Supplementary Data

Evaluating T cell cross-reactivity between tumors and immune-related adverse events with TCR sequencing: pitfalls in interpretations of functional relevance

Cottrell and Zhang, *et. al.*

Table S1. Specimen descriptions and TCR sequencing summary

ID	Description	Location	Time from diagnosis	Procedure	Treatment status	Total TCR Reads*	Productive Rearrangements*	Total Included Reads**	# Nucleotide Clones	# Amino Acid Clones
T_P	Primary tumor	Right kidney	0	Resection	Pretreatment	12637	4195	12588	3908	3843
irAE	Lichenoid dermatitis (irAE)	Right calf skin	9 months	Biopsy	On therapy	4194	1034	4147	3184	3159
T_{M1}	Metastasis	Brain	28 months	Autopsy	Discontinued	1011	797	1008	777	775
T_{M2}	Metastasis	Jejunum	28 months	Autopsy	Discontinued	6969	3493	6945	3315	3279
T_{M3}	Metastasis	Mesentery	28 months	Autopsy	Discontinued	1364	1109	1360	1074	1070
LN1	Mediastinal lymph node	Paratracheal	28 months	Autopsy	Discontinued	12061	10999	12001	10680	10568
LN2	Mediastinal lymph node	Carinal	28 months	Autopsy	Discontinued	34166	29881	34058	29229	28861
LN3	Mediastinal lymph node	Mediastinal	28 months	Autopsy	Discontinued	15895	14353	15842	13901	13764
NK	Normal	Left kidney	28 months	Autopsy	Discontinued	1387	976	1381	949	944
NSB	Normal	Small bowel	28 months	Autopsy	Discontinued	30454	6702	30397	6229	6096
SK	Inflamed skin lesion	Skin	28 months	Autopsy	Discontinued	10522	5770	10493	5529	5470

*From Adaptive Biotechnologies

**Following additional pre-processing to remove non-productive reads

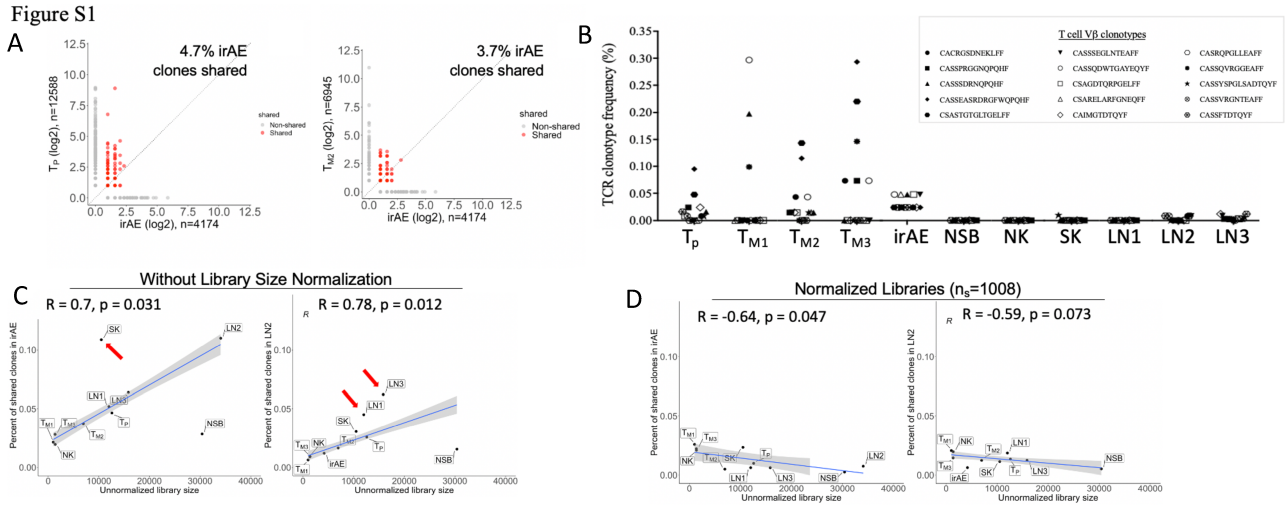


Figure S1. Library size confounding in hypothesis testing with TCR sequencing data. (A) Pairwise comparison demonstrates overlapping T cell clones present in the irAE and primary tumor (4.7% of irAE clones). To focus on T cells recognizing the same antigen, amino acid clonotypes were analyzed exclusively. Each dot represents a unique TCR amino acid sequence, with non-shared clones (gray dots) plotted along the axis of the corresponding specimen and shared clones (red dots) plotted centrally as a function of frequency in each specimen. The number of total TCR sequencing reads is listed for each specimen (n). **(B)** Pairwise comparison demonstrates overlapping T cell clones present in the irAE and one of the progressing metastases (T_{M2}) at the time of autopsy (3.7% of irAE clones). **(C)** Across all specimens, 127 T cell clones were identified in the irAE and at least one tumor specimen(s) and absent in the non-lymphoid normal tissues. **(D)** *Library size confounding* is illustrated by the correlation between library size and the number of shared T cell clones across multiple specimens, illustrated for both the irAE (left) and a lymph node (LN2, right). Notable outliers include specimens from the same tissue site (the seborrheic keratosis and lymph node specimens for the irAE and LN, respectively, red arrows), which show greater sharing than predicted by library size alone. In contrast, the NSB is an outlier that shows less sharing than expected based on the library size for both the irAE and LN2. (**E)** When the library sizes are normalized by down-sampling to the same read count (n_s=1008, with sharing plotted as a function of the original read counts), this positive correlation is eliminated for both the irAE (left) and LN2 (right).

Figure S2

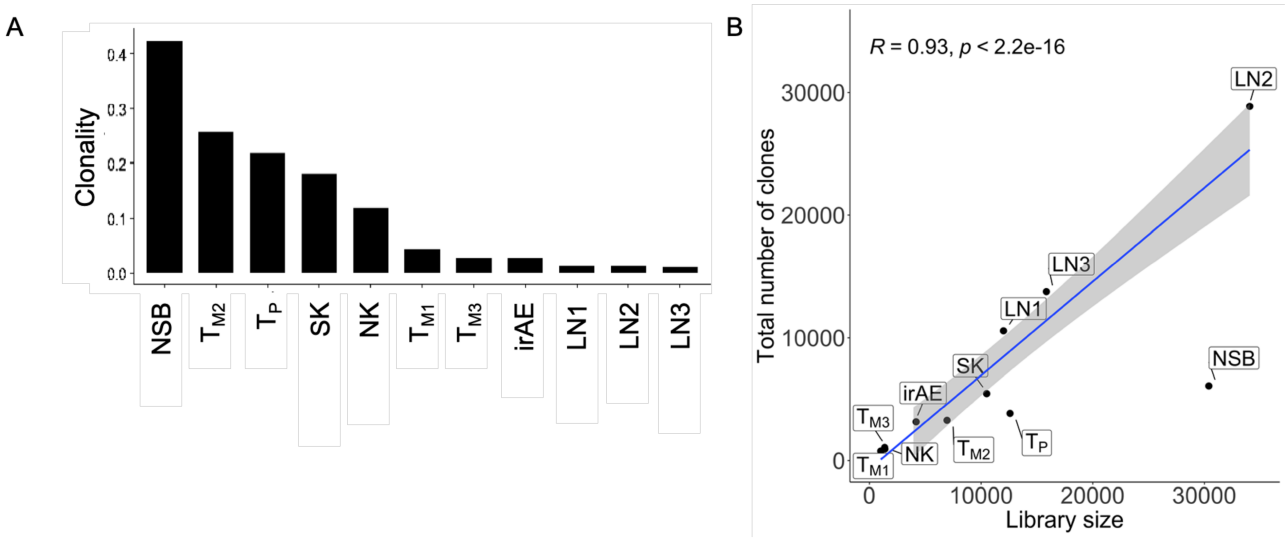


Figure S2. Specimen clonality and the correlation between library size and the number of unique clonotypes. (A) Productive clonality as determined by using the productive amino acid (AA) sequence of the CDR3. **(B)** Correlation of library size and total number of clonotypes. The blue line indicates the robust linear regression line (fitted using R function 'rlm' from 'MASS' package based on M estimator), and the gray area indicates the upper and lower boundaries of the 95% confidence interval.

Figure S3

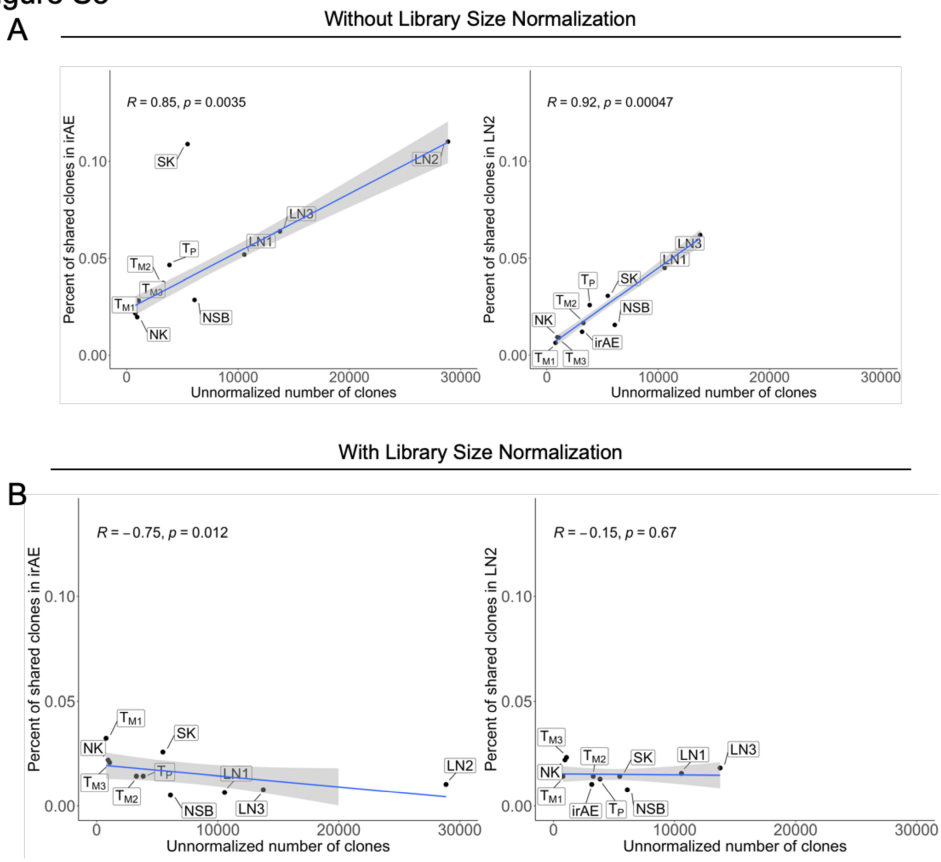


Figure S3. Effects of library size normalization. **(A)** T cell repertoire sharing is highly correlated with the clonal diversity of the specimens, shown for the irAE (left) and LN2 (right). **(B)** Normalizing for the number of clones in each specimen removes this correlation.

Figure S4

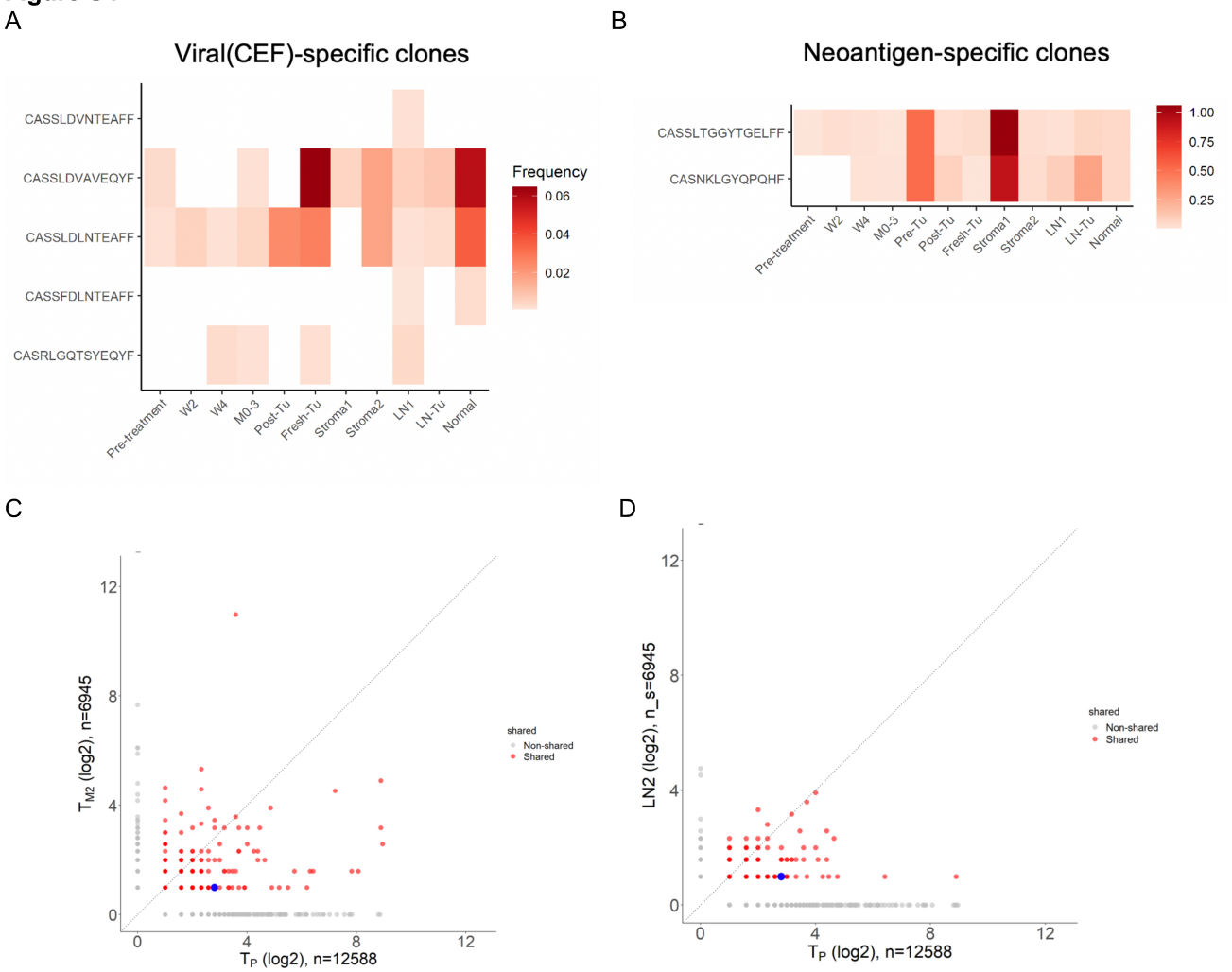


Figure S4. Clonal tracking of antigen-specific T cells. Clonal tracking of CMV, EBV, and flu (CEF)-specific clones **(A)** and neoantigen-specific clones **(B)** across compartments in a non-small cell lung cancer patient treated with anti-PD1 (assay published previously in Forde PM, et al, *N Engl J Med*, 2018 and Danilova et al, *Cancer Immunol Res*, 2018). **(C) and (D)** Clonal tracking of a EBV-specific clone in tumor involved tissues in the patient from our study. The EBV-specific clone CATGTGDSNQPQHF, highlighted in blue, was reported by two independent studies to be EBV-specific in vdjDB database). Clones shared in two specimens were colored in red; non-shared clones were colored as grey.

Figure S5

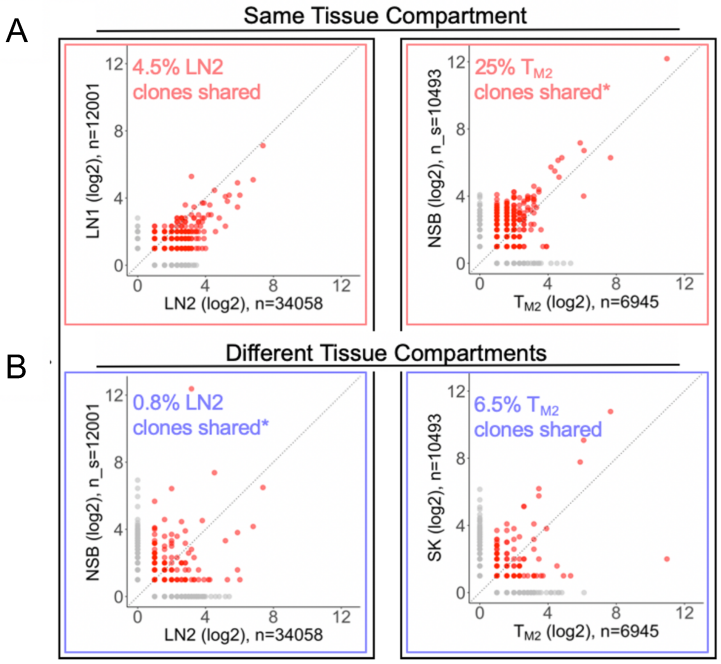
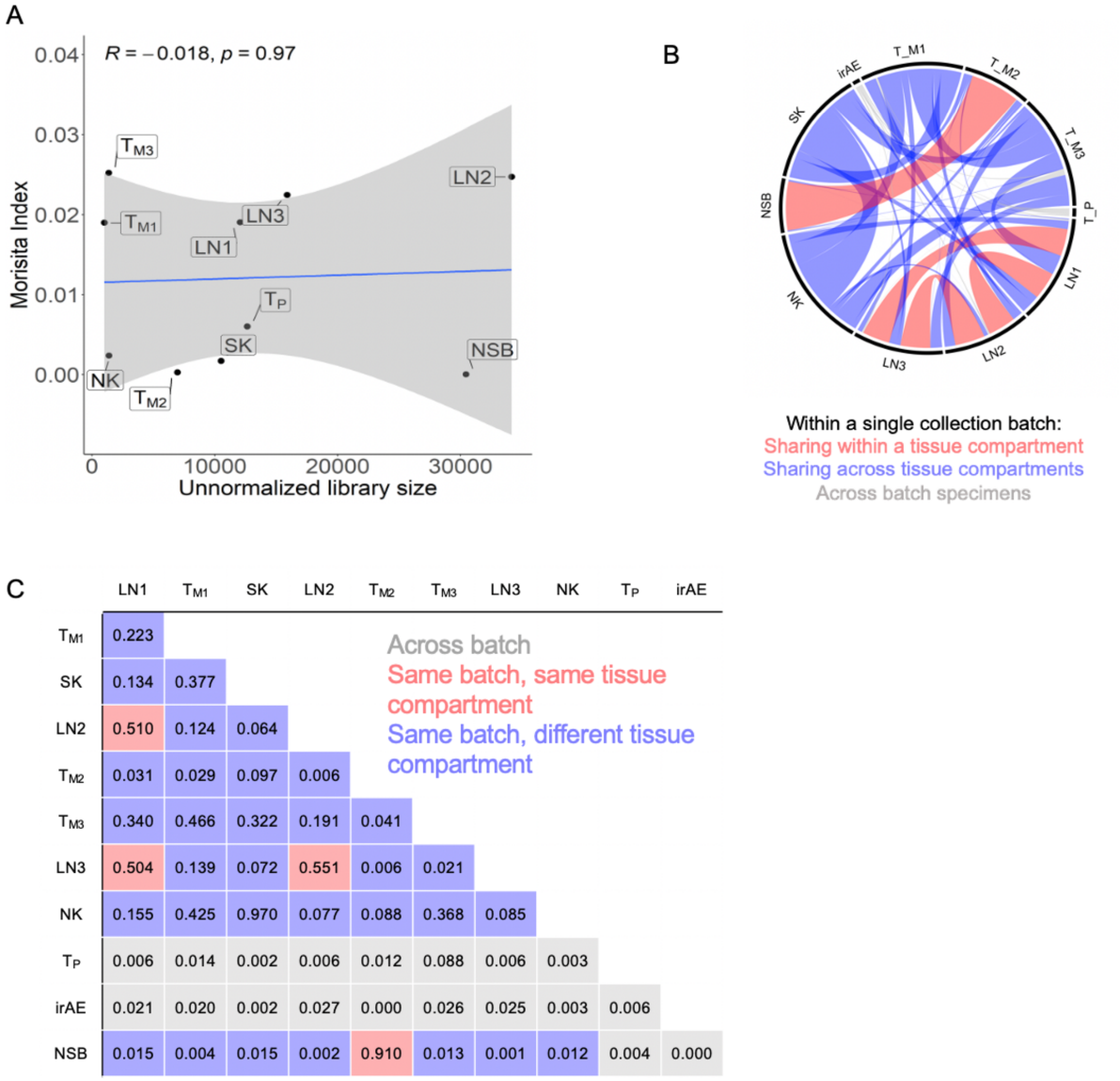


Figure S5 (A) Tissue compartment confounding in hypothesis testing with TCR sequencing data. Following library size normalization (down-sampled specimen reads listed as n_s), significant T cell repertoire overlap can also be identified by comparing samples from the same tissue compartment, shown for two lymph nodes (LN1 and LN2, top left) and two small bowel specimens (NSB and T_{M2}, top right). The percent of clonotypes shared is shown for each comparison. Subsampled comparisons are indicated (*) and the 95% CI for shared clones between NSB and TM2 is 24-26%. (**B**) In contrast, less overlap is seen between specimens from distinct tissue compartments, as shown for LN2 and NSB (bottom left) and skin (SK) and T_{M2} (bottom right). The percent of clonotypes shared is shown for each comparison. Subsampled comparisons are indicated (*) and the 95% CI for shared clones between NSB and LN2 is 0.7-0.9%. No confidence interval is calculated for comparison of non-down-sampled specimens (e.g., SK and T_{M2}).

Figure S6



	LN1	T_{M1}	SK	LN2	T_{M2}	T_{M3}	LN3	NK	T_P	irAE
T_{M1}	0.223									
SK	0.134	0.377								
LN2	0.510	0.124	0.064							
T_{M2}	0.031	0.029	0.097	0.006						
T_{M3}	0.340	0.466	0.322	0.191	0.041					
LN3	0.504	0.139	0.072	0.551	0.006	0.021				
NK	0.155	0.425	0.970	0.077	0.088	0.368	0.085			
T_P	0.006	0.014	0.002	0.006	0.012	0.088	0.006	0.003		
irAE	0.021	0.020	0.002	0.027	0.000	0.026	0.025	0.003	0.006	
NSB	0.015	0.004	0.015	0.002	0.910	0.013	0.001	0.012	0.004	0.000

Figure S6. Batch effect confounding in hypothesis testing with TCR sequencing data. (A) The Morisita overlap index is calculated between the irAE and all other specimens and plotted as a function of library size (total reads). There is no significant correlation observed between the MI and library size. **(B)** The chord diagram illustrates the large number of T cell clones shared among specimens collected at the same timepoint. The width of the bands corresponds to the Morisita Index between the two specimens. Red bands illustrate sharing between specimens from the same tissue compartment and blue bands show sharing across different tissue compartments. The two specimens collected at different timepoints (i.e., different batches, irAE and Tp) are shown in gray. **(C)** The individual pairwise MI calculations are shown for all specimens included in the analysis. Specimens are color coded by batch of collection and tissue compartment to highlight batch and tissue compartment confounding, with the highest MI observed between specimens collected in the same batch and from the same tissue compartment (red).

Figure S7

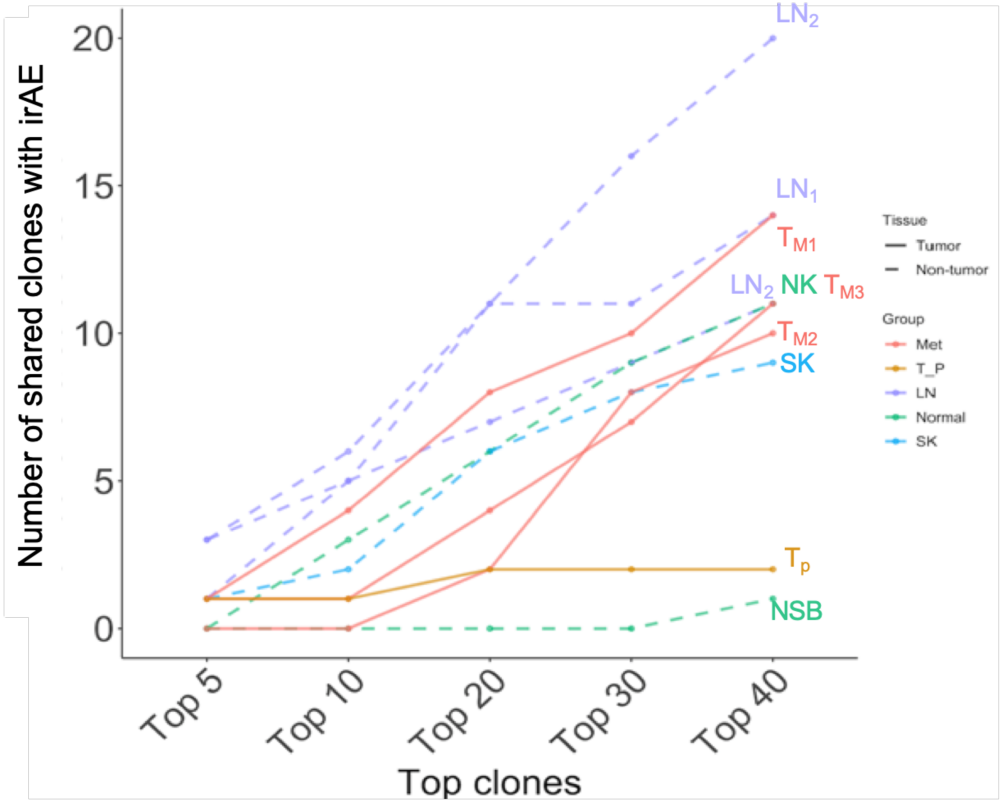


Figure S7. Overlap of the irAE T cell repertoire with highly abundant clones in tumor vs. non-tumor specimens. The number of clones shared with the irAE (y-axis) is shown for the top x clones (x-axis) for each specimen "group". Tumor tissues are indicated with a solid line and non-tumor tissues are indicated with a dotted line. No increase in clonal overlap with the irAE is observed when comparing the most abundant clones in tumor and non-tumor specimens. Sample size is normalized in this analysis by including discrete numbers of unique clones from each specimen (i.e. the top 5, 10, 20, 30 and 40 unique clones).

Figure S8

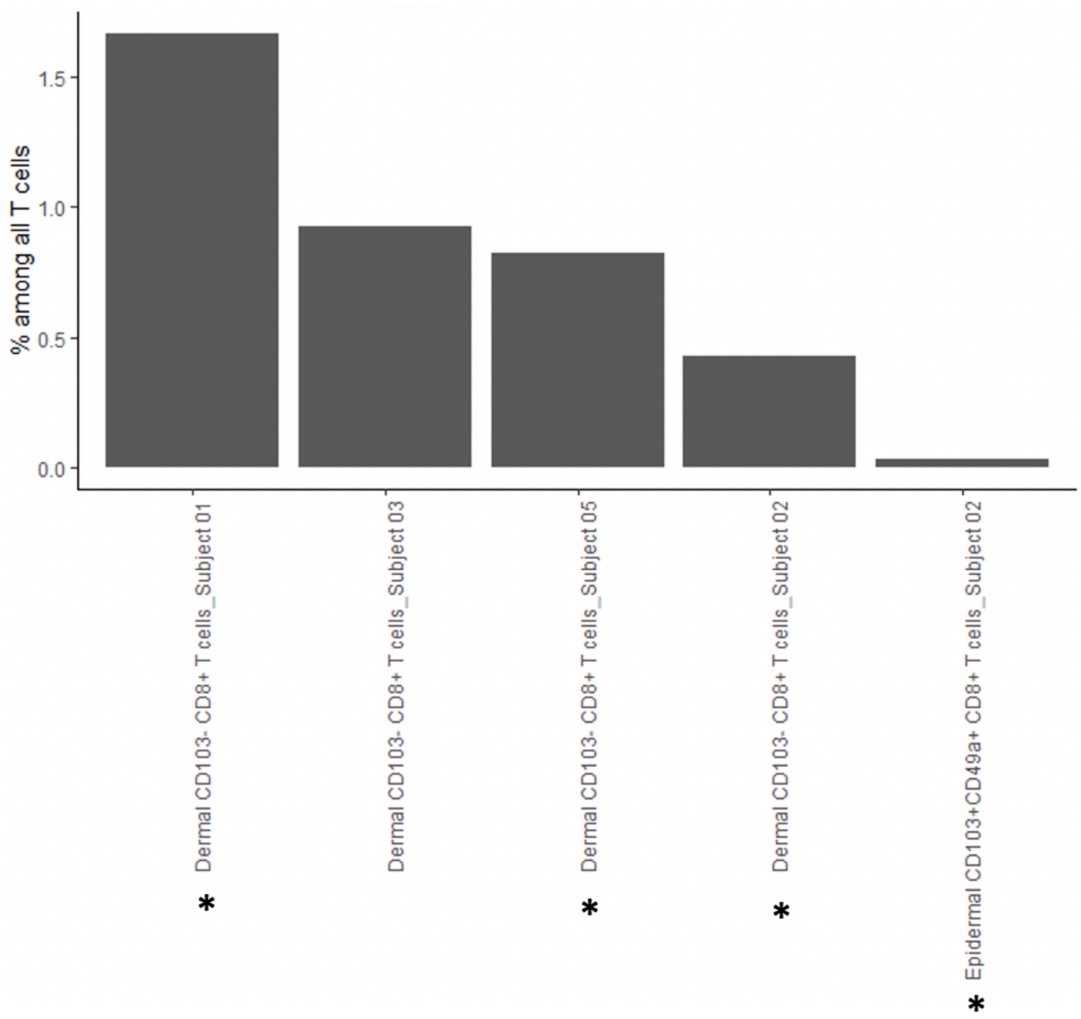


Figure S8. Cumulative abundance of dermal/epidermal clonotypes with the "SSQD" motif in skin from four healthy individuals. Samples with at least 1000 reads were included in the analysis. * individuals having at least one shared clonotypes with the irAE patient in our cohort.

Figure S9

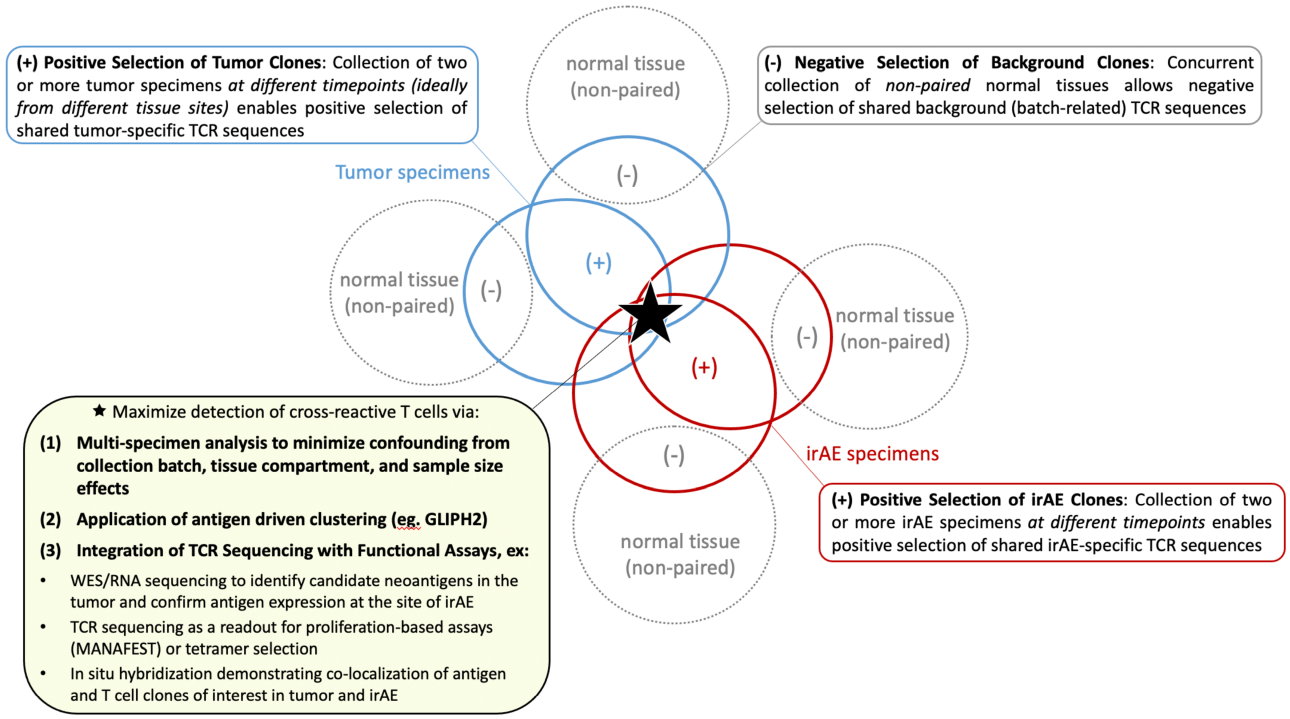


Figure S9. Recommendations for study design and analysis strategies for TCR sequencing-based assessment of cross-reactivity. Appropriate specimen collection enables negative selection of background (batch-effect) clones and positive selection of disease-related clones to enrich for functionally relevant TCR repertoire signatures. It is optimal to collect specimens used for signature enrichment (e.g., tumor-normal or tumor-tumor specimens) from different tissue sites of origin to minimize tissue compartment confounding. Given the practical limitations likely restricting such a specimen collection protocol, the integration of antigen-driven clustering and functional assays to identify tumor reactive T cell clones is a more robust approach for identifying disease relevant T cell signatures for the analysis of multi-specimen TCR sequencing data.