRB1*01:01 network (for class II alleles), I did a second round of training, with the best-performing network's weights used in the initialization step.

2.1.3.2 Transfer learning protocol for binding affinity and HLAp data

To integrate HLAp data into the class I networks, I initially trained each network with binding affinity data as described above, transferred the final weights to a new network, and then continued training with the HLAp data as positive examples augmented with random peptide decoys as negative examples.

2.1.4 Transformation of peptide binding affinities

Predicted binding affinity can be transformed into a range of values well suited for neural network learning by selecting a logarithmic base to match the weakest binding affinity of interest [72]. For most benchmarks in this work, I used the standard upper limit of 50,000 nmol/L, so that predicted binding affinity was

$$y = \max\,(0{,}1 - log_{50k}(IC_{50})$$

For the Bonsack and colleagues data set [8], the upper limit was changed to 100,000 nmol/L because in their experiments, as described in O'Donnell and colleagues [18], binders were defined as peptides with $IC_{50} < 100{,}000$ nmol/L. As binding affinity was determined based on in vitro HLA binding-competition versus a known strong binder (reported $IC_{50} < 50$ nmol/L) experimental IC_{50} values were in the micromolar range.

2.1.5 Performance metrics and selection of final network weights

PPV= NTP/(NTP + NFP), where NTP is the number of true positives and NFP is the number of false positives. I calculated PPV with respect to the top-ranked n peptides, where n is the number of true binders in the ranked list, denoted as PPVn. To minimize overfitting, network training was stopped after 100 epochs but if the best PPVn was reached earlier, network weights from that earlier epoch were used in the final network. Notably, although I chose to optimize the networks on PPVn, an alternative approach could optimize on area under the ROC curve (auROC), Kendall's tau, or Pearson r correlation. For the two alleles in the Immune Epitope Database (IEDB) with the most training examples in their respective class, HLA-A*02:01 for class I and HLA-DRB1*01:01 for class II, training was stopped after 200 epochs.

2.2 Data set collection and curation

2.2.1 Training data sets

The networks were trained with data from a curated version of the IEDB v.2018[73], containing chemical binding affinity measurements for 241,553 peptide–allele pairs covering 217 class I alleles and 96,211 peptide–allele pairs covering 135 class II alleles. Additional training data consisted of 16 class I mon-allelic B-cell line immunopeptidomes [61]. The immunopeptidome data are limited to HLAp binders and were supplemented by decoy random peptides sampled from the human proteome (https://data.mendeley.com/datasets/ 8pz43nvvxh/2).

2.2.2 Benchmark data sets

Kim and colleagues: This benchmark contained 53 MHC class I alleles and 137,654 IC50 measurements published prior to 2009 (training set) and 53 unique MHC class I alleles with 26,888 IC_{50} measurements, published from 2009 to 2013 (test set). Three alleles (HLA-B*27:03, HLA-B*38:01, and HLA-B*08:03) did not contain sufficient training data, and two alleles (HLA-A*46:01 and HLAB*27:03) did not contain any peptides defined as binders in this work (IC50 < 500 nmol/L). Therefore, a total of four alleles (HLA-A*46:01, HLA-B*27:03, HLA-B*38:01, and HLA-B*08:03) were dropped from the analysis. All peptides in this benchmark set consisted of 8 to 11 amino acid residues.

Bonsack and colleagues: This data set contains 475 synthetic peptides derived from model protein sequences HPV16 E6 and E7 tested for binding to 7 alleles (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, HLA-A*24:02, HLA-B*07:02, and HLA-B*15:01). Each peptide was tested in competition-based cellular binding assays with a known high-affinity fluorescein-labeled reference peptide. EBV transformed B-lymphoblastic cells were stripped of

their naturally bound peptides and mixed with serially diluted test peptides and 150 nmol/L of reference peptide. Each synthetic peptide was tested at 8 different concentrations ranging from 780 nmol/L to 100,000 nmol/L. Mixture fluorescence at each synthetic peptide concentration was measured with flow cytometry, and a nonlinear regression analysis was used to find the test peptide concentration that inhibited 50% of the reference peptide binding (IC50). Peptides were classified as binders (IC50≤100,000 nmol/L) or nonbinders (IC50>100,000 nmol/L). Peptides in this independent benchmark set do not have IEDB entries.

Bassani-Sternberg and colleagues, 2015: This data set contains 22,598 unique peptides eluted from 6 cell lines with multiallelic MHCs. Out of the total 6 cell lines, a total of 26 alleles were reported. For each multiallelic cell line, peptide/MHC pairs were found through deconvolution, following the protocol described by Abelin and colleagues [61], with the difference that I used MHCnuggets rather than NetMHCpan 2.8 [74]to predict IC50 values for each peptide–MHC pair. For each cell line, each peptide was initially assigned as a binder to all expressed alleles. Then, for each allele, I filtered out any peptide predicted to bind with IC50 > 1,000 nmol/L to that allele, and with IC50 < 150 nmol/L to any other allele. Peptides found for 6 alleles (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*31:01, and HLA-B*51:01) were selected for allele-specific prediction testing. Trained networks were available for these alleles from all the methods that I compared.

Trolle and colleagues: This data set contains 15,524 unique peptides eluted from soluble HLA (sHLA) transfected HeLa cells, a process that allowed for separating binding peptides to a single MHC allele. This data set reports peptides for 5 MHC alleles. Peptides found for 4 alleles (HLA-A*01:01, HLA-A*02:01, HLA-A24*02, and HLA-B*51:01) were selected for testing. Peptide lengths in this data set range from 8 to 15 amino acid residues.

BST: This benchmark consists of 23,971 HLAp hits for 6 alleles, from Bassani-Sternberg and colleagues [22] and Trolle and colleagues [58] plus 23,947,029 random decoy peptides sampled from the human proteome. Any peptides found to overlap with the training HLAp data [61] were removed.

Jensen and colleagues: This benchmark was designed to assess both allele-specific and rare MHC class II binding affinity predictors. Allele specific prediction was tested with a 5-fold cross-validation experiment on peptides found in IEDB in 2016 but not 2013. Rare allele predictions were tested with the LOMO protocol.

IEDB class I rare alleles: This data set was designed to apply the LOMO protocol to class I alleles. It included 20 "pseudo-rare" alleles with 30 to 100 binding affinity peptide measurements in IEDB.

All data sets used in this work are available at http://dx.doi.org/ 10.17632/8c26kkrfpr.2.

2.3 Evaluation

2.3.1 Model assessment strategy

To accurately assess the performance of MHCnuggets on a variety of MHC–peptide binding prediction tasks, I utilized six benchmark sets: MHC class I alleles, MHC class II alleles, common alleles with a trained model (allele-specific prediction), and rare alleles (pan-allele prediction; Figure 2.2). To compare to the HLA ligand prediction tools from the NetMHC group (NetMHC 3.0, NetMHC 4.0, NetMHCpan 2.0, NetMHCpan 4.0) [16, 17], which can be trained only by their developers, as well as the open-source MHCflurry tools [18], I used multiple benchmarking strategies: (i) independent benchmark test set of peptides not included as training data for any of the methods; (ii) a previously published paired training/testing benchmark;

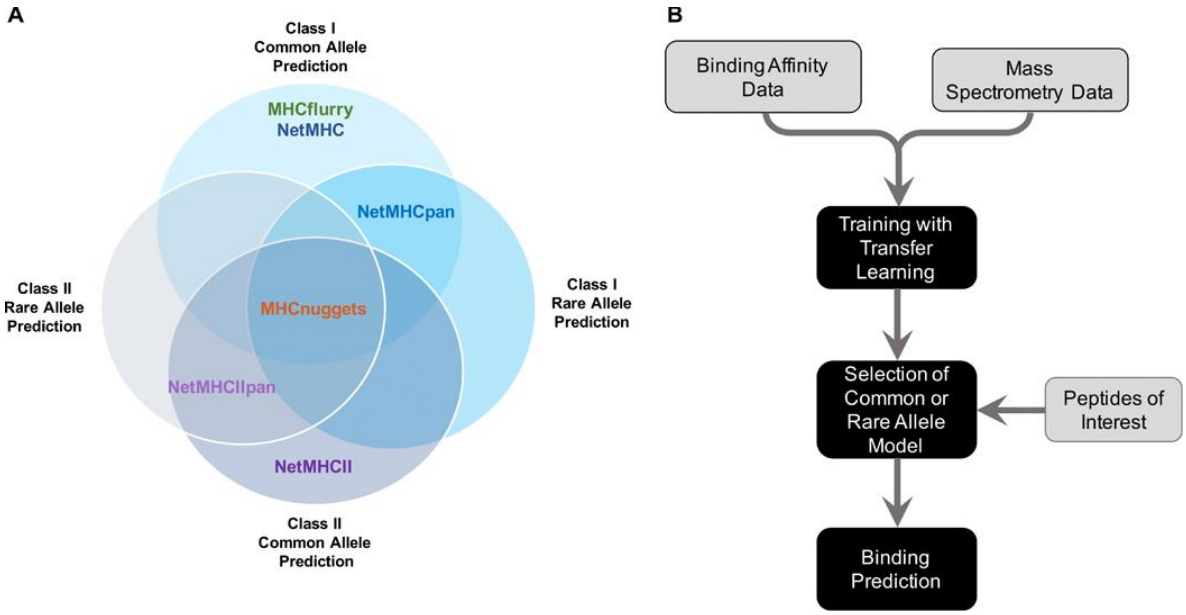

Figure 2.2 MHCnuggets' features. (A) Venn diagram representation of the MHC-peptide binding prediction functions of MHCnuggets and similar tools. (B) Training and MHC allele model selection scheme for MHCnuggets.

(iii) 5-fold cross-validation benchmark; and (iv) leave-one-molecule-out (LOMO) benchmark.

I evaluated six MHC class I predictors on independent binding affinity and HLAp data sets [8, 22, 58]. First, I compared MHCnuggets to several class I predictors that incorporate both binding affinity and HLAp data: MHCflurry 1.2.0, MHCflurry (train-MS), NetMHC 4.0, and NetMHCpan 4.0. Each method was benchmarked using an independent set of MHC-bound peptides identified by mass spectrometry across seven cell lines for six MHC I alleles. For testing, HLAp hits were combined with random decoy peptides sampled from the human proteome in a 1:999 hit–decoy ratio, as described by Abelin and colleagues [61], totaling 23,971,000 peptides. Next, four MHC class I predictors trained only on binding affinity data [MHCnuggets (noMS) and MHCflurry (noMS), NetMHC 3.0 and NetMHCpan 2.0] were evaluated with the Kim and colleagues data set [5], in which each predictor was trained with the BD2009 data and tested on BLIND data. It was possible to compare NetMHC 3.0 and NetMHCpan 2.0 performance on Kim and colleagues, because they have previously published

predicted IC50 values for all peptide–MHC pairs in BLIND. This allowed us to calculate their PPVn, auROC, Kendall's tau, and Pearson r correlations.

Next, I compared MHCnuggets' class II ligand prediction performance with self-reported performance statistics of NetMHC group's MHC class II methods [75]. I used the Jensen and colleagues 5-fold cross-validation benchmark to assess allele specific MHC class II prediction of MHCnuggets and NetMHCII 2.3, for 27 alleles. NetMHCII 2.3 reported the average auROC for five-fold cross-validation, and I report MHCnuggets' PPV for each of the 27 alleles as well as the average auROC, Pearson r, and Kendall's tau correlations.

The LOMO benchmarks are a type of cross-validation designed to estimate the performance of peptide binding prediction with respect to rare MHC alleles. Given training data for n MHC alleles, the data for a single allele are held out and networks are trained for the remaining n 1 alleles. Then for each peptide, predictions were generated by the remaining networks. I designed a LOMO benchmark to evaluate MHC class I rare allele prediction, by selecting 20 alleles with 30 to 100 characterized peptides in the IEDB [73]. For class II rare allele prediction, I used the Jensen and colleagues [75] LOMO benchmark. I was unable to assess rare allele prediction for NetMHC class I methods, as no published results were available. For the NetMHC class II methods, I compared MHCnuggets to their self-reported auROCs.

2.3.2 High-throughput MHCnuggets breaks the MHC ligand prediction plateau

To assess the baseline performance of MHCnuggets allele-specific networks on binding affinity data, I compared my approach with widely used MHC class I ligand prediction methods using two validation sets of binding affinity measurements [5, 8]. I trained and tested

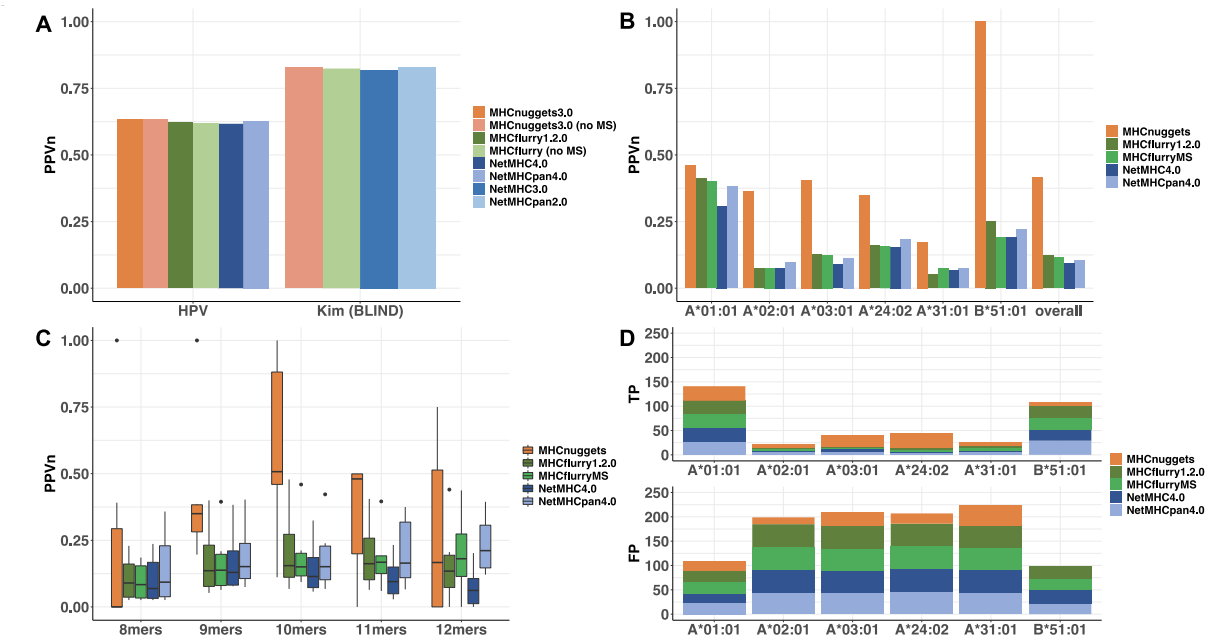

Figure 2.3 MHC class I benchmark comparisons. (A) PPVn for MHC class I allele-specific prediction on binding affinity test sets from Bonsack et al. (7 alleles) and Kim et al. (53 alleles) (5,8) (B) PPVn for MHC class I allele-specific prediction on HLAp BST data set (Bassani-Sternberg et al. and Trolle et al. (7,22)), stratified by allele (6 alleles). (C) PPVn for MHC class I allele-specific prediction on HLAp BST data set (from B) stratified by peptide sequence length. (D) True and false positives for each method on the top 50 ranked peptides from the HLAp BST data set. PPVn = positive predictive value on the top n ranked peptides, where n is the number of true binders. TP=true positives. FP=false positives.

MHCnuggets (noMS) and MHCflurry (noMS) using the Kim and colleagues data set [5] and evaluated the predictions provided by NetMHC 3.0 and NetMHCpan 2.0. I observed that MHCnuggets' performance (PPVn=0.829, auROC=0.924) was comparable with these methods (Figure 2.3A; PPVn of all methods=0.825±0.005, auROC of all methods=0.928±0.0031). MHCnuggets was also comparable (PPVn=0.633, auROC=0.794) to these methods when tested on the Bonsack and colleagues data set [8](PPVn of all methods=0.625±0.008, auROC of all methods=0.77±0.02; Figure 2.3A).

Earlier neoantigen prediction methods focused on class I and trained on binding affinity data from IEDB [76]. More recent work incorporated both binding affinity and HLAp data into network training [14, 18]. I compared MHCnuggets with several class I predictors that used both binding affinity and HLAp data: MHCflurry 1.2.0, MHCflurry (train-MS), NetMHC 4.0, and

NetMHCpan 4.0. I selected the BST HLAp data set [22, 58, 61] as an independent benchmark, as it was not used as training data by any of these methods. For all alleles tested, MHCnuggets achieved an overall PPVn of 0.42 and auROC of 0.82 (Figure 2.3B). On average, MHCnuggets' PPVn was more than three times higher than MHCflurry 1.2.0, MHCflurry (train-MS), NetMHC 4.0, and NetMHCpan 4.0. For all alleles, MHCnuggets predicted fewer binders than other methods, resulting in fewer false-positive predictions. Stratifying by peptide length, MHCnuggets' increased PPVn was most evident for peptides of length 9, 10, and 11 (Figure 2.3C). The length distribution of predicted binders was commensurate with the observed distribution of naturally occurring binders in the HLAp benchmark tests [58].

For some clinical applications, it may be desirable to minimize the number of false positives among a small number of top-scored peptides. I compared PPV of the methods listed above on their top 50 and 500 ranked peptides from the BST data set (six MHC class I alleles). MHCnuggets exhibited the highest PPV in the top 50 for all alleles except HLA-B*51:01 and the highest PPV in the top 500 for all alleles (Figure 2.3D).

2.3.3 Prediction of peptide–MHC binding for class II alleles

I assessed the baseline performance of MHCnuggets class II allele–specific networks on binding affinity data. To enable comparison with the class II methods from the NetMHC group, I used a 5-fold cross-validation benchmark derived from IEDB that was included in the publication describing NetMHCII 2.3 and NetMHCIIpan 3.2 [75]. First, I computed PPVn for each of the 27 allele-specific networks separately (Figure 2.4A; mean PPVn=0.739). Next, I computed the overall auROC, Pearson r, and Kendall's tau correlations for all 27 class II alleles.

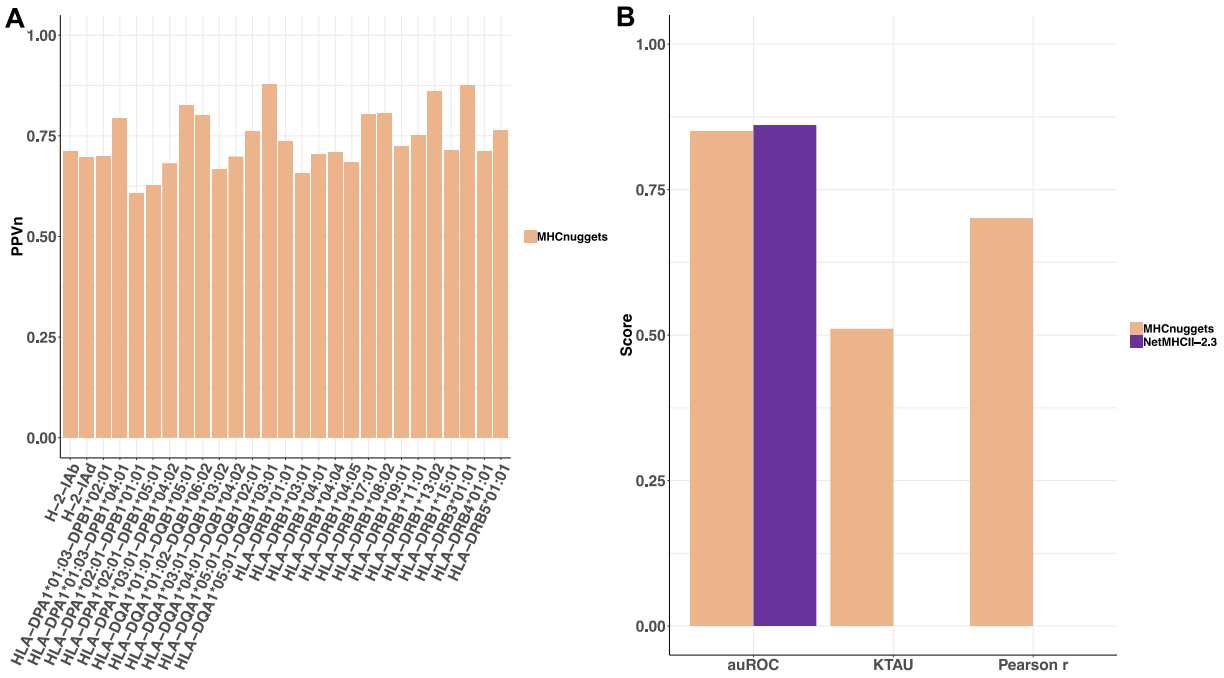

Figure 2.4 MHC class II benchmark comparisons. (A) PPVn for MHC class II allele-specific prediction on binding affinity test set from Jensen et al. (27 alleles, stratified by allele). (B) auROC, K-Tau, Pearson r scores for MHC class II alleles from five-fold cross-validation. NetMHCII2.3 performance is from their self-reported auROC. auROC= area under the receiving operator characteristic curve. K-Tau = Kendall's tau correlation. PPVn = positive predictive value on the top n ranked peptides, where n is the number of true binders.

MHCnuggets overall auROC (0.849) was comparable with that of the NetMHCII 2.3 (0.861) and NetMHCIIpan 3.2 (0.861). Comparison with NetMHC class II methods was limited to overall auROC as published in Jensen and colleagues [75], because their PPVn results are not publicly available (Figure 2.4B).

2.3.4 Prediction of peptide–MHC binding for rare alleles

I estimated performance of those class I and II MHC alleles for which I was unable to train allele-specific networks, using LOMO cross-validation [75]. In this LOMO protocol, MHC-peptide binding was assessed for a well-characterized allele that has been held out from training to approximate prediction performance for a rare allele (Figure 2.5A). For the 20 class I alleles, the mean PPVn was 0.65, and the mean auROC was 0.671. For the 27 class II alleles, the mean

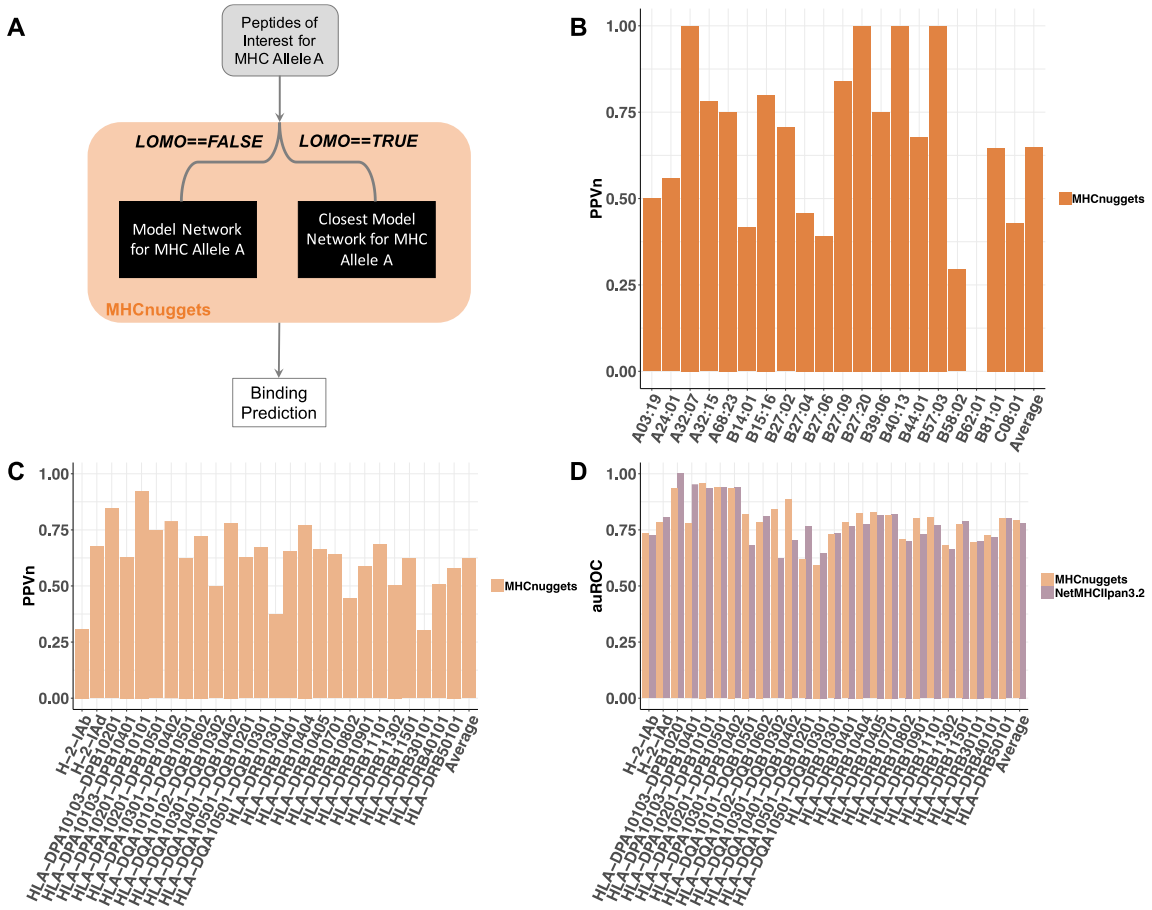

Figure 2.5 MHC class I and II benchmark comparisons to estimate rare allele performance. (A) Schematic representation of leave one molecule out (LOMO) testing. (B) PPVn for MHC class I rare allele prediction on IEDB pseudo-rare alleles binding affinity test set (20 alleles, stratified by allele). (C) PPVn for MHC class II rare allele prediction on binding affinity test set from Jensen et al. (27 alleles, stratified by allele) (39). (D) auROC for MHC class II rare allele prediction on LOMO binding affinity test set from Jensen et al. (27 alleles, stratified by allele) (39). NetMHCIIpan3.2 results are from their self-reported auROC. auROC = area under the receiving operator characteristic curve. PPVn = positive predictive value on the top n ranked peptides, where n is the number of true binders.

PPVn was 0.65 and the mean auROC was 0.792. In comparison, the class II mean auROC of NetMHCIIpan 3.2 was 0.781 (Figure 2.5B–D). NetMHCpan rare allele predictor LOMO test results for class I are not publicly available; therefore, I was unable to compare with them.

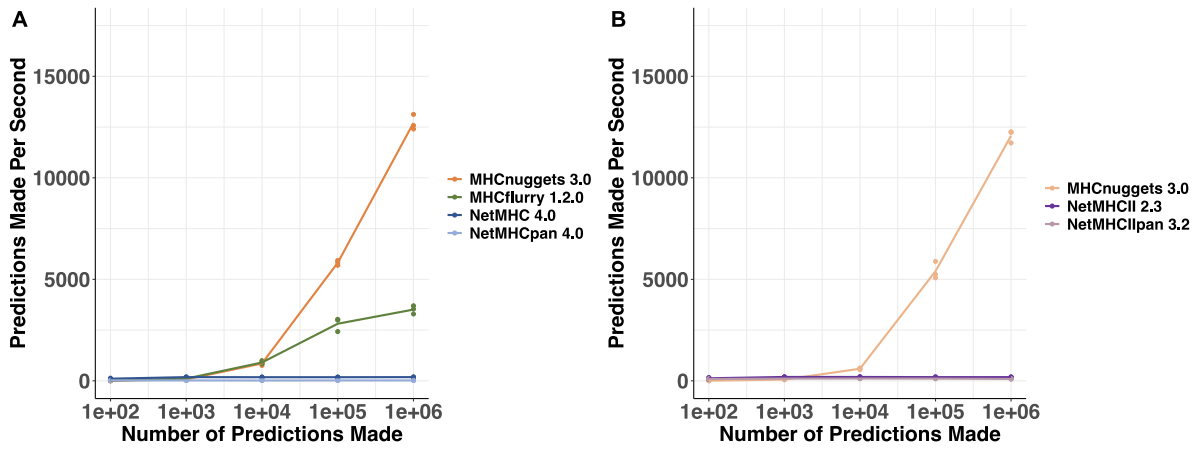

Figure 2.6 Timing and scalability. Runtime benchmark of tested methods using versions available on October 1, 2019 over a range of inputs (up to 1 million peptides). (A) MHC class I prediction. (B) MHC class II prediction

2.4 Runtime analysis

To assess the speed and scalability of the tested methods, I selected 1 million peptides sampled from the Abelin and colleagues data set [61] for class I alleles, and one million peptides sampled from the IEDB (curated data set 2018)[73] for class II alleles. Sampling was done with replacement. For each method listed in Figure 2.1A, networks for three class I MHC alleles (HLA-A*02:01, HLA-A*02:07, and HLA-A*01:01) and three class II MHC alleles (HLA-DRB1*01:01, HLA-DRB1*11:01, and HLA-DRB1*04:01) were used to predict binding over a range of input sample sizes (10^2, 10^3, 10^4, 10^5, and 10^6). All methods were run on a single graphics processing unit (GPU) compute node (one NVIDIA TESLA K80 GPU plus six 2.50 GHz Intel Xeon E5- 2680v3 CPUs, 20 GB memory).

When run on a GPU architecture, MHCnuggets was faster and scaled more efficiently than MHC ligand predictors from the NetMHC family and MHCflurry. Given an input of one million peptides randomly selected from Abelin and colleagues [61], MHCnuggets runtime was 3.62, 69.7, and 624.5 times faster than MHCflurry 1.2.0, NetMHC 4.0, NetMHCpan 4.0, respectively (Figure 2.6A). The improvement was similar for class II peptides, for which an

input of one million peptides to MHCnuggets ran 65.6 times and 126 times faster than NetMHCII 2.3 and NetMHCIIpan 3.2, respectively (Figure 2.6B). As the total number of input peptides was increased from 0 to 1 million, the runtime per peptide plateaued for other methods but decreased exponentially for MHCnuggets.

2.5 Application of MHCnuggets on The Cancer Genome Atlas (TCGA) patients

To assess candidate immunogenic somatic mutations in patients from the The Cancer Genome Atlas (TCGA) cohort, I developed and implemented a basic pipeline based on whole-exome and RNA-seq data (Figure 2.7). My analysis builds upon work from the TCGA PanCancer Analysis teams for drivers [77], mutation calling [78], and cancer immune landscapes [79]. I obtained somatic mutation calls for all cancer types from Multicenter Mutation Calling in Multiple Cancers (MC3; v0.2.8; 7,775 patients). Tumor-specific RNA expression values from Broad TCGA Firehose were standardized across tumor types using the RSEM Z-score [80]. MHC allele calls were obtained from the TCGA cancer immune landscape publication, in which up to 6 MHC class I alleles (HLA-A, HLA-B, and HLA-C) were identified for each patient using OptiType [81]. I included patients for which mutation calls, MHC allele calls, and RNA expression values were available from TCGA. After these considerations, the analysis included 6,613 patients from 26 TCGA tumor types. Six cancer types were not included in my analysis, because 15 or fewer patients met this requirement: lymphoid neoplasm diffuse large B-cell lymphoma, esophageal carcinoma, mesothelioma, skin cutaneous melanoma, stomach adenocarcinoma, and ovarian serous cystadenocarcinoma.

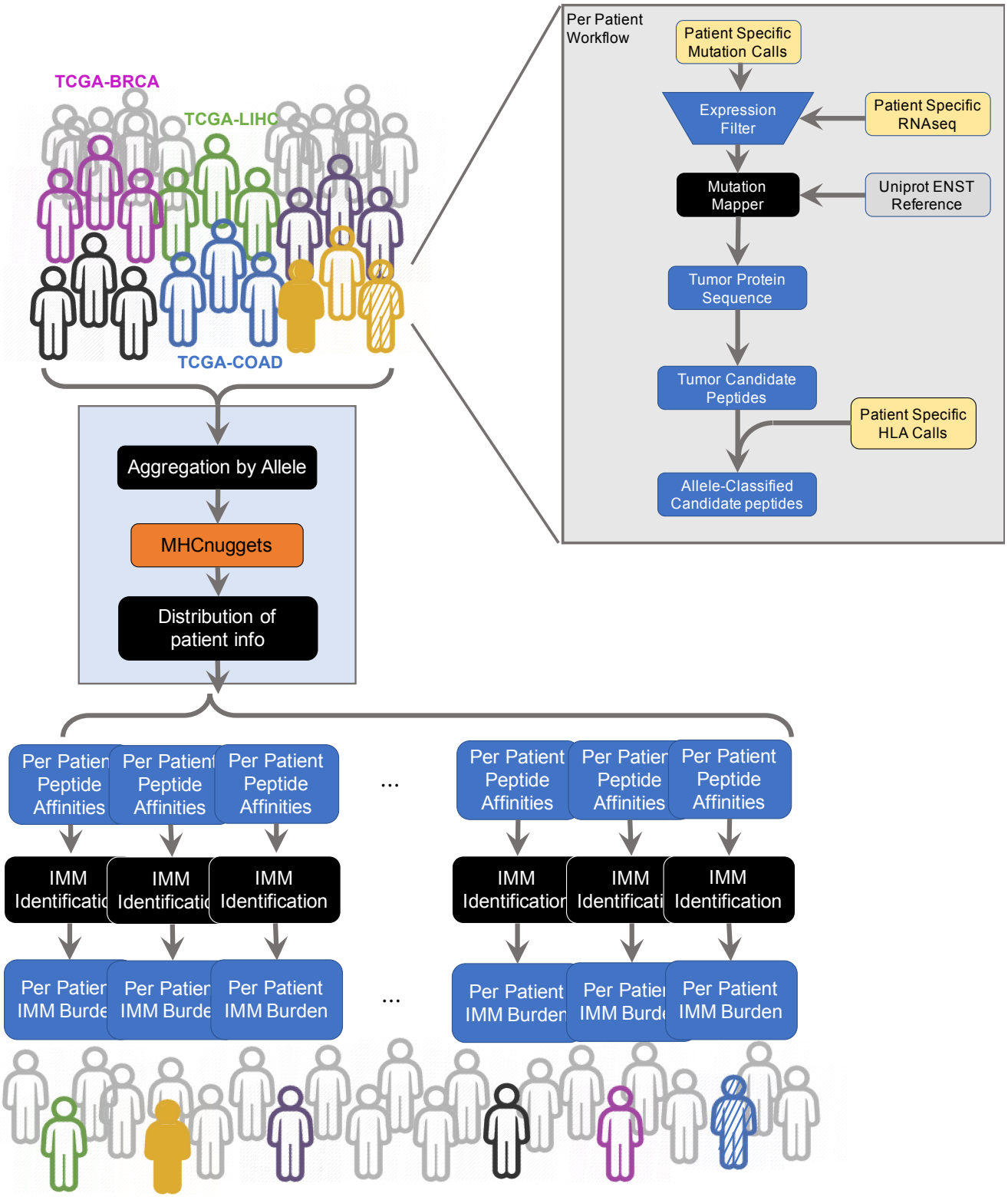

Figure 2.7 TCGA analyses workflow

The somatic missense mutations identified in each patient were filtered to include only those with strong evidence of mutant gene RNA expression in that patient ($Z > 1.0$). For each mutation that passed this filter, I used the transcript assigned by MC3 to pull flanking amino acid residues from the SwissProt database [82], yielding a 21-amino acid residue sequence fragment centered at the mutated residue. All candidate peptides of length 8, 9, 10, and 11 that included the mutated residue were extracted from each sequence fragment. Next binding affinity predictions were generated for each mutated peptide for up to six MHC class I alleles, depending on the patient's HLA genotypes. In total, each somatic mutation was represented by 38 mutated peptides for up to six possible MHC pairings.

I applied a permissive filter to select candidate immunogenic peptides, requiring mutated peptides to have binding affinity of $IC50 < 500$ nmol/L for at least one MHC allele. Somatic missense mutations that generated neoantigens meeting these criteria were considered predicted IMMs. For a given patient, if a mutation was predicted to be a predicted IMM for multiple alleles, it was counted only once using the MHC allele with the lowest predicted IC50. Finally, for each patient, I counted the number of predicted IMMs found in their exome and stratified by tumor type. I then identified predicted IMMs that were harbored by more than one patient.

I identified 101,326 unique predicted IMMs in 26 TCGA cancer types, with a mean of 15.6 per patient. I found that the majority of patients harbored fewer than six predicted IMMs, and 197 patients had none. Seventy-two percent of patients had from one to 10 predicted IMMs, compared with 1.9% of patients with more than 100, and 9 patients with more than 1,000 (Figure 2.8A). Cancer types with the highest number of predicted IMMs were uterine corpus endometrial carcinoma (UCEC), colon adenocarcinoma (COAD), and lung adenocarcinoma (LUAD), all

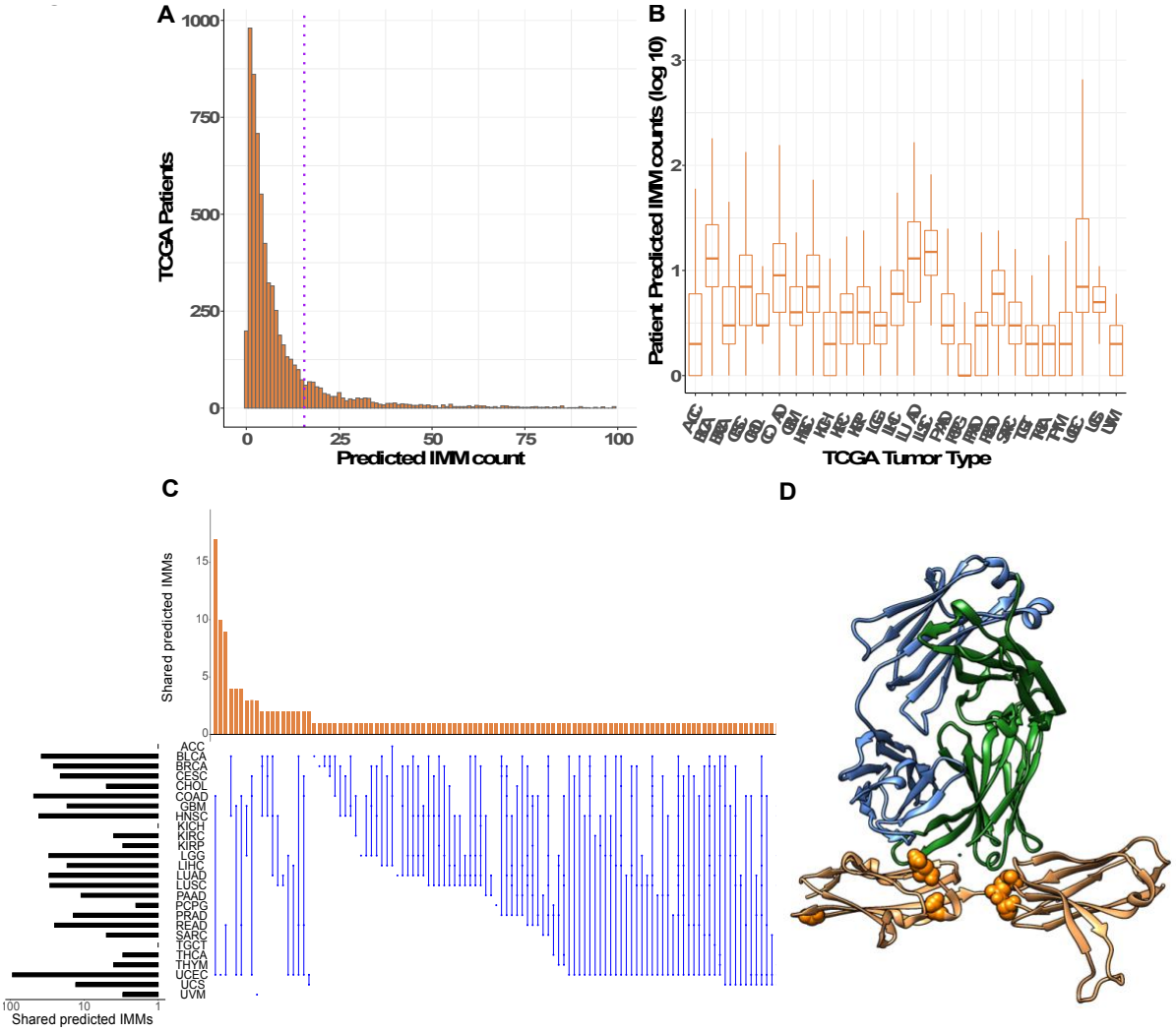

Figure 2.8 MHC class I IMMs in TCGA patients. (A). Number of predicted immunogenic missense mutations (IMMs) identified in 6,613 TCGA patients. Dotted line = mean IMMs per patient (15.6). Note, 123 patients had >100 predicted IMMs but are not included for visual clarity. (B) Number of predicted IMMs by cancer type. (C) IMMs shared by three or more patients and the cancer types in which they occurred. Each row represents a cancer type and each column illustrates the overlap of IMMs seen in a single cancer type or multiple cancer types. For example, the first column shows the number of IMMs shared among patients with colorectal adenocarcinoma (COAD) and uterine corpus endometrial carcinoma (UCEC). Bars to the left show the total number of unique IMMs in each cancer type. *Bar heights reflect count of unique shared IMMs, not total number of patients in which the IMM was observed. Cancer type abbreviations are in Methods. Image generated with UpSetR. (D) Fibroblast growth factor receptor (FGFR3) IMM hot region identified by HotMAPs in bladder cancer (BLCA). IMMs shown and number of BLCA patients with the IMM: p.E216K (1), p.D222N (1), p.G235D (1) p.R248C (3) and p.S249C (24). Except for p.G235D, these IMMs are proximal to the interface of FGFR3 protein and the light and heavy chains of an antibody fragment designed for therapeutic application in bladder cancer (PDB ID: 3GRW) (61). ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC; kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma.

three of which are known for high mutation burden and immunogenicity [79]. UCEC and COAD are also known to have a high frequency of microsatellite instable (MSI) tumors. The lowest number were found in uveal melanoma (UVM), paraganglioma and pheochromocytoma (PCPG), and testicular germ cell tumors (TGCT; Figure 2.8B).

Across all cancer types, I identified 1,393 predicted IMMs harbored by 2 or more patients, of which 167 were identified in 3 or more patients. Of these, 167 only 11.5% occurred exclusively in a single cancer type (Figure 2.8C). The predicted IMMs identified in the largest number of patients were *IDH1* R132H (62), *FGFR3* S249C (24), *PIK3CA* E545K (23), *KRAS* G12D (18), *PIK3CA* E542K (18), *TP53* R175H (18), *TP53* R248Q (18), *TP53* R273C (17), and *KRAS* G12V (16), which are known recurrent oncogenic driver mutations [83, 84]. Of the 1,071 genes harboring predicted IMMs in 2 or more patients, the ones containing the most included *TP53* (68), *CTNNB1* (18), *PIK3CA* (16), *HRAS* (8), *KRAS* (7), *PTEN* (7), *FBXW7* (6), *EGFR* (5), *MDN1* (5), *POLE* (5), *TRRAP* (5), and *VPS13C* (5; Supplementary Table S9C). Six missense mutations harbored by patients in the TCGA cohorts were previously validated by $CD8^+$ T-cell response assays [26, 27, 85]. Of the six missense mutations, *TP53* R248Q, *TP53* Y220C, *TP53* R175H, *TP53* R248W, and *KRAS* G12D were predicted to be IMMs by my MHCnuggets pipeline and were shared by 3 or more of the TCGA patients.

I considered that mutation immunogenicity might be associated with potential driver status of a mutation. Driver status was inferred by CHASMplus [86], a random forest classifier that utilizes a multifaceted feature set to predict driver missense mutations and is effective at identifying both common and rare driver mutations. For each mutation, immunogenicity was represented as a binary response variable and driver status was used as a covariate. Mutations with CHASMplus q-value < 0.01 were considered drivers [86]. 61.7% of the 167 predicted

IMMs shared by 3 or more patients were classified as driver missense mutations by CHASMplus ($q < 0.01$). This percentage is significantly higher than the number of predicted drivers among all TCGA missense mutations (9,821 out of 791,637 or 1.2%). Although many shared IMMs were predicted to be driver missense mutations, the percentage of predicted IMMs predicted to be drivers was ~0.1% of total predicted IMMs in my study. When compared with the OncoKB database of experimentally confirmed driver mutations [87], 53.9% of the shared predicted IMMs identified as "oncogenic" or "likely oncogenic" driver mutations. The percentage is lower (25.7%) if "likely oncogenic" mutations are excluded. I, next, explored the relationship between mutation driver status predicted by CHASMplus and predicted IMM status using logistic regression (R glm package with binomial link logit function). The log odds of being a predicted IMM was significantly decreased for drivers (β=-0.66, Wald test $P < 2e\text{-}16$), which is consistent with previous work suggesting that negative evolutionary selection eliminates MHC class I immunogenic oncogenic mutations early in tumor development [88].

Although I observed a limited number of shared predicted IMMs, I reasoned that protein regions enriched for predicted IMMs could present a therapeutic opportunity in certain cancer types. I then sought to ascertain whether predicted IMMs occurred preferentially in particular gene or protein regions. Using the HotMAPS 1D algorithm v1.2.2 [89], I clustered primary amino acid residue sequence to identify regions where mutations were frequently predicted as IMM, with statistical significance ($q < 0.01$, Benjamini– Hochberg method) [90]. In this analysis, mutations were stratified by cancer type, and I considered enrichment within linear regions of 50 amino acid residues. I identified clusters of residues within protein regions having statistically significant enrichment of predicted IMMs ($q < 0.01$). These included CIC in low-grade glioma (LGG); *NFE2L2* and *FGFR3* (Figure 2.8D) in bladder cancer (BLCA) [91]; *KRAS*

in pancreatic adenocarcinoma (PAAD); *KIT* in TGCTs; *HRAS* in head and neck squamous carcinoma (HNSC); *PTEN*, *POLE*, and *PPP2R1A* in UCEC; and *GNAQ* and *SF3B1* in UVM. Three genes harbored predicted immunogenic regions in more than one cancer type: *TP53* in BLCA, BRCA, HNSC, LGG, and UCEC; *PIK3CA* in HNSC and cervical squamous cell carcinoma (CESC); and CTNNB1 in liver hepatocellular carcinoma (LIHC) and UCEC.

To assess whether the total number of predicted IMMs per patient was associated with changes in tumor immune infiltrates, I performed Poisson regression (R glm package with Poisson link log function). All estimates of immune infiltrates were obtained from Thorsson and colleagues [79, 92]. I fit two univariate models in which the response variable was the predicted IMM count and the covariate was either total leukocyte fraction or fraction of $CD8^+$ T cells. The fitted coefficient $\beta = 0.75$ ($P < 2\times 10^{-16}$, Wald test) indicated that increased predicted IMM load was significantly associated with increased leukocyte fraction in a patient's cancer. In a second model, X was the proportion of $CD8^+$ T cells inferred by CIBERSORT (53). The fitted coefficient $\beta = 5.9$ ($P < 2\times 10^{-16}$, Wald test) indicated that increased predicted IMM load was associated with increased tumor infiltrating $CD8^+$ T cells. Total lymphocyte and (Aggregate3) $CD8^+$ T-cell fractions were estimated in Thorsson and colleagues [79].

Overall, predicted IMMs were almost exclusively private to individual TCGA patients, with only 1,393 predicted IMMs observed in more than 1 patient. Although more than 61% of predicted IMMs shared by more than 2 patients were predicted to be driver mutations, the log odds of immunogenicity decreased for predicted driver mutations, indicating immunogenicity might shape the driver mutation landscape. These findings suggest that IMMs drive tumor immunoediting and may be informative for the interpretation of clinical responses to immunotherapy

Chapter 3 HLA class II immunogenic mutation burden as a predictor of ICB treatment response.

Here, the utility of MHCnuggets in ICB treatment is explored. While tumor mutation burden (TMB) has been the most accepted marker in patient response prediction, I aimed to use MHCnuggets to explore a more biologically relevant but less studied maker—HLA class II immunogenic mutation (IMM). I first extended the MHCnuggets framework by implementing an HLA affinity ranking system. Candidate HLA class II-restricted neoantigens' binding affinities were ranked against those of 100,000 randomly selected human proteome peptides. Immunogenic peptides were defined based on their ranks, and HLA class II IMMs were identified based on their harboring of immunogenic peptides. Next, I evaluated the role of HLA class II immunogenic mutation (IMM) burden in four independent cohorts of patients treated with immunotherapy. I further integrated HLA class II IMMs with HLA germline variation, tumor clonal architecture and investigated the tumor microenvironment composition of tumors harboring a high class II IMM burden. This chapter is based on published work in Shao *et al.* 2022 [93], further information and data are available at https://doi.org/10.1016/j.annonc.2022.03.013.

3.1 MHCnuggets framework rank system extension

3.1.1 Rank algorithm implementation

MHCnuggets' rank prediction algorithm works as follows. When presented with a peptide-HLA pair, it first predicts the binding IC50 of the candidate peptide to the selected HLA. This IC50 is then ranked amongst the predicted IC50s of peptides sampled from the human proteome (HP) to the HLA of interest. These HP IC50 lists, consisting of lists of predicted IC50s of HP peptides against all supported HLAs, were generated as follows.

HP peptides were selected by sampling 100,000 peptides at each common peptide length for the respective HLA class (lengths 8-12 for HLA class I, 12 – 20 for HLA class II). Duplicates were removed, yielding 494,297 peptides in the class I list and 890,671 peptides in the class II list. I predicted the binding IC50 for every peptide to every supported HLA. These HLA-specific IC50 lists are sorted in ascending order. For fast, consistent querying of these lists, I have created accompanying "position dictionaries". These are HLA-specific maps from each unique IC50 to a tuple of the IC50's start and end index within the HP IC50 list. For example, if the first two elements of the HP IC50 list for HLA-A01:01 were 1.4 nM and 1.5nM, the associated entries in the position dictionary would map from 1.4 to (0, 0), and 1.5 to (1, 1) as 1.4 nM starts and ends at position 0, and 1.5 nM starts and ends at position 1.

To reduce the download size of MHCnuggets, the HP IC50 lists are stored in two parts under one pickle, as follows. HP IC50 lists are stored as pickles, with each HLA in the final IC50 pickle having two IC50 lists. The first list is the full first percentile (tightest predicted binders) of the predicted IC50s. This list is used to rank the tightest predicted binders, ensuring high accuracy in ranking candidate peptides that may be immunogenic. The second list is a downsampled version of the HP IC50 lists. Every 25th value of the original class I peptide list

and every 60th value of the class peptide II list was selected to form the downsampled lists. Each HLA has two position dictionaries corresponding to the two IC50 lists. The downsampling was performed to reduce the download size of MHCnuggets and make the package conveniently accessible with GitHub's standard size restrictions (each file < 100 MB). Additionally, I stored the first percentile IC50 value for each HLA as well as the original HP IC50 list length. These values were used in the prediction logic, described below.

When presented with a peptide-HLA pair, I first predicted the binding IC50 of the candidate peptide to the selected HLA. I then compared the candidate peptide's IC50 to the first percentile IC50 value of the HP IC50 list of the same HLA. If the candidate peptide's IC50 was less than or equal to the first percentile value, I performed a binary search of the candidate IC50 within the first percentile HP IC50 list of the HLA of interest. If an exact match was found, I queried the position dictionary to find the start and end index of the IC50, place the candidate peptide's IC50 in the middle of these indices, then computed the rank percentile using this position and the stored original list length of the HP IC50 list. This ensured that percentiles were always computed consistently, with the candidate IC50 in the center of any block of matching IC50s within the HP IC50 list. If no exact match was found, the percentile was calculated using the usual output position of the binary search. Alternatively, if the candidate peptide's IC50 was greater than the first percentile IC50 value of the HP IC50 list of the HLA of interest, the above process was carried out using the downsampled IC50 list rather than the first percentile IC50 list.

Source	Number of Unique Peptides	Number of Source Proteins	Study focus	Rate of presentation
Bassani-sternberg 2016 [94]	15009	2832	All	0.00791
Loffler 2018 [95]	10099	2533	CRC	0.00595
Loffler 2018 [95]	10732	2188	NMC	0.00733
Mommen 2016 [96]	13918	1980	HLA-DR	0.01050
Mutschlechner 2010 [97]	1606	650	HLA-DR	0.00369
Schuster 2017 [98]	17334	3544	Ovarian cancer	0.00730
Strug 2009 [99]	165	106	HLA-DR1	0.00232

Table 3.1 Studies used for HLA class II rate of presentation estimation. CRC=colorectal cancer cell line; NMC=non-malignant colon tissue.

3.1.2 Rank threshold selection

I tried to select a rank threshold that would give a close estimate of the HLA presentation rate when selecting peptides from a pool of human proteins. To do this, I first compiled a list of 6 studies that utilized mass-spectrometry to elute naturally presented peptides on HLA class II molecules. From these studies, the total number of unique peptides and the total number of their source proteins were compiled (Table 3.1).

Then, I computed the number of possible peptides that each of these source proteins can generate (eq1), mimicking the random human peptide generation process of my background list.

$$Total\ Number\ of\ Peptides\ of\ Length\ L\ a\ Protein\ Generates = (Length\ of\ Protein - L) \times Fraction\ of\ Unique\ Peptides \quad (eq1)$$

For the length of protein, I empirically calculated the median length of all the proteins in the Human Proteome (UniProt download 2019) to be 481 amino acids. For my analyses of HLA class II epitopes, I was interested in peptides with length 12-20 amino acids. As such, the calculation of total peptides following eq1 was repeated for each length of 12-20. I also empirically calculated the fraction of unique peptides to be 0.16 by extracting and uniquifying all possible peptide sequence of length 12-20 in the human proteome. Next, I utilized the compiled numbers of source proteins and the total peptides a protein can generate to get the total peptides arose from the source protein pools (eq2).

$$\begin{aligned} &Number\ of\ Total\ Peptides \\ &= Total\ Number\ of\ Peptides\ a\ Protein\ Generates \\ &\times\ Total\ number\ of\ Proteins \quad (eq2) \end{aligned}$$

Finally, I used the number of unique peptides eluted in combination of the total number of peptides from the source protein pool to calculate the presentation rate (eq3).

$$Presentation\ Rate = \frac{Number\ of\ Unique\ Peptides}{Number\ of\ Total\ Peptides\ from\ Source\ Proteins} \quad (eq3)$$

Through this simulation, I found the 6 studies showed very similar peptide presentation rates for HLA class II complexes, ranging from 0.002 to 0.01 (Table 3.1). Source proteins probably only accounted for a small subset of cellular peptides HLA cIass II molecules would encounter, and there are many biological processes that this simulation did not account for. As such, I conservatively selected a rank threshold of 0.001 for my definition of presentation and immunogenicity. Two other thresholds of 0.01 and 0.0001 were also encompassed in this study.

3.2 ICB-treated cohort data compilation

I analyzed whole exome sequence data of 233 patients with NSCLC and melanoma treated with ICB from four published cohorts: the Anagnostou_NSCLC cohort consisted of 89 patients with NSCLC treated with single agent or combination ICB [25]; the Rizvi_NSCLC cohort consisted of 34 patients with NSCLC treated with pembrolizumab [100]; the Anagnostou_Melanoma cohort consisted of 46 patients with melanoma treated with nivolumab or combination nivolumab/ipilimumab as part of the CheckMate 038 clinical trial [101], and the Snyder_Melanoma cohort consisted of 64 patients treated with ipilimumab [102]. Demographics were extracted from the original publications. For the NSCLC cohorts, PD-L1 expression values and smoking status (former smoker, current smoker and never-smoker) were retrieved from the original publications. Clinical responses to ICB treatment were classified as durable clinical benefit (DCB) or non-durable benefit (NDB). For all four cohorts, DCB was defined as radiographic complete response, partial response, or stable disease for a duration longer than 6 months.

3.3 Assessment of immunogenic mutation burden

3.3.1 Somatic mutation calls

Whole exome sequencing (WES) data was analyzed with the Strelka mutation calling pipeline [103]. Mutations that were in common SNP locations based on dbSNP v138 and those with more than one BLAT hit [104] were filtered. The final list of mutations passed the following criteria: mutation allele frequency (MAF)>=10%, distinct mutant fragments>=4, matched normal coverage >=11, and matched normal MAF>=3%. All missense mutations identified in each cohort were considered as potential IMMs. Mutant and wildtype peptide

sequences surrounding the affected amino acid were extracted, filtering out silent and nonsense mutations with varcode [105]. Windowing around the affected amino acid, all possible 12-20mer [94, 106, 107] mutant/reference peptide pairs were selected as potential HLA class II neoepitopes. HLA class I candidate neoepitopes were extracted for all possible 8-12mer mutant/reference peptide pairs [58, 61, 94].

3.3.2 HLA genotyping

Individual's HLA class I germline haplotypes (HLA-A, HLA-B and HLA-C) were computed with OptiType [81], utilizing whole exome sequencing data. Similarly, HLA class II germline haplotypes were computed with combined results of SOAP-HLA [108] and xHLA [109]. Specifically, xHLA was used to determine haplotypes for HLA-DPB1, HLA-DQB1 and HLA-DRB1. SOAP-HLA was used to determine haplotypes for HLA-DPA1 and HLA-DQA1. For HLA-DP and HLA-DQ genes, combining alpha and beta genotypes resulted in three different cases: 1). Both alpha and beta genes were heterozygous, generating four different HLA-DP/DQ combinations. 2). Either alpha or beta gene was heterozygous, generating 2 different HLA-DP/DQ combinations. 3). Both alpha and beta genes were homozygous, generating only 1 HLA-DP/DQ combination. HLA-DQ alpha and beta pairing that resulted as forbidden pairings were excluded for all the samples [110].

3.3.3 Immunogenic mutation (IMM) calls

To compute IMM burdens, I used MHCnuggets [31], and converted predicted HLA binding affinities for each peptide into a rank-based score. Predicted neoepitopes were selected using affinity rank thresholds of <0.01, <0.001 and <0.0001. A threshold of 0.001 was used as an

Models	Stratification Methods	elpd diff	se diff	elpd loo	se elpd loo	p loo	se p loo	looic	se looic	Tumor Type
No Cohort Effect	HLA II IMMs	0.00	0.00	-294.99	13.41	3.66	1.07	589.98	26.82	NSCLC
Fix Cohort Effect	HLA II IMMs	-0.67	0.68	-295.66	13.32	4.68	1.21	591.32	26.63	NSCLC
Variable Cohort Effect	HLA II IMMs	-0.70	0.83	-295.69	13.37	5.11	1.30	591.38	26.75	NSCLC
No Cohort Effect	HLA II IMMs	0.00	0.00	-248.68	20.71	2.29	0.27	497.35	41.42	Melanoma
Fix Cohort Effect	HLA II IMMs	-0.35	0.72	-249.02	20.85	3.15	0.33	498.05	41.71	Melanoma
Variable Cohort Effect	HLA II IMMs	-0.23	0.90	-246.99	20.84	3.34	0.38	498.16	41.66	Melanoma
No Cohort Effect	TMB	0.00	0.00	-293.05	13.19	3.45	0.99	586.11	26.39	NSCLC
Fix Cohort Effect	TMB	-0.58	0.49	-293.63	13.12	4.22	1.04	587.26	26.23	NSCLC
Variable Cohort Effect	TMB	-0.80	0.38	-293.85	13.16	4.44	1.12	587.71	26.31	NSCLC
No Cohort Effect	TMB	0.00	0.00	-250.32	20.59	2.15	0.21	500.63	41.18	Melanoma
Fix Cohort Effect	TMB	-0.48	0.66	-250.79	20.69	3.07	0.27	501.58	41.38	Melanoma
Variable Cohort Effect	TMB	-0.70	0.66	-251.01	20.71	3.31	0.36	502.03	41.42	Melanoma

Table 3.2 Multi-Cohort Bayseian Survival Model Comparison. elpd=expected log pointwise predictive density; se=standard error; p_loo=effective number of parameters; loo=eave-one-out cross-validation approach; looic=-2*elpd_loo; elpd_diff: difference of elpd between the first and second model, when positive then the expected predictive accuracy for the second model is higher. A negative elpd_diff favors the first model. *elpd_diff>4 and elpd_diff>se_diff are criteria for significant model difference; No model difference can be observe in this table

estimate of the actual presentation rate in the tumor microenvironment. IMMs were defined as missense mutations with at least one predicted mutation-association neoantigen (MANA) with an HLA affinity ranking at the 0.001 percentile for a given HLA haplotype. I computed IMM burden values for each tumor by counting IMMs, defined as mutations that produce at least one predicted MANA with affinity ranking as described above.

3.3.4 Statistical Analyses

The two-sided Mann-Whitney U test was used to evaluate differences in IMM burden and TMB with respect to DCB and NDB groups across all cohorts. For survival analysis, high IMM burden tumors were ranked in the fourth quartile (top 25%) of their cohort and low IMM burden individuals ranked in the first three quartiles (bottom 75%). NSCLC and melanoma cohorts were grouped by cancer type and analyzed with Bayesian Survival models [111] to confirm the absence of cohort effects. For each cancer type, IMM burden data was fit to three models: a model with constant cohort effect, a model with cohort effect varying with IMM burdens, and a model with no cohort effects. The fit of the three models to the data was compared in a leave-one-out (loo) fashion and pairwise expected log pointwise predictive density difference (elpd_diff) was reported. Through comparisons of elpd_diff, while the model with no cohort effect was found to be the best fitting model (Table 3.2), all three models were found to be equivalent. As such, downstream survival analyses were done with cohorts of the same cancer types grouped together to increase statistical power. Kaplan-Meier curves were used to visualize differences in survival for patients with high and low IMM burden tumors. Survival differences between groups were compared with a two-sided log-rank test, hazard ratios were calculated by univariate Cox regressions. Multivariate Cox regression was used to evaluate the impact of

established factors on patient survival. Hazard ratios and p values from the Wald test were reported. A p value threshold of 0.05 was used as the indication for statistical significance, and Benjamini-Hochberg FDR correction was applied. For correlation analyses, I used pairwise Spearman correlations, and p values were reported with Benjamini-Hochberg FDR corrections.

3.3.5 HLA class II IMM burden is linked with clinical benefit from ICB

A total of 5,639 HLA class II IMMs were identified in tumor samples from 233 individuals across all four NSCLC and melanoma cohorts. In NSCLC, average (± standard error) HLA class II IMM burden was 8.6±0.95 (Anagnostou_NSCLC) and 10.94±2.65 (Rizvi_NSCLC), while in melanoma, average HLA class II IMM burden was 41.48±8.25 (Anagnostou_Meanoma) and 40.39±5.96 (Snyder_Melanoma). I did not detect a difference in HLA class II IMM burden distributions between cohorts within tumor types (NSCLC cohorts MW p=0.19; Melanoma cohorts MW p=0.50). HLA class II IMM burdens in the NSCLC (median=6 mutations) and melanoma cohorts (median=25 mutations) were an order of magnitude smaller than TMB in NSCLC (median=97 mutations) and melanoma (median=283 mutations). Additionally, HLA class II IMM burdens were smaller than HLA class I IMM burdens in NSCLC (median=13 mutations) and melanoma (median=43 mutations).

In all four cohorts, the HLA class II IMM burden of tumors responding to therapy was consistently higher than that of non-responding tumors (Anagnostou_NSCLC Mann Whitney U test (MW) p=0.00032; Rizvi_NSCLC MW p=0.0096; Anagnostou_Melanoma MW p=0.001; Snyder_Melanoma MW p=0.0011; Figure 3.1A-D). HLA class II IMM burdens were highly correlated with TMB (Pearson R=0.96, P<0.0001), as well as with HLA class I IMM burdens (Pearson r=0.95, p<0.0001) and with HLA class I MANA burdens (Pearson r=0.94, p<0.0001).

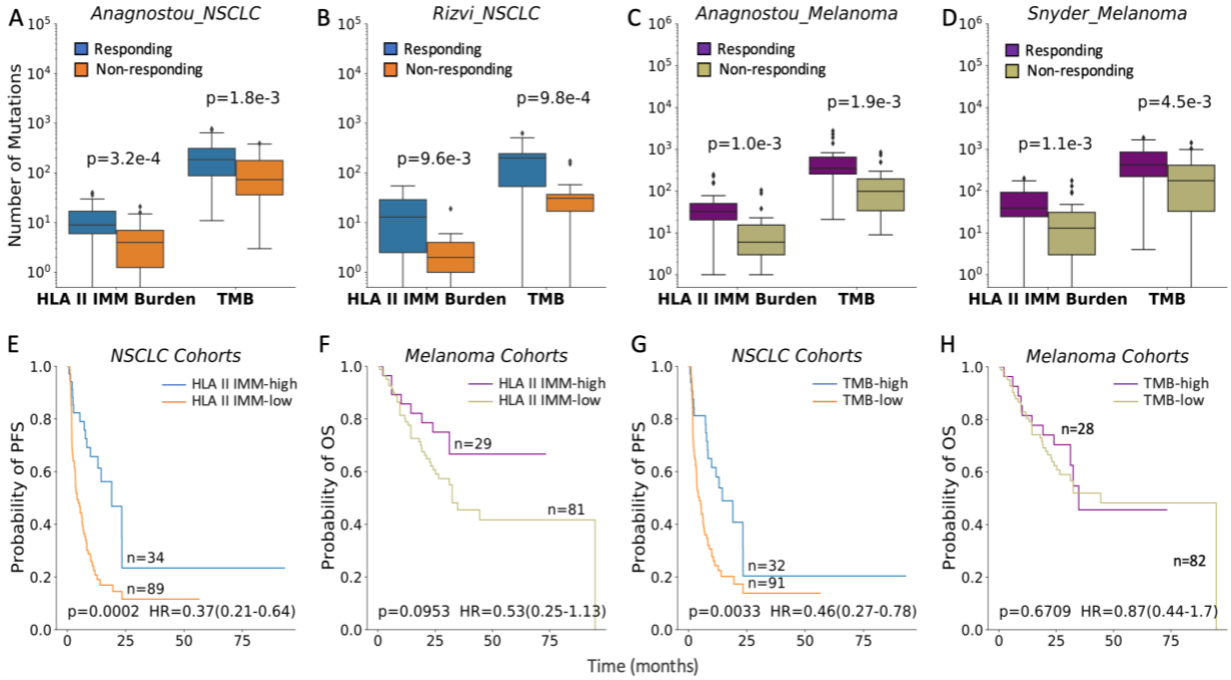

Figure 3.1 HLA class II immunogenic mutation burden is a marker for patient response and survival. (A)-(D) HLA class II IMM burden and TMB both separate responder and non-responder groups in all four cohorts (Anagnostou NSCLC HLA II IMM burden MW P = 0.0003, TMB MW P = 0.0018; Rizvi NSCLC HLA II IMM burden MW P = 0.0096, TMB MW P = 0.00098; Anagnostou melanoma HLA II IMM burden MW P = 0.001, TMB MW P = 0.0019; Snyder melanoma HLA II IMM burden MW P = 0.0011, TMB MW P = 0.0045). HLA class II IMM burden was an order of magnitude smaller than TMB. (E-H) Patients were stratified into high (top 25%) and low (bottom 75%) groups based on their HLA class II IMM burden or TMB for survival analyses. Cohorts of the same tumor type were grouped. Kaplan Meier curves were plotted to show survival differences. (E) High HLA class II IMM burden was associated with longer progression-free survival in NSCLC cohorts (log rank P = 0.0002, HR = 0.37, 95% CI 0.21-0.64). (F) High HLA class II immunogenic mutation burden was trending towards significantly longer OS in melanoma cohorts (log rank P = 0.095, HR = 0.53, CI 0.25-1.13). (G) TMB-high group was significantly associated with longer PFS in NSCLC cohorts (log rank P = 0.0033, HR = 0.46, CI 0.27-0.78). (H) No survival benefit was observed in the TMB-high group of the melanoma cohorts (log rank P = 0.67, HR = 0.87, CI 0.44-1.70). P values were calculated based on the log rank test; a P value of 0.05 was used for significance level.

less than one third of HLA class II IMMs (1653 out of 5,639, 29%) overlapped with HLA class I IMMs.

For survival analysis, patients were first stratified into high burden (top 25%) and low burden groups (bottom 75%) in their respective tumor types; no cohort bias towards HLA class II IMM high or low burden was seen in either tumor type (Fisher's exact test p=0.8235 for NSCLC, p=1.0, for melanoma). Higher HLA class II IMM burden was associated with longer progression-free survival (PFS) in the NSCLC cohorts (log rank p=0.0002, HR=0.37, 95% CI:

0.21-0.64; Figure 3.1E) while a trend towards longer OS was noted in the melanoma cohorts (log rank p=0.095, HR=0.53, 95% CI: 0.25-1.13; Figure 3.1F). High TMB was significantly associated with longer PFS in the NSCLC cohorts (log rank p=0.0033, HR=0.46, 95% CI:0.27-0.78; Figure 3.1G), but not in the melanoma cohorts (log rank p=0.67, HR=0.87, 95% CI:0.44-1.70; Figure 3.1H). The improved prognostic value of HLA class II IMMs in the melanoma cohorts can be explained by re-categorization of 5 responding tumors from the TMB low group into the HLA class II IMM high group (Figure 3.2A). Notably, there was an increased survival benefit of patients with HLA class II IMM burdens in the top quartile (top 25%) compared to the lowest quartile (bottom 25%) in the melanoma cohorts (log rank p=0.01, HR=0.35, CI:0.15-0.82; Figure 3.2B). HLA class I MANA burden was associated with survival in NSCLC but not in melanoma cohorts (NSCLC: log rank p=0.0004, HR=0.37, CI:0.21-0.65; Melanoma: log rank p=0.33, HR=0.71, CI:0.35-1.43). When HLA class I MANA and class II IMM burdens were combined, the association with survival persisted in NSCLC cohorts (HLA class I and class II IMM intersection: NSCLC: log rank p=0.0005, HR=0.37, CI=0.21-0.66; Melanoma: log rank p=0.12, HR=0.55, CI=0.26-1.18; overall HLA class I and class II IMMs: NSCLC: log rank p=0.0003, HR=0.37, CI=0.21-0.65; Melanoma: log rank p=0.12, HR=0.55, CI=0.26-1.18) (Figure 3.2C-D).

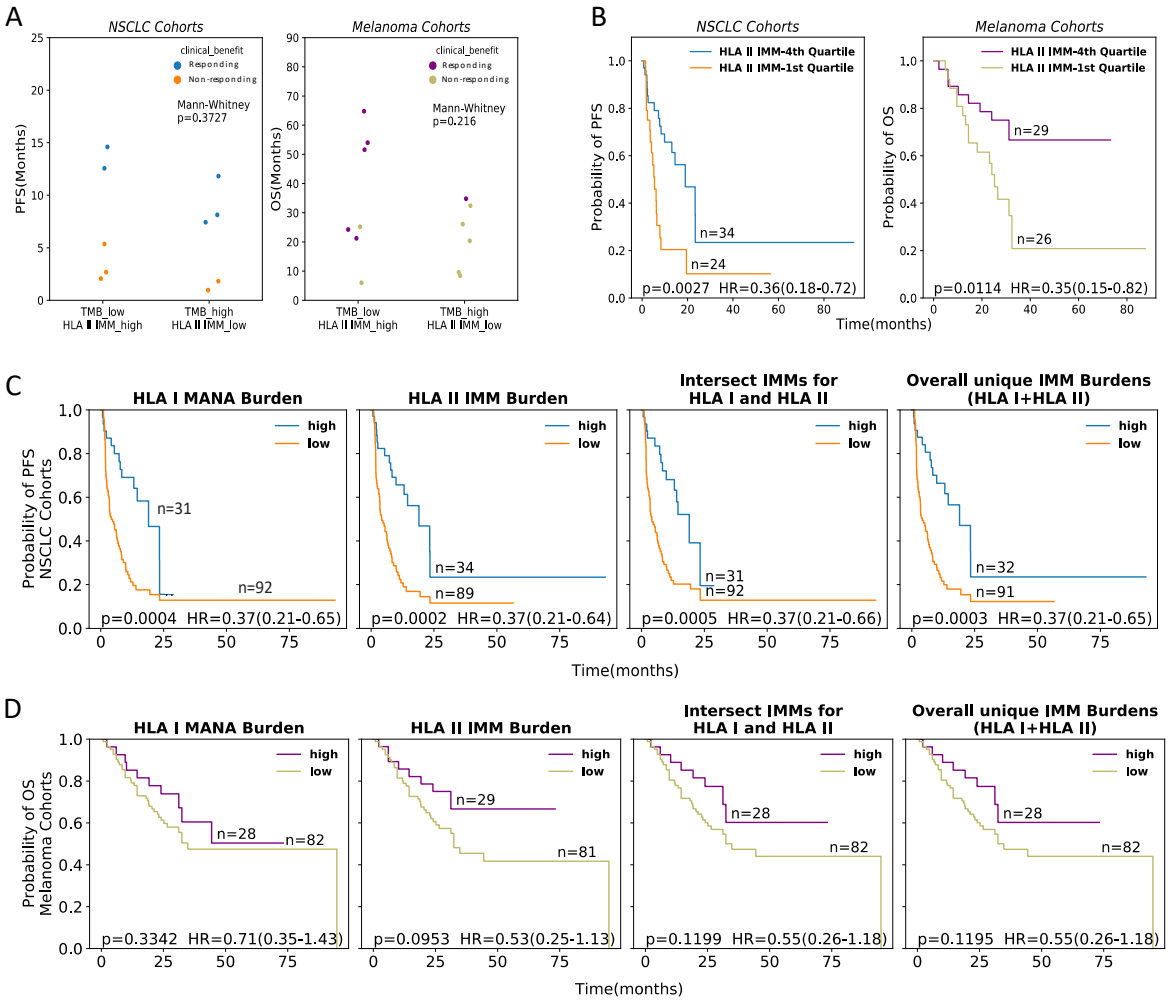

Figure 3.2 Cohort TMB, HLA class I and II IMMs and MANA landscapes and survival analyses A) Samples assigned in different burden groups when using HLA class II IMM burden compared to TMB in NSCLC and melanoma cohorts. Different assignment of 10 samples in melanoma cohorts explained the improvement of survival stratification. B) Significant survival difference was observed when comparing samples with HLA class II IMM burden in the first quartile and samples with HLA II IMM burden in the last quartile for both NSCLC (log rank p=0.0027, HR=0.36, CI:0.18-0.72) and melanoma (log rank p=0.011, HR=0.35, CI:0.15-0.82) cohorts. C-D Survival difference between the high and low burden groups stratified by HLA class I MANA burden, HLA class II IMM burden, Intersect IMM burden of HLA class I and II IIMMs and Overall unique HLA class I and II IMMs in NSCLC and melanoma cohorts. C) Significant survival differences were observed for all stratification methods, while HLA class II IMM burden showed the greatest difference (log rank p=0.0002, HR=0.37, CI: 0.21-0.63) in NSCLC cohorts. D) Survival differences observed were not significant for all stratification methods, but HLA class II IMM burden trended towards longer OS (log rank p=0.0953, HR=0.53, CI:0.25-1.13) in melanoma cohorts.

3.4 Assessment of HLA II IMM clonality and tumor heterogeneity

3.4.1 Mutation clonality calls

To identify clonal and subclonal mutations, the cancer cell fraction of each tumor mutation was estimated. The observed variant allele fraction (V_{exp}) of the mutation was modeled by considering the sample purity (α), the copy number of the mutated locus in tumor (n_T) and normal (n_N), mutation cellular fraction (C) and mutation multiplicity (m).

$$V_{exp} = \frac{m\ C\ \alpha}{\alpha\ n_T + (1-\alpha) n_N}$$

The mutation multiplicity m represents the integer number of mutated copies.

A 95% confidence interval (CI) for the variable V_{exp} was constructed by using the distinct mutant read counts and distinct coverage at of the mutated locus. Substitution of the other known variable gave a confidence interval for the product $m\ C$. Mutations were classified into clonal and subclonal categories based on the following rules. (1) If the CI for $m\ C$ overlaps an integer value, that value is estimated to indicate the multiplicity of the mutation and the mutation is clonal (C=1). (2) If the upper bound of the CI for $m\ C$ is below 1, the multiplicity is set to 1, and the mutation is subclonal, unless the resulting estimate for C is within a tolerance threshold (0.25) of 1. (3) If the CI for $m\ C$ is above 1 and does not overlap any integer values, multiplicity is greater than 1 and m is set such that the confidence interval for C falls within the expected intervals of [0,1] [25, 112]. Forty samples across four cohorts with mutation purity too low to analyze for copy number (<20%) were excluded from this analysis.

3.4.2 Intra-tumoral IMM heterogeneity analyses

Intra-tumoral HLA class I or II IMM heterogeneity was assessed by determining the fraction of subclonal IMMs as follows:

$$IMM\ Heterogeneity = \frac{Number\ of\ Subclonal\ IMMs\ in\ Tumor\ Sample}{Total\ Number\ of\ IMMs\ in\ Tumor\ Sample}.$$

Samples with no IMMs were excluded from downstream analyses and IMM heterogeneity thresholds of 0.05 and 0.1 were considered. This approach is similar to that of McGranahan *et al*. [113].

3.4.3 Impact of HLA II IMM clonality and intra-tumoral heterogeneity

I evaluated the impact of clonal HLA class II IMMs on clinical outcomes by computing the intra-tumor IMM heterogeneity (IMM_{het}) for each tumor. Using estimated mutation cancer cell fractions and excluding samples with low purity (<20%), I assessed HLA class II IMM clonalities for 193 tumor samples and found that 84.6% of identified IMMs were clonal. IMM_{het}, defined as the fraction of subclonal HLA class II IMMs, was similar in the melanoma (6.3%) and NSCLC (5.8%) cohorts. While NSCLC tumors harboring high numbers of clonal IMMs (top25%) had lower levels of IMM_{het} (MW p=0.059), this observation was not apparent for the melanoma tumors analyzed (MW p=0.32). Patients with NSCLC tumors with high clonal HLA class II IMM burden (top 25%) had a significantly longer PFS (log rank p=0.0001, HR=0.3 CI: 0.16-0.58; Figure 3.3A). In considering different IMM_{het} levels, patients with tumors harboring low IMM_{het} and high clonal HLA class II IMM burden (top 25%) had longer PFS in the NSCLC cohorts (IMM_{het} <0.05: log rank p=0.0041, HR=0.25, CI:0.09-0.7; IMM_{het} <0.1: log rank p=0.0009, HR=0.32, CI:0.16-0.65; Figure 3.3B-C). These findings did not reach statistical significance in the melanoma cohorts (clonal HLA class II IMM burden without IMM_{het}: log rank p=0.4, HR=0.72, CI:0.32-1.64; with IMM_{het}<0.05: log rank p=0.27, HR=0.56, CI:0.2-1.57; with IMM_{het}<0.1: log rank p=0.49, HR=0.74, CI:0.31-1.76; Figure 3.3D-F). IMM_{het} alone was not correlated with PFS in the NSCLC cohorts (IMM_{het}<=0.05, Log Rank p=0.83, HR=1.05,

CI:0.65-1.71; IMM_{het}<=0.1, Log Rank p=0.22, HR=0.73, CI:0.44-1.22) or OS in the melanoma cohorts (IMM_{het}<=0.05, Log Rank p=0.84, HR=0.93, CI:0.49-1.79; IMM_{het}<0.1, Log Rank p=0.66, HR=0.84, CI:0.4-1.78).

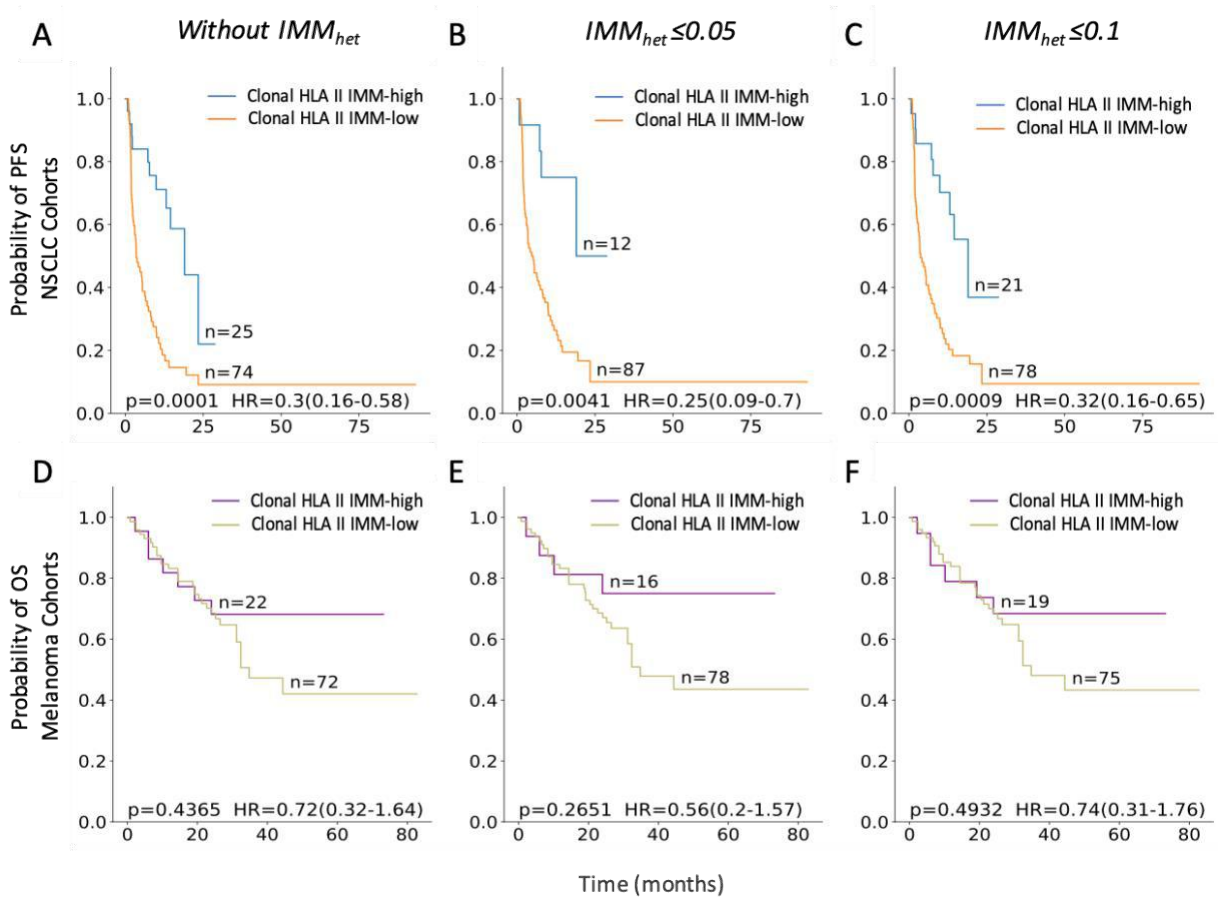

Figure 3.3 Impact of clonality and tumor heterogeneity on patient survival. Clonality of IMMs was determined using mutation cellular fractions. Samples with low purity (a threshold of 20%) were excluded from this analysis. Patients were stratified into high (top 25%) and low (bottom 75%) groups based on their clonal HLA class II IMM burden. For the NSCLC cohorts, a total of 99 samples passed the purity threshold; for melanoma cohorts, a total of 94 samples passed the purity threshold. Intratumoral IMM heterogeneity (IMMhet) was defined as the fraction of subclonal IMMs. (A) High clonal class II IMM burden without IMMhet thresholds showed significant association with progression-free survival (PFS) in NSCLC cohorts [log rank P = 0.0001, HR = 0.3 confidence interval (CI) 0.16-0.58]. (B), (C) NSCLC tumors harbored high clonal HLA class II IMM burden combined with IMMhet thresholds of ≤0.05 or ≤0.1 and had significantly longer PFS: (B) with IMMhet ≤0.05 (log rank P = 0.0041, HR = 0.25, CI 0.09-0.7), (C) with IMMhet ≤0.1 (log rank P = 0.0009, HR = 0.32, CI 0.16-0.65). (D)-(F) Clonal HLA class II IMM burden with or without IMMhet thresholds did not show any association with overall survival (OS) in melanoma cohorts: (D) without IMMhet (log rank P = 0.4, HR = 0.72 CI 0.33-1.56), (E) with IMMhet ≤0.05 (log rank P = 0.21, HR = 0.52, CI 0.18-1.46), (F) with IMMhet ≤0.1 (log rank P = 0.33, HR = 0.65, CI 0.27-1.56).

HLA, human leukocyte antigen; HR, hazard ratio; NSCLC, non-small-cell lung cancer.

3.5 Impact of germline HLA class II variation in combination with HLA class II IMM burden on clinical outcomes

I first considered the impact of patients' HLA class II germline variation combined with HLA class II IMM burden on survival. While I found no correlation between HLA class II IMM burden and the number of HLA class II heterozygous alleles for the HLA-DPA1, DPB1, DQA1, DQB1 or DRB1 genes (NSCLC cohorts: Spearman rho=-0.03, p=0.71; melanoma cohorts: Spearman rho=0.08, p=0.39), the majority of cases analyzed harbored maximal HLA class II germline heterozygosity, defined as 9 or 10 heterozygous HLA class II alleles (NSCLC cohorts: 78 out of 123, 63%; melanoma cohorts 80 out of 110, 72%). Maximal HLA class II germline heterozygosity, when combined with high HLA class II IMM burden (top 25%), conferred longer PFS in the NSCLC (log rank p=0.0002, HR=0.24, CI:0.11-0.54; Figure 3.4A), with a trend towards longer OS in the melanoma cohorts (log rank p=0.12, HR=0.47, CI:0.8-1.26; Figure 3.4B). Next, I controlled for either HLA class II IMM burden or HLA class II allele counts to evaluate their contributions to survival benefit. Within the patient group with maximal HLA class II heterozygosity, those with tumors harboring high HLA class II IMM burden had improved PFS in the NSCLC cohorts (log rank p=0.0001, HR=0.24, CI:0.11-0.52; Figure 3.4C) and OS in melanoma cohorts (log rank p=0.1, HR=0.49, CI:0.2-1.19; Figure 3.4D).

Of the HLA class II loci studied, heterozygosity at the HLA-DP locus (3 or 4 HLA-DP alleles) combined with high HLA class II IMM burden (top 25%) showed an association with longer PFS in the NSCLC cohorts (log rank p=0.0005, HR=0.26, CI:0.12-0.58; Figure 3.4E) and a weak trend for longer OS in the melanoma cohorts (log rank p=0.0994, HR=0.41, CI:014-1.23; Figure 3.4F). When controlled for HLA-DP heterozygosity, I again noted an improved survival in the HLA class II IMM high group compared to the low group in the NSCLC (log rank

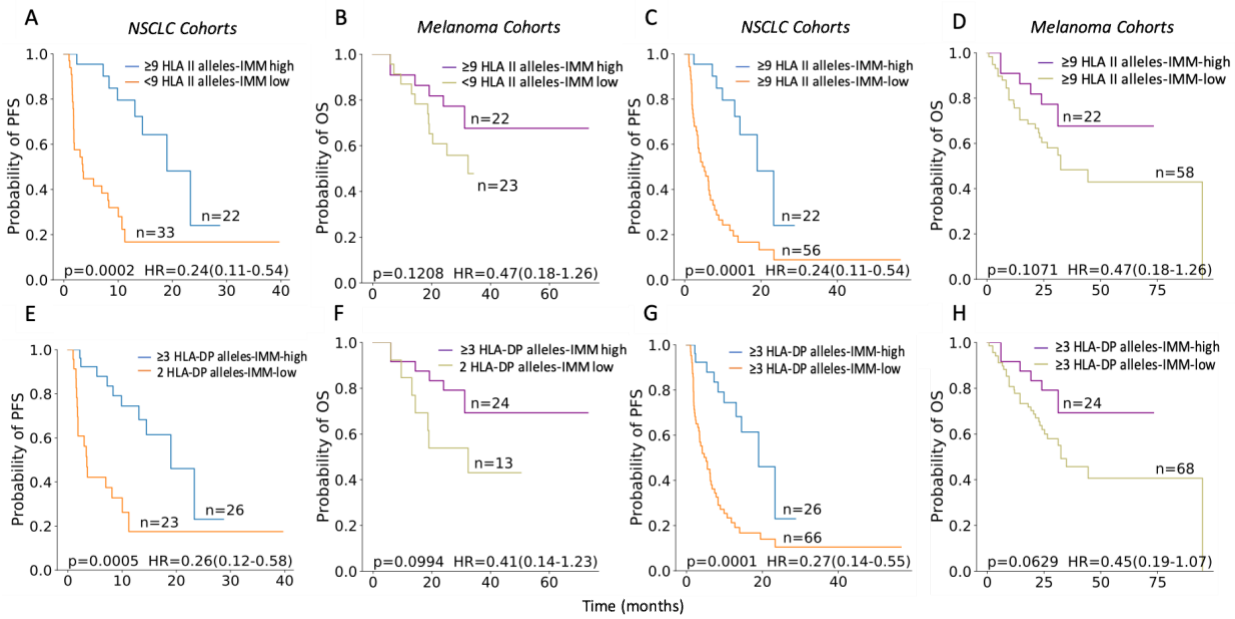

Figure 3.4 Impact of HLA class II alleles in combination with HLA class II immunogenic mutation burden on patient outcome. (A), (B) Patients with the maximum heterozygous HLA class II alleles (9 or 10) and high HLA class II immunogenic mutation (IMM) burdens were found to have longer PFS in NSCLC cohorts (log rank P = 0.0002, HR = 0.24, CI 0.11-0.54), but not with OS in melanoma cohorts (log rank P = 0.12, HR = 0.47, CI 0.18-1.26). (C), (D) Within samples with maximum heterozygous HLA class II alleles, samples with high HLA class II IMM burdens had improved PFS in NSCLC (log rank P = 0.0001, HR = 0.24, CI 0.11-0.54) and OS in melanoma cohorts (log rank P = 0.1, HR = 0.47, CI 0.18-1.26). (E), (F) Maximum heterozygosity in at least one of HLA-DPA1 and HLA-DPB1 genes along with high HLA class II IMM was found to improve patient PFS in NSCLC cohorts (log rank P = 0.0005, HR = 0.26, CI 0.12-0.58) and resulted in a weak trend to longer OS in melanoma cohorts (log rank P = 0.099, HR = 0.41, CI 0.14-1.23). (G), (H) Within samples with more than 3 HLA-DP alleles, those with high HLA class II IMM burden had further improvements in PFS of NSCLC cohorts (log rank = 0.0001, HR = 0.27, CI 0.14-0.55) and OS in melanoma cohorts (log rank P = 0.063, HR = 0.45, CI 0.19-1.07). Log rank P values are reported.

p=0.0001, HR=0.27, CI:0.14-0.55; Figure 3.4G), and melanoma cohorts (log rank p=0.063, HR=0.45, CI:0.19-1.07; Figure 3.4H). HLA-DPB1 or HLA-DPA1 heterozygosity combined with high HLA class II IMM burden was also significantly associated with PFS in NSCLC cohorts (log rank p=0.0039, HR=0.16, CI:0.04-0.66 and log rank p=0.0005, HR=0.26, CI:0.12-0.58 respectively; Figure 3.5A). However, HLA-DPA1 or HLA-DPB1 heterozygosity when combined with HLA class II IMM burden was not associated with OS in the melanoma cohorts (log rank p=0.27, HR=0.46, CI:0.11-1.91 and log rank p=0.17, HR=0.48, CI:0.16-1.43 respectively; Figure 3.5B). Additionally, heterozygosity on its own at any of the loci did not show association with PFS in NSCLC (Figure 3.5C), whereas HLA-DPA1 heterozygosity on its own (log rank

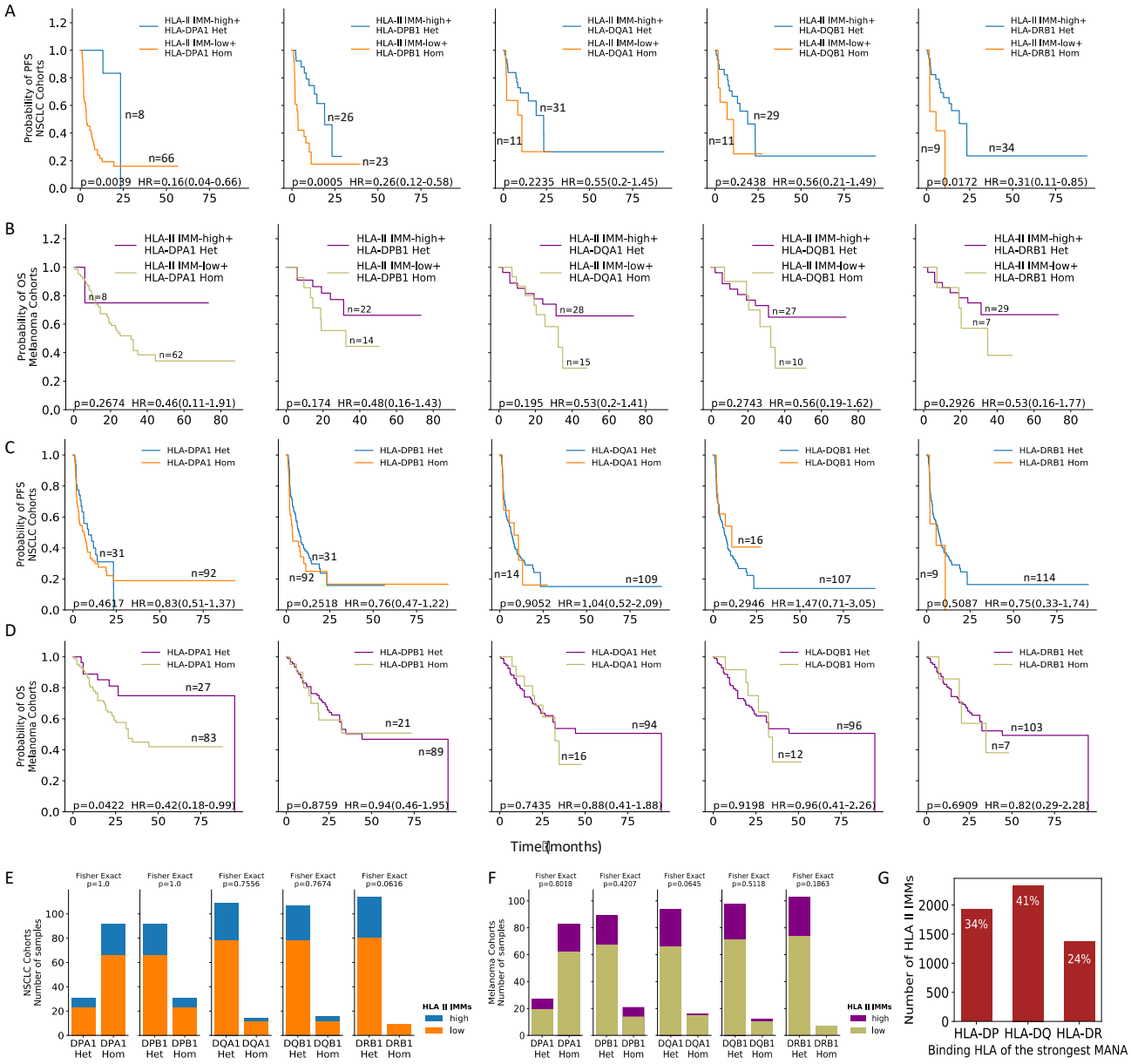

Figure 3.5 Different HLA class II genotype variation's impact on survival. (A),(B). Survival benefits of all five HLA class II genotypes (HLA-DPA1, DPB1, DQA1, DQB1 and DRB1) variation in combination with HLA class II IMM burdens for NSCLC and melanoma cohorts. (C),(D) Survival difference of heterozygous vs homozygous samples for each of the HLA class II genotypes in NSCLC and melanoma cohorts. (E),(F) Number of samples heterozygous or homozygous for all HLA class II genotypes. High HLA class II IMM burden did not co-occurred with HLA class II variation for any of the loci in NSCLC or melanoma cohorts. HLA II IMM high=top 25% burden, HLA II IMM low=75% burden. (G) Number of HLA class II IMMs had the strongest binding by HLA-DP, HLA-DQ and HLA-DR. Of the total 5639 HLA class II IMMs identified, HLA-DQ had the strongest binding at the highest frequency while HLA-DR had the lowest frequency.

p=0.042, HR=0.42, CI:0.18-0.99) was associated with longer OS in the Melanoma cohorts (Figure 3.5D). Co-occurrence of heterozygosity at any of the HLA class II loci and high HLA class II IMM burden was not observed in either NSCLC or Melanoma cohorts (Figure 3.5E-F). Lastly, the binding preferences of HLA class II molecules were investigated. HLA-DQ complexes have previously been found to preferentially bind human proteins compared to HLA-DRs [114]. My findings corroborated this result, as the MANAs associated with the 5,639 HLA class II IMMs had the strongest binding to HLA-DQ alleles (Figure 3.5G).

3.6 HLA II IMM burden compared to known factors in association to survival outcomes

Established factors, such as programmed death-ligand 1 (PD-L1) expression and smoking status, have previously been shown to impact clinical outcomes for patients with NSCLC treated with immune checkpoint blockade [115, 116]. Here, to ensure consistency between the categorical PD-L1 expression values in the Rizvi NSCLC cohort and the continuous values from the Anagnostou NSCLC cohort, I categorized the continuous values into Strong (>= 50% membranous staining), Weak (1-49% membranous staining), Negative (<1% membranous) and Unknown (unassessed) groups. PD-L1 was stained for tumor cells in Anagnostou NSCLC cohort, and it was stained for tumor and immune infiltrating cells in Rizvi NSCLC cohort.

I found that NSCLC tumors with a high HLA class II IMM burden showed positive PD-L1 expression (Fisher exact test p=0.024), but the levels of PD-L1 expression did not monotonically correlate with levels of HLA class II IMM burdens (Spearman rho=0.07, p=0.57). A high HLA class II IMM burden was also correlated with current or former smoking status (Fisher Exact test p=5.47e-5), as well as with number of pack-years in a subset of patients

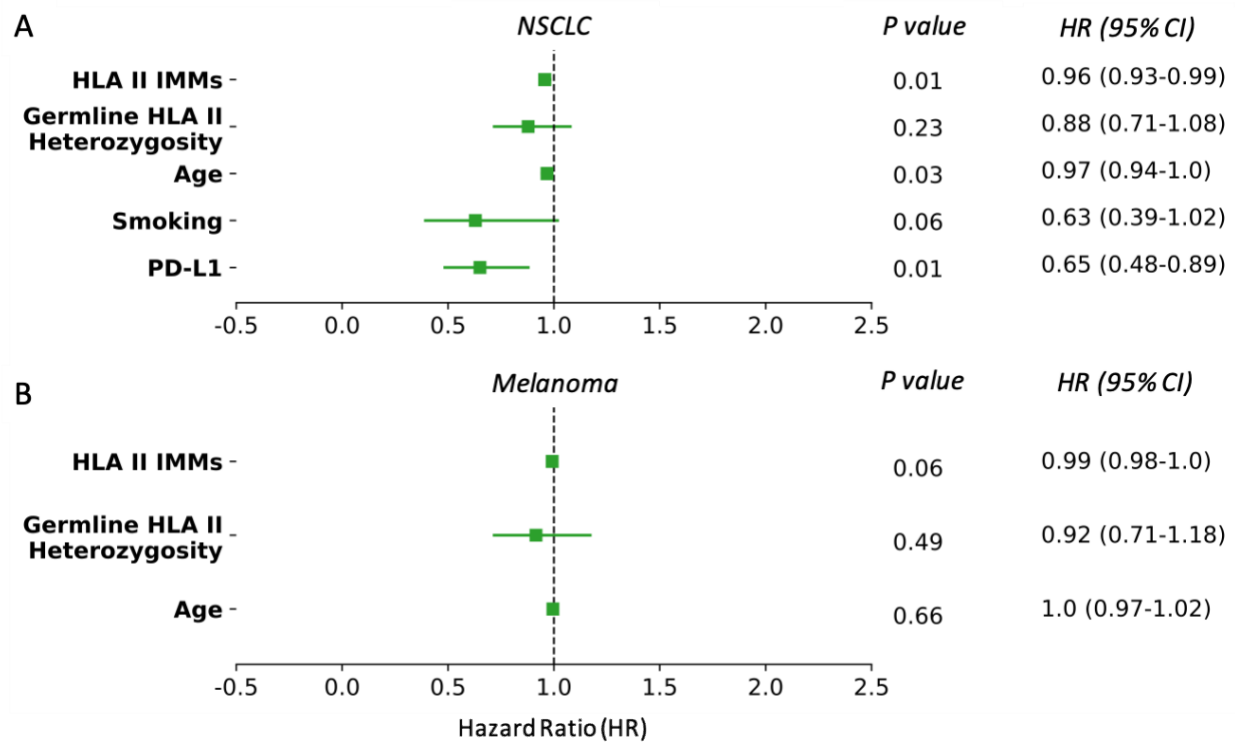

Figure 3.6 Associations between established factors and patient survival. Established factors of age, smoking status, and PD-L1 expression and HLA class II immunogenic mutation (IMM) burdens were analyzed through multivariate Cox survival models to understand their impact on patient survival. Smoking status was categorized into three groups—never smoker = 0, former smoker = 1, and current smoker = 2. PD-L1 expression levels were assigned into four groups: strong (≥50% membranous staining) = 2, weak (1%-49% membranous staining) = 1, negative (<1% membranous staining) = 0, and unassessed. Patients with unassessed PD-L1 expression were excluded in this analysis. HLA class II IMM burdens, as a continuous variable, indicated the number of IMMs a sample had. Germline HLA class II heterozygosity, as a continuous variable, indicated the numbers of unique HLA class II alleles a sample had. (A) HLA class II IMM burden (Wald test P = 0.01, HR = 0.96, CI 0.93-0.99), age (Wald test P = 0.03, HR = 0.97, CI 0.94-1.0) and PD-L1 tumor expression (Wald test P = 0.01, HR = 0.65, CI 0.48-0.89) were found to reduce survival hazard in NSCLC cohorts. (B) Only HLA class II IMM burden had a trend towards significant effects in reducing survival hazard in melanoma cohorts (Wald test P = 0.06, HR = 0.99, CI 0.98-1.0).

(Spearman rho=0.36, p=0.037). I subsequently analyzed the interactions of HLA class II IMM burden, patient demographics and PD-L1 expression and smoking in a multivariate Cox survival model. For the NSCLC cohorts, continuous HLA class II IMM burden (Wald Test p=0.01, HR=0.96, 95% CI:0.93-0.99), age (Wald Test p=0.03, HR=0.97, 95% CI=0.94-1.0) and PD-L1 expression (Wald Test p=0.01, HR=0.65, 95% CI:0.48-0.89) were independently associated with favorable PFS (Figure 3.6A). For the melanoma cohorts, I found a trend towards an independent

association between HLA class II IMM burden and OS (Wald Test p=0.06, HR=0.99, 95% CI:0.98-1.0; Figure 3.6B). The number of heterozygous HLA class II alleles, as a continuous variable, was not independently associated with survival in either cohort type (Figure 3.6).

3.7 Treatment effect

To establish the impact of treatment effects on my analysis, I compared the differential survival benefits overall and in high or low HLA class II IMM burden groups across varied treatments. In NSCLC cohorts, I found a high number of HLA class II IMMs harbored by patients treated with Anti-PD1+chemo (n=2) (ANOVA p=0.0062), but no differential survival benefits (Figure 3.7). For melanoma cohorts, there was no difference in HLA class II IMM burden across treatment groups, but I saw a survival benefit for patients with low HLA class II IMMs treated with Tremelimumab (n=2) (Figure 3.8). However, in both cases, the treatment groups consisted of 2 samples and were not likely to be representative of the cohorts. In terms of standalone or combination effects of Anti-PD1 or of Anti-CTL4 treatments, I did not see any differential survival benefit overall or in high or low HLA class II IMM burden groups (Figure 3.7-3.8).

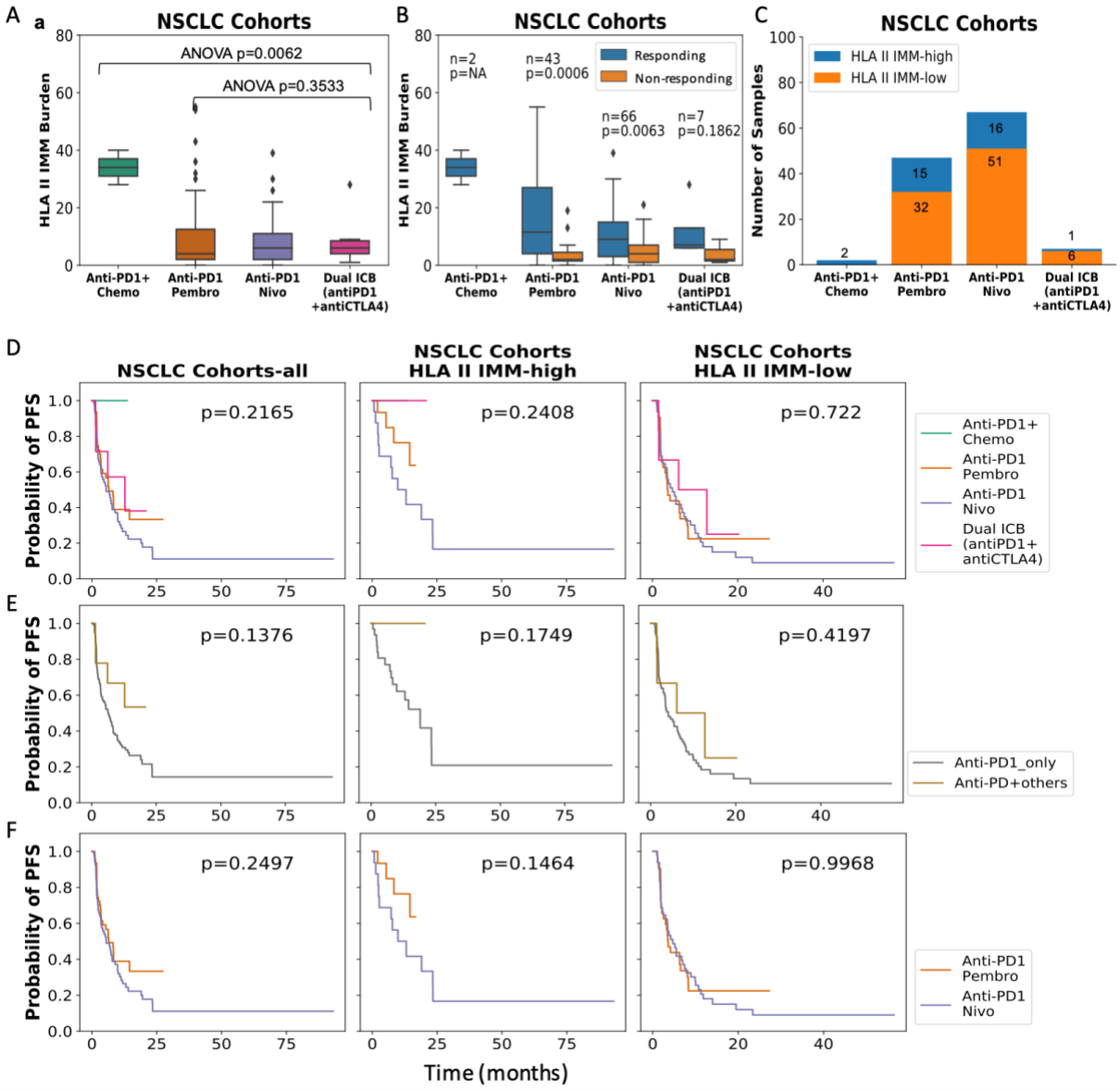

Figure 3.7 Treatment effect of NSCLC cohorts. (A) Higher numbers of HLA class II IMM burdens were observed in samples treated with anti-PD1+chemo (ANOVA p=0.0062). (B) Responders treated with Anti-PD1 (Pembrolizumab and Nivolumab) alone had higher numbers of HLA class II IMMs than Non-responders. (C) Number of samples classified in HLA class II IMM burden high and low groups in different treatment groups. (D) No survival difference was observed of samples that went through different treatment plans. (E) Treatment were separated into Anti-PD1 only vs Anti-PD1+other treatment forms. No survival difference was observed for different treatment categories. (F) Between the two Anti-PD1 treatments (Pembrolizumab and Nivolumab), no survival difference was observed either.

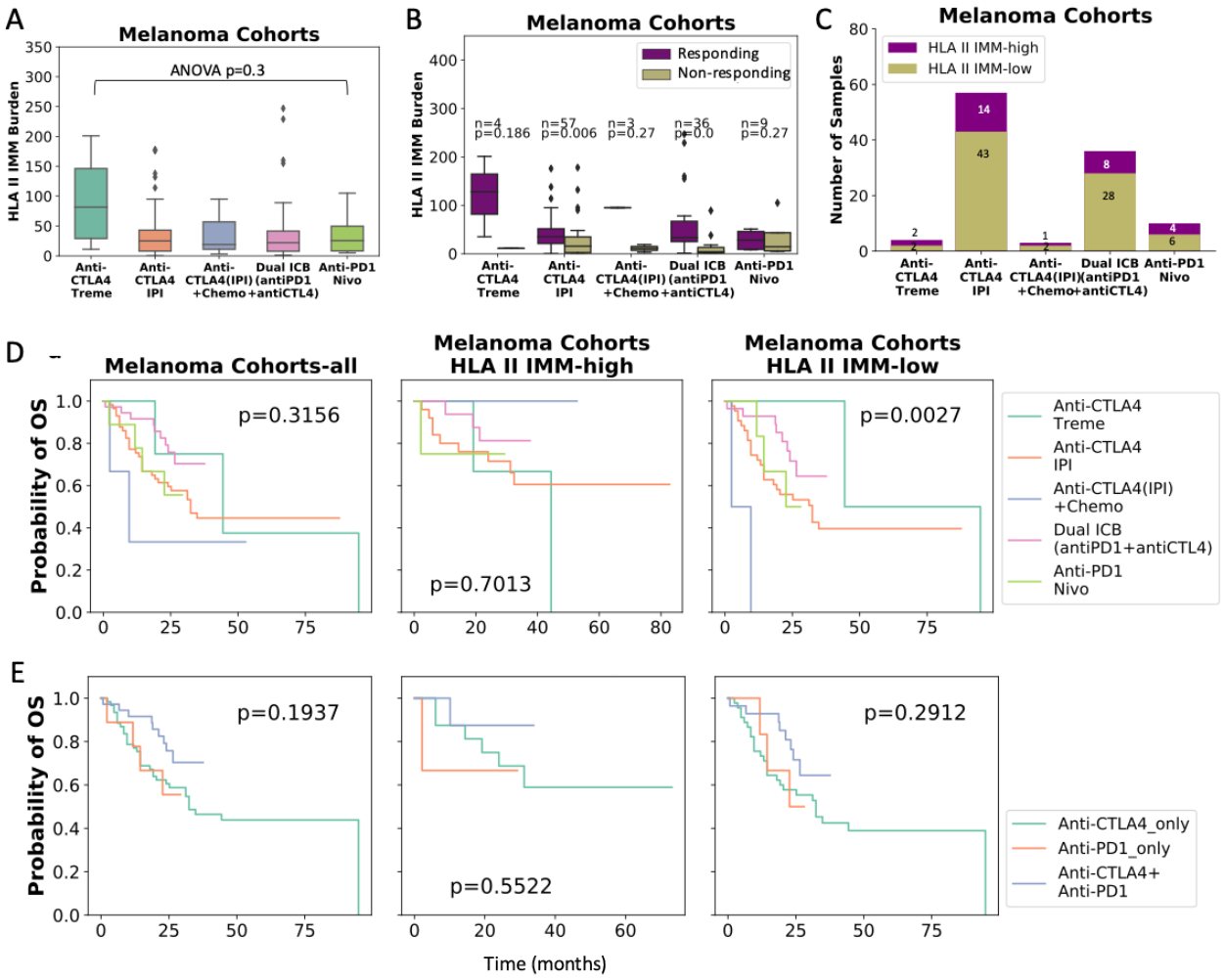

Figure 3.8 Treatment effects of Melanoma Cohorts. (A) No difference of HLA class II IMM burdens was observed in samples with different treatment plans. (B) Responding tumor treated with Anti-CTL4 (ipilimumab) alone (MW p=0.006) and Anti-CTL4+antiPD1 (MW p<0.001) had higher numbers of HLA class II IMMs than non-responding tumors (C) Number of samples classified in HLA class II IMM burden high and low groups in different treatment groups. (D) For HLA class II IMM low samples, those that were treated with tremelimumab (Anti-CTL4, n=2) had lower survival than samples with other treatments (log rank p=0.0027). (E) Treatment were separated into Anti-PD1 only, Anti-CTL4 only and Dual treatments of Anti-PD1+Anti-CTL4. No survival difference was observed for different treatment categories

3.8 High HLA class II IMM burden is associated with an inflamed tumor microenvironment

3.8.1 RNA sequencing Analyses for TIL compositions showed association with high expression levels of CD4+ T cells

RNA sequencing and TCR sequencing data from the Anagnostou Melanoma cohort were used to assess Tumor Infiltration Leukocyte (TIL) compositions. In brief, paired-end RNA sequencing data were aligned with the STAR pipeline [117] and filtered with Picard Tools (https://broadinstitute.github.io/picard/) using the MarkDuplicates filter. Normalized expression data (normalized with RSEM v1.2.30 [80] using the strand-specific mode, reported in transcripts per million (TPM)) were then input into CIBERSORT v1.06 [92] for 22 immune cell types' absolute compositions. Absolute composition values of all 22 immune cell types were first correlated with HLA class II burdens using Spearman correlation, and then they were analyzed for differential comparison between high and low HLA class II IMM burden groups using two-sided Mann-Whitney U test. All statistical analyses were corrected by Benjamini-Hochberg method to minimize false discoveries.

RNA sequence data deconvolution revealed that densities of pre-treatment activated CD4+ memory T-cells and M0 macrophages in the TME were positively correlated with HLA class II IMM burden (Spearman rho=0.39, p=0.01, FDR p=0.12 and Spearman rho=0.45, p=0.003, FDR p=0.08 respectively; Figure 3.9A). Next, immune cell fractions were compared in the TME of tumors with high (top 25%) versus low HLA class II IMM burdens (bottom 75%). Activated memory CD4+ T-cells and T follicular helper cells were significantly more abundant in the TME of high HLA class II IMM burden tumors prior to ICB (MW p=0.022 and p=0.028 respectively; Figure 3.9B). A trend towards higher abundance of CD8+ T-cells and M1 macrophages was also

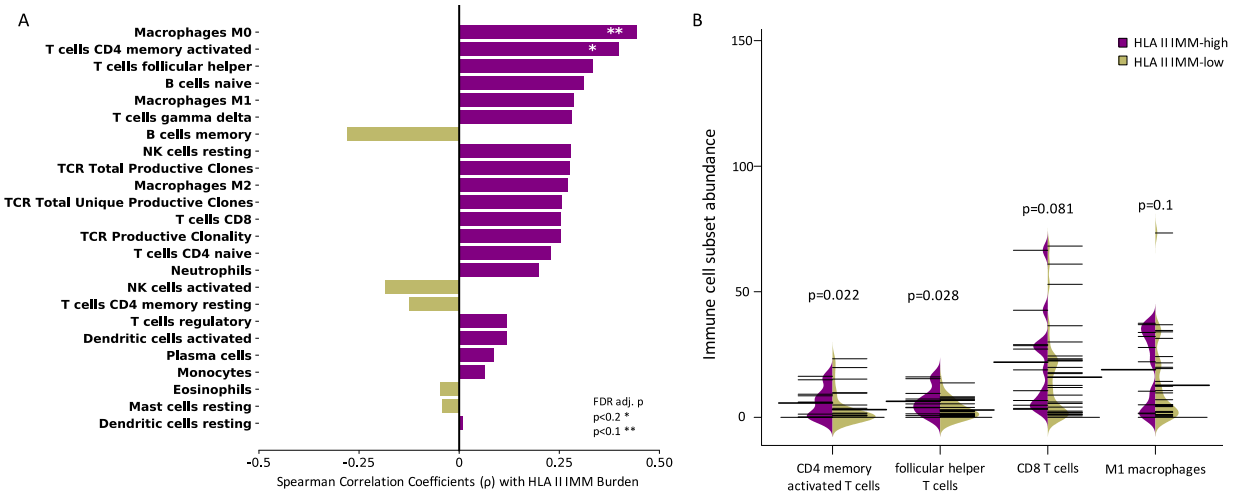

Figure 3.9 Correlation between HLA class II immunogenic mutation burden and pretreatment tumor infiltration lymphocytes in melanoma. (A) Pairwise Spearman correlations between pretreatment immune cell subsets (determined by RNA sequencing data deconvolution or TCR sequencing) and HLA class II IMM load. A positive correlation was found between HLA class II IMM burden and pretreatment M0 macrophages (Spearman's rho = 0.45, P = 0.003, FDR P = 0.08), and CD4+ memory activated T cells (Spearman's rho = 0.39, P = 0.01, FDR P = 0.12). (B) Tumor infiltration lymphocyte levels in high and low HLA class IMM burden groups were compared. Pretreatment CD4+ memory activated T cells (Mann–Whitney P = 0.022) and T follicular helper cells (Mann–Whitney P = 0.028) were found to have a significantly increased expression level in the high burden group. CD8+ T cells (Mann–Whitney P = 0.081) and M1 macrophages (Mann–Whitney P = 0.1) had higher expression levels in high HLA class II IMM burden group trending towards significance. Thin black lines indicate individual samples in the group of interest. Thick black line indicates the average of the group. Purple shading indicates the distributions of HLA class II IMM burden high groups, and the dark khaki shading indicates the distributions of HLA class II IMM burden low groups.

FDR, false discovery rate calculated with Benjamini-Hochberg method; HLA, human leukocyte antigen; IMM, immunogenic mutation; NK, natural killer; TCR, T-cell receptor. ∗FDR adjusted P < 0.2. ∗∗FDR adjusted P < 0.1.

noted in the TME of high HLA class II IMM burden tumors (MW P=0.081 and 0.1 respectively; Figure 3.9B).

3.8.2 Gene Set Enrichment Analysis showed enrichments of TCR and BCR activation signals both at baseline and on treatment in HLA class II IMM high patients

Differential expression analysis was performed using a negative binomial model with the DESeq2 package[118] to compare pre-therapy and on-therapy RNA sequencing data from the Anagnostou-Melanoma cohort. Genes were tested for differential expression based on High (top 25%) and Low (bottom 75%) groups, as defined by HLA class II IMM burden, TMB or HLA

class I MANA burdens. The resulting statistics were used to generate gene level rankings for input to gene set enrichment analysis with fgsea[119]. The scores were defined as -log(p) * sign(fc) where p is the p-value from the negative binomial model and fc is fold-change between the High and Low groups. A selection of 58 gene sets were used, including Hallmark inflammatory response sets and gene sets associated with antigen specific response, TCR/BCR signaling and other type 2 immune response[120-133].

Based on the GSEA, I found an over-representation of interferon-γ response gene sets in the HLA class II IMM high tumors prior to immune checkpoint blockade (FDR p=5.31e-16). Prominent enrichments were also found in gene sets related to BCR/TCR signaling (FDR p=2.03e-13 and 1.06e-11 respectively) and antigen presentation (FDR p=1.12e-7), particularly MHC class II antigen presentation (FDR p=2.59e-6) in pre-treatment tumors with a high HLA class II IMM burden (Figure 3.10). Notably, immune checkpoint blockade induced an upregulation of gene sets related to TCR/BCR signaling and IL2_STAT5 TCR activation in the TME of HLA class II IMM high tumors that persisted during treatment (FDR p=4.91e-4, 4.91e-4, and p=0.041 respectively; Figure 3.10). When patients were separated into TMB high/low groups, TCR signaling (FDR p=4.84e-10 for REACTOME_DOWNSTREAM_TCR_SIGNALING and 3.05e-8 for REACTOME_TCR_SIGNALING), BCR signaling (FDR p=6.99e-5), interferon gamma response/inflammatory response (FDR p=3.05e-8 and 2.51e-4 respectively) and MHC class I and II antigen presentation (FDR p= 2.32e-2 and 2.28e-5 respectively) gene sets were enriched for the TMB high group at baseline. While TCR signaling (FDR p=2.33e-2) enrichment in TMB high group persisted on treatment, proinflammatory cytokine—interferon gamma (FDR p=4.28e-2))—gene set was enriched in the TMB low group on treatment (Figure 3.11). When separating

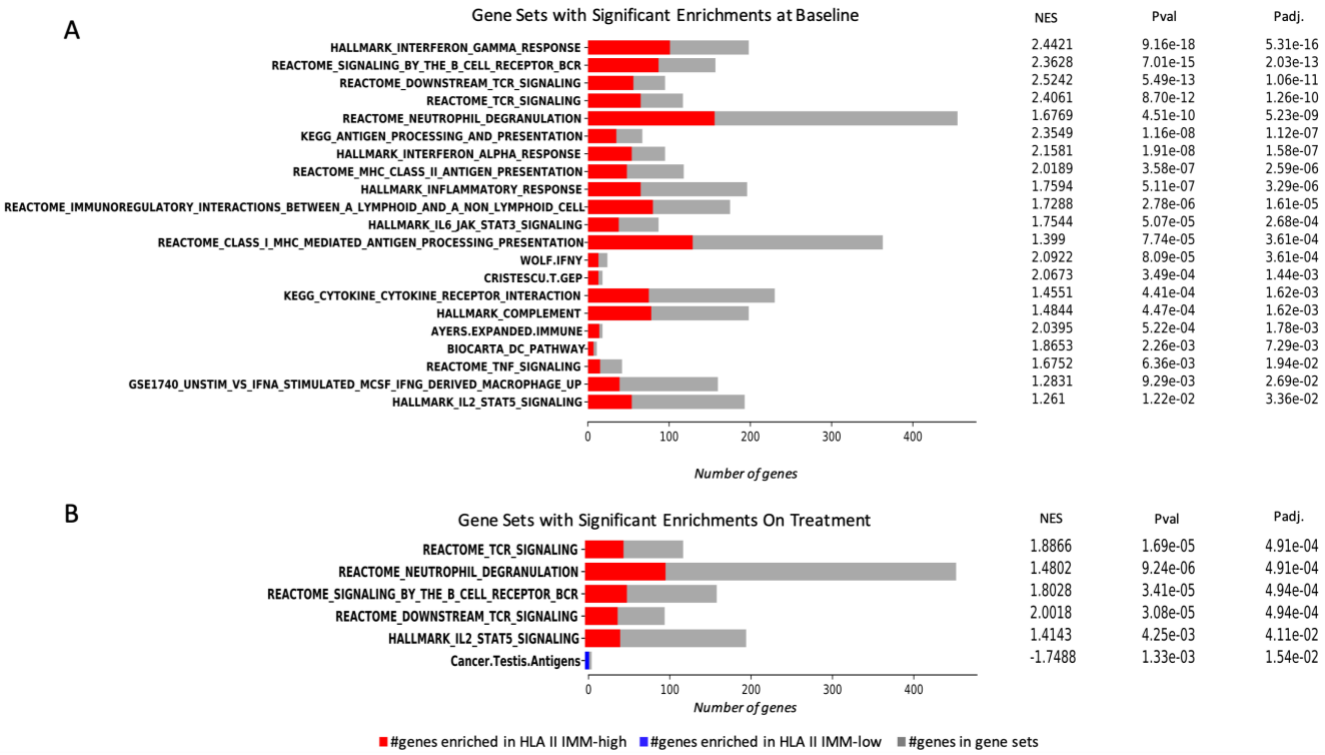

Figure 3.10 Gene set enrichment analyses for HLA class II immunogenic mutation burden-high and -low group. A selection of 58 gene sets related to inflammatory response, antigen presentations, and type 2 immunity were used to assess differences of gene expression in HLA class II IMM burden-high and -low groups. All the gene sets that were significantly enriched in either the high or low groups are shown. The number of genes that overlapped with gene sets overrepresented in the HLA class II IMM burden high group are shown in red, and the number of genes that overlapped with gene sets overrepresented in the HLA class II IMM burden-low group are shown in blue. Overall, 21 gene sets were enriched in the HLA class II IMM burden-high group at baseline. Six gene sets continued to be enriched in this group on treatment. Particularly, strong inflammatory response, antigen presentation, and T-cell receptor/B-cell receptor signaling gene set enrichments can be seen both at baseline and on treatment.

HLA, human leukocyte antigen; IMM, immunogenic mutation; NES, normalized enrichment score; BCR, B-cell receptor; TCR, T-cell receptor; IL, interleukin; JAK, Janus kinase; STAT, signal transducer and activator of transcription 3; MHC, major histocompatibility complex; INFy, interferon gamma; DC, dendritic cell; TNF, tumor necrosis factor; MCSF, macrophage colony-stimulating factor.

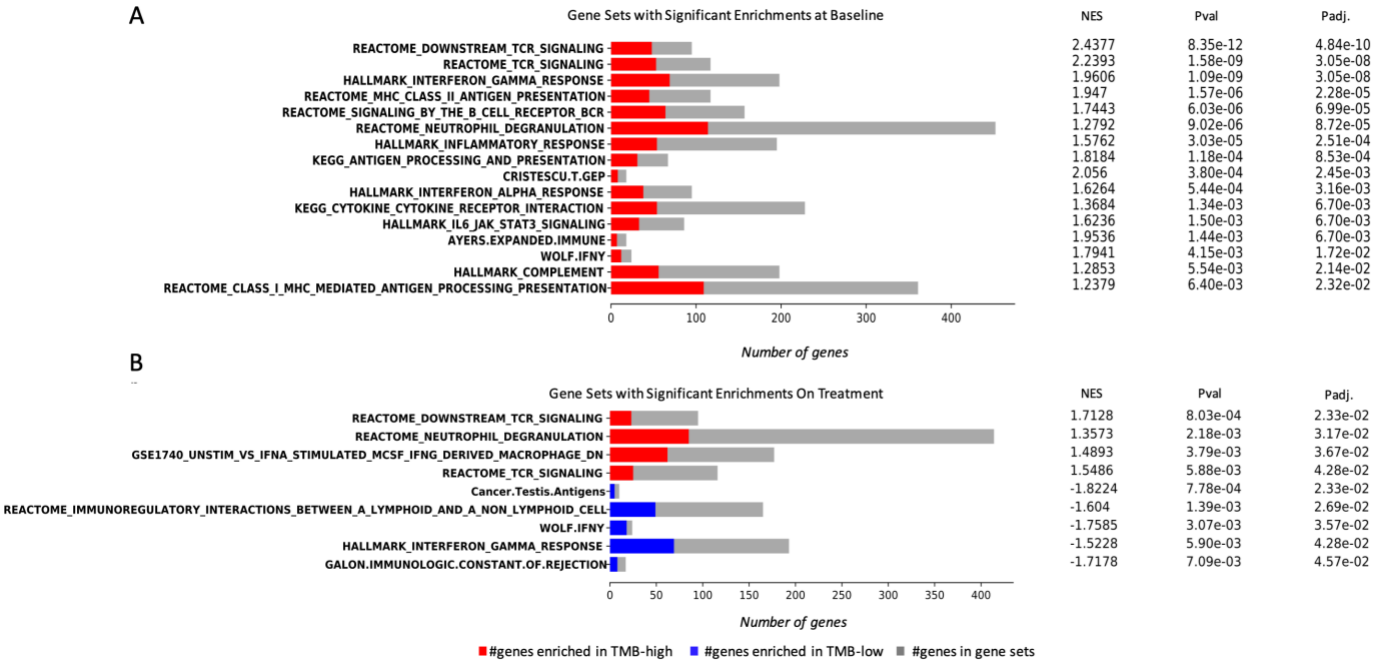

Figure 3.11 Gene set enrichment analyses (GSEA) for tumor mutation burden (TMB) high and low group. A selection of 58 genes sets related to inflammatory response, antigen presentations and type 2 immunity were used to understand the difference of gene expression in TMB high and low groups. All the gene sets that were significantly enriched in either the high or low groups are shown. Number of genes overlapped with gene sets that were overrepresented in the TMB high group are shown in red, and number of genes overlapped with gene sets overrepresented in the TMB low group are shown in blue. Overall, gene sets related to TCR/BCR signaling, MHC I & II antigen presentation and inflammatory response were enriched in the TMB high group at baseline. TCR signaling gene sets enrichment persisted in TMB high group on treatment, but interferon gamma gene response related gene sets were enriched in TMB low group on treatment.

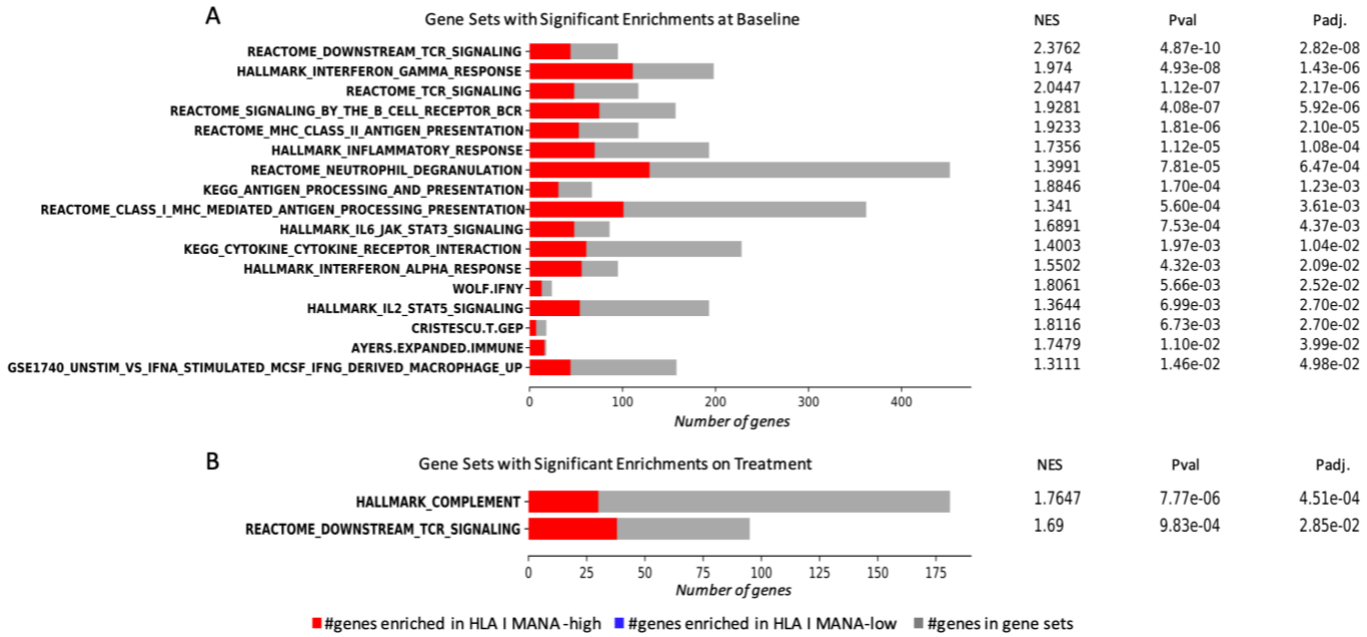

Figure 3.12 Gene Set Enrichment Analyses for HLA I MANA burden high vs low samples. A selection of 55 genes sets related to inflammatory response, antigen presentations and type 2 immunity were used to understand the difference of gene expression in HLA I MANA burden high and low groups. All the gene sets that were significantly enriched in either the high or low groups are shown. Number of genes overlapped with gene sets that were overrepresented in the HLA I MANA high burden group are shown in red, and number of genes overlapped with gene sets overrepresented in the HLA I MANA burden low group are shown in blue. Overall, gene sets related to TCR/BCR signaling, MHC I & II antigen presentation and inflammatory response were enriched in the HLA I MANA high group at baseline. TCR signaling gene set enrichment persisted in HLA I MANA high group on treatment.

patients according to HLA class I MANA loads, I saw similar enrichments for TCR (FDR p=2.82e-8 for REACTOME_DOWNSTREAM_TCR_SIGNALING and 2.17e-6 for REACTOME_TCR_SIGNALING), BCR signaling (FDR p=5.92e-6) and MHC I and II antigen presentation (FDR p=3.61e-3 and 2.10e-5 respectively) at baseline for the high group. Only complement (FDR p=4.51e-4) and TCR signaling (FDR p=2.85e-2) were overrepresented in the HLA class I high MANA load group (Figure 3.12). Collectively, my findings suggest an inflamed TME for tumors harboring a high HLA class II burden with additional induction of adaptive immunity cascades after treatment with immune checkpoint blockade.

3.9 HLA Class II IMM Expression

RNA sequencing data was also used to determine HLA class II IMM expression. Immunogenic mutations were considered expressed if their genomic coordinates were spanned by at least 3 reads as defined in previous publications [101]. An average 32% (0-100%) of the identified HLA class II IMMs were expressed at baseline, and an average of 21% (0-100%) of the identified HLA class II IMMs were expressed on-treatment. As expected, high HLA class II IMM burden indicated higher levels of expressed HLA class II IMMs pre-treatment (MW p=1.26e-6) (Figure 3.13A). Responders also had significantly higher numbers of expressed HLA class II IMMs at baseline (MW P=0.007) (Figure 3.13B). Additionally, patients with high levels of expressed HLA class II IMM burdens trended towards longer overall survival (log rank p=0.14, HR=0.24, CI:0.03-1.89) (Figure 3.13C). Overall, Isaw a slight, but not significant, decrease in numbers of expressed HLA class II IMMs between baseline and on-treatment. This decrease of expressed HLA class II IMMs was comparable between the high and low HLA class II IMM groups when normalized to baseline burdens (MW P=0.19) (Figure 3.13D).

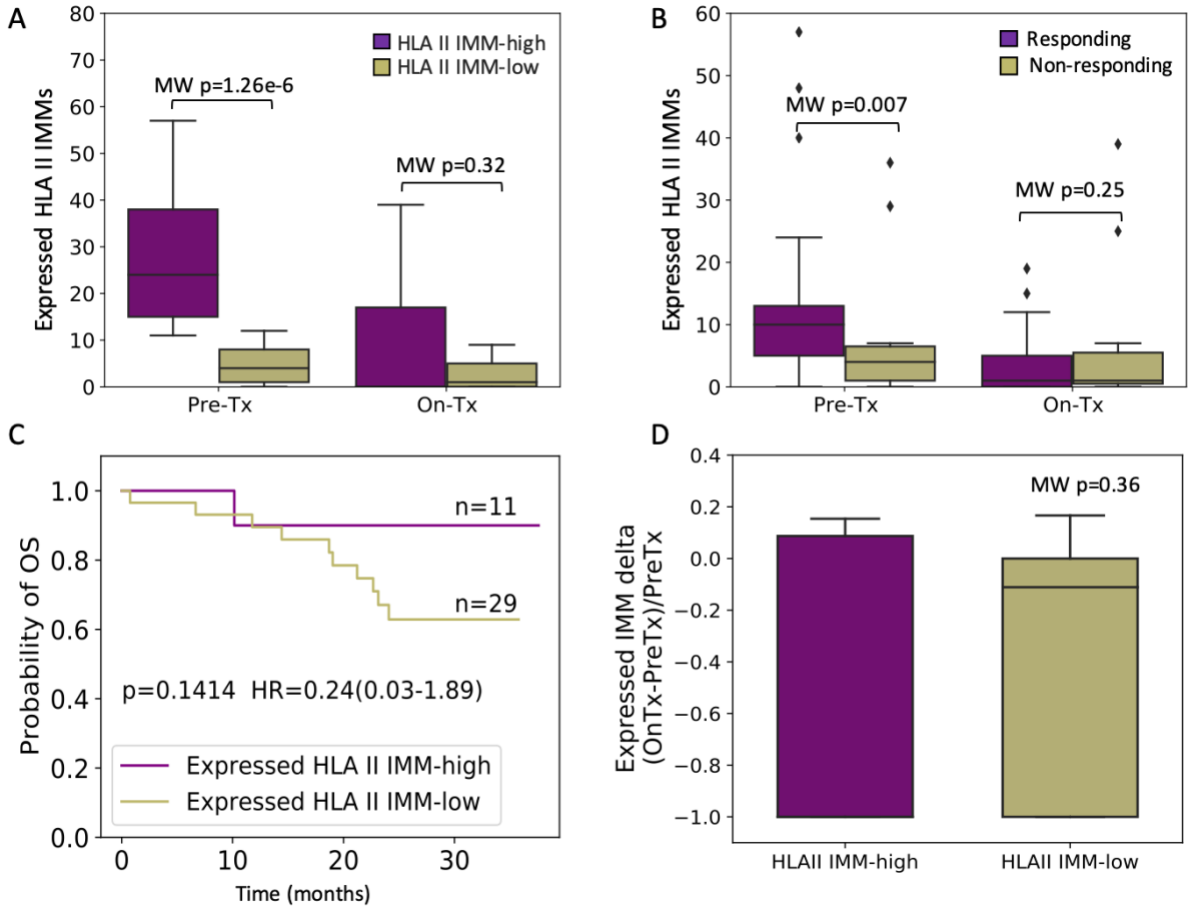

Figure 3.13 Expressed HLA class II IMM burdens. (A) Tumors that had high HLA class II IMMs had significantly high numbers of expressed HLA class II IMM burden pre-treatment (MW p=1.26e-6), but not on-treatment (MW p=0.32). (B) Responding tumors had significantly higher numbers of expressed HLA class II IMMs than non-responding tumors pre-treatment (MW p=0.007). Such difference is not seen on treatment (MW p=0.25). (C) Patients with higher numbers of expressed HLA class II IMMs (top 25%) trended towards longer OS than those with lower numbers of expressed HLA class II IMMs (bottom 75%) (log rank p=0.14, HR=0.25, CI:0.03-1.89). (D) Changes of Expressed HLA class II IMMs normalized to pre-treatment counts are shown between the HLA class II IMM high vs low groups. No difference is seen between the high and low groups (MW p=0.36).

Overall, this work supports the importance of HLA class II immunogenic mutations in anti-tumor immune responses that are reflected in clinical outcomes for patients treated with immune checkpoint blockade. Such efforts may inform selection criteria for the increasing number of patients receiving cancer immunotherapy. Furthermore, my findings may inform mutation selection for cancer vaccine approaches, expanding HLA class I-only neoepitope vaccination strategies.

Chapter 4 Sex-dimorphism assessed through ICB-treated patients' immunogenomic landscape

Sex-based response difference in ICB treated cohorts has been known. Here, I aimed to shed light on such difference in exploring patients' immunogenic landscape using the MHCnuggets framework. First, I evaluated the background association between immunogenic mutation load and HLA zygosity in a sex-dependent manner in NSCLC tumors from TCGA. To explore sex-specific genomic features linked with ICB response, I analyzed whole exome sequence data of two independent cohorts of patients with NSCLC treated with ICBs and assessed differences in the MHC I and II-restricted immunogenic mutation repertoire combined with the germline and somatic HLA class I and II zygosity. My findings highlight sex-specific differences in immunegenomic determinants of response to ICB that may impact clinical decision making. This chapter is based on published work in Scott and Shao *et al.* 2022 [134], further information and data are available at https://doi.org/10.3389/fonc.2022.945798.

4.1 ICB-treated cohort compilation

4.1.1 Cohort characteristics

The primary NSCLC cohort consisted of 89 patients treated with ICB therapy at Johns Hopkins Sidney Kimmel Cancer Center and the Nederlands Kanker Institute; whole exome sequence and clinical metadata were retrieved from the original publication [25]. In addition to sequence annotation of activating EGFR mutations, review of clinical next generation sequencing data was performed to identify ALK, ROS1, and RET rearrangements. ALK rearrangement status was not available for 7 of 89 tumors, ROS1 and RET rearrangement status was not available for 13 tumors, including 2 in patients with no history of tobacco exposure (1 male and 1 female). Whole exome sequence data from a published cohort of 34 NSCLC patients treated with PD-1 blockade (NSCLC validation cohort) were obtained and analyzed to validate key findings from the primary NSCLC cohort[100]. Driver gene fusion analyses were not available for this cohort. Genomic and demographic information of 286 lung adenocarcinoma (LUAD) and 196 lung squamous cell carcinoma (LUSC) samples from The Cancer Genome Atlas (TCGA) were retrieved from the NCI Genomic Data Commons (https://gdc.cancer.gov/about-data/publications/mc3-2017). Clinical annotations of tumors and structural variants including gene fusions were accessed using the TCGA clinical data resource [135].

4.1.2 Definition of objective response to ICB treatments

ICB-treated patients were administered with anti-PD-1 or anti-PD-L1 therapy alone or in combination with anti-CTLA-4 therapy or chemotherapy. Given the challenges with conventional radiologic response assessments that may underestimate the unique patterns and

timing of response to immune targeted therapies, I defined response as durable clinical benefit if complete response, partial response, or stable disease was achieved with a duration of >6 months. Responding and non-responding tumors, therefore, refer to patients attaining durable clinical benefit (DCB) and non-durable clinical benefit (NDB), respectively. Overall survival was used to determine long-term outcome for the primary NSCLC cohort. Progression free survival only was available for the NSCLC validation cohort.

4.2 Identification of putative immunogenic mutations.

4.2.1 Somatic mutation extraction.

Missense somatic mutation calls for both primary and validation cohorts were extracted from the original publications [25, 136]. Missense somatic mutation calls for the TCGA samples were obtained from Multi-Center Mutation Calling in Multiple Cancers (MC3; https://gdc.cancer.gov/about-data/publications/mc3-2017).

4.2.2 HLA genotyping

HLA class I and class II germline genotypes for both primary and validation cohorts were identified as previously described [25, 93]. In brief, using whole exome sequencing, each samples' HLA class I germline haplotype (HLA-A, HLA-B, and HLA-C) were identified with OptiType [81]. For HLA class II haplotypes, an ensemble approach utilizing SOAP-HLA [108] and xHLA [109] was employed such that xHLA was used to determine HLA-DPB1, HLA-DQB1 and HLA-DRB1 haplotypes while SOAP-HLA was used to determine HLA-DPA1 and HLA-DQA1 haplotypes. HLA class I germline genotypes for TCGA NSCLC samples were

obtained from the TCGA landscape publication [79] that utilized OptiType [81], while HLA class II germline genotypes were retrieved from the publication of Marty-Pyke et al., 2018 [137].

4.2.3 Immunogenic mutation (IMM) calls

The burdens of immunogenic mutations (IMM) were computed as described in Chapter 3 and previous publication [93]. In brief, using varcode (https://github.com/openvax/varcode), silent and nonsense mutations were filtered out of each patient's mutation profiles, and mutant peptide sequences surrounding the affected amino acid for all missense mutations were extracted. Windowing around the affected amino acid, 8-11mers were extracted for HLA class I analyses, and 12-20mers, HLA class II analyses. Next, I employed MHCnuggets [31] to obtain the ranks of the binding affinities of all the mutation containing peptides against the respective HLA class I and II haplotypes of the patients[93]. Each candidate peptide's predicted MHC binding affinities were compared against the MHC binding affinities of a list of 100,000 human proteome peptides[93]. Using a rank threshold of 0.01, I considered all the epitopes with predicted binding affinity over this rank immunogenic neoantigens. A putative IMM was thus defined as a missense mutation that contain as least one predicted mutation-associated neoantigens (MANA) fulfilling these criteria.

4.3 HLA loss of heterozygosity analyses

For the primary and validation cohorts, loss of HLA class I germline molecules in the tumors were determined by LOHHLA [138], for which allele specific copy numbers of HLA class I locus were realigned to patient specific reference sequences and corrected by tumor purity and ploidy. Tumor purity and ploidy were assessed in each sample analyzed as described

previously [25]. In brief, somatic copy number profiles were first determined by mapping reads to exonic and intronic regions (bins) of the genome while correcting for region size, CG content and sequence complexity [139]. Next, tumor copy number profiles were compared to a reference panel of matched normal samples to derive copy ratio values. Circular binary segmentation [140] was next applied to copy ratio profiles to determine genomic segment boundaries. Segmental copy ratio values and minor allele frequency of heterozygous single nucleotide polymorphisms (SNPs) overlapping the segment were used to estimate tumor purity and ploidy throughout the genome; all possible combination of tumor purity and ploidy were evaluated for the optimal combination based on maximum likelihood estimation. For HLA class II germline molecule loss in the tumors and all HLA molecule loss in the TCGA samples, the minor allele copy numbers were utilized as previously described [25]. Briefly, loss of heterozygosity (LOH) occurred when minor allele copy number of the overlapping genomic region equaled zero. HLA allele specific copy number information of the TCGA samples were obtained from analyses of SNP6 copy number array data on Synapse (https://www.synapse.org/#!Synapse:syn1710464).

4.4 Mutational signatures extraction

The contribution of smoking-related mutational signatures in the mutational spectra of NSCLC tumors in the primary and validation cohorts were extracted from the original publication [25]. In brief, the deconstructSigs R package was utilized [141] to calculate the contribution of COSMIC smoking signature 4 from all the coding point mutations in their trinucleotide context [142].

4.5 Statistical analysis

Differences between responding and non-responding tumors were evaluated using chi-squared or Fisher's exact test for categorical variables and the Mann-Whitney (MW) test for continuous variables. Where noted, tumors were classified based on their missense tumor mutational burden (TMB) or IMM load as high or low using the second tertile as a cut-off point. Median point estimate and 95% confidence interval (CI) for overall survival and progression free survival were estimated by the Kaplan-Meier method and survival curves were compared through the nonparametric log-rank test. Univariate Cox proportional hazards regression analysis was used to determine the impact of individual parameters on survival outcomes. All p values were based on two-sided testing and differences were considered significant at $p < 0.05$.

4.6 Sex-dependent association between immunogenic mutation load and HLA zygosity

Cancer immunoediting selects tumor clones that escape immune control throughout tumorigenesis and cancer evolution likely in a sex-specific immune context [143, 144]. To investigate sex-based background differences in the immunogenomic landscape of NSCLC tumors independent of therapy, I analyzed HLA zygosity and IMM load in 482 lung cancer tumors from TCGA, including 279 males and 203 females. In males, Idid not detect an association between tumor IMM load and germline HLA class I or class II homozygosity (HLA class I MW p=0.64, HLA class II MW p=0.64; Figures 4.1A-B).

Interestingly, and in contrast with previous findings pointing towards a lack of association between HLA germline diversity and TMB [25], the tumor IMM load in females was positively correlated with germline HLA allele diversity, particularly for HLA class II associated

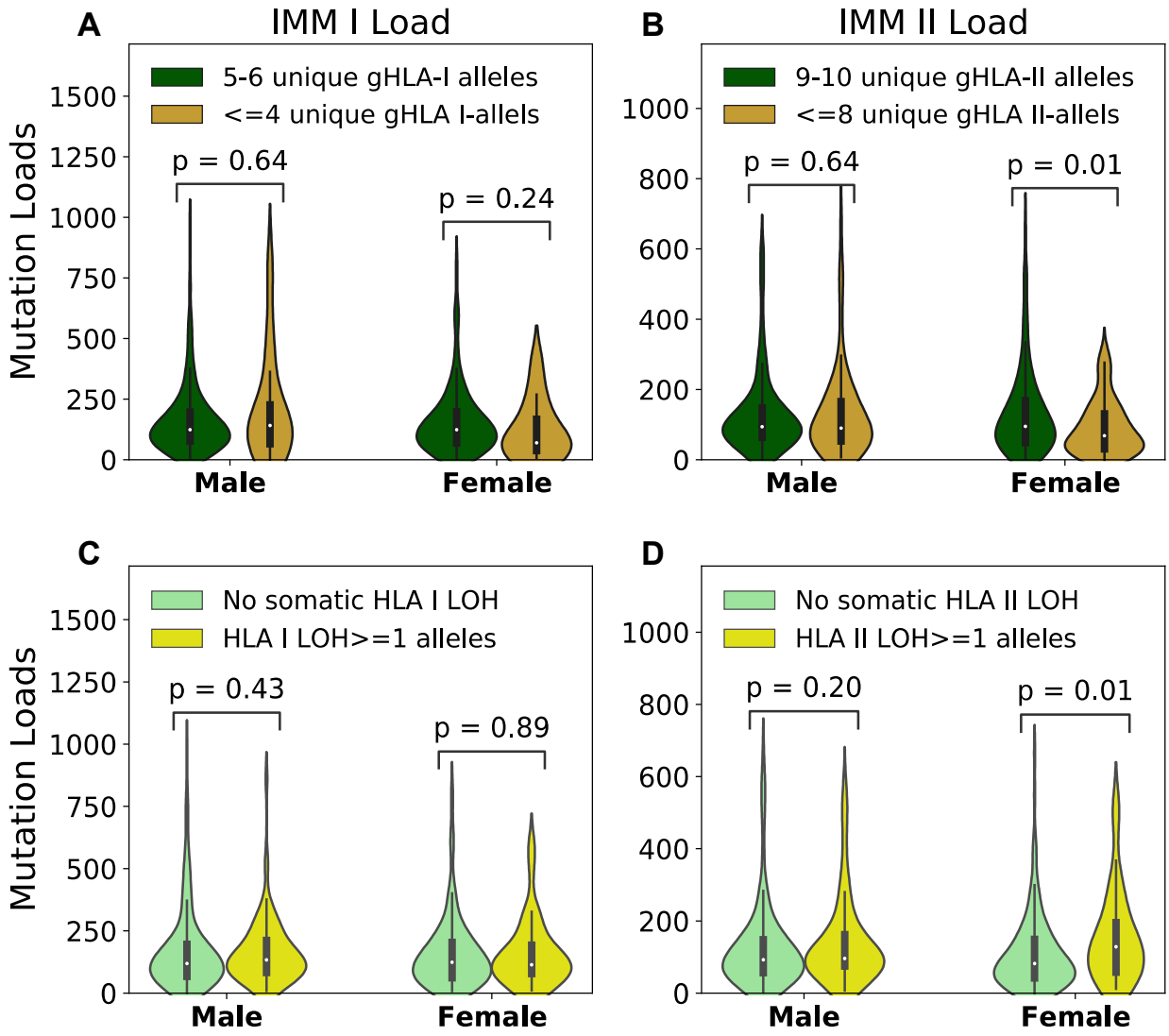

Figure 4.1 Background immunogenic mutation association with HLA diversity in TCGA-NSCLC cohort. (A) No association between class I IMM loads and tumor HLA I diversities was found in either males (MW p=0.64) or females (MW p=0.24). (B) Female tumors with high germline HLA II diversity had significantly higher class II IMM loads (MW p=0.01). Male tumors did not show class II IMM load difference between the high germline HLA II diversity and the low germline HLA II diversity groups (MW p=0.64). (C) No association between loss of heterozygosity (LOH) of HLA I alleles and class I IMM loads were identified in either males (MW p=0.43) or females (MW p=0.89). (D) Female tumors lost ≥1 HLA II alleles had higher class II IMM loads than those with no LOH (MW p=0.01). LOH for HLA II alleles did not associate with class II IMM load difference in male tumors (MW p=0.20).

IMMs. Among female patients with germline homozygosity in two or more HLA class II alleles, the median predicted class II IMM load was 68.5 as compared to 95 among tumors in females with the most diverse HLA class II repertoire (MW p=0.01; Figure 4.1B). In contrast, the median class II IMM load among males with germline HLA-II homozygosity was 90, as compared to 94 among those with more diverse HLA-II alleles (MW p=0.64). When excluding 45 patients (15 male, 30 female) with activating EGFR mutations or ALK, ROS1, or RET gene fusions, this trend persisted in females, with median tumor class II IMM load of 103 in patients with at least 9 unique germline HLA-II alleles, as compared to 88 in patients with a higher degree of germline HLA homozygosity (MW p=0.06).

While germline HLA repertoire contributes to tumor-immune recognition and immunoediting, tumors with high neoantigen burdens may escape immune surveillance by somatic loss of heterozygosity of HLA alleles [138]. In considering the role of somatic loss of heterozygosity in tumor evolution as a means of immune evasion, I next classified tumors with at least one fewer unique HLA allele in tumor as compared to germline and observed that loss of heterozygosity at the HLA-II locus was associated with high class II IMM load only in females (MW p=0.014; Figures 4.1C-D). Again, this association persisted when tumors with EGFR, ALK, ROS1, or RET driver mutations were excluded (MW p=0.02).

4.7 HLA class I and II immunogenic mutation load predict ICB response in females

The primary NSCLC cohort (Anagnostou) included 89 adults with NSCLC treated with ICB therapy, consisting of 46 male and 43 female patients with a median age at ICB treatment initiation of 64 years. The NSCLC validation cohort (Rizvi) included 34 adults with NSCLC

treated with pembrolizumab, consisting of 16 male and 18 female patients with a median age of 62.5 years. Demographic, tumor, and treatment information for both cohorts have been previously published [25, 100]. Two patients in the primary NSCLC cohort were not evaluable for the durable clinical benefit endpoint.

No significant baseline differences were observed between NSCLC tumors of males and females with respect to TMB, class I IMM load, class II IMM load, or mutational smoking signature (for all comparisons MW $p \geqslant 0.8$). Similarly, there was no significant difference in rate of durable clinical benefit (DCB; Fisher's Exact p=1.0), progression free survival (log-rank p=0.94, HR=1.02, 95% CI 0.62-1.68) or overall survival (log-rank p=0.39, HR=1.31, 95% CI 0.7-2.42) between males and females following ICB therapy.

To investigate the difference in immunogenomic features predicting ICB response between males and females, I measured TMB, class I and II IMM load, and mutational smoking signature in responding and non-responding tumors by each sex. In female patients, DCB following ICB therapy was significantly associated with higher TMB (MW p=0.005), class I IMM load (MW p=0.005), class II IMM load (MW p=0.004), and mutational smoking signature (MW p=0.0006), while none of these features were significantly different between responding and non-responding tumors in male patients (Figures 4.2A–D).

In order to incorporate the effect of gender- or sex-biased differences like tobacco exposure and prevalence of tumors with driver mutations that typically do not respond to ICB therapy [145, 146], I further stratified the primary cohort by tobacco history and driver gene mutation status. Six tumors harbored activating EGFR mutations, while no fusions of ALK, ROS1, or RET were detected in those with available testing. As expected, none of the six tumors harboring activating EGFR alterations (5 female, 1 male) demonstrated DCB in response to ICB

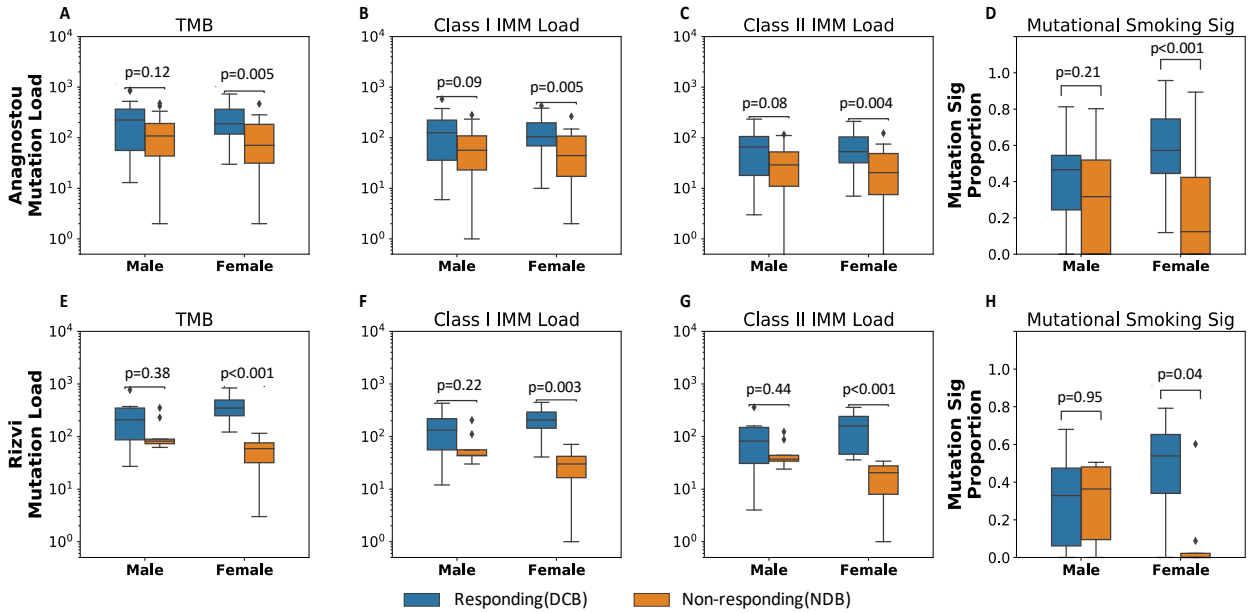

Figure 4.2 Immunogenic mutation load distinguishes responding from non-responding tumors in females who received immune checkpoint blockade. In the primary NSCLC cohort (Anagnostou), (A) female responding tumors harbored significantly higher TMB than female non-responding tumors (MW p=0.005). Similarly, (B) Class I and (C) class II IMM loads separated female response groups (Class I IMM loads MW p=0.005; Class II IMM loads MW p=0.004), but not male response groups (Class I IMM loads: MW p=0.09; Class II IMM loads: MW p=0.08). (D) Smoking mutational signature levels also only differed in female response groups (MW p=0.0006), but not male response groups (MW p=0.21). These results were corroborated in the validation cohort (Rizvi) with immunogenomic features, (E) TMB, (F) Class I IMM loads, (G) Class II IMM loads, and (H) mutational smoking signature.

therapy. Excluding these six patients from the cohort, the somatic mutational features remained predictive of DCB for NSCLC in females but not in males, including TMB (females MW p=0.03, males MW p=0.13), class I IMM load (females MW p=0.03, males MW p=0.10), class II IMM load (females MW p=0.02, males MW p=0.10), and mutational smoking signature (females MW p=0.005, males MW p=0.26). Among patients with a self-reported history of smoking (35 males, 33 females), the mutational smoking signature trended toward association of ICB response only in tumors in females (median signature contribution of 0.57 in the DCB group versus 0.38 for the NDB group, p=0.06). In contrast, there was no difference in mutational smoking signature in male tumors by therapeutic response (median signature contribution of 0.47 in the DCB group versus median signature contribution of 0.46 in the NDB group, p=0.95).

These findings were corroborated in an independent validation cohort (Rizvi) of 34 patients with ICI-treated NSCLC [100]. Among tumors of 18 female patients in the cohort, TMB, class I IMM load, class II IMM load and mutational smoking signature were significantly higher in those attaining DCB as compared to NDB following ICB therapy, including TMB (MW p=0.0009), class I IMM load (MW p=0.003), class II IMM load (MW p=0.0009), and mutational smoking signature (MW p=0.04; Figures 4.2E–H). Among 16 male patients in the cohort, there was no significant difference in these features between responding and non-responding tumors. Similar to the primary cohort, these differences persisted when excluding three patients with tumors harboring EGFR alterations.

4.8 Combined HLA zygosity with immunogenic mutation load predicts ICB response in males

Next, I considered the contribution of antigen presentation capacity in addition to IMM load as associated with differential clinical responses to ICB in males and females. I assessed HLA genetic variation as an indicator of neoantigen presentation capacity and its impact on outcomes by sex. Patient germline and tumor HLA haplotypes were classified as high HLA diversity ($\geqslant$5 unique alleles for HLA class I and $\geqslant$9 unique alleles for HLA class II) or low HLA diversity groups ($\leqslant$4 unique HLA I class alleles or $\leqslant$8 HLA class II alleles). There was no significant correlation with germline or tumor HLA zygosity alone and ICB benefit in either sex.

Even in tumors with maximal HLA heterozygosity, the potential immunogenicity of mutation-associated neoantigens depends on effective presentation by the patient's unique inherited MHC repertoire. In contrast to TMB, the predicted IMM load incorporates MHC affinity to estimate the number of neoantigens likely to stimulate an anti-tumor immune response

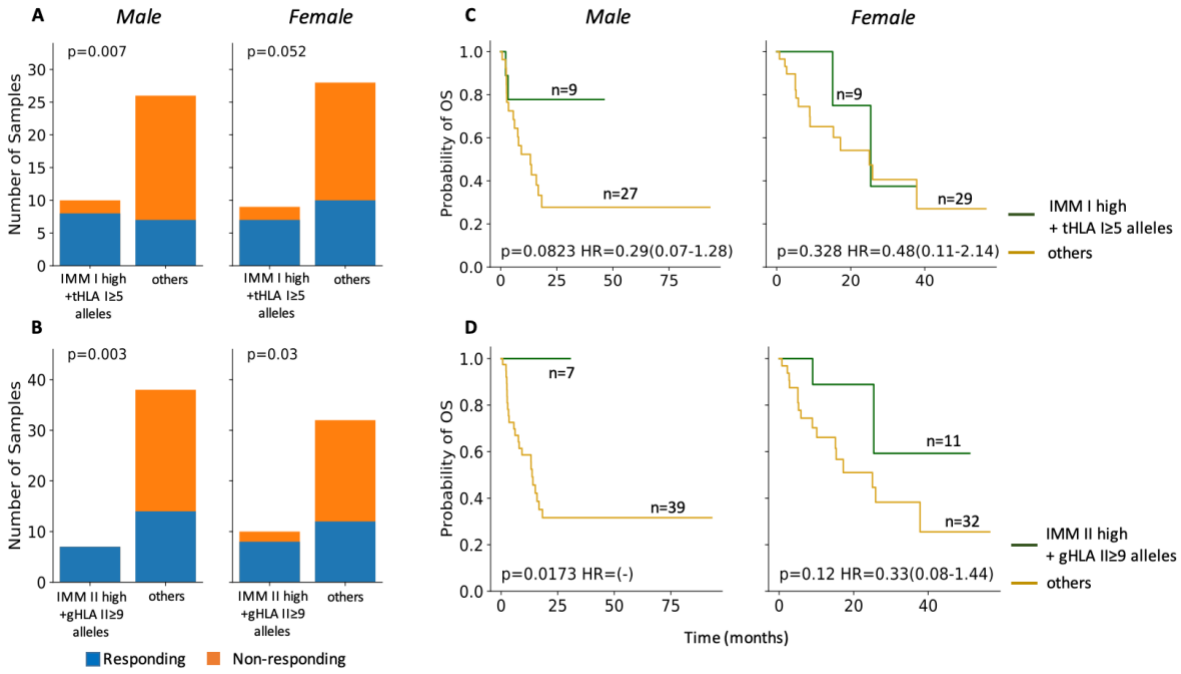

Figure 4.3 HLA heterozygosity combined with immunogenic mutation loads predicted ICB response and survival in males. (A) Male and female tumors with high IMM-I loads and high tumor HLA I diversity (tHLA I ≥5) co-occurred with response to ICB treatments (male Fisher's Exact p=0.02, female Fisher's Exact p=0.05). (B) Both male and female tumors with high HLA II restricted IMMs and high germline HLA II diversity (gHLA II ≥9) co-occurred with ICB response (male: Fisher's Exact p=0.003; female: Fisher's Exact p=0.03). (C) Male patients with high tumor IMM-I loads and high tumor HLA I diversity trended towards longer overall survival (OS) (log-rank p=0.1), while their female counterparts did not (log-rank p=0.33). (D) Combined high tumor IMM-II loads and high germline HLA II diversity was associated with significantly longer OS among males (log-rank p=0.02) but not females (log-rank p=0.12). Hazard ratio (HR) shown with 95% confidence interval.

[93]. In the primary ICB-treated cohort (Anagnostou), response to ICB correlated with combined high IMM load and high HLA diversity in tumors in both males and females (Figures 4.3A-B). Interestingly, significantly improved OS was only observed in males with high IMM load and high HLA diversity, particularly for HLA class II (log-rank p=0.017; Figures 4.3C-D). This finding in males is consistent with my hypothesis that more highly mutated tumors with intact antigen presentation capacity will benefit from improved outcomes following ICB therapy. In females, HLA diversity did not have a combinatorial effect on IMM load for predicting survival after ICB therapy (Figures 4.3C-D); these findings suggest that alternative mechanisms of

immune escape likely contribute to the differential sex-dependent response and outcome following ICB therapy.

In summary, the dimorphic activity of the immune system in males and females strongly influences tumorigenesis by immune selection and has important implications in the growing field of cancer immunotherapy. My data highlight that the interpretation of biomarkers and the mutational landscape of tumors for prediction of ICB therapy response requires consideration of the biological sex context in which the tumor has evolved.

Chapter 5 Discussion

5.1 Computational solution for a more precise immunotherapy patient stratification based on mutation-associated neoantigen identification

In this thesis, I developed MHCnuggets, a flexible open-source platform for MHC–peptide binding prediction that can handle common MHC class I and II alleles as well as rare alleles of both classes, to better understand neoantigen profiles of immunotherapy treated patients. The LSTM network architecture of MHCnuggets can handle peptide sequences of arbitrary length, without shortening or splitting. The single neural network architecture requires fewer hyperparameters than more complex architectures and simplifies network training. In addition, my neural network transfer learning protocols allow for parameter sharing among allele-specific, binding affinity– and HLAp-trained networks. When trained on binding affinity data, MHCnuggets performs as well as other current methods. When trained on both binding affinity and HLAp data, I demonstrate improved PPVn on an independent HLAp test set, with respect to other methods that use both binding affinity and HLAp data. Although PPVn was lowest for the independent HLAp test set for all methods, this result is likely due to systematic differences between training HLAp data (monoallelic B-cell lines) [61] and the test data comprised of seven multiallelic cell lines (HeLA, HTC116, JY, fibroblasts, SupB15, HCC1937, and HCC1143) [22, 58], yielding a more challenging prediction problem. I attribute MHCnuggets' improvement on the independent test set with respect to other methods to optimization of PPVn in my network training protocol and my implementation of transfer learning to integrate information from binding affinity and HLAp measurements. The

performance of all methods is generally highest when both training and test data come from similar binding affinity experiments, but performance improvement on HLAp data is more biologically relevant [147].

I demonstrate improved scalability by comparing the runtime of MHCnuggets on 1 million peptides to comparable methods, and further by processing over 26 million expressed peptide–allele pairs across TCGA samples in under 2.3 hours. I identified 101,326 unique IMMs harbored by patients using 26 cancer types sequenced by the TCGA, based on transcriptional abundance and differential binding affinity compared with reference peptides. These results contrast with a previous report of neoantigens in TCGA patients in several respects. Rech and colleagues [148] applied a minimum expression threshold of 1 RNA-seq read count, an IEDB-recommended combination of neoantigen predictors derived primarily from different versions of NetMHC, and IC50 threshold of 50 nmol/L to identify strong MHC binders. Their approach yielded 495,793 predicted class I classically defined neoantigen peptides (each harboring a single immunogenic mutation) from 6,324 patients in 26 cancer types. In my study, high variability in neoantigen burden across cancer types was observed. The difference between predicted IMM and neoantigen burden in the two studies is likely due to differences in RNA expression threshold and the low false-positive rate of MHCnuggets compared with IEDB-recommended tools.

By extending the MHCnuggets framework to a rank system, I was able to characterize the HLA class II-restricted immunogenic mutation and associated neoantigen landscapes for patients with lung cancer and melanoma receiving immune checkpoint blockade therapies. While HLA class II IMM burden was, overall, an order of magnitude smaller than TMB, IMM burden was more strongly associated with clinical response and survival. The majority of HLA class II

IMMs did not overlap with HLA class I IMMs, suggesting that the former represent a distinct subset of immunogenic mutations that may play a critical role in tumor rejection, complementing HLA class I restricted neoepitope induction of CD8+ T cell cytotoxicity.

While HLA class I neoantigens have been historically at the epicenter of neoantigen-driven tumor rejection, tumor antigen specific CD4+ T-cell activity in the TME has been shown to be a prerequisite for spontaneous and immunotherapy-related tumor rejection [36, 149]. The less stringent sequence and length requirements for neopeptide binding to the HLA class II proteins compared to HLA class I restricted epitopes, further supports the notion that a sizable fraction of HLA class II epitopes may be presented [150]. In murine tumor models, a large fraction of the immunogenic mutanome has been shown to be recognized by CD4+ T-cells and subsequent vaccination with such epitopes may confer sustained tumor clearance [35]. These observations suggest that HLA class II restricted immunogenic mutations and associated neoantigens may have key and non-overlapping -with HLA class I neoepitope- functions in anti-tumor immune responses. HLA class I germline heterozygosity has previously been found to impact ICB treatment outcomes [57]. This notion, however, was challenged by a recent study where no germline HLA variation effects were identified for pembrolizumab treated patients [151]. Here, I found that HLA class II variation combined with HLA class II IMM burden was associated with improved survival for ICB treated patients. However, when HLA class II IMM burden was controlled, HLA class II variation alone was not prognostic. The joint effect of HLA class II variation and HLA class II IMM burden was thus largely contributed by HLA class II IMM burden. Collectively, my findings suggest a role for immunogenic HLA class II restricted mutations and neoantigens in predicting clinical outcomes with immune checkpoint blockade.

Heterozygous HLA-DPs, when combined with HLA class II IMM burdens, showed the strongest combined prognostic effects out of the HLA II genes as suggested by previous studies [57, 101]. Previous identification of HLA-DP alleles' associations with overactive immunity and overrepresentation of endogenous proteins in the epitopes' origins also suggested unique abilities of HLA-DPs in presenting self-peptides [152-155]. HLA-DQ and HLA-DR locI was more heterozygous than HLA-DP, with >90% of the samples heterozygous at those loci. Consistent with previous work [114], the most frequent HLA class II strong binding MANAs were identified with HLA-DQ alleles and the least frequent with HLA-DR alleles. However, HLA-DQ's role in central tolerance in the thymus [114] suggests that HLA-DQ binding MANAs may have already achieved central tolerance. HLA-DRs, on the other hand, bind epitopes originating from acute infections more readily [114]. As such, different HLA class II genotypes demonstrated different binding characteristics and functions, thus should be also considered when understanding patients HLA class II IMM landscapes.

Importantly, in studying the tumor microenvironment of tumors with a high content of HLA class II IMMs by RNA sequence data deconvolution, I identified an increased density of CD4+ memory activated T-cells, T follicular helper cells, and M1 macrophages, which are critical components of successful tumor rejection [156, 157]. Gene set enrichment analyses leveraging transcriptome data prior to- and during immune checkpoint blockade revealed a prominent enrichment in adaptive immunity programs, with increased representation of interferon-γ, HLA class II antigen processing and presentation and TCR/BCR signaling pathway gene sets. Notably, in the NSCLC cohorts, tumors with high HLA class II IMM burden also had higher PD-L1 expression and were encountered in individuals with a smoking history. These findings suggest that HLA class II IMM burden is affecting tumor immune surveillance and

potentially reshaping the tumor microenvironment towards a more inflamed state, ultimately promoting tumor rejection.

Further MANA and IMM analyses through the lens of sex-dimorphism shed lights on intrinsic sex-related immunogenomic difference in NSCLC cohorts. Growing evidence points to a sexual dimorphism in response to immunotherapy across multiple tumors that likely arises from differences in immune surveillance in males and females. During tumorigenesis, immune surveillance shapes tumor evolution and selection pressure drives tumor immune escape. I demonstrated that somatic mutational features, including class I and II predicted IMM load as well as TMB, are associated with durable clinical benefit following ICB therapy in females but not in males. This finding is consistent with recent reports demonstrating that as a predictive biomarker for ICB response, TMB performed better in females than in males, an effect that appears to be independent of smoking signature and oncogenic-driver mutations like EGFR [54, 158].

Antigen presentation capacity in the form of a diverse repertoire of HLA-encoded MHC molecules allows for effective presentation of more tumor-associated neopeptide candidates to stimulate an anti-tumor immune response. As a predictor of response to ICB therapy, HLA-I germline homozygosity in at least one locus and HLA-I loss of heterozygosity have been associated with worse survival and were significantly more negatively prognostic when combined with low TMB [25, 57, 159, 160]. I have previously shown that fewer than five unique tumor HLA class I alleles combined with low TMB was associated with lower CD8+ T cell infiltration and was predictive of worse overall survival in this NSCLC cohort, suggesting HLA loss is an adaptive mechanism for immune evasion in tumors with high mutation burden [25]. In further examining these patterns by sex, I found that the combination of diverse HLA alleles and

high IMM load was predictive of improved survival following ICB therapy in males but not in females, and particularly with respect to HLA class II. This pattern suggests that in females, more immunogenic tumors with high neoantigen burden and intact antigen presentation capacity are under selective pressure to develop alternative mechanisms of immune evasion, favoring immune escape and poor ICB response. In other words, females with already robust tumor-specific immunity will have a lower potential for effect with ICB therapy (i.e. a lower therapeutic index) [161], while males with less baseline anti-tumor response may attain a greater magnitude of therapeutic response.

In further exploring the immunogenomic landscape that evolves in tumors in males and females, I found that maximal HLA class II germline diversity is associated with higher IMM load in females but not in males. This finding indicates that tumors with intact antigen presentation capacity may acquire more novel oncogenic mutations to survive in an environment with more effective anti-tumor immune surveillance. Castro et al. recently demonstrated that females are more likely to accumulate driver mutations early in tumorigenesis that are less effectively presented by their inherited HLA genotype, particularly for HLA class II [144], and thus higher mutation burden may not always reflect increased immunogenicity in females. These data suggest that in an environment of stronger immune selection pressure, tumors in females have developed multiple mechanisms to evade tumor-specific T cell responses and support my finding that combined high IMM and HLA-II heterozygosity were positive predictors of post-ICB survival in males but not females.

Adaptive tumor alterations in antigen presentation machinery, predominately through selection for loss of heterozygosity of HLA alleles, also promotes immune escape by decreased antigen presentation capacity [138]. Loss of heterozygosity at HLA-I loci is common and has

been noted in 30-40% of several published NSCLC cohorts [25, 138, 162]. Somatic HLA-I loss and low TMB has been associated with poor ICB response [57] and correcting for mutation-associated neoantigens presented by lost HLA alleles identified high TMB tumors with poor response to ICB therapy [162]. In analyzing this relationship by sex, I found that HLA-II loss of heterozygosity is also strongly associated with higher IMM load only in tumors in females. Thus, loss of HLA expression as an adaptive mechanism for immune evasion in tumors with high mutational load may be a more prominent influence in females.

In summary, by MHCnuggets, I explored the mutation-associated neoantigens and immunogenic mutation landscape in both NSCLC and melanoma cohorts. By identifying HLA class II IMM burden as a biomarker and by understanding inherent immunogenic sex-difference, these works have help to provide a more precise ICB-treated patient stratification computationally.

5.2 Future work directions

5.2.1 Incorporating further considerations of the biological process of peptide presentation in computational predictions

Indeed, the context of a peptide sequence, such as what sequences are flanking, its source protein and the expression of the source protein, is informative for MHC ligand prediction [21, 61]. This type of information is available only for a limited number of HLAp data sets, which were unavailable to us for training purposes. As more well-characterized HLAp data sets become available, I will further develop MHCnuggets to include these features. Additionally, I was only able to obtain monoalleleic HLAp training datasets for HLA class I alleles. As I and others have demonstrated that incorporating HLAp data does improve peptide presentation prediction, future

development of MHCnuggets should extend HLAp data in the training of HLA class II alleles should they become available.

5.2.2 Considering immunogenic mutations beyond missense mutations

In the current studies, only MANAs derived from missense mutations were considered, therefore this effort did not evaluate putative neoantigens from indels, splice variants, and somatic gene fusions. As those mutations in general can generate "more foreign" antigens, they often could induce a more effective tumor rejection [163]. In the current scheme of understanding immunogenic mutation burdens, the count of missense mutations could potentially overshadow the counts of indels or structural variants. As such, future work of identifying IMMs beyond missense mutations would need to consider assigning proper weights to indels or structural variants when counting mutation burden. Finally, I currently considered all HLA class II immunogenic mutations to be equally contributing to anti-tumor immune responses, future studies will evaluate the contribution of dual HLA class I and II IMMs and highly expressed HLA class II IMM derived neoepitopes in driving clinical outcome in the context of immunotherapy.

5.2.3 Exploring the multiple sex-based variables at play in cancer development and the immune response

Cofactors related to additional hormonal, environmental, and genetic influences are all sex-based variables that should be further explored to understand cancer development. Tumors included in each cohort represented a single time point in tumorigenesis, and all ICB treated patients had advanced disease. Considering recent incorporation of immunotherapy in early stage

and locally advanced NSCLC [164, 165], it will be important to understand the evolution of these sex-based immunogenomic features and their context in different stages of disease. Similarly, continuing to examine these features with consideration of sex-based differences in response to combination chemotherapy and immunotherapy will be an important next step. Finally, though this analysis considered select driver mutations with known decreased response to ICB and female predominance, a multitude of additional targetable and non-targetable mutations have unclear impact on ICB response, such as KRAS, BRAF, MET, HER2, PI3K, PTEN, and DNA repair genes [146]. The complexity of the interaction of these drivers and other co-mutations with sex, smoking exposure, mutational burden and ICB response requires further investigation and emphasizes the need for intricate and individualized modeling for predictive biomarkers.

Bibliography

[1] V. Anagnostou, K. N. Smith, P. M. Forde, N. Niknafs, R. Bhattacharya, J. White, and others, "Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non–Small Cell Lung Cancer," *Cancer Discovery*, 2017.

[2] M. Yarchoan, B. A. Johnson, 3rd, E. R. Lutz, D. A. Laheru, and E. M. Jaffee, "Targeting neoantigens to augment antitumour immunity," *Nat Rev Cancer,* vol. 17, no. 9, pp. 569, Aug 24, 2017.

[3] C. Lundegaard, O. Lund, S. Buus, and M. Nielsen, "Major histocompatibility complex class I binding predictions as a tool in epitope discovery," *Immunology,* vol. 130, no. 3, pp. 309-318, 2010.

[4] M. Andreatta, and M. Nielsen, "Gapped sequence alignment using artificial neural networks: application to the MHC class I system," *Bioinformatics,* vol. 32, no. 4, pp. 511-7, Feb 15, 2016.

[5] Y. Kim, J. Sidney, S. Buus, A. Sette, M. Nielsen, and B. Peters, "Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions," *BMC Bioinformatics,* vol. 15, no. 1, pp. 241-241, 2014.

[6] Y. Kim, J. Sidney, C. Pinilla, A. Sette, and B. Peters, "Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior," *BMC Bioinforma.,* vol. 10, 2009.

[7] T. Trolle, I. G. Metushi, J. A. Greenbaum, Y. Kim, J. Sidney, O. Lund, and others, "Automated benchmarking of peptide-MHC class I binding predictions," *Bioinformatics,* vol. 31, no. 13, pp. 2174-2174, 2015.

[8] M. Bonsack, S. Hoppe, J. Winter, D. Tichy, C. Zeller, M. Küpper, E. Schitter, R. Blatnik, and A. Riemer, "Performance evaluation of MHC class-I binding prediction tools based on an experimentally validated MHC-peptide binding dataset," *Cancer Immunology Research*, 2019.

[9] D. Gfeller, M. Bassani-Sternberg, J. Schmidt, and I. F. Luescher, "Current tools for predicting cancer-specific T cell immunity," *OncoImmunology,* vol. 5, no. 7, pp. e1177691-e1177691, 2016.

[10] X. S. Liu, and E. R. Mardis, "Applications of Immunogenomics to Cancer," *Cell,* vol. 168, no. 4, pp. 600-612, 2017.

[11] N. B. Editorial, "The problem with neoantigen prediction," *Nature Biotechnology,* vol. 35, pp. 97, 02/08/online, 2017.

[12] M. Wieczorek, E. T. Abualrous, J. Sticht, M. Alvaro-Benito, S. Stolzenberg, F. Noe, and C. Freund, "Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation," *Front Immunol,* vol. 8, pp. 292, 2017.

[13] Y. C. Lu, and P. F. Robbins, "Targeting neoantigens for cancer immunotherapy," *Int Immunol,* vol. 28, no. 7, pp. 365-70, Jul, 2016.

[14] V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, and M. Nielsen, "NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data," *The Journal of Immunology*, 2017.

[15] E. Karosiene, M. Rasmussen, T. Blicher, O. Lund, S. Buus, and M. Nielsen, "NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ," *Immunogenetics,* vol. 65, no. 10, pp. 711-724, 2013.

[16] C. Lundegaard, K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen, "NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11," *Nucleic Acids Research,* vol. 36, no. Web-Server-Issue, pp. 509-512, 2008.

[17] M. Nielsen, C. Lundegaard, T. Blicher, K. Lamberth, M. Harndahl, S. Justesen, and others, "NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence," *PLOS ONE,* vol. 2, no. 8, pp. 1-10, 2007.

[18] T. Odonnell, A. Rubinsteyn, M. Bonsack, A. Riemer, and J. Hammerbacher, "MHCflurry: open-source class I MHC binding affinity prediction," *bioRxiv*, 2017.

[19] M. Bassani-Sternberg, and G. Coukos, "Mass spectrometry-based antigen discovery for cancer immunotherapy," *Curr Opin Immunol,* vol. 41, pp. 9-17, Aug, 2016.

[20] T. N. Schumacher, and R. D. Schreiber, "Neoantigens in cancer immunotherapy," *Science,* vol. 348, no. 6230, pp. 69-74, Apr 3, 2015.

[21] B. Bulik-Sullivan, J. Busby, C. D. Palmer, M. J. Davis, T. Murphy, A. Clark, M. Busby, F. Duke, A. Yang, L. Young, N. C. Ojo, K. Caldwell, J. Abhyankar, T. Boucher, M. G. Hart, V. Makarov, V. T. Montpreville, O. Mercier, T. A. Chan, G. Scagliotti, P. Bironzo, S. Novello, N. Karachaliou, R. Rosell, I. Anderson, N. Gabrail, J. Hrom, C. Limvarapuss, K. Choquette, A. Spira, R. Rousseau, C. Voong, N. A. Rizvi, E. Fadel, M. Frattini, K. Jooss, M. Skoberne, J. Francis, and R. Yelensky, "Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification," *Nat Biotechnol*, Dec 17, 2018.

[22] M. Bassani-Sternberg, S. Pletscher-Frankild, L. J. Jensen, and M. Mann, "Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation," *Mol Cell Proteomics,* vol. 14, no. 3, pp. 658-73, Mar, 2015.

[23] K. M. Boehm, B. Bhinder, V. J. Raja, N. Dephoure, and O. Elemento, "Predicting peptide presentation by major histocompatibility complex class I: an improved machine learning approach to the immunopeptidome," *BMC Bioinformatics,* vol. 20, no. 1, pp. 7, Jan 5, 2019.

[24] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[25] V. Anagnostou, N. Niknafs, K. Marrone, D. C. Bruhm, J. R. White, J. Naidoo, K. Hummelink, K. Monkhorst, F. Lalezari, M. Lanis, S. Rosner, J. E. Reuss, K. N. Smith, V. Adleff, K. Rodgers, Z. Belcaid, L. Rhymee, B. Levy, J. Feliciano, C. L. Hann, D. S. Ettinger, C. Georgiades, F. Verde, P. Illei, Q. K. Li, A. S. Baras, E. Gabrielson, M. V. Brock, R. Karchin, D. M. Pardoll, S. B. Baylin, J. R. Brahmer, R. B. Scharpf, P. M. Forde, and V. E. Velculescu, "Multimodal genomic features predict outcome of immune checkpoint blockade in non-small-cell lung cancer," *Nature Cancer,* vol. 1, no. 1, pp. 99-111, 2020/01/01, 2020.

[26] M. R. Parkhurst, P. F. Robbins, E. Tran, T. D. Prickett, J. J. Gartner, L. Jia, G. Ivey, Y. F. Li, M. El-Gamil, A. Lalani, J. S. Crystal, A. Sachs, E. Groh, S. Ray, L. T. Ngo, S. Kivitz,

A. Pasetto, R. Yossef, F. J. Lowery, S. L. Goff, W. Lo, G. Cafri, D. C. Deniger, P. Malekzadeh, M. Ahmadzadeh, J. R. Wunderlich, R. P. T. Somerville, and S. A. Rosenberg, "Unique Neoantigens Arise from Somatic Mutations in Patients with Gastrointestinal Cancers," *Cancer Discov,* vol. 9, no. 8, pp. 1022-1035, Aug, 2019.

[27] E. Tran, M. Ahmadzadeh, Y. C. Lu, A. Gros, S. Turcotte, P. F. Robbins, J. J. Gartner, Z. Zheng, Y. F. Li, S. Ray, J. R. Wunderlich, R. P. Somerville, and S. A. Rosenberg, "Immunogenicity of somatic mutations in human gastrointestinal cancers," *Science,* vol. 350, no. 6266, pp. 1387-90, Dec 11, 2015.

[28] V. Lennerz, M. Fatho, C. Gentilini, R. A. Frye, A. Lifke, D. Ferel, C. Wolfel, C. Huber, and T. Wolfel, "The response of autologous T cells to a human melanoma is dominated by mutated neoantigens," *Proc Natl Acad Sci U S A,* vol. 102, no. 44, pp. 16013-8, Nov 1, 2005.

[29] J. J. Havel, D. Chowell, and T. A. Chan, "The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy," *Nat Rev Cancer,* vol. 19, no. 3, pp. 133-150, Mar, 2019.

[30] T. A. Chan, M. Yarchoan, E. Jaffee, C. Swanton, S. A. Quezada, A. Stenzinger, and S. Peters, "Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic," *Ann Oncol,* vol. 30, no. 1, pp. 44-56, Jan 1, 2019.

[31] X. M. Shao, R. Bhattacharya, J. Huang, I. K. A. Sivakumar, C. Tokheim, L. Zheng, D. Hirsch, B. Kaminow, A. Omdahl, M. Bonsack, A. B. Riemer, V. E. Velculescu, V. Anagnostou, K. A. Pagel, and R. Karchin, "High-Throughput Prediction of MHC Class I and II Neoantigens with MHCnuggets," *Cancer Immunol Res,* vol. 8, no. 3, pp. 396-408, Mar, 2020.

[32] E. Ghorani, R. Rosenthal, N. McGranahan, J. L. Reading, M. Lynch, K. S. Peggs, C. Swanton, and S. A. Quezada, "Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma," *Ann Oncol,* vol. 29, no. 1, pp. 271-279, Jan 1, 2018.

[33] E. M. Van Allen, D. Miao, B. Schilling, S. A. Shukla, C. Blank, L. Zimmer, A. Sucker, U. Hillen, M. H. G. Foppen, S. M. Goldinger, J. Utikal, J. C. Hassel, B. Weide, K. C. Kaehler, C. Loquai, P. Mohr, R. Gutzmer, R. Dummer, S. Gabriel, C. J. Wu, D. Schadendorf, and L. A. Garraway, "Genomic correlates of response to CTLA-4 blockade in metastatic melanoma," *Science,* vol. 350, no. 6257, pp. 207-211, Oct 9, 2015.

[34] D. Miao, C. A. Margolis, N. I. Vokes, D. Liu, A. Taylor-Weiner, S. M. Wankowicz, D. Adeegbe, D. Keliher, B. Schilling, A. Tracy, M. Manos, N. G. Chau, G. J. Hanna, P. Polak, S. J. Rodig, S. Signoretti, L. M. Sholl, J. A. Engelman, G. Getz, P. A. Janne, R. I. Haddad, T. K. Choueiri, D. A. Barbie, R. Haq, M. M. Awad, D. Schadendorf, F. S. Hodi, J. Bellmunt, K. K. Wong, P. Hammerman, and E. M. Van Allen, "Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors," *Nat Genet,* vol. 50, no. 9, pp. 1271-1281, Sep, 2018.

[35] S. Kreiter, M. Vormehr, N. van de Roemer, M. Diken, M. Lower, J. Diekmann, S. Boegel, B. Schrors, F. Vascotto, J. C. Castle, A. D. Tadmor, S. P. Schoenberger, C. Huber, O. Tureci, and U. Sahin, "Mutant MHC class II epitopes drive therapeutic immune responses to cancer," *Nature,* vol. 520, no. 7549, pp. 692-6, Apr 30, 2015.

[36] E. Alspach, D. M. Lussier, A. P. Miceli, I. Kizhvatov, M. DuPage, A. M. Luoma, W. Meng, C. F. Lichti, E. Esaulova, A. N. Vomund, D. Runci, J. P. Ward, M. M. Gubin, R. F. V. Medrano, C. D. Arthur, J. M. White, K. C. F. Sheehan, A. Chen, K. W.

Wucherpfennig, T. Jacks, E. R. Unanue, M. N. Artyomov, and R. D. Schreiber, "MHC-II neoantigens shape tumour immunity and response to immunotherapy," *Nature,* vol. 574, no. 7780, pp. 696-701, Oct, 2019.

[37] S. Paul, D. Weiskopf, M. A. Angelo, J. Sidney, B. Peters, and A. Sette, "HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity," *Journal of immunology (Baltimore, Md. : 1950),* vol. 191, pp. 5831-5839, 2013.

[38] M. Nielsen, and M. Andreatta, "NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets," *Genome Med,* vol. 8, no. 1, pp. 33, Mar 30, 2016.

[39] F. Conforti, L. Pala, V. Bagnardi, T. De Pas, M. Martinetti, G. Viale, R. D. Gelber, and A. Goldhirsch, "Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis," *Lancet Oncol,* vol. 19, no. 6, pp. 737-746, Jun, 2018.

[40] A. Grassadonia, I. Sperduti, P. Vici, L. Iezzi, D. Brocco, T. Gamucci, L. Pizzuti, M. Maugeri-Sacca, P. Marchetti, G. Cognetti, M. De Tursi, C. Natoli, M. Barba, and N. Tinari, "Effect of Gender on the Outcome of Patients Receiving Immune Checkpoint Inhibitors for Advanced Cancer: A Systematic Review and Meta-Analysis of Phase III Randomized Clinical Trials," *J Clin Med,* vol. 7, no. 12, Dec 12, 2018.

[41] Y. Wu, Q. Ju, K. Jia, J. Yu, H. Shi, H. Wu, and M. Jiang, "Correlation between sex and efficacy of immune checkpoint inhibitors (PD-1 and CTLA-4 inhibitors)," *Int J Cancer,* vol. 143, no. 1, pp. 45-51, Jul 1, 2018.

[42] J. A. Pinto, C. S. Vallejos, L. E. Raez, L. A. Mas, R. Ruiz, J. S. Torres-Roman, Z. Morante, J. M. Araujo, H. L. Gomez, A. Aguilar, D. Bretel, C. J. Flores, and C. Rolfo, "Gender and outcomes in non-small cell lung cancer: an old prognostic variable comes back for targeted therapy and immunotherapy?," *ESMO Open,* vol. 3, no. 3, pp. e000344, 2018.

[43] C. Wang, W. Qiao, Y. Jiang, M. Zhu, J. Shao, P. Ren, D. Liu, and W. Li, "Effect of sex on the efficacy of patients receiving immune checkpoint inhibitors in advanced non-small cell lung cancer," *Cancer Med,* vol. 8, no. 8, pp. 4023-4031, Jul, 2019.

[44] H. El-Osta, and S. Jafri, "Predictors for clinical benefit of immune checkpoint inhibitors in advanced non-small-cell lung cancer: a meta-analysis," *Immunotherapy,* vol. 11, no. 3, pp. 189-199, Feb, 2019.

[45] M. Reck, D. Rodriguez-Abreu, A. G. Robinson, R. Hui, T. Csoszi, A. Fulop, M. Gottfried, N. Peled, A. Tafreshi, S. Cuffe, M. O'Brien, S. Rao, K. Hotta, M. A. Leiby, G. M. Lubiniecki, Y. Shentu, R. Rangwala, J. R. Brahmer, and K.-. Investigators, "Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer," *N Engl J Med,* vol. 375, no. 19, pp. 1823-1833, Nov 10, 2016.

[46] J. Brahmer, K. L. Reckamp, P. Baas, L. Crino, W. E. Eberhardt, E. Poddubskaya, S. Antonia, A. Pluzanski, E. E. Vokes, E. Holgado, D. Waterhouse, N. Ready, J. Gainor, O. Aren Frontera, L. Havel, M. Steins, M. C. Garassino, J. G. Aerts, M. Domine, L. Paz-Ares, M. Reck, C. Baudelet, C. T. Harbison, B. Lestini, and D. R. Spigel, "Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer," *N Engl J Med,* vol. 373, no. 2, pp. 123-35, Jul 9, 2015.

[47] H. Borghaei, L. Paz-Ares, L. Horn, D. R. Spigel, M. Steins, N. E. Ready, L. Q. Chow, E. E. Vokes, E. Felip, E. Holgado, F. Barlesi, M. Kohlhaufl, O. Arrieta, M. A. Burgio, J. Fayette, H. Lena, E. Poddubskaya, D. E. Gerber, S. N. Gettinger, C. M. Rudin, N. Rizvi,

L. Crino, G. R. Blumenschein, Jr., S. J. Antonia, C. Dorange, C. T. Harbison, F. Graf Finckenstein, and J. R. Brahmer, "Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer," *N Engl J Med,* vol. 373, no. 17, pp. 1627-39, Oct 22, 2015.

[48] F. Conforti, L. Pala, V. Bagnardi, G. Viale, T. De Pas, E. Pagan, E. Pennacchioli, E. Cocorocchio, P. F. Ferrucci, F. De Marinis, R. D. Gelber, and A. Goldhirsch, "Sex-Based Heterogeneity in Response to Lung Cancer Immunotherapy: A Systematic Review and Meta-Analysis," *J Natl Cancer Inst,* vol. 111, no. 8, pp. 772-781, Aug 1, 2019.

[49] N. Duma, A. Abdel-Ghani, S. Yadav, K. P. Hoversten, C. T. Reed, A. N. Sitek, E. A. L. Enninga, J. Paludo, J. V. Aguilera, K. Leventakos, Y. Lou, L. A. Kottschade, H. Dong, A. S. Mansfield, R. Manochakian, A. A. Adjei, and R. S. Dronca, "Sex Differences in Tolerability to Anti-Programmed Cell Death Protein 1 Therapy in Patients with Metastatic Melanoma and Non-Small Cell Lung Cancer: Are We All Equal?," *Oncologist,* vol. 24, no. 11, pp. e1148-e1155, Nov, 2019.

[50] J. M. Unger, R. Vaidya, K. S. Albain, M. LeBlanc, L. M. Minasian, C. C. Gotay, N. L. Henry, M. J. Fisch, S. M. Lee, C. D. Blanke, and D. L. Hershman, "Sex Differences in Risk of Severe Adverse Events in Patients Receiving Immunotherapy, Targeted Therapy, or Chemotherapy in Cancer Clinical Trials," *J Clin Oncol,* vol. 40, no. 13, pp. 1474-1486, May 1, 2022.

[51] C. C. Whitacre, "Sex differences in autoimmune disease," *Nat Immunol,* vol. 2, no. 9, pp. 777-80, Sep, 2001.

[52] S. L. Klein, and R. Morgan, "The impact of sex and gender on immunotherapy outcomes," *Biol Sex Differ,* vol. 11, no. 1, pp. 24, May 4, 2020.

[53] I. Capone, P. Marchetti, P. A. Ascierto, W. Malorni, and L. Gabriele, "Sexual Dimorphism of Immune Responses: A New Perspective in Cancer Immunotherapy," *Front Immunol,* vol. 9, pp. 552, 2018.

[54] F. Conforti, L. Pala, E. Pagan, V. Bagnardi, T. De Pas, P. Queirolo, E. Pennacchioli, C. Catania, E. Cocorocchio, P. F. Ferrucci, M. Saponara, G. Orsolini, P. Zagami, E. Nicolo, F. De Marinis, G. Tortora, E. Bria, S. Minucci, H. Joffe, P. Veronesi, J. Wargo, R. Rosenthal, C. Swanton, A. Mantovani, R. D. Gelber, G. Viale, A. Goldhirsch, and G. Giaccone, "Sex-Based Dimorphism of Anticancer Immune Response and Molecular Mechanisms of Immune Evasion," *Clin Cancer Res,* vol. 27, no. 15, pp. 4311-4324, Aug 1, 2021.

[55] M. J. Polanczyk, C. Hopke, A. A. Vandenbark, and H. Offner, "Treg suppressive activity involves estrogen-dependent expression of programmed death-1 (PD-1)," *Int Immunol,* vol. 19, no. 3, pp. 337-43, Mar, 2007.

[56] N. J. Rothenberger, A. Somasundaram, and L. P. Stabile, "The Role of the Estrogen Pathway in the Tumor Microenvironment," *Int J Mol Sci,* vol. 19, no. 2, Feb 19, 2018.

[57] D. Chowell, L. G. T. Morris, C. M. Grigg, J. K. Weber, R. M. Samstein, V. Makarov, F. Kuo, S. M. Kendall, D. Requena, N. Riaz, B. Greenbaum, J. Carroll, E. Garon, D. M. Hyman, A. Zehir, D. Solit, M. Berger, R. Zhou, N. A. Rizvi, and T. A. Chan, "Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy," *Science,* vol. 359, no. 6375, pp. 582-587, Feb 2, 2018.

[58] T. Trolle, C. P. McMurtrey, J. Sidney, W. Bardet, S. C. Osborn, T. Kaever, A. Sette, W. H. Hildebrand, M. Nielsen, and B. Peters, "The Length Distribution of Class I-Restricted

T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference," *J Immunol,* vol. 196, no. 4, pp. 1480-7, Feb 15, 2016.

[59] D. P. Kingma, and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR,* vol. abs/1412.6980, 2014.

[60] Y. Gal, "A Theoretically Grounded Application of Dropout in Recurrent Neural Networks," *arXiv:1512.05287*, 2015.

[61] J. G. Abelin, D. B. Keskin, S. Sarkizova, C. R. Hartigan, W. Zhang, J. Sidney, J. Stevens, W. Lane, G. L. Zhang, T. M. Eisenhaure, K. R. Clauser, N. Hacohen, M. S. Rooney, S. A. Carr, and C. J. Wu, "Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction," *Immunity,* vol. 46, no. 2, pp. 315-326, Feb 21, 2017.

[62] M. P. Lefranc, V. Giudicelli, P. Duroux, J. Jabado-Michaloud, G. Folch, S. Aouinti, E. Carillon, H. Duvergey, A. Houles, T. Paysan-Lafosse, S. Hadi-Saljoqi, S. Sasorith, G. Lefranc, and S. Kossida, "IMGT(R), the international ImMunoGeneTics information system(R) 25 years on," *Nucleic Acids Res,* vol. 43, no. Database issue, pp. D413-22, Jan, 2015.

[63] J. Sidney, E. Assarsson, C. Moore, S. Ngo, C. Pinilla, A. Sette, and B. Peters, "Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries," *Immunome Res,* vol. 4, pp. 2, Jan 25, 2008.

[64] D. Gfeller, P. Guillaume, J. Michaux, H. S. Pak, R. T. Daniel, J. Racle, G. Coukos, and M. Bassani-Sternberg, "The Length Distribution and Multiple Specificity of Naturally Presented HLA-I Ligands," *J Immunol,* vol. 201, no. 12, pp. 3705-3716, Dec 15, 2018.

[65] J. Greenbaum, J. Sidney, J. Chung, C. Brander, B. Peters, and A. Sette, "Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes," *Immunogenetics,* vol. 63, no. 6, pp. 325-35, Jun, 2011.

[66] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, and M. Devin, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[67] F. Chollet, and others, "Keras," GitHub, 2015.

[68] J. Hundal, B. M. Carreno, A. A. Petti, G. P. Linette, O. L. Griffith, E. R. Mardis, and M. Griffith, "pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens," *Genome Med,* vol. 8, no. 1, pp. 11, Jan 29, 2016.

[69] J. Hundal, S. Kiwala, J. McMichael, C. A. Miller, H. Xia, A. T. Wollam, C. J. Liu, S. Zhao, Y. Y. Feng, A. P. Graubert, A. Z. Wollam, J. Neichin, M. Neveau, J. Walker, W. E. Gillanders, E. R. Mardis, O. L. Griffith, and M. Griffith, "pVACtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens," *Cancer Immunol Res,* vol. 8, no. 3, pp. 409-420, Mar, 2020.

[70] M. A. Wood, M. Paralkar, M. P. Paralkar, A. Nguyen, A. J. Struck, K. Ellrott, A. Margolin, A. Nellore, and R. F. Thompson, "Population-level distribution and putative immunogenicity of cancer neoepitopes," *BMC Cancer,* vol. 18, no. 1, pp. 414, Apr 13, 2018.

[71] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning." pp. 270-279.

[72] M. Nielsen, C. Lundegaard, P. Worning, S. L. Lauemoller, K. Lamberth, S. Buus, S. Brunak, and O. Lund, "Reliable prediction of T-cell epitopes using neural networks with novel sequence representations," *Protein Sci,* vol. 12, no. 5, pp. 1007-17, May, 2003.

[73] R. Vita, S. Mahajan, J. A. Overton, S. K. Dhanda, S. Martini, J. R. Cantrell, D. K. Wheeler, A. Sette, and B. Peters, "The Immune Epitope Database (IEDB): 2018 update," *Nucleic Acids Res,* vol. 47, no. D1, pp. D339-D343, Jan 8, 2019.

[74] I. Hoof, B. Peters, J. Sidney, L. E. Pedersen, A. Sette, O. Lund, S. Buus, and M. Nielsen, "NetMHCpan, a method for MHC class I binding prediction beyond humans," *Immunogenetics,* vol. 61, no. 1, pp. 1-13, Jan, 2009.

[75] K. K. Jensen, M. Andreatta, P. Marcatili, S. Buus, J. A. Greenbaum, Z. Yan, A. Sette, B. Peters, and M. Nielsen, "Improved methods for predicting peptide binding affinity to MHC class II molecules," *Immunology,* vol. 154, no. 3, pp. 394-406, Jul, 2018.

[76] R. Vita, J. A. Overton, J. A. Greenbaum, J. Ponomarenko, J. D. Clark, J. R. Cantrell, and others, "The immune epitope database (IEDB) 3.0," *Nucleic Acids Research,* vol. 43, no. D1, pp. D405-D405, 2015.

[77] M. H. Bailey, C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, A. Colaprico, M. C. Wendl, J. Kim, B. Reardon, P. K. Ng, K. J. Jeong, S. Cao, Z. Wang, J. Gao, Q. Gao, F. Wang, E. M. Liu, L. Mularoni, C. Rubio-Perez, N. Nagarajan, I. Cortes-Ciriano, D. C. Zhou, W. W. Liang, J. M. Hess, V. D. Yellapantula, D. Tamborero, A. Gonzalez-Perez, C. Suphavilai, J. Y. Ko, E. Khurana, P. J. Park, E. M. Van Allen, H. Liang, M. C. W. Group, N. Cancer Genome Atlas Research, M. S. Lawrence, A. Godzik, N. Lopez-Bigas, J. Stuart, D. Wheeler, G. Getz, K. Chen, A. J. Lazar, G. B. Mills, R. Karchin, and L. Ding, "Comprehensive Characterization of Cancer Driver Genes and Mutations," *Cell,* vol. 173, no. 2, pp. 371-385 e18, Apr 5, 2018.

[78] K. Ellrott, M. H. Bailey, G. Saksena, K. R. Covington, C. Kandoth, C. Stewart, J. Hess, S. Ma, K. E. Chiotti, M. McLellan, H. J. Sofia, C. Hutter, G. Getz, D. Wheeler, L. Ding, M. C. W. Group, and N. Cancer Genome Atlas Research, "Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines," *Cell Syst,* vol. 6, no. 3, pp. 271-281 e7, Mar 28, 2018.

[79] V. Thorsson, D. L. Gibbs, S. D. Brown, D. Wolf, D. S. Bortone, T. H. Ou Yang, E. Porta-Pardo, G. F. Gao, C. L. Plaisier, J. A. Eddy, E. Ziv, A. C. Culhane, E. O. Paull, I. K. A. Sivakumar, A. J. Gentles, R. Malhotra, F. Farshidfar, A. Colaprico, J. S. Parker, L. E. Mose, N. S. Vo, J. Liu, Y. Liu, J. Rader, V. Dhankani, S. M. Reynolds, R. Bowlby, A. Califano, A. D. Cherniack, D. Anastassiou, D. Bedognetti, A. Rao, K. Chen, A. Krasnitz, H. Hu, T. M. Malta, H. Noushmehr, C. S. Pedamallu, S. Bullman, A. I. Ojesina, A. Lamb, W. Zhou, H. Shen, T. K. Choueiri, J. N. Weinstein, J. Guinney, J. Saltz, R. A. Holt, C. E. Rabkin, N. Cancer Genome Atlas Research, A. J. Lazar, J. S. Serody, E. G. Demicco, M. L. Disis, B. G. Vincent, and L. Shmulevich, "The Immune Landscape of Cancer," *Immunity,* vol. 48, no. 4, pp. 812-830 e14, Apr 17, 2018.

[80] B. Li, and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome," *BMC Bioinformatics,* vol. 12, pp. 323, Aug 4, 2011.

[81] A. Szolek, B. Schubert, C. Mohr, M. Sturm, M. Feldhahn, and O. Kohlbacher, "OptiType: precision HLA typing from next-generation sequencing data," *Bioinformatics,* vol. 30, no. 23, pp. 3310-6, Dec 1, 2014.

[82] C. UniProt, "Activities at the Universal Protein Resource (UniProt)," *Nucleic Acids Res,* vol. 42, no. Database issue, pp. D191-8, Jan, 2014.

[83] B. Karakas, K. E. Bachman, and B. H. Park, "Mutation of the PIK3CA oncogene in human cancers," *Br J Cancer,* vol. 94, no. 4, pp. 455-9, Feb 27, 2006.

[84] D. C. Tomlinson, C. D. Hurst, and M. A. Knowles, "Knockdown by shRNA identifies S249C mutant FGFR3 as a potential therapeutic target in bladder cancer," *Oncogene,* vol. 26, no. 40, pp. 5889-99, Aug 30, 2007.

[85] P. Malekzadeh, A. Pasetto, P. F. Robbins, M. R. Parkhurst, B. C. Paria, L. Jia, J. J. Gartner, V. Hill, Z. Yu, N. P. Restifo, A. Sachs, E. Tran, W. Lo, R. P. Somerville, S. A. Rosenberg, and D. C. Deniger, "Neoantigen screening identifies broad TP53 mutant immunogenicity in patients with epithelial cancers," *J Clin Invest,* vol. 129, no. 3, pp. 1109-1114, Mar 1, 2019.

[86] C. Tokheim, and R. Karchin, "CHASMplus reveals the scope of somatic missense mutations driving human cancers," *Cell Systems,* vol. in press, 2019.

[87] D. Chakravarty, J. Gao, S. M. Phillips, R. Kundra, H. Zhang, J. Wang, J. E. Rudolph, R. Yaeger, T. Soumerai, M. H. Nissan, M. T. Chang, S. Chandarlapaty, T. A. Traina, P. K. Paik, A. L. Ho, F. M. Hantash, A. Grupe, S. S. Baxi, M. K. Callahan, A. Snyder, P. Chi, D. Danila, M. Gounder, J. J. Harding, M. D. Hellmann, G. Iyer, Y. Janjigian, T. Kaley, D. A. Levine, M. Lowery, A. Omuro, M. A. Postow, D. Rathkopf, A. N. Shoushtari, N. Shukla, M. Voss, E. Paraiso, A. Zehir, M. F. Berger, B. S. Taylor, L. B. Saltz, G. J. Riely, M. Ladanyi, D. M. Hyman, J. Baselga, P. Sabbatini, D. B. Solit, and N. Schultz, "OncoKB: A Precision Oncology Knowledge Base," *JCO Precis Oncol,* vol. 2017, Jul, 2017.

[88] R. Marty, S. Kaabinejadian, D. Rossell, M. J. Slifker, J. van de Haar, H. B. Engin, N. de Prisco, T. Ideker, W. H. Hildebrand, J. Font-Burgada, and H. Carter, "MHC-I Genotype Restricts the Oncogenic Mutational Landscape," *Cell,* vol. 171, no. 6, pp. 1272-1283 e15, Nov 30, 2017.

[89] C. Tokheim, R. Bhattacharya, N. Niknafs, D. M. Gygax, R. Kim, M. Ryan, D. L. Masica, and R. Karchin, "Exome-Scale Discovery of Hotspot Mutation Regions in Human Cancer Using 3D Protein Structure," *Cancer Res,* vol. 76, no. 13, pp. 3719-31, Jul 1, 2016.

[90] Y. Benjamini, and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 57, no. 1, pp. 289-300, 1995.

[91] J. Qing, X. Du, Y. Chen, P. Chan, H. Li, P. Wu, S. Marsters, S. Stawicki, J. Tien, K. Totpal, S. Ross, S. Stinson, D. Dornan, D. French, Q. R. Wang, J. P. Stephan, Y. Wu, C. Wiesmann, and A. Ashkenazi, "Antibody-based targeting of FGFR3 in bladder carcinoma and t(4;14)-positive multiple myeloma in mice," *J Clin Invest,* vol. 119, no. 5, pp. 1216-29, May, 2009.

[92] A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh, "Robust enumeration of cell subsets from tissue expression profiles," *Nat Methods,* vol. 12, no. 5, pp. 453-7, May, 2015.

[93] X. M. Shao, J. Huang, N. Niknafs, A. Balan, C. Cherry, J. White, V. E. Velculescu, V. Anagnostou, and R. Karchin, "HLA class II immunogenic mutation burden predicts response to immune checkpoint blockade," *Ann Oncol,* vol. 33, no. 7, pp. 728-738, Jul, 2022.

[94] M. Bassani-Sternberg, E. Braunlein, R. Klar, T. Engleitner, P. Sinitcyn, S. Audehm, M. Straub, J. Weber, J. Slotta-Huspenina, K. Specht, M. E. Martignoni, A. Werner, R. Hein, H. B. D, C. Peschel, R. Rad, J. Cox, M. Mann, and A. M. Krackhardt, "Direct

identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry," *Nat Commun,* vol. 7, pp. 13404, Nov 21, 2016.

[95] M. W. Loffler, D. J. Kowalewski, L. Backert, J. Bernhardt, P. Adam, H. Schuster, F. Dengler, D. Backes, H. G. Kopp, S. Beckert, S. Wagner, I. Konigsrainer, O. Kohlbacher, L. Kanz, A. Konigsrainer, H. G. Rammensee, S. Stevanovic, and S. P. Haen, "Mapping the HLA Ligandome of Colorectal Cancer Reveals an Imprint of Malignant Cell Transformation," *Cancer Res,* vol. 78, no. 16, pp. 4627-4641, Aug 15, 2018.

[96] G. P. Mommen, F. Marino, H. D. Meiring, M. C. Poelen, J. A. van Gaans-van den Brink, S. Mohammed, A. J. Heck, and C. A. van Els, "Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity," *Mol Cell Proteomics,* vol. 15, no. 4, pp. 1412-23, Apr, 2016.

[97] S. Mutschlechner, M. Egger, P. Briza, M. Wallner, P. Lackner, A. Karle, A. B. Vogt, G. F. Fischer, B. Bohle, and F. Ferreira, "Naturally processed T cell-activating peptides of the major birch pollen allergen," *J Allergy Clin Immunol,* vol. 125, no. 3, pp. 711-8, 718 e1-718 e2, Mar, 2010.

[98] H. Schuster, J. K. Peper, H. C. Bosmuller, K. Rohle, L. Backert, T. Bilich, B. Ney, M. W. Loffler, D. J. Kowalewski, N. Trautwein, A. Rabsteyn, T. Engler, S. Braun, S. P. Haen, J. S. Walz, B. Schmid-Horch, S. Y. Brucker, D. Wallwiener, O. Kohlbacher, F. Fend, H. G. Rammensee, S. Stevanovic, A. Staebler, and P. Wagner, "The immunopeptidomic landscape of ovarian carcinomas," *Proc Natl Acad Sci U S A,* vol. 114, no. 46, pp. E9942-E9951, Nov 14, 2017.

[99] I. Strug, J. M. Calvo-Calle, K. M. Green, J. Cruz, F. A. Ennis, J. E. Evans, and L. J. Stern, "Vaccinia peptides eluted from HLA-DR1 isolated from virus-infected cells are recognized by CD4+ T cells from a vaccinated donor," *J Proteome Res,* vol. 7, no. 7, pp. 2703-11, Jul, 2008.

[100] N. A. Rizvi, M. D. Hellmann, A. Snyder, P. Kvistborg, V. Makarov, J. J. Havel, W. Lee, J. Yuan, P. Wong, T. S. Ho, M. L. Miller, N. Rekhtman, A. L. Moreira, F. Ibrahim, C. Bruggeman, B. Gasmi, R. Zappasodi, Y. Maeda, C. Sander, E. B. Garon, T. Merghoub, J. D. Wolchok, T. N. Schumacher, and T. A. Chan, "Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer," *Science,* vol. 348, no. 6230, pp. 124-8, Apr 3, 2015.

[101] V. Anagnostou, D. C. Bruhm, N. Niknafs, J. R. White, X. M. Shao, J. W. Sidhom, J. Stein, H. L. Tsai, H. Wang, Z. Belcaid, J. Murray, A. Balan, L. Ferreira, P. Ross-Macdonald, M. Wind-Rotolo, A. S. Baras, J. Taube, R. Karchin, R. B. Scharpf, C. Grasso, A. Ribas, D. M. Pardoll, S. L. Topalian, and V. E. Velculescu, "Integrative Tumor and Immune Cell Multi-omic Analyses Predict Response to Immune Checkpoint Blockade in Melanoma," *Cell Rep Med,* vol. 1, no. 8, pp. 100139, Nov 17, 2020.

[102] A. Snyder, V. Makarov, T. Merghoub, J. Yuan, J. M. Zaretsky, A. Desrichard, L. A. Walsh, M. A. Postow, P. Wong, T. S. Ho, T. J. Hollmann, C. Bruggeman, K. Kannan, Y. Li, C. Elipenahli, C. Liu, C. T. Harbison, L. Wang, A. Ribas, J. D. Wolchok, and T. A. Chan, "Genetic basis for clinical response to CTLA-4 blockade in melanoma," *N Engl J Med,* vol. 371, no. 23, pp. 2189-2199, Dec 4, 2014.

[103] C. T. Saunders, W. S. Wong, S. Swamy, J. Becq, L. J. Murray, and R. K. Cheetham, "Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs," *Bioinformatics,* vol. 28, no. 14, pp. 1811-7, Jul 15, 2012.

[104] W. J. Kent, “BLAT--the BLAST-like alignment tool,” *Genome Res,* vol. 12, no. 4, pp. 656-64, Apr, 2002.
[105] openvax. "VarCode," https://github.com/openvax/varcode.
[106] J. G. Abelin, D. Harjanto, M. Malloy, P. Suri, T. Colson, S. P. Goulding, A. L. Creech, L. R. Serrano, G. Nasir, Y. Nasrullah, C. D. McGann, D. Velez, Y. S. Ting, A. Poran, D. A. Rothenberg, S. Chhangawala, A. Rubinsteyn, J. Hammerbacher, R. B. Gaynor, E. F. Fritsch, J. Greshock, R. C. Oslund, D. Barthelme, T. A. Addona, C. M. Arieta, and M. S. Rooney, “Defining HLA-II Ligand Processing and Binding Rules with Mass Spectrometry Enhances Cancer Epitope Prediction,” *Immunity,* vol. 51, no. 4, pp. 766-779 e17, Oct 15, 2019.
[107] C. Chong, F. Marino, H. Pak, J. Racle, R. T. Daniel, M. Muller, D. Gfeller, G. Coukos, and M. Bassani-Sternberg, “High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferongamma-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome,” *Mol Cell Proteomics,* vol. 17, no. 3, pp. 533-548, Mar, 2018.
[108] H. Cao, J. Wu, Y. Wang, H. Jiang, T. Zhang, X. Liu, Y. Xu, D. Liang, P. Gao, Y. Sun, B. Gifford, M. D'Ascenzo, X. Liu, L. C. Tellier, F. Yang, X. Tong, D. Chen, J. Zheng, W. Li, T. Richmond, X. Xu, J. Wang, and Y. Li, “An integrated tool to study MHC region: accurate SNV detection and HLA genes typing in human MHC region using targeted high-throughput sequencing,” *PLoS One,* vol. 8, no. 7, pp. e69388, 2013.
[109] C. Xie, Z. X. Yeo, M. Wong, J. Piper, T. Long, E. F. Kirkness, W. H. Biggs, K. Bloom, S. Spellman, C. Vierra-Green, C. Brady, R. H. Scheuermann, A. Telenti, S. Howard, S. Brewerton, Y. Turpaz, and J. C. Venter, “Fast and accurate HLA typing from short-read next-generation sequence data with xHLA,” *Proc Natl Acad Sci U S A,* vol. 114, no. 30, pp. 8059-8064, Jul 25, 2017.
[110] K. Y. Jen, and V. G. Cheung, “Identification of novel p53 target genes in ionizing radiation response,” *Cancer Res,* vol. 65, no. 17, pp. 7666-73, Sep 1, 2005.
[111] B. Goodrich, J. Gabry, I. Ali, and S. Brilleman. "rstanarm: Bayesian applied regression modeling via Stan," https://mc-stan.org/rstanarm.
[112] N. Niknafs, V. Beleva-Guthrie, D. Q. Naiman, and R. Karchin, “SubClonal Hierarchy Inference from Somatic Mutations: Automatic Reconstruction of Cancer Evolutionary Trees from Multi-region Next Generation Sequencing,” *PLoS Comput Biol,* vol. 11, no. 10, pp. e1004416, Oct, 2015.
[113] N. McGranahan, A. J. Furness, R. Rosenthal, S. Ramskov, R. Lyngaa, S. K. Saini, M. Jamal-Hanjani, G. A. Wilson, N. J. Birkbak, C. T. Hiley, T. B. Watkins, S. Shafi, N. Murugaesu, R. Mitter, A. U. Akarca, J. Linares, T. Marafioti, J. Y. Henry, E. M. Van Allen, D. Miao, B. Schilling, D. Schadendorf, L. A. Garraway, V. Makarov, N. A. Rizvi, A. Snyder, M. D. Hellmann, T. Merghoub, J. D. Wolchok, S. A. Shukla, C. J. Wu, K. S. Peggs, T. A. Chan, S. R. Hadrup, S. A. Quezada, and C. Swanton, “Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade,” *Science,* vol. 351, no. 6280, pp. 1463-9, Mar 25, 2016.
[114] M. Manczinger, and L. Kemeny, “Peptide presentation by HLA-DQ molecules is associated with the development of immune tolerance,” *PeerJ,* vol. 6, pp. e5118, 2018.
[115] A. A. Davis, and V. G. Patel, “The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors,” *J Immunother Cancer,* vol. 7, no. 1, pp. 278, Oct 26, 2019.

[116] J. Norum, and C. Nieder, "Tobacco smoking and cessation and PD-L1 inhibitors in non-small cell lung cancer (NSCLC): a review of the literature," *ESMO Open,* vol. 3, no. 6, pp. e000406, 2018.

[117] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras, "STAR: ultrafast universal RNA-seq aligner," *Bioinformatics,* vol. 29, no. 1, pp. 15-21, Jan 1, 2013.

[118] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol,* vol. 15, no. 12, pp. 550, 2014.

[119] G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, and A. Sergushichev, "Fast gene set enrichment analysis," *bioRxiv*, pp. 060012, 2021.

[120] M. Ayers, J. Lunceford, M. Nebozhyn, E. Murphy, A. Loboda, D. R. Kaufman, A. Albright, J. D. Cheng, S. P. Kang, V. Shankaran, S. A. Piha-Paul, J. Yearley, T. Y. Seiwert, A. Ribas, and T. K. McClanahan, "IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade," *J Clin Invest,* vol. 127, no. 8, pp. 2930-2940, Aug 1, 2017.

[121] R. Cristescu, R. Mogg, M. Ayers, A. Albright, E. Murphy, J. Yearley, X. Sher, X. Q. Liu, H. Lu, M. Nebozhyn, C. Zhang, J. K. Lunceford, A. Joe, J. Cheng, A. L. Webber, N. Ibrahim, E. R. Plimack, P. A. Ott, T. Y. Seiwert, A. Ribas, T. K. McClanahan, J. E. Tomassini, A. Loboda, and D. Kaufman, "Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy," *Science,* vol. 362, no. 6411, Oct 12, 2018.

[122] J. C. Beltra, S. Manne, M. S. Abdel-Hakeem, M. Kurachi, J. R. Giles, Z. Chen, V. Casella, S. F. Ngiow, O. Khan, Y. J. Huang, P. Yan, K. Nzingha, W. Xu, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, L. M. Schuchter, A. C. Huang, and E. J. Wherry, "Developmental Relationships of Four Exhausted CD8(+) T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms," *Immunity,* vol. 52, no. 5, pp. 825-841 e8, May 19, 2020.

[123] A. Calabro, T. Beissbarth, R. Kuner, M. Stojanov, A. Benner, M. Asslaber, F. Ploner, K. Zatloukal, H. Samonigg, A. Poustka, and H. Sultmann, "Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer," *Breast Cancer Res Treat,* vol. 116, no. 1, pp. 69-77, Jul, 2009.

[124] J. Galon, H. K. Angell, D. Bedognetti, and F. M. Marincola, "The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures," *Immunity,* vol. 39, no. 1, pp. 11-26, Jul 25, 2013.

[125] J. Griss, G. Viteri, K. Sidiropoulos, V. Nguyen, A. Fabregat, and H. Hermjakob, "ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis," *Mol Cell Proteomics,* vol. 19, no. 12, pp. 2115-2125, Dec, 2020.

[126] M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, and M. Tanabe, "KEGG: integrating viruses and cellular organisms," *Nucleic Acids Res,* vol. 49, no. D1, pp. D545-D551, Jan 8, 2021.

[127] O. Khan, J. R. Giles, S. McDonald, S. Manne, S. F. Ngiow, K. P. Patel, M. T. Werner, A. C. Huang, K. A. Alexander, J. E. Wu, J. Attanasio, P. Yan, S. M. George, B. Bengsch, R. P. Staupe, G. Donahue, W. Xu, R. K. Amaravadi, X. Xu, G. C. Karakousis, T. C. Mitchell, L. M. Schuchter, J. Kaye, S. L. Berger, and E. J. Wherry, "TOX transcriptionally and epigenetically programs CD8(+) T cell exhaustion," *Nature,* vol. 571, no. 7764, pp. 211-218, Jul, 2019.

[128] A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J. P. Mesirov, and P. Tamayo, "The Molecular Signatures Database (MSigDB) hallmark gene set collection," *Cell Syst,* vol. 1, no. 6, pp. 417-425, Dec 23, 2015.

[129] F. O. Martinez, S. Gordon, M. Locati, and A. Mantovani, "Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression," *J Immunol,* vol. 177, no. 10, pp. 7303-11, Nov 15, 2006.

[130] C. F. Schaefer, K. Anthony, S. Krupa, J. Buchoff, M. Day, T. Hannay, and K. H. Buetow, "PID: the Pathway Interaction Database," *Nucleic Acids Res,* vol. 37, no. Database issue, pp. D674-9, Jan, 2009.

[131] A. Szanto, B. L. Balint, Z. S. Nagy, E. Barta, B. Dezso, A. Pap, L. Szeles, S. Poliska, M. Oros, R. M. Evans, Y. Barak, J. Schwabe, and L. Nagy, "STAT6 transcription factor is a facilitator of the nuclear receptor PPARgamma-regulated gene expression in macrophages and dendritic cells," *Immunity,* vol. 33, no. 5, pp. 699-712, Nov 24, 2010.

[132] I. Tassiulas, X. Hu, H. Ho, Y. Kashyap, P. Paik, Y. Hu, C. A. Lowell, and L. B. Ivashkiv, "Amplification of IFN-alpha-induced STAT1 activation and inflammatory function by Syk and ITAM-containing adaptors," *Nat Immunol,* vol. 5, no. 11, pp. 1181-9, Nov, 2004.

[133] D. M. Wolf, M. E. Lenburg, C. Yau, A. Boudreau, and L. J. van 't Veer, "Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity," *PLoS One,* vol. 9, no. 2, pp. e88309, 2014.

[134] S. C. Scott, X. M. Shao, N. Niknafs, A. Balan, G. Pereira, K. A. Marrone, V. K. Lam, J. C. Murray, J. L. Feliciano, B. P. Levy, D. S. Ettinger, C. L. Hann, J. R. Brahmer, P. M. Forde, R. Karchin, J. Naidoo, and V. Anagnostou, "Sex-specific differences in immunogenomic features of response to immune checkpoint blockade," *Front Oncol,* vol. 12, pp. 945798, 2022.

[135] J. Liu, T. Lichtenberg, K. A. Hoadley, L. M. Poisson, A. J. Lazar, A. D. Cherniack, A. J. Kovatich, C. C. Benz, D. A. Levine, A. V. Lee, L. Omberg, D. M. Wolf, C. D. Shriver, V. Thorsson, N. Cancer Genome Atlas Research, and H. Hu, "An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics," *Cell,* vol. 173, no. 2, pp. 400-416 e11, Apr 5, 2018.

[136] D. E. Wood, J. R. White, A. Georgiadis, B. Van Emburgh, S. Parpart-Li, J. Mitchell, V. Anagnostou, N. Niknafs, R. Karchin, E. Papp, C. McCord, P. LoVerso, D. Riley, L. A. Diaz, Jr., S. Jones, M. Sausen, V. E. Velculescu, and S. V. Angiuoli, "A machine learning approach for somatic mutation discovery," *Sci Transl Med,* vol. 10, no. 457, Sep 5, 2018.

[137] R. Marty Pyke, W. K. Thompson, R. M. Salem, J. Font-Burgada, M. Zanetti, and H. Carter, "Evolutionary Pressure against MHC Class II Binding Cancer Mutations," *Cell,* vol. 175, no. 2, pp. 416-428 e13, Oct 4, 2018.

[138] N. McGranahan, R. Rosenthal, C. T. Hiley, A. J. Rowan, T. B. K. Watkins, G. A. Wilson, N. J. Birkbak, S. Veeriah, P. Van Loo, J. Herrero, C. Swanton, and T. R. Consortium, "Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution," *Cell,* vol. 171, no. 6, pp. 1259-1271 e11, Nov 30, 2017.

[139] E. Talevich, A. H. Shain, T. Botton, and B. C. Bastian, "CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing," *PLoS Comput Biol,* vol. 12, no. 4, pp. e1004873, Apr, 2016.

[140] A. B. Olshen, E. S. Venkatraman, R. Lucito, and M. Wigler, "Circular binary segmentation for the analysis of array-based DNA copy number data," *Biostatistics,* vol. 5, no. 4, pp. 557-72, Oct, 2004.

[141] R. Rosenthal, N. McGranahan, J. Herrero, B. S. Taylor, and C. Swanton, "DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution," *Genome Biol,* vol. 17, pp. 31, Feb 22, 2016.

[142] L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, I. Australian Pancreatic Cancer Genome, I. B. C. Consortium, I. M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, and M. R. Stratton, "Signatures of mutational processes in human cancer," *Nature,* vol. 500, no. 7463, pp. 415-21, Aug 22, 2013.

[143] R. D. Schreiber, L. J. Old, and M. J. Smyth, "Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion," *Science,* vol. 331, no. 6024, pp. 1565-70, Mar 25, 2011.

[144] A. Castro, R. M. Pyke, X. Zhang, W. K. Thompson, C. P. Day, L. B. Alexandrov, M. Zanetti, and H. Carter, "Strength of immune selection in tumors varies with sex and age," *Nat Commun,* vol. 11, no. 1, pp. 4128, Aug 17, 2020.

[145] K. K. W. To, W. Fong, and W. C. S. Cho, "Immunotherapy in Treating EGFR-Mutant Lung Cancer: Current Challenges and New Strategies," *Front Oncol,* vol. 11, pp. 635007, 2021.

[146] K. Seegobin, U. Majeed, N. Wiest, R. Manochakian, Y. Lou, and Y. Zhao, "Immunotherapy in Non-Small Cell Lung Cancer With Actionable Mutations Other Than EGFR," *Front Oncol,* vol. 11, pp. 750657, 2021.

[147] M. Bassani-Sternberg, C. Chong, P. Guillaume, M. Solleder, H. Pak, P. O. Gannon, L. E. Kandalaft, G. Coukos, and D. Gfeller, "Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity," *PLoS Comput Biol,* vol. 13, no. 8, pp. e1005725, Aug, 2017.

[148] A. J. Rech, D. Balli, A. Mantero, H. Ishwaran, K. L. Nathanson, B. Z. Stanger, and R. H. Vonderheide, "Tumor Immunity and Survival as a Function of Alternative Neopeptides in Human Cancer," *Cancer Immunol Res*, Jan 16, 2018.

[149] N. N. Hunder, H. Wallen, J. Cao, D. W. Hendricks, J. Z. Reilly, R. Rodmyre, A. Jungbluth, S. Gnjatic, J. A. Thompson, and C. Yee, "Treatment of metastatic melanoma with autologous CD4+ T cells against NY-ESO-1," *N Engl J Med,* vol. 358, no. 25, pp. 2698-703, Jun 19, 2008.

[150] P. Y. Arnold, N. L. La Gruta, T. Miller, K. M. Vignali, P. S. Adams, D. L. Woodland, and D. A. Vignali, "The majority of immunogenic epitopes generate CD4+ T cells that

are dependent on MHC class II-bound peptide-flanking residues," *J Immunol,* vol. 169, no. 2, pp. 739-49, Jul 15, 2002.

[151] A. Chhibber, L. Huang, H. Zhang, J. Xu, R. Cristescu, X. Liu, D. V. Mehrotra, J. Shen, P. M. Shaw, M. D. Hellmann, and A. Snyder, "Germline HLA landscape does not predict efficacy of pembrolizumab monotherapy across solid tumor types," *Immunity,* vol. 55, no. 1, pp. 56-64 e4, Jan 11, 2022.

[152] T. Horiki, H. Inoko, J. Moriuchi, Y. Ichikawa, and S. Arimori, "Combinations of HLA-DPB1 and HLA-DQB1 alleles determine susceptibility to early-onset myasthenia gravis in Japan," *Autoimmunity,* vol. 19, no. 1, pp. 49-54, 1994.

[153] E. W. Petersdorf, M. Malkki, C. O'HUigin, M. Carrington, T. Gooley, M. D. Haagenson, M. M. Horowitz, S. R. Spellman, T. Wang, and P. Stevenson, "High HLA-DP Expression and Graft-versus-Host Disease," *N Engl J Med,* vol. 373, no. 7, pp. 599-609, Aug 13, 2015.

[154] G. Diaz, B. Canas, J. Vazquez, C. Nombela, and J. Arroyo, "Characterization of natural peptide ligands from HLA-DP2: new insights into HLA-DP peptide-binding motifs," *Immunogenetics,* vol. 56, no. 10, pp. 754-9, Jan, 2005.

[155] S. Dai, G. A. Murphy, F. Crawford, D. G. Mack, M. T. Falta, P. Marrack, J. W. Kappler, and A. P. Fontenot, "Crystal structure of HLA-DP2 and implications for chronic beryllium disease," *Proc Natl Acad Sci U S A,* vol. 107, no. 16, pp. 7425-30, Apr 20, 2010.

[156] K. Ley, "M1 Means Kill; M2 Means Heal," *J Immunol,* vol. 199, no. 7, pp. 2191-2193, Oct 1, 2017.

[157] M. K. MacLeod, J. W. Kappler, and P. Marrack, "Memory CD4 T cells: generation, reactivation and re-assignment," *Immunology,* vol. 130, no. 1, pp. 10-5, May, 2010.

[158] S. Wang, L. A. Cowley, and X. S. Liu, "Sex Differences in Cancer Immunotherapy Efficacy, Biomarkers, and Therapeutic Strategy," *Molecules,* vol. 24, no. 18, Sep 4, 2019.

[159] A. Abed, L. Calapre, J. Lo, S. Correia, S. Bowyer, A. Chopra, M. Watson, M. A. Khattak, M. Millward, and E. S. Gray, "Prognostic value of HLA-I homozygosity in patients with non-small cell lung cancer treated with single agent immunotherapy," *J Immunother Cancer,* vol. 8, no. 2, Nov, 2020.

[160] F. Perea, M. Bernal, A. Sanchez-Palencia, J. Carretero, C. Torres, C. Bayarri, M. Gomez-Morales, F. Garrido, and F. Ruiz-Cabello, "The absence of HLA class I expression in non-small cell lung cancer correlates with the tumor tissue structure and the pattern of T cell infiltration," *Int J Cancer,* vol. 140, no. 4, pp. 888-899, Feb 15, 2017.

[161] F. Mauvais-Jarvis, H. K. Berthold, I. Campesi, J. J. Carrero, S. Dakal, F. Franconi, I. Gouni-Berthold, M. L. Heiman, A. Kautzky-Willer, S. L. Klein, A. Murphy, V. Regitz-Zagrosek, K. Reue, and J. B. Rubin, "Sex- and Gender-Based Pharmacological Response to Drugs," *Pharmacol Rev,* vol. 73, no. 2, pp. 730-762, Apr, 2021.

[162] J. H. Shim, H. S. Kim, H. Cha, S. Kim, T. M. Kim, V. Anagnostou, Y. L. Choi, H. A. Jung, J. M. Sun, J. S. Ahn, M. J. Ahn, K. Park, W. Y. Park, and S. H. Lee, "HLA-corrected tumor mutation burden and homologous recombination deficiency for the prediction of response to PD-(L)1 blockade in advanced non-small-cell lung cancer patients," *Ann Oncol,* vol. 31, no. 7, pp. 902-911, Jul, 2020.

[163] S. Turajlic, K. Litchfield, H. Xu, R. Rosenthal, N. McGranahan, J. L. Reading, Y. N. S. Wong, A. Rowan, N. Kanu, M. Al Bakir, T. Chambers, R. Salgado, P. Savas, S. Loi, N. J. Birkbak, L. Sansregret, M. Gore, J. Larkin, S. A. Quezada, and C. Swanton, "Insertion-

and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis," *Lancet Oncol,* vol. 18, no. 8, pp. 1009-1021, Aug, 2017.

[164] P. M. Forde, J. Spicer, S. Lu, M. Provencio, T. Mitsudomi, M. M. Awad, E. Felip, S. Broderick, J. Brahmer, and S. J. Swanson, "Abstract CT003: Nivolumab (NIVO)+ platinum-doublet chemotherapy (chemo) vs chemo as neoadjuvant treatment (tx) for resectable (IB-IIIA) non-small cell lung cancer (NSCLC) in the phase 3 CheckMate 816 trial," *Cancer Research,* vol. 81, no. 13_Supplement, pp. CT003-CT003, 2021.

[165] E. Felip, N. Altorki, C. Zhou, T. Csoszi, I. Vynnychenko, O. Goloborodko, A. Luft, A. Akopov, A. Martinez-Marti, H. Kenmotsu, Y. M. Chen, A. Chella, S. Sugawara, D. Voong, F. Wu, J. Yi, Y. Deng, M. McCleland, E. Bennett, B. Gitlitz, H. Wakelee, and I. M. Investigators, "Adjuvant atezolizumab after adjuvant chemotherapy in resected stage IB-IIIA non-small-cell lung cancer (IMpower010): a randomised, multicentre, open-label, phase 3 trial," *Lancet,* vol. 398, no. 10308, pp. 1344-1357, Oct 9, 2021.