ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ANNALS OF ONCOLOGY driving innovation in oncology

ORIGINAL ARTICLE

Leveraging big data of immune checkpoint blockade response identifies novel potential targets

Y. Bareche[1,2†], D. Kelly[3†], F. Abbas-Aghababazadeh[4†], M. Nakano[4], P. N. Esfahani[4], D. Tkachuk[4], H. Mohammad[4], R. Samstein[5], C.-H. Lee[6], L. G. T. Morris[7], P. L. Bedard[3], B. Haibe-Kains[4,8,9,10,11,12,13*‡] & J. Stagg[1,2*‡]

[1]Faculty of Pharmacy, Université de Montréal, Montreal; [2]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Institut du Cancer de Montréal, Montreal; [3]Princess Margaret Cancer Centre, University Health Network, Division of Medical Oncology and Hematology, Toronto; [4]Princess Margaret Bioinformatics and Computational Genomics Laboratory, University Health Network, Toronto, Canada; [5]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York; [6]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York; [7]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, USA; [8]Princess Margaret Cancer Centre, University Health Network, Toronto; Departments of [9]Medical Biophysics; [10]Computer Science, University of Toronto, Toronto; [11]Ontario Institute for Cancer Research, Toronto; [12]Vector Institute for Artificial Intelligence, Toronto; [13]Biostatistics Division, Dalla Lana School of Public Health, Toronto, Canada

Available online 30 August 2022

Background: The development of immune checkpoint blockade (ICB) has changed the way we treat various cancers. While ICB produces durable survival benefits in a number of malignancies, a large proportion of treated patients do not derive clinical benefit. Recent clinical profiling studies have shed light on molecular features and mechanisms that modulate response to ICB. Nevertheless, none of these identified molecular features were investigated in large enough cohorts to be of clinical value.

Materials and methods: Literature review was carried out to identify relevant studies including clinical dataset of patients treated with ICB [anti-programmed cell death protein 1 (PD-1)/programmed death-ligand 1 (PD-L1), anti-cytotoxic T-lymphocyte antigen 4 (CTLA-4) or the combination] and available sequencing data. Tumor mutational burden (TMB) and 37 previously reported gene expression (GE) signatures were computed with respect to the original publication. Biomarker association with ICB response (IR) and survival (progression-free survival/overall survival) was investigated separately within each study and combined together for meta-analysis.

Results: We carried out a comparative meta-analysis of genomic and transcriptomic biomarkers of IRs in over 3600 patients across 12 tumor types and implemented an open-source web application (predictIO.ca) for exploration. TMB and 21/37 gene signatures were predictive of IRs across tumor types. We next developed a *de novo* GE signature (PredictIO) from our pan-cancer analysis and demonstrated its superior predictive value over other biomarkers. To identify novel targets, we computed the T-cell dysfunction score for each gene within PredictIO and their ability to predict dual PD-1/CTLA-4 blockade in mice. Two genes, *F2RL1* (encoding protease-activated receptor-2) and *RBFOX2* (encoding RNA-binding motif protein 9), were concurrently associated with worse ICB clinical outcomes, T-cell dysfunction in ICB-naive patients and resistance to dual PD-1/CTLA-4 blockade in preclinical models.

Conclusion: Our study highlights the potential of large-scale meta-analyses in identifying novel biomarkers and potential therapeutic targets for cancer immunotherapy.

Key words: immunotherapy, transcriptomic, biomarker, meta-analysis, machine learning, scientific software

INTRODUCTION

Immune checkpoint blockade (ICB) with monoclonal antibodies (mAbs) targeting programmed cell death protein 1 (PD-1), programmed death-ligand 1 (PD-L1) or cytotoxic T-lymphocyte antigen 4 (CTLA-4) produces durable survival benefits in a variety of cancers. However, 60%-80% of treated patients do not derive any clinical benefit.[1] Resistance to ICB may be attributed to tumor-intrinsic factors such as genomic alterations, oncogenic signaling and suppressed antigen presentation,[2-4] or tumor-extrinsic factors

*Correspondence to: Prof. John Stagg, CRCHUM, 900 St-Denis Street, Montreal, H2X 0A2, Canada. Tel: +1-1514-890-8000; ext 25170
E-mail: john.stagg@umontreal.ca (J. Stagg).
*Prof. Benjamin Haibe-Kains, The MARS Center, TMDT room 11-310, 101 College Street, Toronto, ON, M5G 1L7, Canada. Tel: +1-416-581-7628
E-mail: bhaibeka@uhnresearch.ca (B. Haibe-Kains).

†These authors contributed equally.
‡These authors are equal senior authors.

such as host immunity and the tumor microenvironment (TME).[5] Identifying patients unlikely to benefit from ICB therapy may help identify critical mechanisms of treatment resistance and support better clinical decision making.

Although PD-L1 expression determined by immunohistochemistry is currently used as a companion diagnostic biomarker,[6,7] it has several limitations, including arbitrary cut-point and discordance between assays in the context of combination treatments.[8] Another recently approved biomarker of ICB therapy consists of tumor mutational burden (TMB) [$\geq$10 mutations per megabase (Mb)] for treatment of unresectable or metastatic solid tumors with pembrolizumab.[9,10] However, TMB assessment requires standardization to overcome the variability associated with various sequencing technologies and bioinformatics interpretation.[11,12] Furthermore, since there is substantial heterogeneity of TMB across cancer types, the development of cancer-specific TMB cut-offs may be more appropriate.[12]

Several gene expression (GE) signatures have also been proposed as candidate biomarkers of ICB response (IR).[13-15] Most transcriptomic signatures, however, have been derived from a single tumor type and a limited number of patients. Power calculations have indeed shown that most molecular studies are generally underpowered and that cohorts of thousands of patients are needed to accurately uncover specific biomarkers of IR.[14,16,17] Accordingly, few potential ICB resistance mechanisms and new therapeutic targets have been discovered.[18-22] Recently, Litchfield et al.[23] reported a large meta-analysis of previously described biomarkers of IR in 1000 patients. This important study, however, strongly focused on the genomic mechanisms of response to ICB and only had GE data for a subset of patients.

Here, we addressed these issues by analyzing the genomic and transcriptomic profiles driving resistance to ICB therapy through a pan-cancer meta-analysis of $>$3600 ICB-treated patients from 26 studies. We evaluated the reproducibility of a compendium of 40 signatures previously described as ICB biomarkers and developed a robust predictive ICB score (PredictIO). Finally, inspired from actors such as the Tumor Immune Dysfunction and Exclusion (TIDE) team (tide.dfci.harvard.edu)[18,24] or the Cancer Research Institute (CRI) iAtlas (www.cri-iatlas.org)[25] that have paved the way for data exploration of ICB-treated patients, we implemented predictIO.ca, an open-source cloud-based web application allowing the scientific community to interactively explore the predictive value of individual genes and gene sets of interest (expression, mutation or copy number) for immuno-oncology therapies.

MATERIALS AND METHODS

Systematic review of published clinical studies

Studies were identified by an electronic search on PubMed using the MeSH terms: (cancer OR tumor) AND ((PD-1 OR PD1) OR (PD-L1 OR PDL1) OR (CTLA-4 OR CTLA4)) AND ((anti-PD-1 OR anti-PD1) OR (anti-PD-L1 OR anti-PDL1) OR (anti-CTLA-4 OR anti-CTLA4)) AND (genomic OR transcriptomic OR mutation) (Figure S1). Eligible studies were in the English language, including at least one type of genomic data and clinical outcome information from advanced solid tumor patients treated with anti-PD-1/L1 and/or anti-CTLA-4 ICB therapy, and were published between January 2015 and September 2020. Genomic data were defined as RNA-sequencing and/or tumor exome or targeted-DNA sequencing. Clinical outcome data included response (according to RECIST or other response criteria), progression-free survival (PFS) or overall survival (OS). Studies with fewer than 20 patients were excluded. Trials of ICB in combination with other treatment modalities such as chemotherapy, targeted therapies and radiation were excluded. There was no minimum duration of follow-up for inclusion. Trials in the neoadjuvant setting were excluded (Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2022.08.084). Genomic signatures selected for evaluation were taken from papers showing association with immunotherapy response or known immune escape mechanism (Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2022.08.084).

Response to ICB treatment

Response (R) was defined as RECIST (v1.1) complete or partial response or stable disease (SD) without PFS event at 6 months. Non-response (NR) was defined as RECIST progressive disease or RECIST SD with a PFS event occurring within 6 months. If RECIST information was not available, patients without any PFS event at 6 months and patients with a PFS event occurring within 6 months were classified as R and NR, respectively. If PFS information was unavailable, patients with RECIST SD as best response were classified as not assessable.

Data processing

Outcome data and mutational and transcriptomic profiles were collected for each study. If not publicly available, individual patient clinical and molecular data information was requested from the corresponding authors. Somatic mutation data were composed of gene-level alteration with mutation context information (i.e. synonymous or non-synonymous).

The processed transcriptomic profile was composed of log2-transformed transcripts per million (TPM) data. When raw sequencing data were available (Fumet.1,[26] Fumet.2,[26] Hugo,[27] Hwang,[28] Jung[29] and Riaz[30]), transcript abundance was quantified from the fastq files through Salmon (v1.4.0),[31] using the grch38 GENCODE transcriptome of reference. Gene-level TPM data were then obtained from the estimated transcript abundance level with tximport (v.1.20.0) R package. Otherwise, raw count/TPM data were downloaded directly from the respective publications (Braun,[32] Jerby_Arnon,[20] Liu,[33] Mariathasan,[19] Miao.1,[14] Snyder,[34] Van_Allen[13]). Fragments per kilobase of transcript per million mapped reads (FPKM) from Nathanson et al.[35] were converted to TPM data using the following

formula: $TPM_i = \left(\frac{FPKM_i}{\sum_j FPKM_j}\right) \times 10^6$.

For each dataset, genes with zero expression value in at least 50% of the samples were filtered out. Because of the presence of variability in clinical follow-up between studies, OS and PFS were censored, respectively, at 36 and 24 months. The data composition of the overall cohort is presented in Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2022.08.084.

After processing each study independently, all the datasets were gathered together into an 'ExpressionSet' R object. For each molecular data type, patients were then compared to each other in order to remove any potential sample coming from duplicated patients within each study and between studies. For the mutational profiles, Cohen's κ score $\geq$0.9 was used to define samples coming from the same patient. For the transcriptomic profile, Spearman correlation $\geq$0.98 was used to define samples coming from the same patient. Samples were removed from the study if they were defined as duplicated for at least one of the two molecular data types. As a result, 90 patients were removed for Miao.2,[36] 12 patients from Braun,[32] 1 from Jung[29] and 1 from Samstein.[9]

Tumor mutational burden computation

As mutation profiles from the curated studies were heterogeneous, with some studies reporting all the mutations found within each sample and others only reporting only non-synonymous mutation, we evaluated non-synonymous TMB across studies. The computation of the TMB per Mb was carried out as defined: $TMB = mut_{ns}/target$, with mut_{ns} and *target* defined, respectively, as the number of non-synonymous mutations and the target size of the sequencing. For each tumor type within the meta-dataset, TMB per Mb was calculated, and the median was used as a custom cut-point to assess low and high TMB (Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2022.08.084).

Gene signature computation

Gene signatures (Supplementary Tables S3-S7, available at https://doi.org/10.1016/j.annonc.2022.08.084) composed of genes without any weight were computed using gene set variation analysis (GSVA)[37] enrichment score. Gene signatures composed of genes associated with specific weights (+1 for increased and −1 for decreased expression) were computed using the weighted mean expression as defined: $Gene\_sig = (\sum Gene_x \times weight_x)/\#_{Gene}$, with $Gene_x$, $weight_x$ and $\#_{Gene}$ corresponding, respectively, to the expression and the weight of the gene X and the total number of genes in the signature. For each investigated study, signatures were only computed if at least 80% of their genes were present in the data. We applied z-score transformation to the genes of each signature before the computation. The COX_IS, ImmunoPhenScore, Innate anti-PD-1 Resistance (IPRES) and TIDE signature scores were computed as described in their original publication.[18,22,27,38] Z-score transformation was applied before the GSVA or the weighted mean computation and after the computation for the four specific signatures.

Gene signature network analysis

In order to assess the level of similarity between the curated signatures, we computed the matrix of overlapping genes defining each signature. Distance between signatures was then assessed using principal component analysis (PCA). Clusters of signatures were then derived by applying the PC1 and PC2 into Affinity Propagation Clustering (APC) using the apcluster (v1.4.8) R package. Within each cluster of at least two signatures, we combined all the genes and computed the Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis using the enrichR (v.3.0) R package.[39]

Gene signature correlation analysis

Spearman correlation between each curated signature was computed within each ICB study from the discovery cohort. The median of each association was then plotted using the corrplot (v.0.84) R package.

De novo signature characterization

We selected the 12 246 genes available in at least 12 of the 14 studies with transcriptomic data and carried out a meta-analysis of the association of these genes with IR (genes with zero expression value in at least 50% of the samples were filtered out in the data filtering step). After correction for multiple testing using the Benjamin–Hochberg method (false discovery rate, FDR), none of the investigated genes were significant after the correction. Based on the methodology presented by Shi et al.,[40] we ranked the genes by the absolute value of the log odds ratio (logOR) and selected the top 100 genes significantly associated with IR (threshold: $|\text{logOR}| \geq 0.23$). Out of the 100 genes (Supplementary Table S8), 77 were associated with response and 23 were associated with resistance to ICB therapy. We defined the PredictIO signature, as the ratio of the GSVA[37] signatures composed of the response and the resistance genes as described: $PredictIO = GSVA\left(Response_{Genes}\right)/GSVA\left(Resistance_{Genes}\right)$.

Composition of the ICB validation cohort

From October 2020 to June 2022, we collected additional independent datasets that were not included in our first literature scan or were freshly released from the scientific community to validate our findings. During this process, we collected seven cohorts of ICB-treated patients (six adjuvant and one neoadjuvant study) with available transcriptomic data (Supplementary Table S9, available at https://doi.org/10.1016/j.annonc.2022.08.084): two melanoma (Puch[41]; Gide[42]), one gastric (Kim[43]), one esophageal (Van Den Ende[44]), one renal cell carcinoma (RCC) (Shiuan[45]), one pan-cancer (INSPIRE[46]) and one pancreatic cohort (Padron[47]). Process RNA-seq data from the Puch[41], Shiuan[45], Van Den Ende[44] and Padron[47] studies were directly obtained from the

corresponding publication. For the Gide[42], INSPIRE[46] and Kim[43] studies, transcript abundance was quantified from the fastq files through the Salmon (v1.4.0)[31] program. Gene-level TPM data were then obtained from the estimated transcript abundance level with tximport (v.1.20.0) R package.

Whole-exome sequencing (WES) reads from the INSPIRE and Kim studies were aligned to the reference human genome GRCh38 using bwa (v.0.7.17)[48] aligner with default parameter settings. Somatic mutation calls were generated from matched normal and tumor BAM files using Strelka (v.2.9.10)[49] variant caller. Somatic mutations with a variant allele frequency >10% and a coverage >20 reads were kept and annotated using snpEFF (v.5.0e).[50]

T-cell dysfunction score analysis

T-cell dysfunction score was computed as described in Jiang et al.[18] Briefly, we computed the average expression of the cytotoxic T lymphocyte (CTL) signature composed of *CD8A*, *CD8B*, *GZMA*, *GZMB* and *PRF1*. From the 29 curated The Cancer Genome Atlas (TCGA) cancer type with at least 100 patients, we identified a trend or a significant association of CTL levels toward a good OS in 7 TCGA cancer types: breast (BRCA), cervical squamous cell carcinoma (CESC), head—neck squamous cell carcinoma (HNSC), ovarian (OV), sarcoma (SARC), skin cutaneous melanoma (SKCM) and uterine corpus endometrial carcinoma (UCEC). We applied a Cox proportional hazards (Cox-PH) survival regression to assess how a specific gene from the PredictIO interacts with the association between CTL levels and survival outcome. The Cox-PH model was also corrected for patients' age and sex clinical information. We extracted the log hazard ratio (logHR) and the corresponding *P* value of the interaction test and defined it as T-cell dysfunction score. For each gene, we derived a pan-cancer T-cell dysfunction score by computing a meta-analysis of the T-cell dysfunction scores in the seven TCGA studies.

Mouse model RNA-sequencing data

Raw RNA-sequencing data of pretreatment AB1 (mesothelioma) and Renca (kidney cancer) tumors in BALB/c mice both treated with anti-CTLA4 and anti-PD-L1 were downloaded from the National Center for Biotechnology Information (NCBI) repository (BioProject: PRJNA481876).[51] This cohort is composed of 24 samples from AB1 pretreatment tumors (12 responders, 12 non-responders) and 24 samples from Renca pretreatment tumors (12 responders, 12 non-responders). The abundance of transcripts was quantified from the fastq files through the Salmon (v1.4.0)[31] program. Gene-level TPM data were then obtained from the estimated transcript abundance level with tximport (v.1.20.0) R package.

Affymetrix (Mouse Genome 430 2.0 Array) GE for 90 melanoma tumors[52] arising from an Hgftg;Cdk4R24C/R24C genetically engineered mouse model treated with an anti-mouse PD-L1 antibody was retrieved from the Gene Expression Omnibus (GEO) portal (GSE172320). Data were normalized using the Robust Multichip Average method. Linear modeling was carried out using Limma's lmFit function.

Processed TPM expression data of 16 orthotopic E0771 tumors [triple-negative breast cancer (TNBC)] from mice treated with an anti-PD-1 antibody were retrieved from Chen et al.[53] The abundance of transcripts was quantified from the fastq files through the Salmon (v1.4.0)[31] program. Gene-level TPM data were then obtained from the estimated transcript abundance level with tximport (v.1.20.0) R package.

Pan-cancer TCGA analysis

TCGA[54] RNA-seq raw count expression data from 29 cancer types were downloaded from the Xena (https://xena.ucsc.edu/) repository.[55,56] Gene-level TPM data were then obtained from the estimated transcript abundance level with tximport (v.1.20.0) R package. This dataset is composed of 10 543 cancer samples from 29 cancer types (TCGA).

TCGA-GTEX analysis

Uniformly realigned and re-called TPM GE dataset for all TCGA[54] and Genotype—Tissue Expression (GTEx)[57] samples was downloaded from the Xena (https://xena.ucsc.edu/) repository.[55,56] This dataset is composed of 9547 cancer samples from 28 cancer types (TCGA) and 7408 healthy samples from 30 tissue types across nearly 1000 individuals (GTEx). Immune cell population estimated from TCGA samples was obtained from Saltz et al.[58,59]

Protein—protein interaction network construction and gene function and pathway enrichment analysis

We investigated gene interaction using the STRING[60] database to assess how proteins from PredictIO were interacting between each other. Only interactions with confidence scores ≥0.15 were kept. Functional gene ontology enrichment analysis was then carried out on the EnrichR web application (https://maayanlab.cloud/Enrichr/).[39]

Single-cell data analysis

Single-cell expression data from the 73 studies for *F2RL1* and *RBFOX2* were retrieved from the TISCH platform (tisch.comp-genomics.org/).[61]

Statistical analysis

Descriptive statistics are reported as percentages for categorical variables and the median and interquartile range (IQR) for continuous variables. Comparisons between two categorical variables were assessed using Fisher's exact test. Differences between two continuous variables were assessed using Wilcoxon rank sum test (with permutation test if the number of samples was <10). Association of specific biomarkers with IR was assessed using a logistic regression model. Association of specific biomarkers with ICB survival (either PFS or OS) was assessed using Cox proportional hazards model to assess the discrimination or separation of a survival model. To improve reproducibility, the results of individual independent studies were pooled using random-effects meta-analysis with inverse

variance weighting in DerSimonian and Laird random-effects models.[62] Heterogeneity across studies was evaluated by using the Q statistic along with I^2 index, which describes the total variation across studies attributable to heterogeneity rather than sampling error.[63-65] Note that I^2 value of >50% along with Cochran's Q statistic $P < 0.05$ represents moderate-to-high heterogeneity.[63] Subgroup analysis was considered to assess the impact of tumor type, sequencing technology and normalization method on the source of moderate-to-high heterogeneity.[62] Potential publication bias was carried out using the funnel plot and the Egger test. No statistically significant publication bias was observed (data not shown). Kaplan—Meier curves and log-rank (Mantel—Cox) tests were conducted for PredictIO (as categorical variable) survival analysis. All analyses were carried out on the R platform (version 3.6.3). When needed, *P* values were corrected for multiple testing using the Benjamini—Hochberg (FDR) method. Associations were deemed statistically significant for *P* values and/or FDR ≤0.05.

PredictIO web application

PredictIO web application is a web-based tool that enables researchers to explore the gene signatures in the 25 studies (Explore feature) and to submit real-time gene signature analysis (Biomarker Evaluation feature).

The Explore feature allows researchers to explore pre-computed gene signatures by simply entering the model and outcome of the pre-computed data. To use the Biomarker Evaluation feature, a user inputs (i) a list of genes, (ii) data types to be analyzed (single-nucleotide variant, copy number alteration or expressions), (iii) clinical data filters such as sex, tumor types and treatment types, and (iv) an email to receive results. Once the analysis is complete, the user receives an email with a link to the analysis results page.

The back-end of the web application is built with Flask, a python-based web application framework, and the front-end is built with ReactJS. The back-end uses scripts written in R to carry out the real-time gene signature analysis, and SendGrid to send analysis completion emails to users. The interactive data visualization features on the front-end are implemented using D3.js and Plotly.js.

RESULTS

Compendium of ICB genomic and clinical data for meta-analysis

We collected and aggregated data from 3648 patients with advanced solid tumors treated with ICB, divided into the discovery and validation cohorts (Supplementary Figure S1, available at https://doi.org/10.1016/j.annonc.2022.08.084). In the discovery cohort, the main tumor types were melanoma (26%), lung (17.9%), clear-cell RCC (16.4%) and urothelial cancer (10.4%). The majority of patients (79.8%) received an anti-PD-(L)1 mAb, 9.8% received an anti-CTLA-4 mAb therapy and 10.4% received a combination of anti-PD-(L)1 and anti-CTLA-4 mAbs. Within the cohort, 35.9% of patients were identified as responders. The median PFS and OS were 4.9 and 18 months, respectively (Table 1). We defined 'logOR' and 'logHR' as the natural logarithm of, respectively, the OR and the HR. OR refers to the ratio of the odds of NR and HR refers to the ratio of the hazard rates of progression or death. Of note, sex was not significantly associated with outcomes (Supplementary Figure S2A, available at https://doi.org/10.1016/j.annonc.2022.08.084) and a weak but significant association of age (as continuous) with IR was observed (logOR = −0.02; 95% confidence interval (CI) −0.03 to −0.01; $P = 0.001$) (Supplementary Figure S2B, available at https://doi.org/10.1016/j.annonc.2022.08.084).

TMB is a strong predictive pan-cancer biomarker of ICB responses

We investigated the pan-cancer predictive value of TMB in our data compendium. TMB was significantly associated with improved IR (logOR = −1.15; 95% CI −1.59 to −0.70; $P < 0.001$; Figure 1A), irrespective of the type of variable

Table 1. Clinicopathological characteristics of the discovery cohort

		Total	EXP	SNV
	Studies, *n* (%)	26 (100)	21 (80.8)	17 (65.4)
	No. of patients	3648	1393	3066
Age, years	Median age (IQR)	61 (54-70)	62 (55-69)	61 (53-70)
	NA (%)	979 (26.8)	802 (57.6)	647 (21.1)
Sex	Female (%)	1360 (37.3)	448 (32.2)	1165 (38)
	Male (%)	1963 (53.8)	674 (48.4)	1697 (55.3)
	NA (%)	325 (8.9)	271 (19.5)	204 (6.7)
Primary	Melanoma (%)	949 (26)	482 (34.6)	737 (24)
	Lung (%)	652 (17.9)	144 (10.3)	540 (17.6)
	Kidney (%)	597 (16.4)	293 (21)	495 (16.1)
	Bladder (%)	379 (10.4)	194 (13.9)	337 (11)
	HNC (%)	171 (4.7)	14 (1)	171 (5.6)
	Unknown (%)	166 (4.6)	39 (2.8)	160 (5.2)
	Colon (%)	130 (3.6)	0 (0)	130 (4.2)
	Brain (%)	117 (3.2)	0 (0)	117 (3.8)
	Ureteral (%)	102 (2.8)	51 (3.7)	97 (3.2)
	Esophagus (%)	83 (2.3)	0 (0)	83 (2.7)
	Breast (%)	53 (1.5)	7 (0.5)	53 (1.7)
	Stomach (%)	46 (1.3)	0 (0)	46 (1.5)
	Pancreas (%)	45 (1.2)	45 (3.2)	0 (0)
	Gastric (%)	45 (1.2)	45 (3.2)	42 (1.4)
	Esophageal (%)	35 (1)	35 (2.5)	0 (0)
Drug treatment	PD-1/PD-L1 (%)	2912 (79.8)	1255 (90.1)	2333 (76.1)
	CTLA-4 (%)	356 (9.8)	66 (4.7)	356 (11.6)
	Combo (%)	380 (10.4)	72 (5.2)	377 (12.3)
Response	No response (%)	1211 (64.1)	737 (66)	900 (63.2)
	Response (%)	677 (35.9)	379 (34)	523 (36.8)
	Total (%)	1888 (51.8)	1116 (80.1)	1423 (46.4)
PFS (months)	No progression (%)	457 (29.0)	193 (24.5)	361 (30.1)
	Progressed (%)	1120 (71.0)	595 (75.5)	840 (69.9)
	Total (%)	1577 (43.2)	788 (56.6)	1201 (39.2)
	Median PFS	4.85	4.11	4.83
	IQR PFS	4.27-5.49	3.58-5.33	4.18-5.51
OS (months)	Alive (%)	1391 (43.2)	377 (35.5)	1296 (44.7)
	Dead (%)	1829 (56.8)	684 (64.5)	1606 (55.3)
	Total (%)	3220 (88.3)	1061 (76.2)	2902 (94.7)
	Median OS	18	15.34	19
	IQR OS	16.52-19.29	13.38-17.58	17.08-20.7

CTLA-4, cytotoxic T-lymphocyte antigen 4; EXP, expression data; HNC, head and neck cancer; IQR, interquartile range; NA, not available; OS, overall survival; PD-1, programmed cell death protein 1; PD-L1, programmed death-ligand 1; PFS, progression-free survival; SNV, single-nucleotide variant.

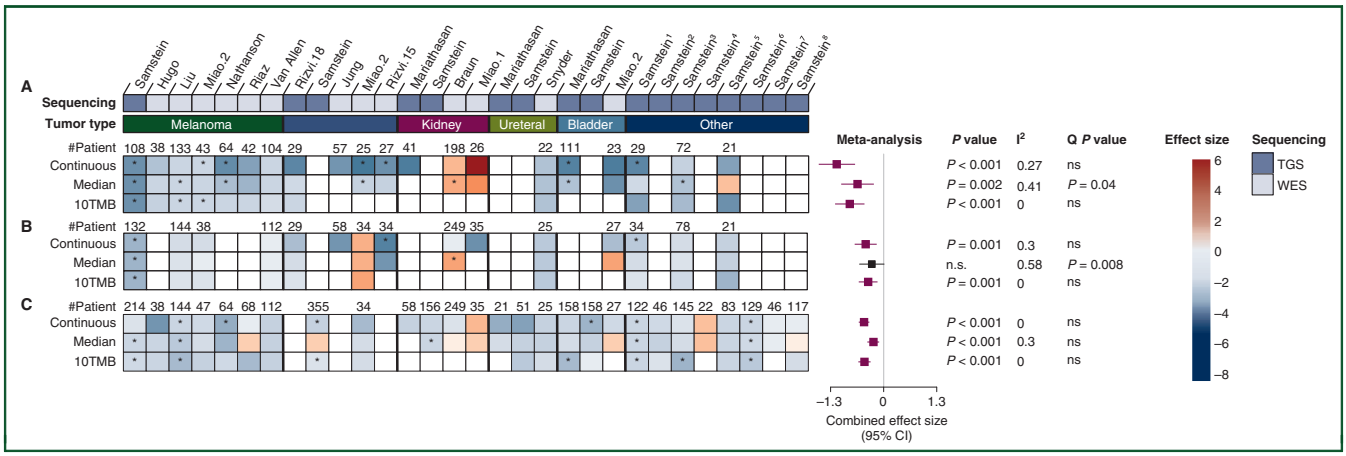

Figure 1. Tumor mutational burden (TMB) association with clinical outcome in immune checkpoint blockade (ICB)-treated patients. Meta-analysis of TMB with ICB response (A), progression-free survival (PFS) (B) and overall survival (OS) (C). The value in the heatmap corresponds to the effect size of each individual study. *$P \leq 0.05$. Forest plot displaying the combined effect size for each clinical outcome (i.e. log odds ratio for the response and log hazard ratio for the PFS/OS) and 95% confidence intervals (CIs) of TMB association with the corresponding clinical outcome. Horizontal bars represent the 95% CIs of effect size. Statistically significant and non-significant associations in the meta-analysis are shown, respectively, in red and black. Missing data are displayed as white boxes. Heterogeneity across studies was evaluated using the I^2 index along with Cochran's Q *P* value. The 'Other' subgroup, from the Samstein[9] study, is composed of #1: unknown; #2: stomach; #3: head and neck; #4: eye; #5: esophagus; #6: colon; #7: breast; #8: brain.
ns, non-significant; TGS, targeted gene sequencing.

used (i.e. continuous, categorical using the median or categorical using Food and Drug Administration-approved 10 TMB/Mb). Higher TMB was also associated with longer PFS and OS (Figure 1B and C). Of note, we observed significant heterogeneity for TMB as a biomarker of PFS (as a categorical variable using the median; $I^2 = 58\%$; Q *P* value = 0.008). Tumor subtype analysis further showed that TMB was not associated with improved response in RCC patients, supporting previous studies.[14,32,66] Together, our results confirmed that TMB is a robust biomarker of IR and survival in different solid malignancies, except for RCC.

Pan-cancer testing of gene signatures associated with ICB responses

We next evaluated the robustness of candidate pretreatment GE signatures previously reported to be associated with increased sensitivity or resistance to ICB therapy (Supplementary Tables S3-S5, available at https://doi.org/10.1016/j.annonc.2022.08.084). We identified five clusters of signatures sharing similar genes (Figure 2A): three clusters related to chemokine, cytokine and T-cell receptor signaling pathways; one related to antigen-presenting machinery; one heterogeneous cluster related to cell cycle, pathways in cancer and extracellular matrix receptor. We then investigated the prognostic value of these signatures in our discovery cohort (Figure 2B and C). All the curated signatures were z-scaled transformed to allow for effect-size comparison. Of note, the IPRES and the Macrophage signatures were not investigated in our meta-analysis due to a low number of genes present in the discovery cohort. Out of 37 remaining curated signatures, 22 (59%) were significantly associated after correction for multiple testing with at least one clinical outcome at pan-cancer level (FDR $<$ 0.05; Figure 2B and Supplementary Tables S6 and S7, available at https://doi.org/10.1016/j.annonc.2022.08.084). Ten ICB-sensitive signatures (i.e. signal transducer and activator of transcription 1 (STAT1), Inflammatory, IRG_Ayers, TIS, TLS, T_cell_inflamed, PDL1, IFNG, CD8_Jiang and CYT) were significantly associated with all three clinical outcomes (i.e. response, PFS and OS). Out of 12 ICB-resistance signatures, 4 (33%) were significantly associated with at least one clinical outcome (PTEN_MITF, T-cell exclusion, C_ECM and COX_IS; Figure 2B). PTEN_MITF was the only signature significantly associated with all three clinical outcomes (Figure 2B). Notably, in contrast to ICB-sensitive signatures, ICB-resistance signatures were not strongly correlated to each other, reflecting more diverse biological pathways (Supplementary Figure S3, available at https://doi.org/10.1016/j.annonc.2022.08.084).

Subgroup analysis showed similar patterns of predictive association of GE signatures in melanoma and lung cancer (Figure 2C). In RCC tumors, however, several GE signatures displayed an opposite predictive association compared to pan-cancer, melanoma or lung cancer (Figure 2C). For instance, antigen-presenting machinery signatures (APM_Thompson and APM_Wang) were significantly associated with worse OS in RCC meta-analysis (Figure 2).

Identification of genes associated with ICB response

To gain more insight into the mechanism of response and resistance to ICB therapy, we carried out a *de novo* meta-analysis of the 12 329 genes present at least in 12 of the 14 transcriptomic curated studies. We ranked the genes based on their logOR and derived signatures from the top 20 up to the top 200 genes to investigate their association with IR (Supplementary Figure S4A, available at https://doi.org/10.1016/j.annonc.2022.08.084). We selected the top 100 genes as the best cut-off (Figure 3A, Supplementary Figure S4A and

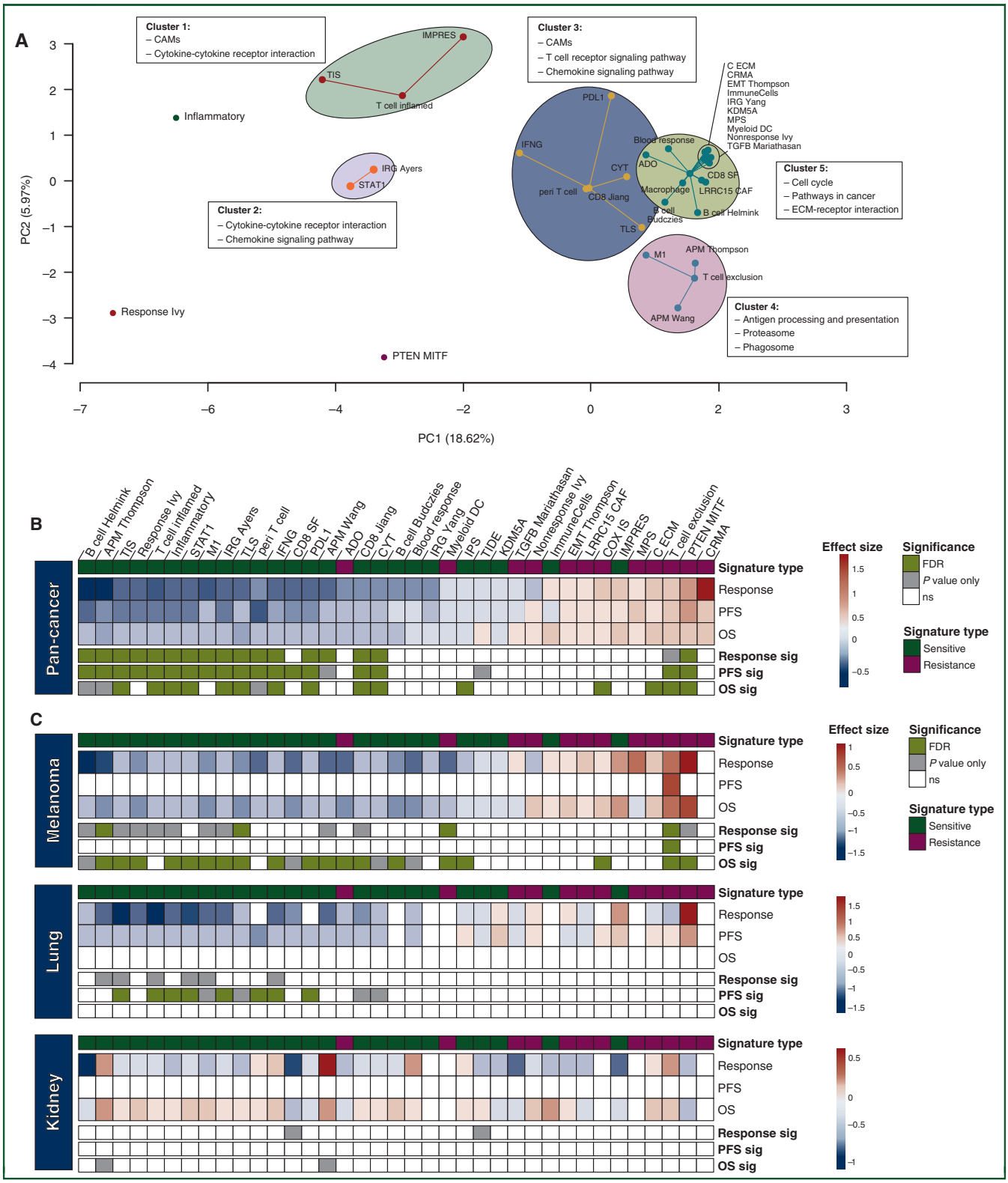

Figure 2. Prognostic landscape of the curated immune checkpoint blockade (ICB) signatures. (A) Principal component analysis (PCA) of the overlapping genes between each curated signature. Clusters were determined using Affinity Propagation Clustering (APC). Kyoto Encyclopedia of Genes and Genomes pathway enriched within each cluster was computed from enrichR. Heatmap summarizing the meta-analysis carried out on the 37 curated signatures associated with clinical outcome in the pan-cancer cohort (B) and the tumor-specific studies (C). Signatures were identified as 'Sensitive' or 'Resistance' based on corresponding publications and displayed, respectively, in green or purple. The value in the heatmap corresponds to the pooled effect size of the meta-analysis of each signature association with response, progression-free survival (PFS) and overall survival (OS). Missing effect-size information is shown as white boxes. *P* values were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) correction. Associations with FDR $\leq$0.05 are shown in black while those with an FDR $>$0.05 and a *P* value $\leq$0.05 are shown in gray. Transcriptomic cohorts used for (A), (B) and (C) analysis: Braun,[32] Fumet.1,[26] Fumet.2,[26] Hugo,[27] Hwang,[28] Jerby_Arnon,[20] Jung,[29] Liu,[33] Mariathasan,[19] Miao.1,[14] Nathanson,[35] Riaz,[30] Snyder,[34] Van_Allen.[13] ECM, extracellular matrix; STAT1, signal transducer and activator of transcription 1.

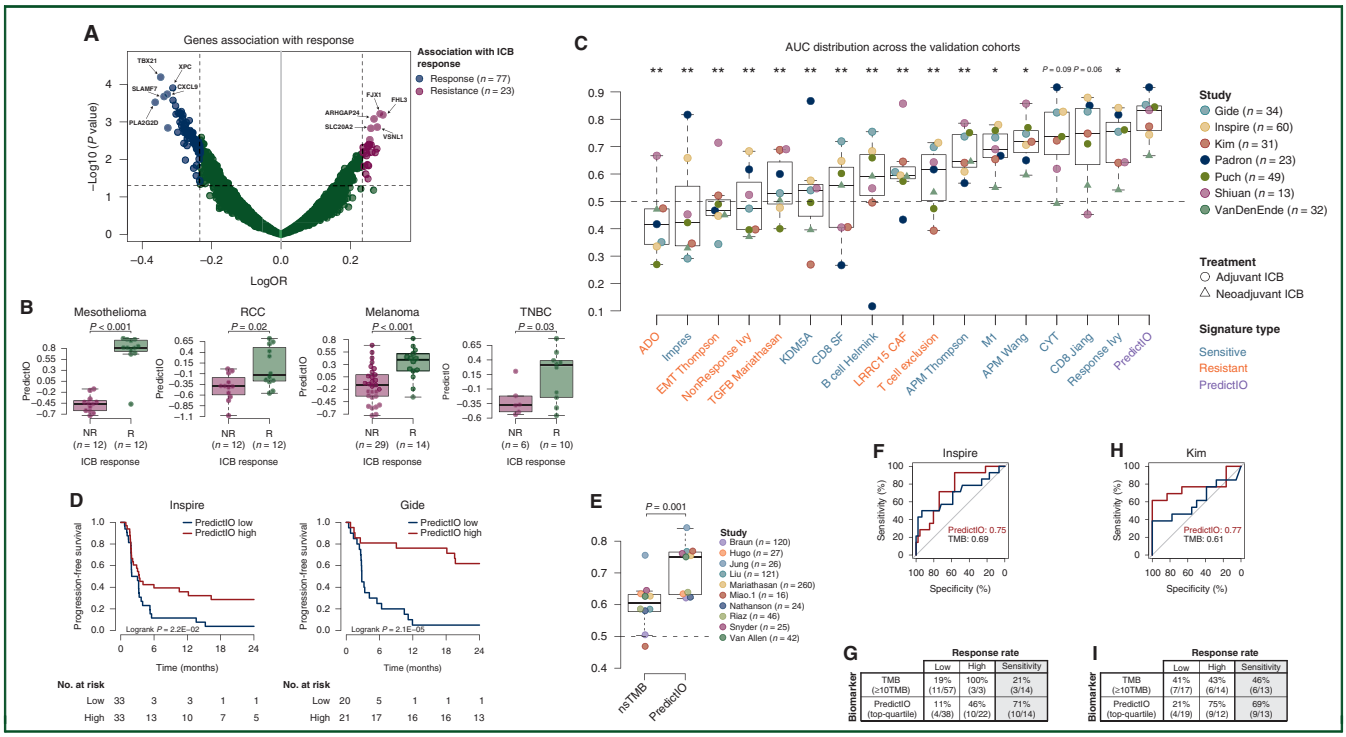

Figure 3. PredictIO: characterization of a *de novo* immune checkpoint blockade (ICB) biomarker. (A) Volcano plot summarizing the meta-analysis of the individual genes associated with response to ICB therapy in the discovery cohort (Braun,[32] Fumet.1,[26] Fumet.2,[26] Hugo,[27] Hwang,[28] Jerby_Arnon,[20] Jung,[29] Liu,[33] Mariathasan,[19] Miao.1,[14] Nathanson,[35] Riaz,[30] Snyder,[34] Van_Allen[13]). The *x*- and y-axis represent, respectively, the log odds ratios (logOR) and the −log10(*P* value). The horizontal bold dashed line represents the *P* value threshold ($P \leq 0.05$) for statistically significant association. The two vertical bold dashed lines represent the logOR threshold ($logOR \geq 0.23$) of the top 100 genes associated with ICB response. Blue and red dots represent, respectively, statistically significant associated genes with better and worse response to ICB therapy. Gray dots represent the non-significant ones. (B) PredictIO difference in expression between non-responder (NR) and responder (R) in AB1 (Mesothelioma) and Renca (RCC) tumors from Zemek et al.,[51] Hgftg;Cdk4R24C/R24C genetically engineered mice (Melanoma) from Meskini et al.[52] and E0771 [triple-negative breast cancer (TNBC)] from Chen et al.[53] Differences were assessed using a two-sided Wilcoxon rank sum test. (C) Area under the curve (AUC) distribution (*y*-axis) of 16 curated signatures and PredictIO available in all of the validation cohorts (Gide,[42] INSPIRE,[46] Kim,[43] Padron,[47] Puch,[41] Shiuan,[45] Van Den Ende[44]). Signatures were ranked based on the median AUC value. Only signatures with available AUC value for all the investigated studies are shown. Boundaries of the box indicate the first and third quartiles of the AUC value. The bold horizontal line indicates the median and the two external horizontal lines show the minimum and maximum values. The horizontal bold dashed line represents the 0.5 AUC value. Signatures were identified as 'Sensitive' or 'Resistant' based on corresponding publications and displayed, respectively, in blue or red. PredictIO is displayed in yellow. Superiority of PredictIO was assessed using a permutation-based one-sided Wilcoxon rank sum test (10 000 permutations). *P* values were corrected for multiple testing using the Benjamini—Hochberg (false discovery rate, FDR) method. *FDR < 0.05; **FDR < 0.01; ns, non-significant. (D) Kaplan—Meier analysis of PredictIO's impact on progression-free survival (PFS) in INSPIRE, Gide and Padron validation cohorts. (E) Comparison of the AUC distribution of the PredictIO and the tumor mutational burden (TMB) (as continuous) in the discovery cohort. Superiority of PredictIO was assessed using a permutation-based one-sided Wilcoxon rank sum test (10 000 permutations). (F) Receiver operating characteristic (ROC) curve of PredictIO and TMB (as continuous) in the INSPIRE validation cohort. (G) Response rate table for PredictIO and TMB (as categorical variable) in the INSPIRE validation cohort. (H) ROC curve of PredictIO and TMB (as continuous) in the Kim validation cohort. (I) Response rate table for PredictIO and TMB (as categorical variable) in the Kim validation cohort.

Table S8, available at https://doi.org/10.1016/j.annonc.2022.08.084); this list is hereafter referred to as the PredictIO signature. Out of the top 100 genes, 77 were associated with ICB sensitivity and 23 were associated with ICB resistance at the pan-cancer level. Enrichment analysis for gene ontology showed that genes associated with ICB sensitivity were associated with immune response regulation (Supplementary Figure S4B and C, available at https://doi.org/10.1016/j.annonc.2022.08.084), whereas ICB-resistance genes were associated with epithelial—mesenchymal transition (EMT) (Supplementary Figure S4B-D, available at https://doi.org/10.1016/j.annonc.2022.08.084).

We then evaluated the predictive value of the PredictIO signature. We defined the PredictIO signature as the ratio of the GSVA[37] signatures composed of the response and the resistance genes as described: $PredictIO = GSVA\ (Response_{Genes})/GSVA(Resistance_{Genes})$.

We firstly investigated whether PredictIO was associated with IR in our discovery cohort and found significant and stronger association with IR than any other investigated signature (Supplementary Figure S4E, available at https://doi.org/10.1016/j.annonc.2022.08.084). We then assessed the prognostic association of our signature in ICB-treated preclinical cancer models. In four distinct mouse models (mesothelioma, RCC, melanoma and TNBC) treated with ICB therapy, PredictIO score obtained from pretreatment samples showed consistent ability to stratify sensitive from resistant mice (Figure 3B), with areas under the curve (AUCs) ranging from 0.78 to 0.95 (Supplementary Figure S4F, available at https://doi.org/10.1016/j.annonc.2022.08.084).

The robustness of the PredictIO gene signature was further validated in seven independent cohorts of ICB-treated cancer patients,[41-46] encompassing 302 patients with melanoma,

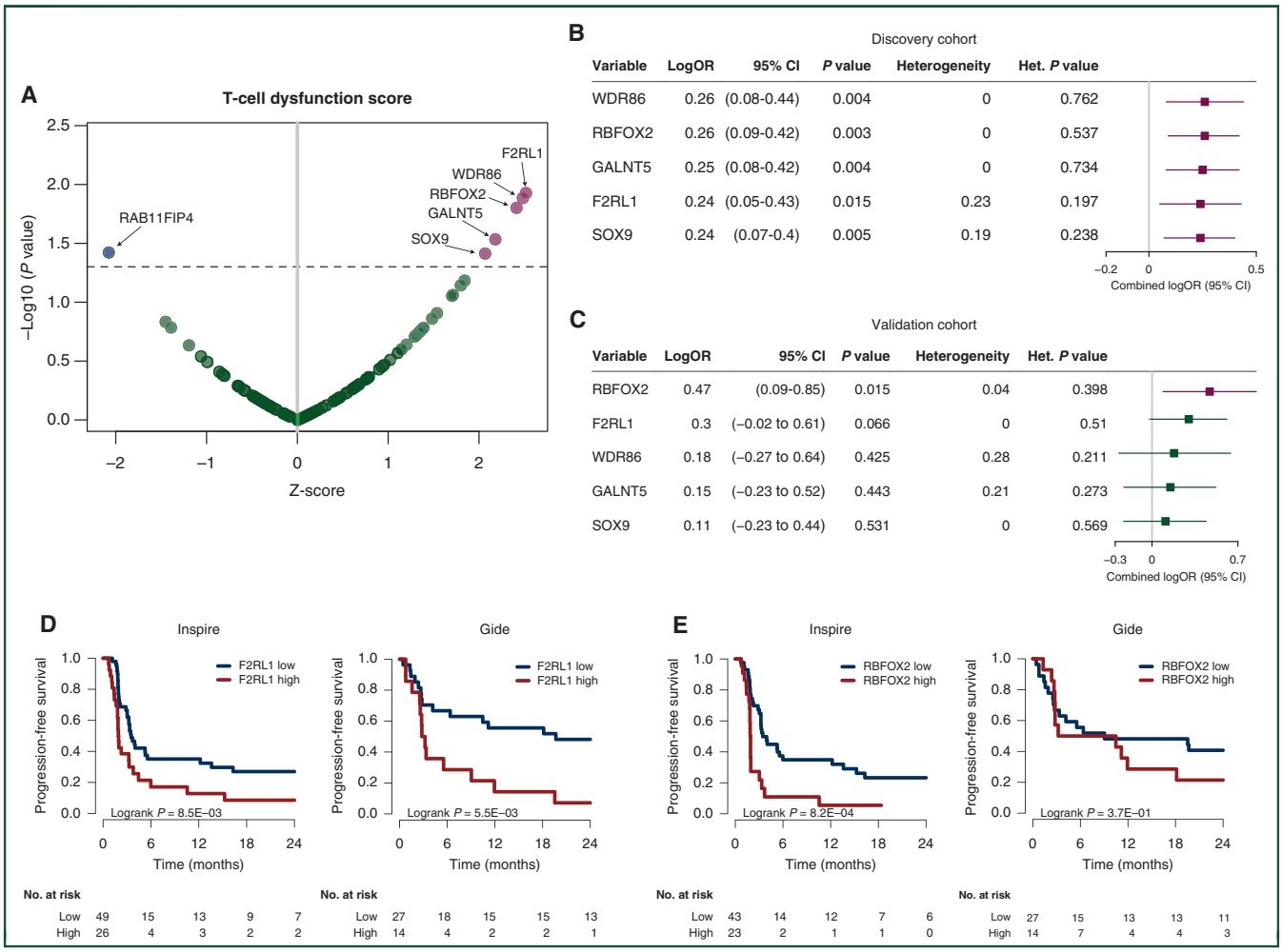

Figure 4. F2RL1 and RBFOX2 as potential therapeutic targets to overcome immune checkpoint blockade (ICB) resistance. (A) Volcano plot summarizing the meta-analysis of the T-cell dysfunction score, from the genes identified in PredictIO, in The Cancer Genome Atlas (TCGA) studies (BRCA—breast, CESC—cervical squamous cell carcinoma, HNSC—head—neck squamous cell carcinoma, OV—ovarian, SARC—sarcoma; SKCM—skin cutaneous melanoma; UCEC—uterine corpus endometrial carcinoma). The *x*- and *y*-axis represent, respectively, the z-score and the −log10(FDR). The horizontal bold dashed line represents the false discovery rate (FDR) threshold (FDR $\leq$ 0.05) for statistically significant association. Blue and red dots represent, respectively, statistically significant associated genes with negative and positive T-cell dysfunction scores. Gray dots represent the non-significant ones. Meta-analysis of the five identified genes with T-cell dysfunction score association with ICB response (B) in the discovery and (C) validation cohorts. Forest plot displaying log odds ratios (logOR) and 95% confidence intervals (CIs). Horizontal bars represent the 95% CIs of effect size. Significant association with survival is shown in red. Kaplan—Meier analysis of F2RL1 (D) and RBFOX2 (E) impact on progression-free survival (PFS) in INSPIRE and Gide validation cohorts.

gastric, esophageal, RCC, pancreatic and other tumor types (Supplementary Table S9, available at https://doi.org/10.1016/j.annonc.2022.08.084). Remarkably, when compared to other candidate GE signatures, PredictIO was the most robust predictor of IR, with significantly higher AUC values (median AUC: 0.83; IQR: 0.76-0.85; Figure 3C), except for CYT and CD8_Jiang signatures, which had a larger variability. In cohorts with available survival data (INSPIRE[46] and Gide[42]), PredictIO further exhibited a strong and significant association with PFS and OS (Figure 3D). Of interest, we investigated the prognostic value of PredictIO through a meta-analysis of >10 000 non-ICB-treated patients from 29 cancer types (Supplementary Figure S5, available at https://doi.org/10.1016/j.annonc.2022.08.084). We found that PredictIO was not significantly associated with prognosis (FDR = 0.097), showing that the observed association with IR is specific to the immunotherapy setting. We also investigated individually the prognostic value of each gene composing PredictIO and found that only 3 out of the 100 genes were significantly associated with prognosis, namely RAB42, FJX1 and SORCS2 (Supplementary Figure S6, available at https://doi.org/10.1016/j.annonc.2022.08.084).

In the discovery cohort, PredictIO also showed a significantly higher AUC (median AUC: 0.73; 95% CI 0.64-0.77) compared to TMB (median AUC: 0.60; 95% CI 0.57-0.64) (P = 0.001; Figure 3E). PredictIO (as continuous) showed higher AUC value than TMB (as continuous) in the pan-cancer INSPIRE trial[46] (Figure 3F) and the gastric Kim trial[43] (Figure 3H). Notably, IR prediction sensitivity was consistently higher for PredictIO (top quartile) versus TMB (10 TMB/Mb) in these two trials (Figure 3G and I) with significant differences in the INSPIRE study (21% versus 71%; two-sided Fisher P = 0.023). Overall, the top-quartile value of PredictIO was able to stratify

patients to identify 71% and 69% of ICB responders, respectively, in the INSPIRE and Kim studies. In contrast, TMB was only able to reach 21% and 46% sensitivity with 10 TMB/Mb as cut-points, respectively, in the INSPIRE and Kim studies.

Taken together, our results highlight the strong and robust prognostic value of the PredictIO gene signature and its potential clinical relevance as compared to the current standard.

Identification of potential targets to overcome ICB resistance

We next sought to identify novel therapeutic targets to overcome ICB resistance. For this, we calculated the T-cell dysfunction score for each gene within the PredictIO signature and compared their individual predictive value in the AB1 and Renca models. Described by Jiang et al.,[18] the T-cell dysfunction score measures the impact of a given gene on the prognostic value of a CTL GE signature (CTL; composed of *CD8A*, *CD8B*, *GZMA*, *GZMB* and *PRF1*) in ICB-naive tumors. T-cell dysfunction scores were computed in TCGA datasets where the CTL signature was significantly associated with better OS (Figure 4). Meta-analysis of these studies identified five PredictIO genes significantly associated with a high T-cell dysfunction score: *F2RL1*, *WDR86*, *RBFOX2*, *GALNT5* and *SOX9* (Figure 4A). Amongst these, *F2RL1, GALNT5, RBFOX2* and *SOX9* were further associated with resistance in at least one of the four ICB-treated preclinical models (Supplementary Figure S7, available at https://doi.org/10.1016/j.annonc.2022.08.084). *RBFOX2* (logOR = 0.47; 95% CI 0.09-0.85; $P = 0.015$) and F2RL1 (logOR = 0.30; 95% CI −0.02 to 0.61; $P = 0.066$) were the only genes associated with worse IR in our discovery and validation cohorts (Figure 4B and C and Supplementary Figure S8A and B, available at https://doi.org/10.1016/j.annonc.2022.08.084).

F2RL1, encoding for the protease-activated receptor-2 (PAR2), was consistently overexpressed in cancer tissues, except for skin (Supplementary Figure S8C, available at https://doi.org/10.1016/j.annonc.2022.08.084). *RBFOX2*, encoding for an RNA-binding protein essential for EMT-driven alternative splicing,[67] was more homogeneously expressed, although significantly increased in some malignancies, such as pancreas, lung, kidney and breast cancer (Supplementary Figure S8D, available at https://doi.org/10.1016/j.annonc.2022.08.084). Further supporting our T-cell dysfunction analysis, *F2RL1* and *RBFOX2* were negatively associated with CD8 T cells and M2 macrophages and positively correlated with M2 macrophages (Supplementary Figure S8E and F, available at https://doi.org/10.1016/j.annonc.2022.08.084). Lastly, we investigated F2RL1 and RBFOX2 expression in 28 tumor types from 72 single-cell studies and identified that these two genes were mainly expressed on malignant and fibroblast cells (Supplementary Figure S8G and H, available at https://doi.org/10.1016/j.annonc.2022.08.084).

Interactive mining of biomarkers for ICB response

To facilitate the investigation of individual genes and gene signatures in ICB-treated patients, we developed PredictIO.ca, an open-source cloud-based web application composed of three interactive modules: Pre-Computed Signatures, Biomarker Evaluation and PredictIO. The first module will allow users to graphically explore the entire signature analysis presented in this manuscript. With the Biomarker Evaluation module, users will investigate the predictive value of their own gene/signatures, computed either using the weighted mean or the GSVA method (depending on if the signature includes coefficient or not), as described in this manuscript. Briefly, users can assess the relationship of their signature with published curated signatures (as shown in Figure 2A) and explore the association of a specific biomarker with IR, PFS and OS in the 26 studies encompassing >3600 patients with RNA-seq and/or targeted gene sequencing/WES available. Amongst these two modules, users can download all the generated figures in high resolution, the tables used to generate them along with the selected sequencing and clinical information to maximize research transparency and reproducibility. Finally, using the last module, clinicians and researchers will be able to upload patients' RNA-seq data to compute the PredictIO score.

DISCUSSION

Assessing the validity and reproducibility of predictive GE signatures of IR remains an open challenge due to limited access to pharmacogenomic data for immunotherapies. In this study, we carried out the largest comparative analysis to date of candidate genomics and transcriptomic biomarkers of IRs in a pan-cancer cohort of over 3600 patients. At the genomic level, we validated that high TMB was significantly associated with better IRs across cancer indications, with the exception of kidney cancer, as previously reported.[14,32,66] At the transcriptomic level, only approximately half of curated GE signatures (22 out of 37; 59%) were significantly associated with pan-cancer IRs. Five ICB-sensitive signatures (APM_Wang, CYT, APM_Thompson, M1 and CD8_Jiang) were significantly associated with pan-cancer response in the discovery cohort and displayed a median AUC >0.70 in validation cohorts.

Yet, our study demonstrated the superior predictive impact of a novel GE signature—PredictIO—derived from our *de novo* pan-cancer analysis. Our premise was that a large pan-cancer analysis would be ideal to uncover critical pathways of ICB resistance, and that small single tumor type cohorts may bias toward the identification of prognostic genes irrespective of their immune impact. PredictIO was more robust and more statistically significant than all the other curated GE signatures in our validation pan-cancer cohort, with the highest median AUC (median AUC: 0.81; 95% CI 0.76-0.86). Notably, PredictIO allowed for better patient stratification of response than TMB in the two phase II studies of pembrolizumab in advanced solid tumors (INSPIRE)[46] and advanced gastric tumors (Kim),[43] with a

sensitivity of ∼70%, as compared with TMB (10 TMB/Mb) that had at best a sensitivity of 46%.

A unique feature of PredictIO, compared to other signatures, is that it is composed of a mix of genes associated with either ICB sensitivity (77%) or resistance (23%). To uncover novel therapeutic targets of cancer immunotherapy, we cross-analyzed the PredictIO signature with two additional datasets: a compendium of TCGA studies to compute T-cell dysfunction scores[18] and transcriptomic profiles of mouse tumors associated with response or resistance to dual PD-1/CTLA-4 blockade.[51-53] Five PredictIO genes were concurrently associated with a high T-cell dysfunction score (*F2RL1*, *WDR86*, *RBFOX2*, *GALNT5 and SOX9*), amongst which only *F2RL1* and *RBFOX2* were associated with worse IR in our discovery and validation cohorts.

RBFOX2 is an RNA-binding protein that promotes alternative splicing. In response to transforming growth factor (TGF)-β signaling, RBFOX2 together with another RNA-binding protein, PCBP1, promote alternate splicing of TAK1 pre-messenger RNA, a required step for TGF-β-induced EMT.[68] Induction of EMT further up-regulates RBFOX2 expression to drive mesenchymal tissue-specific splicing.[67,69] Depletion of RBFOX2 in mesenchymal tumor cells reduces cell invasion and promotes epithelial phenotype.[67,70] Since RBFOX proteins regulate a wide range of neurodevelopmental functions,[71] targeting RBFOX2 may be ill-advised. One could envision, however, targeting the splicing of TAK1 by RBFOX2 with splice-switching oligonucleotides, as proposed by Tripathi et al.[68] This approach has been successfully used to induce TAK1 isoform switching in mouse livers.[72] Furthermore, recent studies demonstrated in preclinical models the benefit of ICB therapy in combination with pharmacological perturbation of splicing modulators, which induce anti-viral immune response through immunogenic splicing-derived neoepitopes.[73,74]

F2RL1 encodes for PAR2, a G-protein-coupled receptor that responds to activated proteases including trypsin, tissue kallikreins, mast cell tryptase, cathepsins, neutrophil elastase and tissue factor (TF)/factor (F) VIIa and Xa complexes involved in coagulation. PAR2 is increased by inflammatory stimuli and the observation that mice deficient in PAR2 are protected from inflammatory arthritis and pronociceptive actions drove the development of selective antagonists for the treatment of rheumatic diseases and chronic pain (NCT04198558).[75-77] In response to inflammatory stimuli, PAR2 has been shown to suppress type I interferon (IFN) responses and STAT1 signaling,[78] and polarizes human macrophages toward an M2-like phenotype.[79] Interestingly, a recent study demonstrated that inhibiting the protease FXa, which activates PAR2, significantly improved ICB therapy in a mouse model of cancer through macrophage reprogramming.[80] PAR2 signaling thus appears as a critical pathway of tumor immune evasion and may represent a novel therapeutic target to overcome ICB resistance. Importantly, PAR2 antagonists are currently in clinical development and could be rapidly tested in combination with ICB in cancer patients.

The other PredictIO genes concurrently associated with a high T-cell dysfunction score were *WDR86*, *GALNT5 and SOX9*. *WDR86* is a poorly characterized gene. *GALNT5* encodes an N-acetylgalactosaminyltransferase that catalyzes glycosylation of Golgi proteins and may be involved in promoting EMT.[81] The high mobility group transcription factor *SOX9* has also been associated with EMT and oncogenesis,[82,83] and may decreased type I IFN responses.[84]

Recognizing the potential clinical value of the generated data for the scientific community, we developed PredictIO (PredictIO.ca), an open-source cloud-based web application allowing users to investigate the association of individual genes and signatures with IR and ICB resistance. Other platforms, such as the one developed by the TIDE team (tide.dfci.harvard.edu)[18,24] or the Cancer Research Institute (CRI) iAtlas (www.cri-iatlas.org),[25] already allow users to investigate the predictive value of their biomarker of interest separately within each study. Here, we proposed a more clinically impactful approach granting users the ability to carry out comprehensive meta-analyses of response, PFS and OS of their biomarker of interest (expression, mutation or copy number). Altogether this web application will bring high-impact knowledge to the community which could lead to the development of new clinical targets aiming to improve response in ICB therapy.

Our study has several limitations. Our pan-cancer cohort was composed of heterogeneous studies with different ICB drug therapies and variability in patient population including in terms of lines of prior therapy. Moreover, these studies were composed of distinct sequencing technology and bioinformatics processing. In our meta-analysis, we did not observe any significant heterogeneity from these last two factors. The main source of heterogeneity was originating from the various cancer types. Our curated dataset was mainly composed of melanoma, lung and kidney cancer and there is a need to pursue curating new studies to increase data for other tumor types. Another limitation of this study is the fact that we mainly focused on the transcriptomic profiles and left out some putative interesting genomic alterations. Several previously published studies[13,14,32,85,86] had already investigated the genomic alterations associated with response or resistance to ICB with only limited success. Finally, association between gene(s) and IR does not prove a causative role of these genes.

In conclusion, our study highlights the relative value and limitations of previously reported GE signatures in their ability to predict IR, and demonstrated the power of *de novo* pan-cancer meta-analysis of GE profiles in order to bring forth novel predictive biomarkers and uncover potential novel therapeutic targets in immuno-oncology.

FUNDING

This work was supported by a research grant (Immunotherapy network; iTNT) from the Terry Fox Research Institute (JS and BHK) (no grant number); by a research grant (ODAC competition) funded by Génome Québec, the

Ministere de l'Economie et de l'Innovation du Quebec, IVADO, the Canada First Research Excellence Fund and Oncopole (JS) (no grant number); the Princess Margaret Cancer Foundation (BHK); the Genome Canada Bioinformatics and Computational Biology (BHK); and the Canadian Institutes for Health Research (BHK). JS is a researcher of CRCHUM, which receives support from the Fonds de recherche Québec—Santé (FRQS). YB was supported by the Jean-Guy Sabourin Research Chair in Pharmacology of Université de Montréal, and by a postdoctoral fellowship from FRQS. JS is supported by a new investigator award from the Canadian Institutes of Health Research (CIHR). PLB is a co-contact principal investigator for NCI UM1 grant [grant number CA186644]. The funding sources had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

DISCLOSURE

JS is a permanent SAB member and owns stocks of Surface Oncology. BHK is a consultant for Code Ocean. PLB reports research grant (to institution) for AstraZeneca, Bristol MyersSquibb, GlaxoSmithKline, Roche/Genentech, Novartis, Merck, Zymeworks, SeaGen, Takeda, Lilly, Pfizer, Nektar Therapeutics, Mersana, Bicara Therapeutics, Amgen; (uncompensated) advisory for Lilly, Amgen, Merck, SeaGen, Gilead. All other authors have declared no conflicts of interest.

DATA SHARING

Resource availability

Lead contact. Further information and requests for resources and reagents should be addressed to: Benjamin Haibe-Kains (Benjamin.Haibe-Kains@uhnresearch.ca) and John Stagg (john.stagg@umontreal.ca).

Data availability

https://doi.org/10.5281/zenodo.6142357

Code availability

Code for analysis scripts → https://github.com/bhklab/PredictIO
Code for web-app → https://github.com/bhklab/PredictIO-webapp

Research reproducibility

A complete software environment through CodeOcean containing all necessary data and code to reproduce the analysis and figures described in this manuscript is available at https://codeocean.com/capsule/3860745/tree.

REFERENCES

1. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19(3):133-150.
2. Pitt JM, Vétizou M, Daillère R, et al. Resistance mechanisms to immune-checkpoint blockade in cancer: tumor-intrinsic and -extrinsic factors. *Immunity*. 2016;44(6):1255-1269.
3. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell*. 2017;168(4):707-723.
4. Vitale I, Shema E, Loi S, Galluzzi L. Intratumoral heterogeneity in cancer progression and response to immunotherapy. *Nat Med*. 2021;27(2):212-224.
5. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013;19(11):1423-1437.
6. Schmid P, Adams S, Rugo HS, et al. Atezolizumab and nab-paclitaxel in advanced triple-negative breast cancer. *N Engl J Med*. 2018;379(22):2108-2121.
7. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med*. 2012;366(26):2443-2454.
8. Doroshow DB, Bhalla S, Beasley MB, et al. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol*. 2021;18(6):345-362.
9. Samstein RM, Lee C-H, Shoushtari AN, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet*. 2019;51(2):202-206.
10. Subbiah V, Solit DB, Chan TA, Kurzrock R. The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) $\geq$10: a decision centered on empowering patients and their physicians. *Ann Oncol*. 2020;31(9):1115-1118.
11. Chan TA, Yarchoan M, Jaffee E, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol*. 2019;30(1):44-56.
12. McGrail DJ, Pilié PG, Rashid NU, et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann Oncol*. 2021;32(5):661-672.
13. Van Allen EM, Miao D, Schilling B, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015;350(6257):207-211.
14. Miao D, Margolis CA, Gao W, et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science*. 2018;359(6377):801-806.
15. Cristescu R, Mogg R, Ayers M, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science*. 2018;362(6411):eaar3593.
16. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505(7484):495-501.
17. Hegde PS, Chen DS. Top 10 challenges in cancer immunotherapy. *Immunity*. 2020;52(1):17-35.
18. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. 2019;24(10):1550-1558.
19. Mariathasan S, Turley SJ, Nickles D, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018;554(7693):544-548.
20. Jerby-Arnon L, Shah P, Cuoco MS, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell*. 2018;175(4):984-997.e24.
21. Cabrita R, Mitra S, Sanna A, et al. The role of PTEN loss in immune escape, melanoma prognosis and therapy response. *Cancers (Basel)*. 2020;12(3):742.
22. Bonavita E, Bromley CP, Jonsson G, et al. Antagonistic inflammatory phenotypes dictate tumor fate and response to immune checkpoint blockade. *Immunity*. 2020;53(6):1215-1229.e8.
23. Litchfield K, Reading JL, Puttick C, et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell*. 2021;184(3):596-614.e14.
24. Fu J, Li K, Zhang W, et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med*. 2020;12(1):1-8.
25. Eddy JA, Thorsson V, Lamb AE, et al. CRI iAtlas: an interactive portal for immuno-oncology research. F1000Res. 2020;9:1028.

26. Fumet JD, Richard C, Ledys F, et al. Prognostic and predictive role of CD8 and PD-L1 determination in lung tumor tissue of patients under anti-PD-1 therapy. *Br J Cancer*. 2018;119(8):950-960.
27. Hugo W, Zaretsky JM, Sun L, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell*. 2016;165(1):35-44.
28. Hwang S, Kwon AY, Jeong JY, et al. Immune gene signatures for predicting durable clinical benefit of anti-PD-1 immunotherapy in patients with non-small cell lung cancer. *Sci Rep*. 2020;10(1):643.
29. Jung H, Kim HS, Kim JY, et al. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat Commun*. 2019;10(1):4278.
30. Riaz N, Havel JJ, Makarov V, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell*. 2017;171(4):934-949.e15.
31. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419.
32. Braun DA, Hou Y, Bakouny Z, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. *Nat Med*. 2020;26(6):909-918.
33. Liu D, Schilling B, Liu D, et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat Med*. 2019;25(12):1916-1927.
34. Snyder A, Nathanson T, Funt SA, et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: an exploratory multi-omic analysis. *PLoS Med*. 2017;14(5):e1002309.
35. Nathanson T, Ahuja A, Rubinsteyn A, et al. Somatic mutations and neoepitope homology in melanomas treated with CTLA-4 blockade. *Cancer Immunol Res*. 2017;5(1):84-91.
36. Miao D, Margolis CA, Vokes NI, et al. Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet*. 2018;50(9):1271-1281.
37. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
38. Charoentong P, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248-262.
39. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14(1):128.
40. Shi L, Jones WD, Jensen RV, et al. The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics*. 2008;9(suppl 9):S10.
41. Cui C, Xu C, Yang W, et al. Ratio of the interferon-γ signature to the immunosuppression signature predicts anti-PD-1 therapy response in melanoma. *NPJ Genom Med*. 2021;6(1):7.
42. Gide TN, Quek C, Menzies AM, et al. Distinct immune cell populations define response to anti-PD-1 monotherapy and anti-PD-1/anti-CTLA-4 combined therapy. *Cancer Cell*. 2019;35(2):238-255.e6.
43. Kim ST, Cristescu R, Bass AJ, et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med*. 2018;24(9):1449-1458.
44. van den Ende T, de Clercq NC, van Berge Henegouwen MI, et al. Neoadjuvant chemoradiotherapy combined with atezolizumab for resectable esophageal adenocarcinoma: a single-arm phase II feasibility trial (PERFECT). *Clin Cancer Res*. 2021;27(12):3351-3359.
45. Shiuan E, Reddy A, Dudzinski SO, et al. Clinical features and multiplatform molecular analysis assist in understanding patient response to anti-PD-1/PD-L1 in renal cell carcinoma. *Cancers (Basel)*. 2021;13(6):1475.
46. Cindy Yang SY, Lien SC, Wang BX, et al. Pan-cancer analysis of longitudinal metastatic tumors reveals genomic alterations and immune landscape dynamics associated with pembrolizumab sensitivity. *Nat Commun*. 2021;12(1):5137.
47. Padrón LJ, Maurer DM, O'Hara MH, et al. Sotigalimab and/or nivolumab with chemotherapy in first-line metastatic pancreatic cancer: clinical and immunologic analyses from the randomized phase 2 PRINCE trial. *Nat Med*. 2022;28(6):1167-1177.
48. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
49. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-1817.
50. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. 2012;6(2):80-92.
51. Zemek RM, De Jong E, Chin WL, et al. Sensitization to immune checkpoint blockade through activation of a STAT1/NK axis in the tumor microenvironment. *Sci Transl Med*. 2019;11(501):eaav7816.
52. El Meskini R, Atkinson D, Kulaga A, et al. Distinct biomarker profiles and TCR sequence diversity characterize the response to PD-L1 blockade in a mouse melanoma model. *Mol Cancer Res*. 2021;19(8):1422-1436.
53. Chen IX, Newcomer K, Pauken KE, et al. A bilateral tumor model identifies transcriptional programs associated with patient response to immune checkpoint blockade. *Proc Natl Acad Sci U S A*. 2020;117(38):23684-23694.
54. Hoadley KA, Yau C, Hinoue T, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*. 2018;173(2):291-304.e6.
55. Vivian J, Rao AA, Nothaft FA, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol*. 2017;35(4):314-316.
56. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675-678.
57. Ardlie KG, DeLuca DS, Segrè AV, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-660.
58. Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity*. 2018;48(4):812-830.e14.
59. Saltz J, Gupta R, Hou L, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep*. 2018;23(1):181-193.e7.
60. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362-D368.
61. Sun D, Wang J, Han Y, et al. TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res*. 2021;49(D1):D1420-D1430.
62. Borenstein M, Higgins JPT. Meta-analysis and subgroups. *Prev Sci*. 2013;14(2):134-143.
63. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-560.
64. Lin L, Chu H, Hodges JS. Analysis : reducing the impact of outlying studies. *Biometrics*. 2017;73(1):156-166.
65. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Stat Med*. 1991;10(11):1665-1677.
66. McDermott DF, Huseni MA, Atkins MB, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med*. 2018;24(6):749-757.
67. Braeutigam C, Rago L, Rolke A, Waldmeier L, Christofori G, Winter J. The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. *Oncogene*. 2014;33(9):1082-1092.
68. Tripathi V, Shin JH, Stuelten CH, Zhang YE. TGF-β-induced alternative splicing of TAK1 promotes EMT and drug resistance. *Oncogene*. 2019;38(17):3185-3200.
69. Venables JP, Brosseau JP, Gadea G, et al. RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol Cell Biol*. 2013;33(2):396-405.

70. Shapiro IM, Cheng AW, Flytzanis NC, et al. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet*. 2011;7(8):e1002218.
71. Gehman LT, Meera P, Stoilov P, et al. The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes Dev*. 2012;26(5):445-460.
72. Zhou D, Shao Q, Fan X, et al. Regulation of Tak1 alternative splicing by splice-switching oligonucleotides. *Biochem Biophys Res Commun*. 2018;497(4):1018-1024.
73. Bowling EA, Wang JH, Gong F, et al. Spliceosome-targeted therapies trigger an antiviral immune response in triple-negative breast cancer. *Cell*. 2021;184(2):384-403.e21.
74. Lu SX, De Neef E, Thomas JD, et al. Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell*. 2021;184(15):4032-4047.e31.
75. Tsai CC, Chou YT, Fu HW. Protease-activated receptor 2 induces migration and promotes Slug-mediated epithelial-mesenchymal transition in lung adenocarcinoma cells. *Biochim Biophys Acta Mol Cell Res*. 2019;1866(3):486-503.
76. Jiang Y, Zhuo X, Fu X, Wu Y, Mao C. Targeting PAR2 overcomes gefitinib resistance in non-small-cell lung cancer cells through inhibition of EGFR transactivation. *Front Pharmacol*. 2021;12:625289.
77. Sun L, Li PB, Yao YF, et al. Proteinase-activated receptor 2 promotes tumor cell proliferation and metastasis by inducing epithelial-mesenchymal transition and predicts poor prognosis in hepatocellular carcinoma. *World J Gastroenterol*. 2018;24(10):1120-1133.
78. Weithauser A, Bobbert P, Antoniak S, et al. Protease-activated receptor-2 regulates the innate immune response to viral infection in a coxsackievirus B3-induced myocarditis. *J Am Coll Cardiol*. 2013;62(19):1737-1745.
79. García-González G, Sánchez-González A, Hernández-Bello R, et al. Triggering of protease-activated receptors (PARs) induces alternative M2 macrophage polarization with impaired plasticity. *Mol Immunol*. 2019;114:278-288.
80. Graf C, Wilgenbus P, Pagel S, et al. Myeloid cell-synthesized coagulation factor X dampens antitumor immunity. *Sci Immunol*. 2019;4(39):eaaw8405.
81. Detarya M, Sawanyawisuth K, Aphivatanasiri C. The O-GalNAcylating enzyme GALNT5 mediates carcinogenesis and progression of cholangiocarcinoma via activation of AKT/ERK signaling. *Glycobiology*. 2020;30(5):312-324.
82. Vasaikar S, Huang C, Wang X, et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell*. 2019;177(4):1035-1049.e19.
83. Li T, Huang H, Shi G, et al. TGF-β1-SOX9 axis-inducible COL10A1 promotes invasion and metastasis in gastric cancer via epithelial-to-mesenchymal transition. *Cell Death Dis*. 2018;9(9):849.
84. Jiang X, Xu LG. SOX9 negatively regulates the RLR antiviral signaling by targeting MAVS. *Virus Genes*. 2022;58(2):122-132.
85. Zaretsky JM, Garcia-Diaz A, Shin DS, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N Engl J Med*. 2016;375(9):819-829.
86. McGranahan N, Rosenthal R, Hiley CT, et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell*. 2017;171(6):1259-1271.e11.