Supplementary Information

A bilateral tumor model identifies transcriptional programs associated with patient response to immune checkpoint blockade

Ivy X. Chen[a], Kathleen Newcomer[b], Kristen E. Pauken[c,d], Vikram R. Juneja[c,d], Kamila Naxerova[e], Michelle W. Wu[a], Matthias Pinter[a,2], Debattama R. Sen[c,f], Meromit Singer[b,c,g,3,1], Arlene H. Sharpe[c,d,g,1], and Rakesh K. Jain[a,1]

Affiliations:

[a] *Edwin L. Steele Laboratories, Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114.*

[b] *Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, 02215.*

[c] *Department of Immunology, Blavatnik Institute, Harvard Medical School, Boston, MA, 02115.*

[d] *Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, 02115.*

[e] *Center for Systems Biology, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114.*

[f] *Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, 02215.*

[g] *Broad Institute of MIT and Harvard, Cambridge, MA, 02139.*

[1] *Correspondence to: jain@steele.mgh.harvard.edu (RKJ), arlene_sharpe@hms.harvard.edu (AHS) and msinger@ds.dfci.harvard.edu (MS)*

[2] *Present address: Division of Gastroenterology & Hepatology, Department of Internal Medicine III, Medical University of Vienna, Vienna, Austria*

[3] *Present address: Guardant Health*

MATERIALS AND METHODS

Cell lines and cell culture: E0771 (TNBC) or E0771-GFP cells were cultured in RPMI-1640 medium + 10% FBS (Gibco); B16F10 (melanoma) cells were cultured in DMEM medium +10% FBS (Gibco, supplemented with 4.5 g/L glucose). All cells were cultured in a 37°C humidified incubator, with 5% CO_2.

Mouse tumor models: All animal experiments were performed using 6-8 weeks old female C57BL/6 mice. *Cxcr3*$^{-/-}$ mice were obtained from Jackson Laboratory and were in the C57BL/6 background. Orthotopic E0771 breast tumors derived from C57BL/6 strain were generated by implanting 200,000 cells into the third mammary fat pad of a mouse. For bilateral tumor models, 200,000 cells were implanted at the same time into both third mammary fat pads of a mouse. For histology imaging, 200,000 E0771-GFP tumor cells were implanted into the third mammary pat pad of a mouse and analyzed. The melanoma model was generated by subcutaneously implanting 200,000 B16-F10 tumor cells into left and right flanks of a mouse. Tumor sizes were measured with a caliper. All animal procedures were carried out following the Public Health Service Policy on Humane Care of Laboratory Animals and approved by the Institutional Animal Care and Use Committee of Massachusetts General Hospital.

Histology: Mice bearing orthotopic E0771 tumors were split into time- and size-matched (~75 mm^3) treatment groups. Tumor sizes were measured with a caliper. The mice were then treated with an anti-PD-1 antibody (200 μg, RMP 1-14, BioXcell) or a control IgG antibody (200 μg, 2A3, BioXcell) on day 8, 11, and 14. Tumors were isolated at different timepoints during the treatment period (days following treatment initiation): day 0 (pre-treatment), 2, 5, and 8. The tumors were removed, fixed in 4% formaldehyde in PBS (30 min/mm diameter of tissue), incubated in 30% sucrose in PBS overnight at 4 °C, and frozen in optimal cutting temperature compound (OCT) (Tissue-Tek). Frozen sections (20 μm thick) were blocked with 5% normal donkey serum (NDS) and immunostained with primary antibodies. CD8 was stained (BioLegend, 53.6-7, 1:100 dilution) and slides were counterstained with DAPI (Vector Labs).

Histological image analysis: Stained sections were imaged with a confocal microscope (Olympus) and quantified by selecting eight random fields (four interior, four periphery) at 20x oil magnification. Numbers of CD8+ T cells were analyzed based on the positive DAPI staining through a custom MATLAB script using built-in image processing functions. To obtain mosaic images, a slide scanner was used (Zeiss AxioScan Z1). Identical analysis settings and image thresholds were used for all tumors, and thresholds were determined using negative control stains utilizing only a secondary antibody.

Flow cytometry analysis: Mice bearing orthotopic E0771 breast tumors were split into treatment groups, time- and size-matched for treatment (75 mm^3 in E0771). The mice were then treated with an anti-PD-1 antibody (200 µg, RMP 1-14, BioXcell) or a control IgG (200 µg, 2A3, BioXcell) on day 8, 11, and 14. Tumors were isolated at different timepoints during the treatment period (days following treatment initiation): day 0 (pre-treatment), 2, 5, and 8. Prior to tissue collection, mice were anesthetized and cardiac perfused with 10 mL of PBS to remove excess red blood cells. Breast tumor tissues were then harvested, minced, digested, and incubated at 37 °C for 1 hour with DMEM media containing collagenase type 1A (1.5 mg/mL), hyaluronidase (1.5 mg/mL), and DNase (2 mg/mL). The digestion mixtures were filtered through 70-µm cell strainers. Single-cell suspensions were incubated with rat anti-mouse CD16/CD32 mAb for 10 minutes at 4 °C, and then stained with fluorochrome conjugated antibodies. The stained cell suspensions were washed and resuspended in cold buffer (1% BSA, 0.1% sodium azide in PBS). Fixable dye (eFluro780, eBioscience) was used for gating of viable cells. Flow cytometry data was obtained using an LSRII flow cytometer (Becton Dickinson) and analyzed using FACSDiva software. The double/aggregated events were gated out using forward scatter area (FSC-A) vs. forward scatter width (FSC-W) and side scatter area (SSC-A) vs. side scatter width (SSC-W). Various combinations of the following monoclonal anti-mouse antibodies were used: CD4 (RM4-5, BioLegend), CD45 (30-F11, BioLegend), CD11b (M1/70, BioLegend), CD3 (17A2, BioLegend), CD25 (PC61, BioLegend), CD8a (53-6.7, BioLegend), FoxP3 (FJK-16s, eBioscience), CD279 (29F.1A12, BioLegend), CD274 (10F.9G2, BioLegend), CD152 (UC10-4B9, BioLegend), CXCR3 (CXCR3-173, BioLegend), CD19 (6D5, BioLegend), granzyme B (GZ11, BioLegend), Tnf-α (MP6-XT22, eBioscience), IFN-γ (XMG1.2, BioLegend), IL-2 (JES6-5H4, BioLegend), Ki67 (B56, BD Biosciences), ICOS (15F9, BioLegend).

Statistical analysis: Statistical analyses were performed using GraphPad Prism software. The data are presented as means with standard errors (s.e.m.). Groups were compared using one-way ANOVA and unpaired Student's t-test as indicated in each experiment. Tumor growth curves were compared using two-way ANOVA with Sidak's multiple comparisons test. In pairwise comparisons within studies where multiple comparisons were made, *P* values were adjusted using Holm-Bonferroni correction.

RNA-sequencing: Mice bearing orthotopic E0771 tumors were then treated with an anti-PD-1 antibody (200 µg, RMP 1-14, BioXcell) or a control IgG (200 µg, 2A3, BioXcell) on day 8, 11, and 14. At day 15, the tumors were then excised and single cell suspensions were made. The suspensions were diluted in PBS buffer with 2% BSA and 1 mM EDTA and the CD45+CD3+CD8+ T cell population was sorted on FACS Aria 2. RNA was extracted using Myone Silane Dynabeads (Thermo Fisher Scientific). The RNA fragments were bar-coded using 8-bp barcodes in conjunction with

standard Illumina adaptors. Agencourt AMPure XP bead cleanup (Beckham Coulter/Agencourt) and 14 PCR cycles were used to amplify the samples. Sequencing was carried out on a HiSeq 2000 (Illumina). Using the aligner Salmon (http://salmon.readthedocs.io/en/latest/salmon.html) under default filtering settings, sequencing reads (fastqs) were aligned to, and count estimates calculated for, GenCode-annotated mouse (mm12, vM9) transcripts.

Differential expression assessment and ranking: The R packages 'DESeq2' and 'sva' were used in conjunction to account for batch effects in the population RNA sequencing counts and to call differentially expressed genes. The function 'DESeq' was run with the design '~ Batch + Response'. Normalized counts from the resulting 'DESeqDataSet' object were then used in sva, subsetting to only include genes with mean expression greater than one. 'svaseq' recommended using four of the SV vectors produced by 'svaseq' to correct for batch effects. The design formula for the 'DESeqDataSet' function was set to include the correction vectors provided by 'sva' in order to correct for batches across experiments. 'DESeq' was then run to identify differentially expressed genes with alpha set to 0.05. Genes that had an adjusted p-value below 0.05, and a fold change greater than 1.5 were considered differentially expressed. A mouse-derived early-therapeutic-stage responder signature was defined as the 85 genes that were upregulated in the responders. A mouse-derived early-therapeutic-stage non-responder signature was defined as the 69 genes that were upregulated in the non-responders.

For visualizations we used the quantile normalized TPM values, log2(TPM+1) adjusted, and batch corrected by regressing out the batch-associated vectors output by 'sva' (https://support.bioconductor.org/p/87508/). The corrected TPM was used to generate the principle component analysis with the 'PCA' function in the 'FactoMineR' R package. The corrected TPM values for the 154 DE genes were centered and scaled by gene, and 'aheatmap' from the R package 'NMF' was used to create the heatmap of differentially expressed genes.

TIDE analysis: The batch-corrected TPM matrix was centered and the gene names were translated to Entrez. The data was then uploaded onto the TIDE *(1)* website and run with a cancer type of 'other' and 'yes' for previous immunotherapy. The results were then plotted using 'ggplot2'.

Enrichment analysis tests: All genes were ranked by their p-value and by their fold change. The two ranking values were then aggregated to create a single ranking by taking the mean of the p-value and fold-change rankings *(2)*. The ranked list was then used in the preranked analysis of GSEA to search for significant associations with gene signatures. Default settings were used with the exceptions: permutations was set to 100, the enrichment statistic set to 'classic', and the max size set to 2500 (see Table S4 for parameters). The signature sets used were all GO terms, Kegg and Reactome pathways, and immune signatures from MSigDB, respectively groups 'c5', 'c2', and

'c7'. Gene signatures derived from the literature were also analyzed as cited in the figures and manuscript.

Identifying human patient cohorts: To identify patient subsets from the METABRIC database *(3)* that were high or low for expression of a given gene signature we computed for each patient a weighted signature score. Given the gene-by-patient RNA matrix R (including I genes and J patients) and a gene signature $G \subset \{1, \dots, I\}$, a 'supergene score' S for each patient $j \in \{1, \dots, J\}$ is computed as:

$$S_j = \sum_{i \in G} R_{ij}$$

The Pearson correlation C between each gene and the supergene is computed. A weighted score P for each patient is computed as:

$$P_j = \sum_{i \in G} R_{ij} \cdot C_i$$

The mean and standard deviation of the patient scores can then be calculated, and used to identify patients that have high or low expression of G.

Gene signatures used in survival plots: Patients from the METABRIC dataset *(3)* that had RNA-seq data were downloaded (1904 patients). RNA-seq data was centered and scaled by gene. Responder and non-responder mouse-derived gene signatures were used to generate human-matched signatures by transferring gene names from the mouse to human annotation. Genes that did not have a known human equivalent were removed.

To obtain gene signatures for responders and non-responders that have been adjusted to the human landscape, each gene's correlation to the patients' human-matched mouse-derived signature was computed and ranked, and the top ranking genes were used to construct human-adjusted gene signatures. This was done selecting either the top 10 genes or the genes 3 standard deviations above the mean to establish comprehensive human-adjusted signatures for responders and non-responders.

Each of the signatures generated was used separately to compute patient scores as decribed above in section "Identifying human patient cohorts". Patients with more than 1.5 standard deviations above or -1.5 standard deviations below the mean patient score were considered as cohorts in the survival analysis. These cohorts were then plotted comparatively using survival data from METABRIC.

References

1. P. Jiang, *et al.*, Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).

2. A.-L. Boulesteix, M. Slawski, Stability and aggregation of ranked gene lists. *Brief. Bioinformatics* **10**, 556–568 (2009).

3. C. Curtis, *et al.*, The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).

Figure S1

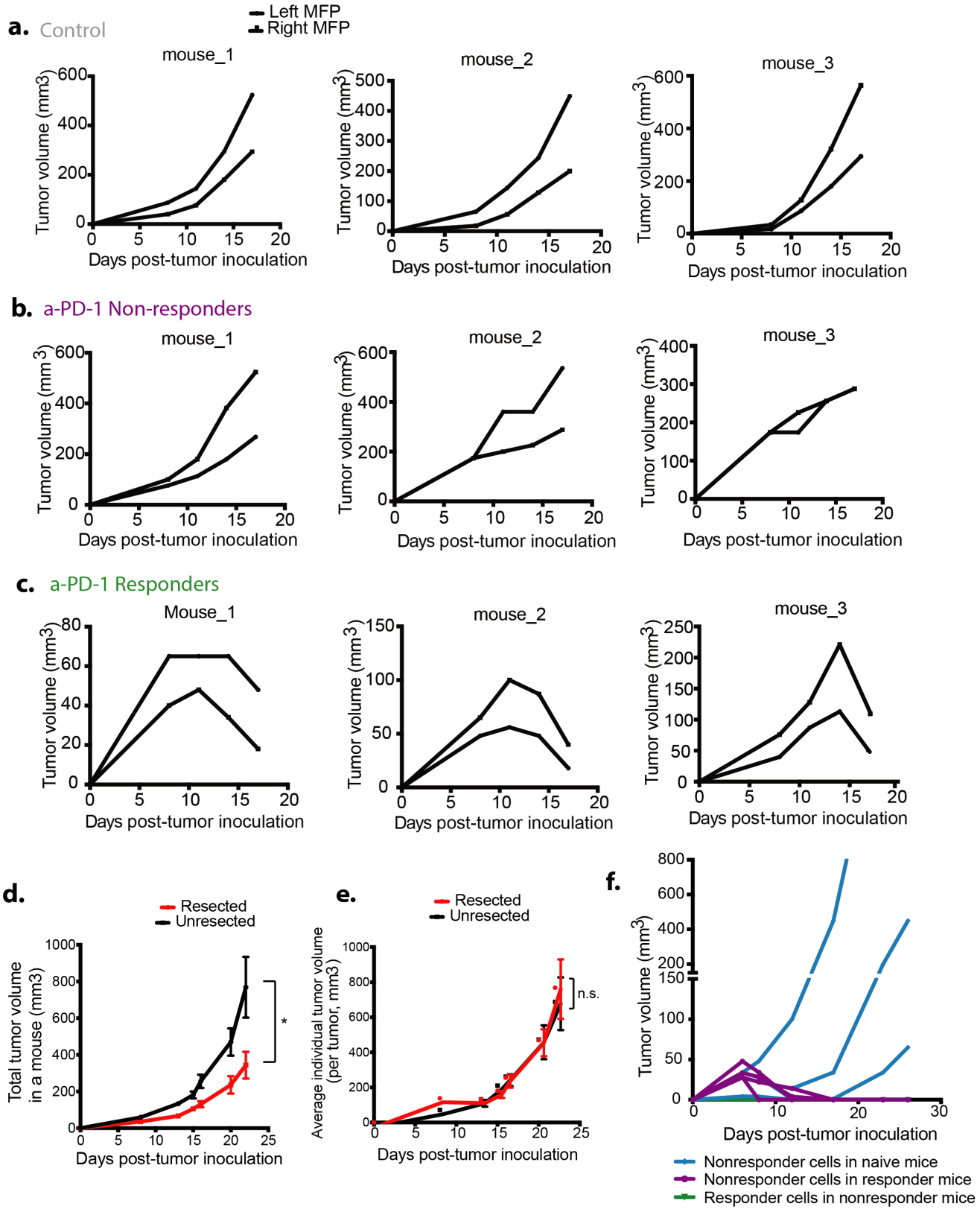
a. Control
Left MFP
Right MFP
mouse_1
mouse_2
mouse_3
Tumor volume (mm3)
Days post-tumor inoculation
b. a-PD-1 Non-responders
mouse_1
mouse_2
mouse_3
Tumor volume (mm3)
Days post-tumor inoculation
c. a-PD-1 Responders
Mouse_1
mouse_2
mouse_3
Tumor volume (mm3)
Days post-tumor inoculation
d.
Resected
Unresected
Total tumor volume in a mouse (mm3)
Days post-tumor inoculation
*
e.
Resected
Unresected
Average individual tumor volume (per tumor, mm3)
Days post-tumor inoculation
n.s.
f.
Tumor volume (mm3)
Days post-tumor inoculation
Nonresponder cells in naive mice
Nonresponder cells in responder mice
Responder cells in nonresponder mice

Fig. S1: Bilateral tumors match in their response to ICB in a breast cancer model. The same numbers of E0771 breast tumor cells are implanted orthotopically into the left and right mammary fat pad (MFP) of each mouse.

(a-b) Mice were time- and size-matched (~75 mm^3) 8 days post-tumor inoculation and were treated with anti-PD-1 mAb or IgG (control) on day 8, 11, 14. N = 10. (A) Representative tumor growth curves for mice receiving IgG control. (b-c) Representative tumor growth curves for mice receiving anti-PD-1. Both tumors in each mouse display the same growth trend of either (b) progression on ICB or (c) response and tumor regression.

(d) The average total tumor burden observed for mice with two tumors (unresected) is higher than for mice with one tumor (resected). N = 3-4. $P < 0.005$ by two-way ANOVA with Sidak's multiple comparisons test; Error bars indicate s.e.m.

(e) The average individual tumor volume is not significantly (n.s) different between mice with resected versus unresected tumors. N = 3-4. By two-way ANOVA with Sidak's multiple comparisons test; Error bars indicate s.e.m.

(f) Tumor growth of tumor slurry extracted from responding or non-responding tumors. Nonresponding cells were implanted into a treatment-naïve mouse or a responder mouse. Responding cells were implanted into a non-responder mouse. N=3-4.

Figure S2

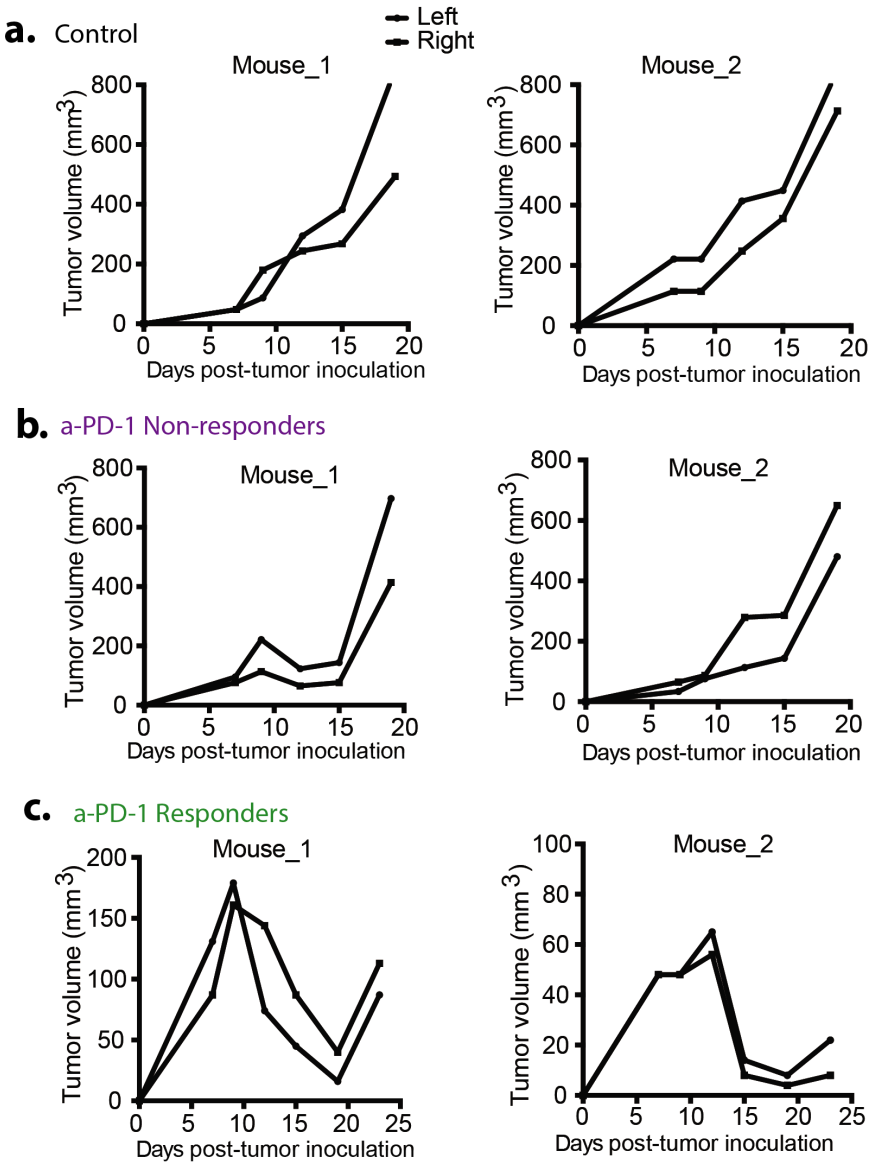

Fig S2. Bilateral tumors match in their response to ICB in a melanoma model. The same number of B16-F10 melanoma tumor cells were implanted subcutaneously into the left and right flank of each mouse. Mice were time- and size-matched 8 days post-tumor inoculation and were treated with anti-PD-1 mAb or IgG (control) for 3 doses every three days following treatment initiation. N = 10.

(a) Representative tumor growth curves for mice receiving IgG control.

(b-c) Representative tumor growth curves for mice receiving anti-PD-1 in (b) non-responders and (c) responders.

Figure S3

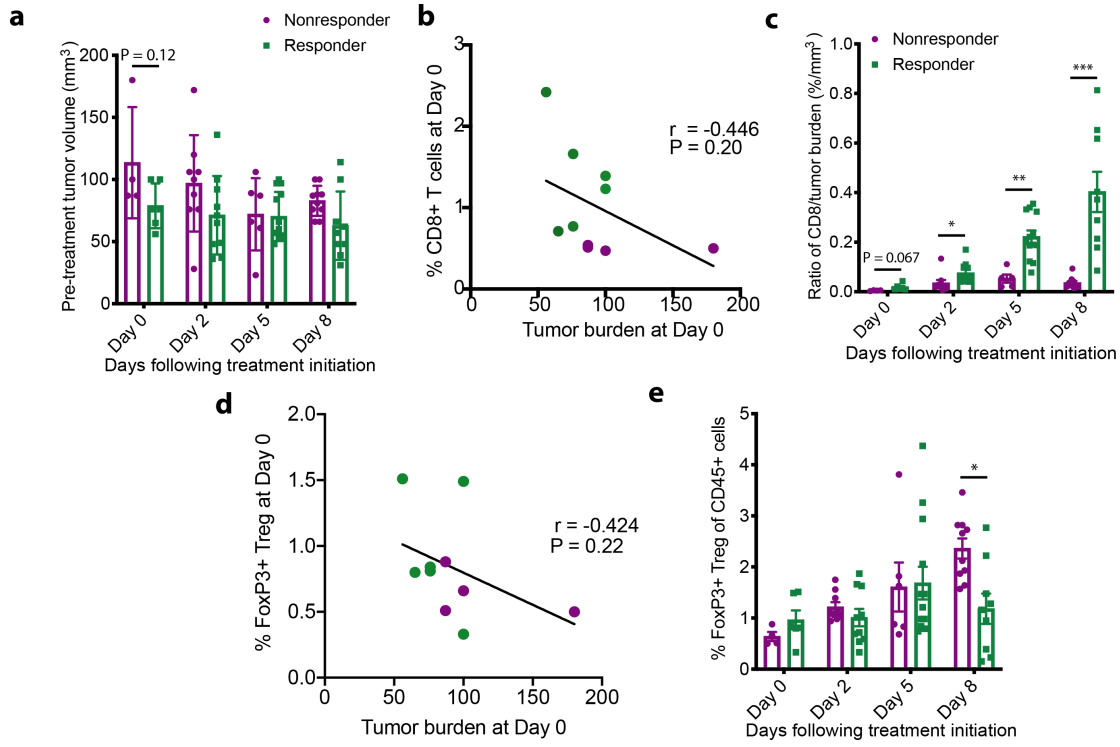

Fig S3. Tumor burden and Treg infiltration in tumors of responders and non-responders.

Mice were treated with α-PD-1 mAb or IgG (control) on day 8, 11, and 14 after tumor inoculation with E0771 BC cells. Tumors were collected throughout the treatment period at different time points (days following treatment initiation): day 0 (no-treatment, N=4-6), 2 (after one dose, N= 9-11), 5 (after two doses, N=6-10), and 8 (after three doses, N=9-10) to evaluate treatment progression.

(a) Baseline tumor volume (pre-treatment at day 0 before treatment initiation) of each cohort of mice collected at different time points. No statistical significance between the tumor burdens was observed. By Student's *t*-test; Error bars indicate s.e.m.

(b) Pearson correlation analysis between tumor burden and intratumoral CD8+ T cells at day 0 (r = -0.446, P = 0.20).

(c) Ratio of CD8+ T cells to tumor burden. * P < 0.05, ** P < 0.005, *** P < 0.0005, by Student's *t*-test. Error bars indicate s.e.m.

(d) Pearson correlation analysis between tumor burden and intratumoral Foxp3+ Treg cells at day 0 (r = -0.424, P = 0.22).

(e) CD45+CD4+Foxp3+ Treg fraction in the tumors. No differences in Foxp3+ cell population during earlier treatment time points, but there was a significant decrease in FoxP3+ cell population in the responding tumors after 3 doses of treatment at day 8 (* P < 0.05, Student's *t*-test). Error bars indicate s.e.m.

Figure S4

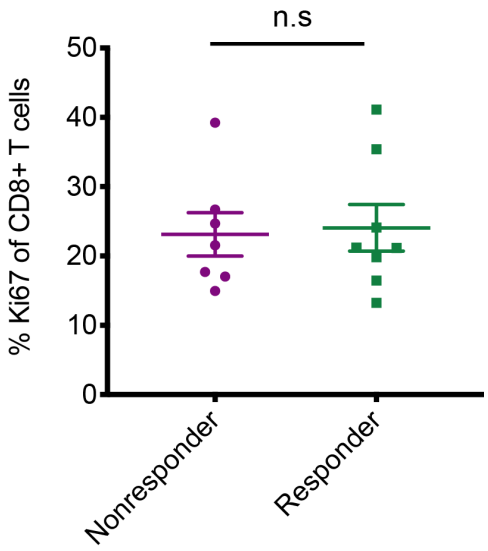

Fig S4. Tumors resected prior to ICB do not differ in Ki67 frequencies of CD8+ T cells between responding and non-responding mice. Ki67 frequencies measured by flow-cytometry at day 0 (before α-PD-1 treatment). P = 0.84 by Student's t-test. Error bars indicate s.e.m.

Figure S5

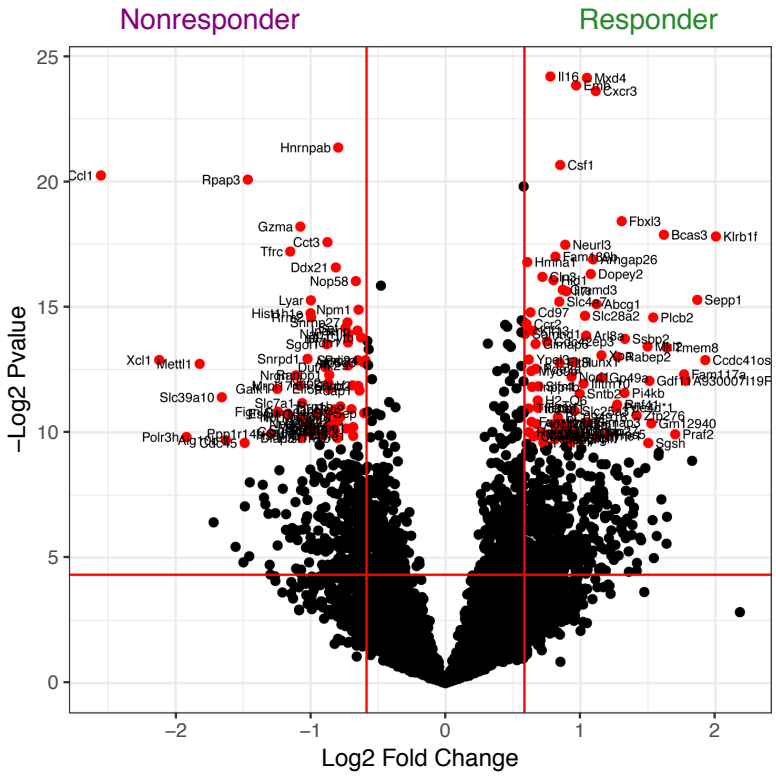

Fig S5. Differential expression across responding and non-responding mice shown for transcriptional profiles from CD8+ T cells of early resected tumors.

Figure S6

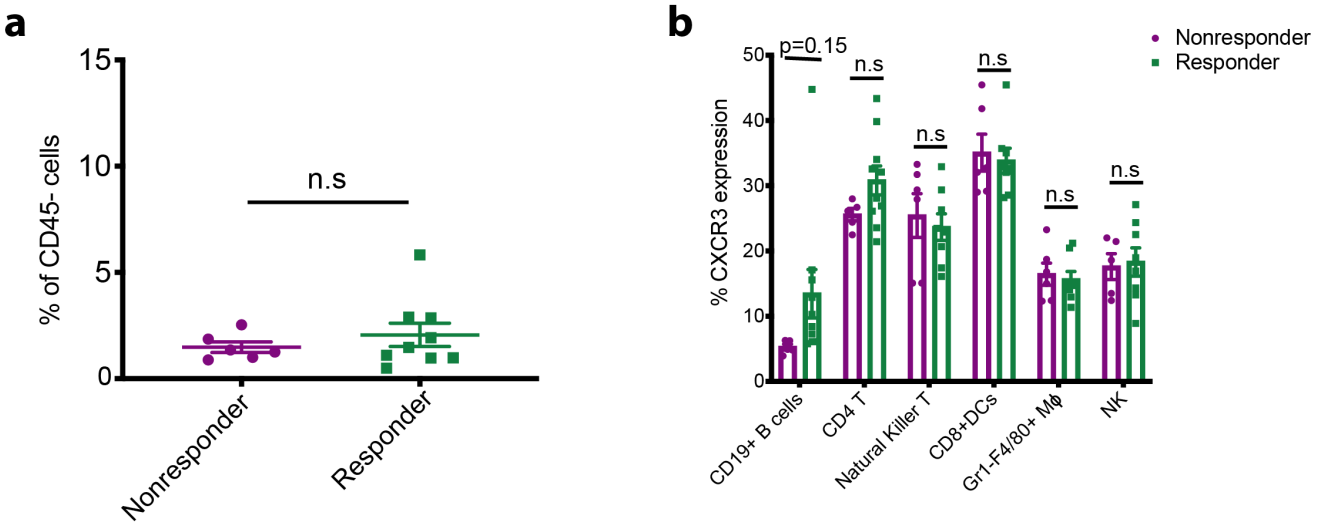

Fig S6. CXCR3 expression across various cell types. Orthotopic E0771 breast tumors received anti-PD-1 mAb on day 8, 11, and 14 after tumor inoculation. The same number of tumor cells were implanted into the left and right mammary fat pad (MFP) of each mouse and one tumor was resected at day 8 prior to treatment for analysis and the other was monitored as responder or non-responder. N=6-9. (a) Flow cytometry analysis of CXCR3 expression in CD45- cells. (b) Expression of CXCR3 on immune cell populations. By Student's *t*-test; Error bars indicate s.e.m.

Figure S7

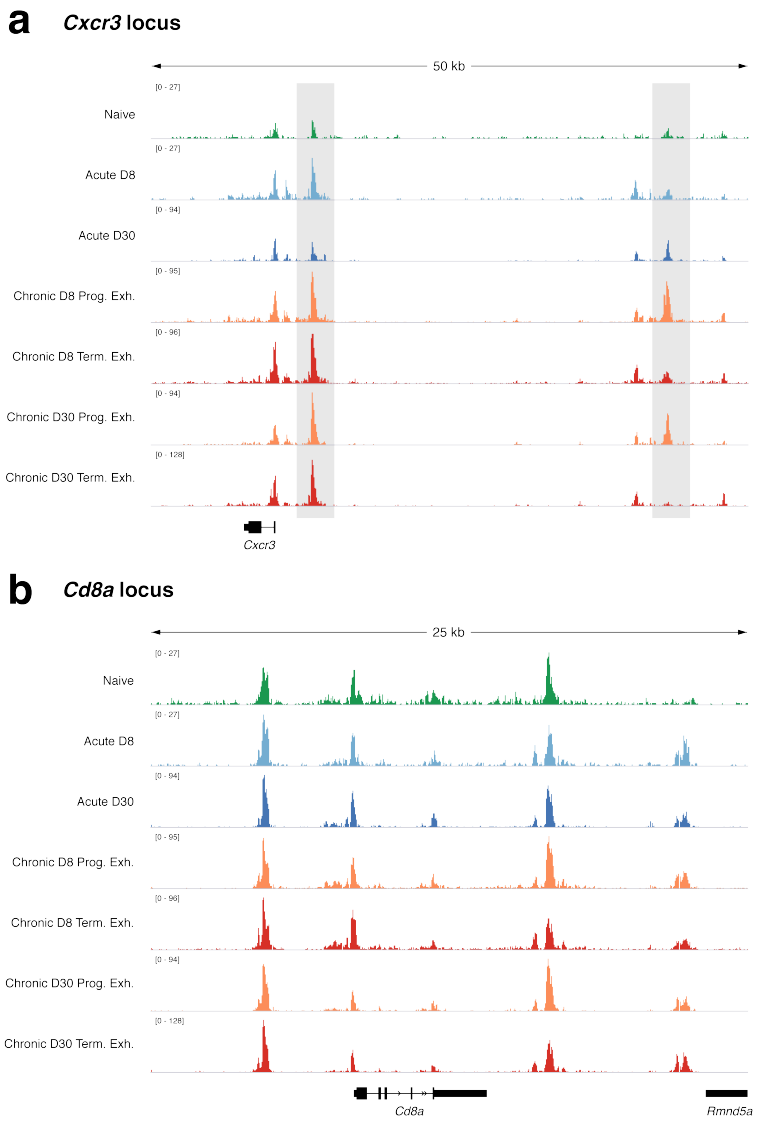

Fig S7. Open chromatin regions in the *Cxcr3* locus in naïve, effector, memory, and exhausted CD8+ T cell subsets. Representative ATAC seq tracks at the (a) *Cxcr3* locus and (b) *Cd8a* locus from CD8+ T cells sorted from either naïve mice or mice at day 8 or day 30 post-infection with acute (Armstrong) or chronic (Clone 13) lymphocytic choriomeningitis virus (LCMV). ATAC seq data accessed from (55, 57, 58) and all samples were normalized to the *Cd8a* locus (b) as a control for the visualization with IGV 2.7.0. peak viewer. The grey bar in (a) indicates peaks that show differences between groups. Abbreviations used: “Prog. Exh.” = progenitor exhausted subset, sorted as CD44+ PD-1+ SlamF6+ Tim-3-, and “Term. Exh.” = terminal exhausted subset, sorted as CD44+ PD-1+ SlamF6- Tim-3+. CD8+ T cells from naïve and acute LCMV mice were sorted as bulk CD8+ T cell populations.

Figure S8

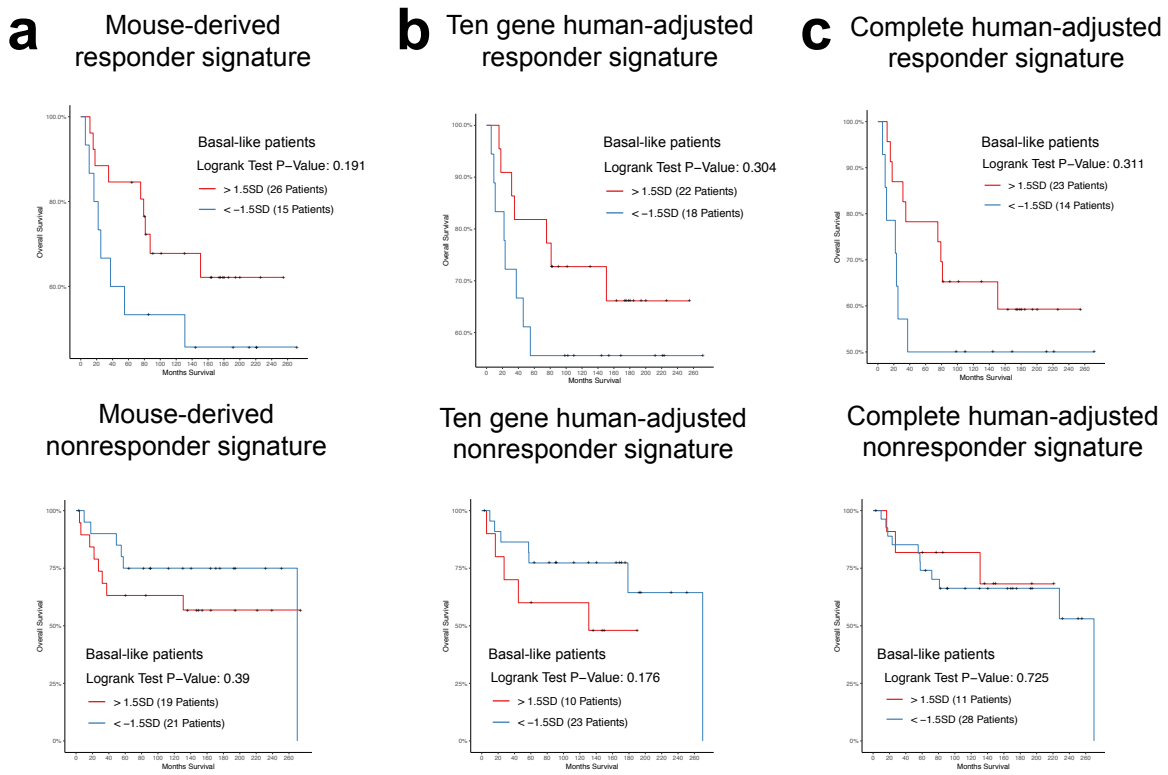

Fig S8. Mouse derived early-therapeutic-stage responder / non-responder gene signatures trend towards being predictive of patient survival and response to ICB in basal-like breast cancer patients. (A-C) An analysis of association of BC patient survival (METABRIC database, restricted to basal-like cancers) with high or low scores for each of the responder and non-responder signatures (Methods) restricted to patients with basal-like breast cancer. (A) Partitioning BC patients with basal-like cancers by their high (red) or low (blue) expression of the mouse-derived responder signature (top) and non-responder signature (bottom). (B-C) Following the adjustment of the mouse-derived early-therapeutic-stage signatures to the human BC landscape (Methods), partitioned BC patients with basal-like cancers by expression of the ten gene responder (B, top) or nonresponder (B, bottom) signatures or the complete responder signature (C, top) or nonresponder (C, bottom) signature.