

Supplemental Methods

Immunohistochemistry

The following antibodies were used for IHC staining: mouse monoclonal anti-PD-L1 antibody (E1L3N, Cell Signaling Technology; 1:200) and mouse polyclonal anti-CD3 antibody (Agilent, 1:400). Five-micron sections were baked at 60°C for an hour, followed by deparaffinization, rehydration, and epitope retrieval using the Dako PT Link platform (Agilent). IHC staining was carried out on the Dako Link 48 autostainer (Agilent), with antibody incubation for 60 minutes, amplification using Envision FLEX mouse linkers, and visualization using the Envision Flex High-sensitivity visualization system (Agilent).

DNA isolation, hybrid capture, library preparation, and sequencing

We cut TMAs into 20-micron sections and scraped each core with a blade into a 1.5 low adhesion microcentrifuge tube for DNA extraction using the Qiamp DNA FFPE Tissue Kit Cat# 56404 following the manufacturer's protocol. DNA concentration was measured by QUBIT DSDNA HS ASSAY (Thermo Fisher Cat# Q32854).

Agilent SureSelect XT HS part# G9704E was used to generate the libraries from 42 "hot", "cold" and benign samples following the manufacturer's protocol. 22 to 70 ng of input gDNA was used per sample. We performed 13 cycles of pre-hybridization PCR for all samples to amplify, index and barcode individual libraries prior to hybrid capture. Two to three libraries (500 ng of amplified DNA per library) were grouped into 18 pools corresponding to 11 patients, for pooled capture reactions. Hybrid capture was performed using a custom bait design that covers the coding regions of 800 cancer-related genes plus the entire locus for a subset of tumor suppressors and oncogenes (Design ID: 3183291; 56762 probes; 3.9Mb capture size). We performed 10 cycles of PCR to amplify post-capture material. QC was performed prior and post hybrid capture using the Agilent TapeStation platform.

Post-capture libraries were combined in equimolar ratios into two separate pools based on their unique library indices. The two pools were sequenced on an Illumina NovaSeq SP FlowCell with an XP2 loading kit, with 150 cycles paired-end (2x150). Pass-filter reads were demultiplexed into FASTQ files, trimmed using Agilent SureCallTrimmer version 4.0.1 and aligned at the lane-

level to the human genome (version b37+decoy) with BWA-mem version 0.7.17. SAM files were sorted, duplicate-marked and base quality score recalibrated using the Genome Analysis Toolkit (GATK) version 4.1.3.0. Lane-level BAM files were merged using Picard and duplicate-marked again using GATK. Capture efficiency and coverage were determined using GATK CollectHSMetrics.

Somatic mutations were identified using GATK MuTect2, first by calling all normal samples in tumor-only mode to generate a panel-of-normals, and then running MuTect2 again in tumor-normal (paired) mode also against the panel-of-normals. VCF files were filtered by MuTect2 to remove false-positives, artifacts, and orientation-bias mutations. Germline mutations were identified using GATK HaplotypeCaller in cohort mode, combining and genotyping each VCF file using GATK CombineGVCFs and GenotypeGVCFs, respectively. The unfiltered VCF file was recalibrated for point mutations and insertions/deletions using VariantRecalibrator against HapMap 3.3, 1000 Genomes Omni 2.5, 1000 Genomes Phase 1 SNPs, dbSNP version 138, and the Mills/1000 Genomes gold-standard indel set. Filtered somatic and germline VCF files were annotated with Oncotator (version 1, April052016 data corpus). All mutations were individually inspected and curated using the Integrated Genome Viewer (IGV). Tumor mutation burden (TMB) was calculated by determining the number of point mutations at 5% or greater variant allele fraction from all coding regions of the genome covered by the capture library and dividing it by the number of bases of the coding regions of the genome covered by the capture library.

Somatic copy number alterations (SCNA) were inferred from the panel-based sequencing data using the GATK SCNA pipeline. First, read counts were determined from each tumor or normal BAM file using GATK CollectReadCounts, and a panel-of-normals was used to denoise the read counts of all tumor samples using GATK CreateReadCountPanelOfNormals and GATK DenoiseReadCounts, respectively. Next, b-allele frequencies were determined for all tumor and normal BAM files using GATK CollectAllelicCounts. Finally, GATK ModelSegments was used to combine b-allele data and smoothed read count for each tumor/normal pair to generate segmentation (SEG) files with whole-genome copy number estimates. Copy numbers were called using GATK CallCopyRatioSegments and modeled with regions of heterozygosity with GATK PlotModeledSegments. The final modeled SEG file was then processed using GISTIC 2.0 (GenePattern module version 6.15.28) to call gene-level gains or losses (threshold 0.1 for single-copy, 1.3 for two-copy). All gene-level copy number calls were manually inspected and

curated using IGV. Percent genome altered (PGA) was calculated by adding the lengths of all genomic segments estimated to have greater than 0.1 or less than -0.1 \log_2 copy number ratios, and dividing it by the length of the human genome.

FASTQ files from each of the 489 prostate cancer TCGA cases were downloaded from the NCI Genomic Data Commons, aligned with STAR version 2.7.0f and gene expression summaries were estimated using RSEM. Transcripts per million (TPM) values for each gene were processed using single-sample geneset enrichment analysis against the mSigDB Oncogenic Gene Signatures genesets. A two-sided t-test was used to compare the average ssGSEA score for each geneset in the CHD1-deletion or CHD1-WT group, and then P values were adjusted using the method of Benjamini and Hochberg.

Multiplex immunofluorescence (IF) analysis

IF panel 1: A seven-plex IF assay was performed on 4- μ m FFPE sections, using Leica Bond Rx autostainer. Briefly the staining consists of sequential tyramine signal amplified immunofluorescence labels for each target, and a DAPI counterstain. Each labeling cycle consists of application of a primary antibody, a secondary antibody conjugated to horse radish peroxidase (HRP), and an opal fluorophore (Opal 690, Opal 570, Opal 540, Opal 620, Opal 650 and Opal 520, Akoya Biosciences), respectively. The stained slides were scanned on a Perkin Elmer Vectra 3 imaging system (Akoya Biosciences) and analyzed using Halo Image Analysis platform (Indica Labs). Each single stained control slide is imaged with the established exposure time for creating the spectral library. We ran an algorithm learning tool utilizing the Halo image software training for the gland and stroma regions, and subsequently completed cell segmentation. The thresholds for the antibodies were set respectively, based on the staining intensity, by cross reviewing more than 20 images. Cells with the intensity above the setting threshold were defined as positive.

IF panel 2: The assay was performed on a Bond RX Autostainer (Leica Biosystems) using the Opal multiplex IHC system (PerkinElmer/Akoya Biosciences Cat# NEL871001KT), as described in detail previously (31) (DOI: [dx.doi.org/10.17504/protocols.io.bjbzkip6](https://doi.org/10.17504/protocols.io.bjbzkip6)). Whole slide multispectral images were acquired at 10x magnification using Vectra 3.0 (PerkinElmer/Akoya Biosciences) and visually inspected by using Phenochart 1.0.9 (PerkinElmer/Akoya Biosciences, RRID:SCR_019160) to select five intra-tumoral CD8⁺ T cell-enriched regions of

interest (ROIs; area of each ROI = 669 μm x 500 μm) per case. In one case, additional ROIs were taken to ensure that a minimum of 100 CD8⁺ cells would be analyzed. ROIs were scanned at 20x magnification, imported into InForm v2.2.0 (PerkinElmer, RRID:SCR_019161) and deconvoluted using a multispectral library built with single stain slides. Image analysis was performed using HALO v2.1.1637.18 (Indica Labs, RRID:SCR_018350) and the Indica Labs High-Plex FL v2.0 module. A unique algorithm was created for each case, and its accuracy validated through visual inspection by a pathologist (M. Ficial) with extensive experience in image analysis.

IF panel 3: We deparaffinized sections in successive incubations with xylene and decreasing concentrations (100, 95, 75, 50, 0%) of EtOH. Antigen retrieval utilized Abcam 100x TrisEDTA Antigen Retrieval Buffer (pH = 9) heated under high pressure. We washed sections in PBS + 0.1% Tween20 before antibody staining. We blocked sections for 30 min with a solution of 10% goat serum in 1x PBS + 0.1% Tween20. before staining with primary and secondary antibodies. Primary antibodies were used at a concentration of 1:100 (MHC-II) or 1:150 (CD8, TCF1) and incubated for 1 h at room temperature. Secondary antibodies were used at a concentration of 1:250 (A488, A568) or 1:500 (A647) and incubated for 30 min at room temperature. We collected IF images using a Zeiss Z.1 Slide Scanner equipped with a Colibri 7 Flexible Light Source. We used Zeiss ZenBlue software for post-acquisition image processing. We analysed whole-slide images using CellProfiler and custom R and python scripts, as previously described (32). We examined 100 x 100 μm regions and evaluate whether antigen-presenting cell (APC) niches are present or not, by the definition outlined in the main body of the manuscript, and calculated percentage of tumor containing APC niches as the number of regions with APC niches divided by the total number of regions containing DAPI⁺ cells. CellProfiler was used to identify cellular objects within images and measure fluorescence intensity. Custom R and python scripts were used to perform quality control and remove imaging artifacts, measure distance between cellular objects, and calculate cellular density.

Supplemental Figure 1: (A) Analyzer output of TCR sequencing and (B) percentage of productive rearrangements from the top 10 rearrangements (blue) versus all others (grey)

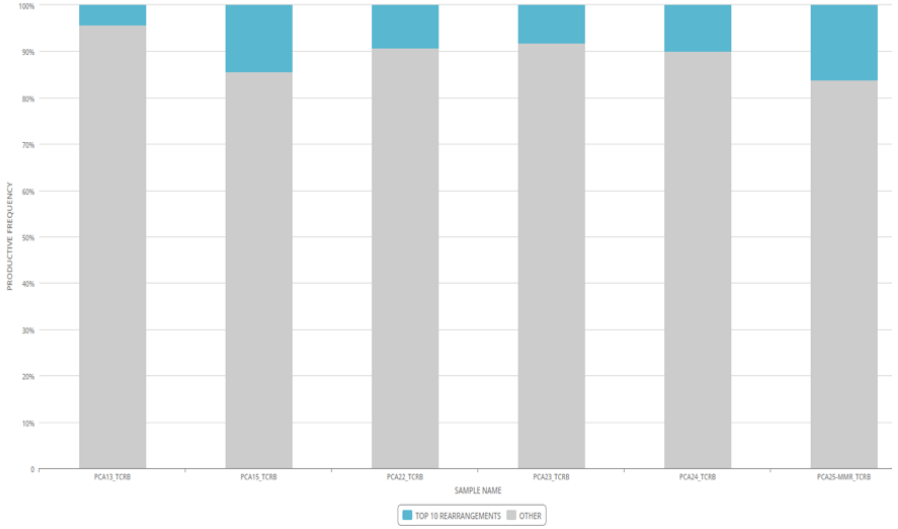
A.

6 Total Samples 74,300 Total Templates (Sum) 5.8084% Max Productive Frequency 0.1091 Max Productive Clonality

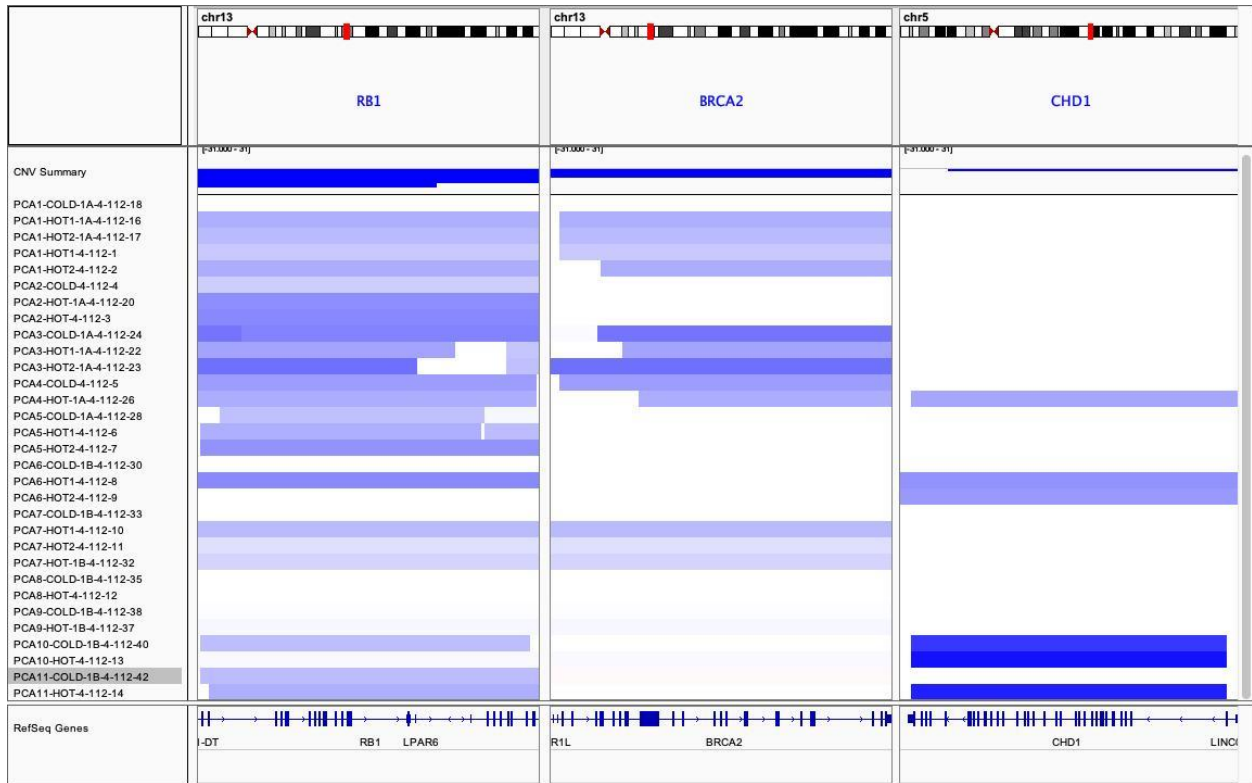
Project Settings: Exclude DJ Rearrangements for B-Cells

Sample Name	Total Templates (Sum)	Total Productive Templates (Sum)	Fraction Productive	Rearrangements	Productive Rearrangements	Productive Clonality	Max Productive Frequency	Locus
PCA15_TCRB	7,132	5,699	0.7991	3,681	2,951	0.1001	4.386735%	TCRB
PCA13_TCRB	25,447	21,554	0.847	16,198	13,645	0.0462	0.793356%	TCRB
PCA23_TCRB	10,096	8,606	0.8524	6,739	5,662	0.0589	1.626772%	TCRB
PCA25-MMR_TCRB	11,918	9,228	0.7743	6,970	5,521	0.1063	5.808409%	TCRB
PCA22_TCRB	4,185	3,376	0.8067	3,037	2,480	0.048	2.162322%	TCRB
PCA24_TCRB	15,522	12,609	0.8123	7,302	5,956	0.1091	2.63304%	TCRB

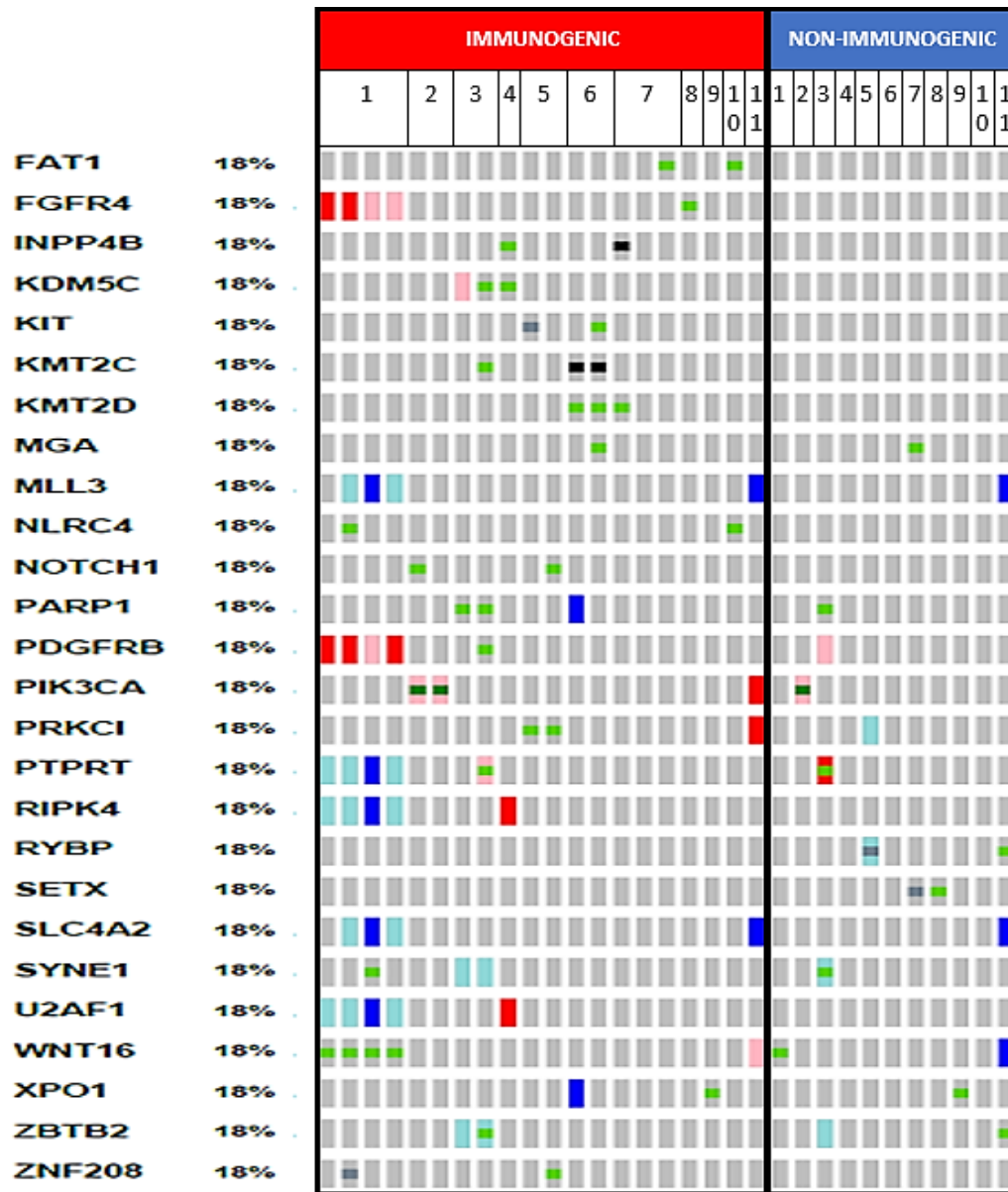
B.



Supplemental Figure 2: Copy-number variation in *RB1*, *BRCA2*, and *CHD1* in individual patients and tumor foci (“cold” indicating non-inflamed focus, “hot” indicating inflamed focus)

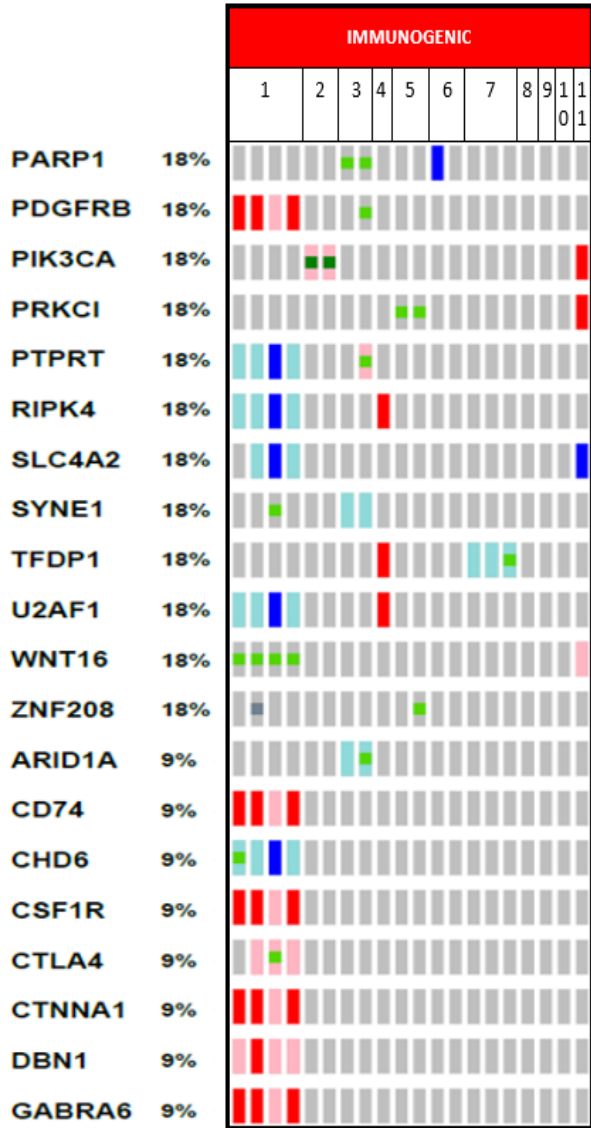


Supplemental Figure 3: Complete genomic alterations present in immunogenic and non-immunogenic foci (without first panel showing top alterations, shown in main **Figure 4**)



Genetic Alteration

- Missense Mutation (putative driver)
- Missense Mutation (unknown significance)
- Truncating Mutation (putative driver)
- Truncating Mutation (unknown significance)
- Amplification
- Gain
- Deep Deletion
- Shallow Deletion
- No alterations



Supplemental Figure 5: Most common alterations in non-immunogenic foci

