Poster No. 211

Understanding contribution and independence of multiple biomarkers for predicting response to atezolizumab

Parantu Shah[1], Jan Kaczmarczyk[2], Alexander Myronov[2], Katarzyna Gruba[2], Piotr Stępniak[2], Alex Rolfe[1]

[1]EMD Serono Research & Development Inc, Billerica MA; [2]Ardigen S.A., Kraków, Poland

BACKGROUND

- Biomarkers for accurately predicting response to anti–PD-L1 immunotherapy from baseline samples are lacking due to biological and technical reasons[1]
- Response to anti–PD-1 and anti–PD-L1 agents may be dependent on multiple factors, including tumor characteristics, tumor microenvironment, and the status of the host immune system. Different biomarkers are expected to measure different aspects of these interactions[2]
- For example, PD-L1 expression by immunohistochemistry (IHC) measures the immune suppressive environment; tumor mutation burden (TMB) and neo-antigen burden (NB) measure tumor foreignness; immune phenotype may provide estimates of immune cell infiltration into the tumor
- Several features are associated with response but do not provide good classification performance[1,2]
- No published studies have extensively compared and integrated multiple biomarkers of response. Moreover, the costs and complexities added by generation and integration of multiple biomarkers are not clear

Hypothesis

- Combined information from multiple biomarkers will improve prediction of response to anti–PD-L1 therapies

OBJECTIVES

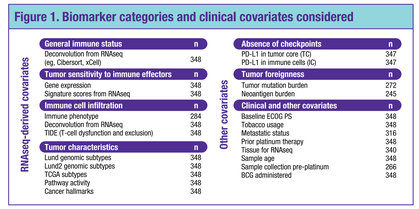
- Understand the predictive value of biomarkers in a univariate manner and rank different biomarkers to allow prioritization in data generation
- Generate sparse signatures of response from ultra–high-dimensional gene expression and gene signatures from RNAseq profiles, understand their predictive value, and identify signatures that may help to understand the tumor biology
- Quantitate improvements in response prediction using systematic combinations of biomarkers
- Understand biomarkers suitable for distinguishing patients with stable disease (SD) from patients with progressive disease (PD)
- Understand whether added complexity, due to integration of multiple biomarkers, is justified
- Provide guidelines for future biomarker studies for checkpoint inhibitor therapies

METHODS

Data set and procedure for model building

- We analyzed and integrated biomarkers measured in pretreatment tumor samples from IMvigor210,[3] a single-arm phase 2 trial investigating the clinical activity of PD-L1 blockade with atezolizumab in metastatic urothelial cancer (**Figure 1**)[4]
- Additional biomarkers were derived from RNAseq data by scoring different signatures (eg, pathways and cancer hallmarks) and immune content deconvolution (eg, ESTIMATE, xCell, CIBERSORT, TIDE, Immunophenoscore, IMPRES)
- Patients with complete or partial response (CR/PR) were considered responders; patients with SD or PD were considered nonresponders
- Evaluation of predictive and prognostic properties of biomarkers (both single and multivariate) was performed throughout using repeated (5×) 5-fold cross-validation. Prediction metrics are reported from the test set in each case (**Table 1**)
- Area under the ROC curve (AUC) statistic was used to report predictive performance

Figure 1. Biomarker categories and clinical covariates considered



BCG, Bacille Calmette Guerin; **ECOG PS**, Eastern Cooperative Oncology Group performance status; **TCGA**, The Cancer Genome Atlas. Biomarkers categorized according to Blank et al. Science 2016[2]

Procedure for building penalized classification models from RNAseq data (gene and signature level)

- To remove gene signatures with prognostic value, we used survival association of genes and gene signatures from bladder cancer, lung cancer, and melanoma cohorts from TCGA
- We compared performance of regularized logistic regression, support vector machine (SVM), and random forest for binary classification and chose regularized logistic regression for sparse model building as being the best
- For models from gene expression and signature scores, 4 different methods of feature selection were applied (**Table 2**) before building penalized classification models
- For models trained on RNAseq gene expressions and signatures, there were 238-239 patients in each training set and 59-60 in each test set

Unpenalized models were built for integrating biomarkers by considering gene expression as a single covariate

- For the models combining various types of information (eg, NB with RNAseq), the size of the considered sets of patients were smaller as patients with missing values were dropped
- Logistic regression, extreme gradient boosting (XGB), and extreme trees (XT) were used to build models that integrated multiple biomarkers

RESULTS

Biomarker performance in stratifying response

- NB (and highly correlated TMB; r=0.85) was the best performing biomarker of response (AUC of 0.77) among those analyzed (**Table 1**)
- Models derived from gene expression signatures performed the second best (AUC 0.71; **Tables 1** and **2**)
- None of the examined biomarkers could distinguish patients with SD from patients with PD with high AUC (**Table 1**)
- Metastatic disease status had moderate predictive power (AUC 0.65) for stratifying response

Performance of penalized classification models derived from gene expression and gene signatures

- While inferior to TMB/NB in response prediction (**Table 2**), these models did not provide a better performance than an AUC of approximately 0.7 obtained for single genes or signatures
- Metastatic disease status models performed similarly to those models built from gene expression data
- Certain gene signatures were frequently present during the model building process, highlighting their biological importance (**Figure 2**) and independent contributions to response

Comparison with PD-L1 IHC

- 4 biomarkers stratified response better than PD-L1 IC expression (**Table 1**)
- PD-L1 IC had better AUC to stratify responders vs nonresponders than to stratify SD from PD
- Models derived from both gene signatures and gene expression outperformed PD-L1 IHC (**Table 1** and **2**)

Table 1. NB/TMB have the best power to predict response groups, but no biomarker satisfactorily distinguished patients with SD from patients with PD

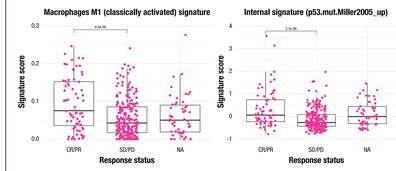
Biomarker	Responder vs Nonresponder		SD vs PD	
	n	AUC	n	AUC
Neoantigen burden	216	0.767	163	0.495
Tumor mutation burden	234	0.727	173	0.493
Gene signatures	298	0.712	230	0.583
Metastatic disease status	271	0.648	219	0.576
Gene expression	298	0.646	230	0.474
PD-L1 immune cell level	297	0.616	229	0.543
Immune cell status at enrollment	298	0.605	230	0.477
Lund2 subtype	298	0.601	230	0.561
Lund subtype	298	0.583	230	0.534
Tissue	292	0.570	225	0.517
Baseline ECOG PS	298	0.580	230	0.544
Tobacco	298	0.558	230	0.452
Pre-platinum	227	0.549	180	0.456
Immune phenotype	244	0.524	183	0.552
TCGA subtype	298	0.514	230	0.426
Sex	298	0.537	230	0.485
PD-L1 tumor cell level	297	0.494	229	0.513
Intravesical BSG	298	0.467	230	0.502

Table 2. AUC of classification models built using gene signatures (left panel) and genes (right panel) with different feature selection methods

Gene Signature Models		Gene Expression Models	
Method	AUC	Method	AUC
RECIST SVM	0.712	LPC	0.66
Nonprognostic	0.687	Univariate selection	0.65
LPC	0.665	WGCNA	0.63
Univariate	0.663	All genes	0.603

LPC, lassoed principal component; **WGCNA**: weighted gene coexpression network analysis

Figure 2. Score distribution for top 2 signatures frequently present in regularized classification models from gene signatures



Rationale for and outcome of biomarker integration experiments

- Examination of individual markers and pairs suggested that while single markers were enriched for response groups, combining information from multiple biomarkers may help stratify response better (**Figure 3**)
- For example, a high threshold of TMB enriched for responders (**Figure 3A**)
- However, the genomically unstable (GU) subtype showed responders at every IC level (**Figure 3B**), and a combination of lymph node-only metastasis status with IC2-positive status enriched for responders better than individual markers (**Figure 3C**), suggesting increased information to stratify response
- Unfortunately, systematic combinations of biomarkers with logistic regression (LogReg) and tree-based models (XGB and XT) showed very little (AUC 0.81) increase in predictive power (**Table 3**)

Figure 3. Combinations of biomarkers may better stratify response

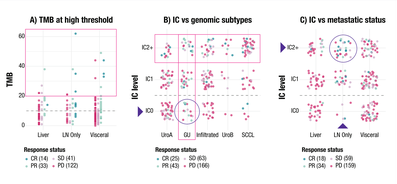


Table 3. AUC for combination of biomarkers

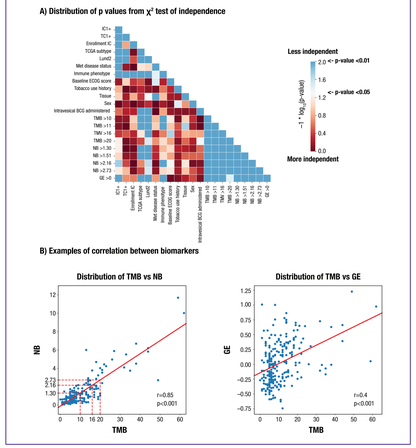
Input Features	LogReg		XGB		XT	
	ROC	HR	ROC	HR	ROC	HR
TMB+NB+ECOG+GE	0.801	1.547	0.793	1.551	**0.815**	1.9
log_TMB+log_NB+ECOG+GE	**0.812**	1.82			0.811	1.867
log_TMB+log_NB+ECOG	**0.804**	1.811			0.796	1.793
TMB+NB+ECOG	0.794	1.742	0.799	1.347	**0.803**	1.889
NB+ECOG+GE	0.797	1.83	0.778	1.627	**0.798**	1.82
NB+ECOG	0.783	1.829	0.785	1.363	**0.796**	1.807
log_TMB+log_NB+GE	0.776	1.33			**0.785**	1.485
TMB+ECOG+GE	**0.782**	1.715	0.72	1.413	0.765	1.856
TMB+NB+GE	0.776	1.397	0.759	1.39	**0.78**	1.552
log_NB+GE	0.773	1.358			**0.778**	1.541
log_TMB+log_NB	**0.776**	1.506			0.768	1.483
TMB+NB+PD-L1(IHC)+GE	0.765	1.364	0.758	1.403	**0.775**	1.586
NB+GE	0.774	1.562	0.76	1.472	**0.775**	1.565
TMB+NB+Lund2_GU+GE	**0.774**	1.332	0.759	1.383	0.767	1.512
log_NB	**0.773**	1.564			0.761	1.428
NB	**0.773**	1.751	0.756	1.382	0.77	1.467
TMB+NB	**0.772**	1.455	0.773	1.272	0.769	1.49
TMB+Lund2_GU	**0.77**	1.195	0.742	1.265	0.764	1.491
TMB+NB+Lund2_GU	**0.769**	1.293	**0.77**	1.396	0.765	1.48
TMB+Lund2_GU+GE	**0.769**	1.532	0.734	1.329	0.767	1.567
log_TMB+GE	**0.764**	1.23			0.762	1.444
TMB+GE	**0.764**	1.356	0.728	1.273	0.761	1.568

GE, gene expression

(In)dependence of biomarkers in predicting response

- The χ^2 tests for independence between pairs of biomarkers suggest that the lack of improvement may be due to presence of correlation between examined biomarkers (**Figure 4A**)
- For example, TMB and NB were highly correlated (**Figure 4B**; left panel). Moderate correlation was found between TMB/NB and probabilities for response from gene signature models (**Figure 4B**; right panel). Similarly, PD-L1 IHC levels (TC and IC) were also not completely independent, and genomic subtypes were not independent
- This may explain the limited gain after combining some biomarkers
- Interestingly, at the TMB threshold of ≥10, the contribution of TMB appaered to be independent

Figure 4. Biomarkers of interest may not provide independent information for predictive models



LIMITATIONS OF THE STUDY

- Additional predictive biomarkers could be derived from the raw sequencing read information accompanying the study. For example:
  - Genome-wide information on mutation and changes in chromosomal arm and focal changes in copy number, which have been shown to be better predictors of response than TMB/NB only
  - Information about changes in HLA locus (genotype, expression, and mutation) that need specialized software to show association with response and survival
  - Repertoire of B- and T-cell receptors
  - Microsatellite instability
- Unfortunately, we could not add the additional biomarkers due to restrictions on raw data access. Hopefully, future studies will include information on and impact of these biomarkers in order to predict response of biomarker combinations

CONCLUSIONS

- NB/TMB had the best AUC for response prediction and could be the most attractive biomarker to measure in ongoing monotherapy and combination immunotherapy trials. This is in line with current knowledge from the literature
- Combinations of multiple biomarkers using 3 different methods did not improve response prediction significantly compared with NB
- The lack of improvement may be attributed to lack of independence of biomarkers and reduced number of samples
- The examined biomarkers did not distinguish PD patients from SD patients, suggesting natural grouping of patients with SD and patients with PD as nonresponders
- Published algorithms for prediction of response (IMPRES, Immunophenoscore, and TIDE) did not provide good stratification
- Models from gene expression signatures were the second best predictors of response. Gene expression also provides additional information that may improve survival prediction
- Therefore, generation of paired WEX/RNAseq data could be attractive for baseline response prediction for checkpoint inhibitors
- The search for an ideal immunotherapy biomarker is not over. A single biomarker may not be sufficient to predict response to immunotherapies. Integration of independent biomarkers based on biological mechanisms is required

REFERENCES

1. Havel JJ, et al. *Nat Rev Cancer.* 2019;19(3):133-50.
2. Blank CU, et al. *Science.* 2016;352(6286):658-60.
3. Balar AV, et al. *Lancet.* 2017;389(10064):67-76.
4. Mariathasan S, et al. *Nature.* 2018;554(7693):544-48.

ACKNOWLEDGMENTS

The authors thank the patients and their families who were part of the IMvigor210 trial and Roche/Genentech for making the data publicly available for analysis. Medical writing support was provided by ClinicalThinking Inc, Hamilton, NJ. The study was funded and directed by EMD Serono R&D Inc, Billerica, MA, (EMD Serono is a business of Merck Healthcare KGaA, Darmstadt, Germany) and carried out at Ardigen S.A., Kraków, Poland. All relevant disclosures will be published with the abstract on the ASCO meeting library website (https://meetinglibrary.asco.org/).

Correspondence: Parantu Shah, parantu.shah@emdserono.com

Abstract No. 2567. Presented at the ASCO 2019 Annual Meeting, May 31–June 4, 2019; Chicago, IL, USA