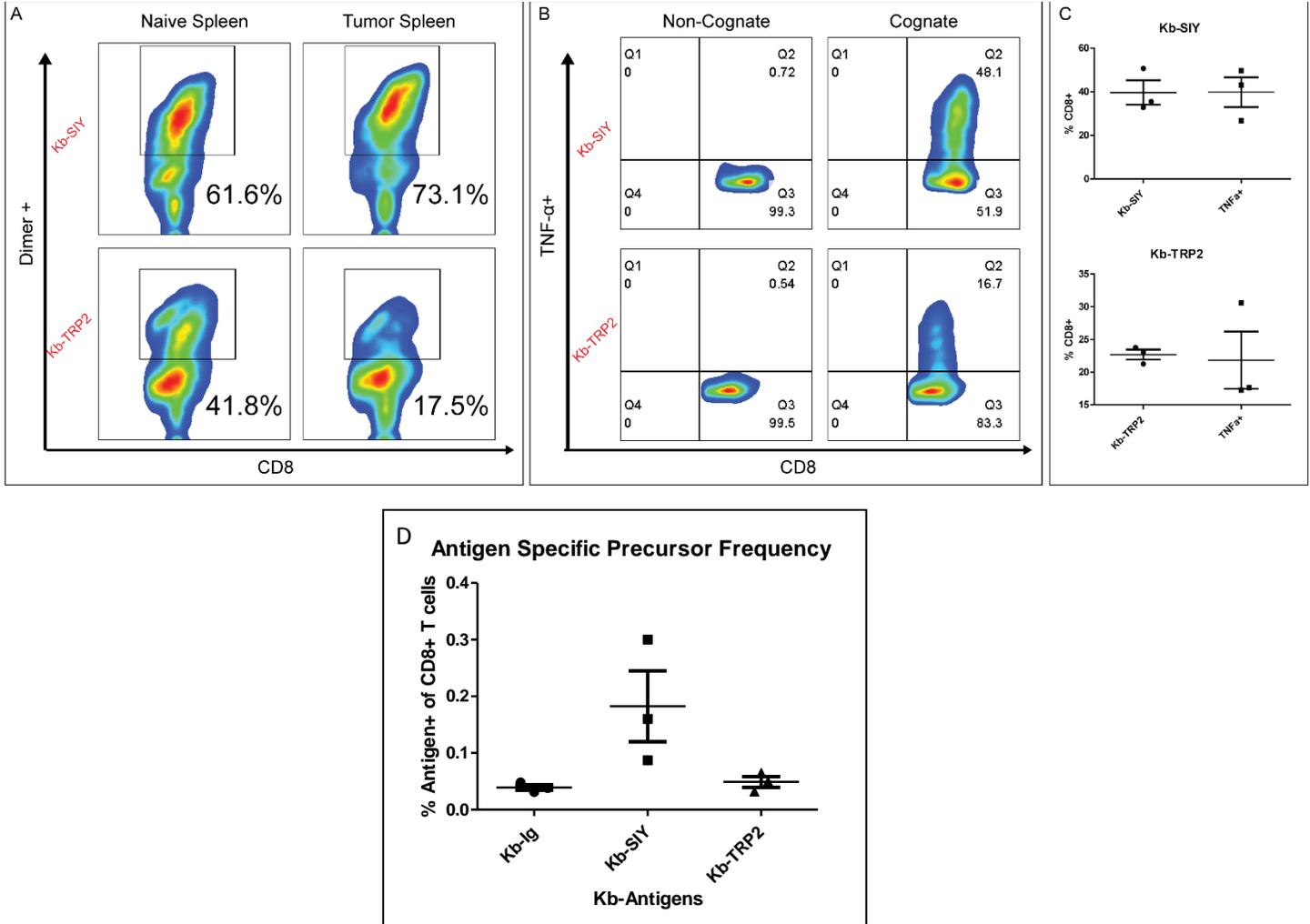


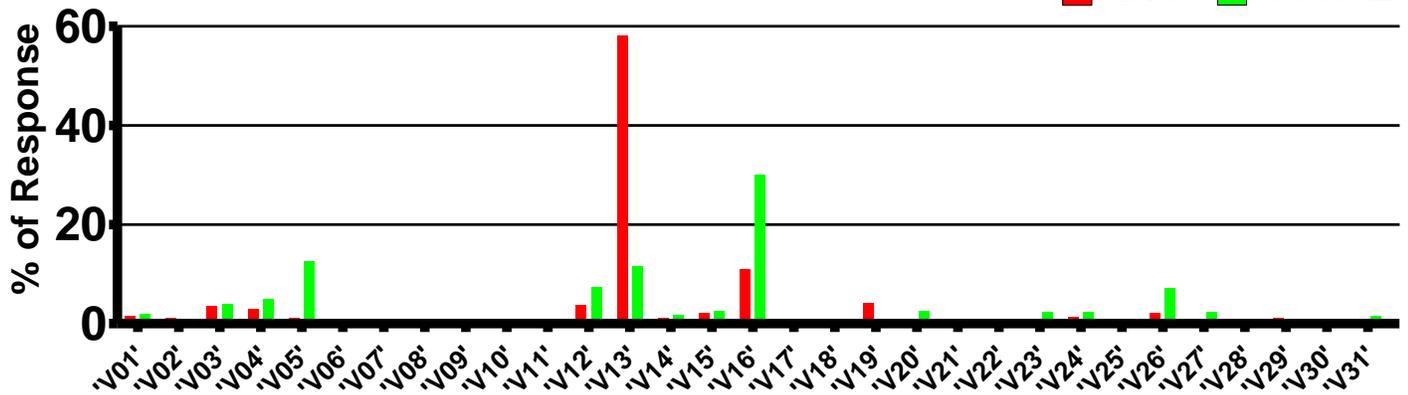
Supplementary Figure 1. In order to set thresholds for motif detection, two variables were optimized on a Naïve Adult B6 CD8 Repertoire (taken from Adaptive Biotechnologies ImmunoSeq Sample Data). Motif detection was completed across a range of phylogenetic distances and frequency thresholds and number of motifs detected was monitored. Since our analysis was looking for dominant motifs above what is present in an unexpanded population, we chose a frequency threshold of 0.03 and Phylogenetic Distance threshold of 0.35, at which 0 motifs were detected in Naïve B6 background.



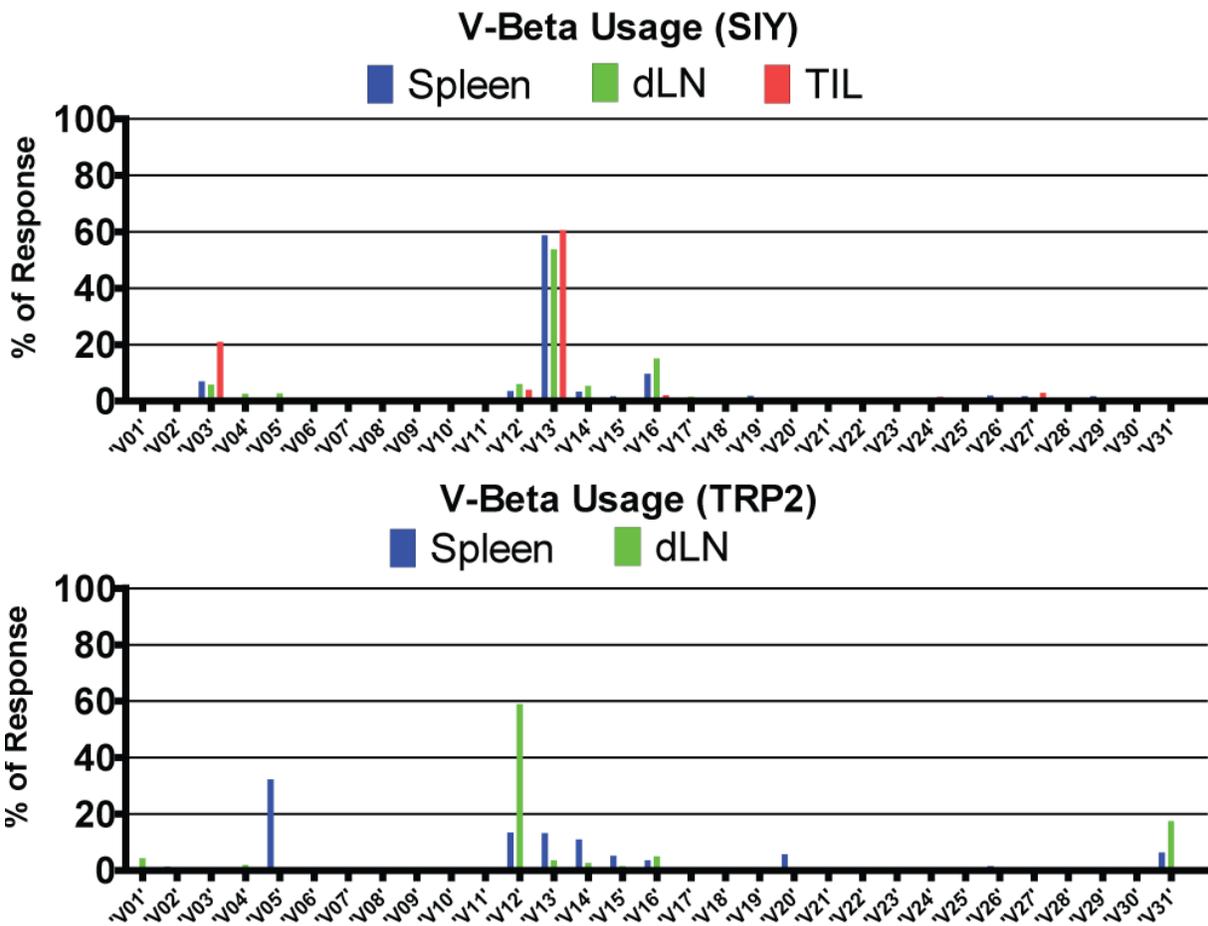
Supplementary Figure 2. A) FACS analysis of antigen-specific CD8 T cells on D7 in Naïve and Tumor-bearing lymphoid organs. B) ICS staining of antigen-specific CD8 T Cells confirming specificity and functionality. C) Comparison of Dimer+ and TNF+ CD8. D) Antigen-specific CD8 T cells staining from splenic CD8 T cells directly ex vivo compared to unloaded Kb-Ig staining. N = 3, Statistical 2-tailed T-test.

V-beta Usage of CD8 Response

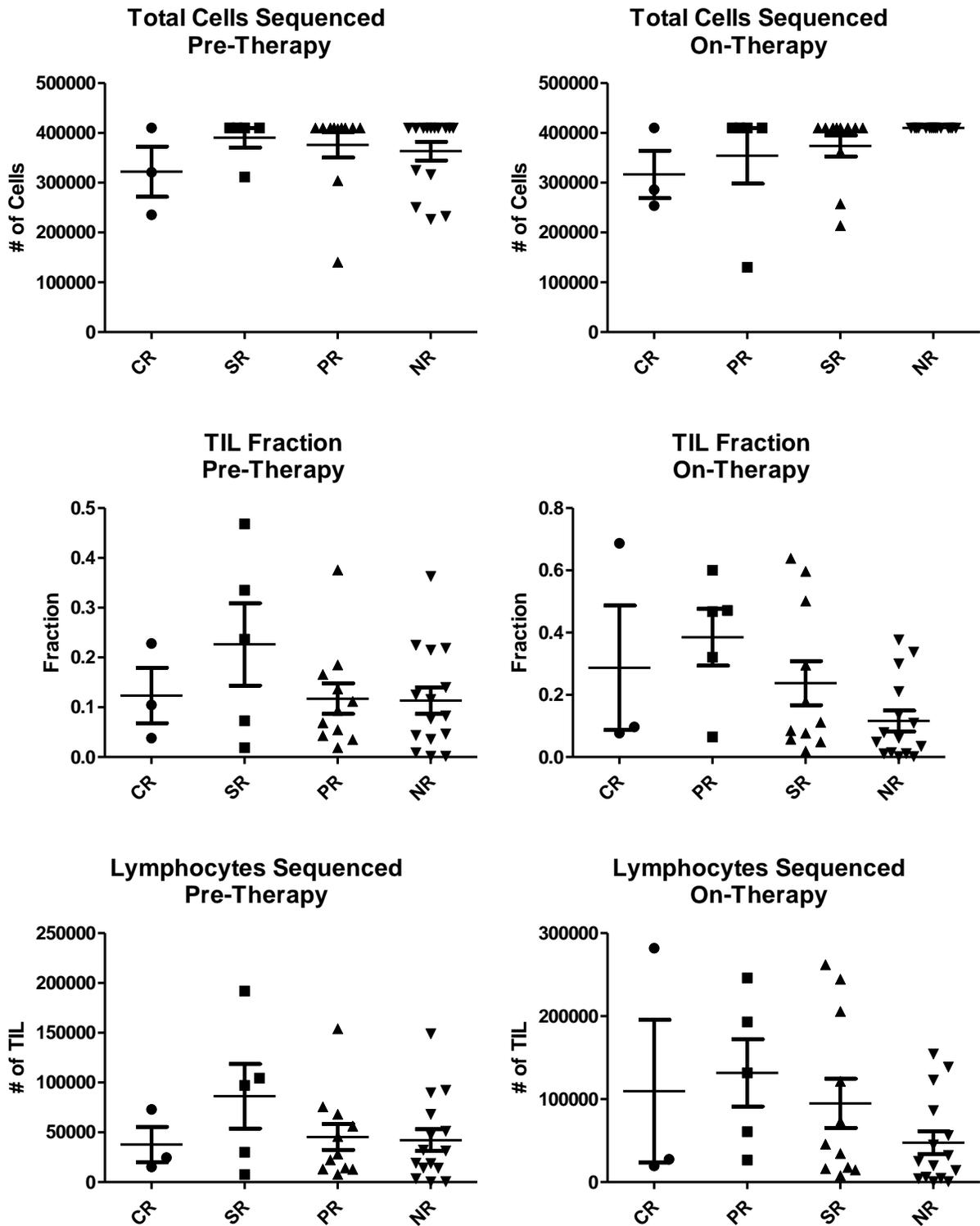
■ SIY ■ TRP2



Supplementary Figure 3. V-beta usage for Naïve Kb-SIY & Kb-TRP2 Response



Supplementary Figure 4. V-beta usage for Tumor-Bearing Kb-SIY & Kb-TRP2 Response in Various Lymphoid Organs



Supplementary Figure 5. In order to determine number of tumor-infiltrating lymphocytes that were sequenced for each patient, Adaptive reported amount of total DNA in nanograms that underwent sequencing and based on the assumption of 6.5pgDNA per cell, we were able to calculate the number of total cells that underwent

sequencing. Furthermore, Adaptive calculated a %TIL metric based on number of non-recombined to recombined sequence reads. With this information, we were able to deduce the number of starting lymphocytes that were sequenced for each patient.

A)

$$\text{Sequence Distance} = \left(1 - \frac{\text{Score}_{12}}{\text{Score}_{11}}\right) \left(1 - \frac{\text{Score}_{12}}{\text{Score}_{22}}\right)$$

$$\text{Mapped Sequence Distance} = \frac{1}{1 + \left[\left(1 - \frac{\text{Score}_{12}}{\text{Score}_{11}}\right) \left(1 - \frac{\text{Score}_{12}}{\text{Score}_{22}}\right)\right]}$$

	Sequence 1	Sequence 2	Sequence 3	Sequence 4	Sequence 5	Sequence 6	Sequence 7
Sequence 1		Blue	Blue	Blue	Blue	Blue	Blue
Sequence 2			Blue	Blue	Blue	Blue	Blue
Sequence 3				Blue	Blue	Blue	Blue
Sequence 4					Blue	Blue	Blue
Sequence 5						Blue	Blue
Sequence 6							Blue
Sequence 7							

B)

```

for i=1:size(DistanceMatrix3,1);
    for j=1:size(DistanceMatrix3,2);
        if i==j
            if Reads(i)==1
                ScorePreOut(i,j)=0;
            else
                ScorePreOut(i,j)=DistanceMatrix3(i,j)*combntns(Reads(i),2);
            end
        else
            ScorePreOut(i,j)=DistanceMatrix3(i,j)*Reads(i)*Reads(j);
        end
    end
end
end

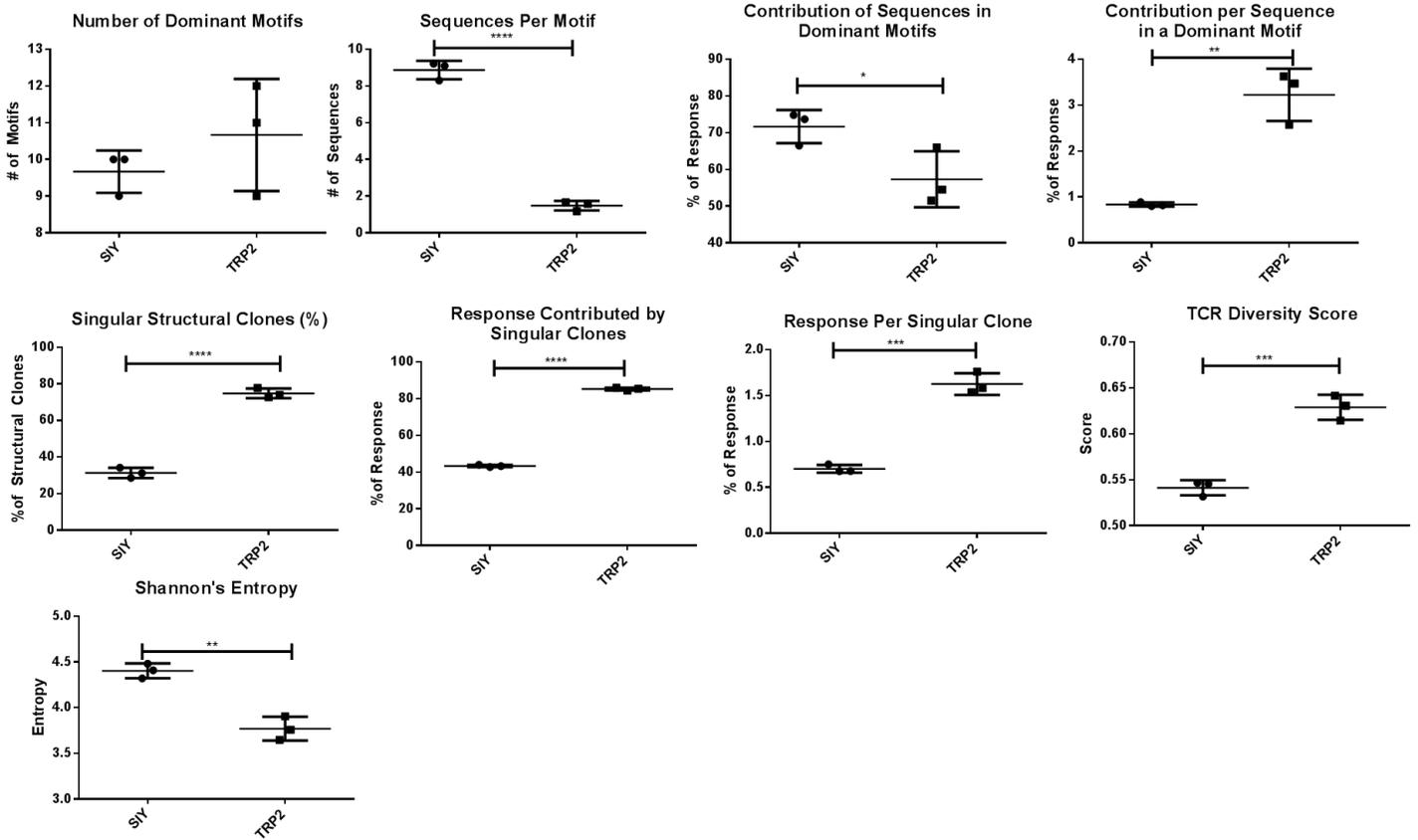
```

```
TCRDiversityScore=1-(sum(sum(ScorePreOut))/combntns(sum(Reads),2));
```

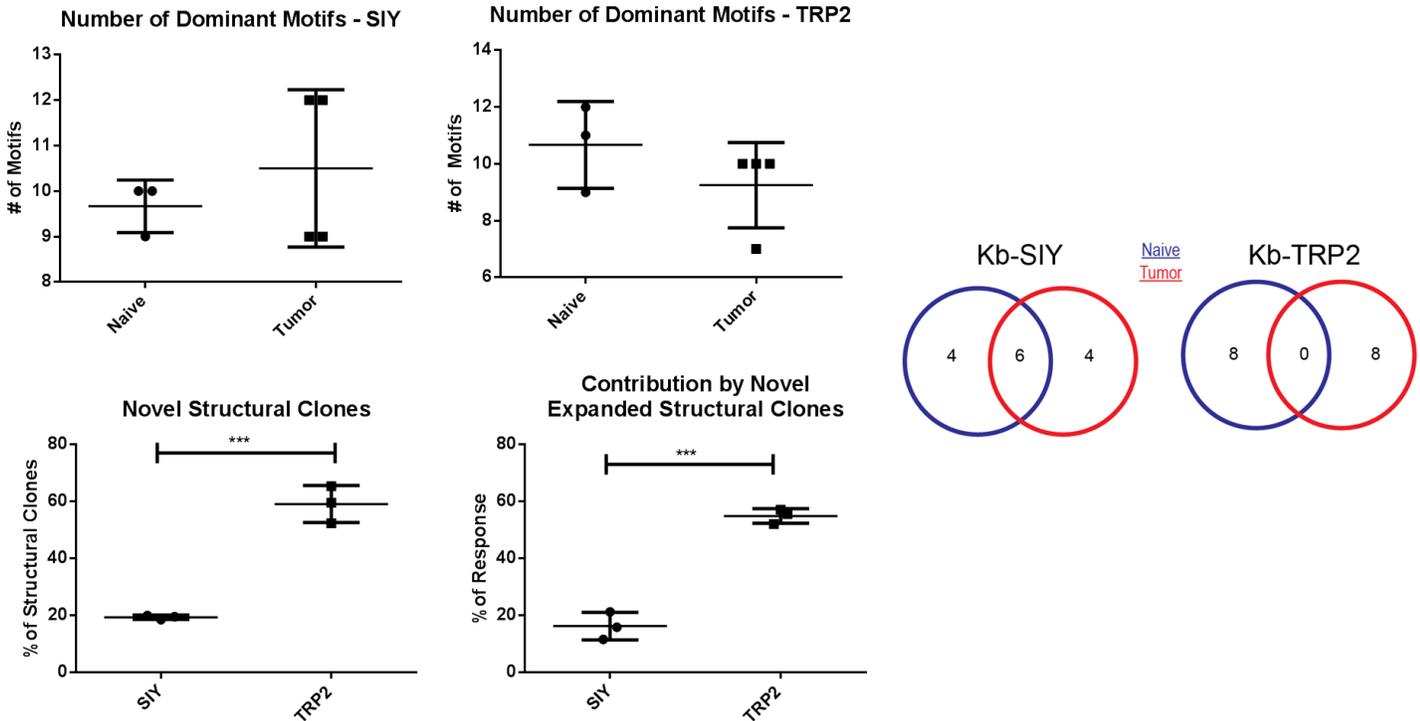
Supplementary Figure 6. Calculation of TCR Diversity Score. A) Initially, all pair-wise distances are calculated with sequence distance based on global alignment scores. Sequence distance is then converted to a mapped sequence distance with values between 0 and 1. This Mapped Sequence Distance is defined as 0 being infinite sequence difference and 1 being identify. B) The distance matrix that is calculated is then weighted by the number of reads. The purpose of this step is to determine the average distance between every cell in the analysis. The piece of code iterates through the distance matrix calculated in A. If $i=j$ (meaning that we are examining the same sequence against itself) and there exists only 1 read of that sequence, the new matrix calculation is 0, meaning there is no need to calculate a distance between a read and itself. If there is more than 1 read, then the new matrix entity calculation is the distance of the sequence and itself multiplied by the total number of all possible combinations of those reads. For example, if there are 10 reads of a given sequence, this new matrix entity is calculated as 1 (sequence distance) * 45 (all possible combinations of 10 reads). In all other

cases, where the two sequences are different, the new matrix entity is calculated as the distance between those two sequences multiplied by the number of reads of the first sequence multiplied by the reads of the second sequence. Finally, this new matrix is summed and divided by the number of possible combinations of all reads. This number is then subtracted by 1 to give the final Mapped Sequence Distance where 0 represents identity and 1 represents infinite difference.

Naïve SIY vs. TRP2 Repertoire



Naïve vs. Tumor-Bearing SIY & TRP2 Repertoire



Supplementary Figure 7. Duplicates of Murine Experiments. Corresponding duplicate figures to Figures 2&3 in main manuscript showing differences in SIY/TRP2 repertoire in naïve and tumor-bearing setting.

The screenshot displays the ImmunoMapQuant software interface. At the top center is the ImmunoMAP logo, which features a stylized tree structure with colored nodes. Below the logo, the interface is divided into several sections:

- Select Files:** A list box containing two files: "SIY_Control_Spleen_3" and "SIY_Spleen_Tumor_2".
- ImmunoMap Parameters:**
 - Select Sequences for Analysis by:** A dropdown menu with options: "Frequency Cut (%)", "Fraction of Response (%)", "Number Of Unique CDR3", and "Number of Reads". The "Number of Reads" option is selected, with a value of "25" entered in the adjacent text box.
 - Homology Threshold:** A text box containing ".35".
 - Cluster Frequency Threshold:** A text box containing "1".
 - Scoring Matrix:** A dropdown menu with options: "BLOSUM62", "PAM10", "DAYHOFF", and "GONNET". "PAM10" is selected.
 - Gap Penalty:** A text box containing "30".
- Single File Analysis:**
 - A large button labeled "Run ImmunoMap".
 - A "Save Table to CSV" button.
 - A table with the following data:

	Filename	# of Unique CDR3	Shannon Entropy
1	SIY_Cont...	6	0.7995
2	SIY_Sple...	5	0.7499
 - A dropdown menu for "SIY_Control_Spleen_3.tsv" showing a list of motifs:
 - Motif 1 - 8.5587%
 - Motif 2 - 6.1875%
 - Motif 3 - 3.8162%
 - Motif 4 - 3.5198%
 - Motif 5 - 3.4087%
 - A "Visualize Dominant Motifs" button.
- Multiple File Analysis:**
 - A "View Weighted Repertoire Dendrogram" button.
 - Two sample selection dropdowns: "Sample #1" (SIY_Control_Sp...) and "Sample #2" (SIY_Control_S...).
 - A "Compare Two Repertoires" button.
 - Four empty text boxes for results:
 - Novel Repertoire (% of Sequences)
 - Novel Repertoire (% of Response)
 - # of Motifs (Sample 1)
 - # of Motifs (Sample 2)
 - A "Homologous Motifs" table:

	1	2
1		
2		
3		
4		

Supplementary Figure 8. ImmunoMap Graphical User Interface & Instructions for Use.

- 1) User selects files for analysis by pressing the 'Select Files' button and importing tsv files of TCRSeq Data exported by Adaptive Biotechnologies. If one has a different source of TCRSeq, one can submit a tsv file with the first column being nucleotide sequences, the second column being amino acid sequence, and the third column being number of counts.
- 2) After importing the files, one can set a variety of ImmunoMap parameters including how much of the file to use, structural homology thresholds, cluster frequency thresholds as well as parameters such as the scoring matrix used and the gap penalty.
- 3) After parameters have been set, one highlights the files under the 'Select Files' button they want to analyze at once. At this point, the user can press 'Run Immunomap' and get a table of all relevant ImmunoMap metrics for their respective files. For each file, they can view the multiple alignments of their dominant motifs by selecting the file and the motif in the window to the right and pressing 'Visualize Dominant Motifs.' Finally, one can save a csv file with all the summarized ImmunoMap metrics by pressing the 'Save Table to CSV' button. This file can be opened by Microsoft Excel using a comma as the delimiter.
- 4) If one desires to compare multiple files by visualizing them by the Weighted Repertoire Dendrograms, they can highlight the desired files for visualization in the window beneath the 'Select Files' button and press 'View Weighted Repertoire Dendrogram.'
- 5) Finally, if one desires to compare two repertoires by seeing the percent of structural overlap as well as shared dominant motifs, one should select the two files they want to compare in the drop menu's provided and press the 'Compare Two Repertoires' button.