

RESEARCH ARTICLE

Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis

Alexandra Snyder^{1,2}*, Tavi Nathanson³, Samuel A. Funt^{1,2}, Arun Ahuja³, Jacqueline Buros Novik³, Matthew D. Hellmann^{1,2}, Eliza Chang³, Bulent Arman Aksoy³, Hikmat Al-Ahmadie⁴, Erik Yusko⁵, Marissa Vignali⁵, Sharon Benzeno⁵, Mariel Boyd¹, Meredith Moran¹, Gopa Iyer^{1,2}, Harlan S. Robins^{5,6}, Elaine R. Mardis⁷, Taha Merghoub¹, Jeff Hammerbacher^{3i,1/2}, Jonathan E. Rosenberg^{1i,1/2}, Dean F. Bajorin^{1,2i,1/2}

1 Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, **2** Department of Medicine, Weill Cornell Medical College, New York, New York, United States of America, **3** Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, **4** Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, **5** Adaptive Biotechnologies, Seattle, Washington, United States of America, **6** Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **7** Department of Pediatrics, Institute for Genomic Medicine, The Research Institute at Nationwide Children's Hospital, The Ohio State University College of Medicine, Columbus, Ohio, United States of America

These authors contributed equally to this work.

^{i,2} These authors are joint senior authors on this work.

* snyderca@mskcc.org

OPEN ACCESS

Citation: Snyder A, Nathanson T, Funt SA, Ahuja A, Buros Novik J, Hellmann MD, et al. (2017) 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* 14(5): e1002309. <https://doi.org/10.1371/journal.pmed.1002309>

Academic Editor: John D. Minna, University of Texas Southwestern Medical Center at Dallas, UNITED STATES

Received: November 2, 2016

Accepted: April 26, 2017

Published: May 26, 2017

Copyright: © 2017 Snyder et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available at <http://doi.org/10.5281/zenodo.546110>.

Funding: This work was supported by a grant from the Ludwig Center for Cancer Research (<http://www.ludwigcancerresearch.org/>) and the NIH/NCI Cancer Center Support Grant P30 CA008748. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Abstract

Background

Inhibition of programmed death-ligand 1 (PD-L1) with atezolizumab can induce durable clinical benefit (DCB) in patients with metastatic urothelial cancers, including complete remissions in patients with chemotherapy refractory disease. Although mutation load and PD-L1 immune cell (IC) staining have been associated with response, they lack sufficient sensitivity and specificity for clinical use. Thus, there is a need to evaluate the peripheral blood immune environment and to conduct detailed analyses of mutation load, predicted neoantigens, and immune cellular infiltration in tumors to enhance our understanding of the biologic underpinnings of response and resistance.

Methods and findings

The goals of this study were to (1) evaluate the association of mutation load and predicted neoantigen load with therapeutic benefit and (2) determine whether intratumoral and peripheral blood T cell receptor (TCR) clonality inform clinical outcomes in urothelial carcinoma treated with atezolizumab. We hypothesized that an elevated mutation load in combination with T cell clonal dominance among intratumoral lymphocytes prior to treatment or among peripheral T cells after treatment would be associated with effective tumor control upon

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests. ERM is a member of the Editorial Board of PLOS Medicine. SAF declared stock ownership in Kite Pharmaceuticals. MDH declared research grants from BMS and Genentech and paid consultancy from Genentech, Merck, BMS, AstraZeneca, Janssen and Neon. EY declared employment at Adaptive Biotechnologies. MV declared employment by Adaptive Biotechnologies Corp. with salary and stock options. SB declared ownership of stocks or shares and paid employment at Adaptive Biotechnologies. HR declared employment, equity ownership, patents, and royalties with Adaptive Biotechnologies. JER declared consultancy for Roche/Genentech, Bristol Myers Squibb, Oncogenex, Eli Lilly, AstraZeneca, Merck, Sanofi and Boehringer-Ingelheim, and stock in Merck and Illumina. DFB declared consultancy for Bristol Myers Squibb, Genentech-Roche, Pfizer, Merck, Novartis and Eli-Lilly; research support from Bristol Myers Squibb, Genentech-Roche, Merck, Novartis and Amgen; and travel support from Eli-Lilly, Genentech and Merck. AS declared research grants from BMS and paid consultancy for BMS and Driver. JER declared research grants from Neon and paid consultancy for AstraZeneca.

Abbreviations: AUC, area under the curve; BCG, Bacillus Calmette-Guérin; BQSR, Base Quality Score Recalibration; DCB, durable clinical benefit; DCB-OS, overall survival greater than 12 months; DDR, DNA damage response; FDR, false discovery rate; FFPE, formalin-fixed paraffin embedded; GSEA, Gene Set Enrichment Analysis; HR, hazard ratio; IC, immune cell; OS, overall survival; PBMC, peripheral blood mononuclear cell; PD-L1, programmed death-ligand 1; PFS, progression-free survival; RNA-seq, RNA sequencing; TCR, T cell receptor; TCR-seq, T cell receptor sequencing; TIL, tumor-infiltrating T lymphocytes; VAF, variant allele frequency; WES, whole exome sequencing.

treatment with anti-PD-L1 therapy. We performed whole exome sequencing (WES), RNA sequencing (RNA-seq), and T cell receptor sequencing (TCR-seq) of pretreatment tumor samples as well as TCR-seq of matched, serially collected peripheral blood, collected before and after treatment with atezolizumab. These parameters were assessed for correlation with DCB (defined as progression-free survival [PFS] >6 months), PFS, and overall survival (OS), both alone and in the context of clinical and intratumoral parameters known to be predictive of survival in this disease state.

Patients with DCB displayed a higher proportion of tumor-infiltrating T lymphocytes (TIL) ($n = 24$, Mann-Whitney $p = 0.047$). Pretreatment peripheral blood TCR clonality below the median was associated with improved PFS ($n = 29$, log-rank $p = 0.048$) and OS ($n = 29$, log-rank $p = 0.011$). Patients with DCB also demonstrated more substantial expansion of tumor-associated TCR clones in the peripheral blood 3 weeks after starting treatment ($n = 22$, Mann-Whitney $p = 0.022$). The combination of high pretreatment peripheral blood TCR clonality with elevated PD-L1 IC staining in tumor tissue was strongly associated with poor clinical outcomes ($n = 10$, hazard ratio (HR) (mean) = 89.88, HR (median) = 23.41, 95% CI [2.43, 506.94], $p(\text{HR} > 1) = 0.0014$). Marked variations in mutation loads were seen with different somatic variant calling methodologies, which, in turn, impacted associations with clinical outcomes. Missense mutation load, predicted neoantigen load, and expressed neoantigen load did not demonstrate significant association with DCB ($n = 25$, Mann-Whitney $p = 0.22$, $n = 25$, Mann-Whitney $p = 0.55$, and $n = 25$, Mann-Whitney $p = 0.29$, respectively). Instead, we found evidence of time-varying effects of somatic mutation load on PFS in this cohort ($n = 25$, $p = 0.044$). A limitation of our study is its small sample size ($n = 29$), a subset of the patients treated on IMvigor 210 (NCT02108652). Given the number of exploratory analyses performed, we intend for these results to be hypothesis-generating.

Conclusions

These results demonstrate the complex nature of immune response to checkpoint blockade and the compelling need for greater interrogation and data integration of both host and tumor factors. Incorporating these variables in prospective studies will facilitate identification and treatment of resistant patients.

Author summary

Why was this study done?

- A new type of cancer treatment called checkpoint blockade therapy activates the immune system to fight cancer.
- When these therapies work, patients with advanced disease can experience long-lasting disease control or even cures.
- However, most patients will not experience these benefits, and it is crucial to identify these patients in advance so that we can develop better treatments for them.

What did the researchers do and find?

- i In this study, we studied 29 patients with advanced bladder cancers treated with a checkpoint blockade drug called atezolizumab.
- i We examined features of the tumor and the immune system, as well as clinical features.
- i We found that these features were related to each other, and to the success of therapy, in various ways.
- i Patients who had a diverse repertoire of T cells in their blood tended to survive longer. Patients who had poor clinical prognostic factors, like having cancer that had traveled to their liver, tended to have worse survival.

What did the research findings mean?

- i This study demonstrates that we need to take the tumor, immune system, and clinical picture into account if we are to improve the efficacy of immune-mobilizing therapies in cancer.
- i Some patients may be too sick to benefit from checkpoint blockade therapy, despite, in some cases, having biomarkers in their tumors that would predict benefit.

Introduction

Atezolizumab has demonstrated responses in 15%–25% of patients with advanced urothelial carcinoma and improved survival compared to historical expectations [1,2]. Similar to predictive factor analyses in melanoma, colon cancer, and non-small cell lung cancer studies with other checkpoint blockade agents, Rosenberg and colleagues reported a statistically significant association between mutation load and response to atezolizumab in urothelial cancer patients [2]. However, mutation load in the atezolizumab study was predicted based on an estimate using a targeted panel and not with whole exome sequencing (WES). Similar to findings from prior studies, the association between this predicted mutation load and outcomes in patients with urothelial cancer was not dichotomous; there were tumors from patients with elevated mutation load that did not respond to therapy, and vice versa. Additionally, positive programmed death-ligand 1 (PD-L1) staining of infiltrating immune cells by immunohistochemistry was associated with, but poorly predicted, response. A statistical model suggested that both PD-L1 staining and mutation load impacted the likelihood of response. However, the authors did not recommend their clinical use.

Collectively, studies to date imply that a combination of immune parameters are necessary to gain further precision in determining the likelihood of benefit from these immunotherapies and that a single biologic marker will be insufficient. There have been few attempts to integrate molecular and immunologic data from patients treated with checkpoint blockade and their tumors. Consequently, we performed whole exome, RNA, and T cell receptor (TCR) sequencing (TCR-seq) of tumor samples from patients treated with atezolizumab, as well as TCR-seq of matched, serially collected peripheral blood.

Methods

Ethics statement

All research involving human participants was approved by the authors' Institutional Review Board (MSKCC IRB), and all clinical investigation was conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent was obtained from the participants.

Analysis plan

The analyses that were and were not included in the prespecified analysis plan are detailed in [S2 Text](#).

Patients and clinical characteristics

All patients had locally advanced or metastatic urothelial carcinoma and were treated at Memorial Sloan Kettering Cancer Center ($n = 29$) on protocol NCT02108652 [2]. All patients initiated therapy in 2014, were treated with atezolizumab 1,200 mg IV every 3 weeks, and provided written consent according to Institutional Review Board-approved protocols permitting tissue and blood collection, sequencing, and correlative studies. Patient tumor samples were assessed prospectively and centrally (by HistoGeneX, Brussels, Belgium) for PD-L1 expression by immunohistochemistry with the SP142 assay (Ventana, AZ, USA) [1]. The PD-L1 tumor-infiltrating immune cell (IC) status was defined by the percentage of PD-L1⁺ cells in the tumor microenvironment: IC0 (<1%), IC1 (1% but <5%), and IC2/3 (≥5%), as defined in the original study. Six patients had multiple samples evaluated for PD-L1 IC status; for 4 of these patients, the sample used for PD-L1 IC status in this analysis was the same as the sample that was whole exome sequenced. For the remaining 2, the PD-L1 status of 1 patient's tumor (1994) used a separate primary tumor sample that agreed in status with a metastatic sample; the other patient's tumor (6229) used a metastatic sample site that agreed in status with another metastatic sample site. Smoking status was evaluated using previously completed self-reported smoking questionnaires or review of medical records. One patient was excluded because the patient did not consent to correlative studies beyond PD-L1 testing that was performed as part of the clinical trial.

Tumor and blood samples

All tumor tissue used for sequencing was obtained prior to dosing with atezolizumab. Tumor samples used for whole exome sequencing were all formalin-fixed paraffin-embedded (FFPE). The presence of tumor tissue in the sequenced samples was confirmed by examination of a representative hematoxylin and eosin-stained slide by a genitourinary pathologist (H.A.). Peripheral blood mononuclear cells (PBMCs) were isolated and stored as previously described [3]. PBMCs were collected pretreatment and during treatment.

Clinical efficacy analysis

Tumor responses to atezolizumab were evaluated by CT scan every 9 weeks for the first 12 months following day 1 of cycle 1. After 12 months, tumor assessments were performed every 12 weeks. The response evaluation criteria in solid tumors (RECIST) version 1.1 was used to define objective clinical responses by the institutional radiologist.

DNA extraction and high-throughput TCR sequencing

Tumor samples from patients 0522 and 6800 were excluded from tumor TCR analyses after failing quality control. Patients 0979, 7592, and 8214 did not have available tumors for TCR sequencing and were therefore excluded from tumor TCR analyses as well. Genomic DNA was purified from total PBMCs and tumor samples using the Qiagen DNeasy Blood extraction kit. The TCR CDR3 regions were amplified and sequenced using immunoSEQ1 (Adaptive Biotechnologies, Seattle, WA), as previously described [4]. In brief, bias-controlled V and J gene primers were used to amplify rearranged V(D)J segments for high-throughput sequencing at approximately 20X coverage. After correcting sequencing errors via a clustering algorithm, CDR3 segments were annotated according to the International ImMunoGeneTics Collaboration [5] to identify the V, D, and J genes that contributed to each rearrangement. A mixture of synthetic TCR analogs in each PCR was used to estimate the absolute template abundance (i.e., the number of cells bearing each unique TCR sequence) from sequencing data, as previously described [6]. To determine TIL content in FFPE samples as a T cell fraction, we amplified several housekeeping genes and quantitated their template counts to determine the amount of DNA usable for TCR sequencing. ImmunoSEQ then amplifies and sequences the molecules with rearranged TCR chains. Because the immunoSEQ assay aligns sequences to the IMGT database, sequences are annotated as complete VDJ rearrangements or nonproductive rearrangements (a stop codon or out of frame CDR3 region was generated during VDJ recombination in 1 of the alleles); all downstream analysis in this work proceeded with complete, productive sequences. To estimate the number of starting templates that were in the sample, the number of sequence reads for each TCR sequence is measured. Synthetic control templates were also spiked into each sample, thereby enabling quantitation of input TCR templates from the read counts. For each sample, Shannon entropy was also calculated on the clonal abundance of all productive TCR sequences in the data set. Shannon entropy was normalized to the range by dividing Shannon entropy by the logarithm of the number of unique productive TCR sequences in the data set. This normalized entropy value was then inverted to produce the clonality metric. Those T cell clones whose frequencies differed between samples from a given subject taken at different time points, or between cell populations (e.g., between total PBMCs and tumor), were computationally identified as previously described [9]. The input data consisted of the absolute abundance for each TCR clone in each sample. Fisher's exact test was used to compute a *p*-value for each clone across the 2 samples against the null hypothesis that the population abundance of the clone is identical in the 2 samples. We corrected for multiple testing to control the false discovery rate (FDR) using the Benjamini-Hochberg procedure and employed a significance threshold of 0.01 on adjusted *p*-values.

Whole exome sequencing

Twenty-six FFPE-derived tumor and frozen PBMC-derived normal paired samples were sequenced by exome hybrid capture, using the IDT xGen Whole Exome Panel (<https://www.idtdna.com/pages/products/nextgen/target-capture/xgen-lockdown-panels/xgen-exome-panel>) and standard protocols. Briefly, each sample was used to create a barcoded Illumina library, tumor samples were pooled at optimal multiplex to create an equimolar pool into the hybrid capture reaction, which was performed according to the manufacturer's suggested protocol. Similarly, normal samples were pooled and introduced to the hybrid capture reaction. Following the recovery of captured library fragments, PCR amplification was performed, the resulting pools of fragments were quantitated using qPCR (Kapa Bio) and sequenced in

separate lanes by paired-end 150 bp reads, using the Illumina HiSeq 4000. Whole exome sequencing results for 1 sample (for patient 4072) were excluded after failing to meet coverage requirements.

Somatic variant calling

DNA sequencing data for the tumor and normal samples were aligned to the GRCh37 reference using *bwa-mem* (v. 0.7.10) with default settings. The resulting BAMs were processed through Picard MarkDuplicates and the GATK (v. 3.5i₂10) pipeline, including Base Quality Score Recalibration and Indel Realignment. Single nucleotide variants (SNVs) were called from Mutect (v. 1.1.6) and Strelka (v. 1.0.14) with default settings. Variants from either call were included and the variants calls were further filtered to those with depth (in normal and tumor samples) ≥ 7 reads, $>10\%$ tumor variant allele frequency (VAF), and $\geq 3\%$ normal VAF [10]. Mutations per megabase was computed by normalizing the number of mutations by the number of exonic loci with ≥ 7 reads in normal and tumor samples, calculated using Pageant (<https://github.com/hammerlab/pageant>).

Variants were annotated as missense variants by Varcode (v. 0.5.10, <https://github.com/hammerlab/varcode>) and PyEnsembl (v. 1.0.3, <https://github.com/hammerlab/pyensembl>) using Ensembl Release 75 and annotated as deleterious using PolyPhen (v. 2.2.2). DNA damage response (DDR) genes were gathered from [11,12].

Mutational signatures were inferred from the somatic mutation calls using deconstructSigs (v 1.6.0).

Human leukocyte antigen typing

Human leukocyte antigen (HLA) types for each patient were computed from the normal sequencing data using OptiType (v. 1.0.0).

RNA sequencing

RNA was extracted from 26 FFPE tumor samples and evaluated for quality and quantity using the Agilent RNA pico chip. Each sample was prepared for sequencing by constructing an Illumina Tru-Seq Stranded RNA kit, according to the manufacturer's protocol. The resulting libraries were amplified by PCR, quantitated, pooled, and processed through a hybrid capture intermediate using the IDT xGen Exome reagent (as above). The captured fragments were quantitated, diluted, and were sequenced using 2 x 150 bp paired-end reads on the Illumina HiSeq 4000.

The RNA sequencing (RNA-seq) data were aligned to the GRCh37 reference in Ensembl Release 75 using STAR (v. 2.4.1d), and transcript quantification was performed using kallisto (v. 0.42.3). The STAR alignment was only used for identifying variant-supporting reads in the RNA. For gene-level analysis, the transcript quantifications were aggregated to the gene level using tximport (<http://f1000research.com/articles/4-1521/v1>).

Expressed mutations and neoantigens were computed using Isovar (v. 0.0.6, <https://github.com/hammerlab/isovar>), based on the RNA reads overlapping each mutation. Sleuth (v. 0.28.1) was used for differential expression analysis, and Gene Set Enrichment Analysis (GSEA) was used for pathway enrichment analysis. ESTIMATE was used to quantify immune and stromal scores from RNA-seq data.

Neoantigen calling

Neoantigens were computed from all nonsynonymous mutations using Topiary (v. 0.1.0, <https://github.com/hammerlab/topiary>) and NetMHCCons (v. 1.1) with HLA alleles calculated by

OptiType. As with expressed mutations, expressed neoantigens were those supported in the RNA with at least 3 uniquely-mapped reads matching the cDNA sequence.

Statistical analysis

All statistical analysis was performed in Python and R (v. 3.3.1). Cohorts (v. 0.4.0, <https://github.com/hammerlab/cohorts>) and Biokepi (<https://github.com/hammerlab/biokepi>) were used to orchestrate the analysis. The Mann Whitney and Fisher's Exact test were performed using the Python scientific computing library, SciPy (v. 0.18.1). Kaplan-Meier curves were computed with Lifelines (v. 0.9.1.0). Survival and logistic regression models were estimated using PyStan (v. 2.12.0.0), and the Stan statistical computing software (v. 2.12.0). Survival analyses utilized a proportional hazards piecewise exponential model with a random walk baseline hazard. The analysis for presence of a time-varying covariate effect was performed in R using survival (v. 2.39.5) to look for the association of scaled Schoenfeld residuals with log(time), whereas the estimation of the time-varying covariate effect was performed using Stan. This analysis estimated the covariate effect at each timepoint with a random-walk prior. In some cases, alternative specifications of models written in Stan were interrogated as sensitivity analyses; see the project's GitHub repository (<https://github.com/hammerlab/multi-omic-urothelial-anti-pdl1>) and [S3 Text](#) for details.

All analysis code is available at <https://github.com/hammerlab/multi-omic-urothelial-anti-pdl1> for open access by readers.

Mutation calls, TCR-seq, and RNA-seq data are available at <http://doi.org/10.5281/zenodo.546110>. Additional data are available at <https://github.com/hammerlab/multi-omic-urothelial-anti-pdl1>.

Results

Patient characteristics

Twenty-nine patients with metastatic urothelial cancer from a single institution, treated with atezolizumab, as part of a single-arm phase II study (IMvigor 210, NCT 02108652), were included in the analyses. The patients displayed characteristics typical of the metastatic urothelial cancer population studied in IMvigor 210: 25 of 29 were males with urothelial cancers of bladder origin, and 18 of 29 had a reported prior smoking history ([Table 1](#)). Patients had an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 and had 0 to 3 prior regimens of chemotherapy. Of this group, 25 patients had sufficient tumor tissue for WES, 26 for RNA-seq, and 24 for TCR-seq. Twenty-nine had a pretreatment peripheral blood sample on which TCR-seq could be performed; 24 had 1 pretreatment and at least 1 posttreatment peripheral blood collection.

Intratumoral and peripheral TCR features associate with durable clinical benefit

The importance of T cells to the anti-tumor response has long been known [15]; the relevance of intratumoral and peripheral TCR clonality to the anti-tumor response is an area of active study. A single previous study of melanoma patients treated with anti-PD-1 therapy demonstrated that patients whose tumors featured both high levels of tumor-infiltrating T lymphocytes (TIL) along with high TIL clonality were more likely to experience radiographic response to therapy [8]. A separate study examined the peripheral TCR repertoire in anti-cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) treated patients with prostate cancer or melanoma and found that clonotype stability was associated with response [16]. To our knowledge, no

prior study has reported both intratumoral and peripheral TCR clonality in a single population treated with checkpoint blockade therapy.

We performed TCR-seq of tumors and PBMCs at serial time points in our cohort. Due to limitations in sample availability (Methods), this analysis included tumors from 24 patients and peripheral blood from 29 patients, including pretreatment samples in all patients, and between 1 and 8 total time points. A median of 141,255 (range 43,052–335,087) cells were analyzed per peripheral blood sample, including 82,636 (range 24,095–207,860) unique TCRs,

Table 1. Baseline characteristics.

Variable	Durable Clinical Benefit (n = 9)	No Durable Benefit (n = 20)
Median age (range)	66 (57–74)	72 (46–81)
Male sex	9 (100%)	16 (80%)
Site of primary tumor		
Bladder	7 (78%)	19 (95%)
Upper tract	2 (12%)	1 (5%)
ECOG performance status		
0	0 (0%)	1 (5%)
1	9 (100%)	19 (95%)
History of tobacco use		
No	4 (44%)	7 (35%)
Yes	5 (56%)	13 (65%)
Hemoglobin concentration <10 gm/dL	1 (11%)	3 (15%)
Albumin lower limit of normal	2 (22%)	5 (25%)
Metastatic sites at baseline		
Visceral [^]	3 (33%)	16 (80%)
Liver	2 (22%)	9 (45%)
Lymph node only	6 (67%)	4 (20%)
Number of previous systemic regimens in the metastatic setting		
0	5 (56%)	3 (20%)
1	2 (22%)	15 (70%)
2	1 (11%)	1 (5%)
3	1 (11%)	1 (5%)
Previous neoadjuvant or adjuvant chemotherapy, with progression within 12 months	3 (33%)	2 (10%)
Time since previous chemotherapy < 3 months	2 (29%)	5 (26%)
Prognostic risk group for previously treated patients*		
Low	3 (42%)	7 (37%)
Intermediate	2 (29%)	5 (26%)
High	2 (29%)	7 (37%)
Prognostic risk group for previously untreated patients [†]		
Low	0 (0%)	0 (0%)
Intermediate	2 (100%)	0 (0%)
High	0 (0%)	1 (100%)
Intravesical BCG administered (%)	2 (22%)	10 (50%)

[^]Visceral metastasis defined as liver, lung, bone, or any non-lymph node or soft tissue metastasis.

*Based on [13].

[†]Based on [14].

Abbreviations: BCG, Bacillus Calmette-Guérin; ECOG, Eastern Cooperative Oncology Group.

<https://doi.org/10.1371/journal.pmed.1002309.t001>

with a median clonality of 0.080 (range 0.014–0.37) and a median T cell proportion of 0.31 (range 0.082–0.64) in the tumors, the corresponding values included 1,402 (range 63–133,167) T cells, 1,086 (range 67–56,273) unique TCRs, clonality of 0.096 (range 0.033–0.34), and T cell proportion of 0.097 (range 0.0098–0.33).

In our patient group, we first asked whether there was an association between outcome and either TIL clonality or TIL proportion, or with clonality in the peripheral blood. Consistent with the data from Tumei and colleagues [8], tumors from patients who experienced a durable clinical benefit (DCB) exhibited a higher TIL proportion than those patients who experienced progressive disease, with a median of 0.21 (range 0.049–0.35) tumors from patients who had progression-free survival (PFS) greater than 6 months versus 0.069 (range 0.0098–0.24) tumors from patients who did not ($n = 24$, Mann-Whitney $p = 0.047$, Fig 1A). The consistency of results is notable given that Tumei and colleagues used a different methodology (IHC) to assess TIL proportion than was used in our study. However, TIL proportion was not associated with continuous PFS ($n = 24$, log-rank $p = 0.32$) or OS ($n = 24$, log-rank $p = 0.26$) when split by median proportion. TIL clonality alone was not significantly associated with DCB ($n = 24$, Mann-Whitney $p = 0.10$, Fig 1A), continuous PFS (split by median clonality, $n = 24$, log-rank $p = 0.51$), or continuous OS (split by median clonality, $n = 24$, log-rank $p = 0.47$). Tumors with less than the median TIL proportion or TIL clonality, considered jointly as 1 feature, were less likely to display DCB (Fig 1B, 25% of patients with DCB versus 81% of patients without DCB, $n = 24$, Fisher's Exact $p = 0.021$). Considering TIL proportion and TIL clonality separately, when using median as a threshold, did not result in a significant difference in terms of DCB in either case (S1A Fig). It remains unclear whether TIL clonality adds to TIL proportion in its association with DCB in this study (TIL proportion and TIL clonality versus TIL proportion alone, $n = 24$, log-likelihood $p = 0.100$).

We next examined pretreatment peripheral blood clonality and its relationship to DCB. Because a diverse TCR repertoire in circulation may increase the likelihood that a tumor-specific T cell population is present, we hypothesized that TCR clonality would be inversely associated with response. We found that low pretreatment peripheral TCR clonality was associated with improved PFS (split by median clonality, $n = 29$, log-rank $p = 0.048$), overall survival (OS, split by median clonality, $n = 29$, log-rank $p = 0.011$), and OS greater than 12 months (DCB-OS, $n = 29$, Mann-Whitney $p = 0.0061$; Fig 1C, 1D and 1E), although not with DCB (Fig 1F, $n = 29$, Mann-Whitney $p = 0.25$).

Finally, we explored the relationship between intratumoral and peripheral TCR clonality. Individual T cell clones present in tumors can be tracked in the peripheral blood during treatment (examples in S1B Fig). Expansion of tumor-associated TCRs occurred in the peripheral blood in all patients (Fig 1G). However, a more pronounced expansion of intratumoral TCR clones was observed in DCB patients at 3 weeks after initiation of treatment (second dose of therapy) ($n = 22$, Mann-Whitney $p = 0.022$, Fig 1H) that was not significant at 6 weeks after therapy initiation ($n = 20$, Mann-Whitney $p = 0.17$, S1C Fig). Interestingly, all patients with low pretreatment peripheral TCR clonality and high TIL clonality survived greater than 12 months (DCB-OS, S1D Fig).

Association of tumor genetic features with PFS or OS

To further examine intratumoral factors associated with therapeutic efficacy, we performed WES on 25 FFPE archived tumor samples. Mean target coverage was 129 (range 44–914) tumors and 73 (range 59–914) normal tissue. SNVs were identified and annotated as silent, missense, or nonsense mutations (Fig 2A). There was no significant association between median missense mutation load and DCB (median mutations per megabase 3.24 [range 0.038–14.2]

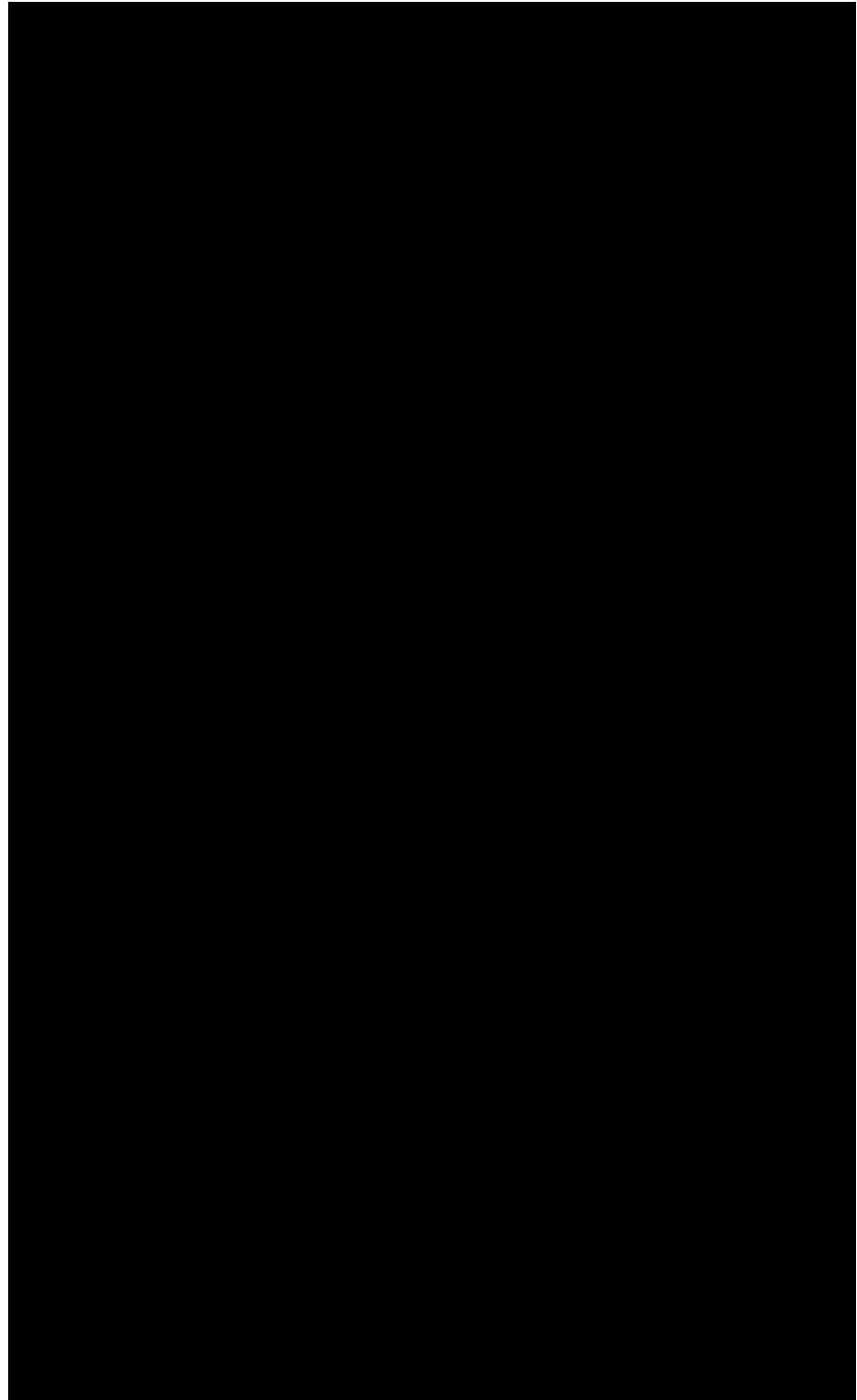


Fig 1. T cell receptor (TCR) clonality and treatment response. (A) Tumor-infiltrating T lymphocytes (TIL) proportion alone was associated with durable clinical benefit (DCB), with a median of 0.21 (range 0.049; 0.28) in tumors from patients who had DCB versus 0.069 (range 0.0098; 0.20) in tumors from patients who did not ($n = 24$, Mann-Whitney $p = 0.047$). TIL clonality alone was not significantly associated with DCB, with a median of 0.12 (range 0.047; 0.3) in tumors from patients with DCB and a median of 0.092 (range 0.033; 0.12) in tumors from patients who did not have DCB.

0.22) in tumors from patients without DCB ($n = 24$, Mann-Whitney $p = 0.10$). (B) Tumors with less than the median TIL proportion or TIL clonality, considered jointly as 1 feature, were less likely to display DCB (Fig 1B, 25% of patients with DCB versus 81% of patients without DCB, $n = 24$, Fisher's Exact $p = 0.021$). (C) Patients with pretreatment peripheral TCR clonality less than the median exhibited improved progression-free survival (PFS, $n = 29$, log-rank $p = 0.048$). (D) Patients with pretreatment peripheral TCR clonality less than the median exhibited improved overall survival (OS, $n = 29$, log-rank $p = 0.011$). (E) There was a significant association between TCR clonality in the peripheral blood prior to initiating treatment and overall survival greater than 12 months (DCB-OS) (DCB-OS: TCR clonality 0.060 [range 0.022 to 0.20]; OS less than 12 months: 0.15 [range 0.031 to 0.35], $n = 29$, Mann-Whitney $p = 0.0061$). (F) There was no significant association between pretreatment peripheral TCR clonality and DCB (DCB: TCR clonality 0.068 [range 0.027 to 0.20]; non-DCB: 0.14 [range 0.022 to 0.35], $n = 29$, Mann-Whitney $p = 0.25$). (G) Expansion of TCR clones found in TIL (orange bars) occurred in the peripheral blood 3 weeks after initiating treatment in all patients. (H) The number of TCR clones found in TIL that expanded in the peripheral blood 3 weeks after initiating treatment was 8.00 (range 4.00 to 12.00) in patients with DCB and 2.50 (range 1.00 to 4.00) in non-DCB patients ($n = 22$, Mann-Whitney $p = 0.022$).

<https://doi.org/10.1371/journal.pmed.1002309.g001>

11.46] in patients with DCB compared to 0.45 [range 0.019 to 9.90] those without DCB, $n = 25$, Mann-Whitney $p = 0.22$, Fig 2B). There was also no significant association between missense mutation load and DCB-OS ($n = 25$, Mann-Whitney $p = 0.37$, S2A Fig). In a survival analysis for time to disease progression or mortality, the estimated hazard ratio associated with increase in missense SNV count per megabase was 0.92 (95% CI [0.78, 1.09]). These results are not surprising given that the present sample size ($n = 25$) is underpowered to detect an effect of magnitude similar to that observed by Rosenberg and colleagues [2] (power = 0.2, assuming median of 12.4 versus 6.4 mutations per megabase among patients with DCB versus non-DCB response).

When filtering to expressed mutations, we found a median of 0.79 (range 0.00 to 3.36) expressed mutations per megabase for patients with DCB and a median of 0.16 (range 0.00 to 3.34) expressed mutations per megabase for patients without DCB ($n = 25$, Mann-Whitney $p = 0.26$, S2B Fig). Consistent with the known importance of specific variant calling pipelines to output [17,18], we found that different filtering techniques impacted the association with DCB (S1 Table). Missense mutation load, when counting only mutations that were removed after postprocessing (via Base Quality Score Recalibration [BQSR] and depth/VAF filtering), was predictive of response ($n = 25$, Mann-Whitney $p = 0.0078$).

One hypothesis for explaining the association between mutation load and outcome to treatment with checkpoint blockade is the generation of neoantigens, altered peptides presented by the major histocompatibility complex that are capable of eliciting an antitumor T cell response and are more common with increased mutation load. After performing in silico HLA typing (Methods), we examined predicted neoantigens that were 8 to 11 amino acids in length, resulting from the missense mutations of patients treated with atezolizumab. There was no significant association between predicted neoantigens per megabase and either DCB or DCB-OS. Patients with DCB had a median 4.58 (range 0.037 to 39.48) predicted neoantigens per megabase, while patients without DCB had 1.35 (range 0.00 to 20.20) ($n = 25$, Mann-Whitney $p = 0.55$, S2C Fig and S2D Fig). Filtering of predicted neoantigens to focus only on those expressed in RNA (Methods) also demonstrated no significant association between expressed predicted neoantigens and clinical benefit with atezolizumab ($n = 25$, Mann-Whitney $p = 0.29$, Fig 2C and S2B Fig). Here, too, we acknowledge the limitations in statistical power to detect associations due to the sample size of our study.

The association between mutation load and response likelihood strengthens over time

Given that the mutation load and outcomes were weakly associated in the complete IMvigor210 dataset and not statistically significantly associated in this cohort, we embarked upon

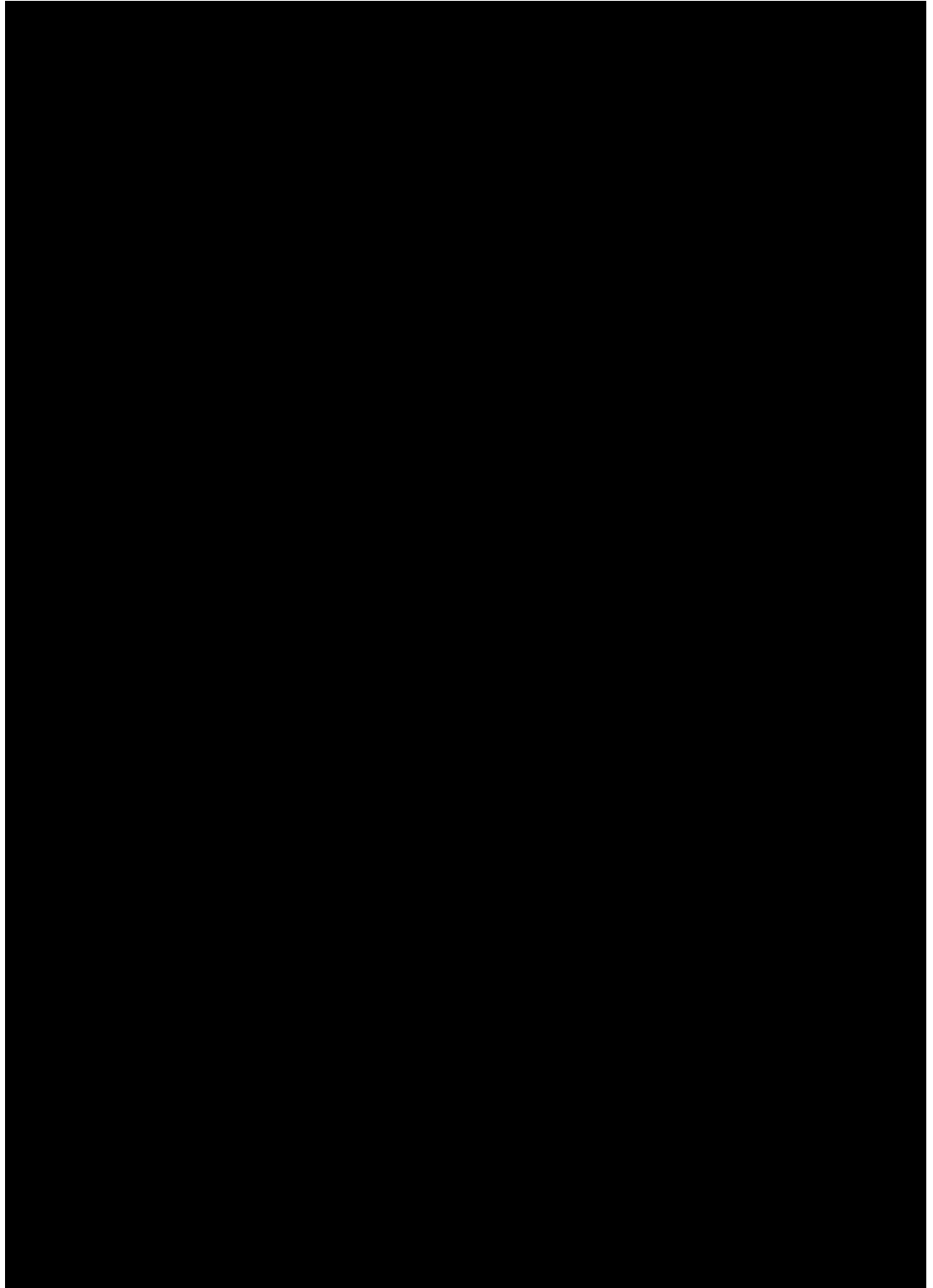


Fig 2. Single nucleotide variants (SNVs) and treatment response. (A) SNVs, premature stop codons, transversions, mutations in start or stop codons and splice site variants, as well as transitions and transversions were called for all samples. (B) Median mutations per megabase of 3.24 (range 0.038 to 11.46) tumors from patients who progressed at or after 6 months as compared to 0.45 (range 0.019 to 1.99) in those who progressed in less than 6 months ($n = 25$, Mann-Whitney $p = 0.22$). (C) Median expressed neoantigens in tumors from patients who progressed at or after 6 months was 1.32 (range 0.00 to 6.06) versus 0.29 (range 0.00 to 0.67) in those who progressed before 6 months ($n = 25$, Mann-Whitney $p = 0.29$).

<https://doi.org/10.1371/journal.pmed.1002309.g002>

an exploration of additional factors, including tumor microenvironmental and systemic measures, which may modify the importance of this variable or independently affect outcomes.

To this end, we examined the time-varying association between mutation load and PFS to see if mutation load had a differential association with early hazards in contrast to late hazards. We found evidence of time-varying effects of somatic mutation load on PFS in this cohort ($n = 25$, $p = 0.044$, see [Methods](#)). We estimated these effects to consist of a stronger association of somatic mutation load with reduced mortality or disease progression more than 3 months after treatment ($n = 11$, hazard ratio (HR) = 0.69, 95% CI [0.38, 0.99]) as compared to that during the first 3 months ($n = 25$, HR = 0.91, 95% CI [0.75, 1.07]; [Fig 3A](#)). This effect estimate yielded a p -value for interaction of 0.1, which does not contradict the test for presence of the effect ($n = 25$, $p = 0.044$) because that test is better powered. When a similar analysis was performed for time-varying association with OS, the evidence in support of the existence of time-varying effects was similar ($n = 25$, $p = 0.082$; notable, despite $p > 0.05$, given the PFS results above). In terms of the estimate of these effects, patients who survived longer than 3 months exhibited a stronger association between the number of somatic mutations per megabase and a lower risk of subsequent mortality ($n = 11$, HR = 0.80, 95% CI [0.60, 1.00]) as compared to those who survived less than or equal to 3 months ($n = 25$, HR = 1.02, 95% CI [0.79, 1.22], [Fig 3B](#)). This suggests that the time-varying effect is not likely an artifact of differential association with survival versus progression. Looking at the Kaplan-Meier estimates of PFS among patients with mutation load per megabase above and below the median value of 1.03, it is apparent that there is very little separation of these 2 populations until approximately 3 months after treatment and that there is a high frequency of both progression and mortality events at this time ([Fig 3C](#)). While we report these results using a prespecified threshold of 3 months, in a nonparametric analysis we found that the reduction in risk associated with somatic mutation load increased steadily over time without the emergence of a clear inflection point ([S3C Fig](#); see [S3 Text and S3D Fig](#) for further model interrogation). That said, we note that the 95% confidence intervals around the hazard ratios prior to 3 months include 1 (all p -values > 0.05 ; aggregate $p(\text{HR} > 1) = 0.21$), while those after 3 months are significantly less than 1 (all p -values < 0.05 ; aggregate $p(\text{HR} > 1) = 0.020$; $p = 0.07$ for interaction comparing aggregate HRs). This suggests that our threshold of 3 months, which was selected based on clinical experience, may be a convenient summary of these results.

These data suggest that in patients with rapidly progressive disease, factors other than mutation load likely determine their outcome. This observation is not surprising in that clinical factor analysis of this disease state has identified a heterogeneous population of patients, with 5 clinical factors distinguishing those likely to experience a rapid and early death from those more likely to survive longer [13]. We hypothesized that such patients might simply be too clinically and systemically unwell to mount the necessary immune response, despite some of them harboring tumor biomarkers thought to confer a likelihood of DCB, including elevated mutation load. When we examined the 5-factor score in this subset relative to the rest of the dataset, we found that, indeed, patients who survived less than or equal to 3 months exhibited a significantly higher 5-factor score (3.00 (range 2.00; 4.00)) contrast to 1.50 (range 0.00; 4.00) in patients who survived longer than 3 months ($n = 26$, Mann-Whitney $p = 0.018$, [S3A Fig](#)). Patients surviving less than 3 months were much more likely to have liver metastases: 100% in patients surviving less than or equal to 3 months and 22% in patients surviving longer than 3 months ($n = 29$, Fisher's Exact $p = 0.00097$, [S3B Fig](#)). There were no significant differences in these patients with respect to Bacillus Calmette-Guérin (BCG) exposure ($n = 29$, Fisher's Exact $p = 0.20$), missense SNV load ($n = 25$, Mann-Whitney $p = 0.26$), and pretreatment peripheral TCR clonality ($n = 29$, Mann-Whitney $p = 0.12$). Keeping in mind the limited sample size of this cohort, these data suggest that there is a subset of nearly end-stage patients

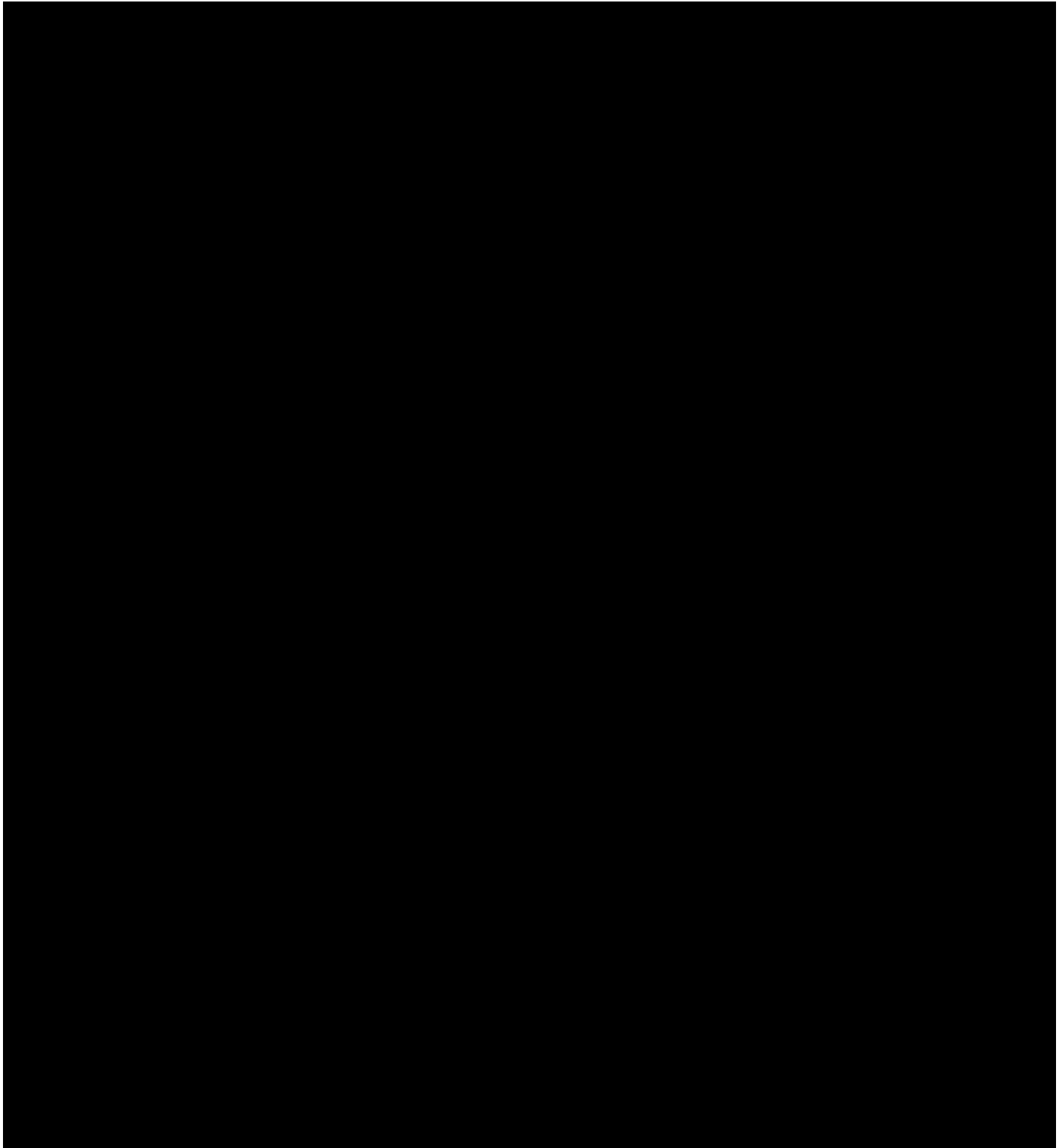


Fig 3. Time-dependent relationship between mutation load and treatment response. (A) There was a stronger association between somatic mutation load and progression-free survival (PFS) for events occurring more than 3 months following therapy (blue box: hazard ratio [HR] = 0.69, 95% CI [0.38, 0.99]), as compared to those in the first 3 months (red box: HR = 0.91, 95% CI [0.75, 1.07]). (B) There was a stronger association between somatic mutation load and overall survival (OS) for events occurring more than 3 months following therapy (blue box: HR = 0.80, 95% CI [0.60, 1.00]), as compared to those in the first 3 months (red box: HR = 1.02, 95% CI [0.79, 1.22]). (C) Kaplan-Meier estimates for PFS among patients with missense single nucleotide

variants (SNVs) per megabase above the median observed value of 1.03 (red) and those with counts below the median value (blue). Time (Days) is plotted on a log-scale. For context, the frequency of observed events (progression and/or mortality) is plotted below the x-axis among patients with missense SNV per megabase above (red) and below (blue) the median value, with the approximate schedule of follow-up scans per the study protocol (see [Methods](#)) shown as vertical dotted lines.

<https://doi.org/10.1371/journal.pmed.1002309.g003>

with cancer in whom clinical variables may negate immunological response, despite the presence of 1 or more favorable tumor-associated biomarkers. The inclusion of these clinical variables is warranted in future studies.

Examination of the tumor microenvironment shows evidence for adaptive immunity and suppression in responding tumors

Several studies have suggested that an inflammatory microenvironment, tumor, or IC PD-L1 expression increase the likelihood of response to checkpoint blockade. As seen in the published IMVigor 210 cohort, PD-L1 IC expression was significantly associated with DCB in this subset ($n = 29$, Spearman $\rho = 0.48$, $p = 0.0083$, [S4A Fig](#)). We quantified immune infiltration from RNA-seq using ESTIMATE [19]. The immune score, while associated with the TIL proportion estimated through TCR-seq ([S4B Fig](#)), was estimated to be 764.37 (range –1195.08 to 1509.65) in patients with DCB and 263.49 (range –1100.78 to 1734.28) in patients without DCB but was not significantly different ($n = 26$, Mann-Whitney $p = 0.33$, [S4C Fig](#)). When we performed GSEA using the Hallmark Geneset [20], we did not observe any differentially expressed gene sets between tumors from patients with DCB versus no DCB. Furthermore, RNA expression of *PD-L1* did not correlate with reported IC PD-L1 staining level ($n = 26$, Spearman $\rho = 0.045$, $p = 0.83$, [S4D Fig](#)). We did not observe a difference in tumor MHC class I expression according to DCB ([S4E Fig](#), *HLA-A*: $n = 26$, Mann-Whitney $p = 0.26$, *HLA-B*: $n = 26$, Mann-Whitney $p = 0.36$, *HLA-C*: $n = 26$, Mann-Whitney $p = 0.24$).

Given that such agnostic approaches did not reveal a clear association between tumor microenvironment factors and response, we pursued a hypothesis-driven approach examining the genes that show up-regulation at the cell surface during T cell exhaustion. When categorized by DCB, there was no significant difference in expression of such genes, including *CTLA-4*, *TIGIT*, *HAVCR2* (*TIM-3*), or *LAG-3* [21]. When grouped by PD-L1 staining, we found low expression of all markers in the PD-L1 low group (IC0), as expected. However, in the PD-L1 high group (IC2), *HAVCR2* exhibited significantly higher expression in tumors from patients who experienced DCB than in those who did not ([S4F Fig](#)). Interestingly, of the 4 IC2 tumors among *HAVCR2*-high patients (*HAVCR2* expression greater than the median), 2 had missense SNV loads at or below the median (2 and 57); the other 2 had 180 and 412 SNVs. Additionally, although Rosenberg and colleagues [2] found that luminal cluster II showed a significantly higher response rate among the 4 subtypes of RNA expression from The Cancer Genome Atlas (TCGA), no significant association was found here between the 4 clusters and DCB ($n = 20$, Fisher's Exact $p = 0.36$) ([S4G Fig](#)) nor between the luminal/basal subcategorization and DCB ($n = 20$, Fisher's Exact $p = 1.00$), possibly due to sample size.

Relative importance of somatic, immune, and clinical factors in resistance and response to PD-L1 blockade

Unanswered questions that arise from the many studies of biomarker correlates of checkpoint blockade response are whether measures such as mutation load, PD-L1 staining, and others reflect the same underlying biology or if each confers an independent effect on outcome?

When examined in conjunction with mutation load, the greater the expression of PD-L1, the more negative the association of mutation load with hazard (i.e., higher mutation load was

associated with longer survival). Among patients with tumors showing little-to-no expression of PD-L1 (IC0 rated), each unit increase in missense SNV count per megabase was associated with a negligible change in hazard ($n = 4$, HR = 1.43, 95% CI [0.75, 2.98]). Among patients with tumors expressing PD-L1 at moderate or high levels (IC1 or IC2 staining), missense SNV count per megabase was associated with lower risk for disease progression or mortality (among IC1: $n = 11$, HR = 0.75, 95% CI [0.47, 1.14]; among IC2: $n = 10$, HR = 0.73, 95% CI [0.48, 1.06]). Although our limited sample size precludes making an assertion that mutation load is associated with survival in any particular subgroup (e.g., when looking among IC1 and IC2 tumors alone), our data do support the presence of an interaction among these variables ($p = 0.046$ for interaction; [S5A Fig](#)). Given the plausibility of the finding that somatic mutation load may correlate better with survival among patients with an inflamed tumor microenvironment, the addition of somatic mutation load to PD-L1 IC staining warrants further study.

We found a similar, albeit weaker, interaction effect when looking at the association of somatic mutation load (missense SNV count per megabase) and PFS, according to the presence/absence of liver metastasis prior to treatment administration ($p = 0.14$ for interaction). Among patients without liver metastasis, somatic mutation load was associated with a lower risk for disease progression or mortality ($n = 16$, HR = 0.73, 95% CI [0.50, 1.02], [S5B Fig](#)) than patients with liver metastasis ($n = 9$, HR = 0.96, 95% CI [0.66, 1.37], [S5B Fig](#)).

To our surprise, although both PD-L1 staining and mutation load were each associated with response in the original study [2], these variables did not correlate with each other ([Fig 4A](#)). Furthermore, pretreatment peripheral TCR clonality did not correlate with mutation load ([Fig 4B](#)). The lack of association between these variables suggests that each might have an independent or semi-independent role in determining the likelihood of response to therapy. TCR clonality and infiltration did, however, correlate with PD-L1 IC score: those tumors with higher clonality or higher infiltration also featured higher PD-L1 staining ($p = 0.02$ and $p = 0.01$, respectively, [Fig 4C and 4D](#)).

In an analysis to see whether the association between pretreatment peripheral TCR clonality and PFS varied by PD-L1 IC score, we found some evidence of an interaction ($p = 0.015$ for interaction; [Fig 4E](#)). Among patients with low levels of PD-L1 expression, there was little association between pretreatment peripheral TCR clonality and PFS (among IC0: $n = 4$, HR [mean] = 1.86, HR [median] = 1.55, 95% CI [0.50, 4.99], $p(\text{HR} > 1) = 0.21$; among IC1: $n = 11$, HR [mean] = 0.69, HR [median] = 0.58, 95% CI [0.15, 1.84], $p(\text{HR} < 1) = 0.19$). Among patients with high levels of PD-L1 expression, by comparison, we observed almost complete separation of PFS according to pretreatment peripheral TCR clonality (among IC2: $n = 10$, HR [mean] = 89.88, HR [median] = 23.41, 95% CI [2.43, 506.94], $p(\text{HR} > 1) = 0.0014$; [Fig 4E](#)). Similar results were seen in analyses with respect to OS and in a logistic regression analysis for DCB ([S5C Fig](#), [S5D Fig](#), [S5E Fig](#)).

To resolve the hypothesis that those patients with low peripheral TCR clonality simply were healthier, we examined the association between 5-factor score and pretreatment peripheral TCR clonality and did not find such an association ($n = 26$, Spearman rho = 0.25 $p = 0.22$, [S5F Fig](#)).

In a multivariate survival model for time to disease progression or mortality, which allows the effect of each biomarker to vary according to intratumoral PD-L1 IC score, we find that the correlation of each intratumoral, peripheral, or clinical biomarker with disease progression or mortality is relatively independent of the others ([Fig 4F](#), [S5G Fig](#)). Perhaps with the notable exception of the association between liver metastatic status and time to progression or survival, the correlation of each intratumoral or peripheral biomarker with outcome is strongest in the group with the highest levels of IC PD-L1 expression ([S2 Table](#)).

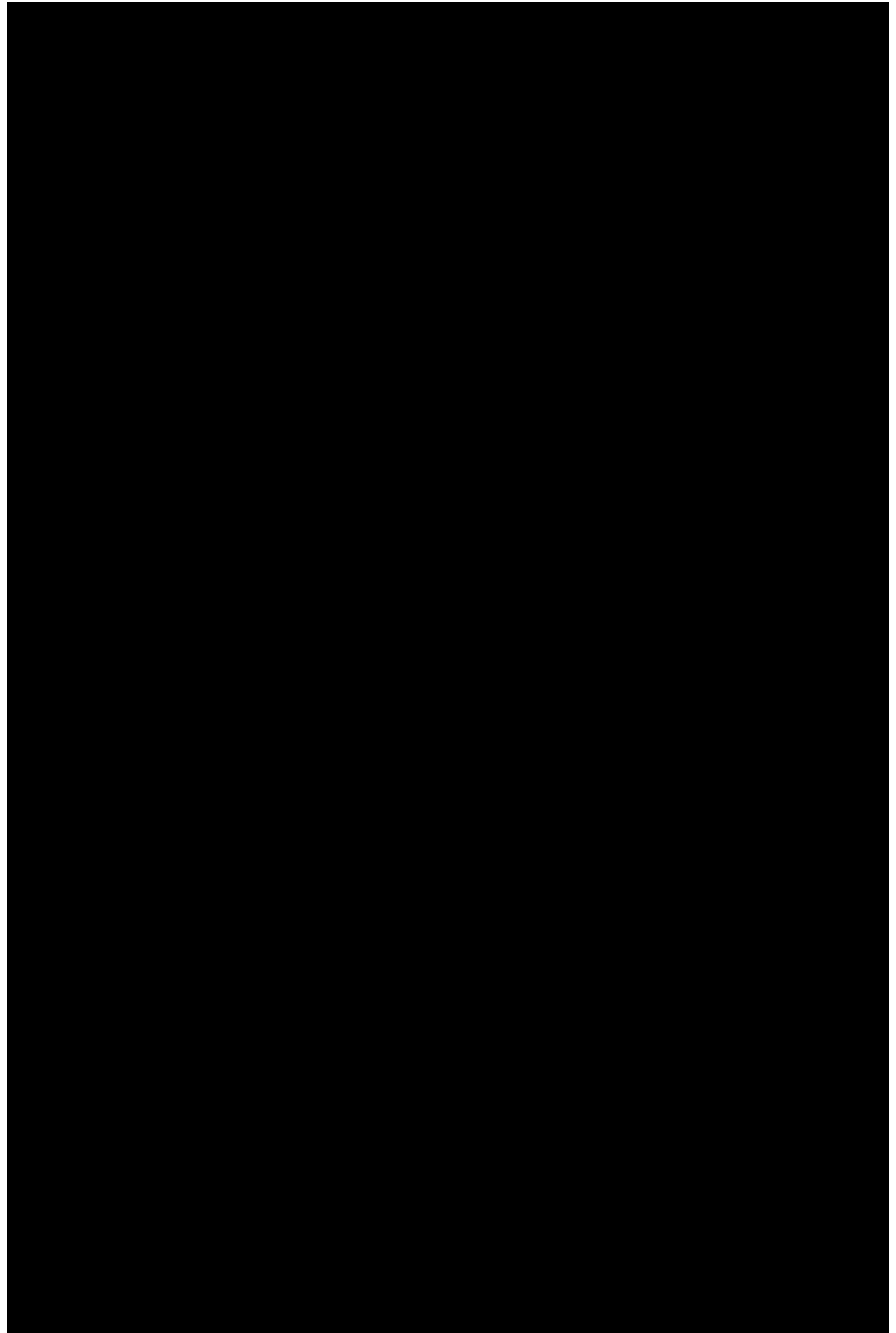


Fig 4. Associations between measured somatic, immune, and clinical variables. (A) Although both programmed death-ligand 1 (PD-L1) staining and mutation load each were weakly associated with response, these variables were not correlated with each other ($n = 25$, Spearman $\rho = 0.14$ $p = 0.51$). (B) Pretreatment peripheral T cell receptor (TCR) clonality did not correlate with mutation load ($n = 25$, Pearson $r = 0.0017$ $p = 0.99$). (C) Tumor-infiltrating T lymphocytes (TIL) proportion as estimated by TCR sequencing was

associated with PD-L1 immune cell (IC) staining ($n = 24$, Spearman $\rho = 0.51$ $p = 0.010$). (D) TIL clonality was associated with PD-L1 IC staining ($n = 24$, Spearman $\rho = 0.48$ $p = 0.017$). (E) Hazard associated with log (pretreatment peripheral TCR clonality) by level of IC PD-L1 expression (IC0, IC1, or IC2). (F) Multivariate survival analysis of various clinical, peripheral, and intratumoral biomarkers for association with time to disease progression or mortality, utilizing a varying-coefficient model, which allows the hazard associated with a 1-unit increase in a biomarker's value to vary according to the level of intratumoral PD-L1 expression (IC score).

<https://doi.org/10.1371/journal.pmed.1002309.g004>

Discussion

The treatment of previously incurable metastatic solid tumors with checkpoint blockade agents has led to dramatic success in a minority of patients, a finding that has generated substantial excitement in the field, with associated correlative studies and drug development. Here, we undertook the in-depth characterization of tumors and peripheral blood from 29 patients treated on IMvigor 210, a Phase II study in which 310 patients were treated with the anti-PD-L1 agent atezolizumab. In this cohort of patients, we illustrate the importance of host immune factors, including intratumoral and peripheral TCR clonality, infiltration, and expansion, to clinical outcomes. We did not find a significant association between mutation or expressed neoantigen load and PFS or DCB (defined as PFS > 6 months). However, we did observe a time-dependent relationship between mutation load and outcome, wherein a relationship between mutation burden and outcome could only be detected in those patients surviving greater than 3 months. This analysis implies that patients who experienced rapid progression may display systemic indicators of immune deficiency despite elevated mutation load in the tumors. Calculation of the hazard ratios for each measured biomarker and clinical factor underscores the concept that a complex interaction of both host and tumor variables determines whether a patient will experience clinical benefit from anti-PD-L1 therapy. Although the overall study found significant associations between mutation load as measured by the Foundation Medicine targeted sequencing panel and radiographic response [2], there was no statistically significant association between mutation load and DCB or survival in the patient subset studied here, despite the similarity of our study population to the parent study. This contrast may be due to a combination of factors. First, though statistically significant, the association in the overall study was not categorical: as in other studies of mutation load, this factor alone was not predictive of response. Second, we have less power to detect this association in our smaller subset compared with the larger studied cohort. Third, standardized definitions and calculations of mutation load do not exist as of yet: each published study has used differing methodologies [2,10,22,23]. Indeed, in this study, depending on the method used, the association between mutation load and clinical outcomes varied from $p < 0.08$ to $p > 0.4$ (area under the curve [AUC] values and p -values in S1 Table). To illustrate the fickle nature of defining mutation load, counting only the mutations excluded by BQSR, as opposed to only those remaining after BQSR, showed a significant association with DCB. Together, these findings underscore the need for improved and standardized mutation calling methods. The weak association of mutation load with DCB and the lack of such standardization render this biomarker unfit for application to individual patients at present. Furthermore, if validated in another dataset, this analysis implies that a clinical and immunological state may exist in patients with advanced cancer, such that patients with very rapidly progressing disease and expected death in less than 3 months do not respond despite the presence of positive intratumoral biomarkers.

In an attempt to deepen our understanding of the biology of response and resistance, we studied additional factors. We found that even in this small dataset, TCR clonality below the median in the peripheral blood prior to treatment, expansion of tumor-associated TCR in the periphery 3 weeks after initiating treatment, and higher TIL proportion are all associated with

clinical benefit. These data suggest that TCR-seq provided additional insights into response and resistance beyond mutation load and PD-L1 staining. With respect to biomarker development, our study implies that noninvasive metrics such as pretreatment peripheral TCR clonality and known prognostic features such as the presence of liver metastases may be worthy of further study in urothelial cancer patients treated with PD-L1 blockade.

From a mechanistic perspective, these findings imply an important relationship between circulating and intratumoral immunity upon PD-L1 blockade. We hypothesize that low TCR clonality in the peripheral blood prior to treatment increases the likelihood that a patient harbors 1 or several clones capable of tumor recognition, whether of neoantigens or tumor-associated antigens. The expansion of tumor-associated TCRs in the peripheral blood underscores the continuity of the tumor and blood compartments, and suggests that the activity of PD-L1 blockade may involve circulating T cells more than was previously thought. Indeed, this raises the possibility that anti-tumor T cells may home from the periphery to the tumor before later recirculating.

Finally, though limited in power by the small sample size, we attempted to integrate the importance of the studied variables. This analysis demonstrated both hypothesized and unexpected interactions. For example, while mutation load seemed to be associated with outcome more significantly in PD-L1 IC1 and IC2 tumors, high PD-L1 IC staining in the setting of high peripheral TCR clonality was associated with a substantial hazard for poor outcome. Given the significance of PD-L1 expression in mediating response to anti-PD-L1 therapy, the presence of these interactions may argue in their favor as predictive rather than prognostic biomarkers. Further analysis is required to elucidate the role of these biomarkers in mediating response to checkpoint blockade.

This study has several limitations. The patients under study were treated at a single institution and represent a small subset of the overall study, limiting statistical power. As a single-arm Phase II study, there is no control arm for comparison. Tumor samples were FFPE and were not collected immediately prior to treatment initiation. Only 1 sample per patient was utilized, which does not necessarily capture the heterogeneity of each tumor. Finally, a number of analyses were performed on a small number of patients without an independent validation cohort; although most analyses were prespecified, there is a risk of Type II error, and no adjustments were made for multiple testing.

In conclusion, we have demonstrated the potential value of pursuing an integrated study of somatic, immune, and clinical features in order to elucidate the biological mechanisms underlying response to checkpoint blockade and ultimately improve our ability to practice precision medicine. We hope this work will motivate further multi-omics studies of response to checkpoint blockade.

Supporting information

S1 Table. Choice of depth and variant allele frequency (VAF) filtering, as well as whether or not to run Base Quality Score Recalibration (BQSR), resulted in differences in predictive value for mutation load. We chose to optimize for precision, choosing the highlighted set of filters/BQSR. Precision is defined as the fraction of filtered missense mutations in Integrated Mutation Profiling for Actionable Cancer Targets (IMPACT) genes that were actual IMPACT panel variants. Recall is defined as the fraction of actual IMPACT panel variants that were found in the filtered missense results.
(DOCX)

S2 Table. Summary of results from multivariate survival analysis of various clinical, peripheral, and intratumoral factors to estimate their independent association with hazard

for disease progression or mortality (progression-free survival [PFS]) and for mortality (overall survival [OS]) according to the level of intratumoral programmed death-ligand 1 (PD-L1) expression (immune cell [IC] grade). Results are summarized as median and 95% posterior intervals.
(DOCX)

S1 Fig. (A) Twenty-five percent of patients with durable clinical benefit (DCB) had less than the median tumor-infiltrating T lymphocytes (TIL) proportion versus 63% of patients without DCB ($n = 24$, Fisher's Exact $p = 0.19$); similarly, 25% of patients with DCB had less than the median TIL clonality versus 63% of patients without DCB ($n = 24$, Fisher's Exact $p = 0.19$). (B) T cell receptor (TCR) overlap between the pretreatment and 3-week posttreatment peripheral blood in 1 patient with limited clinical benefit (progression-free survival [PFS] = 37 days) and 1 patient with DCB (complete response [CR] at 630 days after starting treatment). The association between pretreatment peripheral blood TCR sequences (x -axis) and posttreatment peripheral blood TCR sequences (y -axis) is overlaid with the presence of tumor-associated T cell clones. Gray indicates TCRs present only in the peripheral blood; blue indicates TCRs present in the tumor and blood; orange indicates TCRs present in the tumor and expanded in the blood with treatment. (C) There was no significant expansion of TIL-associated TCR clones between pretreatment (3.00 [range 1.00; 9.00]) and 6 weeks posttreatment (2.00 [range 1.00; 8.00]) ($n = 20$, Mann-Whitney $p = 0.17$). (D) The combination of high pretreatment TIL and low pretreatment peripheral blood TCR clonality were predictive of DCB ($n = 24$, Fisher's Exact $p = 0.0069$) and overall survival greater than 12 months (DCB-OS) ($n = 24$, Fisher's Exact $p = 0.014$). For DCB, a logit model combining both was more predictive than peripheral blood ($n = 24$, log-likelihood $p = 0.00029$) or TIL ($n = 24$, log-likelihood $p = 0.00051$) clonality alone. For DCB-OS, both combined were more predictive than TIL ($n = 24$, log-likelihood $p = 0.0029$) clonality alone.
(DOCX)

S2 Fig. (A) No significant association between the number of missense single nucleotide variants (SNVs) per megabase and overall survival greater than 12 months (DCB-OS), with 2.13 (range 0.038; 11.46) tumors from those patients who survived greater than 12 months versus 0.48 (range 0.019; 9.90) those who did not ($n = 25$, Mann-Whitney $p = 0.37$). (B) No significant difference was found between median expressed neoantigens in tumors from patients who survived greater than or equal to 12 months: 1.31 (range 0.00; 6.06) versus 0.35 (range 0.00; 5.30) those who survived less than 12 months ($n = 25$, Mann-Whitney $p = 0.36$). (C) No significant difference was found between median predicted neoantigens per megabase: 4.58 (range 0.037; 39.48) tumors from patients with durable clinical benefit (DCB) as compared to 1.35 (range 0.00; 20.22) those who progressed in less than 6 months (no DCB) ($n = 25$, Mann-Whitney $p = 0.55$). (D) No significant difference was found between median predicted neoantigens per megabase in tumors from those patients who survived greater than 12 months (3.56 [range 0.037; 39.48]) compared to 1.37 (range 0.00; 20.22) those who did not (no DCB-OS) ($n = 25$, Mann-Whitney $p = 0.81$).
(DOCX)

S3 Fig. (A) Patients who survived less than 3 months (red box) exhibited a significantly higher 5-factor score (3.00 [range 2.00; 4.00]) compared to 1.50 [range 0.00; 4.00] patients who survived >3 months; blue box; $n = 26$, Mann-Whitney $p = 0.018$). (B) Patients who survived less than or equal to 3 months (red box) were more likely to have liver metastases (100% in patients who survived less than or equal to 3 months and 22% in patients who survived longer than 3 months, $n = 29$, Fisher's Exact $p = 0.00097$). (C) The hazard ratio for each mutation per

megabase, estimated at each unique failure time. Red box plots summarize 50% and 95% posterior intervals for each observed failure/censor time, with median values shown in green. Time (Days) is plotted on a log-scale. Estimates are not independent from one another since the model utilizes a random-walk parameterization to allow the variance in hazard over time to be modeled flexibly. (D) Posterior predicted intervals for progression-free survival (PFS), which are drawn from the survival model estimating the time-varying effect of mutation count on PFS. Intervals are shown for patients with missense single nucleotide variants (SNVs) per megabase above the median and those with counts below the median value (blue), for illustrative purposes. This cutpoint was not used in the model; missense SNV per megabase was included as a continuous covariate. Lines are drawn at median values of the posterior predictive distribution, with 50% credible intervals shown in the shaded regions. Time (Days) is plotted on a log-scale. (DOCX)

S4 Fig. (A) Programmed death-ligand 1 (PD-L1) immune cell (IC) staining as reported by the sponsor in the published study [2], and outcomes in our cohort were significantly associated in this substudy ($n = 29$, Spearman rho = 0.48 $p = 0.0083$). (B) ImmuneScore was associated with tumor-infiltrating T lymphocytes (TIL) proportion ($n = 24$, Spearman rho = 0.47 $p = 0.022$). (C) There was no association between ImmuneScore and durable clinical benefit (DCB) (DCB: 764.37 [range -1195.08 to 1509.65]; no DCB: 263.49 [range -1100.78 to 1734.28]) ($n = 26$, Mann-Whitney $p = 0.33$). (D) PD-L1 expression as measured by RNA sequencing was not associated with PD-L1 IC level ($n = 26$, Spearman rho = 0.045 $p = 0.83$). Tumor cell PD-L1 staining was not available. (E) Human leukocyte antigen (HLA) Class I expression was not associated with DCB (HLA-A: $n = 26$, Mann-Whitney $p = 0.26$, HLA-B: $n = 26$, Mann-Whitney $p = 0.36$, HLA-C: $n = 26$, Mann-Whitney $p = 0.24$). (F) Expression of other inhibitory markers, in particular *HAVCR2* (also known as *TIM-3*), was higher in DCB patients in the IC2 group. (G) No association was found between TCGA RNA Subtype and response in this substudy ($n = 20$, Fisher's Exact $p = 0.36$). (DOCX)

S5 Fig. (A) The hazard associated with log(missense single nucleotide variant [SNV] count per megabase) by level of immune cell (IC) (IC0, IC1, or IC2) programmed death-ligand 1 (PD-L1) expression. (B) The hazard associated with log(missense SNV count per megabase) by presence or absence of liver metastasis at enrollment. (C) The association of peripheral T cell receptor (TCR) clonality prior to treatment with time to mortality (overall survival [OS]) varies according to IC (IC0, IC1, or IC2) PD-L1 expression. (D) The association of peripheral TCR clonality prior to treatment with durable clinical benefit (DCB) varies according to IC (IC0, IC1, or IC2) PD-L1 expression. (E) The association of peripheral TCR clonality prior to treatment with DCB (OS) varies according to IC (IC0, IC1, or IC2) PD-L1 expression. (F) There was no significant relationship between 5-Factor score and pretreatment TCR clonality ($n = 26$, Spearman rho = 0.25 $p = 0.22$). (G) Multivariate survival analysis of various clinical, peripheral, and intratumoral biomarkers for association with time to mortality (OS), utilizing a varying-coefficient model, which allows the hazard associated with a 1-unit increase in a biomarker's value to vary according to the level of intratumoral PD-L1 expression (IC score). Note that the x -axis has been truncated at a value of 10 for clarity, even though this results in the exclusion of some estimated hazard ratio (HR) values (specifically that for pretreatment peripheral TCR clonality among IC2 patients). (DOCX)

S6 Fig. (A) No significant association between the number of missense single nucleotide variants (SNVs) found on MSK-IMPACT and durable clinical benefit (DCB) (DCB 0.13 (range

0.00 [range 0.00; 0.31] versus no DCB 0.046 [range 0.00; 0.37] ($n = 25$, Mann-Whitney $p = 0.42$). (B) No significant association between the number of missense SNV found on MSK-IMPACT and overall survival (OS) (survival greater than 12 months 0.093 [range 0.00; 0.31] versus less than 12 months 0.074 [range 0.00; 0.37] ($n = 25$, Mann-Whitney $p = 0.78$). (C) There was no significant difference in apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APO-BEC) signature found in tumors from patients with progression-free survival (PFS) DCB (0.19 [range 0.00; 0.56]) compared to no DCB (0.00 [range 0.00; 0.46]) ($n = 25$, Mann-Whitney $p = 0.23$). (D) There was a significant correlation between missense SNV count and APOBEC signature mutations ($n = 25$, Pearson $r = 0.40$ $p = 0.048$). (E) There was no significant association between FGFR3 mutations or expression ($n = 26$, Mann-Whitney $p = 0.39$). (F) There was no significant association between MYC expression and outcome measured by PFS DCB ($n = 26$, Mann-Whitney $p = 0.87$). (G) There was no significant association between DDR mutations rated as ≥ 2 possibly pathogenic and DCB ($n = 25$, Mann-Whitney $p = 0.20$). (H) Univariate association of exonic SNV, missense SNV, and neoepitope load with DCB, with (blue bars) and without (red bars) filtering by expression. (I) Univariate association of exonic SNV, missense SNV, and neoepitope load with OS greater than 12 months, with (blue bars) and without (red bars) filtering by expression. (J) Univariate association of exonic SNV, missense SNV, and neoepitope load with PFS, showing results with (blue bars) and without (red bars) filtering by expression. (K) Univariate association of exonic SNV, missense SNV, and neoepitope load with OS, showing results with (blue bars) and without (red bars) filtering by expression. (L) Univariate association of expressed/total ratio for exonic SNV, missense SNV, and neoantigen loads with PFS. (M) Univariate association of expressed/total ratio for exonic SNV, missense SNV, and neoantigen loads with OS. (N) Univariate association of expressed/total ratio for exonic SNV, missense SNV, and neoantigen loads with DCB (PFS greater than 12 months). (O) Univariate association of expressed/total ratio for exonic SNV, missense SNV, and neoantigen loads with OS greater than 12 months. (P) There was no significant association between pack years of reported smoking history and DCB ($n = 29$, Mann-Whitney $p = 0.87$). (Q) There were significant associations between PFS ($n = 29$, log-rank $p = 0.024$) and OS ($n = 29$, log-rank $p = 0.018$) and the presence of liver metastasis. There was a significant association between the presence of visceral metastases and poor OS ($n = 29$, log-rank $p = 0.020$). (S) There was not a significant association between 5-factor score and OS ($n = 26$, log-rank $p = 0.13$). (T) No significant association between the number of expressed missense SNV per megabase and DCB (DCB 0.79 [range 0.00; 3.34] versus no DCB: 0.16 [range 0.00; 3.34]) ($n = 25$, Mann-Whitney $p = 0.26$). (DOCX)

S1 Text. STROBE criteria.

(DOCX)

S2 Text. Prespecified analysis plan.

(DOCX)

S3 Text. Supplementary analyses.

(DOCX)

Acknowledgments

We acknowledge Carol Aghajanian and Jedd Wolchok for their mentorship of AS.

Author Contributions

Conceptualization: AS SAF JH JER DFB.

Data curation: MB MM TN JBN EC BAA.

Formal analysis: TN AA JBN EC BAA EY MV.

Funding acquisition: AS JER DFB.

Investigation: HAA SB MB MM GI TM JER DFB.

Methodology: AS TN AA JBN MDH EC BAA EY MV SB HSR ERM JH.

Project administration: AS JH JER DFB.

Resources: AS JH HSR JER DFB.

Software: TN AA JBN EC BAA HSR.

Supervision: JER DFB.

Validation: TN JBN.

Visualization: AS TN AA JBN EY.

Writing original draft: AS TN SAF AA JBN.

Writing review & editing: AS TN SAF AA JBN MDH EC BAA HAA EY MV SB MB MM GI HSR ERM JH JER DFB TM.

References

1. Powles T, Eder JP, Fine GD, Braithen FS, Loriot Y, Cruz C, et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature*. 2014; 515: 558–562. <https://doi.org/10.1038/nature13904> PMID: 25428503
2. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet*. 2016; 387: 1909–1919. [https://doi.org/10.1016/S0140-6736\(16\)00561-4](https://doi.org/10.1016/S0140-6736(16)00561-4) PMID: 26952546
3. Postow MA, Yuan J, Kitano S, Lesokhin AM, Wolchok JD. Markers for anti-cytotoxic T-lymphocyte antigen 4 (CTLA-4) therapy in melanoma. *Methods Mol Biol*. 2014; 1102: 83–95. https://doi.org/10.1007/978-1-62703-727-3_6 PMID: 24258975
4. Robins HS, Campregher PV, Srivastava SK, Wachter A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009; 114: 4099–4107. <https://doi.org/10.1182/blood-2009-04-217604> PMID: 19706884
5. Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IMGT—the international ImMunoGeneTics information system—25 years on. *Nucleic Acids Res*. 2015; 43: D413–D422. <https://doi.org/10.1093/nar/gku1056> PMID: 25378316
6. Wu D, Emerson RO, Sherwood A, Loh ML, Angiolillo A, Howie B, et al. Detection of minimal residual disease in B lymphoblastic leukemia by high-throughput sequencing of IGH. *Clin Cancer Res*. 2014; 20: 4540–4548. <https://doi.org/10.1158/1078-0432.CCR-13-3231> PMID: 24970842
7. Hsu MS, Sedighim S, Wang T, Antonios JP, Everson RG, Tucker AM, et al. TCR Sequencing Can Identify and Track Glioma-Infiltrating T Cells after DC Vaccination. *Cancer Immunol Res*. 2016; 4: 412–418. <https://doi.org/10.1158/2326-6066.CIR-15-0240> PMID: 26968205
8. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*. 2014; 515: 568–571. <https://doi.org/10.1038/nature13954> PMID: 25428505
9. DeWitt WS, Emerson RO, Lindau P, Vignali M, Snyder TM, Desmarais C, et al. Dynamics of the cytotoxic T cell response to a model of acute viral infection. *J Virol*. 2015; 89: 4517–4526. <https://doi.org/10.1128/JVI.03474-14> PMID: 25653453
10. Rizvi N., Hellmann M, Snyder, Et al A. Mutational Landscape Determines Sensitivity to Programmed Cell Death-1 Blockade in Non-Small Cell Lung Cancer. *Science*. 2015;

11. Kauffmann A, Rosselli F, Lazar V, Winnepeninckx V, Mansuet-Lupo A, Dessen P, et al. High expression of DNA repair pathways is associated with metastasis in melanoma patients. *Oncogene*. 2008; 27: 5651–5731. <https://doi.org/10.1038/sj.onc.1210700> PMID: 17891185
12. Tuteja N, Singh MB, Misra MK, Bhalla PL, Tuteja R. Molecular mechanisms of DNA damage and repair: progress in plants. *Crit Rev Biochem Mol Biol*. 2001; 36: 337–397. <https://doi.org/10.1080/20014091074219> PMID: 11563486
13. Sonpavde G, Pond GR, Rosenberg JE, Bajorin DF, Choueiri TK, Necchi A, et al. Improved 5-Factor Prognostic Classification of Patients Receiving Salvage Systemic Therapy for Advanced Urothelial Carcinoma. *J Urol*. 2016; 195: 2771–2811. <https://doi.org/10.1016/j.juro.2015.07.111> PMID: 26292040
14. Bajorin DF, Dodd PM, Mazumdar M, Fazzari M, McCaffrey JA, Scher HI, et al. Long-term survival in metastatic transitional-cell carcinoma and prognostic factors predicting outcome of therapy. *J Clin Oncol*. 1999; 17: 3173–3181. <https://doi.org/10.1200/JCO.1999.17.10.3173> PMID: 10506615
15. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoeediting. *Immunity*. 2004; 21: 137–148. <https://doi.org/10.1016/j.immuni.2004.07.017> PMID: 15308095
16. Cha E, Klinger M, Hou Y, Cummings C, Ribas A, Faham M, et al. Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci Transl Med*. 2014; 6: 238ra70. <https://doi.org/10.1126/scitranslmed.3008211> PMID: 24871131
17. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat Commun*. 2015; 6: 10001. <https://doi.org/10.1038/ncomms10001> PMID: 26647970
18. Qiu P, Pang L, Arreaza G, Maguire M, Chang KC, Marton MJ, et al. Data Interoperability of Whole Exome Sequencing (WES) Based Mutational Burden Estimates from Different Laboratories. *Int J Mol Sci*. 2016; 17.
19. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013; 39: 782–795. <https://doi.org/10.1016/j.immuni.2013.10.003> PMID: 24138885
20. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011; 27: 1739–1740. <https://doi.org/10.1093/bioinformatics/btr260> PMID: 21546393
21. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. 2016; 352: 189–196. <https://doi.org/10.1126/science.aad0501> PMID: 27124452
22. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science*. 2015; 350: 207–211. <https://doi.org/10.1126/science.aad0095> PMID: 26359337
23. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med*. 2014; 371: 2189–2199. <https://doi.org/10.1056/NEJMoa1406498> PMID: 25409260