Aster Insights™

www.AsterInsights.com

APRIL, 2023

WHITE PAPER

INDEPENDENT CONFIRMATION OF THE PROGNOSTIC SIGNIFICANCE OF AN IMMUNE-RELATED SIGNATURE IN CLEAR CELL RENAL CELL CARCINOMA UTILIZING REAL WORLD CLINICOGENOMIC DATA

Authors:

Michael Radmacher, PhD

Daniel Elgort, PhD

Oliver A. Hampton, PhD, MS

Phaedra Agius, PhD

ABSTRACT

We utilized Aster Insights' clinicogenomic database to assess the prognostic significance of a signature of six immune-related genes in clear cell renal cell carcinoma first reported by Zhou et al in 2022. Our validation set was comprised of 1,060 clear cell renal cell carcinoma patients, each of whom had RNA-seq profiling of a sample collected from the primary tumor and associated longitudinal clinical data. In our validation set a risk score derived from the six-gene signature was an independent prognostic factor for overall survival in both univariate and multivariate modeling, the latter of which also considered age, sex, and stage, confirming findings in the original report. We were unable to definitively confirm the signature's previously reported association with response to immunotherapy. In a subset of 163 patients from our cohort who started immune-checkpoint inhibitor therapy within five years of RNA-seq, we observed no association between risk score and response to immune checkpoint inhibitor therapy. There was, however, a marginal association between the risk score and real-world progression-free survival. Our study demonstrates the utility of Aster Insights' real-world data for independent validation of reported gene expression signatures and their association with patient outcomes.

Introduction

The rapid increase in cancer-specific gene expression signatures and advancement of immunotherapy strategies in oncology is precipitating the need for clinicogenomic databases to allow for verification of published findings in additional datasets and to augment related discoveries in the field. Growing bodies of real-world data (RWD) present an opportunity in this direction. To demonstrate the utility of Aster Insights' real-world clinical and molecular data for such a purpose, we sought to confirm characteristics of a previously reported gene signature for clear cell renal cell carcinoma (ccRCC) in a larger, independent cohort.

ccRCC is the most common histological subtype of kidney cancer, accounting for approximately 75% of kidney cancer cases (Choueiri & Motzer, 2017). In recent years the use of immune checkpoint inhibitors (ICI) in the treatment of metastatic and advanced ccRCC has rapidly increased, resulting in improvement in clinical benefit and survival (Tung & Sahu, 2021). Zhou et al reported a six immune-related gene signature in ccRCC that was an independent prognostic factor for overall survival (OS) and significantly associated with response to ICI therapy (Zhou et al, 2022).

In this retrospective, observational study, we used Aster Insights' clinicogenomic data to evaluate characteristics of the reported signature. Aster Insights is a patient-first, oncology-focused health informatics solutions company that seeks to accelerate discovery, development, and delivery of more personalized therapies to improve patient outcomes. Aster Insights is well-positioned to support both academic research, by driving scientific collaborations and expanding cancer knowledge (Morris et al, 2022; Hoskins et al, 2023), and pharmaceutical research, by expediting time-to-market drug discovery (Ayers et al, 2019; Koomen et al, 2021). Aster Insights generates real-world, longitudinal clinical data together with molecular data (whole exome sequencing and RNAseq) in order to improve cancer care by accelerating precision medicine.

Results

Characteristics of ccRCC Cohort and Association with Risk Score

We sought to confirm characteristics of a previously reported immune-related signature in ccRCC using Aster Insights' AVATAR data. An Avatar is a representation of a particular patient's linked molecular and longitudinal clinical data in de-identified form (see Methods). Aster Insights' study cohort was comprised of 1060 Avatars diagnosed with ccRCC between 2004 and 2021 (Table 1). Median age at diagnosis was 61.1 years and the majority of Avatars were male (63%) and had stage III or IV disease (61% among those with non-missing stage).

TABLE 1. Clinical characteristics of full study cohort and stratified by risk group.

Variable*		Full Study Cohort (n = 1060)		Low Risk (n = 530)		High Risk (n = 530)		*p*-value
Age at Diagnosis	*years*	61.1	(21.6, 88.7)	60.8	(21.6, 88.7)	61.5	(24.7, 86.7)	0.27
Sex	*Male*	670	(63%)	297	(56%)	373	(70%)	<0.0001
Pathological Stage	*I*	249	(31%)	164	(43%)	85	(21%)	<0.0001
	II	62	(8%)	31	(8%)	31	(7%)	
	III	334	(42%)	154	(40%)	180	(43%)	
	IV	154	(19%)	36	(9%)	118	(29%)	
	Missing	*261*		*145*		*116*		
Risk Score		-11.2	(-93.3, 29.1)					

* Median and range are reported for continuous variables; counts and frequencies are reported for categorical variables.

The previously reported immune-related ccRCC signature is comprised of six genes (AIM2, CD5L, CSF1, HHLA2, IRF6, and TIMP3) (Zhou et al, 2022). The original study reported a formula for summarizing expression of these genes as a risk score (see Methods). We calculated the risk score for every Avatar in our study cohort and used the median value to categorize Avatars into low- and high-risk groups. Risk groups were similar in age, but the high-risk group was disproportionately male and was more likely to have late-stage disease (Table 1).

The relationship between relative expression of individual genes and the risk score was as expected based on the risk score formula with, for example, higher expression of AIM2 and CSF1 associated with higher risk score (Figure 1). The one exception is CD5L, which had very low expression across Avatars in our study cohort and was not correlated with risk score.

FIGURE 1

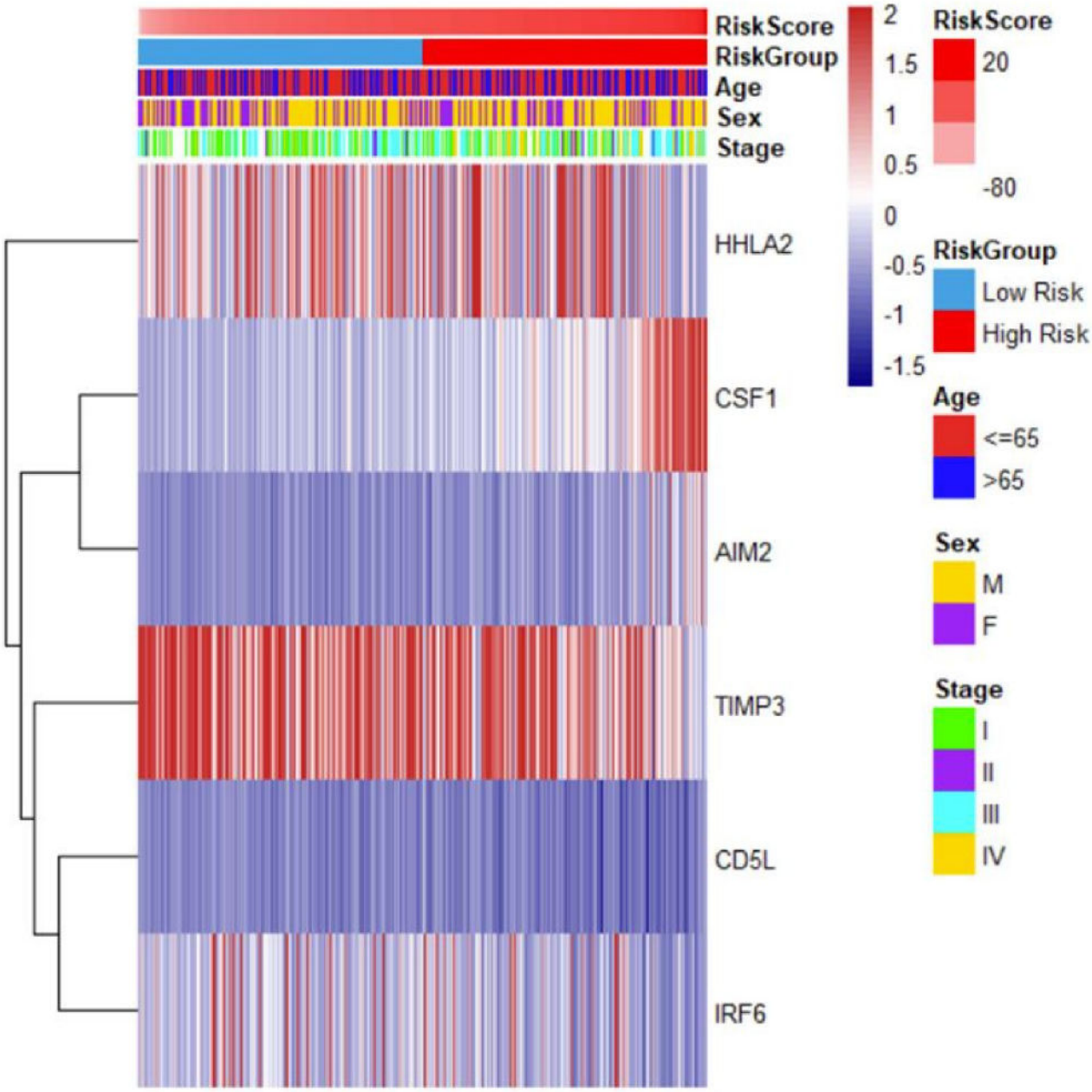

Figure 1: Heat map of relative expression of the six genes comprising the signature and risk score. Clinical characteristics are also displayed. Avatars are ordered left to right by increasing risk score. Expression values are scaled for each gene. Genes are clustered using Pearson's correlation as the similarity metric.

Association of Risk Score with OS

We observed a strong association between OS and the risk score groups, with poorer survival in the high-risk group ($p < 0.0001$; Figure 2A). Notably, this association was maintained in the subset of Avatars with stage III or IV disease but not in the subset with stage I or II disease (Figures 2B-C). Zhou et al. demonstrated the same pattern of prognostic associations in their study (Zhou et al, 2022).

FIGURE 2

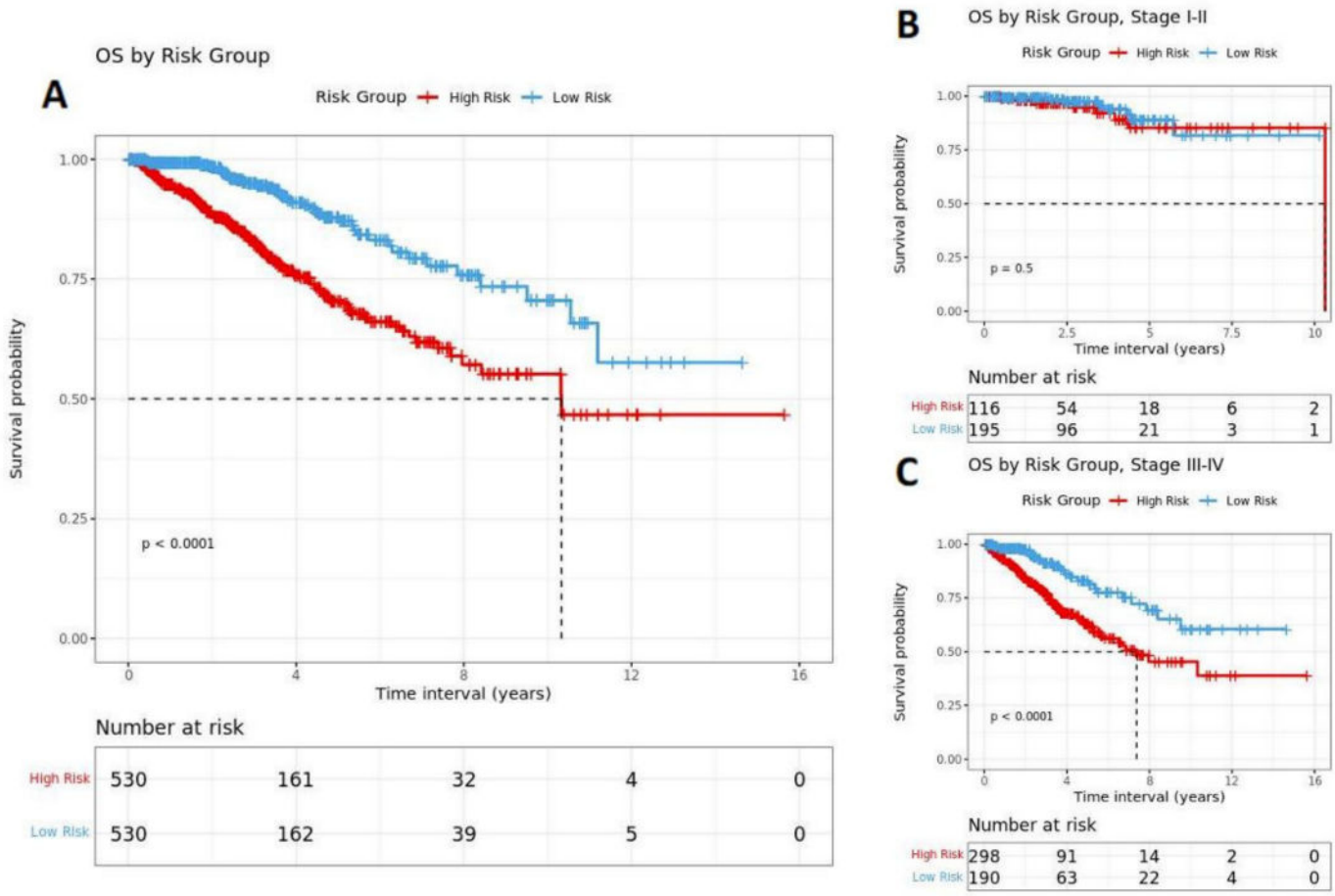

Figure 2: Overall survival by risk group in (A) the full study cohort, (B) the subset with stage I or II disease, and (C) the subset with stage III or IV disease.

To further investigate the prognostic significance of risk score in the context of other potentially prognostic clinical characteristics, we performed univariate and multivariate Cox regression analysis (Table 2). Age, pathological stage, and risk score were significant in univariate modeling; gender was not. In a multivariate model containing each of these factors, risk score retained its significance, confirming the risk score's independent prognostic significance. Zhou et al. reported the same associations in their study (Zhou et al, 2022).

TABLE 2. Cox regression analysis of risk score and clinical characteristics.

	Univariate Models			Multivariate Model		
	Hazard Ratio	**95% CI**	***p*-value**	**Hazard Ratio**	**95% CI**	***p*-value**
Age	1.027	(1.012, 1.042)	2.73E-04	1.024	(1.006, 1.042)	0.007
Sex (M)	1.249	(0.888, 1.756)	0.201	0.803	(0.548, 1.177)	0.261
Stage	1.981	(1.606, 2.444)	1.73E-10	1.729	(1.388, 2.155)	1.05E-06
Risk Score	1.054	(1.036, 1.073)	1.88E-09	1.046	(1.025, 1.068)	1.45E-05

Association of Risk Score with Response to ICI-Therapy and rwPFS in the ICI-Treated Subset

Zhou et al. demonstrated their risk score was associated with response to ICI therapy. Specifically, they applied their signature to 56 patient samples from the IMvigor210 cohort, a phase-II single-arm study investigating response to atezolizumab, and noted significantly lower risk scores among patients with a response (Zhou et al, 2022). To confirm this reported association with response to ICI, we identified a subset of 163 Avatars from our study cohort who had ICI-treatment within five years after RNA-seq specimen collection with available follow-up beyond initiation of treatment (Table 3). The most prevalent ICI therapy in this subset was nivolumab and ipilimumab received in combination. The ICI subset was more predominately male and more predominately comprised of late-stage disease than the full study cohort. The ICI subset was divided into low- and high-risk groups based on the median value of the risk score within the subset.

TABLE 3. Clinical characteristics for the ICI-treated subset.

Variable*		ICI-treated subset (n = 163)	
Age as Diagnosis	*years*	60.7	(24.7, 83.9)
Sex	*Male*	121	(74%)
Pathological Stage	*I*	9	(6%)
	II	5	(4%)
	III	55	(39%)
	IV	71	(51%)
	Missing	*23*	
First IO regimen	Atezolizumab	1	(1%)
	Ipilimumab	1	(1%)
	Pembrolizumab	30	(18%)
	Nivolumab	59	(36%)
	Nivolumab + Ipilimumab	72	(44%)
Risk Score		-5.5	(-45.3, 15.1)

* Median and range are reported for continuous variables; counts and frequencies are reported for categorical variables.

Using available clinical records in Aster Insights' database, we found favorable responses during time on ICI therapy for only nine Avatars. An additional 14 Avatars were on ICI therapy for at least two years; we considered these as favorable responses, too. Fifteen other Avatars were continuing ICI-treatment at last follow-up and had not yet been on therapy for two years; the response for these Avatars was treated as missing. All remaining Avatars (125) were considered non-responders, resulting in a response rate of 15.5%. There was no significant difference in response rate between risk groups (14.9% in low-risk vs. 16.2% in high-risk; p = 1).

Recognizing that positive responses to treatment may be under-represented in our clinical records, we also performed real-world progression-free survival (rwPFS) analysis, which does not rely on reporting of positive responses to treatment but, rather, the timing of a negative response (i.e., progression; see Methods for details of defining rwPFS). Risk group showed a marginal association with rwPFS ($p = 0.086$; Figure 3A). We speculated that risk score may vary over time; in such a case, scores measured closer to the time of ICI therapy initiation would likely be more relevant for predicting rwPFS. We therefore performed a subset comparison focused on Avatars who initiated ICI-therapy within 6 months of RNA-seq specimen collection. For the subset comparison, we still observed a marginal prognostic association between risk group and rwPFS ($p = 0.074$; Figure 3B).

FIGURE 3

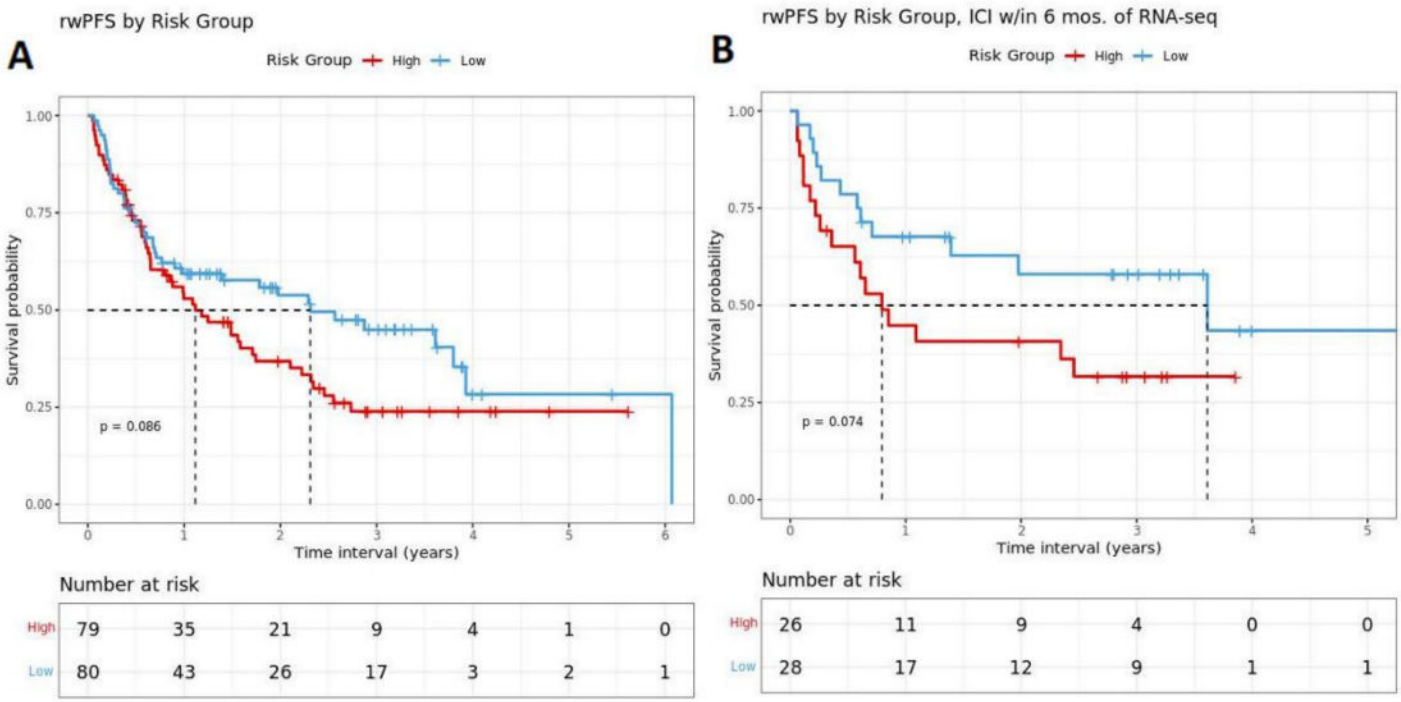

Figure 3. Real-world progression-free survival by risk group in (A) the ICI-treated subset and (B) the more restricted subset ICI-treated within 6 months of collection of the specimen used to calculate risk score.

Conclusions

We were able to recapitulate the major findings of the ccRCC gene expression signature first reported by Zhou et al, most notably a strong association with overall survival independent of other prognostic factors. While we were unable to confirm an association with clinical response to ICI therapy, our data set allowed for investigation of rwPFS endpoints and we observed a marginal association between risk group and rwPFS. Importantly, the marginal association with rwPFS was maintained and even slightly strengthened when we restricted analysis to patients who received ICI therapy within 6 months of RNA-seq, even though statistical power for that comparison was diminished by the reduction in sample size. This is an example of the necessity for careful consideration of timing of events and subset analyses and/or cohort stratification when using RWD. Our study demonstrates the utility of Aster Insights' real-world clinicogenomic data for independent validation of reported gene expression signatures and their association with patient outcomes.

The Aster Insights clinical data model comprises the following clinical data, as tables: (1) cytogenetic abnormalities, (2) diagnosis, (3) family history, (4) imaging, (5) labs, (6) medications, (7) metastatic disease, (8) outcomes, (9) patient history, (10) patient master, (11) physical assessment, (12) radiation, (13) stem cell transplant, (14) surgery biopsy, (15) tumor marker, (16) tumor marker flow panel, (17) tumor sequencing, (18) vital status and (19) a clinical molecular linkage file.

The Aster Insights molecular data model comprises the following: (1) tumor whole exome sequencing (bam and vcf) files, (2) normal whole exome sequencing (bam and vcf) files, (3) RNASeq expression (gene and transcript) files, (4) QC and metrics files, (5) MSI/TMB file, (6) CNV file, (7) gene fusion file (RNASeq-derived). Aster Insights will soon provide (8) Structural variants.

Aster Insights offers two strategic partnerships with the pharmaceutical industry: (1) data licensing, in which industry partners receive both clinical and molecular data or (2) SEARCH, in which Aster Insights' bioinformatics team partners with industry to answer specific inquiries with deliverances of reports (but not raw data).

Contact our Business Development team at Sales@AsterInsights.com for more information about our unique clinical and molecular dataset, and discover how we can help enable new insights for researchers and accelerate target identification and drug discovery using advanced patient cohorts.

Learn more at www.AsterInsights.com, and follow us on LinkedIn and Twitter.

Materials and Methods

Data Sources

We used data from Aster Insights' AVATAR database (February 2023 release) which regularly receives specimens and data submissions from its Oncology Research Information Exchange Network® (ORIEN) partners. ORIEN is a cancer precision medicine initiative initially developed by the Moffitt Cancer Center (Caligiuri et al, 2016; Fenstermacher et al, 2011). It has evolved into a consortium research network of 18 U.S. cancer centers. All ORIEN members utilize a standard protocol: Total Cancer Care® (TCC). TCC is a prospective cohort study with whole-exome tumor sequencing, RNA sequencing, germline sequencing, and lifetime follow up. Nationally, over 350,000 participants have enrolled. As part of the TCC study, participants agree to have their clinical data followed over time, to undergo germline and tumor somatic sequencing, and to be contacted in the future if an appropriate clinical trial becomes available (Dalton et al, 2018). Creation of Avatars occurs for a subset of TCC enrollees. Each Avatar is a representation of a particular patient's linked molecular and longitudinal clinical data in de-identified form. Avatars are commercially available from Aster Insights.

RNA Sequencing & Bioinformatics

Aster Insights' Avatar specimens undergo RNA extraction. For frozen and OCT tissue RNA extraction, QIAGEN RNeasy plus mini kit is performed, generating 216 base pair (bp) average insert size. For formalin-fixed paraffin-embedded (FFPE) tissue, Covaris Ultrasonication FFPE DNA/RNA kit is utilized to extract RNA, generating 165 bp average insert size. Aster Insights' RNA sequencing (RNA-seq) is performed using the Illumina TruSeq RNA Exome with single library hybridization, cDNA synthesis, library preparation, sequencing (100 bp paired reads at Hudson Alpha, 150 bp paired reads at Fulgent) to a coverage of 100M total reads / 50M paired reads. Raw reads of RNA-seq data are saved in a fastq format.

The RNA-seq analysis pipeline was developed by Aster Insights. First, adapter sequences are trimmed by Bbduk software using paired-end read option. STAR (v.2.7.3a, https://github.com/alexdobin/STAR) is then used to align reads on the GRCh38/hg38 human reference genome. Sequencing and alignment quality are assessed using RNA-SeQC 2 (v 2.4.2, https://github.com/getzlab/rnaseqc). Finally, RSEM (v1.3.1, https://github.com/deweylab/RSEM) is used to calculate expression values (TPM) at the gene level based on transcriptome alignment on the reference annotation of GeneCode build version 32.

RNA-seq data for our study cohort were adjusted for batch effects related to preservation method (e.g., formalin-fixed vs. frozen) using the ComBat algorithm of the sva package (v3.34.0, https://doi.org/doi:10.18129/B9.bioc.sva). TPM values were log2 transformed prior to ComBat normalization (a constant value of 1 was added before log transformation to avoid zeros as input to log-transformation) and exponentiated after normalization to revert to a linear scale.

Cohort Selection

Cohort eligibility criteria included: 1) a ccRCC diagnosis, 2) RNA-seq data available for analysis from the primary tumor, 3) RNA-seq specimen collected no more than 1 month prior to and no more than 1 year after the ccRCC diagnosis, 4) RNA-seq specimen collected prior to January 1, 2022, and 4) if the Avatar received immunotherapy then RNA-seq specimen collected prior to the initiation of immunotherapy. Criteria for a ccRCC diagnosis were a diagnosis of kidney cancer (topography code C64) with one of the following morphology codes: 8310/3, 8311/3, 8312/3, 8316/3, 8317/3, 8318/3, or 8319/3 (World Health Organization, 2013). In situations where an Avatar had more than one qualifying RNA-seq specimen, the earliest collected specimen was used.

A subset of the cohort receiving immune checkpoint inhibitor (ICI) treatment was evaluated for response to therapy. The eligibility criteria for this subset included: 1) initiation of therapy with one of the following ICIs after RNA-seq specimen collection: atezolizumab, ipilimumab, nivolumab, or pembrolizumab, 2) initiation of therapy within five years of RNA-seq specimen collection, 3) initiation of therapy prior to January 1, 2022, and 4) patient follow-up beyond the date of initiation of therapy.

Calculation of Risk Scores
The Zhou et al signature was comprised of six genes: CSF1, CD5L, AIM2, TIMP3, IRF6, and HHLA2 (Zhou et al, 2022). The authors developed a risk score, *r*, that is a linear combination of the expression value of these six genes, as follows:

$$r = 0.403 \times E_{CSF1} - 0.291 \times E_{CD5L} + 0.235 \times E_{AIM2} - 0.341 \times E_{TIMP3} - 0.188 \times E_{IRF6} - 0.206 \times E_{HHLA2}$$

where E*i* represents expression of gene *i*. We calculated this risk score for every Avatar in our study cohort, using ComBat-normalized TPM values for the measurement of gene expression. The risk score was used to form low- and high-risk groups defined by the median value of the risk score; continuous values of the risk score were also used for analysis as appropriate.

Outcome Measures
Primary study outcome measures were overall survival (OS) and, for the ICI-treated subset, response to therapy and real-world progression-free survival (rwPFS). We measured OS from the time of diagnosis.

For response to therapy, Avatars were categorized as responders if: 1) a response ("No evidence of Disease", "Complete Response/Remission (CR)", or "Objective Response (OR)") was noted in available clinical records during the duration of first ICI therapy or 2) the Avatar was on their first ICI therapy for at least two years. Avatars who were not categorized as responders and who were continuing ICI-treatment at last follow-up and had not yet been on therapy for two years were considered to have missing response status. All other Avatars were categorized as non-responders.

The validity of deriving progression-free survival endpoints from real-world data has recently been demonstrated (Griffith et al, 2019). For our implementation of rwPFS, the following were considered progression events: 1) an indication of progression in available clinical records ("No Response (NR)/Stable Disease (SD)", "Progression", "Progressive Disease (Prog/PD)", "Recurrence", "Regional Recurrence", or "Distant systemic recurrence of an invasive tumor only"), 2) an indication that the ICI therapy was stopped due to progression, 3) other evidence of metastatic disease, 4) an indication of therapeutic surgery in available clinical records, and 5) death. To filter out potential pseudo-progression events (Tenold et al, 2020), we excluded progression events that were subsequently followed by evidence of a response to the ICI therapy. Furthermore, we excluded from survival analysis any Avatars who had a progression event, including death, within two weeks of the initiation of ICI therapy (and who did not have record of a later positive response) as these events occurred before the treatment had a chance to be effective.

Statistical Analyses
Differences in a continuous covariate between groups were compared using the Wilcoxon rank sum, two-sample t, or Kruskal-Wallis test, as appropriate. Associations of a categorical covariate with group were determined by the chi-sqaure or Fisher's exact test, as appropriate. Differences in survival between groups were examined by Kaplan-Meier analysis and assessed by the log-rank test. To assess the prognostic significance of the risk score in the context of other potential prognostic factors, univariate and multivariate Cox regression analysis were performed using the survival package in R (v3.5-3).

R software version 3.6.1 was used for all statistical analyses. Two-sided p-values were calculated and p-values < 0.05 were considered statistically significant.

References

Ayers M, Nebozhyn M, Cristescu R, McClanahan TK, Perini R, Rubin E, Cheng JD, Kaufman DR, Loboda A. (2019). Molecular Profiling of Cohorts of Tumor Samples to Guide Clinical Development of Pembrolizumab as Monotherapy. Clinical Cancer Research, 25:1564-1573.

Caligiuri MA, Dalton WS, Rodriguez L, Sellers T, Willman CL. (2016). Orien: reshaping cancer research and treatment. Oncology Issues, 31(3):62-66.

Choueiri TK, Motzer RJ. (2017). Systemic therapy for metastatic renal-cell carcinoma. New England Journal of Medicine, 376:354–66.

Dalton WS, Sullivan D, Ecsedy J, Caligiuri MA. (2018). Patient Enrichment for Precision-Based Cancer Clinical Trials: Using Prospective Cohort Surveillance as an Approach to Improve Clinical Trials. Clinical Pharmacology & Therapeutics, 104(1):23-26.

Fenstermacher DA, Wenham RM, Rollison DE, Dalton WS. (2011). Implementing personalized medicine in a cancer center. The Cancer Journal, 17(6):528-536.

Griffith SD, Tucker M, Bowser B, Calkins G, Chang C-H. (2019). Generating Real-World Tumor Burden Endpoints from Electronic Health Record Data: Comparison of RECIST, Radiology-Anchored, and Clinician-Anchored Approaches for Abstracting Real-World Progression in Non-Small Cell Lung Cancer. Advances in Therapy, 36:2122-2136.

Hoskins EL, Samorodnitsky E, Wing MR, Reeser JW, Hopkins JF, Murugesan K, Kuang Z, Vella R, Stein L, Risch Z, Yu L, Adebola S, Paruchuri A, Carpten J, Chahoud J, Edge S, Kolesar J, McCarter M, Nepple KG, Reilley M, Scaife C, Tripathi A, Single N, Huang RSP, Albacker LA, Roychowdhury S. (2023). Pan-cancer Landscape of Programmed Death Ligand-1 and Programmed Death Ligand-2 Structural Variations. JCO Precision Oncology, 7:e2200300.

Koomen DC, Meads MB, Magaletti DM, Guingab-Cagmat JD, Oliveira PS, Fang B, Liu M, Welsh EA, Meke LE, Jiang Z, Hampton OA, Tungesvik A, De Avila G, Alugubelli RR, Nishihori T, Silva AS, Eschrich SA, Garrett TJ, Koomen JM, Shain KH. (2021). Metabolic Changes Are Associated with Melphalan Resistance in Multiple Myeloma. Journal of Proteome Research, 20(6):3134-3149.

Morris BB, Smith JP, Zhang Q, Jiang Z, Hampton OA, Churchman ML, Arnold SM, Owen DH, Gray JE, Dillon PM, Soliman HH, Stover DG, Colman H, Chakravarti A, Shain KH, Silva AS, Villano JL, Vogelbaum MA, Borges VF, Akerley WL, Gentzler RD, Hall RD, Matsen CB, Ulrich CM, Post AR, Nix DA, Singer EA, Larner JM, Stukenberg PT, Jones DR, Mayo MW. (2022). Replicative Instability Drives Cancer Progression. Biomolecules, 12(11):1570.

Tenold M, Ravi P, Kumar M, Bowman A, Hammers H, Choueiri TK, Lara PN. (2020). Current Approaches to the Treatment of Advanced or Metastatic Renal Cell Carcinoma. American Society of Clinical Oncology Educational Book, 187-196.

Tung I, Sahu A. (2021). Immune Checkpoint Inhibitor in First-Line Treatment of Metastatic Renal Cell Carcinoma: A Review of Current Evidence and Future Directions. Frontiers in Oncology, 11:707214.

World Health Organization. (2013). International Classification of Diseases in Oncology (ICD-O), 3rd ed.

Zhou L, Fang H, Yin M, Long H, Weng G. (2022). Novel immune-related signature based on immune cells for predicting prognosis and immunotherapy response in clear cell renal cell carcinoma. Journal of Clinical Laboratory Analysis, 36:e24409.