

# Prediction of lung cancer immunotherapy response via machine learning analysis of immune cell lineage and surface markers

Alex N. Mueller<sup>a</sup>, Samantha Morrissey<sup>b</sup>, Hunter A. Miller<sup>c</sup>, Xiaoling Hu<sup>b</sup>, Rohit Kumar<sup>a,d</sup>,  
Phuong T. Ngo<sup>a,d</sup>, Jun Yan<sup>b,c,d,e,\*</sup> and Hermann B. Frieboes<sup>c,d,f,g,\*</sup>

<sup>a</sup>*School of Medicine, University of Louisville, Louisville, KY, USA*

<sup>b</sup>*Division of Immunotherapy, Department of Surgery, University of Louisville, Louisville, KY, USA*

<sup>c</sup>*Department of Pharmacology and Toxicology, University of Louisville, Louisville, KY, USA*

<sup>d</sup>*UofL Health – Brown Cancer Center, University of Louisville, Louisville, KY, USA*

<sup>e</sup>*Department of Surgery, University of Louisville, Louisville, KY, USA*

<sup>f</sup>*Center for Predictive Medicine, University of Louisville, Louisville, KY, USA*

<sup>g</sup>*Department of Bioengineering, University of Louisville, Louisville, KY, USA*

Received 6 December 2021

Accepted 13 May 2022

## Abstract.

**BACKGROUND:** Although advances have been made in cancer immunotherapy, patient benefits remain elusive. For non-small cell lung cancer (NSCLC), monoclonal antibodies targeting programmed death-1 (PD-1) and programmed death ligand-1 (PD-L1) have shown survival benefit compared to chemotherapy. Personalization of treatment would be facilitated by *a priori* identification of patients likely to benefit.

**OBJECTIVE:** This pilot study applied a suite of machine learning methods to analyze mass cytometry data of immune cell lineage and surface markers from blood samples of a small cohort ( $n = 13$ ) treated with Pembrolizumab, Atezolizumab, Durvalumab, or Nivolumab as monotherapy.

**METHODS:** Four different comparisons were evaluated between data collected at an initial visit (baseline), after 12-weeks of immunotherapy, and from healthy (control) samples: healthy vs patients at baseline, Responders vs Non-Responders at baseline, Healthy vs 12-week Responders, and Responders vs Non-Responders at 12-weeks. The algorithms Random Forest, Partial Least Squares Discriminant Analysis, Multi-Layer Perceptron, and Elastic Net were applied to find features differentiating between these groups and provide for the capability to predict outcomes.

**RESULTS:** Particular combinations and proportions of immune cell lineage and surface markers were sufficient to accurately discriminate between the groups without overfitting the data. In particular, markers associated with the B-cell phenotype were identified as key features.

**CONCLUSIONS:** This study illustrates a comprehensive machine learning analysis of circulating immune cell characteristics of NSCLC patients with the potential to predict response to immunotherapy. Upon further evaluation in a larger cohort, the proposed methodology could help guide personalized treatment selection in clinical practice.

Keywords: lung cancer, immunotherapy, machine learning, immune cells, cell markers

\*Corresponding authors: Jun Yan, Department of Surgery, CTRB 319, University of Louisville, Louisville, KY 40202, USA. Tel.: +502 852 3628; Fax: +502 852 2123; E-mail: jun.yan@louisville.edu.

Hermann B. Frieboes, Department of Bioengineering, Lutz Hall 419, University of Louisville, Louisville, KY 40292, USA. Tel.: +502 852 3302; Fax: +502 852 6806; E-mail: hbfrie01@louisville.edu.

## 1. Introduction

Challenges facing immunotherapy blockade therapy were recently reviewed [1]. In particular for non-small cell lung cancer (NSCLC), monoclonal antibodies Pembrolizumab and Nivolumab targeting programmed death-1 (PD1) and Atezolizumab and Durvalumab targeting programmed death ligand-1 (PD-L1) [2–4], have shown limited but effective objective response rates (ORR) of  $\sim 20\%$  [5–8]. Treatment options could be improved if responding patients could be identified *a priori* to a particular treatment [9]. However, therapeutic indication of approved checkpoint inhibitors remains limited. PD-1 or PD-L1 expression by itself cannot accurately predict patient response [10]. Similarly, PD-1 mRNA expression is not predictive of outcomes while patients are on anti-PD-1 medication [11]. To address these shortcomings, an analysis of effective predictive biomarkers is indicated [12].

Evaluation of immune cell lineage and surface markers may provide a better understanding of the effects of anti-PD-1 and anti-PD-L1 therapy in NSCLC patients. However, these molecular data are often large and complex to analyze. Mathematical models based on statistics and machine learning algorithms applied to intricate datasets may allow for detection of complex associations that may not be identifiable by simpler analyses such as correlations and linear regressions. Different methods have been employed to build such predictive models for application to NSCLC computed tomography (CT) imaging [13], oral cancer [14], and coronary heart disease [15], among others [16–19]. Recently, we applied machine learning to metabolomic profiles of NSCLC patient tissue samples to predict disease control and progressive disease groups as response to first-line therapy [20].

This pilot study applies a suite of machine learning models to analyze immune cell surface and lineage expression data from a small set of NSCLC patients treated with monoclonal antibodies at both initial and 12-week visit timepoints. The primary study goal is to determine which patients will respond to treatment based on their relative immune cell numbers and proportions of cell surface and lineage marker expression. Secondary goals include identifying which features are potentially important in prediction and to help elucidate immune cell response changes post therapy. Longer term, differences between responders and non-responders could lead to evaluation of potential therapeutic targets.

## 2. Materials and methods

### 2.1. Sample collection

All specimens were collected following approved Internal Review Board protocols at University of Louisville Hospital (IRB 15.0273) from patients with known NSCLC. The samples were obtained with the understanding and written consent of each subject. The study conforms with the Code of Ethics of the World Medical Association (Declaration of Helsinki), printed in the British Medical Journal (18 July 1964). Demographic information, including age, sex, race, smoking history, personal history of malignancy, and relevant family history were recorded. Samples were collected by the clinical team, blinded to the research analysis.

### 2.2. Patient characteristics

Patient data were collected at the UofL Health – Brown Cancer Center. There were six healthy controls, thirteen NSCLC patients, four of which did not respond to treatment, and nine who had no cancer growth after 12-weeks of treatment. Further classifying the NSCLC, eleven of the patients had adenocarcinoma and two had squamous cell carcinoma. One patient had a previous history of cancer, which was successfully treated. Mean patient age was 66 with most patients lying in 53–78 range; only patients of white and African American races were represented.

### 2.3. Evaluation of response

Data were collected at an initial visit and after 12-weeks at follow-up. Output variable was a two-level factor on whether the cancer progressed (non-responders) or stabilized/regressed (responders), determined via computed tomography (CT).

### 2.4. Human PBMC isolation

Whole blood samples were centrifuged at 1600 rpm for 10 min. Plasma was aspirated and aliquoted into 1 mL Eppendorf tubes and immediately stored at  $-80^{\circ}\text{C}$  until future use. Remaining cell layers were diluted with an equal volume of complete RPMI1640. Blood suspension was layered over 5 mL of Ficoll-Paque (Cedarlane Labs, Burlington, ON) in a 15 mL conical tube. Samples were then centrifuged at 2,000 rpm for 30 min at room temperature (RT) without brake. Mononuclear cell layer was then transferred to a new 15 mL conical tubes and washed with complete RPMI 1640. Cell pellet was resuspended in 3 mL of RPMI1640 and counted for sample processing.

## 2.5. Mass cytometry sample preparation and data collection

Cell expression and lineage markers were collected using Maxpar Direct Immune Profiling kit (Fluidigm, SKU 201325) and time-of-flight mass cytometry (cyTOF) techniques. PBMCs were isolated as described above. Cells were stained according to manufacturer's instructions. Briefly,  $0.8\text{--}1.5 \times 10^6$  cells per patient sample were washed twice with Maxpar Cell Staining Buffer. They were incubated with 5  $\mu\text{l}$  Human Fc block (Human TruStain FcX, Biolegend) for 10 min. Cells were then directly transferred to prepared tubes containing antibody pellet and stained for 30 min. Cells were washed and fixed in 1.6% formaldehyde for 10 min at RT, and then incubated overnight in 125 nM of Intercalator-Ir (Fluidigm) at 4°C.

Prior to acquisition, samples were washed twice with Cell Staining Buffer and kept on ice until acquisition. Cells were resuspended at a concentration of 1 million cells/mL in Cell Acquisition Solution containing a 1/9 dilution of EQ 4 Element Beads (Fluidigm). Samples were acquired on a Helios (Fluidigm) at an event rate of < 500 events/sec. After acquisition, data were normalized using bead-based normalization in the CyTOF software. Data were gated to exclude residual normalization beads, debris, dead cells and doublets, leaving DNA<sup>+</sup>CD45<sup>+</sup>Cisplatin<sup>low</sup> events for subsequent clustering and high dimensional analyses.

The study was limited to the markers provided in this kit, and because a kit was used, determining that some markers are not useful in outcome prediction may not reduce the materials, time, or cost of testing a sample. This differentiation will, however, provide insight to the downstream effects of immunotherapy treatment. Markers collected are shown in Supplementary Table 1. Maxpar changed their kit during the course of the study, so only samples detected by both versions of the kit were considered for analysis. Supplementary Table 2 and Supplementary Table 3 summarize the roles and cell populations affiliated with each of the markers used.

## 2.6. R cytometry libraries

R packages flowCore, FlowWorkspace, openCyto, ggCyto, and flowTrans were downloaded from Bioconductor open-source software for bioinformatics and were instrumental in performing storage, handling, gating, and transformation of the mass cytometry data.

## 2.7. Preprocessing: Mass cytometry normalization

Median calibration bead expression was computed for each set of data collected by CyTOF. The variance

among the calibration bead expression was found to be insignificant, so no normalization based on calibration beads was performed.

## 2.8. Data transformation for improved gating

To better identify distributions in the data collected from mass cytometry, hyperbolic arcsin transform was performed on the data. flowTrans was used to optimize the parameters  $a$  and  $b$  in the formula  $a \sinh(\text{data} * b + a)$  to achieve more distinct separations that help with automated gating [21].

## 2.9. Gating strategy to identify immune cells

Several layers of gating are required to attain useable data by removing doublets and dead cells so their expression does not influence later determination of cell types [22]. Approach is summarized in Supplementary Fig. 1:

1. Determine intact cells by gating cells with high DNA 191Ir and low signal in an empty channel (in this case 140Ce) Calibration beads should be visible in empty channels only – data with high DNA 191Ir and high bead signal should not be used.
2. DNA 191Ir and 193Ir should be strongly correlated – upper right is left off because those are likely cell doublets; this will not remove them all, but limits them.
3. Cells with lots of metal (and doublets) show higher event length, so filter those out; cell size may correlate with event length, but it is not causative.
4. High Rh indicates cell was dead before the test so it should not be used; CD45 marks immune cells.

## 2.10. Gating for immune cells

Gating was automated by specifying minimum and maximum thresholds for trough detection. This separated little to no expression from high expression of a given marker. Examples are shown in Supplementary Fig. 2 for 1D and 2D gating. In initial analysis, B-cell markers commonly good predictive features, so B-cells were chosen as the predominant cell investigated. Following initial gating to isolate immune cells, rest of the gating was performed as specified by Supplementary Table 4 on hyperbolic arcsin transformed data.

## 2.11. Forming input variables

Gating was automated by specifying minimum and maximum thresholds for trough detection. This sepa-

rated little to no expression from high expression of a given marker. Examples are shown in Supplementary Fig. 2 for 1D and 2D gating. In initial analysis, B-cell markers commonly good predictive features, so B-cells were chosen as the predominant cell investigated. Following initial gating to isolate immune cells, rest of the gating was performed as specified by Supplementary Table 4 on hyperbolic arcsin transformed data.

## 2.12. Model type and parameters

The models chosen for analysis include Multilayer Perceptron (MLP), Elastic Net (EN), Random Forest (RF), and Partial Least Squares Discriminant Analysis (PLSDA). These models represent machine learning algorithms that have been successfully applied to two-class outcome prediction problems. Logistic regression models were made using one feature at a time. Each type of model has parameters built into its architecture that can be varied as part of the optimization process. Every parameter setting was chosen via experimentation using a range of values. Models were trained using the caret R package [23]. Some settings and reporting measures were adjusted for incompatibility with model types.

## 2.13. Model validation

0.632 Bootstrapping was chosen as validation method. It involves many iterations of sampling the data with replacement to create a large group of sets. Each iteration has data that was not selected, and these samples are known as out-of-bag samples used for validation. Set of samples from each iteration is used to train the models and validate the results. Number of sampling iterations used for the validation was determined by experimentation. When model coefficients stabilized, the lower limit on iterations was found. Ten thousand iterations were found to be sufficient for each dataset. Some of these iterations were discarded when out-of-bag samples did not have at least one sample from each of the two outcome variables.

## 2.14. Assessing model performance

Performance metrics including AUROC (area under the receiver operating characteristic (ROC) curve), sensitivity, sensitivity, balanced accuracy, and kappa were recorded for every trained model, while AUROC was used as metric for validation. Pracma R package function *trapz* was used to compute AUROC trapezoidal

estimate. The *trainControl* function of caret, which sets up many parameters for the *train* function, has a parameter called *selectionFunction*. It was set to “oneSE” so that the selected model after training was with the AUROC within one squared error of the best. This was done to avoid model overfitting.

## 2.15. Feature importance

The *varImp* function from caret R package was used to determine which variables were most important in making the prediction for each model. With a large set of features, each model could select a different set and still have great performance. To add another metric for identifying important variables, Spearman’s rho statistic was used to determine rank-based association. Features that have low *p*-value for Spearman’s rho, are selected by multiple models, and have high variable importance from the *varImp* function were marked as key features.

The *varImp* function from caret R package was used to determine which variables were most important in making the prediction for each model. PLSDA weights are a function of the reduction of the sums of squares across the number of PLS component. Elastic Net variable importance is the absolute value of the coefficients of the tuned model. For Random Forest *varImp* calculates the mean of the (scaled) class-specific decreases in accuracy and reports it for each of the classes. MLP and other models that do not have a predefined method use a filtering approach where an ROC curve analysis is conducted on each feature [23].

## 2.16. Feature selection

Regularized Random Forest (RRF) and Partial Least Squares (PLS) methods were used for feature selection with the RF and MLP models. The models were made with an increasing number of top variables until a peak model performance was found. The other models did not require feature selection because it is intrinsic to them. Accordingly, the six models used for predictive analysis were: RF, RRF-RF, RRF-MLP, PLSDA, PLS-MLP, and EN, where models with two names (e.g., RRF-RF) denote first the feature selection method hyphenated with the predictive model. Corresponding optimal model parameters are in Supplementary Table 5.

# 3. Results

Patient characteristics are summarized in Table 1. A flow diagram of study workflow is in Fig. 1, describing the training of the models and their validation.

Table 1  
Patient characteristics

| Patient # | Sex | Age | NSCLC subtype                               | Stage | T  | N | M  | Therapy       | Treatment Naïve |
|-----------|-----|-----|---------------------------------------------|-------|----|---|----|---------------|-----------------|
| 1         | M   | 62  | Adenocarcinoma-Bronchioloalveolar Carcinoma | 1     | 1b | 0 | –  | Pembrolizumab | Y               |
| 2         | F   | 53  | Adenocarcinoma – NOS                        | 4     | 3  | 2 | 1b | Pembrolizumab | Y               |
| 3         | F   | 73  | Squamous cell carcinoma                     | 3     | 4  | 3 | 0  | Atezolizumab  | N               |
| 4         | F   | 78  | Adenocarcinoma – NOS                        | 4     | 2  | 2 | 1a | Pembrolizumab | N               |
| 5         | F   | 56  | Adenocarcinoma – NOS                        | 4     | 3  | 2 | 1a | Pembrolizumab | Y               |
| 6         | F   | 72  | Adenocarcinoma – NOS                        | 4     | 1  | 2 | 1b | Atezolizumab  | N               |
| 7         | F   | 56  | Adenocarcinoma – NOS                        | 4     | –  | – | 1b | Atezolizumab  | N               |
| 8         | M   | 95  | Adenocarcinoma – NOS                        | 4     | 4  | 0 | 1  | Pembrolizumab | Y               |
| 9         | F   | 37  | Adenocarcinoma – NOS                        | 1     | 2a | 0 | 0  | Nivolumab     | N               |
| 10        | M   | 69  | Squamous cell carcinoma                     | 4     | 2a | 0 | 1a | Pembrolizumab | N               |
| 11        | F   | 63  | Adenocarcinoma – NOS                        | 4     | 1b | 0 | 1a | Pembrolizumab | Y               |
| 12        | F   | 71  | Adenocarcinoma – NOS                        | 3a    | 3  | 1 | –  | Durvalumab    | Unknown         |
| 13        | F   | 78  | Adenocarcinoma – NOS                        | 4     | 1a | 3 | 1  | Pembrolizumab | Unknown         |

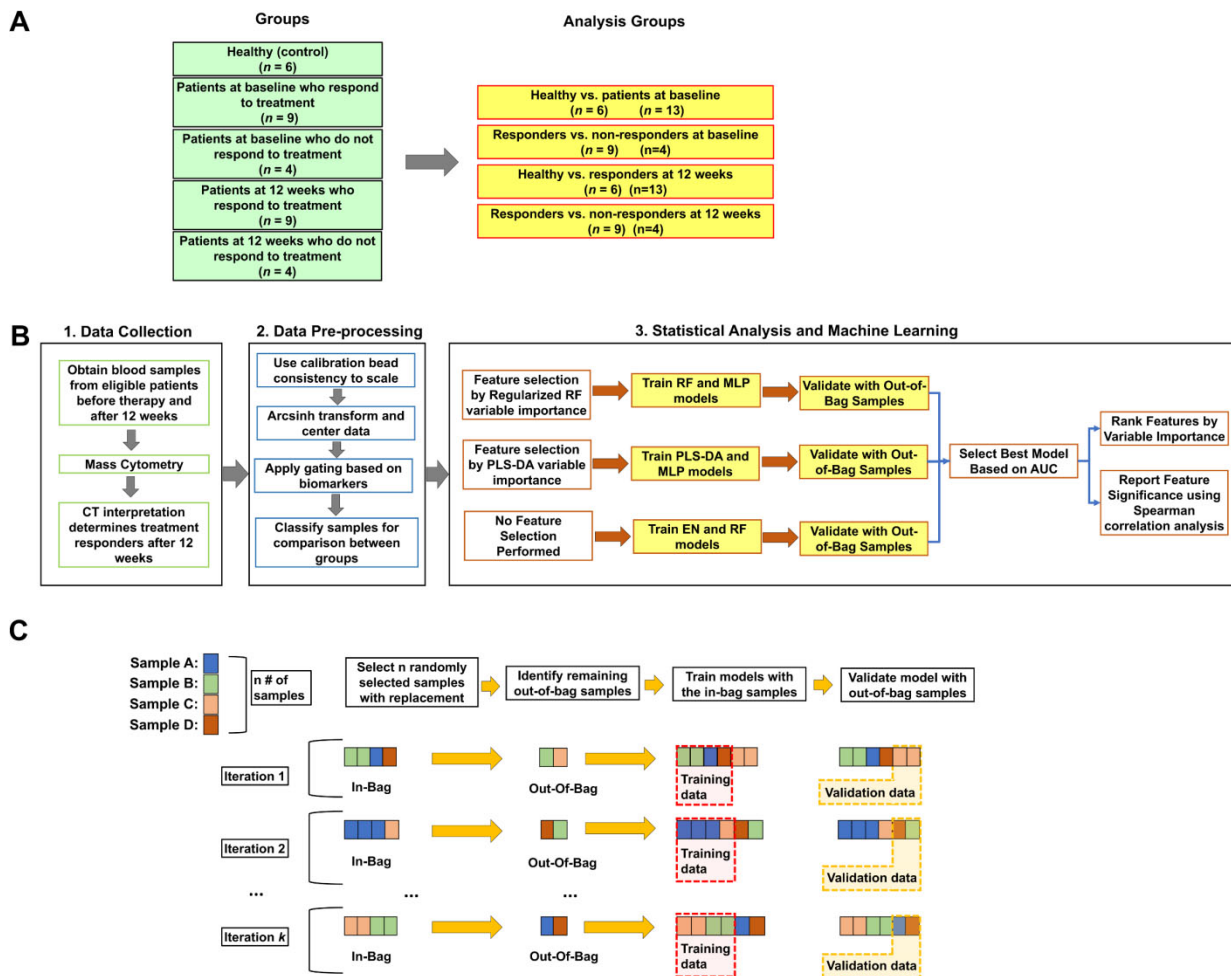


Fig. 1. Study Workflow. (A) Study profile. Of 13 patients eligible for the study and placed on anti-PD-1 therapy, 9 patients had regressed or stable disease, and 4 had disease progression at a twelve-week follow-up. 3 non-responders were on Pembrolizumab and 1 was on atezolizumab. (B) Maxpar Direct Immune Profiling Kit for mass cytometry was used to determine cell expression and lineage markers. Variable importance was used to select feature subsets before training PLS-DA, RF and MLP models. (C) Visual example of model training and validation. For bootstrapping validation, n randomly selected samples are drawn with replacement from a population with n samples. Samples that are drawn are known as in-bag samples and are used as the training data for the models. Any unselected data is known as out-of-bag and is used for model validation. This process is repeated k iterations. Results of each model are the averages of the validations across all iterations. Colour figure online.

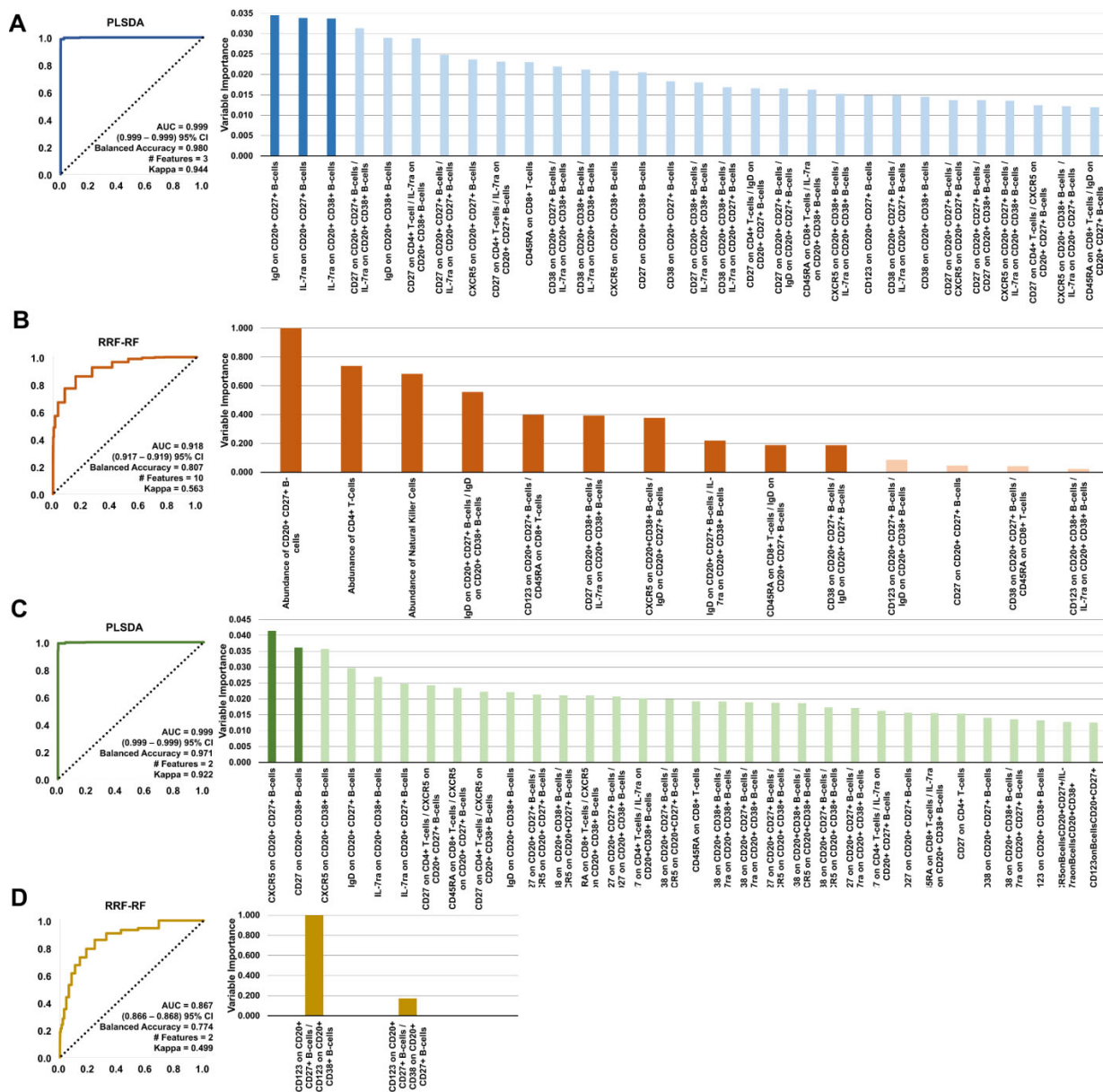


Fig. 2. Classification model results: (A) Healthy; (B) Baseline Responders vs Non-Responders at baseline; (C) Healthy vs 12-week Responders; (D) Responders vs Non-Responders at 12 weeks. Each classification shows AUROC for highest performing model and variable importance graph ranking. Darker bars indicate subset of features needed for the best model to perform as shown in ROC plots. Colour figure online.

The following four patient comparisons were evaluated between data collected at an initial visit (baseline), after 12-weeks of immunotherapy, and from healthy (control) samples: healthy vs patients at baseline, Responders vs Non-Responders at baseline, Healthy vs 12-week Responders, and Responders vs Non-Responders at 12-weeks. The algorithms Random Forest (RF), Partial Least Squares Discriminant Analysis (PLS-DA), Multi-Layer Perceptron (MLP), and Elastic Net (EN)

were applied to find features differentiating between these groups and provide for the capability to predict outcomes. Figure 2 reports the highest performing models for each of the patient classifications: Healthy vs. patients at baseline with PLS-DA (AUC = 0.999, (0.999–0.999) 95% CI), Responders vs. Non-Responders at baseline with RRF-RF (AUC = 0.918 (0.917–0.919) 95% CI), Healthy vs. Responders at 12 weeks with PLS-DA (AUC = 0.999 (0.999–0.999) 95% CI), and

Table 2  
Model outcome summaries. Features used to build the best performing models are indicated in the selected features column

| Comparison                               | Best model | Selected features                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|------------------------------------------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Healthy vs baseline                      | PLS-DA     | IgD on CD20+ CD27+ B-cells<br>IL-7R $\alpha$ on CD20+ CD27+ B-cells<br>IL-7R $\alpha$ on CD20+ CD38+ B-cells                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Responders vs non-responders at baseline | RRF-RF     | Abundance of CD20+ CD27+ B-cells<br>Abundance of Natural Killer (NK) Cells<br>Abundance of CD4+ T-cells<br>CD123 on CD20+ CD27+ B-cells/CD45RA on CD8+ T-cells<br>CD27 on CD20+ CD27+ B-cells/IL-7R $\alpha$ on CD20+ CD38+ B-cells<br>CD38 on CD20+ CD27+ B-cells/IgD on CD20+ CD27+ B-cells<br>CD45RA on CD8+ T-cells/IgD on CD20+ CD27+ B-cells<br>CXCR5 on CD20+ CD38+ B-cells/IgD on CD20+ CD27+ B-cells<br>IgD on CD20+ CD27+ B-cells/IgD on CD20+ CD38+ B-cells<br>IgD on CD20+ CD27+ B-cells/IL-7R $\alpha$ on CD20+ CD38+ B-cells |
| Healthy vs 12-week responders            | PLS-DA     | CXCR5 on CD20+ CD27+ B-cells<br>CD27 on CD20+ CD38+ B-cells                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Responders vs non-responders at 12-weeks | RRF-RF     | CD123 on CD20+ CD27+ B-cells/CD123 on CD20+ CD38+ B-cells<br>CD123 on CD20+ CD27+ B-cells/CD38 on CD20+ CD27+ B-cells                                                                                                                                                                                                                                                                                                                                                                                                                      |

Responders vs. Non-Responders at 12 weeks with RRF-RF (AUC = 0.867 (0.866–0.868) 95% CI). Supplementary Fig. 3 summarizes the performance measures (AUC, balanced accuracy, and kappa) as a function of features retained for each classification for the supervised learning methods (PLS-DA, RRF-RF, RRF-MLP, PLS-MLP). ROC curves for all models evaluated are in Supplementary Fig. 4. Figure 2 further shows highest ranking features for the best performing models, while Fig. 3 shows features with a significant effect of group in terms of relative abundance (Wilcoxon rank-sum test,  $p \leq 0.05$ ). Ranking of features for all evaluated models based on variable importance is in Supplementary Table 6 through Supplementary Table 21.

For Healthy vs Baseline comparison (Figs 2A and 3A), PLS-DA was the best performing model, and it only needed the three features of IgD on B-cells that are CD20+ CD27+, IL-7R $\alpha$  on B-cells that are CD20+ CD27+, and IL-7R $\alpha$  on B-cells that are CD20+ CD38+ to make accurate predictions. These features were statistically significant when comparing their correlations to the outcome.

Comparison of Responders vs Non-Responders at baseline (Figs 2B and 3B) evaluated all samples before immunotherapy. RRF-RF was the best performing model with 10 features selected as important. Five of these were significant with Wilcoxon rank-sum test in addition to being selected by variable importance: amount of CD20+ CD27+ B-cells, Natural Killer cells, CD4+ T-cells, CD123 on B-cells that are CD20+ CD27+/CD45RA expression on CD8+ T-cells, and IgD on B-cells that are CD20+ CD27+/IgD on B-cells that are CD20+ CD38+.

Healthy subjects and 12-week Responders were best predicted by the PLS-DA model (Figs 2C and 3C), with only two features: CXCR5 on B-cells that are CD20+ CD27+ and CD27 on B-cells that are CD20+ CD38+. Lastly, comparison of Responders vs Non-Responders at 12-weeks required only two features with RRF-RF (Figs 2D and 3D): CD123 expression on CD20+ CD27+ B-cells/CD123 expression on CD20+ CD38+ B-cells and CD123 on CD20+ CD27+ B-cells/CD38 on CD20+ CD27+ B-cells were the only important features selected by the model. The most important features selected by various models between group comparisons are summarized in Table 2.

#### 4. Discussion

This proof-of-concept study applied a variety of machine learning approaches and feature selection methods to predict response of NSCLC patients to any one of nivolumab, pembrolizumab, atezolizumab, and durvalumab immunotherapies based on mass cytometry data of blood-borne immune cell lineage and surface markers. Four different patient comparisons were evaluated: healthy vs baseline, responders vs non-responders at baseline, healthy vs 12-week responders, and responders vs non-responders at 12-weeks. Homogenizing the groups based on the immunotherapy is another route for analysis that could provide essential information to better understand how each treatment affects the immune system differently, but there were insufficient samples to perform that analysis in this study. Some research has been done comparing each of the



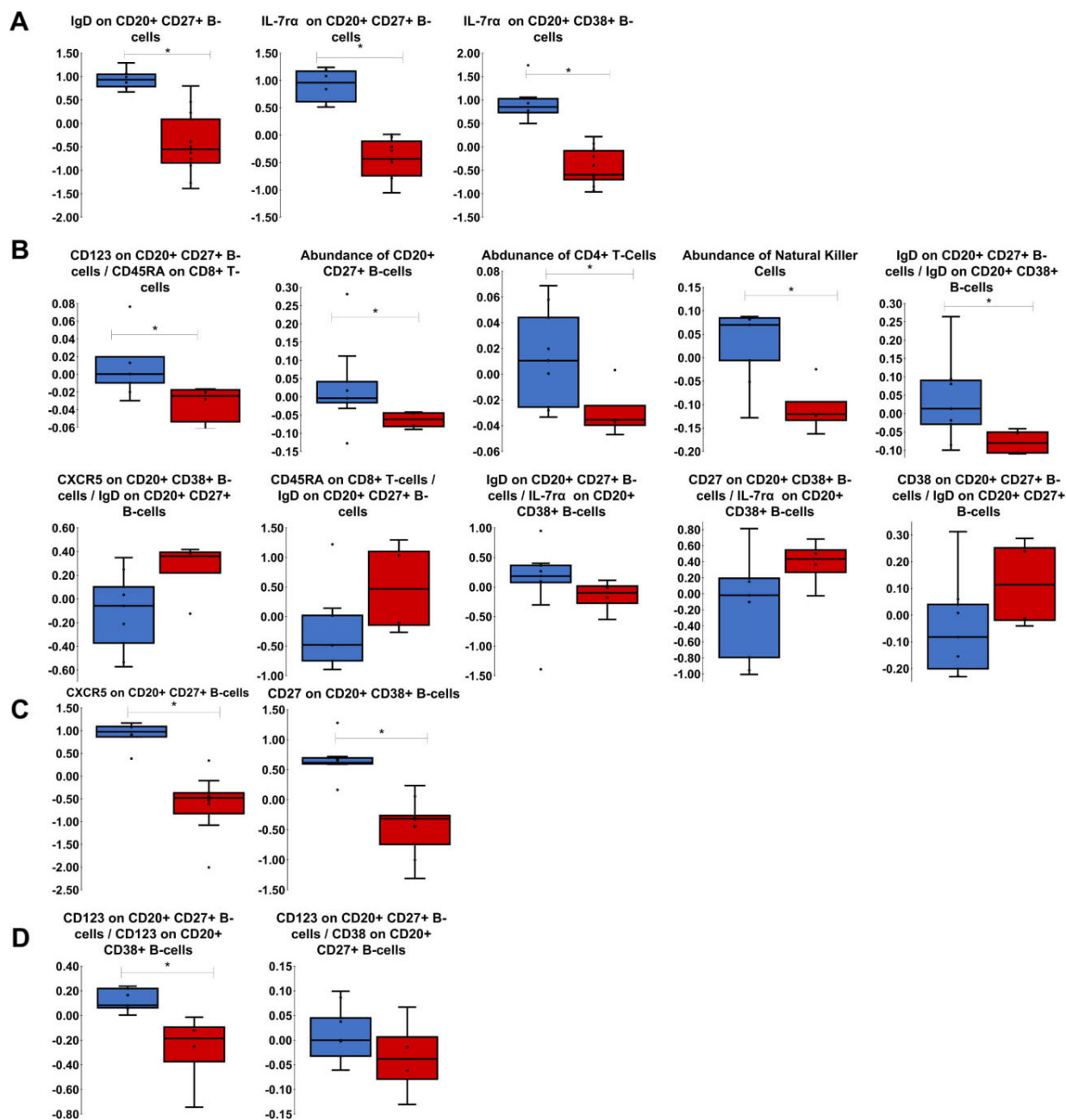


Fig. 3. Relative abundance of key features for each comparison. Each box represents 1<sup>st</sup> and 3<sup>rd</sup> quartiles. Bands within represent the median and x is the mean. Ends of whiskers are maximum and minimum, with points outside being outliers. *P*-values found by Wilcoxon rank-sum test (\**p* ≤ 0.05). (A) Healthy (Blue) vs. Patient (Red) at baseline; (B) Responder (Blue) vs. Non-Responder at baseline (Red); (C) Healthy (Blue) vs. 12-week Responder (Red); (D) Responder (Blue) vs. Non-Responder (Red) at 12 weeks. Colour figure online.

PD-1/PD-L1 inhibitors, and in many ways, there is no significant difference in NSCLC progression between nivolumab, pembrolizumab, atezolizumab, and durvalumab [24,25]. One difference that was noted is that durvalumab had a better overall response rate than atezolizumab [24]. The 95% confidence interval on the

hazard ratio, however, nearly expanded to 1 by reaching 0.98, so further investigation may be warranted. The differences in the side effect profiles of each treatment certainly prove that they affect the immune system differently, but since the cancers do not progress differently, we suspect these features of the immune system



were insignificant to the model predictions. For this reason and because they all affect the same receptor/ligand combination, the four immunotherapies were considered interchangeable in this analysis. The study adhered to CHARMS checklist [26] to ensure the integrity of the prediction models. Although the number of samples was small, the number of features obtained through the study methodology was sufficient to accurately differentiate between the samples without overfitting to the data. A traditional method of validation, k-fold cross-validation, was unable to be used because its limitations with small sample sizes. Bootstrapping validation, as used in this study does not have the same restrictions, and with high enough iterations, simulates outcomes comparable to k-fold cross-validation. More samples would be needed for an additional layer of external validation.

Regularized Random Forest feature selection followed by a Random Forest model (RRF-RF) and partial least squares discriminant analysis (PLSDA) had the highest AUC ( $> 0.99$ ) when predicting whether a sample was from a NSCLC patient or a healthy control. They also had AUC  $> 0.85$  when separating cancer patients from each other based on who responded to treatment. Further, this study implemented proportions of expressions of immune cell markers as features and found that some of these proportions provided for statistically significant separation between groups. Interestingly, several proportion features were found to be in the top ten predictor variables for each model. To our knowledge, proportions of immune cell markers have not been previously considered as a feature to predict patient outcomes. Further exploration of these proportions may elucidate interactions between immune and cancer cells in the immunotherapy response of NSCLC.

The cell types chosen for evaluation this study (Supplementary Table 4) represent fundamental cells that make up the innate and adaptive immune systems. The role of cytotoxic CD8<sup>+</sup> T-cells in the adaptive immune response against cancer has been extensively studied, and advances have been done to modify these cells to better target specific cancers [27]. Interestingly, many of the features used to build the best performing models in this study were related to CD20<sup>+</sup> B-cells, suggesting a fundamental role for these cells in the response to monoclonal antibody therapy. These results are consistent with recent findings highlighting the potential function of these cells in lung cancer immunobiology [28,29]. CD20<sup>+</sup> B-cells include most types of B-cells except for normal plasma B-cells. Proliferation of CD20<sup>+</sup> B-cells serves many purposes in the adaptive immune response

such as antigen presentation and eventually antibody creation. Recently, CD20<sup>+</sup> B-cells were found to help predict prognosis in colorectal cancer [30]. In contrast, features based on antigen presenting cells (APCs) such as dendritic cells, monocytes, and macrophages did not become part of the best performing models. Antigen presenting cells help transport foreign antigens to secondary lymphoid centers and trigger T-cells to initiate the adaptive response. Like cytotoxic T-cells, some studies have focused on modifying these cells as part of cancer immunotherapy [31].

Higher IL-7R $\alpha$  expressions on CD27<sup>+</sup> and CD38<sup>+</sup> B-cells were important features that separated healthy from patients at baseline (Fig. 3A). IL-7 is a key element in B cell development, including specification and commitment of lymphoid stem cells to the B lineage, survival and proliferation of B cell progenitors, and maturation during the pro-B to pre-B transition [32]. Binding of IL-7R by IL-7 has been shown to increase IL-6 secretion [33], which promotes antibody production by B cells [34]. Our study examined IL-7R on some B-cell populations and found expression was diminished in NSCLC in comparison to healthy samples. Less IL-7R expression suggests that these B-cells are less able to proliferate [35], and to mount a humoral defense against the cancer.

One of the highest ranked features in the Baseline Responders vs Baseline Non-Responders comparison was the abundance of NK cells (Fig. 3B). Patients who saw improvement with immunotherapy had more NK cells than those who had cancer progression. It is known that the NSCLC tumor microenvironment suppresses NK cell function [36]. Higher abundance of NK cells prior to treatment suggests that immune cell function may be less altered by the tumor and thus immunotherapy is able to trigger an effective response. Another highly ranked feature in this patient comparison was the abundance of CD4<sup>+</sup> T cells, which was also higher in Baseline Responders. The multifaceted role of these cells in the anti-cancer immune response was recently reviewed [37], highlighting their potential in the rational design of immunotherapies. Further evaluation of the surface and lineage expression markers on this cell type could help target CD4<sup>+</sup> T-cells to improve patient outcomes.

Several proportions of surface markers between immune cells were found to be significantly higher in Responders than in Non-Responders at baseline (Fig. 3B), including CD123 on CD20<sup>+</sup> CD27<sup>+</sup> B-cells/CD45RA on CD8<sup>+</sup> T-cells and IgD on CD20<sup>+</sup> CD27<sup>+</sup> B-cells/IgD on CD20<sup>+</sup> CD38<sup>+</sup> B-cells. Other propor-

tions that were essential for prediction include CD27 on CD20+ CD27+ B-cells/IL-7R $\alpha$  on CD20+ CD38+ B-cells, CD38 on CD20+ CD27+ B-cells/IgD on CD20+ CD27+ B-cells, CD45RA on CD8+ T-cells/IgD on CD20+ CD27+ B-cells, CXCR5 on CD20+ CD38+ B-cells/IgD on CD20+ CD27+ B-cells, IgD on CD20+ CD27+ B-cells/IL-7R $\alpha$  on CD20+ CD38+ B-cells. Achieving an immunological understanding of the mechanisms underlying these interactions requires further investigation.

CXCR5 on CD20+ CD27+ B-cells was found to be higher in healthy patients in comparison to 12-week Responders (Fig. 3C). CXCR5 is a chemoattractant that guides B-cells to secondary lymphoid organs [38]. It was previously reported that CXCR5 expression in NSCLC patients was often elevated, and the cancer expresses the ligand for its receptor [39]. Since patients in this study had low expression of CXCR5 after responding to treatment, it is possible that Pembrolizumab influences this interaction between B-cells and NSCLC. CD27 on CD20+ CD38+ B-cells was also found to be higher in healthy compared to 12-week Responders (Fig. 3C). CD27 is involved in the co-stimulation of B and T-cells, which is why it has been recently investigated as a potential immunotherapy target for cancers [40]. Varlilumab is a CD27 agonist and has been studied for the treatment of many solid tumors in combination to PD-1 targeted therapies. Evidence was previously provided for this treatment while concluding that further trials may be needed [40]. Further investigation of CD27 signaling could provide more information about the mechanisms underlying an immune response to NSCLC.

The proportion of CD123 on CD20+ CD27+ B-cells to CD123 on CD20+ CD38+ B-cells and CD123 on CD20+ CD27+ B-cells to CD38 on CD20+ CD27+ B-cells was higher in Responders than in Non-Responders at 12 weeks (Fig. 3D). CD123 is the IL3 receptor, and it is known to be upregulated in proliferating B-cells and induce inflammatory cytokine secretion [41]. It has been studied for its role in dendritic cell response to NSCLC [42], but its expression on B-cells has not. Studies have also shown that CD123 expression on B-cells is related to certain subtypes of leukemias and lymphomas [43, 44], but further investigation is now warranted for its role in B-cell response to NSCLC.

Combinations of immunotherapy with radiotherapy have been studied, but it remains unclear which patients would benefit from combination treatment [45]. Chemotherapy options have been compared to antibody treatments for determination of which would be a more

successful first line treatment. Docetaxel has been studied as an alternative to Pembrolizumab in NSCLC, and it was concluded that Pembrolizumab offered overall longer survival [46]. Although combination therapies have become standard of care, this study focused on patients on monoclonal antibody monotherapies to isolate relevant features for analysis.

The features highlighted in this study should be investigated further to assess their involvement in NSCLC immunotherapy response and prediction of patient outcomes. Variables other than immune cell lineage and surface markers may also contribute to these outcomes. Of note, some features found in this study may have no clinical significance, and their importance to prediction might be accredited to random chance based on the small sample size or to unknown mechanisms of interaction in the immune response. Limitations of this study include the small number of samples and the existence of potential confounding factors such as age, ethnicity, and type of immunotherapy. These limitations could cause the models to have poor generalization to patient populations with dissimilar characteristics, and because of this, the models should be validated with additional data sets. This validation would strengthen the argument for use of monoclonal antibody monotherapy on NSCLC patients based on particular immune profile characteristics.

## Acknowledgments

Authors acknowledge the UofL Health – Brown Cancer Center Biorepository and the Brown Cancer Center Clinical Trial office for their support of this project. HF acknowledges partial support by the National Institutes of Health/National Cancer Institute grant R15CA203605. JY acknowledges partial support by the National Institutes of Health/National Cancer Institute grant R01CA213990. CyTOF was performed in the Functional Immunomics Core supported by NIH P20GM135004.

## Author contributions

Conception: AM, SM, HM, RK, PN, JY, HF.  
Interpretation or analysis of data: AM, SM, HM, XH, JY, HF.  
Preparation of the manuscript: AM, SM, HM, JY, HF.  
Revision for important intellectual content: AM, SM, HM, JY, HF.  
Supervision: JY, HF.

## Supplementary data

The supplementary files are available to download from <http://dx.doi.org/10.3233/CBM-210529>.

## References

- [1] G. Morad, B.A. Helmink, P. Sharma and J.A. Wargo, Hallmarks of response, resistance, and toxicity to immune checkpoint blockade, *Cell* **184** (2021), 5309–5337.
- [2] H.T. Lee, J.Y. Lee, H. Lim, S.H. Lee, Y.J. Moon, H.J. Pyo, S.E. Ryu, W. Shin and Y.S. Heo, Molecular mechanism of PD-1/PD-L1 blockade via anti-PD-L1 antibodies atezolizumab and durvalumab, *Sci Rep* **7** (2017), 5532.
- [3] K. Leventakos and A.S. Mansfield, Advances in the treatment of non-small cell lung cancer: Focus on nivolumab, pembrolizumab, and atezolizumab, *BioDrugs* **30** (2016), 397–405.
- [4] F. Passiglia, A. Galvano, S. Rizzo, L. Incorvaia, A. Listi, V. Bazan and A. Russo, Looking for the best immune-checkpoint inhibitor in pre-treated NSCLC patients: An indirect comparison between nivolumab, pembrolizumab and atezolizumab, *Int J Cancer* **142** (2018), 1277–1284.
- [5] S.N. Gettinger, L. Horn, L. Gandhi, D.R. Spigel, S.J. Antonia, N.A. Rizvi, J.D. Powderly, R.S. Heist, R.D. Carvajal, D.M. Jackman, L.V. Sequist, D.C. Smith, P. Leming, D.P. Carbone, M.C. Pinder-Schenck, S.L. Topalian, F.S. Hodi, J.A. Sosman, M. Sznol, D.F. McDermott, D.M. Pardoll, V. Sankar, C.M. Ahlers, M. Salvati, J.M. Wigginton, M.D. Hellmann, G.D. Kolli, A.K. Gupta and J.R. Brahmer, Overall survival and long-term safety of nivolumab (anti-programmed death 1 antibody, BMS-936558, ONO-4538) in patients with previously treated advanced non-small-cell lung cancer, *J Clin Oncol* **33** (2015), 2004–2012.
- [6] E.B. Garon, N.A. Rizvi, R. Hui, N. Leighl, A.S. Balmanoukian, J.P. Eder, A. Patnaik, C. Aggarwal, M. Gubens, L. Horn, E. Carcereny, M.J. Ahn, E. Felip, J.S. Lee, M.D. Hellmann, O. Hamid, J.W. Goldman, J.C. Soria, M. Dolled-Filhart, R.Z. Rutledge, J. Zhang, J.K. Linceford, R. Rangwala, G.M. Lubiniecki, C. Roach, K. Emancipator, L. Gandhi and K. Investigators, Pembrolizumab for the treatment of non-small-cell lung cancer, *N Engl J Med* **372** (2015), 2018–2028.
- [7] L. Horn, S.N. Gettinger, M.S. Gordon, R.S. Herbst, L. Gandhi, E. Felip, L.V. Sequist, D.R. Spigel, S.J. Antonia, A. Balmanoukian, P.A. Cassier, B. Liu, M. Kowanzet, C. O'Hear, M. Fasso, W. Grossman, A. Sandler and J.C. Soria, Safety and clinical activity of atezolizumab monotherapy in metastatic non-small-cell lung cancer: Final results from a phase I study, *Eur J Cancer* **101** (2018), 201–209.
- [8] H. Yang, K. Shen, C. Zhu, Q. Li, Y. Zhao and X. Ma, Safety and efficacy of durvalumab (MEDI4736) in various solid tumors, *Drug Des Devel Ther* **12** (2018), 2085–2096.
- [9] A. Hindman, Cost of anti-cancer therapies demands new strategies for decision-making, shared responsibility, communication, *Oncology Issues* **34** (2019), 54–57.
- [10] A. Filipovic, G. Miller and J. Bolen, Progress toward identifying exact proxies for predicting response to immunotherapies, *Front Cell Dev Biol* **8** (2020), 155.
- [11] G.T. Gibney, L.M. Weiner and M.B. Atkins, Predictive biomarkers for checkpoint inhibitor-based immunotherapy, *The Lancet Oncology* **17** (2016), e542–e551.
- [12] T. Shukuya and D.P. Carbone, Predictive markers for the efficacy of anti-PD-1/PD-L1 antibodies in lung cancer, *Journal of Thoracic Oncology* **11** (2016), 976–988.
- [13] L.A. Hunter, Y.P. Chen, L. Zhang, J.E. Matney, H. Choi, S.F. Kry, M.K. Martel, F. Stingo, Z. Liao and D. Gomez, NSCLC tumor shrinkage prediction using quantitative image features, *Computerized Medical Imaging and Graphics* **49** (2016), 29–36.
- [14] S.-W. Chang, S. Abdul-Kareem, A.F. Merican and R.B. Zain, Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods, *BMC bioinformatics* **14** (2013), 170.
- [15] Z. Li, L. Yang, Y. Liu, H. Xu, S. Wang, Y. Liu, H. Zhao, J. Xie, S. Gao and S. Gao, Anti-inflammatory and antioxidative effects of Dan-Lou tablets in the treatment of coronary heart disease revealed by metabolomics integrated with molecular mechanism studies, *Journal of Ethnopharmacology* **240** (2019), 111911.
- [16] P. Milanez-Almeida, A.J. Martins, R.N. Germain and J.S. Tsang, Cancer prognosis with shallow tumor RNA sequencing, *Nature Medicine* **26** (2020), 188–192.
- [17] J. Das, K.M. Gayvert, F. Bunea, M.H. Wegkamp and H. Yu, ENCAPP: Elastic-net-based prognosis prediction and biomarker discovery for human cancers, *BMC genomics* **16** (2015), 1–13.
- [18] I.S. Hofer, C. Lee, E. Gabel, P. Baldi and M. Cannesson, Development and validation of a deep neural network model to predict postoperative mortality, acute kidney injury, and reintubation using a single feature set, *npj Digital Medicine* **3** (2020), 1–10.
- [19] R. Jeena and A. Suresh Kumar, Stroke risk assessment using ridge regression model, in: *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, 2018, pp. 1–3.
- [20] H.A. Miller, X. Yin, S.A. Smith, X. Hu, X. Zhang, J. Yan, D.M. Miller, V.H. van Berkel and H.B. Frieboes, Evaluation of disease staging and chemotherapeutic response in non-small cell lung cancer from patient tumor-derived metabolomic data, *Lung Cancer* **156** (2021), 20–30.
- [21] G. Finak, J. Manuel-Perez and R. Gottardo, flowTrans: Parameter optimization for flow cytometry data transformation, *R package version 1.24.0* (2010).
- [22] A. Cossarizza, H.D. Chang, A. Radbruch, A. Acs, D. Adam, S. Adam-Klages, W.W. Agace, N. Aghaeepour, M. Akdis and M. Allez, Guidelines for the use of flow cytometry and cell sorting in immunological studies, *European Journal of Immunology* **49** (2019), 1457–1973.
- [23] M. Kuhn, Building predictive models in R using the caret package, *Journal of Statistical Software* **28** (2008), 1–26.
- [24] J. Chen, J. Wang and H. Xu, Comparison of atezolizumab, durvalumab, pembrolizumab, and nivolumab as first-line treatment in patients with extensive-stage small cell lung cancer: A systematic review and network meta-analysis, *Medicine* **100** (2021).
- [25] T.-R. Peng, F.-P. Tsai and T.-W. Wu, Indirect comparison between pembrolizumab and nivolumab for the treatment of non-small cell lung cancer: A meta-analysis of randomized clinical trials, *International Immunopharmacology* **49** (2017), 85–94.
- [26] K.G. Moons, J.A. de Groot, W. Bouwmeester, Y. Vergouwe, S. Mallett, D.G. Altman, J.B. Reitsma and G.S. Collins, Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist, *PLoS Med* **11** (2014), e1001744.
- [27] C. Herbel, N. Patsoukis, K. Bardhan, P. Seth, J.D. Weaver and V.A. Boussiotis, Clinical significance of T cell metabolic reprogramming in cancer, *Clinical and translational medicine*

- 5 (2016), 1–23.
- [28] A.J. Patel, A. Richter, M.T. Drayson and G.W. Middleton, The role of B lymphocytes in the immuno-biology of non-small-cell lung cancer, *Cancer Immunol Immunother* **69** (2020), 325–342.
- [29] T.L. Leong and V.L. Bryant, B cells in lung cancer-not just a bystander cell: A literature review, *Transl Lung Cancer Res* **10** (2021), 2830–2841.
- [30] S. Edin, T. Kaprio, J. Hagstrm, P. Larsson, H. Mustonen, C. Bckelman, K. Strigrd, U. Gunnarsson, C. Haglund and R. Palmqvist, The prognostic importance of CD20+ B lymphocytes in colorectal cancer and the relation to other immune cell subsets, *Scientific Reports* **9** (2019), 1–9.
- [31] L.J. Eggermont, L.E. Paulis, J. Tel and C.G. Figdor, Towards efficient cancer immunotherapy: Advances in developing artificial antigen-presenting cells, *Trends in Biotechnology* **32** (2014), 456–465.
- [32] K. Takatsu, Cytokines involved in B-cell differentiation and their sites of action, *Proc Soc Exp Biol Med* **215** (1997), 121–133.
- [33] M. Iwata, L. Graf, N. Awaya and B. Torok-Storb, Functional interleukin-7 receptors (IL-7Rs) are expressed by marrow stromal cells: Binding of IL-7 increases levels of IL-6 mRNA and secreted protein, *Blood* **100** (2002), 1318–1325.
- [34] O. Dienz, S.M. Eaton, J.P. Bond, W. Neveu, D. Moquin, R. Noubade, E.M. Briso, C. Charland, W.J. Leonard, G. Ciliberto, C. Teuscher, L. Haynes and M. Rincon, The induction of antibody production by IL-6 is indirectly mediated by IL-21 produced by CD4+ T cells, *J Exp Med* **206** (2009), 69–78.
- [35] M.R. Clark, M. Mandal, K. Ochiai and H. Singh, Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling, *Nature Reviews Immunology* **14** (2014), 69–80.
- [36] I. Cremer, W.H. Fridman and C. Sauts-Fridman, Tumor microenvironment in NSCLC suppresses NK cells function, *Oncoimmunology* **1** (2012), 244–246.
- [37] R.E. Tay, E.K. Richardson and H.C. Toh, Revisiting the role of CD4(+) T cells in cancer immunotherapy-new insights into old paradigms, *Cancer Gene Ther* **28** (2021), 5–17.
- [38] G. Muller, U.E. Hopken and M. Lipp, The impact of CCR7 and CXCR5 on lymphoid organ development and systemic immunity, *Immunol Rev* **195** (2003), 117–135.
- [39] R. Singh, S. Singh, W.E. Grizzle and J.W. Lillard, Expression and functional role of CXCR5 and CXCL13 in non-small cell lung carcinoma (NSCLC), *AACR* (2011).
- [40] A.M. Starzer and A.S. Berghoff, New emerging targets in cancer immunotherapy: CD27 (TNFRSF7), *ESMO Open* **4** (2019), e000629.
- [41] B. Renner, S. Metz, A.-M. Metzger, S. Neumayer, K. Schmidbauer, Y. Talke, S. Buchtler, D. Halbritter and M. Mack, Expression of IL-3 receptors and impact of IL-3 on human T and B cells, *Cellular Immunology* **334** (2018), 49–60.
- [42] B. Stankovic, H.A.K. Bjrhovde, R. Skarshaug, H. Aamodt, A. Frafjord, E. Miller, C. Hammarstrm, K. Beraki, E.S. Bkkevold and P.R. Woldbk, Immune cell composition in human non-small cell lung cancer, *Frontiers in Immunology* **9** (2019), 3101.
- [43] M. Djokic, E. Bjrkklund, E. Blennow, J. Mazur, S. Sderhll and A. Porwit, Overexpression of CD123 correlates with the hyperdiploid genotype in acute lymphoblastic leukemia, *Haematologica* **94** (2009), 1016.
- [44] U. Testa, E. Pelosi and A. Frankel, CD 123 is a membrane biomarker and a therapeutic target in hematologic malignancies, *Biomarker Research* **2** (2014), 1–11.
- [45] E.C. Ko and S.C. Formenti, Radiotherapy and checkpoint inhibitors: A winning new combination? *Therapeutic Advances in Medical Oncology* **10** (2018), 1758835918768240.
- [46] R.S. Herbst, P. Baas, D.-W. Kim, E. Felip, J.L. Prez-Gracia, J.-Y. Han, J. Molina, J.-H. Kim, C.D. Arvis and M.-J. Ahn, Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): A randomised controlled trial, *The Lancet* **387** (2016), 1540–1550.