

Multicellular immune ecotypes within solid tumors predict real-world therapeutic benefits with immune checkpoint inhibitors

Received: 4 February 2025

Accepted: 2 October 2025

Published online: 13 November 2025

 Check for updates

Xuefeng Wang  , Tingyi Li¹, Islam Eljilany  , Alex Soupir  , Michael Radmacher³, Phaedra Agius³, Vineeth Sukrithan⁴, Aakrosh Ratan  , Martin McCarter  , Bodour Salhia⁷, Howard Colman⁸, Abdul R. Naqash⁹, Igor Puzanov  , Susanne Arnold  , Julian A. Marin-Acevedo¹², Shridar Ganesan¹³, Michelle Churchman  , Patrick Hwu¹⁴, Paulo C. Rodriguez  , William S. Dalton³, George J. Weiner¹⁶ & Ahmad A. Tarhini  

Immune checkpoint inhibitors (ICIs) have transformed cancer treatment, yet predicting patient response remains a major challenge. Carcinoma ecotypes, which capture the cancer-immune interactions, show promise as prognostic biomarkers but remain untested in real-world settings. We compile and analyze the ORIEN Avatar ICI cohort of 1610 patients with matched gene expression data from a broader dataset of 14,997 individuals. Using EcoTyper-based immunophenotyping, we define ecotypes and assess their prognostic value across cancers, with a focused analysis in melanoma. Distinct cell states and ecotypes are consistently associated with survival outcomes across cancer types. We further develop a melanoma-specific ICI predictive model and validate it using data from the phase III ECOG-ACRIN E1609 trial as well as in external harmonized melanoma datasets. Together, these findings establish an ecotype-based framework and provide real-world evidence for their translational utility as clinically actionable biomarkers with prognostic and predictive value to guide ICI therapy.

Cancer continues to be a major global health challenge, responsible for nearly 10 million deaths in 2020 alone¹. Immunotherapies, particularly immune checkpoint inhibitors (ICIs), are showing great potential as new treatments. They work by reinvigorating exhausted T cells and bolstering anti-tumor immunity^{2,3}. However, responses to ICIs remain highly variable across different cancers and patients, largely due to the complex interactions within the tumor microenvironment (TME)^{4–6}. While previous research has largely concentrated on individual components of the TME, such as overall immune infiltration score, PD-L1 expression, and cytotoxic activity (CYT)^{7–9}, recent advances in computational deconvolution techniques have emphasized the critical role of co-occurring cellular states in shaping therapeutic outcomes¹⁰.

It is increasingly recognized that both cancer and immune compartments exhibit significant heterogeneity, and their dynamic interactions play a pivotal role in shaping therapeutic outcomes. Localized immune responses can prevail within specific regions of the tumor microenvironment. However, these local victories may not translate into broader clinical benefits if the immune response fails to expand or persist systemically. Rather than viewing the TME as a static collection of individual cell types, deep immunophenotyping methods such as EcoTyper reveal that immune and stromal cells organize into distinct, co-occurring cellular states that interact dynamically to shape tumor behavior and immune responsiveness^{7,10,11}. For instance, fibroblastic and epithelial-like cells co-occurring with immune-suppressive

macrophages create an environment that hinders T cell infiltration, leading to immunosuppression and resistance to therapy. In contrast, ecosystems enriched with activated CD8+ T cells, mature dendritic cells, and proinflammatory monocytes foster immune activation, resulting in better responses to ICIs. Similarly, ecosystems characterized by central memory T cells and naïve immune populations are linked to prolonged survival and durable responses to immunotherapy. On the other hand, ecosystems dominated by myeloid-derived suppressor cells (MDSCs), M2-like macrophages, and immune-suppressive fibroblasts actively inhibit T cell function and contribute to immunotherapy resistance¹⁰. While these mechanisms are well established individually, their collective prognostic significance, especially in the context of interactions between co-existing cellular states, has not been extensively examined on a broader population scale.

In the original EcoTyper framework, Luca et al.¹⁰ identified ten carcinoma ecotypes (CE1–CE10) by analyzing patterns of cell-state co-occurrence across over solid tumors from 16 cancer types. These ecotypes captured distinct TME architectures with specific biological and clinical associations. CE1 and CE2 were broadly lymphocyte-deficient or immunologically cold, enriched for stromal or epithelial programs and linked to poor prognosis. CE3 was enriched for immunosuppressive myeloid cell states and associated with microsatellite instability and unfavorable survival. CE4 was characterized by myogenesis and a depletion of B cell receptor signaling¹², and was more prevalent in head and neck cancer and prostate cancer. CE5 and CE6 appeared as more intermediate TME ecosystems, with CE5 associated with smoking-related mutations and CE6 enriched for normal-like tissue elements. CE7 and CE8 showed modest immune infiltration; CE7 was linked to age-related signatures and CE8 to breast cancer subtype signatures¹³, and both were more frequent in patients with metastases¹⁴. CE9 and CE10 represented the most immunogenic ecotypes and enriched for activated CD8+ T cells and B cells; they were consistently associated with better patient survival outcomes across multiple cohorts. CE9, in particular, exhibited strong IFN- γ signaling and outperformed other transcriptomic features in predicting immunotherapy benefit in the original study. Follow-up analyses showed that CE10 is more enriched for naïve B and T cells, whereas CE9 contains a higher proportion of activated or exhausted B/T cell subsets¹³. Importantly, CE1, CE9, and CE10 demonstrated high spatial aggregation and strong potential for forming cellular niche interactions within the TME.

Many existing studies, such as The Cancer Genome Atlas (TCGA), emphasize samples with high tumor content and often underrepresent immune and stromal components that are critical to cancer progression and patient outcomes. ICI clinical trials and preclinical studies typically include highly selective patient populations, which limits the broader applicability of their findings. These constraints highlight the need for analyses that integrate real-world data to capture the full diversity of tumor microenvironments and patient characteristics seen in clinical practice. To address this gap, we leverage the ORIEN Avatar dataset, a part of the Total Cancer Care (TCC) initiative that has enrolled over 400,000 patients across 18 cancer centers. This resource provides high-quality molecular profiles and clinical outcomes from patients treated with modern therapies. From the July 2023 dataset freeze, we identify a subcohort of 1610 ICI-treated patients with matched tumor molecular data (including RNA-seq and whole exome sequencing) and detailed clinical annotation. This dataset offers greater scale and clinical representativeness, capturing ecotype distributions across contemporary patient populations and treatment landscapes.

In this work, we apply EcoTyper to the ORIEN Avatar ICI cohort to characterize tumor-immune ecotypes and evaluate their prognostic value in real-world settings. We further develop a melanoma-specific prognostic signature and validate it across multiple

independent cohorts. Our results support the use of composite risk scores based on ecotypes as predictive biomarkers for ICI response, complementing and in some cases outperforming classical immune signatures.

Results

Study population

Founded in 2014, ORIEN real-world data (RWD) includes a cohort of over 400,000 patients who have consented for lifetime participation in the Total Cancer Care (TCC) protocol. The Avatar program has enrolled over 24,000 patients across 30 cancer types. All samples in the Avatar program were analyzed with high-quality molecular profiling of tumors, including whole exome sequencing (WES) and transcriptome sequencing (RNAseq). Each patient contributes around 325 clinical data points, which are updated every six months, covering 18 clinical domains such as outcomes, treatments, demographics, and medical history. The clinical and molecular data are processed by a centralized team responsible for performing rigorous QA/QC, harmonizing datasets across the network, and linking clinical-genomic patient profiles. The latest version includes patients up to July 31, 2023. In a comprehensive examination of the clinical characteristics of participants, we identified 1610 patients treated with immune checkpoint inhibitors (ICIs) from a cohort of 14,997 individuals with pre-treatment sample collection in the ORIEN Avatar database. This real-world ICI cohort is categorized as follows: melanoma ($n=161$), head and neck squamous cell carcinoma (H&N) ($n=240$), non-small cell lung cancer (NSCLC) ($n=270$), bladder cancer ($n=151$), and kidney cancer ($n=311$) as presented in Table 1 and Supplementary Fig. 1. The mean \pm standard deviation (SD) age across the cohort was 61.53 (12.37) years. Specifically, melanoma patients presented a slightly younger demographic with a mean \pm SD age of 59.44 ± 14.30 years at the time of data collection, while NSCLC patients were the oldest with a mean \pm SD age of 63.51 ± 9.58 years. Men and non-Hispanic participants were predominant (61.30% and 93.42%, respectively) across the cohort in all cancer types. The analysis of tumor origin highlighted a noteworthy prevalence of primary tumors in all cancer types, except for a prominent 63.35% of melanoma cases that were identified as metastatic.

The data shows potential gaps in cancer staging information, or that some participants were selected regardless of their stages at the time of initial diagnosis, since over half (51.99%) had stages listed as unknown. However, all subjects included who had advanced metastatic stages necessitating treatment with ICIs. Therefore, various treatments were reported across the cohort, with pembrolizumab being the most administered drug (49.01%). As expected, melanoma patients received ipilimumab at a significantly higher rate (21.74%) compared to other cancer groups.

Global distribution of identified cell-states and multicellular ecotypes

Focusing on cell states and multicellular ecotypes in the discovery cohort, the heatmap in Fig. 1A shows the distribution of multicellular communities or carcinoma ecotypes (CEs) across all samples within this real-world ORIEN ICI cohort, which comprises 1610 patients with 1732 RNAseq samples. Of these, 1254 patients (78%) were successfully assigned to specific ecotypes.

These assignments covered all ten previously defined ecotypes across carcinomas (CE1–10)¹⁰. A notable trend was observed, where CE1 tumors represent more lymphocyte-deficient, and CE10 tumors represent more proinflammatory CEs. The ecotypes with the highest number of tumors assigned were CE1 and CE6 (a non-neoplastic tissue-enriched cell subtype), followed by CE8 through CE10. This distribution showed slight differences from the findings in the TCGA pan-cancer discovery cohort ($n=4729$), where CE1 and CE8 were the most prevalent¹⁰.

Table 1 | Baseline clinical characteristics of patients included in the ORIEN ICI cohort

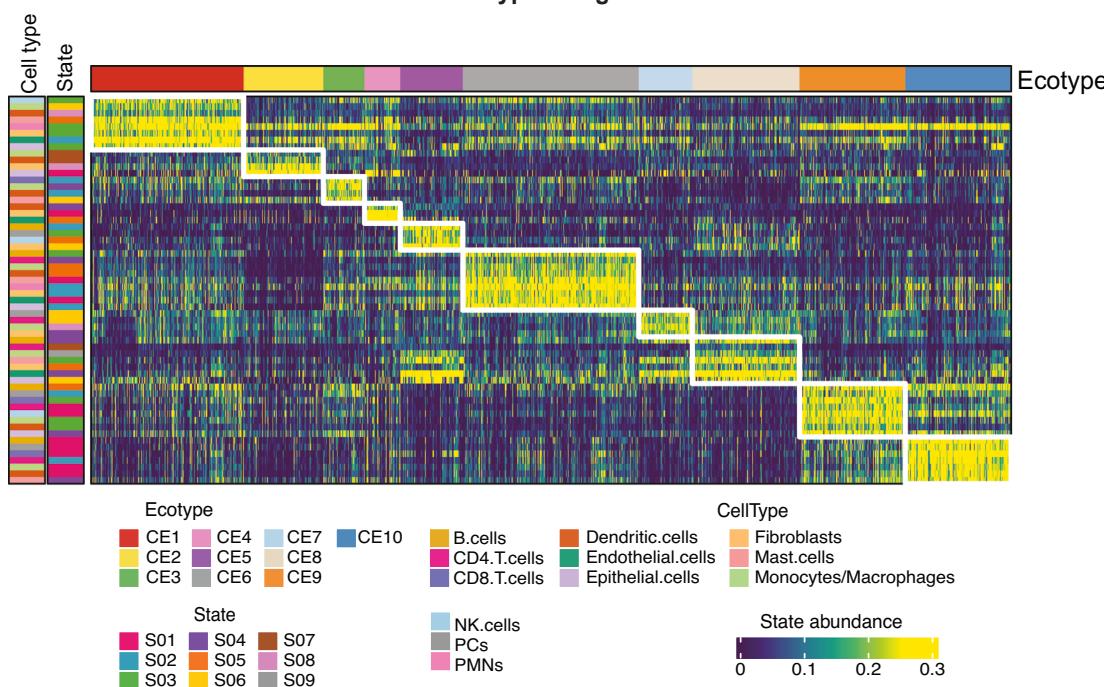
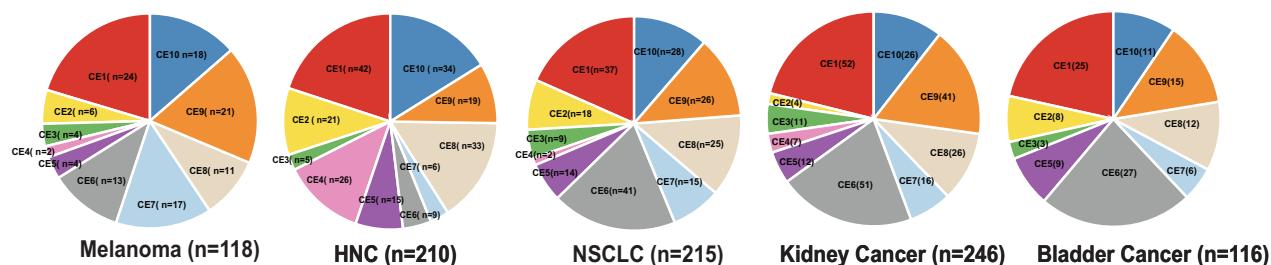
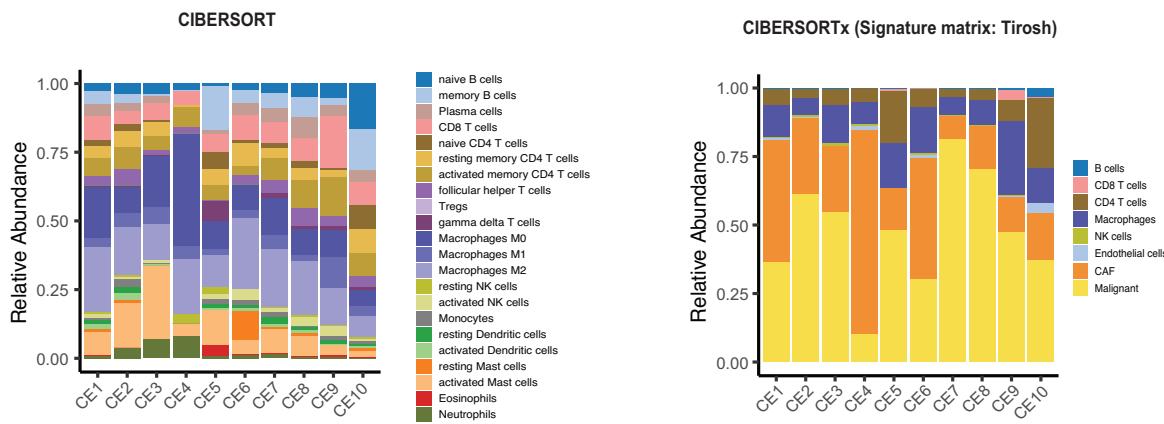
	Overall (N = 1610)	Melanoma (N = 161)	H&N (N = 240)	NSCLC (N = 270)	Others* (N = 942)
Age at study					
Mean (SD)	61.53 (± 12.37)	59.44 (± 14.30)	61.19 (± 10.82)	63.51 (± 9.58)	61.39 (± 13.00)
Gender, n (%)					
Women	38.70	38.75	17.92	46.67	41.72
Men	61.30	61.25	82.08	53.33	58.28
Ethnicity, n (%)					
Hispanic	4.72	8.08	2.92	2.92	5.73
Non-Hispanic	93.42	90.06	96.25	96.25	91.93
Unknown	1.86	1.86	0.83	0.83	2.34
Tumor origin, n (%)					
Primary	65.53	35.40	60.80	60.00	73.35
Metastasis	30.12	63.35	31.67	35.56	22.40
Both	4.35	1.24	7.50	4.44	4.25
LDH (serum), U/L					
Mean (SD)	182.5 (± 127.65)	195.44 (± 92.87)	194.19 (± 86.51)	248.20 (± 181.14)	219.23 (± 126.52)
Missing %	55.90	45.96	54.58	64.07	55.73
Karnofsky/ECOG PS n(%)					
EPS 0/ KPS 100%	12.05	12.42	6.67	13.33	17.95
EPS 1/ KPS 80-90%	10.31	1.86	10.83	18.89	9.13
EPS 2/ KPS 60-70%	2.23	0.62	1.67	3.33	2.33
EPS 3/ KPS 40-50%	0.19	9.32	0.42	0.74	0.00
Unknown	49.01	32.92	56.25	47.41	47.94
Cancer Stage, n (%)					
Stage 0	0.19	0.00	0.00	0.00	0.32
Stage I	6.40	3.11	5.00	12.22	5.63
Stage II	8.20	9.32	2.92	14.81	7.54
Stage III	15.84	14.90	7.50	14.33	18.47
Stage IV	17.33	10.56	24.17	18.52	16.45
Unknown	51.99	62.11	60.42	40.00	51.59
Immuno-oncology therapy, n (%)					
Ipilimumab	3.11	21.74	0.42	0.00	1.49
Nivolumab	26.02	22.36	27.08	21.11	27.71
Pembrolizumab	49.01	39.75	67.08	58.52	43.31
Avelumab	1.30	0.00	0.00	0.37	2.12
Atezolizumab	6.34	0.621	1.67	7.41	8.17
Cemiplimab	0.56	0.00	0.83	0.00	0.74
Durvalumab	2.55	0.00	0.83	10.37	1.17
Ipilimumab+ Nivolumab	10.99	14.29	2.08	2.22	15.29
Ipilimumab+ Nivolumab/ Pembrolizumab	0.06	0.62	0.00	0.00	0.00
Nivolumab-Relatlimab-rmbw	0.06	0.62	0.00	0.00	0.00

*Others: includes all other cancer types except melanoma, H&N and NSCLC. Note the sum of patients across individual cancer types exceeds N = 1610 because some patients were assigned to more than one cancer type due to multiple tumor or disease classifications in the real-world dataset.

ICI immune checkpoint inhibitor, H&N head and neck, NSCLC non-small cell lung cancer, SD standard deviation.

A comparison of the distribution of ecotypes results across the five major cancer types in the cohort as illustrated in Fig. 1B: melanoma, H&N, NSCLC, kidney cancer, and bladder carcinoma revealed that 80% of H&N patients, 80% of NSCLC patients, 75% of kidney cancer patients, 72% of patients with bladder cancer, and 70% of melanoma patients were successfully assigned to a pre-defined ecotype. In each cancer type, the previously known pan-cancer prognostic ecotypes CE1- CE2 and CE9- CE10 were found in around or

more than half of the patients, indicating the translational potential of these four ecotypes for predicting ICI outcomes. Among these five cancer types, H&N exhibited the most distinct cancer-specific distribution, with a higher number of samples assigned to CE4 and notably fewer in CE6. This pattern aligns with the discovery cohort, where a higher prevalence of CE4, potentially linked to myogenesis, was noted in both H&N and prostate cancer (and in older male patients).

A**Pan-cancer Ecotype Assignment****B****Ecotype assignment (RNAseq) across major cancer types****C****Relationship with cellular composition (melanoma)**

To help elucidate the cellular basis of ecotype status, Fig. 1C presents the average cellular composition of the ten CEs in melanoma, estimated using CIBERSORT⁷. The corresponding cellular composition profiles for the other major cancer types in this ICI cohort are shown in Supplementary Fig. 2. The left panel of Fig. 1C shows the cellular abundances derived from the original CIBERSORT method with the

LM22 immune cell reference. Most trends observed are consistent with the original EcoTyper study. As expected, CE9 and CE10 displayed a more immune-infiltrated or proinflammatory phenotype, with CE9 showing the highest CD8⁺ T cell content and CE10 enriched for naïve B cells. Notably, CE4 exhibited depletion of B cells and the highest abundance of M0 macrophages, while CE5 had increased levels of

Fig. 1 | Distribution of CEs in the pan-Cancer ORIEN IO cohort. **A** This heatmap displays the abundance profiles of cell states and CEs identified across all cancer types in the ORIEN IO cohort. Only samples assigned to pre-defined CEs (CE1-CE10) are included. Each column in the heatmap corresponds to RNA sequencing samples ($n=1340$), and each row shows the expression of cell state marker genes within a recovered ecotype. **B** Pie charts illustrate the distribution of the ten CEs within five specific cancer groups: melanoma, head and neck cancer (H&N), non-small cell lung

cancer (NSCLC), kidney cancer, and bladder cancer. Each segment indicates the number of assigned patients per ecotype group—abbreviations: CE, carcinoma ecotype; S, cell state. **C** Average cellular composition of the ten melanoma ecotypes ($n=118$ melanoma samples). The left panel shows immune cell proportions estimated using the original CIBERSORT method (using LM22 as reference). The right panel shows cellular composition estimated using CIBERSORTx with a melanoma-specific single-cell RNA-seq reference (Tirosh dataset¹⁵).

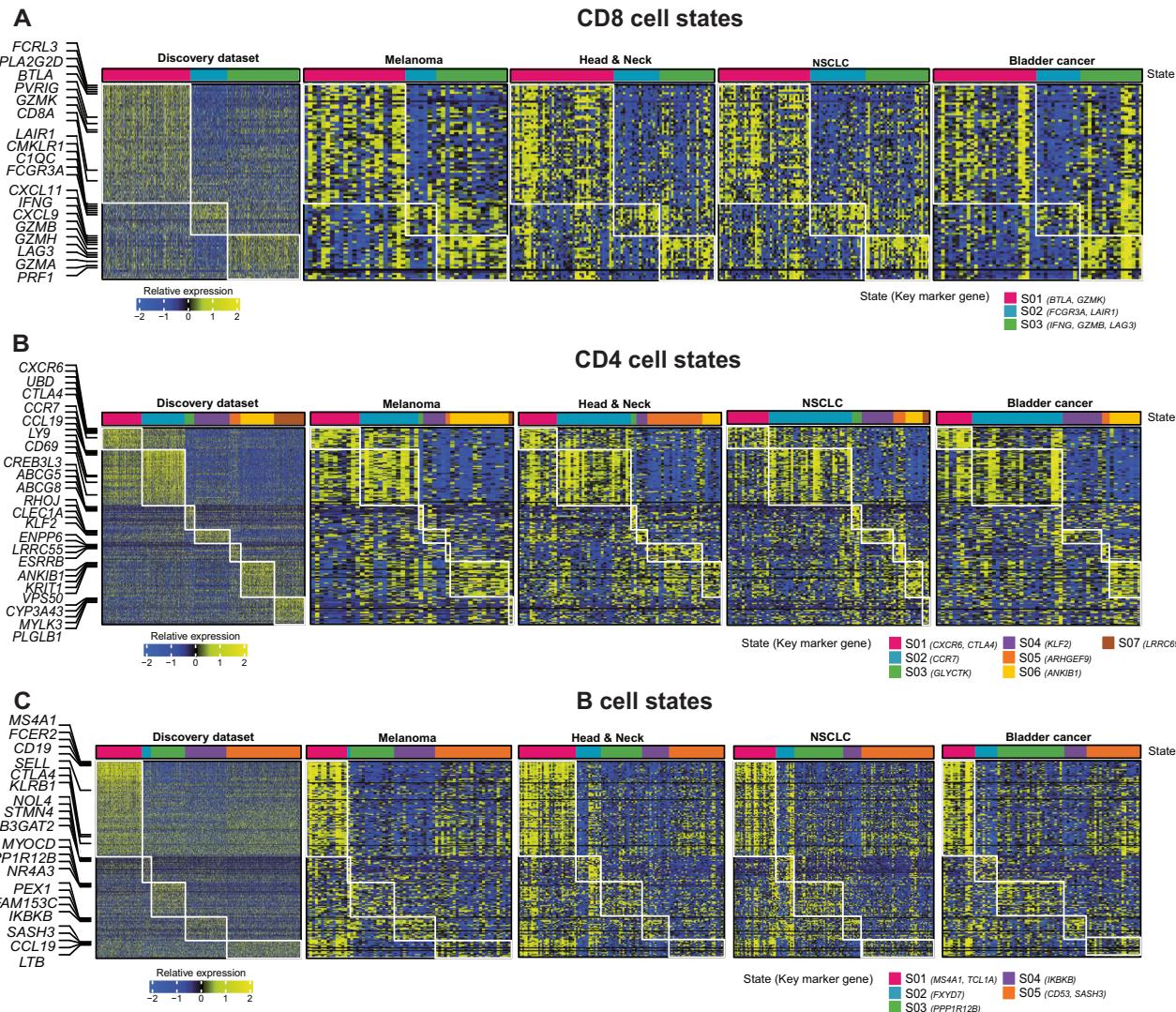


Fig. 2 | Cell state expression across major cancer types. The figure presents heatmaps displaying detailed expression profiles of tumor-associated cell states, focusing on CD4 (A), CD8 (B), and B cells (C). The first heatmap in each panel details the expression of marker genes for each cell state, using data from TCGA datasets as a reference. Subsequent heatmaps illustrate the expression profiles in each major cancer type. Each row corresponds to a cell-state marker gene, and each

column represents RNA sequencing samples. The number of RNAseq samples with assigned cell states are as follows: For CD8 states (A), Melanoma $n=53$, HNC $n=89$, NSCLC $n=99$, and Bladder $n=57$. For CD4 cell states (B), Melanoma $n=53$, HNC $n=74$, NSCLC $n=83$, and Bladder $n=52$. For B cell states (C), Melanoma $n=70$, HNC $n=115$, NSCLC $n=143$, and Bladder $n=80$.

memory B cells. We further applied CIBERSORTx using a reference built from melanoma single-cell RNA-seq data¹⁵, as shown in the right panel of Fig. 1C. Although this reference captures fewer immune cell types and subtypes, it allows for more accurate estimation of malignant and stromal components in solid tissues. Using this approach, we found that CE1 was characterized by a higher abundance of cancer-associated fibroblasts, potentially contributing to its immunosuppressive features. In addition, CE5 and CE10 had the highest proportions of CD4⁺ T cells, and CE9 and CE10 remained enriched for B cell and T cell populations, confirming their TME-hot characteristics.

Distribution of specific immune cell states across cancer types

As shown in Fig. 2, we further investigated the distribution of specific immune cell states identified from the gene expression data from the discovery dataset and in melanoma, H&N, NSCLC, and bladder cancer. These analyses focused on different cell types and their potential role in the immune system. Looking at Fig. 2A, it is apparent that the distribution of the three identified CD8 T cell states, CD8-S01 (naive/central memory T cells), CD8-S02 (late-stage effectors), and CD8-S03 (exhausted/effector memory), is consistent across the four cancer types examined. Note that CD4-S01 and

CD4-S03 are key components of the two proinflammatory communities CE10 and CE9, respectively.

In the CD4 cell type (Fig. 2B), however, only the distributions of the two cell states CD4-S01 (exhausted/effector memory) and CD4-S02 (naïve/central memory) remain consistent across the four cancers, accounting for half of the patients with an assigned CD4 CS assigned. These two states define the two proinflammatory ecotypes. It is observed that CD4-S05 was more enriched in H&N cancer, while CD4-S06 is more enriched in melanoma, with prominence in the specific immune environments of these cancers. Overall, NSCLC displayed a broader mixture of CD4 states, specifically S03-S06. This diversity is presumably attributed to the histological heterogeneity within NSCLC, including adenocarcinoma and squamous cell subtypes.

Regarding the distribution of B cell states, as portrayed in Fig. 2C, relatively similar patterns are observed across the cancers, with certain exceptions; there was a higher prevalence of B-S05 (activated B cells) in melanoma and NSCLC and a relatively higher occurrence of B-S03 (normal-enriched) in bladder cancer. In addition, in the analysis of dendritic cells shown in Supplementary Fig. 3, DC-S01 (myeloid cDC1) was prevalent across all four cancers, especially in melanoma and H&N. Meanwhile, DC-S05, identified as mature/normal-enriched, was more abundant in bladder cancer and NSCLC. These findings, especially the observed variations in non-CD8 immune cells, emphasize the significance of accounting for the complete range of immune cell states and interactions beyond the conventional categories of immune hot and immune cold.

Prognostic ecotypes for ICI cohort

The clinical significance of carcinoma CS co-occurrence networks, or CEs, has been previously established by the high concordance between the identified ecotypes and prognosis across various solid tumor types¹¹. Notably, CE9, characterized by IFN- γ signaling and highly active anti-tumor immune activity, demonstrated potential in predicting ICI response in metastatic melanoma trial data¹⁰. As a primary objective of this study, we conducted a thorough correlative analysis of treatment outcomes concerning all CEs within the ORIEN ICI cohort, aiming to evaluate their predictive value in a real-world setting objectively. In the ORIEN real-world dataset, we lacked specific tumor response information, i.e., complete response (CR)/partial response (PR)/stable disease (SD)/progressive disease (PD); therefore, we utilized OS from the ICI medication start date as the main outcome for the prognostic evaluation. The Kaplan-Meier curves in Fig. 3A compare the estimated OS probability associated with each of the 10 CEs identified in the ORIEN ICI cohort—across all patients who received ICIs. We understand that these patterns can be confounded by the cancer types and their distribution of CEs, but we still want to explore if well-established ecotypes, such as CE9 and CE10, can consistently exhibit similar effects at a pan-cancer level.

Overall, CE9 exhibited the most favorable survival outcomes compared to all other ecotypes, closely followed by CE6 and CE10. Compared to the survival pattern observed in the TCGA discovery cohort, the favorable performance of CE6 is less expected than that of the two other proinflammatory CEs. CE2, a lymphocyte deficiency CE, contributed to the least favorable patient survival outcome, followed by CE4 (myogenesis-associated CE). The remaining ecotype groups (CE1, CE3, CE5, CE7, and CE8) exhibited intermediate risk and did not show clear separations in their survival curves. The survival rates were further analyzed by grouping the ten identified ecotypes into three eco-groups: CE9 & CE10 (the proinflammatory group), CE1 & CE2 (the lymphocyte-deficient group), and CE3-8 (Others) (Fig. 3B). The Kaplan-Meier curves for the three groups show clear statistical separation ($p < 0.001$, log-rank test). Among the groups, CE9&10 exhibited the most favorable survival outcomes, whereas CE1&2 had the poorest. We also performed Kaplan-Meier analyses of ecotypes based on real-world progression-free survival (rwPFS), as shown in Supplementary

Fig. 4, which revealed a similar pattern to the OS outcome as shown in Fig. 3A, B.

We further performed a stratified survival analysis based on the ICI treatment regimen. As shown in Supplementary Fig. 5A, patients were categorized into two subgroups: those receiving anti-PD-(L)1-based regimens and those treated with anti-CTLA-4-based therapies, including both monotherapy and combination therapy. Despite the sample size limitation, particularly for anti-CTLA4 patients, we observed ecotype-outcome associations consistent with those shown in Fig. 3B across both subgroups. We further evaluated ecotype-outcome associations within two additional clinically relevant subgroups: HPV status in HNSCC and smoking status in NSCLC. In both analyses, the ecotype-based survival trends remained consistent with the overall findings (as shown in Supplementary Fig. 5B, C).

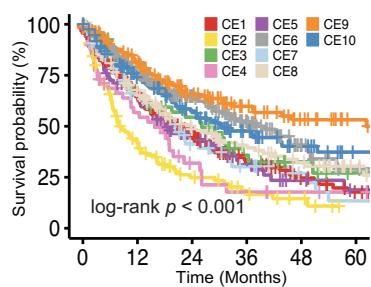
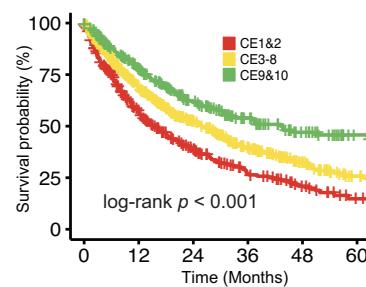
Interestingly, we found that CE6, classified as a normal-like ecotype in the original EcoTyper framework, was associated with favorable survival in our ICI-treated pan-cancer cohort (Fig. 3A). This finding contrasts with the original study, where CE6 showed variable associations with OS across TCGA cancer types and tended toward unfavorable outcomes in limited ICI-treated datasets. Several factors may explain this discrepancy. First, the CE6 survival trend in our cohort may reflect cancer-type composition, because nearly half of the CE6-assigned samples in our data come from kidney cancer, which generally exhibited better survival (Supplementary Fig. 6). Second, CE6 in our dataset was notably enriched among primary cases. These differences in sample source and clinical context may contribute to the different prognostic behavior of CE6.

To better illustrate the ecotype-specific survival patterns, in Fig. 3C, we additionally analyzed the population-level survival trajectories from two to five years for the four ecotypes (CE1, CE7, CE9, and CE10) that had the highest number of assigned patients among melanoma cases. As expected, there was a notable decrease in the number of patients experiencing no events (alive) over three years for CE1, while CE9, indicative of a favorable prognosis, shows a much smaller decrease. After that, we explored the prognostic implications of CD4 T cell states, which are crucial for immune regulation and the activation of cytotoxic T cells. It can be seen from the data in Fig. 3D that CD4-S01 and CD4-S02 exhibited significant positive survival outcomes in melanoma and H&N, consistently showing beneficial effects in the other two cancer types as well. Finally, CD4-S03 (marked by gene *GLYCTK*) was associated with poorer survival outcomes in melanoma, NSCLC, and bladder cancer, although the mechanisms underlying CD4-S03 remain to be elucidated.

Multi-ecotype prognostic signature

To develop and optimize a melanoma-specific prognostic EcoTyper model using the ORIEN ICI melanoma data, we performed regularized Cox regression with repeated subsampling (random elastic net), which ultimately selected five ecotypes: CE1, CE5, CE7, CE9, and CE10. The final melanoma-ICI patient risk score (**EcoRIS**) model was then constructed using a post-selection multivariable Cox regression as follows: $1.13 \times \text{CE1} + 1.49 \times \text{CE5} + 1.39 \times \text{CE7} - 2.11 \times \text{CE9} - 4.86 \times \text{CE10}$ (more details available in “Methods”). As an initial step in evaluating the performance of the EcoRIS prognostic signature, we conducted multivariable Cox regression analyses in the ORIEN melanoma ICI cohort, adjusting for key clinical covariates including age, sex, and sample type (primary vs. metastatic). As shown in Supplementary Fig. 7A, the EcoRIS signature remained significantly associated with both overall survival and rwPFS. No other clinical covariates were significant, as expected, but this multivariable analysis helped validate the observed associations against potential confounding effects.

The predictive accuracy of the EcoRIS model was assessed using two external melanoma datasets: E1609 (a phase III trial involving patients with resected cutaneous melanoma treated with ipilimumab or high-dose interferon alfa) study¹⁶ and a harmonized ICI datasets¹⁷⁻²⁰

A Pan-Cancer Ecotype**B Pan-Cancer Eco-group**

	CE 1&2 (n=323)	CE 9&10 (n=278)	CE 3-8 (n=653)
Events (n)	210	115	342
Median (95% CI), Months	15.74 (12.20, 19.93)	43.24 (30.80, NA)	26.10 (22.15, 30.54)
HR (95% CI)	2.17 (1.73, 2.73)	1.49 (1.21, 1.46)	
log rank p-value	2.62 × 10 ⁻¹¹	0.00022	

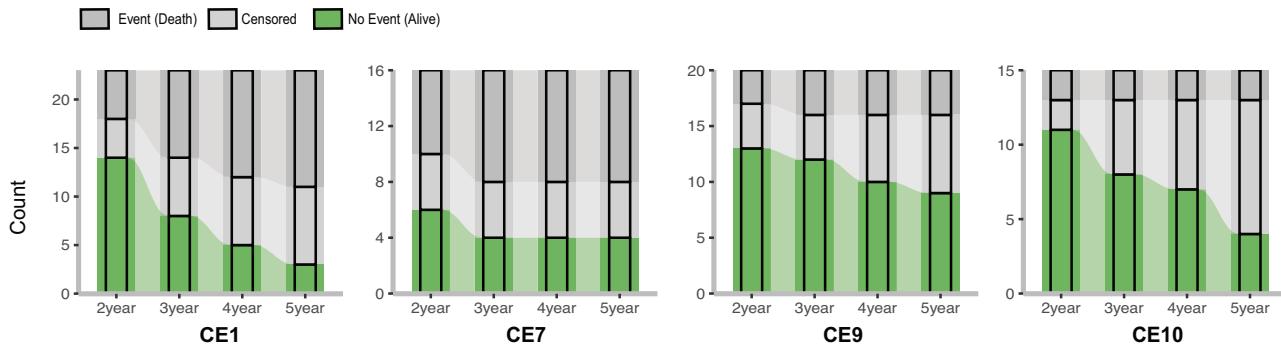
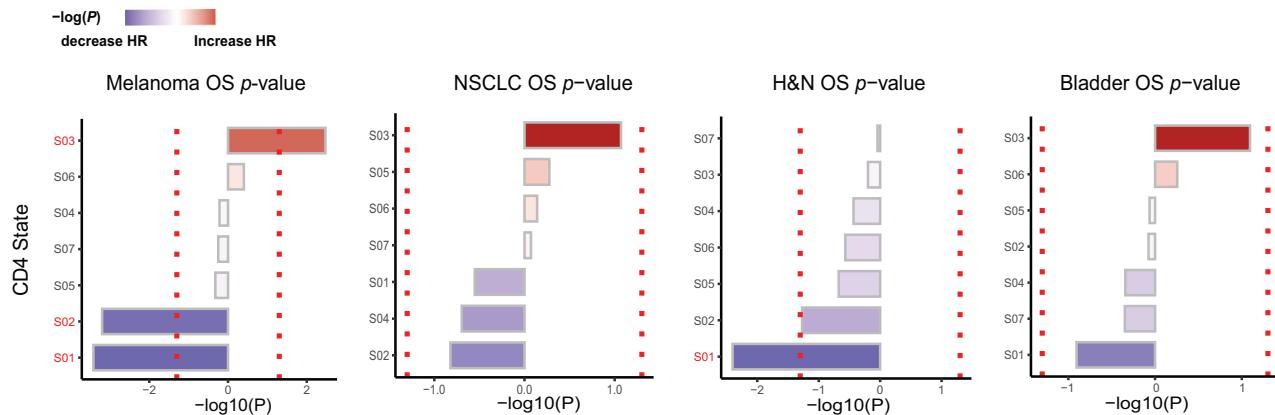
C Melanoma Ecotype survival**D CD4 cell state survival**

Fig. 3 | Associations between CEs, cell states, and survival across major cancer types. **A** Kaplan-Meier curves display patient overall survival across ten CEs for overall survival (OS) within the ORIEN ICI cohort ($n = 1254$) ($p = 2.2 \times 10^{-16}$, log-rank test) **(B)** Kaplan-Meier survival curves with regrouping ten CEs into three subgroups: CE1-CE2, CE9-CE10, and others (CE3-8) ($p = 6.2 \times 10^{-11}$, log-rank test). The risk table presents comparisons between CE 1&2 vs CE 9&10 and CE 3-8 vs CE 9&10, with CE 9&10 serving as the reference group. The table includes the number of patients at baseline, median survival in months with 95% confidence intervals (CI), hazard ratios (HR) with 95% CI, and associated pairwise log-rank p -values (Holm-

adjusted). **C** Bar plots showing survival event trends among melanoma patients over 2 to 5 years of treatment, grouped by CE1 ($n = 23$), CE7 ($n = 16$), CE9 ($n = 20$), and CE10 ($n = 15$). **D** Survival associations of CD4-specific cell states in melanoma ($n = 161$), NSCLC ($n = 240$), and H&N ($n = 270$), and bladder ($n = 152$), with color coding for better survival outcomes (blue bar) or worse survival outcomes (red bar). The x-axis represents $-\log(P)$ value from the univariable Cox regression, the Y-axis characterizes cell state (S), and red dotted lines indicate a p -value of 0.05. CE carcinoma ecotype, H&N head and neck cancer, HR hazard ratio, NSCLC non-small cell lung cancer OS overall survival.

comprising 334 patients with melanoma treated with ICIs. The prognostic value of the 5-CEs risk model across the entire E1609 melanoma cohort was presented in Fig. 4A. Stratification by the median of risk scores revealed that patients on either Ipi or HDI- α with lower predicted scores exhibited significantly better survival outcomes than those in the high-risk score group (log-rank test, $P < 0.001$). The same trend was observed in the subgroup of patients treated only with Ipi. In the multivariable Cox regression setting, as illustrated in the forest plots, only CE9 remained significant in the all-patients cohort and

Ipi-treated cohort [Hazard ratio (HR) (95% CI); 0.04 (0.00 to 0.45), 0.02 (0.00 to 0.44), respectively], suggesting its potential value as a protective single-CE predictor.

We examined two representative melanoma tissue samples from the E1609 cohort by retrieving the matched Hematoxylin and eosin stain (H&E) and immunohistochemistry (IHC)-stained slides, as presented in Fig. 4C. One sample had a low melanoma-ICI patient risk score, and another one was classified under CE9. Notably, both samples displayed structures of lymphoid aggregation, which are

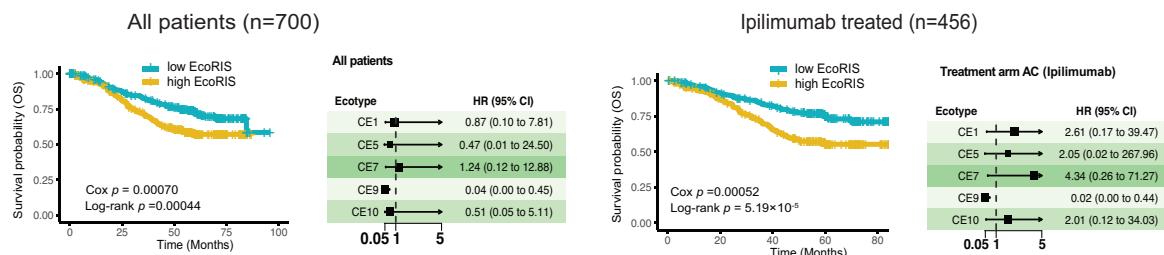
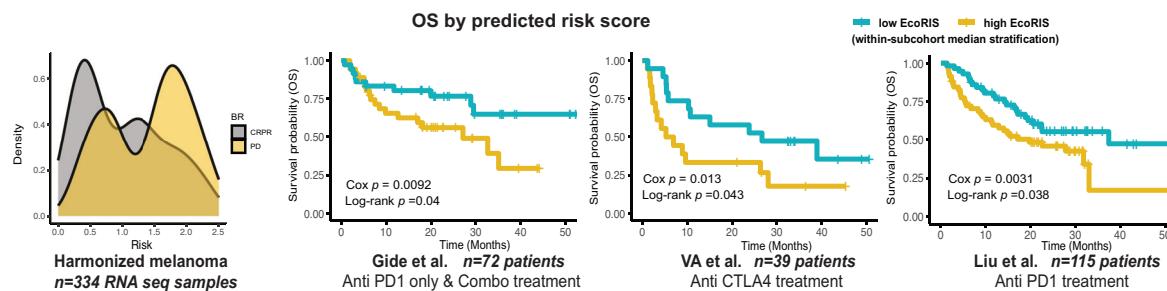
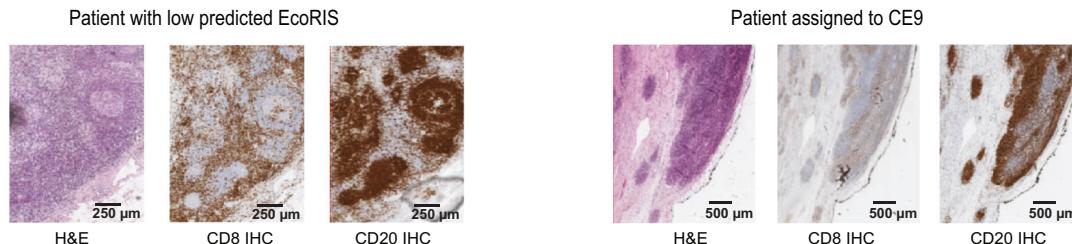
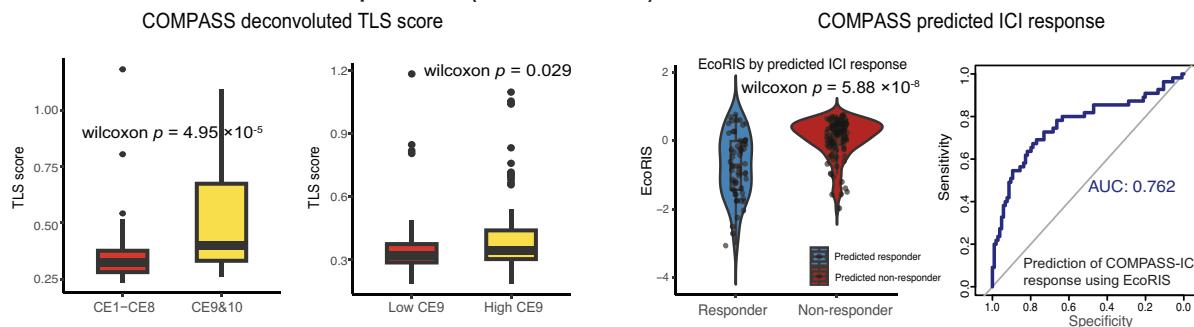
A Validation with E1609**B Validation with public melanoma dataset (Harmonized melanoma ICI patients)****C Representative examples of H&E and IHC-stained slides (E1609)****D COMPASS deconvolution & prediction (ORIEN IO cohort)**

Fig. 4 | Validation of EcoTyper risk scores (EcoRIS) on external melanoma datasets. **A** Prognostic validation of EcoRIS in the E1609 dataset (n = 700): Kaplan-Meier overall survival curves compare patients categorized into low- (n = 350) and high-EcoRIS (n = 350) scores (based on the median). The forest plot describes five component ecotypes associated with HR. Analyses were conducted on the entire E1609 cohort (left) and ipilimumab-treated sub-arm (n = 456) (right). **B** EcoRIS scores and its prognosis in the harmonized melanoma dataset (n = 334)²⁰. Kaplan-Meier curves for overall survival over 50 months with log-rank p-values reported in individual datasets: Gide et al. (n = 72)¹⁷, Van Allen et al. (n = 39)¹⁹, and Liu et al. (n = 115)¹⁸. Stratification is based on within-cohort median scores. CE carcinoma ecotype, CR/PR complete response (CR)/partial response (PR), CTLA4 cytotoxic T-lymphocyte associated protein 4, H&E Hematoxylin, and eosin stain, IHC immunohistochemistry, OS overall survival, PD progressive disease, PD1

programmed cell death protein 1. **C** Representative slides of melanoma tissue stained with H&E and IHC for markers CD8 and CD20, shown from left to right. Images represent a patient with a low predicted risk score and another assigned to CE9. **D** Association of COMPASS-predicted TLS across ecotypes and concordance with COMPASS-predicted treatment response with EcoRIS. The left panel of Fig. 4D shows box plots of COMPASS-predicted TLS across ecotypes ($p = 4.9 \times 10^{-5}$, Wilcoxon test, n = 78(CE-CE8) vs. 35(CE9&10); $p = 0.029$, Wilcoxon test, n = 79 (Low CE9 group) vs. 80 (High CE9 group)). The box plot in the right panel of Fig. 4D displays demonstrates the concordance between COMPASS-predicted treatment response and EcoRIS ($p = 5.88 \times 10^{-8}$, Wilcoxon test, n = 55 responders vs. n = 104 non-responders). In all box plots, the central line marks the median, the lower and upper hinges denote the first and third quartiles, and the whiskers extend to 1.5 × the interquartile range (IQR).

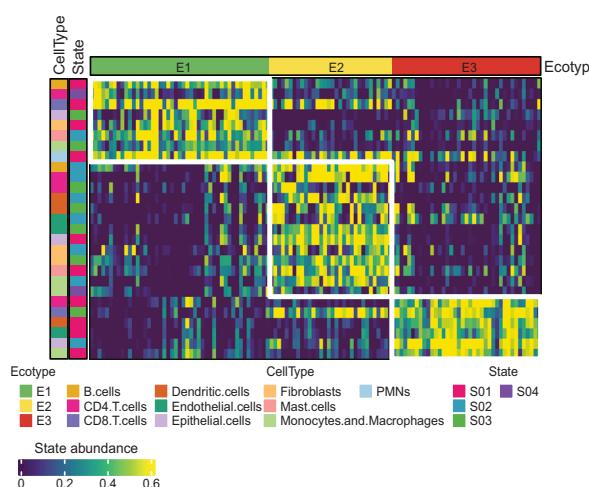
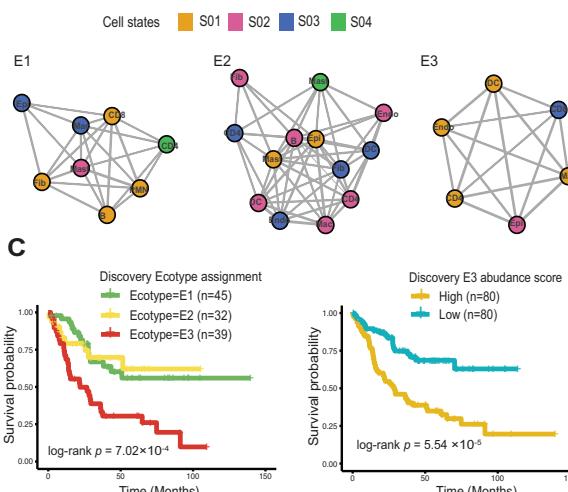
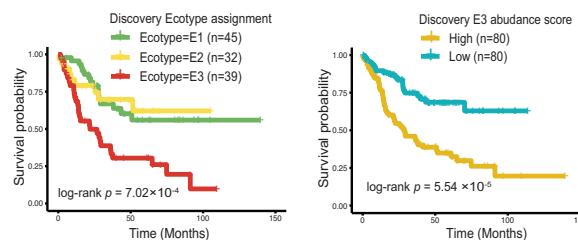
A

Fig. 5 | De-novo discovery and characterization of melanoma-specific cell states. **A** The heatmap displays the distribution of cell states and three de novo ecotypes among melanoma RNAseq samples within assigned ecotypes ($n=118$) in the ORIEN ICI cohort. Each column in the heatmap corresponds to RNA sequencing samples, and each row shows the expression of cell state marker genes within a recovered ecotype. **B** Diagrams illustrate co-occurrence patterns between cell

B**C**

types and states within the three patient groups with identified ecotypes. **C** Kaplan-Meier survival curves showing overall survival for three patient subgroups corresponding to the discovered ecotypes (E1 $n=45$, E2 $n=32$, and E3 $n=39$; left) and for patients stratified by E3 abundance score (low-E3 $n=80$ vs high-E3 $n=80$; right), with exact log-rank p -values reported in the plot—abbreviation: E ecotype; S cell state.

frequently associated with better ICI outcomes, representing potential tertiary lymphoid structures, as indicated by the IHC results for CD8 (CD8+ T cells) and CD20 (B cells). In the harmonized melanoma datasets^{17–20} (Fig. 4B), the patient with progressive disease (PD) displayed a bimodal risk score distribution, with the highest peak predominantly on the high-risk end. In the anti-PD-1-treated cohort from Gide et al.¹⁷, a significance separation in overall survival was observed between the low and high-risk groups. This trend was similarly observed in both in the VA et al.¹⁹ dataset, which included patients who received anti-CTLA4 therapy, and with Liu et al.¹⁸ cohort of patients receiving anti-PD1 therapy.

Because direct ICI response outcomes are unavailable in the ORIEN dataset, we leveraged COMPASS²¹, a foundation model for predicting ICI response from transcriptomic profiles. In our analysis, we deployed a multi-stage fine-tuned model developed on melanoma patients for pembrolizumab-prediction, to infer binary ICI response. As shown in Fig. 4D, the left panel demonstrates that the tertiary lymphoid structures (TLS) score inferred by COMPASS embedding was significantly elevated in tumors assigned to CE9 and CE10. We also observed an association between the TLS score and the corresponding CE9 ecotype score, suggesting enriched TLS-like activity in these proinflammatory ecotypes. In the right panel, EcoRIS scores were significantly higher in the COMPASS-predicted non-responder group, and the area under the curve (AUC) for using EcoRIS to predict response status was 0.762. These findings provide an additional layer of validation supporting the predictive value of the EcoRIS model.

Intriguingly, in the final EcoRIS model, CE10 carried a stronger regression coefficient than CE9 (-4.86 vs -2.11), despite CE9 being the most ICI-associated ecotype from the original EcoTyper study. To better understand this, we evaluated the univariable associations between individual ecotype fractions and overall survival in the ICI-treated melanoma cohort. As shown in Supplementary Fig. 7B, the results were consistent with the multivariable model and aligned with survival trends observed in Fig. 3C, where CE10 demonstrated a more pronounced predictive effect. These findings highlight a potentially underrecognized prognostic value for CE10 in real-world ICI-treated patients. One possible explanation is that CE10 represents a more naïve or recovering cytotoxic immune state, as supported by recent

evidence¹³, whereas CE9 is enriched for exhausted or dysfunctional CD8⁺ and CD4⁺ T cell populations. This distinction is especially relevant in real-world ICI cohorts, where most patients have received prior therapy. In addition, CE10 has been found more frequently localized in lymph nodes rather than within the tumor core, which could influence its prognostic behavior in metastatic diseases such as melanoma.

De novo ecotypes were discovered based on the ORIEN ICI dataset

As part of our exploratory analysis, we further performed the de novo discovery of cell states and ecotypes using the transcriptome data in the ORIEN ICI dataset. Based on CIBERSORTx, we first estimated the abundance of cell types and generated cell-type-specific gene expression profiles (GEPs) across nine cell types. Subsequently, employing the EcoTyper framework, we identified distinct transcriptional cell states within these profiles. Based on the ecotype discovery analysis pipeline, three distinct ecotypes were detected based on 27 different cell states. These newly identified ecotypes are labeled E1, E2, and E3, with 47, 32, and 39 RNAseq samples assigned to them, respectively (Fig. 5A). Ecotype E1 was characterized by CD4-S04, which was related to gene CTLA4 and B-S01, CD8-S01, and others. On the contrary, Ecotype E2 consisted of various cell states, with S02 being the dominant one. Finally, Ecotype E3 was identified by the co-occurrence of CD8-S03 and EPI-S02 (Fig. 5B).

The survival analysis in Fig. 5C indicates that patients categorized under Ecotype E3 experienced significantly poorer survival outcomes than those in E1 and E2 (log-rank test, $P < 0.0001$). In addition, patients with higher continuous E3 score (above the median) had significantly worse survival (log-rank test, $P < 0.0001$), while higher continuous scores for E1 and E2 were associated with better survival (Supplementary Fig. 8A). To further explore the transcriptional landscape of the de novo melanoma-specific ecotypes (E1–E3), we examined their relationship with previously defined pan-cancer ecotypes and associated cell states. As shown in Supplementary Fig. 8B, C, E3, associated with poor ICI outcomes, showed markedly reduced representation of CE9 and CE10, and was instead enriched for CE7. We also assessed the dominant CD4⁺ T cell states within each de novo ecotype and found clear distinctions: E1 was characterized by CD4-S04, a state expressing

genes that are reminiscent of immune-activated states; E2 was enriched for CD4-S02 (naïve/central memory), and E3 for CD4-S01 (exhausted/effector memory). These CD4⁺ T cell compositions may partly explain the prognostic differences among the de novo ecotypes and further support their biological relevance in ICI-treated melanoma. Overall, the distinct profile of Ecotype E3 could complement established prognostic ecotypes such as CE9, enhancing the profiling of immune interactions from derived real-world ICI data.

Discussion

In this study, we utilized the EcoTyper framework and ORIEN Avatar ICI cohort data to investigate how immune cell states and tumor ecotypes influence response to immunotherapies in a real-world setting. Public datasets such as TCGA were valuable for training EcoTyper but were based on samples selected for high tumor purity and diverse cancer types²², limiting their generalizability due to selective inclusion criteria and outdated treatment modalities. In the original EcoTyper study, the prognostic relevance of ecotypes for ICI outcomes was tested only in limited validation cohorts, such as IMvigor, a bladder cancer study of patients receiving anti-PD-L1 therapy¹⁰. In contrast, the ORIEN Avatar cohort offers a much larger sample size with high-quality molecular and clinical data collected across multiple cancer centers. Leveraging this resource allowed us to evaluate the prognostic value of ecotypes in an independent, real-world setting and generate clinically relevant insights into immunotherapy response.

In the current study, the prognostic value of 10 ecotypes at the pan-cancer level was evaluated and subsequently categorized into three subgroups: an immune-deficient group (CE1 and CE2), a proinflammatory group (CE9 and CE10), and an intermediate-risk group (CE3-8). As expected, the immune-deficient group was linked to shorter survival, whereas the proinflammatory group showed significantly better outcomes. These results indicate that ecotype-based prognostic signatures exist at a pan-cancer level. In cancer-type-specific analyses, the prognostic strength of the ecotype-based signature was most evident in melanoma and non-small cell lung cancer (NSCLC), followed by head and neck cancer and bladder cancer (Supplementary Fig. 4). In contrast, the prognostic association was weaker in kidney cancer, possibly reflecting its distinct tumor microenvironment or the distribution of ecotypes in this disease.

We constructed a composite prognostic risk model, termed EcoRIS, in the melanoma cohort using ecotype abundance scores. Because over 20% of patients were not assigned to specific ecotypes, the model incorporates all samples through continuous scores. This approach is advantageous as it preserves more information even for ambiguous ecotypes like CE5 and CE8, which share similar cell states. The final model selected CE1, CE5, CE7, CE9, and CE10. Among these, CE1 and CE7 were notably associated with TGF beta signaling. CE1 exhibited the strongest positive association with UV response downregulation and hypoxia, while CE9 demonstrated significant negative associations with both pathways. CE1 and CE9 were also the ones to which most melanoma patients were assigned. To account for clinical heterogeneity, we performed multivariable Cox regression analyses adjusting for key covariates, including age, sex, and tumor sample type. The EcoRIS signature remained significantly associated with both OS and real-world PFS, and adjustment for clinical covariates, although not individually significant, confirmed the robustness of these associations.

Two notable differences emerged when comparing our findings with the original EcoTyper signatures. First, CE6 was unexpectedly associated with favorable survival in our ICI-treated pan-cancer cohort, where the original study reported variable and often unfavorable associations with prognosis in non-ICI contexts. In our analysis, nearly half of the CE6-assigned tumors originated from kidney cancer, contributing to better prognosis. CE6 was also more common in metastatic or recurrent tumors, suggesting cancer type prevalence and

tissue origin may shape its prognostic behavior. These observations highlight the context-dependent performance of ecotypes in real-world ICI-treated populations. Second, in our EcoRIS model, CE10 showed a stronger protective effect than CE9, despite CE9 being the ICI-responsive in the original EcoTyper. One explanation is that CE10 may reflect a naïve cytotoxic immune state, whereas CE9 is enriched for exhausted or dysfunctional immune populations. This distinction may be especially relevant in real-world data where prior therapies alter the immune landscape. In summary, CE9 and CE10 remained highly immunogenic, yet their contributions to favorable outcomes vary by context, warranting further mechanistic investigation.

We successfully validated the risk model for using an independent cohort from the E1609 phase III trial¹⁶ in melanoma and a harmonized cohort from three independent ICI studies^{17,19,23}, confirming that lower predicted risk scores are associated with improved survival and better response to immunotherapy. In the E1609 cohort, H&E slides available from a subset of patients with low EcoRIS scores and CE9 assignment showed possible lymphoid aggregation, which may indicate enrichment of immune-related functions at the spatial level. In the data with therapy response, patients with higher risk scores were more likely to progress.

The predictive potential of major cell states and communities across all major cancer types was also investigated. Unlike the extensively studied CD8 T cells and B cell states, which largely define immune hot or cold states, CD4⁺ T cells exhibit a broader spectrum, with more than half of their states lacking clear mechanisms. Survival analysis showed that the CD4-S01 and CD4-S03 were the most consistent states across cancers, with S01 associated with improved survival and S03 with poorer survival. These findings suggest that CD4⁺ T-cell states may serve as a complementary marker to established ecotypes and highlight the need for deeper functional analyses.

In addition to recovered cell states and ecotypes, we performed de novo discovery using the melanoma ORIEN ICI data and identified three melanoma-specific ecotypes: E1, E2, and E3. E1 was characterized by CD4-S04, associated with the immune checkpoint gene *CTLA4*, together with B-S01 and CD8-S01, and was linked to favorable survival outcomes. E3, identified by the co-occurrence of CD8-S03 and EPI-S02, was strongly associated with shorter survival. E2 contained a diverse mixture of cell states but showed limited prognostic relevance compared with E1 and E3. The major immune cell states and their signature genes for these three de novo ecotypes are shown in Supplementary Fig. 9. Overall, the distinct profile of E3 may complement established prognostic ecotypes such as CE9, offering additional resolution for evaluating tumor immune functional status.

This study has several limitations. First, the real-world nature of the ORIEN ICI cohort makes it challenging to obtain categorized RECIST responses and comprehensive details on all prior and ongoing treatments, which may impact our findings. Second, different therapy combinations, such as neoadjuvant chemotherapy regimens, may introduce bias and should be further explored in larger cohorts. Third, in calculating survival outcomes, we used the start date of the first ICI medication, which does not account for the duration or effects of prior therapies, although it aligns with conventions used in ICI clinical trials. Fourth, in real-world datasets, the timing of the treatment relative to the biospecimen is often ambiguous, particularly in patients with complex treatment histories or multiple lines of therapy. This limitation complicates the interpretation of transcriptional profiles and ecotype assignments. Finally, differences between primary and metastatic tumor samples may also influence immune profiles and ecotype classifications. The limited number of matched primary-metastatic pairs restricted our ability to systematically assess this effect. The melanoma cohort used to develop EcoRIS included a majority of metastatic samples, which may affect the model's generalizability to cohorts with different sample distributions.

In conclusion, our findings support the use of cell state co-existence patterns, identified through ecotype-based profiling, as biomarkers for predicting immunotherapy outcomes. Leveraging the scale and diversity of the ORIEN ICI cohort, we systematically evaluated the prognostic significance of established ecotypes in a setting that reflects routine clinical practice, an element largely absent from prior studies. Future work integrating ecotypes with emerging spatial omics data may improve tissue specificity and spatial resolution, allowing more precise characterization of immune niches and resistance mechanisms to guide targeted drug development.

Methods

Ethics statement and patient Cohort

This research incorporated a retrospective examination of clinical data and gene expression profiles from consenting patients collected through the Total Cancer Care® (TCC) Protocol (NCT03977402) and Avatar® project conducted within the Oncology Research Information Exchange Network (ORIEN), which includes 18 collaborating cancer centers participating in TCC^{24,25}. Subjects participating in TCC provided written informed consent to allow the use of their tumor and blood biospecimens for genomic and transcriptomic analyses and corresponding clinical data as part of the standard clinical practice to manage their disease. The study was conducted in accordance with the ethical standards of the Declaration of Helsinki, along with approval by the Institutional Review Board (IRB) at each participating institution (Advara IRB # Pro00014441). The ORIEN Avatar cohort in our study consisted of 14,997 individuals, of whom 1610 patients were treated with ICIs and constituted our target population. The ICI medication group included Ipilimumab, Nivolumab, Dostarlimab, Pembrolizumab, Avelumab, Atezolizumab, Cemiplimab, and Durvalumab. The supplementary materials provide additional information on the ICI cohorts used for validation. In addition, part of the validation analyses was performed using the ECOG-ACRIN E1609 cohort (NCT01274338), a randomized phase III study for patients with resected high-risk melanoma.

Overall survival and real-world progression free survival

The primary outcome in the analysis of the ORIEN Avatar cohort was overall survival (OS), defined from the initiation of ICI medication to death. We also evaluated real-world progression-free survival (rwPFS) as a secondary outcome. Progression events in rwPFS were defined as: annotated progression/recurrence in clinical records, annotation of therapy stopped due to progression, identification of new metastases, or death, with right censorship at the date of last contact for patients without a progression event.

RNA Sequencing

The procedure for RNA sequencing for the ORIEN Avatar project was conducted according to methodologies outlined in a white paper previously released (<https://www.asterinsights.com/white-paper/renal-cell-carcinoma-rwd-data/>). The necessary data on RNA expression were sourced from the ORIEN database, necessitating the download of multiple FASTQ files for further examination. Quantifying gene expression involved multiple technical steps. The initial phase employed Bbduk software (version 38.96) to remove adapter sequences from RNA-seq reads, available at <https://sourceforge.net/projects/bbmap/>²⁶. This was followed by aligning the trimmed reads to the human reference genome (GRCh38/hg38) using STAR software (version 2.7.3a), accessible at <https://github.com/alexdobin/STAR>²⁷. The integrity of the RNA samples was evaluated using the RNA-Seq Quality Control (RNA-SeQC) software (version 2.3.2), found at <https://github.com/getzlab/rnaseqc>²⁸. The computation of gene expression levels was performed using the Transcripts Per Million (TPM) metric, following alignment with the GeneCode build version 32 reference annotation through the RNA-Seq by Expectation Maximization (RSEM)

software (version 1.3.1), available at <https://github.com/deweylab/RSEM>²³. For the purposes of analysis, TPM values were transformed to a logarithmic scale ($\log_2 [TPM + 1]$). Any batch effects were adjusted using the ComBat method in the sva package (version 3.34.0), <https://doi.org/10.18129/B9.bioc.sva>²⁹. The original RNA sequencing data are available upon request as outlined in Data Availability.

EcoTyper framework

EcoTyper was conducted to deconvolute the RNA-seq data from the ORIEN ICI cohort and investigate cell states and ecotypes. Using our bulk RNA-seq data, cell states were first identified and categorized within each cell type, which was identified in the original discovery pan-cancer TCGA dataset by Luca et al.¹⁰ Subsequently, the algorithm establishes CS co-occurrence models to define cellular communities referred to as ecosystems. In this study, we focused on assessing the prognostic significance of 10 CEs that have been previously defined. The identified prognostic CEs in the ORIEN melanoma cohort were then independently validated in external cohorts of patients with melanoma treated with ICI¹⁶⁻¹⁹.

Statistical analyses

Our statistical analysis plan involved the following steps. Firstly, it began with a univariable Cox regression analysis to assess the individual prognostic significance of ecotype-groups. The associations between each eco-group and OS as survival outcomes were quantified using Cox regression coefficients and corresponding *p*-values. Secondly, multivariate Cox regression analysis was conducted to assess the independent prognostic value of the selected eco-groups while adjusting for potential confounders. Thirdly, to build the ecotype-based risk model (EcoRIS), we used regularized Cox regression for survival data as implemented in the glmnet package. In the first step, we performed 100 rounds of random resampling with an 80:20 training-to-test split to identify the optimal elastic net mixing parameter (alpha), which controls the balance between the LASSO and ridge penalties. In the second step, using the selected alpha, we applied cross-validation via cv.glmnet to determine the optimal regularization parameter (lambda). The final set of ecotypes was selected based on their selection frequency across the 100 runs of regularized Cox regression on random sampled 80% subsets, using the optimized alpha and lambda values. Following feature selection, a standard Cox proportional hazards regression model was then refit using top-selected ecotypes on the full dataset to estimate their coefficients and construct the EcoRIS model for the ICI risk prediction. Fourthly, model performance was quantitatively assessed using standard survival analysis metrics, including Harrell's concordance index (C-index). Confidence intervals (CIs) and hypothesis testing were used to assess the statistical significance of model improvements compared to baseline or competing models. Finally, cross-validation techniques, such as k-fold cross-validation, were employed to estimate the generalization error of the prognostic model and fine-tune model hyperparameters for optimal performance. All statistical analyses were performed using R version 4.2.3 and relevant packages.

COMPASS immunotherapy foundation model

To infer ICI response in the absence of direct clinical outcomes in the ORIEN dataset, we utilized COMPASS²¹, a recently developed foundation model for transcriptome-based prediction of immunotherapy response. For our analysis, we applied the MSFT-Pembro model, a version of the COMPASS pre-trained model that was multi-stage fine-tuned specifically for pembrolizumab response prediction in melanoma. This model was trained in three stages: (1) self-supervised pre-training on tumors from TCGA to learn pan-cancer transcriptome representations; (2) fine-tuning on all ICI-treated cohorts excluding the target datasets (pembrolizumab-treated melanoma); and (3) final model tuning using 120 pembrolizumab-treated melanoma samples.

Normalized gene expression values (TPM) of 15,672 protein-coding genes were used as input, and the model output a response probability for each ORIEN-ICI melanoma patient. We then derived a binary response prediction by selecting the class (responder vs. non-responder) with the higher predicted probability for each patient. In addition to the predicted ICI response, COMPASS produces a 44-dimensional latent embedding (concept space) that captures key immune-related signals, including T cell exhaustion, cytotoxic T cells, TGF- β signaling, and tertiary lymphoid structures (TLS). In our study, we leveraged these outputs to access the association between TLS and ecotypes, and to examine the concordance between the COMPASS-predicted ICI response and the EcoRIS signature we developed.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The ORIEN ICI RNA-seq and clinical dataset, which was previously generated, is available under restricted access through a closed instance of cBioPortal; access requests can be submitted via the affiliated institutional ORIEN Program Managers (OPMs) via <https://researchdatarequest.orienavatar.com> or by contacting ResearchDataRequest@AsterInsights.com. The E1609 data sets analyzed in the present study will be made available in the National Clinical Trials Network/ NCI Community Oncology Research Program (NCTN/ NCORP) Data Archive (<https://nctn-data-archive.nci.nih.gov>; RRID: SCR_014708). The harmonized melanoma ICI data are available at <https://github.com/ParkerICI/MORRISON-1-public>. The multi-stage fine-tuned COMPASS model used in this study is available at <https://www.immuno-compass.com/download/> under the name “Pembrolizumab Model”. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided in this paper.

Code availability

No custom code or software were used in this study.

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Gotwals, P. et al. Prospects for combining targeted and conventional cancer therapy with immunotherapy. *Nat. Rev. Cancer* **17**, 286–301 (2017).
3. Hegde, P. S., Karanikas, V. & Evers, S. The where, the when, and the how of immune monitoring for cancer immunotherapies in the era of checkpoint inhibition. *Clin. Cancer Res.* **22**, 1865–1874 (2016).
4. Chevrier, S. et al. An immune atlas of clear cell renal cell carcinoma. *Cell* **169**, 736–749 (2017).
5. Horning, S. J. A new cancer ecosystem. *Science* **355**, 1103 (2017).
6. Chen, X. & Song, E. The theory of tumor ecosystem. *Cancer Commun.* **42**, 587–608 (2022).
7. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
8. Lee, J. H. et al. Transcriptional downregulation of MHC class I and melanoma de-differentiation in resistance to PD-1 inhibition. *Nat. Commun.* **11**, 1897 (2020).
9. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
10. Luca, B. A. et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* **184**, 5482–5496 (2021).
11. Rolong, A., Chen, B. & Lau, K. S. Deciphering the cancer microenvironment from bulk data with EcoTyper. *Cell* **184**, 5306–5308 (2021).
12. Li, Y. et al. Clinical and molecular characterizations of HNSCC patients with occult lymph node metastasis. *Sci. Rep.* **15**, 25263 (2025).
13. Rajagopal, P. S. et al. Population-specific patterns in assessing molecular subtypes of young black females with triple-negative breast cancer. *NPJ Breast Cancer* **11**, 1–9 (2025).
14. Moreno, C. S. et al. Integrated genomic analysis of primary prostate tumor foci and corresponding lymph node metastases identifies mutations and pathways associated with metastasis. *Cancers* **15**, 5671 (2023).
15. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
16. Tarhini, A. A. et al. Phase III study of adjuvant Ipilimumab (3 or 10 mg/kg) versus high-dose interferon Alfa-2b for resected high-risk melanoma: North American intergroup E1609. *J. Clin. Oncol.* **38**, 567–575 (2020).
17. Gide, T. N. et al. Distinct immune cell populations define response to anti-PD-1 monotherapy and anti-PD-1/anti-CTLA-4 combined therapy. *Cancer Cell* **35**, 238–255 (2019).
18. Liu, D. et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916–1927 (2019).
19. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
20. Campbell, K. M. et al. Prior anti-CTLA-4 therapy impacts molecular characteristics associated with anti-PD-1 response in advanced melanoma. *Cancer Cell* **41**, 791–806 (2023).
21. Shen, W. et al. Generalizable AI predicts immunotherapy outcomes across cancers and treatments. *medRxiv*, <https://doi.org/10.1101/2025.05.01.25326820> (2025).
22. Cancer Genome Atlas, N. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
23. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
24. Fenstermacher, D. A., Wenham, R. M., Rollison, D. E. & Dalton, W. S. Implementing personalized medicine in a cancer center. *Cancer J.* **17**, 528–536 (2011).
25. Caligiuri, M. A., Dalton, W. S., Rodriguez, L., Sellers, T. & Willman, C. L. *Orien. Oncol. Issues* **31**, 62–66 (2016).
26. Bushnell, B., Rood, J. & Singer, E. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056 (2017).
27. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
28. DeLuca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
29. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Bioinformatics* **8**, 118–127 (2007).

Acknowledgements

We thank the patients, their family members, as well as research staff supporting the conduct of the Total Cancer Care protocol (NCT03977402). This work was supported in part by ORIEN NOVA Team Science Award (to A.T.); NIH grant R01DE030493 (to X.W.); and the Biostatistics and Bioinformatics Shared Resource (BBSR) at the H. Lee Moffitt Cancer Center & Research Institute, an NCI-designated Comprehensive Cancer Center (P30-CA076292). ECOG-ACRIN E1609 trial used in validation was supported in part by the NCI under award numbers: U10CA180820, U10CA180794, UG1CA233180, UG1CA233331, UG1CA233178, UG1CA233324, UG1CA233191, U10CA180888,

UG1CA180821, and UG1CA233184. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

Conceptualization, X.W. and A.T.; Data Curation, X.W., T.L., M.R., P.A., and A.T.; Formal analysis, T.L. and X.W.; Software, A.S and T.L.; Visualization, X.W. and T.L.; intellectual contributions to study design and review of results, V.S., A.T., M.M., B.S., H.C., A.R.N., I.P., S.A., J.A.M-A, S.G., M.C., P.H., P.C.R, W.S.D., G.J.W., and A.T.; Writing-original draft, T.L., X.W., A.T., and I.E.; Funding Acquisition: A.T.; All the authors reviewed and edited the manuscript.

Competing interests

VS declares receiving research funding from Eli Lilly and Company and Crinetics Pharmaceuticals. Consulting for GE Health. Advisory Board member of Exelixis. Receiving speaking fees from Lantheus Pharmaceuticals. MM declares contracted research grants with Merck, Taiho, and the National Comprehensive Cancer Network. IP declares Program Leader (1.2 cal months) for NIH/NCI Cancer Center Support Grant (P30CA016056); consulting with Nouscom, lovance, Nektar, and Regeneron; stock owner with Ideaya, Inc. HC declares advisory board/consultant: Best Doctors/Teladoc, Orbus Therapeutics, Bristol Myers Squibb, Regeneron, Novocure, and PPD/Chimerix; research funding (site PI/institutional contract): Orbus, GCAR, Array BioPharma, Karyopharm Therapeutics, Nuvation Bio, Bayer, Bristol Myers Squibb, and Sumitomo Dainippon Pharma. To date, S.M.A. is a member of ASCO COI. A.A.T. declares contracted research grants with institutions from Bristol Myers Squibb, Genentech-Roche, Regeneron, Sanofi-Genzyme, Nektar, Clinigen, Merck, Acrotech, Pfizer, Checkmate, OncoSec, Scholar Rock, InflaRx GmbH, Agenus; personal consultant/advisory board fees from Bristol Myers Squibb, Merck, Easai, Instil Bio, Clinigen, Regeneron, Sanofi-Genzyme, Novartis, Partner Therapeutics, Genentech/Roche, BioNTech, Concert AI, AstraZeneca outside the submitted work. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-65016-3>.

Correspondence and requests for materials should be addressed to Xuefeng Wang or Ahmad A. Tarhini.

Peer review information *Nature Communications* thanks Leng Han, who co-reviewed with Yuan Liu, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. ²Departments of Cutaneous Oncology and Immunology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. ³Aster Insights, Hudson, FL, USA. ⁴Department of Internal Medicine, Ohio State University and Arthur G James Comprehensive Cancer Center, Columbus, OH, USA. ⁵Center for Public Health Genomics, School of Medicine, University of Virginia, Charlottesville, VA, USA. ⁶Department of Surgery, University of Colorado Cancer Center, Aurora, CO, USA. ⁷Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. ⁸Department of Neurosurgery, School of Medicine, University of Utah, Salt Lake City, UT, USA. ⁹Division of Medical Oncology, Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA. ¹⁰Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA. ¹¹Department of Medical Oncology, University of Kentucky Markey Cancer Center, Lexington, KY, USA. ¹²Medical Oncology, Indiana University Melvin and Bren Simon Comprehensive Cancer Center, Indianapolis, IN, USA. ¹³Rutgers Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, NJ, USA. ¹⁴H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. ¹⁵Department of Immunology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA. ¹⁶Department of Internal Medicine, Carver College of Medicine, University of Iowa Health Care, Iowa City, IA, USA.  e-mail: Xuefeng.Wang@moffitt.org; Ahmad.Tarhini@moffitt.org