Explanatory machine learning modeling of a seven-molecular marker panel for immune checkpoint inhibition response per RECIST v1.1 in metastatic urothelial carcinoma

Research Project- Matthew Li

Introduction

Metastatic urothelial carcinoma (mUC) remains a devastating cancer to treat, with a 5-year survival rate of only 5%. The primary treatment regimens for mUC consist of platinum-based chemotherapy such as cisplatin and carboplatin. With this chemotherapy treatment, the survival rate for mUC is on average 12-15 months, in comparison to 3-6 months without treatment. Regardless of platinum-based chemotherapy being the standard-of-care for mUC, the Eastern Cooperative Oncology Group has determined that roughly 60% of patients are not eligible for this treatment. Thus, many patients require alternative modalities of cancer treatment.

Immune checkpoint inhibitors (ICIs) are an alternative treatment modality, with the FDA approving the following three different forms of ICIs: PD-1 inhibitors, PDL-1 inhibitors, and CTLA-4 inhibitors. In particular, atezolizumab is a PD-L1 inhibitor that is a monoclonal immunoglobulin antibody that blocks the interaction between PD-L1 and PD-1, thus allowing T cells to continue to kill cancer cells. In mUC, immunotherapy has seen survival rates that exceed platinum-based chemotherapy, with an average rate of 2 years in 2021, and can result in tumor remission.

While ICIs can be folded into the standard treatment regimen for mUC patients, only a small subset of patients see benefits from care in terms of tumor size. Therefore, it is important to understand the factors that are key in predicting ICI response. Common factors used in response include tumor mutational burden, PD-L1 expression, and other varying molecular markers. Additionally, clinical factors, such as patient demographics, tumor mutational burden (TMB), tumor neoantigen burden (TNB), and PD-L1 protein expression, are also used to determine the course of treatment .

One challenge for predicting ICI results per tumor shrinkage has been integrating clinical factors with gene expression, and the importance of each trait individually. This study will focus solely on the importance of molecular markers and investigates their impact on ICI response.

To measure this importance under realistic constraints, a L2- penalized logistic regression model will be employed using the IMvigor210 dataset. This approach seeks to isolate predictive signals to purely gene expression within the tumor, and test their efficacy in determining ICI response.

Methods and materials

Dataset selection

We conducted a retrospective exploratory modeling study using the RNA-seq data from IMvigor210 patient-level RNA-seq as the sole dataset. IMvigor210 is a study of PD-L1 blockade, where patients received 1200mg of atezolizumab every three weeks, and the tumor results were evaluated via Response Evaluation Criteria in Solid Tumors (RECIST v1.1) for standardization. Patients were grouped into binary values of 1 or 0, where 1 represents a positive response to ICI treatment, and 0 represents either no or a worsening response; both were assessed by the RECIST v1.1 scale.

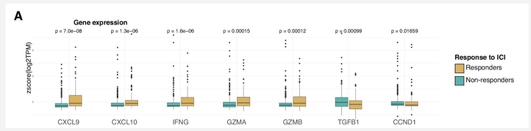
Dataset preparation

In order to normalize data, a log2(count+1) transformation was first conducted on the dataset in R. Reference the figure below for further information on the dataset setup.

$$\begin{bmatrix} patient_1 & \sum \text{molecular marker}_1 & \cdots & outcome_1 \\ patient_2 & \sum \text{molecular marker}_2 & \cdots & outcome_2 \\ patient_3 & \sum \text{molecular marker}_3 & \cdots & outcome_3 \\ \vdots & \vdots & \ddots & \vdots \\ patient_n & \sum \text{gene}_n & \cdots & outcome_n \end{bmatrix}$$

Predictive molecular markers

We focused on a compact preselected gene panel that reflects the main pathways implicated in ICI response. Per Boll et al., differential gene expression measured by RNAseq between patient responses to ICI includes the following genes: CXCL9, CXCL10, CXCL11, CXCL13, interferon-γ, granzyme A, granzyme B, and CCND1. These predictive molecular markers were used as the input feature for the model, with differential expression shown below. The reference differential between cancer and healthy patients is shown below:



Logistic regression model

A logistic regression model with L2 regularization and z-score balanced weights was used to predict each molecular marker's coefficient/weight. The model's base equation is as follow:

$$\text{Logit } P = \log\left(\frac{P}{1-P}\right) = \beta_0 + \sum_{n=1}^{7} k_n \beta_n \qquad \hat{y} = \begin{cases} 1 & \text{if } P \geq 0.5 \\ 0 & \text{if } P < 0.5 \end{cases}$$

Here, beta represents the seven preselected molecular markers, k represents the respective coefficients for each marker, and P delineates a range of probabilities for RECIST 1.1 between 0 and 1. These probabilities will be converted to binary values in the formula above. Note that y hat represents the modified predicted ICI response, and P represents the probability. .

This logistic regression model was created in Python using scikit-learn's LogisticRegression package with L2 penalty, libilinear solver, and balanced class weights. To ensure fairness and increase honesty in the accuracy measurements, a 10-fold stratified cross-validation approach was used; here, the dataset was split into 10 roughly equal folds with a maintained responder-to-non-responder ratio . For each fold, 9 folds were used to train the logistic regression model, and 1 fold was used to test the model's weights. This process was repeated such that each fold served once as the testing data, meaning every patient contributed to the model's accuracy report.

From each fold, the probability (P) value and a binary prediction value (y hat) were generated. Using this information, an accuracy report that measures mean ROC AUC, PR AUC, accuracy, and balanced accuracy was computed, alongside the respective standard deviations

Results

Immune Checkpoint Inhibitor

Using the 10-fold CV logistic regression model, the following metrics were reported below. The ROC AUC score suggests that the model distinguished responders from nonresponders approximately 73% of the time. The PR AUC is a subset of ROC AUC that focuses more on the prevalence of responders. The prevalence in the easierdata IMvigor210 phase II trial has a baseline PR-AUC of .114. The model achieved a PR AUC of .445, indicating it is four times more accurate than random chance at distinguishing responders from nonresponders per RECIST v1.1. Accuracy and balanced accuracy scores reported share the same mean with varying standard deviation values. The balanced accuracy ensures class balance in the unbalanced IMvigor210 dataset. The mean value of .683 represents the model correctly predicting the correct response 68.3% of the time. Reference the table below for extrapolated values.

Coefficient values

The coefficient values below correspond to their respective gene within the 10 CV fold logistic regression model. Here, CXCl9 is shown to be the strongest predictor, with a positive association. This aligns with previously reported values of CXCL9-driven T-cell activity being a known predictor for ICI therapeutics. Additionally, IFNG is also shown to be a strong predictor with a positive association. CXCL10 is less stable, but is still positively correlated. The remaining predictors, including GZMA, GZMB, TGFB1, and CCND1, have a negative association, all of which correspond to known tumor microbiology.

$$p = \frac{1}{1 + e^{-(-0.58 + 0.73\,\text{CXCL9} + 0.42\,\text{CXCL10} + 0.61\,\text{IFNG} - 0.44\,\text{GZMA} - 0.11\,\text{GZMB} - 0.36\,\text{TGFB1} - 0.29\,\text{CCND1})}}$$

Model's Accuracy	Numerical Value
ROC AUC	0.73
PR AUC	0.45
Accuracy	0.68
Balanced Accuracy	0.68

Marker/Intercept	Numerical Value
Intercept (θ_0)	-0.58
CXCL9	0.73
CXCL10	0.42
IFNG	0.61
GZMA	-0.44
GZMB	-0.11
TGFB1	-0.36
CCND1	-0.29

Conclusion

Here, we find that an L2 penalized logistic regression model that contains 7 pre-specified molecular markers can achieve moderate average discrimination in atezolizumab-treated mUC patients from the IMvigor210 dataset. Which implies that there is a link between molecular markers and ICI treatment in the IMvigor210 dataset. These findings are purely exploratory and highlight both the promise and the limitations of using small-panel molecular predictors in immunotherapy.

Discussion

In this exploratory analysis of atezolizumab-treated mUC patients in IMvigor210, we evaluated if a small gene panel on a realistically small sample size could yield a meaningful discrimination of ICI response through L2 penalized logistic regression. Here, we observed that using a prespecified panel of 7 molecular markers through a 10-fold stratified cross-validation was able to discriminate between responders and non-responders (ROC AUC = 0.734). Additionally, PR AUC was noted to be 0.45, which exceeds the baseline of .113. The model's weights or coefficients across the 7 different molecular markers mostly align with the current consensus in predicting their response to atezolizumab. The genes associated with both interferon signaling and cytotoxic immune activation show positive associations with response probabilities, while genes linked to immune suppression or cell-cycle progression show negative associations, which aligns with the current scientific understanding.

Several limitations are present in this model. Notably, the small sample size and low event-per-variable ratio, which contributes to the risk of overfitting and the large standard deviations presented between the CV folds. Additionally, we analyze only the reduced IMvigor210 dataset without an external validation dataset. A secondary validation trial is crucial for the generalizability of the model's weights. Thus, our findings are limited to the IMvigor210 dataset. While we performed an internal cross-validation to better measure the accuracy of an internal model, it does not substitute for the validation of an external dataset.

The future direction of this research would be to evaluate this molecular marker panel under an independent dataset to increase the honesty of the model's accuracy, as well as possibly incorporating clinical factors, whether that be in the same or a distinct regression model. Finally, more complex machine learning pipelines, such as LogitDA, can be employed and improved upon to create more generalizable results.

Review of Literature

Bolek, H., et al. (2024). Meta-analysis of platinum chemotherapy combinations with immunotherapy in metastatic urothelial carcinoma.The Oncologist, 29*(11), 999–1002.https://doi.org/10.1093/oncolo/oyae243

Stecca, C., et al. (2021). Metastatic urothelial cancer: A rapidly changing treatment landscape. *Therapeutic Advances in Medical Oncology, 13*, 17588359211047352. https://doi.org/10.1177/17588359211047352

Boll, L. M., et al. (2025). Predicting immunotherapy response of advanced bladder cancer through a meta-analysis of six independent cohorts. *Nature Communications, 16*(1), 1213. https://doi.org/10.1038/s41467-025-56462-0

Mariathasan, S., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature, 554*(7693), 544–548. https://doi.org/10.1038/nature25501

Piyawajanusorn, C., et al. (2025). Predicting atezolizumab response in metastatic urothelial carcinoma patients using machine learning on integrated tumour gene expression and clinical data. *NPJ Precision Oncology, 9*(1), 170. https://doi.org/10.1038/s41698-025-00969-8

Acknowledgements

Big thanks to Dr. Ayman Reffai and Mrs. Leventhal for all of your help with this project!