

Background

The advent of immune sequencing has allowed scientists and clinicians to understand the interaction of the immune system with various pathologies in a completely novel way. Its initial applications opened our understanding to the depth and breadth of the intrinsic T Cell Receptor (TCR) and B Cell Receptor (BCR) repertoire. Further applications of immune sequencing have helped aid physician-scientists in vaccination development and tracking progression of disease in lymphoid malignancies. Of recent interest, sequencing efforts have been placed on understanding immune responses responsible for controlling oncological malignancies, understanding the antigen-specific as well as the tumor infiltrating lymphocyte (TIL) repertoire.

With the abundance of this new data, there has arisen a need to analyze TCR sequences in a biologically meaningful way. Current methods of analysis have fallen short of providing an intuitive understanding of the repertoire for largely two major reasons: (1) they ignore sequence relatedness and instead focus on diversity as a function of number of sequences and frequency (i.e. Shannon Entropy) and (2) methods that attempt to assess similarities between different repertoires apply stringent criteria where exact TCR clonotypes (whether at the nucleotide or amino acid level) are examined to assess similarity.

To address these gaps in immune repertoire analysis, we have developed ImmunoMap, a set of sequencing analysis based algorithms to visualize and quantify immune repertoire diversity and allow assessment of similarity between TCR sequences and display the scope of diversity between different repertoires. The novelty of this approach lies in the fact that it combines information about frequency and relatedness of TCR sequences through phylogenetic and sequence analysis to determine relatedness of a given repertoire within itself as well as to other repertoires. Furthermore, when comparing different repertoires and determining their similarity, instead of requiring sequences to be identical to establish similarity, this approach requires a certain level of sequence homology instead; with the underlying assumption that highly homologous sequences will have the same ability to bind cognate antigen. This biologically relevant view of the T cell repertoire attempts to reconcile structure and function of TCR sequences, allowing us to understand antigen-specific CD8 T cell responses.

Using ImmunoMap on endogenous responses to Kb-TRP2, a self-peptide-MHC complex, and exogenous epitope Kb-SIY a foreign peptide, akin to a viral antigen or frame-shift mutation in cancer where a completely novel epitope is produced, gives a descriptive analysis of clonal T cell populations that have different levels of negative selection and repertoire diversity. Additionally, this new approach to T cell response has been applied to analyze the characteristics of the TCR repertoire in patients undergoing checkpoint blockade therapy.

Overview of ImmunoMAP

Weighted Phylogenetic Trees: In order to visualize the immune response, our algorithms initially create weighted phylogenetic trees; combining information about sequence relatedness with frequency. The distance from the end of the branches denote phylogenetic distance as determined through sequence homology, and size of the circles at the end of the branches denotes frequency of the read. Colors can either denote V-beta usage or be used to overlay and compare different weighted phylogenetic trees.

Dominant Motif Analysis: When examining antigen-specific responses, it was demonstrated that there is an abundance of sequences that can expand to one epitope. In order to better parse the many sequences that are detected in antigen-specific expansion, we created clusters based on phylogenetic distance and then examined clusters that met certain frequency thresholds, and called them dominant motifs.

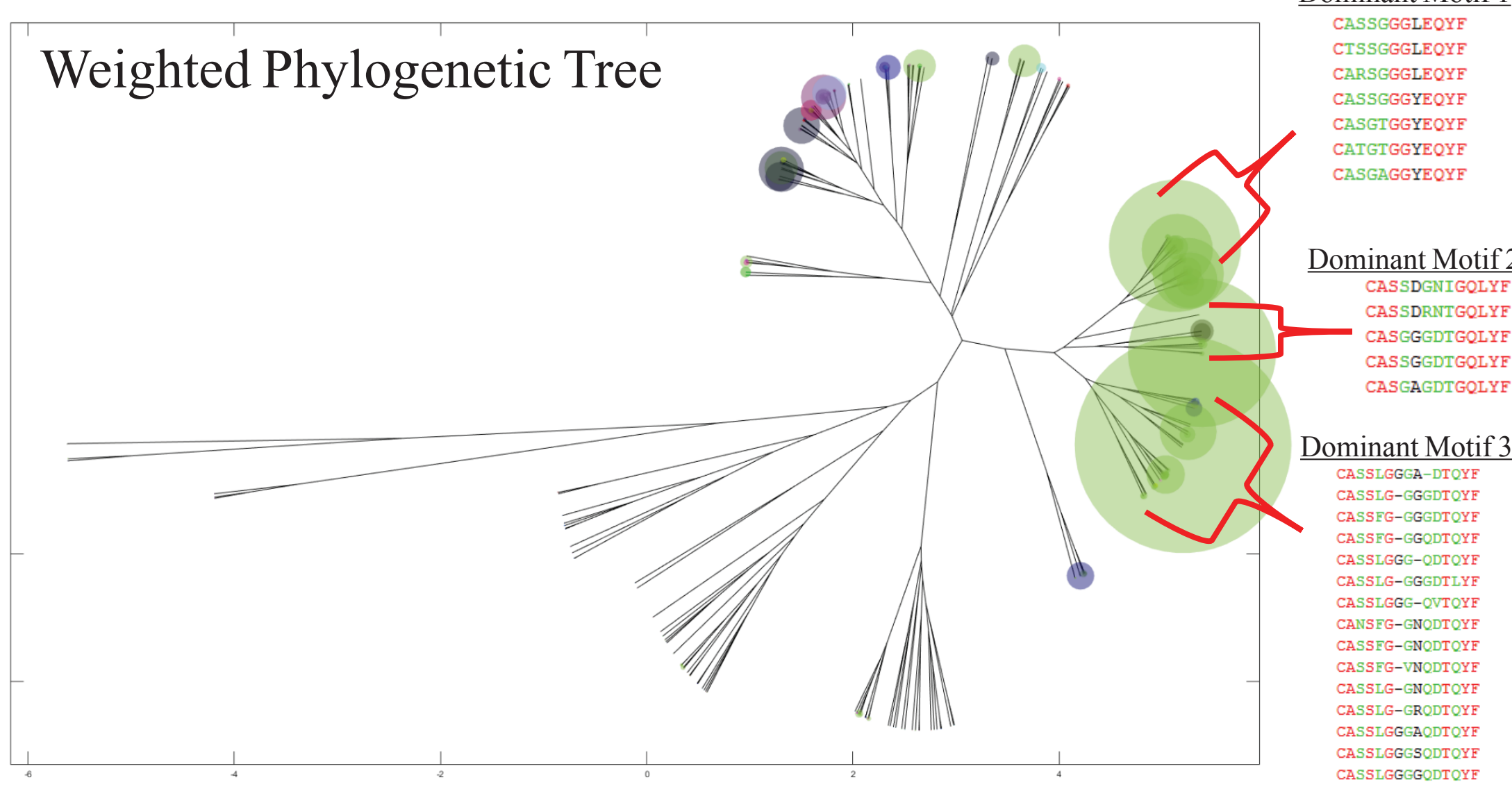
TCR Diversity Score: In an effort to additionally quantify the entire diversity of the TCR repertoire, we created a metric that compared the relatedness of the entire response. The score uses the phylogenetic distance between all combination of sequences in the response to output a mapped phylogenetic distance score between 0 and 1, where a score of 1 would correspond to all sequences in response being identical and 0 would correspond to all sequences being infinitely different.

Sequencing Data (Sequence, Frequency, VDJ Usage)

$$\text{Score}_{12} = \text{Sequence Alignment Score (Sequence 1, Sequence 2)}$$
$$\text{Score}_{11} = \text{Sequence Alignment Score (Sequence 1, Sequence 1)}$$
$$\text{Score}_{22} = \text{Sequence Alignment Score (Sequence 2, Sequence 2)}$$

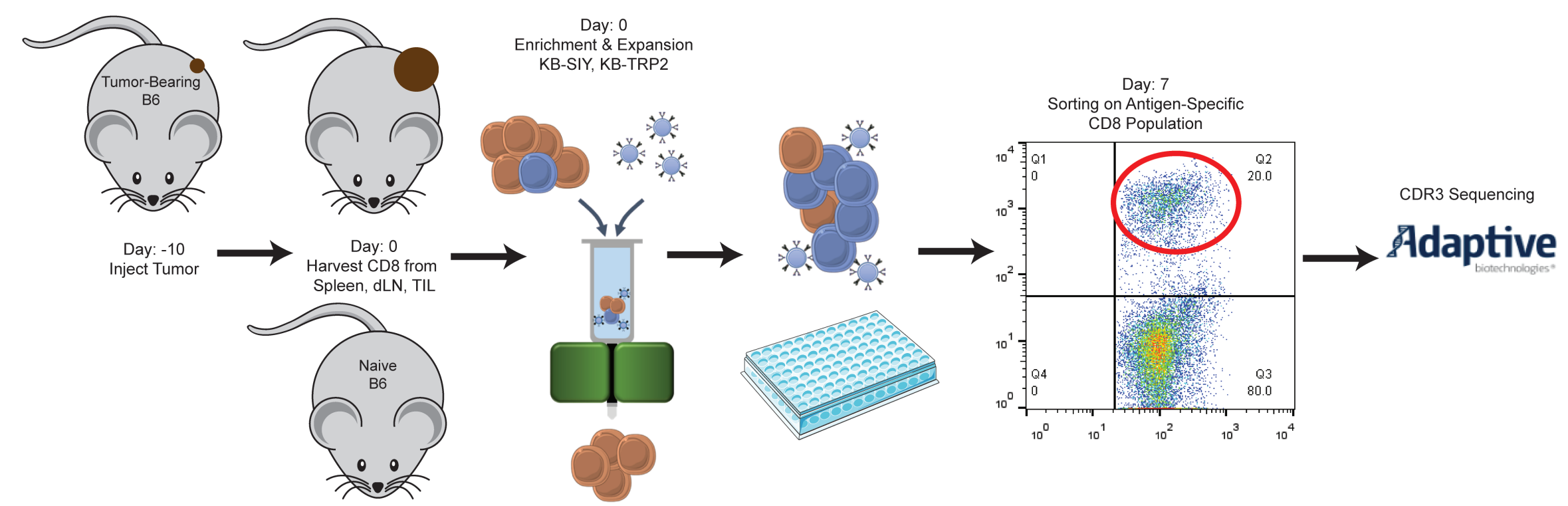
$$\text{Phylogenetic Distance} = \left(1 - \frac{\text{Score}_{12}}{\text{Score}_{11}}\right) \left(1 - \frac{\text{Score}_{22}}{\text{Score}_{22}}\right)$$

CSARDGTGNVGT CASGGDTLYF
| | | | | | | | | | | |
CASGGDTLYF-F CASGGNTLYF
0.2481 **0.9884**

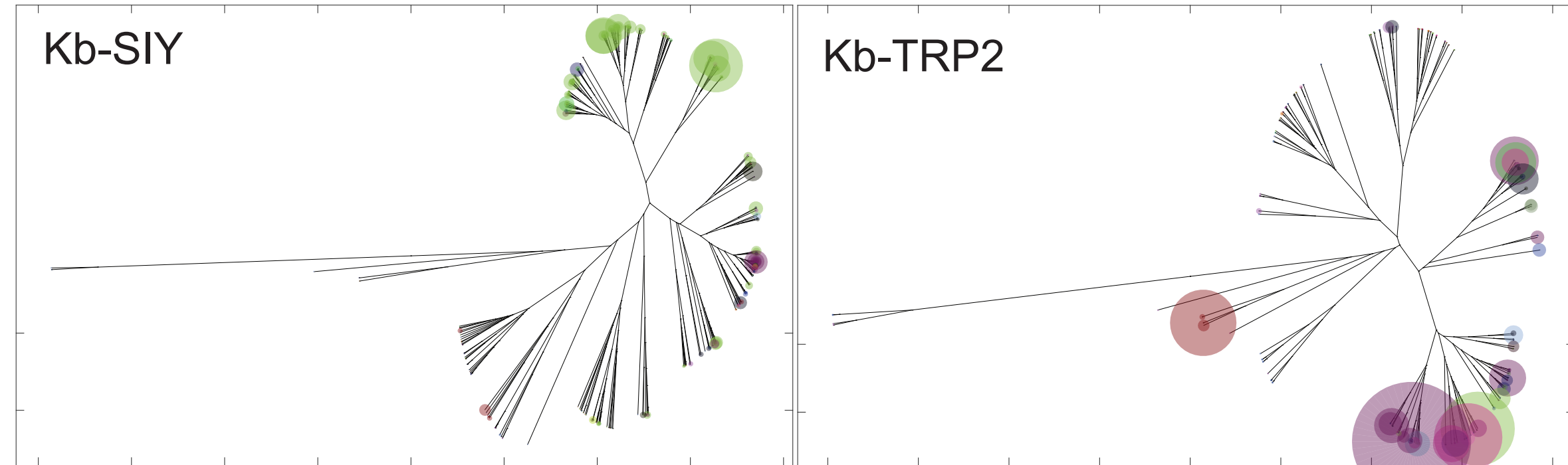


Antigen-Specific Responses in Naive and Tumor-Bearing Settings

Enrichment + Expansion of Antigen-Specific CD8 (Kb-SIY: foreign antigen, Kb-TRP2: self-antigen)



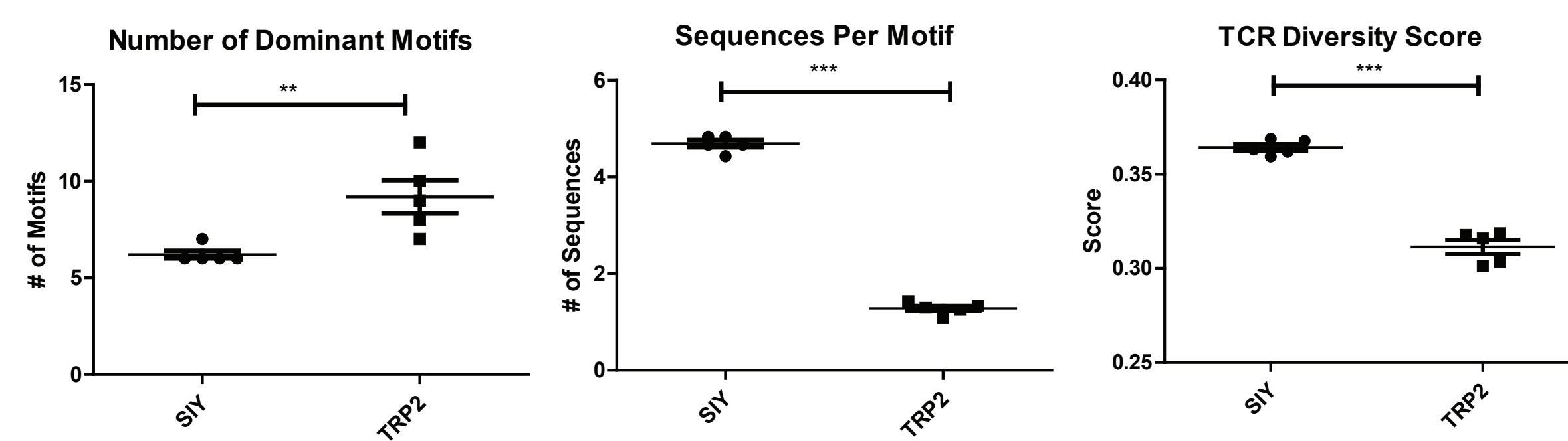
Weighted Phylogenetic Trees - Naive Response



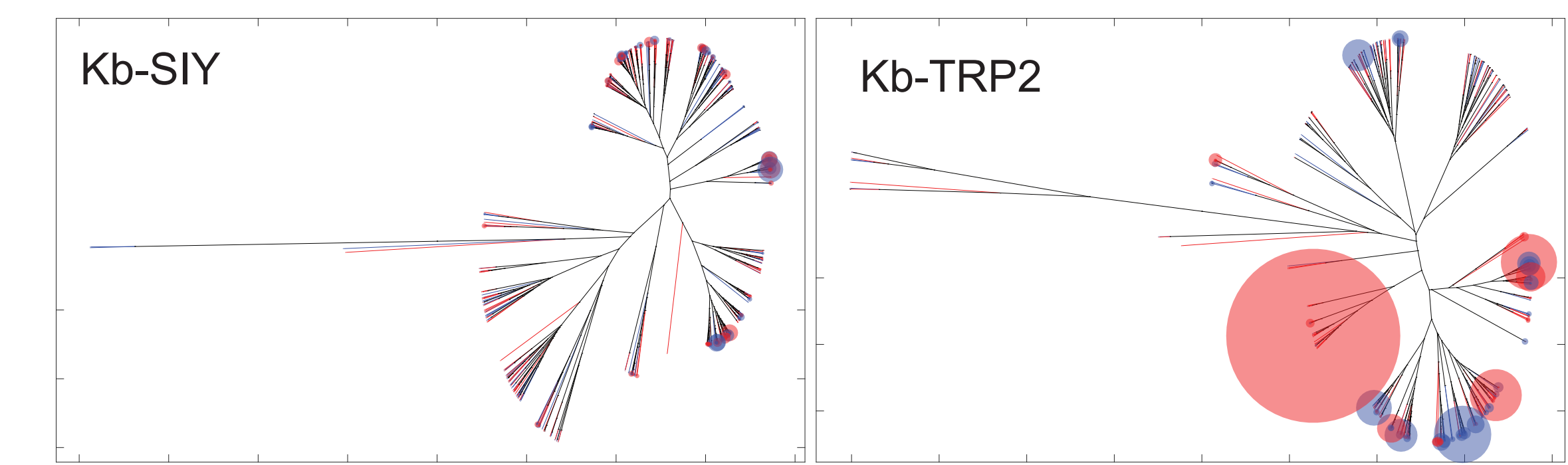
Dominant Motif Analysis - Naive Response

Kb-SIY Contribution to Response (%)	Motif						Kb-TRP2 Contribution to Response (%)	Motif						
	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6		Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6	
3.37	10.65	7.83	4.46	4.67	3.50	10.65	5.60	9.50	3.07	5.66	6.10			

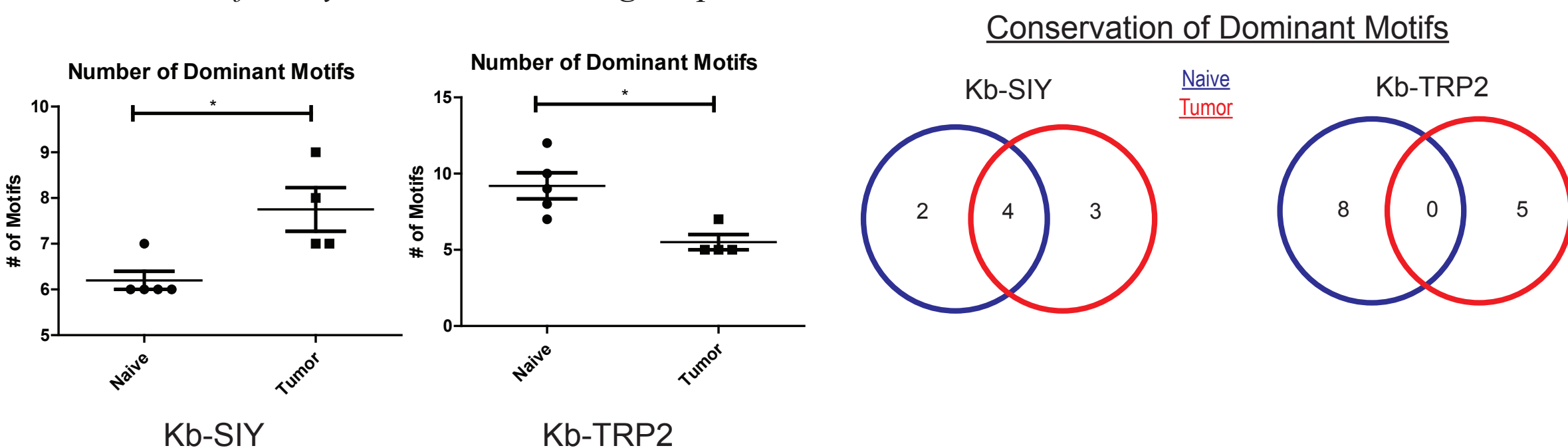
Kb-SIY Contribution to Response (%)	Motif		Kb-TRP2 Contribution to Response (%)	Motif		
	Motif 7	Motif 8		Motif 7	Motif 8	
	3.07	4.05	3.07	4.05		



Weighted Phylogenetic Trees - Tumor-Bearing Response (Blue = Naive, Red = Tumor-Bearing)

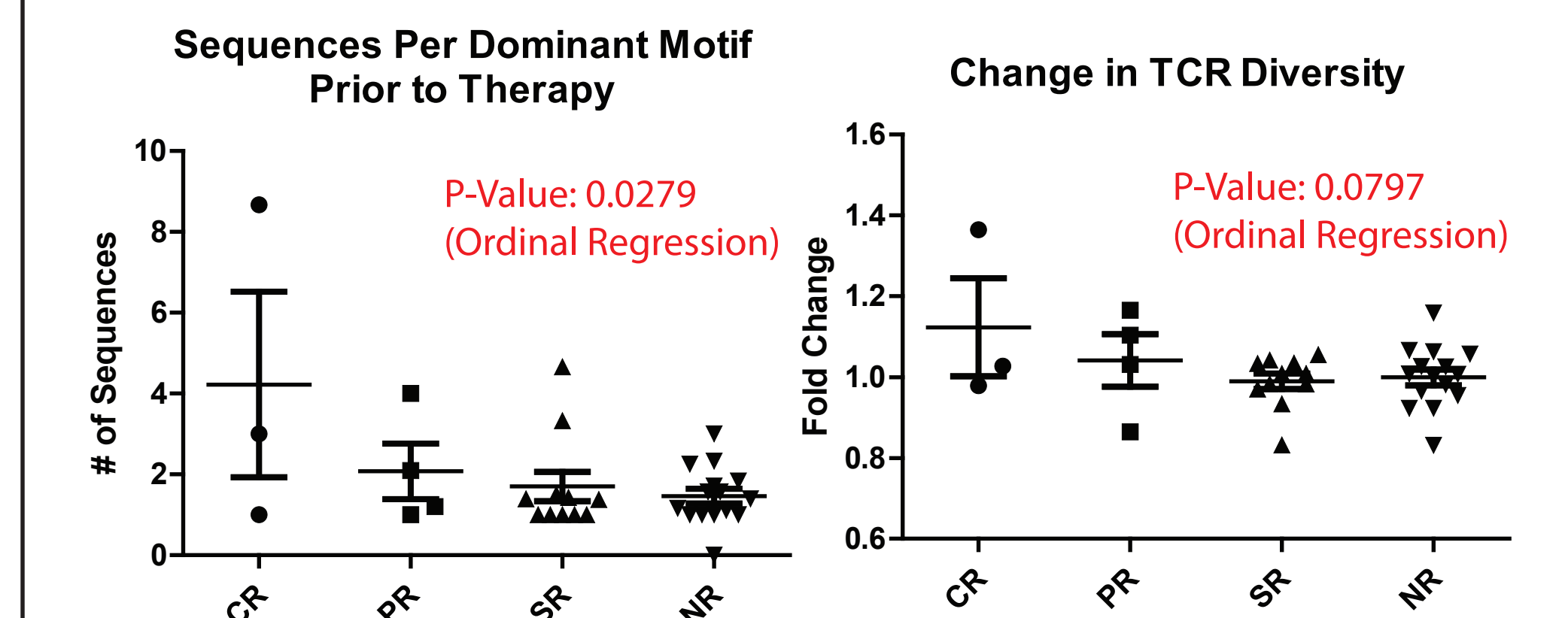
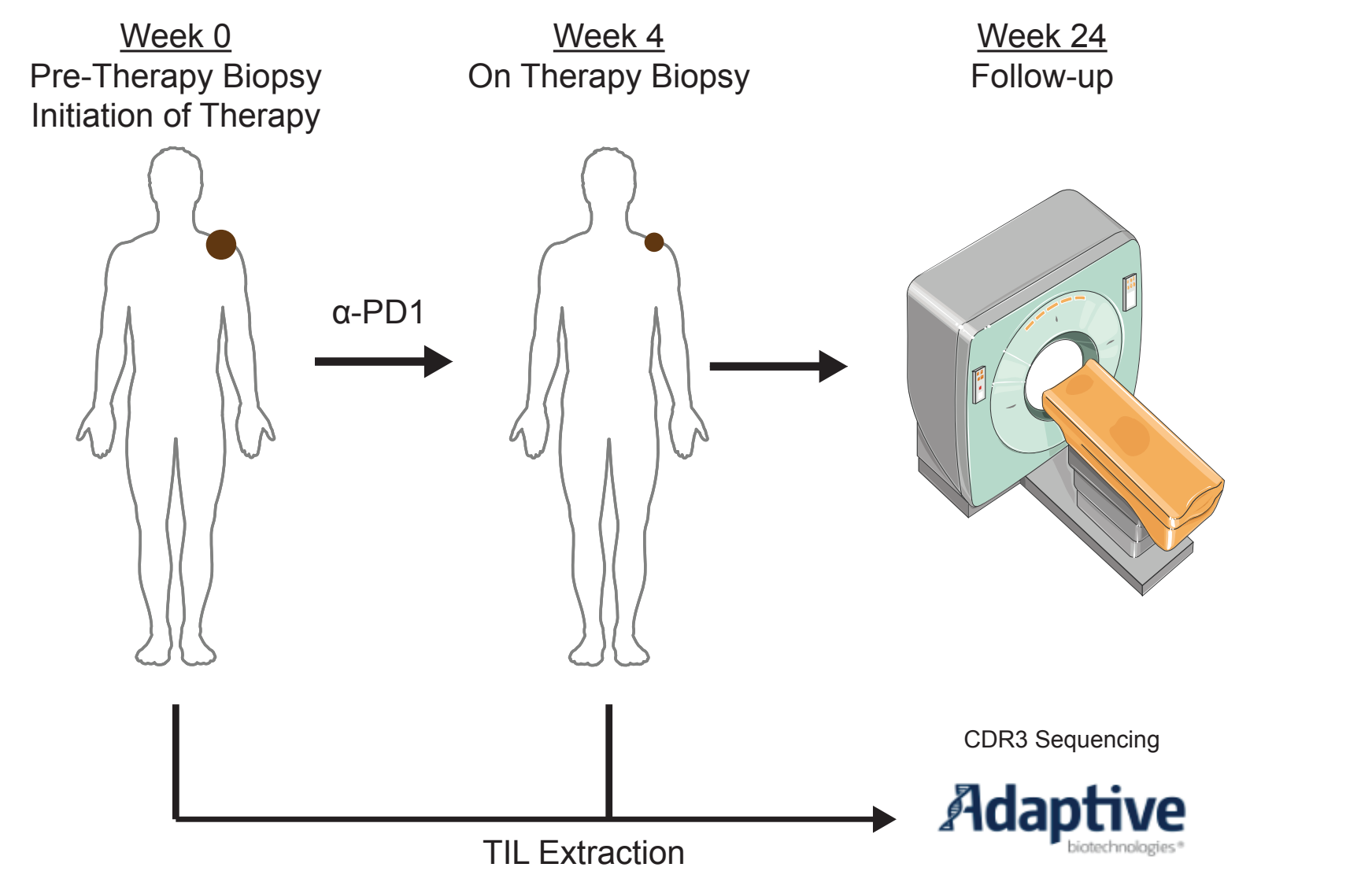


Dominant Motif Analysis - Tumor-Bearing Response



Results

Analysis of Tumor-Infiltrating Lymphocytes from Melanoma Pts. Undergoing anti-PD1



Conclusions

Here we have presented a novel bioinformatics approach to analyze TCR repertoire sequence data in order to elucidate novel understandings of antigen-specific CD8 responses and characteristics that differentiate responders to checkpoint blockade. Combining information about sequence homology and frequency has allowed a more intuitive appreciation of TCR repertoire characteristics through the use of these novel weighted phylogenetic trees. Furthermore, examining homologous sequences as opposed to identical sequences captures an understanding that reconciles the structure and function of the repertoire.

Applying this novel approach to understanding antigen-specific responses has yielded a novel understanding of the diversity of the endogenous repertoire to foreign vs self antigens as well as demonstrated differential effects of immune pressure in the presence of tumor. In clinical settings, ImmunoMAP was able to identify key repertoire characteristics that differentiated responders from non-responders on anti-PD1 therapy.

ImmunoMap not only has the potential to have clinical implications of being able to monitor patients closely on therapy, knowing their likelihood to respond, but this type of analysis can be crucial in understanding the basic science mechanisms of response and how to alter current immune therapies to treat a larger cohort of patients.

Acknowledgements

Jonathan Schneck Lab
Timothy Chan Lab
Biomedical Engineering Training Program
Supported By:
SR01CA108835