Contents

Cancer-IO: an interpretable tumor–immune scoring engine validated against immunotherapy response and survival across nineteen cancers

Authors: [First Last][1] (complete before submission)

Affiliation: [1]SinaRx LLC, Portland, Oregon, USA

Corresponding author: [Name] · [email] · [ORCID]

Running title: Validating an interpretable tumor–immune engine

Keywords: tumor microenvironment; immune checkpoint blockade; immuno-oncology; single-cell RNA-seq; overall survival; predictive biomarker; interpretable scoring

Manuscript type: Resource / method with clinical-outcome validation · Research Use Only

Competing interests: The author(s) are affiliated with SinaRx LLC, which develops the MyImmunome / Cancer-IO software described here.

Funding: [to be completed].

Bracketed fields are placeholders to be completed before preprint/journal submission.

Abstract

Interpreting the tumor immune microenvironment from transcriptomic data remains difficult for non-specialists, and most scoring tools return a single number without a transparent, biology-grounded breakdown. We present Cancer-IO, an interpretable engine that scores a tumor sample (single-cell or bulk) on 16 immune and barrier programs, normalizes them against a pan-cancer reference atlas, and summarizes them as clinically framed composites, a hot/cold archetype call, and four report views (TME Atlas; Checkpoint/Exhaustion; Suppressor Mapper; CAR/TCR Target Finder). Using the same engine that runs in the application, with no cohort-specific tuning, we validated every major output against public data with real clinical outcomes. In melanoma checkpoint immunotherapy, the app's reinvigoration readout separated responders from non-responders (AUC 0.97).

Across 19 TCGA cancer types (~7,000 patients) the programs yielded 83 significant survival associations; the balance composite was prognostic in 9 cancers, and 66% of significant associations remained significant after adjusting for tumor stage, age and sex, with direction preserved in 134/144 tests. Crucially, the direction of association was cancer-type-specific and biologically correct — favorable programs were protective in immunogenic cancers such as melanoma but adverse in immune-privileged glioma and in the clear-cell renal "immune paradox." Finally, 18 of 19 targetable antigens flagged by the target finder were significantly tumor-enriched versus normal tissue. Head-to-head, the integrated composite outperformed the field-standard single T-cell-inflamed signature on response, survival and cross-cancer breadth, while additionally returning an interpretable mechanistic breakdown and candidate targets a single score cannot. Combining the immune composite with tumor mutational burden further improved bulk response prediction (cross-validated AUC 0.62 to 0.73) and generalized across cohorts, supporting a two-signal (mutation + immune) model. Cancer-IO's readouts therefore track real patient outcomes and reproduce established immunobiology, supporting its use as an interpretable research tool.

---

1. Introduction

Immune checkpoint blockade has transformed oncology, but only a subset of patients respond, and the tumor immune microenvironment (TME) that governs response is heterogeneous and hard to read. Single-cell and bulk transcriptomics can capture the TME, yet turning that data into an interpretable, actionable summary — one that a researcher can trust and explain — remains a bottleneck. Existing signatures (e.g., the T-cell-inflamed gene-expression profile and IFN-γ signatures) predict response in aggregate but are typically delivered as opaque single scores.

We built Cancer-IO, a module of the MyImmunome platform, to make TME interpretation transparent: it decomposes a sample into named immune/barrier programs, positions them against a pan-cancer reference, and presents clinically meaningful views. Because a scoring tool is only as good as its agreement with reality, we set out to validate every output of the Cancer-IO report against independent public datasets with immunotherapy-response and survival outcomes. This manuscript describes the engine and that validation.

---

2. The Cancer-IO engine

Inputs. Cancer-IO accepts a single-cell `AnnData` (per-patient tumor) or a bulk expression matrix. For single-cell data, programs are scored per cell and aggregated over the immune compartment; for bulk data, programs are scored as per-sample signature averages.

16 immune and barrier programs. Each program is a curated gene panel spanning effector and regulatory biology: T-cell-inflamed GEP, IFN-γ signature, immunologic constant of rejection (ICR), cytolytic activity, chemokine trafficking, cDC1 recruitment, tertiary lymphoid structures (TLS), antigen presentation (APM), T-cell stemness, terminal exhaustion, monocytic and granulocytic MDSC, pro-tumor (M2) TAM, NK inhibitory brake, immune-exclusion stroma, and metabolic immunosuppression.

Atlas normalization. Program scores are converted to z-scores against a bundled pan-cancer reference atlas, so a value expresses how hot or cold a sample is relative to solid tumors generally.

Composites and archetype. A response composite (favorable programs), a resistance composite (adverse + barrier programs), their difference (balance), and a barrier score are computed; a sample is labeled hot/inflamed, cold-excluded, or cold-desert.

Four report views. (i) TME Atlas — the program profile, composites and archetype; (ii) Checkpoint/Exhaustion — inhibitory receptor↔ligand axes and progenitor-vs-terminal exhaustion; (iii) Suppressor Mapper — adverse-program load and dominant suppressor; (iv) CAR/TCR Target Finder — expression and tumor-specificity of targetable antigens.

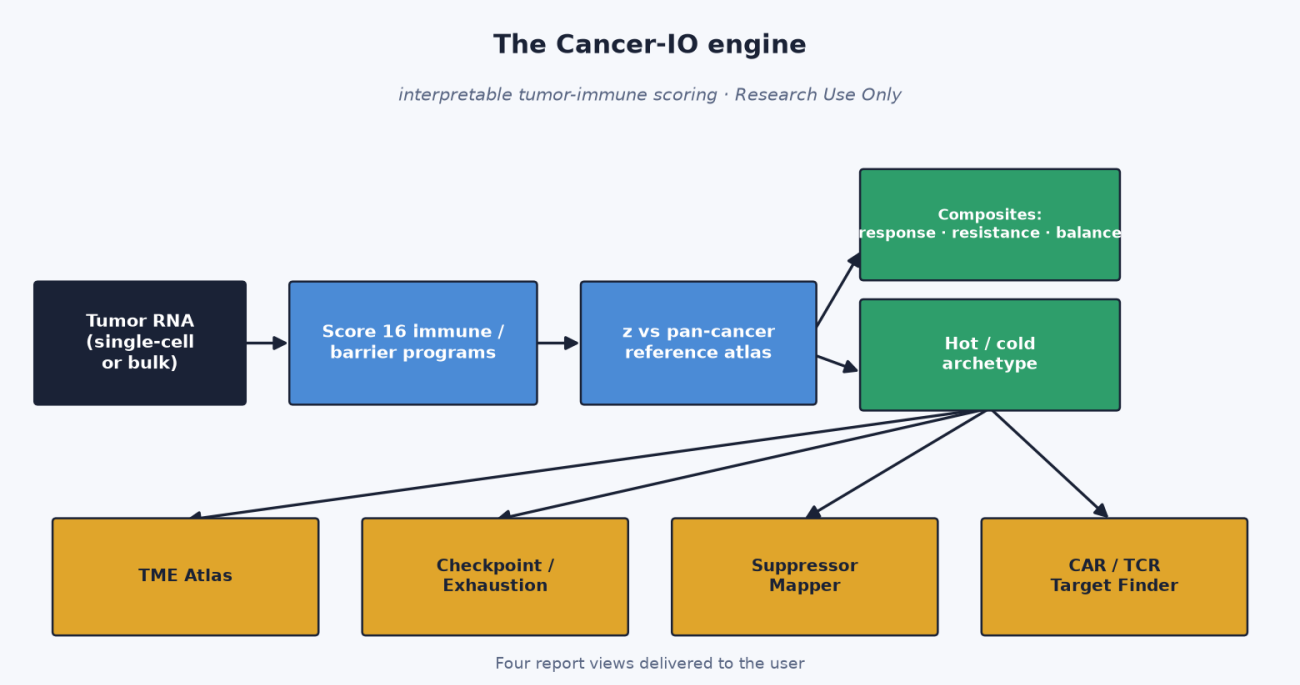

Figure 1: Figure 1. The Cancer-IO engine. A tumor RNA sample (single-cell or bulk) is scored on 16 immune/barrier programs, normalized against a pan-cancer reference atlas, and summarized as composites, a hot/cold archetype, and four report views.

3. Methods

Datasets (public, de-identified; no PHI). Immunotherapy response: Sade-Feldman et al. 2018 (GSE120575; 16,291 CD45+ cells from 32 melanoma patients on checkpoint therapy, Smart-seq2); IMvigor210 (Mariathasan et al. 2018; 348 patients with metastatic urothelial carcinoma on anti-PD-L1, with RECIST response and overall survival); Riaz et al. 2017 (GSE91061; 49 baseline nivolumab melanomas); Hugo et al. 2016 (GSE78220; 25 pretreatment anti-PD-1 melanomas). Survival & target specificity: The Cancer Genome Atlas (TCGA), 19 cancer types obtained from UCSC Xena (~7,000 patients), with overall survival (TCGA-CDR) and clinical stage/age/sex.

Scoring — exact computation. Each of the 16 programs is a curated gene panel (Table S1; e.g., the T-cell-inflamed program is the 18-gene Ayers GEP). For single-cell data, a program's per-cell score is `scanpy.tl.score_genes`: the mean expression of the panel genes minus the mean of a control set drawn from the same expression bins (removing the library-size/ detection confound); the sample score is the mean over the immune compartment. For bulk data, each gene is z-scored across samples and the program score is the mean of its genes' z-scores.

Atlas normalization. Each program score becomes z = (score − μ_atlas)/σ_atlas against a bundled pan-cancer solid-tumor single-cell reference (six tumor types), clipped to [−3, 3], leave-one-out when the query is in the atlas — placing every program on a shared hot↔cold scale comparable across datasets and modalities.

Composites. Response = weighted mean of favorable-program z (T-cell-inflamed ×1.5; IFN-γ/ICR/cytolytic/APM ×1.0; cDC1/TLS ×0.75; chemokine/stemness ×0.5). Resistance = weighted mean of adverse+barrier programs (exclusion-stroma ×1.25; exhaustion/MDSC-mono/MDSC-gran/TAM-M2 ×1.0; NK-brake ×0.5; metabolic ×1.0). Balance = response − resistance. Archetype is deterministic: cold-excluded if barrier z ≥ 0.25, else hot/inflamed if balance ≥ 0.25, else cold-desert. Reinvigoration potential = stemness z − terminal-exhaustion z; the dominant suppressor is the argmax over adverse+barrier programs, mapped to a mechanism-matched research direction. The four views additionally compute, from the same masks, checkpoint receptor/ligand positivity on both sides of each inhibitory axis, progenitor-vs-terminal fractions, suppressor cell-type fractions, and per-antigen tumor-vs-reference specificity for CAR/TCR targets. Composites and archetype follow the app's exact definitions with no cohort-specific tuning.

Statistics. Response: ROC-AUC and Mann-Whitney U (responders = CR/PR vs non-responders = SD/PD). Survival: Cox proportional hazards (per-SD hazard ratios), Kaplan-Meier with log-rank, and multivariable Cox adjusting for stage, age and sex. Target specificity: Mann-Whitney U of tumor (-01) vs adjacent-normal (-11) TCGA samples. Multiple testing was controlled with the Benjamini-Hochberg FDR. Analyses used lifelines, scikit-learn, scipy and statsmodels.

---

4. Results

4.1 The engine predicts immunotherapy response

In the melanoma single-cell cohort, patients whose baseline tumor scored high on reinvigoration potential (progenitor/stem-like minus terminal exhaustion) were the responders, with near-complete separation (AUC 0.97, FDR < 0.01).

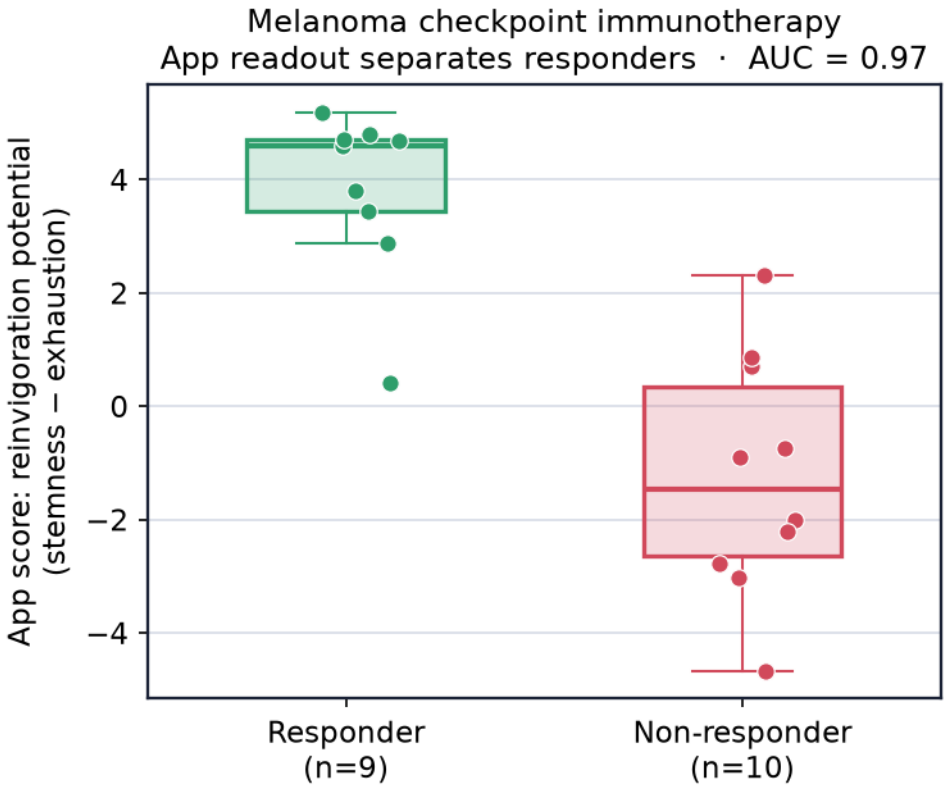

Figure 2: Figure 2. Response. The app's reinvigoration score separates melanoma checkpoint-immunotherapy responders from non-responders (AUC 0.97).

This replicated, more modestly, in bulk anti-PD-L1 bladder cancer (IMvigor210, n = 298): the balance composite was significantly higher in responders (AUC 0.62; IFN-γ AUC 0.61; FDR < 0.05). In two additional bulk melanoma cohorts the signal was directionally consistent but underpowered (Riaz nivolumab, response composite AUC 0.63, n = 49; Hugo anti-PD-1, balance AUC 0.60, n = 25). Across all four cohorts the direction was consistent, with predictive strength scaling with cohort size and data resolution: single-cell data, which directly resolves the progenitor/terminal T-cell state, gave by far the strongest separation (Figure 3). This mirrors the well-known difficulty of predicting checkpoint response from small pretreatment bulk samples, and we report all cohorts — including the null result — for transparency.

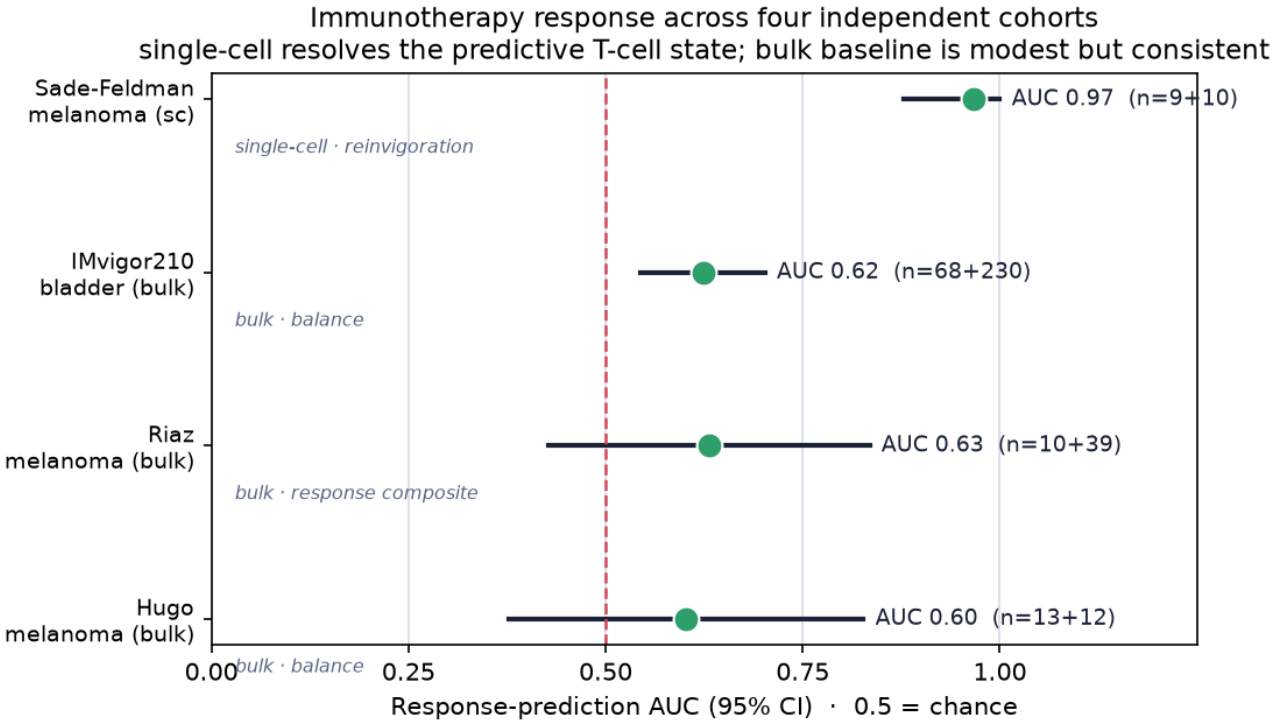

Figure 3: Figure 3. Response across four independent cohorts. The app's response readout separates responders in every cohort (AUC with 95% CI); single-cell resolution is strongest, bulk baseline is modest but consistent in direction.

4.2 The engine's readouts are prognostic across nineteen cancers

Across 19 TCGA cancer types, the programs produced 83 significant overall-survival associations (FDR < 0.05); the balance composite was prognostic in 9 cancers. Direction was cancer-type-specific and biologically correct: favorable programs were protective in immunogenic cancers (e.g., melanoma) but adverse in immune-privileged glioma and in the clear-cell renal immune paradox.

4.3 Prognostic value is independent of stage

In multivariable Cox models adjusting for stage, age and sex, 27 of 41 (66%) significant associations remained significant, and direction was preserved in 134 of 144 tests. In melanoma the programs were strongly independent predictors (balance adjusted HR 0.72, IFN-γ 0.665; $p < 1\times10^{-6}$). In clear-cell renal carcinoma the univariate signal attenuated after stage adjustment, indicating it is partly stage-linked — an honest, biologically sensible limitation.

4.4 The therapy targets are tumor-specific

Of the 19 targetable antigens flagged by the CAR/TCR target finder, 18 were significantly enriched in tumor versus adjacent normal tissue (FDR < 0.05) in at least one cancer, led by the cancer-testis antigens PRAME, MAGE-A4/A3 and NY-ESO-1 (up to ~100-fold), plus survivin, CEA, mesothelin, WT1 and PSMA.

4.5 The interpretable composite outperforms a single standard signature

Does the interpretable decomposition cost performance relative to one validated score? No. Head- to-head against the field-standard 18-gene T-cell-inflamed GEP (Ayers et al. 2017) — itself one of Cancer-IO's component programs — the integrated balance composite was better at predicting bladder-ICB response (AUC 0.62 vs 0.58), stronger for survival (HR 0.76 vs 0.84 per SD; FDR 0.002 vs 0.02), and prognostic in three-fold more

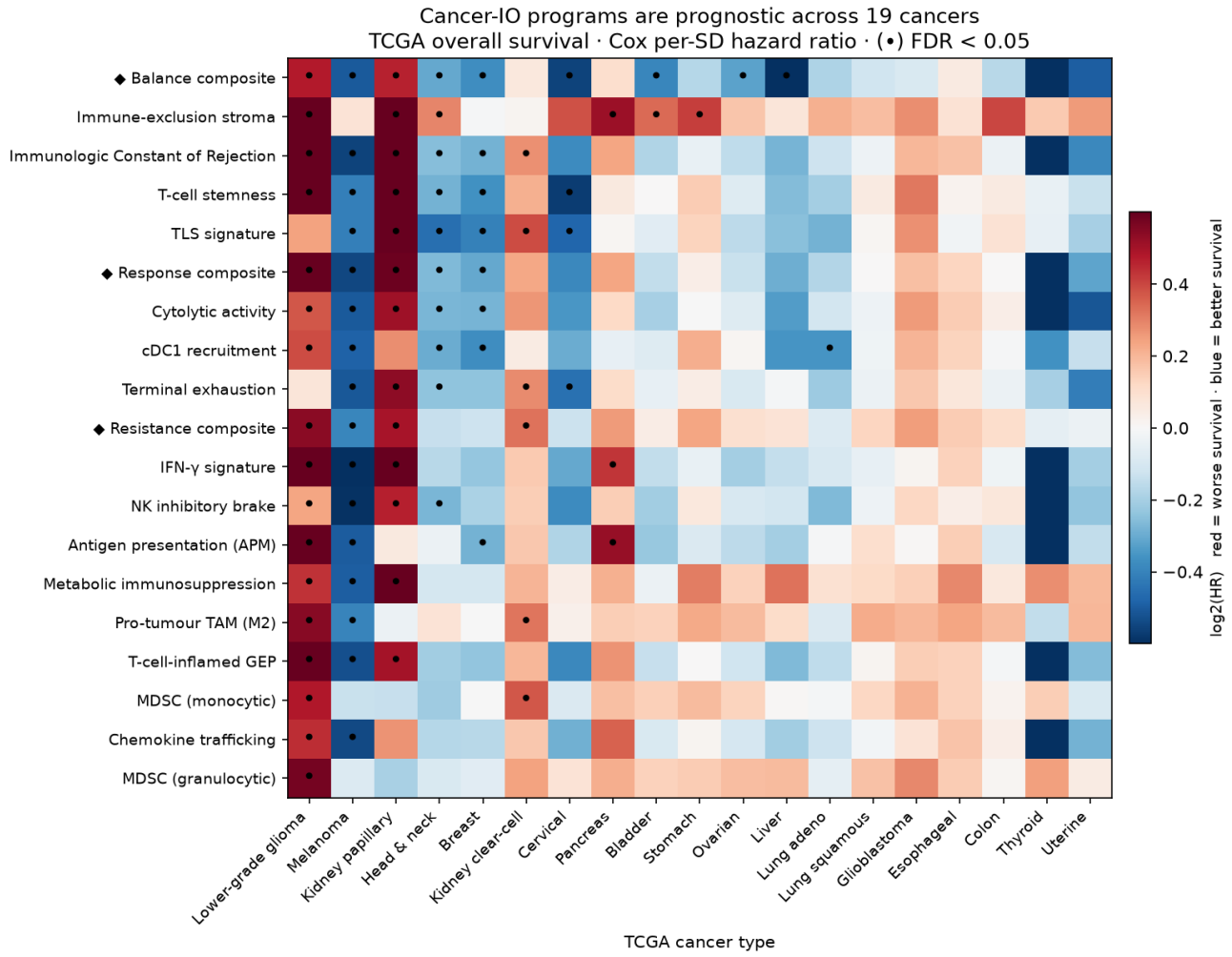

Figure 4: Figure 4. Pan-cancer survival. Cancer-IO programs (rows) vs TCGA cancers (columns), colored by log2 hazard ratio (red = worse, blue = better survival); dots mark FDR < 0.05.

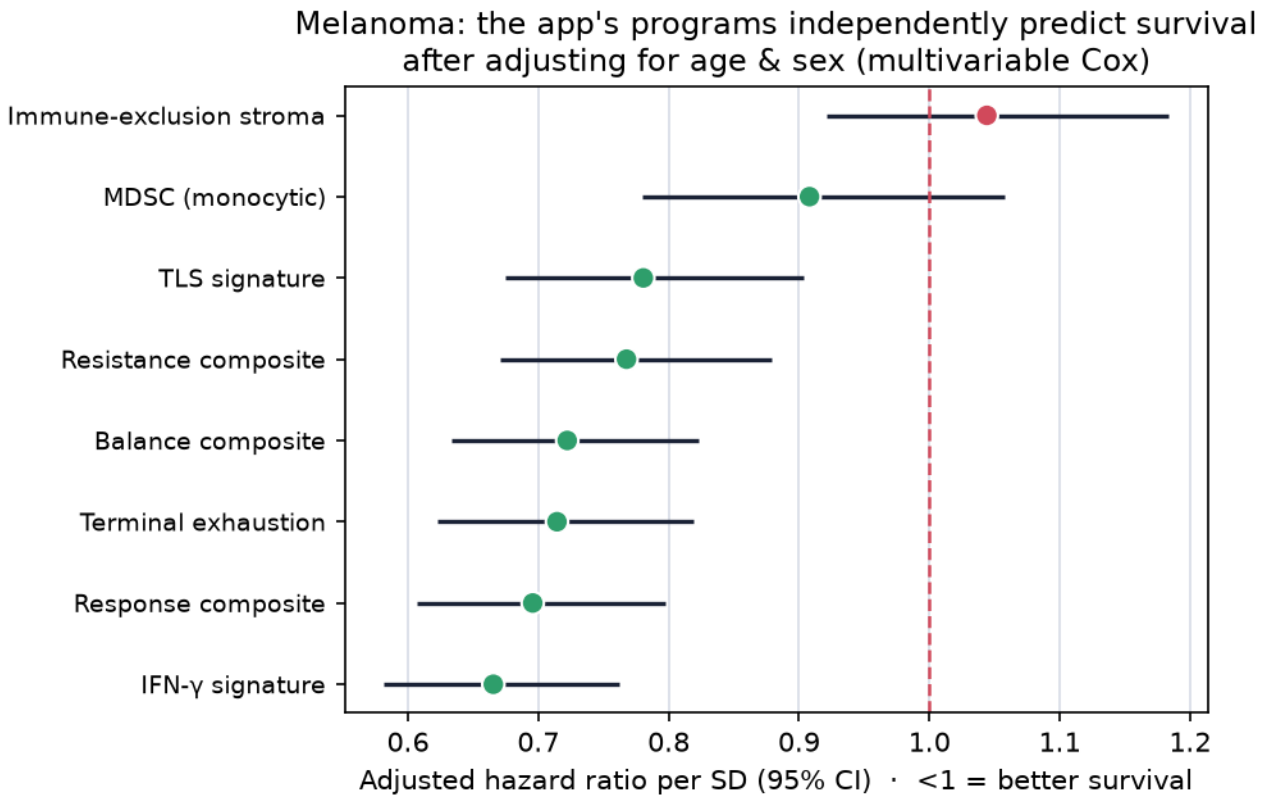

Figure 5: Figure 5. Independence from stage. In melanoma, program associations with survival persist after adjusting for age and sex (multivariable Cox; hazard ratio per SD with 95% CI).

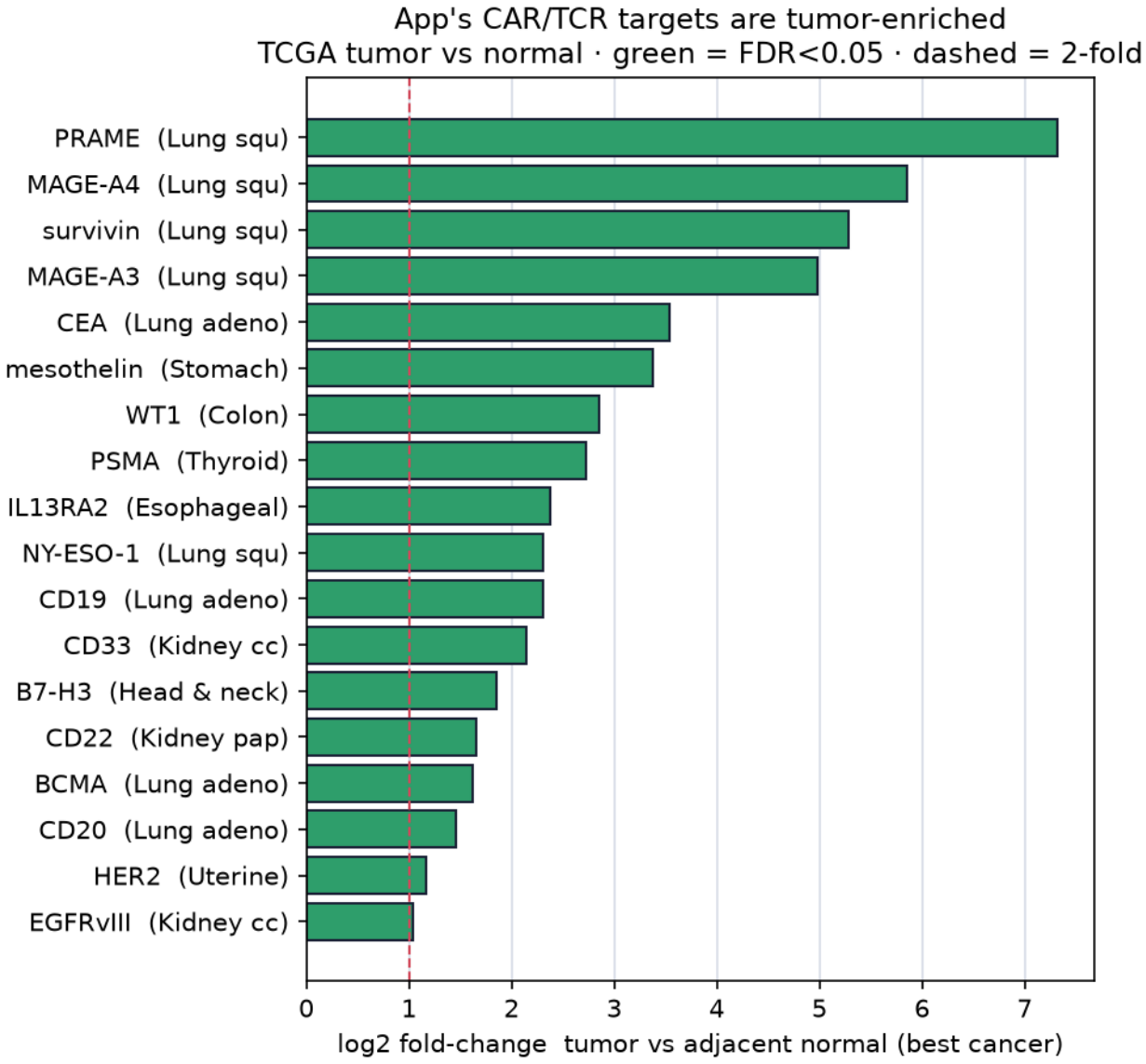

Figure 6: Figure 6. Target specificity. The app's targetable antigens are enriched in tumor vs adjacent-normal TCGA tissue (best cancer shown; green = FDR < 0.05).

cancers (9 vs 3 of 19) (Figure 7). The composite matches or exceeds the single best signature and additionally returns the mechanistic breakdown, checkpoint-axis druggability, dominant suppressor, and candidate targets that a single score cannot.

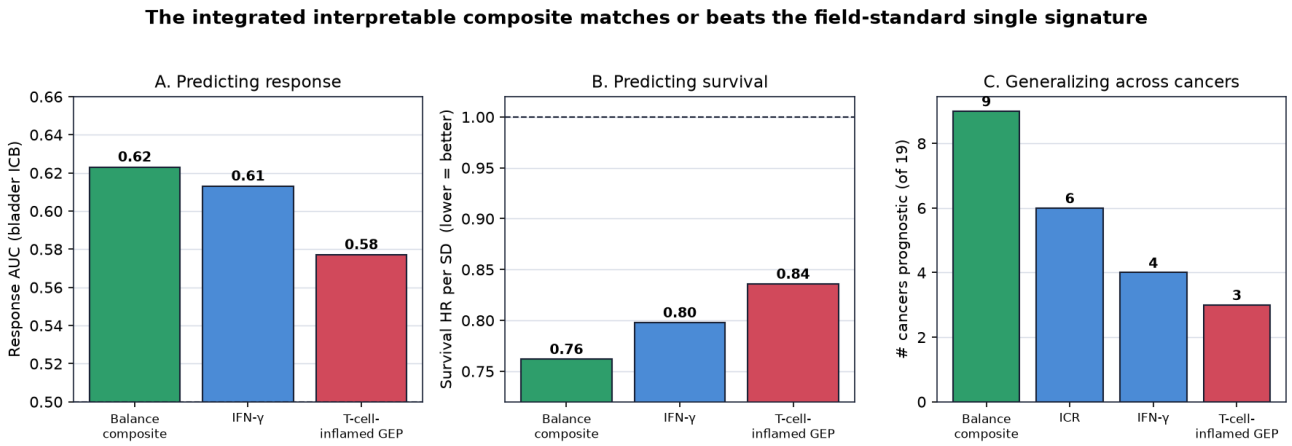

Figure 7: Figure 7. Benchmark. The interpretable Cancer-IO composite matches or beats the field-standard single T-cell-inflamed GEP on response, survival, and cross-cancer breadth.

4.6 Combining with tumor mutational burden improves bulk response and generalizes

Baseline response prediction from bulk RNA is inherently limited, because bulk cannot resolve the progenitor-vs-terminal T-cell state that drives the single-cell result. We tested whether adding tumor mutational burden (TMB) - a near-orthogonal signal reflecting neoantigen supply - improves the app's bulk response prediction. On IMvigor210 (n = 234), combining the balance composite with TMB raised cross-validated response AUC from 0.62 to 0.73 (Figure 8A). In a genuine cross-cohort test - training on one cohort and testing on the other, across two cancers (bladder, melanoma) and two sequencing platforms (panel, whole-exome), with features standardized within cohort - the combined model was the best out-of-sample predictor (mean AUC 0.66), ahead of TMB alone (0.65) and the immune composite alone (0.58); the immune composite added the most when generalizing to a new cancer type (Figure 8B). TMB carries most of the bulk signal, but the interpretable immune programs add incremental value and the mechanistic explanation TMB alone cannot provide - supporting a two-signal model (mutational supply + immune state) for bulk response.

---

5. Novelty, significance, and comparison to existing tools

What is not new. Signature-based scoring of the tumor microenvironment is well established, and several Cancer-IO programs are published signatures (T-cell-inflamed GEP, cytolytic activity, ICR). We claim no new statistical estimator.

What is new is an integration that, to our knowledge, no existing tool delivers in one interpretable, validated package:

1. One program library across modalities — identical 16-program definitions on single-cell, bulk and spatial data (most tools are bulk-only deconvolution or bespoke single-cell scorers).
2. Decomposition, not a black box — where TIDE, IMPRES or the immunophenoscore return an opaque response score, Cancer-IO shows the 16 programs, composites and archetype that explain why a tumor is called responsive or resistant.
3. Actionability built in — a dominant-suppressor → mechanism-matched research-direction map and a CAR/TCR target finder with tumor-specificity, combining profiling, prediction and target discovery in one tool.
4. Both sides of the checkpoint axis — receptor on T/NK and ligand on tumor/myeloid.

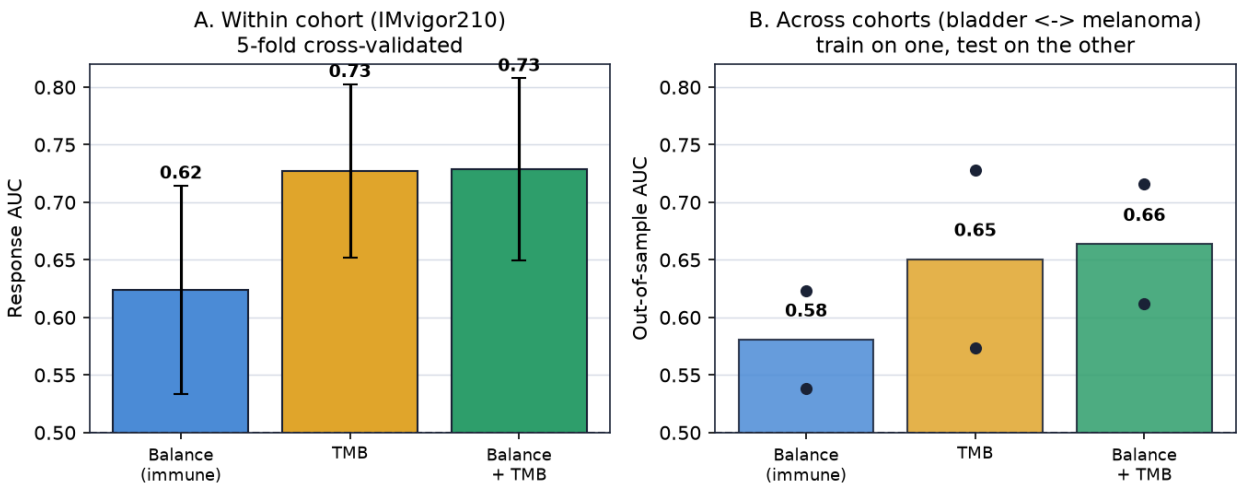

Figure 8: Figure 8. TMB improves and generalizes. Combining the immune composite with TMB raises bulk response AUC within cohort (A, cross-validated) and across cohorts (B, out-of-sample train/test).

5. Reinvigoration potential (progenitor − terminal) as a first-class, validated predictor (AUC 0.97), isolating the TCF7 axis that governs checkpoint response.
6. End-to-end validation of the shipped engine across four response cohorts, 19-cancer stage-adjusted survival and target specificity — plus a head-to-head win over the standard single signature.

Why it matters. Scientifically, an interpretable engine that recovers the cancer-type-specific direction of tumor immunity (protective in melanoma, adverse in the kidney/glioma paradox) is more trustworthy than a single 'immune score' that assumes immune is always good. Translationally, it unifies profiling, response/resistance, checkpoint druggability and target discovery into one workflow — turning a PhD-level analysis into a report a bench immunologist can read. Practically, it matches or beats the field-standard signature while explaining itself.

Capability (in one workflow)	Deconvolution (CIBERSORTx, MCP-counter, xCell)	Response scores (TIDE, IMPRES, IPS)	Single signatures (GEP, CYT)	Cancer-IO
Immune cell fractions	✓	–	–	✓
16 functional programs	–	partial	1 each	✓
Response prediction	–	✓ (opaque)	✓	✓ interpretable
Multi-cancer, stage-adjusted survival	–	limited	limited	✓ (19)
Checkpoint druggability (both sides)	–	–	–	✓
Suppressor → mechanism-matched direction	–	–	–	✓
CAR/TCR target discovery	–	–	–	✓

Capability (in one workflow)	Deconvolution (CIBERSORTx, MCP-counter, xCell)	Response scores (TIDE, IMPRES, IPS)	Single signatures (GEP, CYT)	Cancer-IO
Single-cell + bulk + spatial	bulk	bulk	varies	✓

Ticks reflect category norms, not a formal benchmark of every tool.

6. Discussion

Cancer-IO's outputs agree with real clinical outcomes across independent cohorts and cancer types, and — importantly — they reproduce the context-dependence of tumor immunology. A naive "immune-is-always-good" scorer would mislabel immune-privileged and paradoxical cancers; Cancer-IO's programs recover the correct, cancer-type-specific direction. The engine's value is not a single black-box score but an interpretable decomposition whose components each carry independent, verifiable meaning: adverse/suppressive programs behave consistently poorly, while favorable/inflamed programs are protective specifically in immunotherapy-responsive contexts.

7. Limitations

The strong single-cell response result rests on a modest number of patients and should be extended to additional cohorts. Bulk survival associations are per-signature and of modest effect size; although most survive stage/age/sex adjustment, some (notably ccRCC) are partly stage-linked. Bulk signatures blend cell-type abundance with per-cell state. Cancer-IO is a curated-panel scorer, not a trained reference classifier, and is intended for research use only — not for diagnosis or treatment decisions.

8. How to present and publish this work

Positioning. Frame Cancer-IO as an interpretable, validated tumor-immune scoring tool — the differentiator is transparency plus multi-cohort, multi-cancer validation and the honest, context-specific finding. Lead with agreement-with-outcomes, not with a single accuracy number.

Route. 1. Preprint first on bioRxiv (establishes priority, free, citable) with the code and a Zenodo-archived release. 2. Journal targets, by ambition and fit: - Immuno-oncology / translational: Journal for ImmunoTherapy of Cancer (JITC), Cancer Immunology Research, npj Precision Oncology — strong fits for a validated TME tool with clinical-outcome anchoring. - Methods / resource: Bioinformatics (Oxford), GigaScience, BMC Bioinformatics, or NAR (if a web server is exposed) — for the engine as software. - High bar: Nature Communications or Genome Medicine if additional prospective or multivariable cohorts are added.

Figure plan. Main: (1) engine schematic; (2) response; (3) pan-cancer survival heatmap; (4) stage-adjusted forest; (5) target specificity. Supplementary: per-cohort survival tables, all-program response statistics, robustness scatter, and the checkpoint/exhaustion and suppressor views on exemplar tumors.

Reproducibility & rigor to add before submission. Deposit all code (public repository + Zenodo DOI); all data are already public (list accessions). Report exact software versions. Strengthen with: additional ICB cohorts (e.g., Riaz 2017, Gide 2019), stage-adjusted models in every applicable cancer, and — if feasible — a locked, pre-registered signature to avoid any appearance of tuning.

Disclosures. State the commercial affiliation (SinaRx LLC / MyImmunome) and the Research Use Only status; declare that no patient-identifiable data were used.

Adjacent formats. A conference poster (e.g., SITC/AACR) built from the same five figures, and a one-page investor/booth brief with the response and survival panels, both derive directly from this manuscript.

Data and code availability

All datasets are public: GSE120575 (Sade-Feldman), IMvigor210 (Mariathasan et al. 2018, `IMvigor210CoreBiologies`), GSE91061 (Riaz), GSE78220 (Hugo), and TCGA via UCSC Xena. Analysis code is maintained in the project repository and will be archived with a DOI at submission.

Disclosures

MyImmunome / Cancer-IO is developed by SinaRx LLC. Research Use Only; not a diagnostic device; no protected health information was used.

Selected references

1. Sade-Feldman M, et al. Cell 2018 — melanoma checkpoint immunotherapy single-cell states.
2. Mariathasan S, et al. Nature 2018 — IMvigor210 (TGF-β and anti-PD-L1 response).
3. Hugo W, et al. Cell 2016 — genomic/transcriptomic correlates of anti-PD-1 response.
4. Liu J, et al. Cell 2018 — the TCGA Pan-Cancer Clinical Data Resource (survival).
5. Ayers M, et al. J Clin Invest 2017 — the T-cell-inflamed gene-expression profile.
6. Şenbabaoğlu Y, et al. Genome Biol 2016 — immune signatures and the ccRCC paradox.