Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Metastatic pattern accounts for the apparent predictive value of immune gene signatures in urothelial carcinoma treated with checkpoint inhibitors

Karen Baghdasaryan

karen.baghdasaryan1@edu.ysu.am

Yerevan State University

Research Article

Keywords:

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11621980/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Metastatic pattern accounts for the apparent predictive value of immune gene signatures in urothelial carcinoma treated with checkpoint inhibitors

Karen Baghdasaryan[1*]

[1*]Yerevan State University, Alex Manoogian 1, Yerevan, 0025, Armenia.

Corresponding author(s). E-mail(s): karen.baghdasaryan1@edu.ysu.am;

Abstract

Purpose: Transcriptomic immune signatures are widely proposed as predictors of response to immune checkpoint inhibitors in metastatic urothelial carcinoma, but are almost always evaluated against chance rather than against the markers a clinician already has. We asked what such a signature adds to information that is already measured.

Methods: We analyzed IMvigor210 ($\boldsymbol{n = 348}$ with pre-treatment tumor RNA sequencing; 298 evaluable for response). Six published immune signatures and two pre-specified control signatures were compared with two baselines: routine clinical variables (metastatic pattern, ECOG performance status, prior platinum), and those plus PD-L1 immunohistochemistry and tumor mutational burden. Discrimination was estimated as out-of-fold area under the curve over 50 repeats of stratified five-fold cross-validation; incremental value was tested by likelihood-ratio test. Overall survival was analyzed by Cox regression with Harrell's C-index.

Results: The clinical baseline reached an AUC of 0.697 (95% CI 0.619–0.767) and the extended baseline 0.778 (0.708–0.844). tumor mutational burden alone reached 0.720 (0.639–0.796) and exceeded every immune signature in a paired comparison ($\boldsymbol{\Delta}$AUC $\boldsymbol{-0.122}$ to $\boldsymbol{-0.189}$). No immune signature added measurable information to either baseline ($\boldsymbol{\Delta}$AUC $\leq$ $\boldsymbol{+0.003}$; likelihood-ratio $\boldsymbol{p = 0.06}$–$\boldsymbol{0.99}$), whereas a proliferation control did. Metastatic pattern explained 6.8–11.0% of signature variance and accounted for nearly all of the attenuation seen on adjustment.

Conclusion: Established immune signatures carried no detectable incremental value over markers already measured in this cohort.

Keywords: Urothelial carcinoma, Immune checkpoint inhibitors, Gene expression profiling, tumor mutational burden, Biomarkers

1 Introduction

Transcriptomic measures of the tumor immune microenvironment — CD8 T-effector scores, the T-cell-inflamed gene expression profile (GEP-18), interferon-γ and cytolytic activity scores — are widely proposed as predictors of response to immune checkpoint inhibitors (ICIs) [1]. Their reported discrimination is modest, and it is almost always quantified against chance rather than against the markers already in use. Prognostic-model methodology holds that a candidate marker should be judged by what it adds to an established baseline, not by whether it beats a coin toss [2]. Demonstrating that requires three things an association analysis does not: a baseline built from variables a clinician already has, an estimate of discrimination not fitted on the data it is scored on, and a formal comparison of the nested models. A marker can be strongly associated with outcome and still fail all three, if its information is already in the baseline.

In metastatic urothelial carcinoma (mUC) that baseline is well characterised. Performance status and the pattern of metastatic spread are established determinants of ICI outcome [3, 4], and nodal-confined disease carried a markedly higher response rate than visceral disease in the registrational cohorts [5, 6]. Two molecular markers, PD-L1 immunohistochemistry and tumor mutational burden (TMB), are already measured in this setting [7].

Yet the molecular analyses of those same cohorts largely omit the clinical variables, benchmark new signatures against other molecular markers, or report combined models without isolating what the molecular component contributes [7–10]; we return to these studies in the Discussion. We therefore asked the converse of the usual question: given what is already measured, does an immune signature add anything? We addressed it in IMvigor210, the cohort in which several of these signatures were developed or validated, using pre-specified control signatures to establish that the procedure can detect added value when it exists.

2 Materials and methods

2.1 Cohorts

IMvigor210 (atezolizumab, platinum-treated and cisplatin-ineligible mUC) [5, 6] was obtained from the IMvigor210CoreBiologies package; counts were normalised by the supplied size factors and $\log_2$-transformed. All samples were taken before atezolizumab. Of 348 patients with RNA sequencing, 50 were not evaluable for response and were excluded, leaving 298. GSE176307 [11], a real-world cohort of patients with mUC treated with atezolizumab, pembrolizumab, nivolumab, durvalumab or avelumab, was obtained from Gene Expression Omnibus using the authors' gene-level TPM matrix; 87 patients were evaluable for response, of whom 16 responded. It was used only to test whether the response gradient across metastatic pattern replicates,

since its covariate recording and event count cannot support an incremental-value analysis.

2.2 Signatures

Six immune signatures (CD8 T-effector, antigen-presentation machinery, immune checkpoint, cytolytic activity, GEP-18, IFN-γ-6) and two control signatures (EMT/stroma, proliferation) were scored as the unweighted mean log-expression of their member genes; gene lists are given in Online Resource 1. This departs from some original definitions — GEP-18, for instance, was published as a weighted sum normalised to housekeeping genes on the NanoString platform [1] — so these are re-implementations rather than the original assays. The controls were pre-specified as specificity checks: a confounder of immune biology should act on the immune signatures and not on stromal or proliferative ones.

2.3 Baselines

Baseline A comprised metastatic pattern, ECOG performance status and receipt of prior platinum. Metastatic pattern was coded in three levels: visceral, nodal or soft-tissue-confined, and not recorded. The third level was retained rather than excluded because its missingness is strongly informative — objective response rate was 59.3% in those 27 evaluable patients against 19.2% in the remainder — so a complete-case analysis would discard a group that differs systematically in outcome. What that category represents clinically cannot be determined from the released data; at least one of its members was biopsied in the liver, so it is not simply disease without distant spread. Baseline B added PD-L1 immunohistochemistry on tumor-infiltrating immune cells and TMB.

2.4 Statistical analysis

The response outcome was RECIST complete or partial response versus stable or progressive disease. Discrimination was estimated as out-of-fold area under the receiver operating characteristic curve (AUC) over 50 repeats of stratified five-fold cross-validation. Because the clinical model takes few distinct covariate patterns and therefore yields many tied predictions, the bootstrap was run within each repeat on that repeat's own predictions; averaging predictions across repeats breaks those ties with fold-to-fold noise and spuriously favors the continuous signature models. Incremental value was additionally tested by in-sample likelihood-ratio test, which is biased *in favor* of the larger model. Variance in signature scores explained by metastatic pattern was quantified by bias-corrected ω^2 with 2000-fold bootstrap intervals; ω^2 is used in place of η^2, which is biased upward, and can take small negative values when the true effect is zero. Overall survival was modelled by Cox regression, with Harrell's C-index for discrimination. Whether each signature's effect depended on mutational burden was tested by adding a signature-by-TMB product term to baseline B. Analyses were complete-case and the analyzed n is given with each result; the reduction from 298 to 234 patients for baseline B reflects missing PD-L1 and TMB, and baseline

A was therefore re-estimated on those 234 patients so that the two baselines are compared on identical cases. No correction for multiplicity was applied; eight signatures were tested against two baselines and two endpoints, and the p values should be read accordingly. Analyses used Python 3.12 with scikit-learn, statsmodels and lifelines.

2.5 Pre-specification and deviation

A liver-metastasis (yes/no) contrast was specified in advance as the exposure. *This contrast failed*: variance explained was approximately zero in both cohorts (ω^2 0.001–0.015), indistinguishable from the control signatures. Inspection of IMvigor210 alone showed the signal lay in the nodal-confined stratum, which motivated the coding used here. The replication cohort was then examined once. The exposure is therefore refined, not pre-specified, although the refinement used only the discovery cohort.

3 Results

3.1 Discrimination

Baseline A ($n = 298$, 68 responders) discriminated response with an out-of-fold AUC of 0.697 (95% CI 0.619–0.767). Baseline B ($n = 234$, 61 responders) reached 0.778 (0.708–0.844); on those same 234 patients baseline A reached 0.678 (0.591–0.759), so the gain from adding PD-L1 and TMB is approximately 0.10 AUC on identical patients.

Among the markers measured singly, TMB reached 0.720 (0.639–0.796) and PD-L1 0.564 (0.477–0.654). The six immune signatures ran from 0.538 to 0.618: IFN-γ-6 0.618, CD8 T-effector 0.609, antigen-presentation machinery 0.562, GEP-18 0.560, cytolytic activity 0.558 and immune checkpoint 0.538 (Fig. 1a). They therefore discriminated better than PD-L1 immunohistochemistry but worse than TMB: in a paired comparison on identical patients, TMB exceeded every immune signature (ΔAUC -0.122 to -0.189, all bootstrap intervals excluding zero).

3.2 Incremental value

Adding an immune signature to baseline A changed discrimination by -0.011 to $+0.003$ AUC, with likelihood-ratio p values of 0.11–0.99. Adding one to baseline B changed it by -0.006 to $+0.000$, with p values of 0.06–0.95 (Fig. 1b). Full numbers with intervals are given in Online Resource 1.

The proliferation control behaved differently, adding over both baselines (ΔAUC $+0.028$, $p < 0.001$ over A; $+0.007$, $p = 0.023$ over B). The procedure therefore detects added value of this magnitude when it is present, and the null result for the immune signatures is not simply a failure of the method.

3.3 Dependence on mutational burden

A signature could still be useful by identifying patients in whom mutational burden misleads — a tumor with high burden but no T-cell infiltrate, say. Adding a signature-by-TMB product term to baseline B gave no interaction approaching significance for

any of the six ($p = 0.42$–0.98), so the signatures neither modify the effect of mutational burden nor mark a subgroup in which it behaves differently.

3.4 What accounts for the apparent predictive value

Metastatic pattern explained 6.8–11.0% of the variance in the six immune signatures (CD8 T-effector $\omega^2 = 0.110$, 95% CI 0.060–0.169) while both controls were null (-0.001 and $+0.003$). Objective response rate fell monotonically across the three levels: 59.3%, 37.3% and 15.0% (Fig. 2a). Separating liver from other visceral disease sharpened the gradient further, to 59.3%, 37.3%, 18.0% and 9.9%.

Adjustment attenuated every immune signature on both endpoints. For response, the odds ratio per standard deviation fell from 1.52 to 1.22 for CD8 T-effector and from 1.27 to 1.01 for GEP-18; adjustment for metastatic pattern alone accounted for nearly all of this (to 1.24 and 1.02 respectively), and ECOG explained little signature variance ($\omega^2 \leq 0.031$, against 0.068–0.110 for metastatic pattern). For overall survival (348 patients, 232 events), hazard ratios moved from 0.81–0.90 to 0.88–0.97. The direction persisted — every adjusted point estimate remained below unity and IFN-γ-6 reached $p = 0.060$ — but no signature added discrimination to the clinical model (C-index 0.627 for the clinical model, at most 0.634 with a signature added; likelihood-ratio $p \geq 0.060$).

Neither control attenuated. On response both moved further from unity (proliferation 1.85 to 1.90; EMT/stroma 0.82 to 0.73) and on survival neither moved (0.90 to 0.91; 1.09 to 1.09) (Fig. 2b, c).

3.5 Alternative explanations and sensitivity

Biopsy tissue is an obvious alternative explanation, since a lymph-node biopsy samples an organ composed largely of lymphocytes. Tissue of origin was indeed associated with metastatic category ($\chi^2 = 34.8$, $p = 5 \times 10^{-4}$), but it explained far less signature variance than metastatic pattern did (ω^2 0.007–0.033 against 0.068–0.110). Restricting to the 168 evaluable patients biopsied in the bladder, which holds tissue constant, left the clinical baseline at 0.691 (0.581–0.789) and the incremental values null (largest $+0.022$, $p = 0.089$).

Conclusions were also unchanged when the 27 patients with no metastatic status recorded were excluded (clinical baseline 0.641; incremental values -0.013 to $+0.005$), when liver was separated from other visceral disease (clinical baseline 0.712), and when genes were standardised across samples before averaging rather than averaged on the log scale (Online Resource 1).

In GSE176307 [11] the response gradient replicated: 11 of 38 patients without visceral metastases responded (28.9%) against 5 of 49 with visceral disease (10.2%), odds ratio 0.28, $p = 0.048$. That cohort cannot test incremental value: performance status was recorded for only 72 of the 87 evaluable patients, and those 72 contain just 11 responders, too few to fit an adjusted model. The incremental-value result therefore rests on IMvigor210 alone: what replicates here is the exposure–outcome gradient, not the null finding.

4 Discussion

In this cohort, six established immune signatures added no detectable information to markers that are already measured, and TMB alone discriminated response better than any of them. Much of their apparent predictive value tracks the pattern of metastatic spread — a variable read from routine staging.

Whether metastatic pattern confounds or mediates the relationship cannot be resolved here. If a T-cell-inflamed phenotype itself restrains visceral dissemination, then metastatic pattern lies on the causal path from tumor immunity to outcome, and adjusting for it removes real signal. The result is therefore a constraint on validation rather than a refutation of signature biology: a marker that adds nothing to what is already measured has not been shown to be useful, whichever causal structure holds.

This question has not been put to these data in this form. The principal molecular analysis of IMvigor210 compared nested models by likelihood-ratio test, but the models being compared were molecular throughout; performance status and metastatic site do not enter it [7]. The signature most recently developed in this cohort is benchmarked against PD-L1 and TMB, without clinical covariates [8]. The nearest published comparison asks whether clinical data improve a gene-expression model and reports that it does, but its two arms are expression alone and expression plus clinical data — there is no clinical-only arm against which the expression component can be judged, and metastatic pattern is not among the clinical features used [9]. A further mUC study integrates molecular patterns with clinical data and reports the combined model, again without isolating the molecular contribution [10]. The asymmetry is the point: the baseline against which molecular markers are measured is usually another molecular model.

That TMB outperformed every immune signature on identical patients is consistent with the molecular analysis of this cohort [7], and sets the bar a transcriptomic marker has to clear. It is not proposed here as a better biomarker: its own discrimination, 0.72, falls well short of what would support use alone.

Two features of the analysis bear on how the adjusted estimates should be read. Odds and hazard ratios are non-collapsible: adding a strong prognostic variable moves a conditional estimate away from unity even in the absence of confounding. That is exactly what the control signatures do. The immune signatures move towards unity against that pull, which makes their attenuation harder to explain as an artefact of conditioning. Second, the likelihood-ratio test and discrimination can part company — the EMT/stroma control reached $p = 0.041$ against baseline A while adding +0.009 AUC — which is why both are reported here.

The clinical context has also moved. Atezolizumab is no longer used in this indication, and first-line practice is now enfortumab vedotin with pembrolizumab, given without biomarker selection. These cohorts therefore speak to how biomarkers are validated rather than to a current treatment decision. The methodological point is not specific to atezolizumab: a signature proposed for any ICI regimen in this disease should be shown to add to performance status, metastatic pattern, PD-L1 and TMB before it is advanced as a selection tool.

Five limitations are material. The trial is single-arm, so predictive and prognostic effects cannot be separated; a randomised dataset such as IMvigor211 would be

required. The incremental-value analysis rests on one adequately powered cohort, the replication cohort having only 11 responders. The meaning of the unrecorded metastatic category is unknown, and the result should be read both with and without it, as reported. The confidence intervals on ΔAUC span roughly ± 0.03, so the data exclude a large incremental gain but not a small one; the positive control shows the procedure detects gains of about $+0.03$. Finally, the signatures are re-implementations on RNA sequencing rather than the original assays, and a weighted, platform-matched implementation might behave differently.

5 Conclusion

Six established immune signatures carried no detectable incremental value over markers already measured in metastatic urothelial carcinoma, and tumor mutational burden alone discriminated better than any of them. Demonstrating incremental value over a clinical and molecular baseline, rather than association with response alone, should be a minimum reporting standard for biomarker studies in this disease.

Authors' contribution

K Baghdasaryan: Protocol/project development, Data analysis, Manuscript writing/editing.

Statements and Declarations

Funding. No funding was received for conducting this study.

Competing interests. The author has no relevant financial or non-financial interests to disclose.

Ethics approval. This study analyzed de-identified data from published clinical trial and observational cohorts that are available in the public domain. No new human participants were recruited and no identifiable data were accessed, so institutional ethics approval was not required.

Consent to participate and consent to publish. Not applicable; all data were previously collected under the consent procedures of the original studies and are publicly released in de-identified form.

Data availability. IMvigor210 is available in the IMvigor210CoreBiologies R package. GSE176307 is available from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176307).

References

[1] Ayers M, Lunceford J, Nebozhyn M et al (2017) IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. J Clin Invest 127:2930–2940. https://doi.org/10.1172/JCI91190

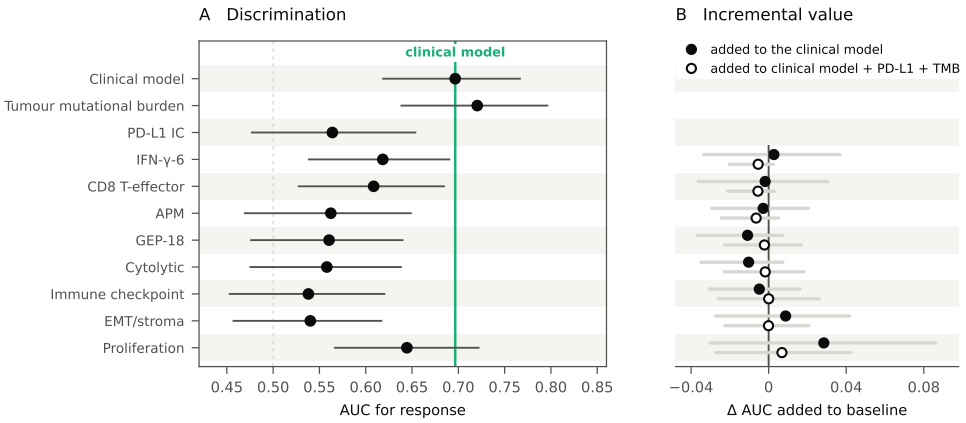


Fig. 1 Immune signatures add nothing to markers already measured. **a** Out-of-fold AUC for RECIST response (points; 95% bootstrap confidence intervals). The top three rows are the markers already in use: the routine clinical model (metastatic pattern, ECOG performance status, prior platinum), tumor mutational burden, and PD-L1 immunohistochemistry. The last two rows, EMT/stroma and proliferation, are pre-specified control signatures rather than candidate biomarkers. The green line marks the clinical model; the dashed line marks chance. **b** Change in AUC when each signature is added to the clinical model (filled circles) or to the clinical model plus PD-L1 and mutational burden (open circles); rows align with panel **a**

[2] Steyerberg EW, Vickers AJ, Cook NR et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21:128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

[3] Bellmunt J, Choueiri TK, Fougeray R et al (2010) Prognostic factors in patients with advanced transitional cell carcinoma of the urothelial tract experiencing treatment failure with platinum-containing regimens. J Clin Oncol 28:1850–1855. https://doi.org/10.1200/JCO.2009.25.4599

[4] Tumeh PC, Hellmann MD, Hamid O et al (2017) Liver metastasis and treatment outcome with anti-PD-1 monoclonal antibody in patients with melanoma and NSCLC. Cancer Immunol Res 5:417–424. https://doi.org/10.1158/2326-6066.CIR-16-0325

[5] Rosenberg JE, Hoffman-Censits J, Powles T et al (2016) Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet 387:1909–1920. https://doi.org/10.1016/S0140-6736(16)00561-4

[6] Balar AV, Galsky MD, Rosenberg JE et al (2017) Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet 389:67–76. https://doi.org/10.1016/S0140-6736(16)32455-2

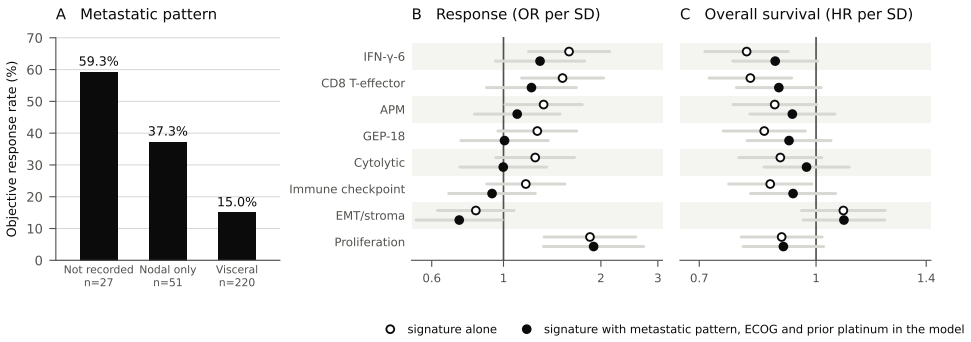


Fig. 2 Metastatic pattern accounts for the apparent predictive value. **a** Objective response rate by the three-level metastatic variable. **b** Odds ratio for response and **c** hazard ratio for overall survival, per standard deviation of signature, from a model containing the signature alone (open circles) and from one that also contains metastatic pattern, ECOG performance status and prior platinum (filled circles); horizontal lines are 95% confidence intervals. Each point is the effect estimate of the signature itself, not the discrimination of the combined model. Every immune signature moves towards unity once the clinical variables are present; neither control does

[7] Mariathasan S, Turley SJ, Nickles D et al (2018) TGFβ attenuates tumor response to PD-L1 blockade by contributing to exclusion of T cells. Nature 554:544–548. https://doi.org/10.1038/nature25501

[8] Langfelder P, Lin ET, Tsai YT, Cha TL, Shieh GS (2025) Gene signature for response prediction to immunotherapy and prognostic markers in metastatic urothelial carcinoma. Front Immunol 16:1607222. https://doi.org/10.3389/fimmu.2025.1607222

[9] Piyawajanusorn C, Ghislat G, Ballester PJ (2025) Predicting atezolizumab response in metastatic urothelial carcinoma patients using machine learning on integrated tumor gene expression and clinical data. npj Precis Oncol 9. https://doi.org/10.1038/s41698-025-00969-8

[10] Váradi M, Horváth O, Soós E et al (2025) Combining molecular patterns and clinical data for better immune checkpoint inhibitor prediction in metastatic urothelial carcinoma. Cancer Immunol Immunother 74:370. https://doi.org/10.1007/s00262-025-04224-8

[11] Rose TL, Weir WH, Kim WY et al (2021) Fibroblast growth factor receptor 3 alterations and response to immune checkpoint inhibition in metastatic urothelial cancer: a real world experience. Br J Cancer 125:1251–1260. https://doi.org/10.1038/s41416-021-01488-6

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarytables1.xlsx