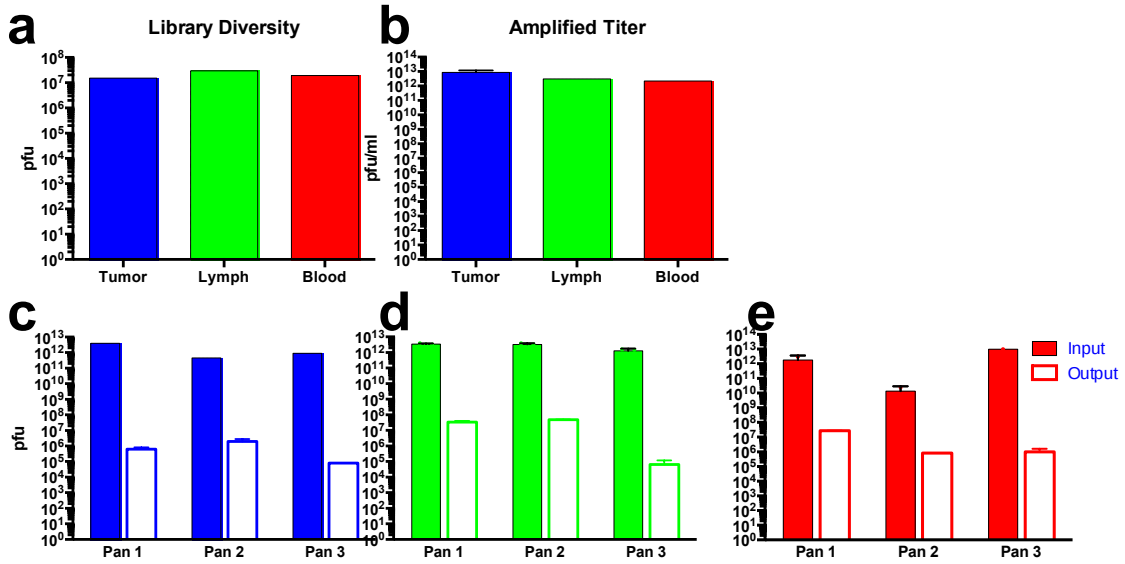
Cancer Immunology, Immunotherapy (submitted in 2014) -- Leah Novinger et al.
Supplemental Figures and Tables



Supplemental Figure 1. Comparison of three scFv libraries derived from the same breast cancer patient. Library diversity **(a)** and amplified size **(b)** were compared for tumor (blue), axillary lymph node (green), and blood (red) libraries created from the same breast cancer patient. Library diversity was based on the overall number of transformants. Library size was the biological titer of the library calculated following amplification. Input (solid bars) and output phage(open bars) from each round of panning are shown for tumor, lymph node, and blood libraries **(c-e)**.

Supplemental Table 1. Summary of pre-enrichment scFv sequenced for heavy chain genetic analysis. Only unique sequences were included in the analysis; duplicate, short, poor quality, unidentifiable (no results) and unknown sequences were removed from the analysis.

Heavy Chain Sequences							
Source	Used in Analysis	Discard					Total
		Duplicate	Short	Poor Sequence	No Results	Unknown	
Tumor	35	0	0	7	4	1	47
LN	56	5	0	16	26	2	105
Blood	45	5	0	2	8	4	64

Supplemental Table 2. Summary of pre-enrichment scFv sequenced for light chain genetic analysis. Only unique sequences were included in the analysis; duplicate, short, poor quality, unidentifiable (no results) and unknown sequences were removed from the analysis.

Light Chain Sequences							
Source	Used in Analysis	Discard					Total
		Duplicate	Short	Poor Sequence	No Results	Unknown	
Tumor	36	1	0	2	1	2	42
LN	72	1	0	10	10	0	93
Blood	44	2	0	12	1	0	59

Supplemental Table 3. Summary of post-enrichment scFv sequenced for heavy chain genetic analysis. Only unique sequences were included in the analysis; duplicate, short, poor quality, unidentifiable (no results) and unknown sequences were removed from the analysis.

Heavy Chain Sequences							
Source	Used in Analysis	Discard					Total
		Duplicate	Short	Poor Sequence	No Results	Unknown	
Tumor	27	3	112	17	20	11	190
LN	94	7	31	18	36	4	190
Blood	3	85	1	0	4	0	93

Supplemental Table 4. Summary of post-enrichment scFv sequenced for light chain genetic analysis. Only unique sequences were included in the analysis; duplicate, short, poor quality, unidentifiable (no results) and unknown sequences were removed from the analysis.

Light Chain Sequences							
Source	Used in Analysis	Discard					Total
		Duplicate	Short	Poor Sequence	No Results	Unknown	
Tumor	31	3	110	15	20	11	190
LN	131	7	5	11	32	4	190
Blood	4	85	0	0	4	0	93

Supplemental Table 5. Summary of heavy chain variable gene segment frequency in tumor, lymph node, or blood libraries. **(a)** Frequency of each gene segment in the library based on total number of gene segments analyzed (indicated in first column for each library). The frequency is reported with the exact confidence interval. **(b)** Overall P values for each segment are reported. If the overall p value < 0.05, the exact p values for pairwise comparisons are shown under the overall p value.

a. Source	VH1	VH2	VH3	VH4	VH5	VH6
Tumor n=35	14% (5% - 30%)	0% (0%-10%)	63% (45%-79%)	23% (10%-40%)	0% (0% - 10%)	0% (0% - 10%)
Lymph Node n=56	9% (3% - 20%)	0% (0% - 6%)	54% (40%-67%)	25% (14%-38%)	11% (4%-22%)	2% (0%-10%)
Blood n=45	13% (5% - 7%)	0% (0% - 8%)	78% (63%-89%)	9% (2%-21%)	0% (0% - 8%)	0% (0% - 8%)

b. Statistical Analysis	VH1	VH2	VH3	VH4	VH5	VH6
Overall p-value	0.6550	1.000	0.0432	0.1052	0.0095	1.0000
T vs LN			0.5140		0.0784	
LN vs B			0.0131		0.0320	
B vs T			0.2128		1.0000	

Supplemental Table 6. Summary of heavy chain joining gene segment frequency in tumor (T), lymph node (LN), or blood (B) libraries. **(a)** Frequency of each gene segment in the library based on total number of gene segments analyzed (indicated in first column for each library). The frequency is reported with the exact confidence interval. **(b)** Overall P values for each segment are reported. If the overall p value < 0.05, the exact p values for pairwise comparisons are shown under the overall p value.

a. Source	JH1	JH2	JH3	JH4	JH5	JH6
Tumor n=35	9% (2%-23%)	3% (0%-15%)	26% (14%-43%)	43% (26%-61%)	6% (1%-19%)	14% (5%-30%)
Lymph Node n=56	0% (0% - 6%)	2% (0%-10%)	34% (22%-48%)	27% (16%-40%)	9% (3% -0%)	29% (17%-42%)
Blood n=45	0% (0% - 8%)	0% (0% - 8%)	31% (18%-47%)	53% (38%-68%)	4% (0%-15%)	11% (3%-24%)

b. Statistical Analysis	JH1	JH2	JH3	JH4	JH5	JH6
Overall p-value	0.0160	0.7255	0.7294	0.0251	0.6859	0.0565
T vs LN	0.0539			0.1685		
LN vs B	1.0000			0.3766		
B vs T	0.0797			0.0079		

Supplemental Table 7. Summary of κ light chain variable gene segment frequency in tumor (T), lymph node (LN), or blood (B) libraries. **(a)** Frequency of each gene segment in the library based on total number of gene segments analyzed (indicated in first column for each library). The frequency is reported with the exact confidence interval. **(b)** Overall P values for each segment are reported. No significant differences in gene frequency were identified among these populations.

a. Source	Vκ1	Vκ2	Vκ3	Vκ4	Vκ5
Tumor n=24	33% (16% - 55%)	0% (0% - 14%)	17% (5% - 37%)	50% (29% - 71%)	0% (0% - 14%)
Lymph Node n=39	23% (11% - 40%)	0% (0% - 9%)	38% (23% - 55%)	38% (23% - 55%)	0% (0% - 9%)
Blood n=38	21% (10% - 37%)	5% (0% - 18%)	34% (18% - 49%)	37% (22% - 54%)	3% (0% - 14%)

b. Statistic Analysis	Vκ1	Vκ2	Vκ3	Vκ4	Vκ5
Overall p-value	0.5672	0.1939	0.1760	0.5571	0.6139

Supplemental Table 8. Summary of λ light chain variable gene segment frequency in tumor (T), lymph node (LN), or blood (B) libraries. **(a)** Frequency of each gene segment in the library based on total number of gene segments analyzed (indicated in first column for each library). The frequency is reported with the exact confidence interval. **(b)** Overall P values for each segment are reported. No significant differences in gene frequency were identified among these populations.

a. Source	Vλ1	Vλ2	Vλ3	Vλ4	Vλ6	Vλ8	Vλ9
Tumor n=12	17% (2% - 48%)	25% (5% - 57%)	25% (5% - 57%)	17% (2% - 48%)	8% (0% - 38%)	0% (0% - 26%)	8% (0% - 38%)
Lymph Node n=33	27% (13% - 46%)	36% (20% - 55%)	27% (13% - 46%)	0% (0%- 11%)	6% (0%- 20%)	3% (0%- 16%)	0% (0%- 11%)
Blood N=6	50% (12%- 88%)	33% (4%- 78%)	0% (0%- 46%)	0% (0%- 46%)	0% (0%- 46%)	0% (0%- 46%)	17% (0%- 64%)

b. Statistical Analysis	Vλ1	Vλ2	Vλ3	Vλ4	Vλ6	Vλ8	Vλ9
Overall p-value	0.3709	0.9027	0.4812	0.0635	1.0000	1.0000	1.0000

Supplemental Table 9. Distribution of nonclonal independent and clonal heavy chain sequences in each library. This table shows in tabular form the number of individual heavy chains shown in the Venn diagram in Figure 2. Clonal heavy chains shared a junctional region with another heavy chain, whereas nonclonal independent heavy chains did not.

	Tumor Library	Lymph Node Library	Blood Library
Nonclonal Independent Heavy Chains	29 (72.5%)	45 (71.4%)	31 (68.9%)
Clonal Heavy Chains	11(27.5%)	18 (28.6%)	14 (31.1%)

Supplemental Table 10. Distribution of nonclonal independent and clonal light chain sequences in each library. This table shows in tabular form the numbers of individual light chains shown in the Venn diagram in Figure 3. Clonal light chains shared a junctional region with another light chain, whereas nonclonal light chains did not.

	Tumor library	Lymph node library	Blood library
Nonclonal Independent Light Chains	28 (73.7%)	49 (75.4%)	39 (90.7%)
Clonal Light Chains	10 (26.3%)	16 (24.6%)	14 (9.3%)

Supplemental Table 11. Number of groups of clones with heavy chain sequences shared between libraries. This table shows in tabular form the clonal groups shown in the overlapping circle portions of the Venn diagram in Figure 2. The observed number of groups of heavy chain sequences shared between the tumor and lymph node libraries was significantly more than expected by chance using a chi-square goodness of fit test [chi square (df=2) = 12.00, exact p-value and point probability = 0.004].

	Tumor and lymph node libraries	Tumor and blood libraries	Lymph node and blood libraries
Observed	6	0	0
Expected	2	2	2

Supplemental Table 12. Number of groups of clones with light chainsequences shared between libraries. This table shows in tabular form the clonal groups shown in the Venn diagram in Figure 3. The observed number of groups of light chain sequences shared between the tumor and lymph node libraries (first row) was significantly more than expected by chance (second row) using a chi-square goodness of fit test [chi square (df=2) = 9.997, exact p-value and point probability = 0.01].

	Tumor and lymph node libraries	Tumor and blood libraries	Lymph node and blood libraries
Observed	5	0	0
Expected	1.67	1.67	1.67

Supplemental Table 13. Clonal heavy and light chain groups identified among scFv selected from each library after enrichment for binding to autologous tumor lysates.

Type of Group	Tumor Library	Lymph Node	Blood Library
Heavy Chain	0	9	0
Light Chain	1	9	0

Supplemental Table 14. Clonal heavy and light chain groups shared between scFv from different libraries after enrichment for binding to autologous tumor lysates.

Type of Group	Tumor and Lymph Node Libraries	Tumor and Blood Libraries	Lymph Node and Blood Libraries
Heavy chain	5	0	0
Light chain	5	0	0