A Multifactorial Model of T Cell Expansion and Durable Clinical Benefit in Response to a PD-L1 Inhibitor

Mark DM Leiserson[1,2], Vasilis Syrgkanis[1], Amy Gilson[1], Miroslav Dudik[3], Samuel Funt[4], Alexandra Snyder[4,5], Lester Mackey[1]

[1]*Microsoft Research, New England*

[2]*University of Maryland, College Park, Department of Computer Science*

[3]*Microsoft Research, New York*

[4]*Memorial Sloan Kettering Cancer Center*

[5]*Adaptive Biotechnologies*

ABSTRACT

Background: Immunotherapies such as checkpoint inhibitors have become a major success in treating patients with late-stage cancers [1], yet the minority of patients benefit [2]. PD-L1 staining and mutation burden are the leading biomarkers associated with response, but each is an imperfect predictor. We seek to address this challenge by developing a multifactorial model for response to anti-PD-L1 therapy.

Methods: We train a model to predict fine-grained immune response in patients after treatment based on 36 clinical, tumor, and circulating molecular features of each patient collected prior to treatment. Our predicted immune response is then used to anticipate durable clinical benefit (DCB) of treatment. We analyze the bladder cancer patient data of Snyder et al. [3] using the elastic net high-dimensional regression procedure [4] and assess accuracy using leave-one-out cross-validation (LOOCV).

Results: In held-out patients, the elastic net procedure explains 80% of the variance in immune response as measured by the log number of T cell clones in the tumor that expand in the blood post-therapy. Moreover, if patients are triaged according to held-out predicted expansion, only 34% of non-DCB patients need be treated to ensure that 100% of DCB patients are treated. In contrast, using PD-L1 staining or mutation load alone, one must treat at least 75% of non-DCB patients to ensure that all DCB patients receive treatment. The final elastic net model fit to the entire patient cohort retains 20 of the 36 input features, a mix of clinical, tumor, and circulating patient attributes. In addition, each class of features contributes significantly to the estimated accuracy of the procedure: when tumor features are excluded as inputs, the variance explained in held-out patients drops to 26%; when circulating features are excluded, the explained variance drops to 13%; and when clinical features are excluded, the explained variance drops to 0%.

Conclusions: In a multifactorial model, tumor, circulating, and clinical features all contribute significantly to the accurate prediction of fine-grained immune response. Moreover, the accurate prediction of this immune response may improve our ability to anticipate clinical benefit.

Introduction

Immunotherapies such as checkpoint inhibitors have become a major success in treating patients with late-stage cancers [1], in many cases leading to durable responses [2]. The basis for this success is that, in some cases, the somatic mutations present in cancer cells allow the immune system's T cells to distinguish cancer from normal cells, in part because mutations may lead to the presentation of *neoantigens* on the cancer cell surface by the major histocompatibility complex. However, many cancers develop mechanisms for suppressing the immune system, e.g., by expressing *checkpoint molecules*. The promise of checkpoint inhibitors is predicated on counteracting checkpoint molecules to unleash the immune system to selectively kill cancer cells.

Despite checkpoint inhibitors' unprecedented successes, there is an urgent need to improve prediction of patient response to checkpoint inhibitor immunotherapy. Response rates vary across patients, and known biomarkers for response such as high mutation load [5,6] are not predictive for every patient [7]. Thus, predicting response is critical for identifying patients who are likely or unlikely to benefit, anticipating adverse responses to treatment [8], and accelerating the development of new treatments. Further, effective models for predicting response may point to molecular features that can be measured and monitored through non-invasive treatments.

A key challenge for predicting response is modeling features of the immune system and cancer simultaneously. Recently, clinicians have begun to collect a wealth of molecular tumor and immune system data before and during immunotherapy, but researchers have yet to model how molecular and clinical features interact to affect response.

To address this challenge, we develop a multifactorial model for response to checkpoint inhibitors. Our approach uses the elastic net [4] -- a machine learning method for regression that automatically selects informative features from the data -- and models clinical, tumor, and immune system features simultaneously. We applied our model to data of Snyder et al. [3], who measured mutations and gene expression in the tumor and T cell receptor (TCR) sequences in the tumor and blood in 29 urothelial cancers treated with anti-PD-L1. Rather than model the clinical response of each patient directly, we modeled the response of each patient's immune system and used the predicted immune responses to stratify patients based on expected clinical benefit. By modeling the immune response, we have the advantage of predicting fine-grained, molecular measurements that are predictive of clinical response.

Our model predicts immune response with high accuracy, and that predicted immune response is associated with durable clinical benefit (DCB). In held-out patients, our model explains 80% of the variance in the immune response as measured by the log number of T cell clones in the tumor that expand in the blood post-therapy. The magnitude of the predicted expansion is associated with DCB, as 100% of held-out patients with DCB have predicted scores above the median score of held-out patients without DCB. Notably, mutation load alone does not demonstrate significant association with DCB. We also evaluated the importance of the tumor, immune system, and clinical features to our model. Our model requires all three feature classes

to make accurate predictions and can explain at most 26% of the held-out patient variance when any one of the classes is removed. Taken together, these results show that models for immune response may be useful for predicting clinical response to immunotherapy and that non-invasive TCR sequencing in the blood may be an effective way to monitor patient response.

Methods

All of our analyses were conducted in Python 3, and Jupyter notebooks that replicate our experiments will be made publicly available.

Patient Data

We used the patient data collected by Snyder et al. [3]. For the data collection details and the Institutional Review Board approval see [3].

Multifactorial modeling of clonal expansion

In our first analysis, we develop a predictive model of the log number of tumor-infiltrating lymphocyte (TIL) clones that expanded in the blood three weeks after each patient's initial immunotherapy treatment. We chose to model TIL clone expansion, as it is a finer-grained and more immediate measurement of patient response than standard clinical, and, yet still exhibits positive association with durable clinical benefit [3]. Our analysis is based on the 21 patients in which this clonal expansion was recorded.

Nineteen features of each patient collected prior to treatment were used as inputs to the predictive model: Age, Albumin < 4, Baseline neutrophil to lymphocyte ratio, Time since last chemotherapy, missense snv count, expressed missense snv count, neoantigen count, expressed neoantigen count, T-cell fraction, Clonality, Diversity, Productive Unique TCRs (cnt), T-cell fraction tumor, Clonality tumor, Diversity tumor, Top Clone Freq(%), Number of chemo regimens total, Prior BCG, and 5-factor score. For the one patient who did not have a 5-factor score recorded, we substituted the patient's 2-factor score. We encoded Prior BCG and Albumin < 4 as binary features, with values in {0,1}. For each of the remaining input variables, we provided both the raw feature value, x, and a log(1+x) transform of the feature value as inputs to our models to capture nonlinear relationships between inputs and the target.

We trained two machine learning models to predict log clonal expansion from our input features: the *elastic net [4]*, a high-dimensional linear regression procedure designed to reduce overfitting and be robust to irrelevant features, and *random forests* [9], a highly-nonlinear regression procedure based on averaging the predictions of many randomly constructed decision trees. We fit the elastic net using the Python scikit-learn [10] ElasticNetCV function with feature normalization, candidate L_1 ratio hyperparameters [.1, .5, .7, .9, .95, .99], hyperparameters selected using leave-one-out cross-validation, and a maximum number of iterations of 100000.

We fit the random forest model using the Python scikit-learn RandomForestRegressor with 1000 decision trees. Prior to fitting each regression model, missing input feature values were imputed using the median non-missing feature value in the training set.

Leave-one-out error analysis of expansion predictions

To estimate the effectiveness of our learning pipelines at predicting the clonal expansion of previously unseen patients, we conducted a leave-one-out cross-validation (LOOCV) analysis. Specifically, for each of the 21 patients in turn, we withheld that individual's data from the training set, fit each of our learning pipelines on the remaining 20 patients, and formed a prediction of the held-out patient's log clonal expansion using each of the learned models. We then compared each learned model's predictions with the observed log clonal expansion for the held-out patient and computed the squared error. By computing the average of this held-out squared error across all patients, we obtain an unbiased estimate of each learning pipeline's error in predicting previously unseen patient expansion. Saria et al. [11] use a similar leave-one-out analysis to assess predictive models of pre-term infant illness. We moreover compute a measure of variance explained in held-out patients by computing one minus the ratio of LOOCV mean squared error (MSE) to empirical variance of log clonal expansion. Finally, following [12], we conduct a permutation test of whether the LOOCV error is significantly larger when patient immune responses are permuted in our cohort uniformly at random.

Feature importance

To assess the degree to which different classes of features contribute to the predictive accuracy of our learning pipelines, we assigned each input feature to a category ('Clinical', 'Circulating', or 'Tumor') and, for each category, repeated our LOOCV analysis with features from that category excluded from the model.

Implications for durable clinical benefit

Snyder et al. [3] previously demonstrated a positive association between the number of TIL clones that expanded in the blood two weeks after treatment and the durable clinical benefit of cancer immunotherapy. Here, a treatment is said to have durable clinical benefit (DCB) for a patient if the patient experiences progression free survival for at least six months after treatment. To assess whether our learning pipelines are also predictive of durable clinical benefit for previously unseen patients, we compare the distribution of held-out expansion predictions for those patients who did and did not experience DCB.

Results

Leave-one-out error analysis of expansion predictions

The multifactorial elastic net model explained 80% of variance in held-out patients with a LOOCV mean squared error of 0.173. The variance of log number of TIL clones expanded in our dataset was 0.88, and this number is equivalent to the mean-squared error achieved by the baseline of predicting the mean log clonal expansion for all patients. A scatter plot of elastic-net-predicted versus ground-truth log expansion for each held-out patient can be found in Figure 1. Figure 2 displays the results of our permutation test: after conducting 1000 random permutations of patient responses, we obtained a Monte Carlo p-value of less than 0.001. The random forest model performed far worse with a LOOCV mean-squared error (0.886) that exceeded the baseline MSE, indicating severe overfitting. For this reason, we focus on the elastic net model in the remainder of our analyses.

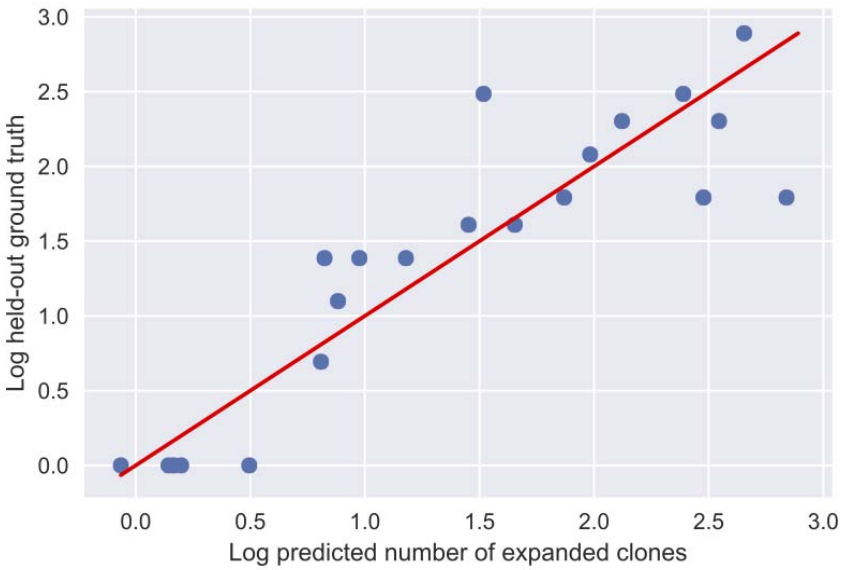

Figure 1: Predicted log TIL expansion versus ground-truth log TIL expansion for patients held out using LOOCV. Predictions are formed using the elastic net.

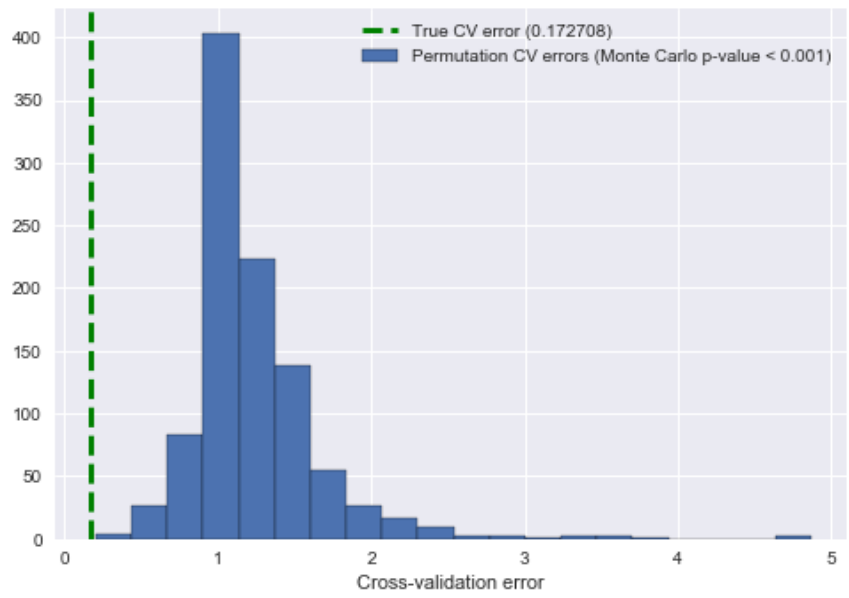

Figure 2: Distribution of LOOCV error when patient responses are permuted uniformly at random 1000 times. The overlaid dotted line displays the LOOCV error obtained on the original dataset.

Feature importance

The elastic net procedure automatically performs feature selection by setting the coefficients of some input variables to zero. The final elastic net model fit to the entire training set retains 20 of the 36 input features, a mix of clinical, tumor, and circulating patient attributes (see Table 1). When we repeat our LOOCV error analysis using only clinical and circulating features (that is, excluding all tumor features), the variance explained in held-out patients drops from 80% to 26%. When we repeat the LOOCV error analysis using only clinical and tumor features (excluding all circulating features), the explained variance drops to 13%. Finally, when we repeat the LOOCV error analysis using only circulating and tumor features (excluding all clinical features), the explained variance drops to 0%; that is, the model performs worse than the baseline in this scenario. These findings indicate that clinical, tumor, and circulating features are all contributing significantly to the predictive accuracy of the elastic net learning pipeline.

Table 1: Learned elastic net coefficients and feature types

Feature name	**Elastic net coefficient**	**Feature type**
expressed_neoantigen_count	0.682021	Tumor
log_neoantigen_count	-0.647034	Tumor
Top Clone Freq(%)	0.548716	Circulating
Time since last chemotherapy	-0.539601	Clinical

Age	-0.52979	Clinical
log_Productive Unique TCRs (cnt)	0.396964	Circulating
Albumin < 4	-0.356917	Clinical
T-cell fraction	-0.322786	Circulating
T-cell fraction_tumor	0.295661	Tumor
log_factor score	0.223579	Clinical
log_Number of chemo regimens total	-0.212176	Clinical
log_Diversity_tumor	-0.18734	Tumor
Prior BCG	0.147289	Clinical
Diversity_tumor	-0.125397	Tumor
neoantigen_count	-0.100804	Tumor
factor score	0.065281	Clinical
log_Top Clone Freq(%)	0.062521	Circulating
missense_snv_count	-0.054105	Tumor
log_Baseline neutrophil to lymphocyte ratio	0.027796	Clinical
Clonality_tumor	0.011165	Tumor
Clonality	0	Circulating
Baseline neutrophil to lymphocyte ratio	0	Clinical
expressed_missense_snv_count	0	Tumor
log_Clonality_tumor	0	Tumor
log_T-cell fraction_tumor	0	Tumor
log_expressed_neoantigen_count	0	Tumor
Diversity	0	Circulating
Productive Unique TCRs	0	Circulating

(cnt)		
log_T-cell fraction	0	Circulating
log_Clonality	0	Circulating
Number of chemo regimens total	0	Clinical
log_Age	0	Clinical
log_Time since last chemotherapy	0	Clinical
log_missense_snv_count	0	Tumor
log_expressed_missense_snv_count	0	Tumor
log_Diversity	0	Circulating

Implications for durable clinical benefit

Figure 3 compares the distributions of the predicted number of expanded TIL clones for held-out patients who did and did not experience durable clinical benefit (DCB). This display should be contrasted with the DCB-stratified distributions of standard biomarkers like missense single-nucleotide variant (SNV) count, expressed neoantigen count, and PD-L1 staining (Figures 4, 5, and 6 respectively). Notably, 100% of patients who experienced durable clinical benefit have predicted expansion scores above the sixty-sixth percentile prediction for patients without durable clinical benefit. This indicates that, if patients are triaged according to predicted expansion, only 34% of non-DCB patients need be treated to ensure that 100% of DCB patients are treated. In contrast, using PD-L1 staining alone, one must treat at least 75% of non-DCB patients to ensure that all DCB patients receive treatment. Using missense SNV count or expressed neoantigen count alone, one must treat at least 87.5% of non-DCB patients to ensure that all DCB patients are treated. Each of these biomarkers is known to be associated with DCB in cancer immunotherapy patients [13]. However, in this dataset, each biomarker alone is a relatively poor discriminator of DCB and non-DCB patients. Nevertheless, the elastic net learning pipeline is able to integrate missense SNV count, expressed neoantigen count, and several other patient features into a composite biomarker with greater DCB discriminating ability.

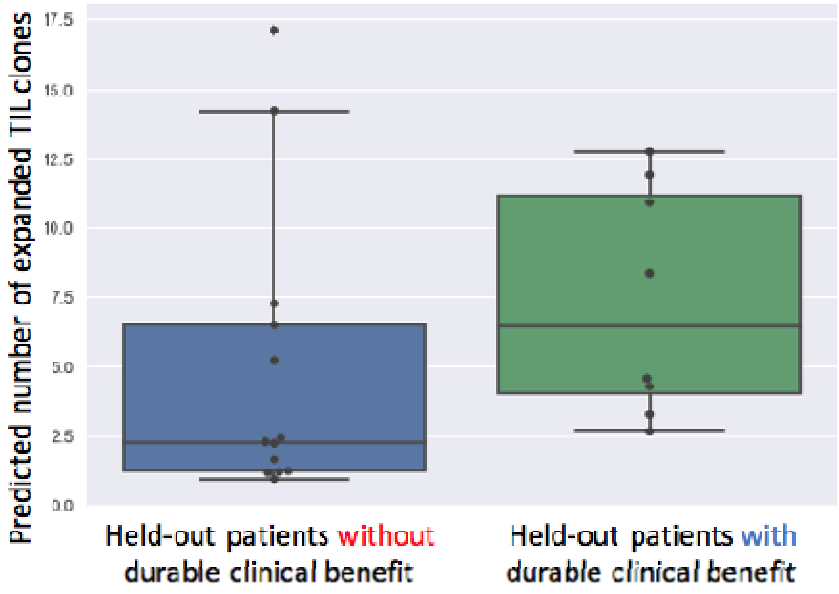

Figure 3: Distributions of predicted number of expanded TIL clones in patients with and without clinical benefit.

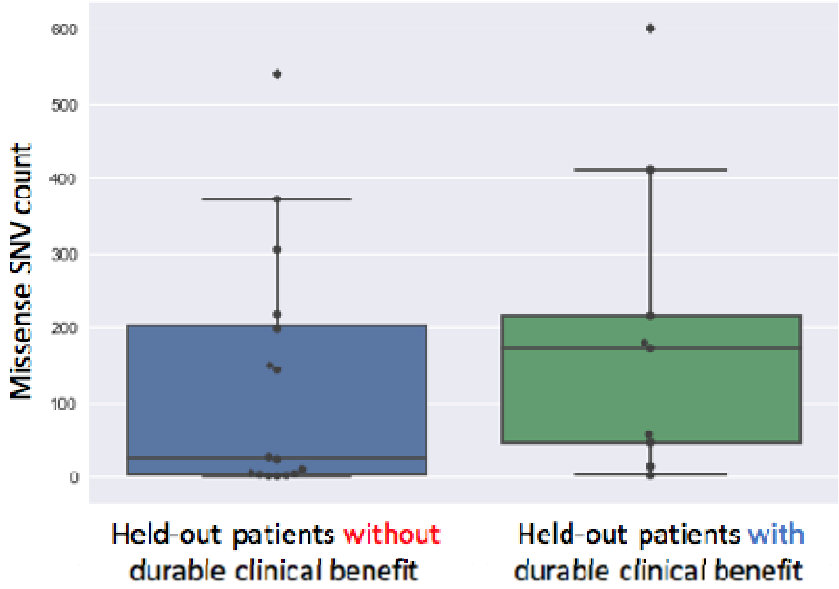

Figure 4: Distributions of missense SNV count in patients with and without clinical benefit.

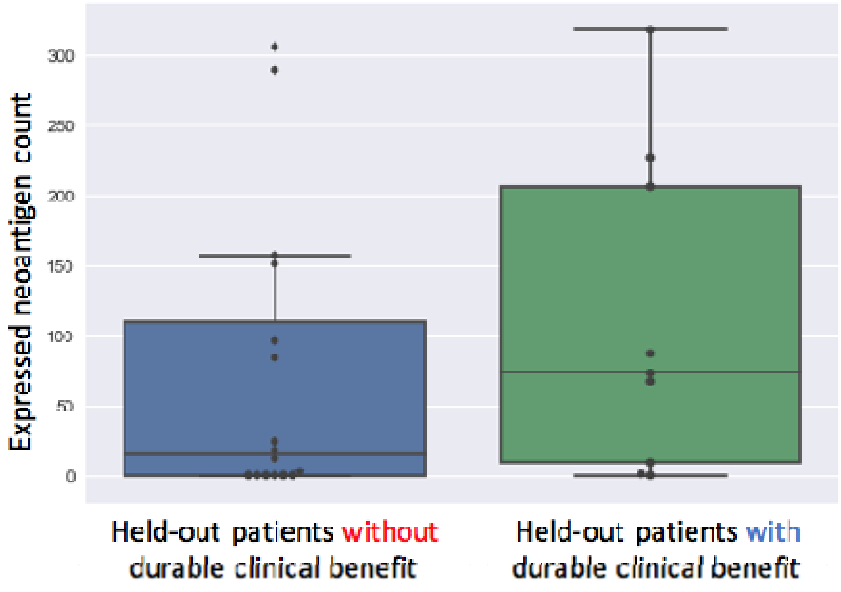

Figure 5: Distributions of expressed neoantigen count in patients with and without clinical benefit.

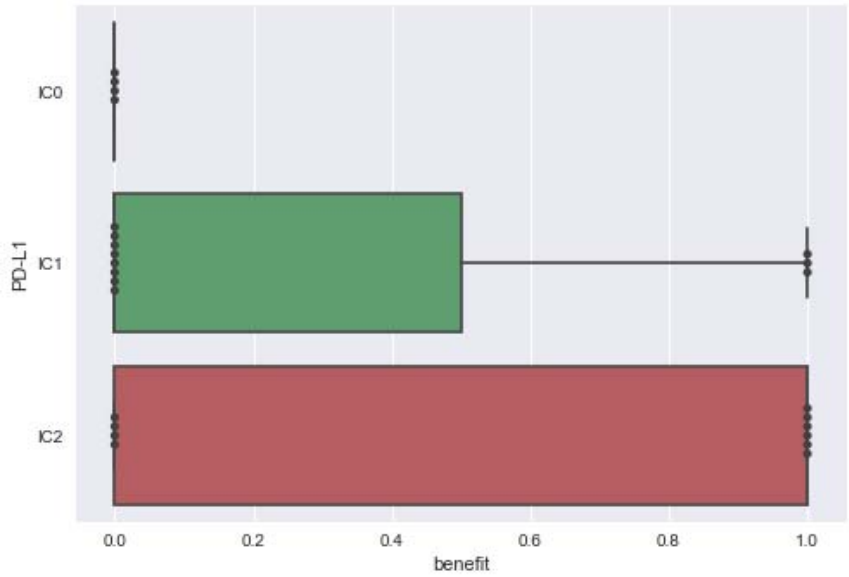

Figure 6: Distributions of PD-L1 in patients with and without clinical benefit.

Discussion

We have introduced a multifactorial model for predicting response to checkpoint inhibitor immunotherapy. Our model integrates tumor, clinical, and immune features to predict a measure of immune response. Notably, we train our model to predict a fine-grained intermediate measure of immune response which is associated with the coarser-grained clinical response.

We demonstrate and evaluate our model on a dataset of urothelial cancers from Snyder et al. [3]. We find that our model can predict the number of tumor infiltrating T cell clones that expand in the blood post-therapy with high estimated accuracy. In addition, if the patients in our cohort are triaged according to held-out predicted expansion, only 34% of non-DCB patients need be treated to ensure that 100% of DCB patients are treated. Moreover, we find that our model achieves the highest LOOCV accuracy when tumor, circulating, and clinical features are all included, demonstrating that they provide complementary information. We next intend to validate our model fully out-of-sample, on a previously unseen cohort of bladder patients.

We anticipate integrative models of tumor, immune system, and clinical features such as those introduced here will be increasingly useful as TCR sequencing becomes routine in the clinical practice. Non-invasive measurement of immune system activity in the blood has the potential to transform many areas of cancer care, including early detection, precision oncology, and monitoring response. Predictive models of the peripheral immune system, and predictive models of clinical response from peripheral immune system features, will be crucial to realizing these goals.

References

1. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012;12: 252–264. doi:10.1038/nrc3239

2. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. Cell. 2017;168: 707–723. doi:10.1016/j.cell.2017.01.017

3. Snyder A, Nathanson T, Funt SA, Ahuja A, Buros Novik J, Hellmann MD, et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. PLoS Med. 2017;14: e1002309. doi:10.1371/journal.pmed.1002309

4. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. Blackwell Publishing Ltd; 2005;67: 301–320. doi:10.1111/j.1467-9868.2005.00503.x

5. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371: 2189–2199. doi:10.1056/NEJMoa1406498

6. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science. 2015;350: 207–211. doi:10.1126/science.aad0095

7. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science. 2017;357: 409–413. doi:10.1126/science.aan6733

8. Michot JM, Bigenwald C, Champiat S, Collins M, Carbonnel F, Postel-Vinay S, et al.

Immune-related adverse events with immune checkpoint blockade: a comprehensive review. Eur J Cancer. 2016;54: 139–148. doi:10.1016/j.ejca.2015.11.016

9. Breiman L. Random Forests. Mach Learn. Kluwer Academic Publishers; 2001;45: 5–32. doi:10.1023/A:1010933404324

10. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12: 2825–2830. Available: http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

11. Saria S, Rajani AK, Gould J, Koller D, Penn AA. Integration of early physiological responses predicts later illness severity in preterm infants. Sci Transl Med. 2010;2: 48ra65. doi:10.1126/scitranslmed.3001304

12. Ojala M, Garriga GC. Permutation Tests for Studying Classifier Performance. J Mach Learn Res. 2010;11: 1833–1863. Available: http://www.jmlr.org/papers/volume11/ojala10a/ojala10a.pdf

13. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol. 2016;17: e542–e551. doi:10.1016/S1470-2045(16)30406-5