Supplemental Figure 1. Error rate estimation. We used high frequency clonal expansions to derive an error rate based on how many sequences should be considered erroneous. Because our sequences are approximately all the same length this is a reasonable procedure. We show two examples with a benchmark sequence and those that are considered errors. The type of error is noted as well as the position of the error(s). The count of high frequency clonotypes was not corrected; only the erroneous sequences removed. Under the assumption that 1 of every 50 sequences is incorrect. Every fifty singletons (clonotype observed once) could have one erroneous counterpart. The same would hold for every 25 doubletons, etc.

Supplemental Figure 2. Relation between CDR3 and NDN length. The relation of CDR3 length (column) to NDN length (row) is shown for A. Clonotypes, and B. Observations. The data are shaded to represent relative frequency of CDR3 length for each NDN length. The sum for NDN length is given in the bottom row. The sum for each CDR3 length is given in the right-most column.

Supplemental Figure 3. Amino acid usage based on NDN start position. The amino acid utilization is shown for clonotypes whose NDN starts at: A. CDR3 position 2, B CDR3 position 3, C. CDR position 4, and D. CDR3 position 5. The data are shown as relative frequency of each amino acid for each start. The data values are shaded to show relative frequency. Amino acids that are V encoded are boxed. The start column for each panel was also shown in Figure 5B. The amino acids are identified in the left-most column and are in the same order as in Figure 5.

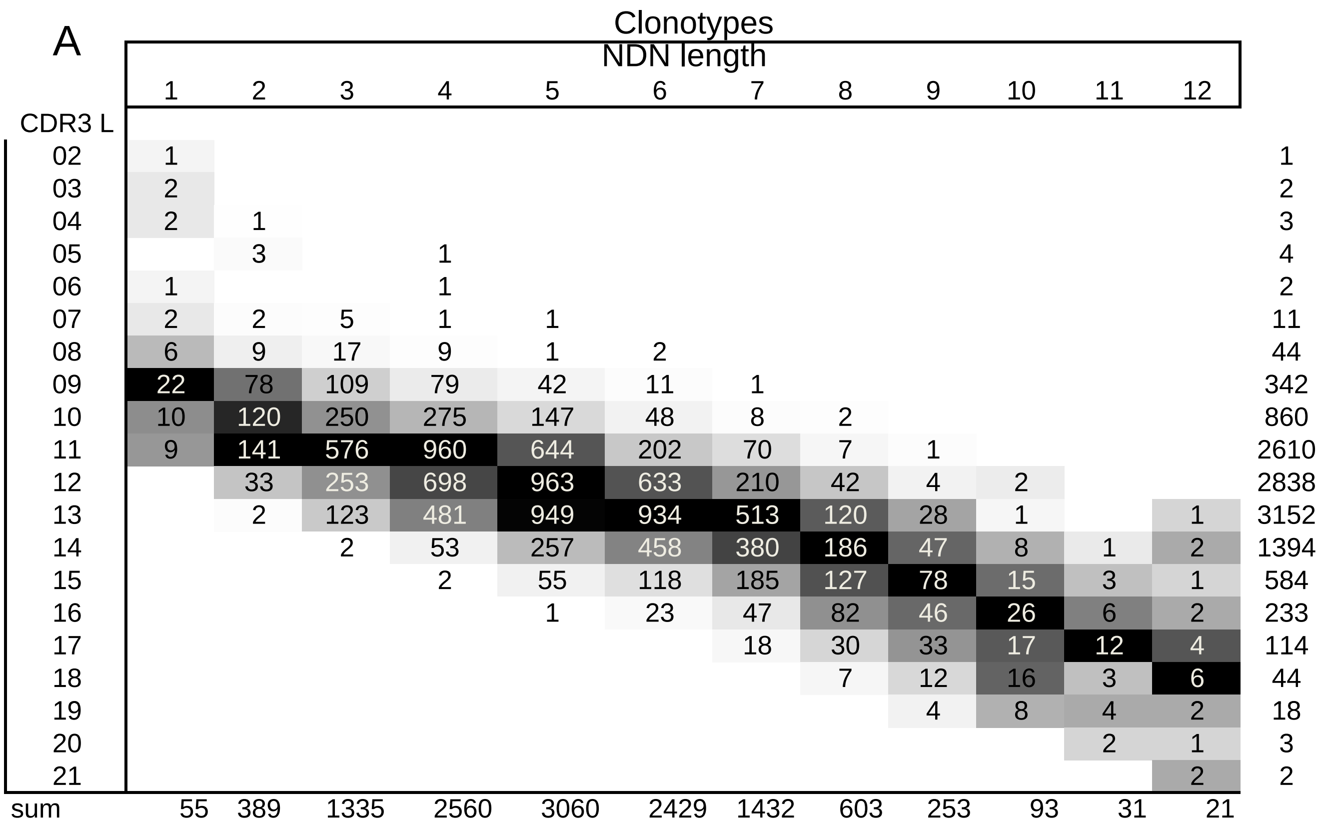
Supplemental Figure 4. Rank frequency data. For each panel the X-axis is the natural log of the rank and the Y axis is the natural log of the rank frequency. The slope, intercept and R^2 value for each trend line are shown in each panel except for A. **A. Rank frequency plot of 2aa motif data.** As described in the text the doublet motif data could not be fit to a line. **B. Doublet motifs from clonotypes of CDR3 length 11.** The data are further divided into subsets based on NDN start position. The NDN start positions (2 to 5) are identified in each panel. **C. Doublet motifs from clonotypes of CDR3 length 12 further divided by NDN start position. D. Doublet motifs of clonotypes of CDR3 length 13 further divided by NDN start position. E. Doublet motifs of clonotypes of CDR3 length 14 further divided by NDN start position.**

Supplemental Figure 1

Clonotype Name	J	Count	Type	Change	Region
sMMAYe.1111B19S1B21L10	2S1	1731		Benchmark	
sMLAYe.1611B19S1B21L10	2S1	1	transversion	ATG to CTG	NDN
sMMAYe.1121B19S1B21L10	2S1	2	transition	GCT to GCC	NDN
sMMAYe.1141B19S1B21L10	2S1	1	transversion	GCT to GCG	NDN
sMMTYe.1111B19S1B21L10	2S1	1	transition	GCT to ACT	NDN
sMMVYe.1111B19S1B21L10	2S1	4	transition	GCT to GTT	NDN
sMTAYe.1411B19S1B21L10	2S1	6	transition	ATG to ACG	NDN
sMVAYe.1411B19S1B21L10	2S1	3	transition	ATG to GTG	NDN
sMMAYEq.11111B19S1B21L10	2S1	1	transition	GAG to GAA	J
sMMAYEQf.111121B19S1B21L10	2S1	1	transition	CAG to CAA	J
sMMAYEOf.111122B19S1B21L10	2S1	3	transition	CAG to TAG*	J
sMMAYEQLf.1111224B19S1B21L10	2S1	2	transition	TTC to CTC	J
sMMAYEQSf.1111222B19S1B21L10	2S1	5	transition	TTC to TCC	J
sMMAYERf.111124B19S1B21L10	2S1	1	transition	CAG to CGG	J
sMMAYGq.11114B19S1B21L10	2S1	5	transition	GAG to GGG	J
sMMAMSSSf.1111651B19S1B21L10	2S1	1	frame-shift	GAG to AG	J
			sum errors	37	2.1%
sFRSGq.2662B19S1B15L11	1S5	1903		Benchmark	
sSFRSGq.62662B19S1B15L11	1S5	3	transition	AGT to AGC	V
sGFRSGq.12662B19S1B15L11	1S5	1	transition	AGT to GGT	V
aSSFRSGq.652662B19S1B15L11	1S5	1	transition	AGT to AGC	V
aGSFRSGq.152662B19S1B15L11	1S5	5	transition	AGT to GGT	V
sFRSGq.2661B19S1B15L11	1S5	1	transition	AGG to GGG	NDN
sFGSGq.2462B19S1B15L11	1S5	1	transition	GGC to GGT	NDN
sFRGDq.2622B19S1B15L11	1S5	2	transition x 2	AGCGGC to **G**GCG**A**C	NDN
sFRGGq.2622B19S1B15L11	1S5	5	transition	AGC to GGC	NDN
sFRSDq.2662B19S1B15L11	1S5	11	transition	GGC to GAC	NDN
sFRSSq.2666B19S1B15L11	1S5	1	transition	GGC to AGC	NDN
sFRSGPp.26624B19S1B15L11	1S5	2	transversion	CAG to CCG	J
sFRSGRp.26624B19S1B15L11	1S5	6	transition	CAG to CGG	J
sFRSGQPRh.2662224B19S1B15L11	1S5	1	transition	CAG to CGG	J
sFRSGQPQRf.26622223B19S1B15L11	1S5	1	transition	CAT to CGT	J
sFRSGQPQHf.26622221B19S1B12L11	1S2	3	transition	CAT to CAC	J
sFRSGQPQAf.26622223B19S1B15L11	1S5	1	frame-shift	CAT to GCAT	J
sFRSGQPSYf.26622261B19S1B15L11	1S5	1	frame-shift	CCC to CC	J
sFRSGPAPAf.26643233B19S1B15L11	1S5	1	frame-shift	CCC to CCCG	J
			sum errors	47	2.4%

Supplemental Figure 2

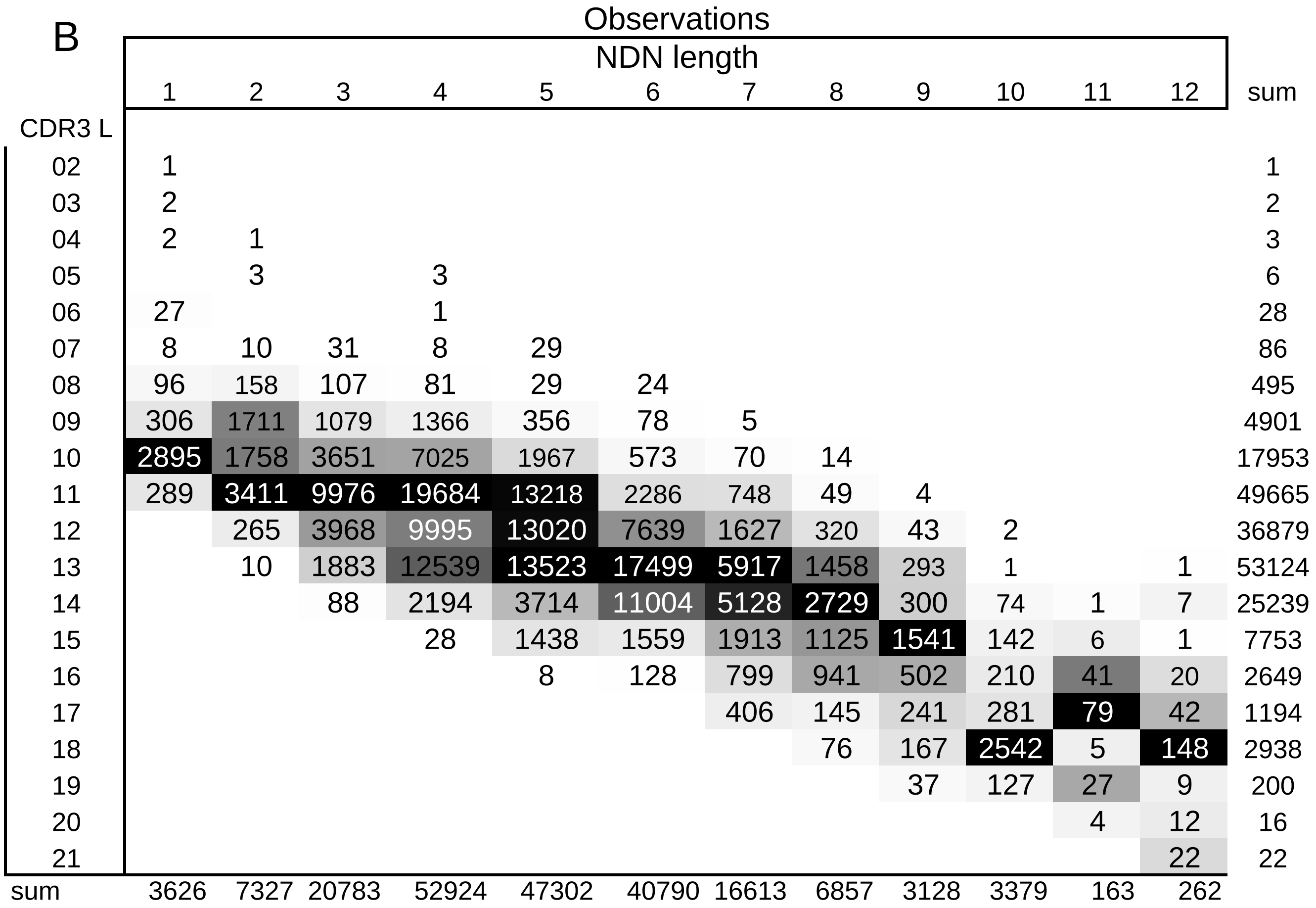
A



Clonotypes

CDR3 L	1	2	3	4	5	6	7	8	9	10	11	12	
	NDN length												
02	1												1
03	2												2
04	2	1											3
05		3		1									4
06	1			1									2
07	2	2	5	1	1								11
08	6	9	17	9	1	2							44
09	22	78	109	79	42	11	1						342
10	10	120	250	275	147	48	8	2					860
11	9	141	576	960	644	202	70	7	1				2610
12		33	253	698	963	633	210	42	4	2			2838
13		2	123	481	949	934	513	120	28	1		1	3152
14			2	53	257	458	380	186	47	8	1	2	1394
15				2	55	118	185	127	78	15	3	1	584
16					1	23	47	82	46	26	6	2	233
17							18	30	33	17	12	4	114
18								7	12	16	3	6	44
19									4	8	4	2	18
20											2	1	3
21												2	2
sum	55	389	1335	2560	3060	2429	1432	603	253	93	31	21	

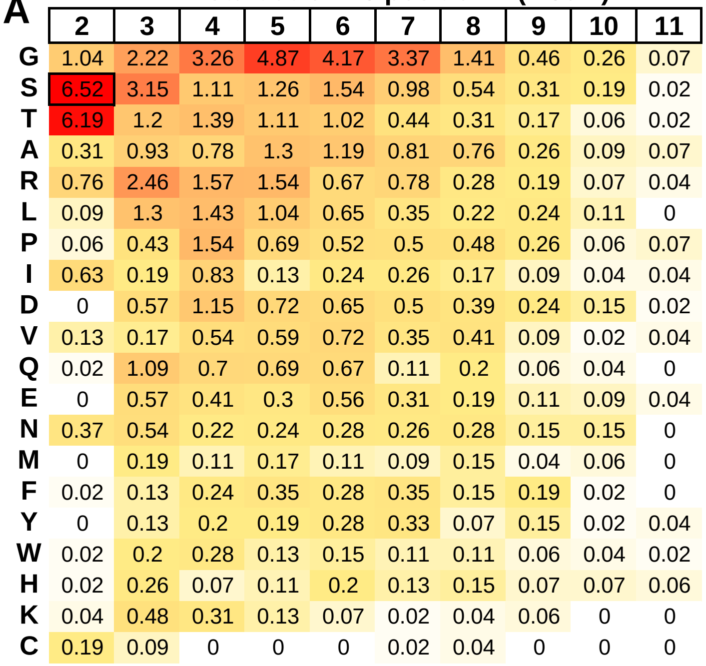
B



Observations

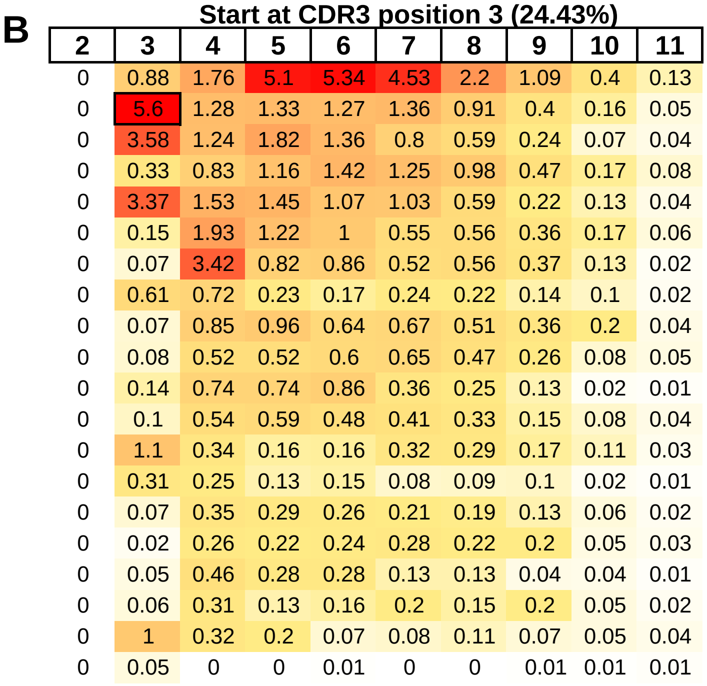
CDR3 L	1	2	3	4	5	6	7	8	9	10	11	12	sum
	NDN length												
02	1												1
03	2												2
04	2	1											3
05		3		3									6
06	27			1									28
07	8	10	31	8	29								86
08	96	158	107	81	29	24							495
09	306	1711	1079	1366	356	78	5						4901
10	2895	1758	3651	7025	1967	573	70	14					17953
11	289	3411	9976	19684	13218	2286	748	49	4				49665
12		265	3968	9995	13020	7639	1627	320	43	2			36879
13		10	1883	12539	13523	17499	5917	1458	293	1		1	53124
14			88	2194	3714	11004	5128	2729	300	74	1	7	25239
15				28	1438	1559	1913	1125	1541	142	6	1	7753
16					8	128	799	941	502	210	41	20	2649
17							406	145	241	281	79	42	1194
18								76	167	2542	5	148	2938
19									37	127	27	9	200
20											4	12	16
21												22	22
sum	3626	7327	20783	52924	47302	40790	16613	6857	3128	3379	163	262	

Supplemental Figure 3



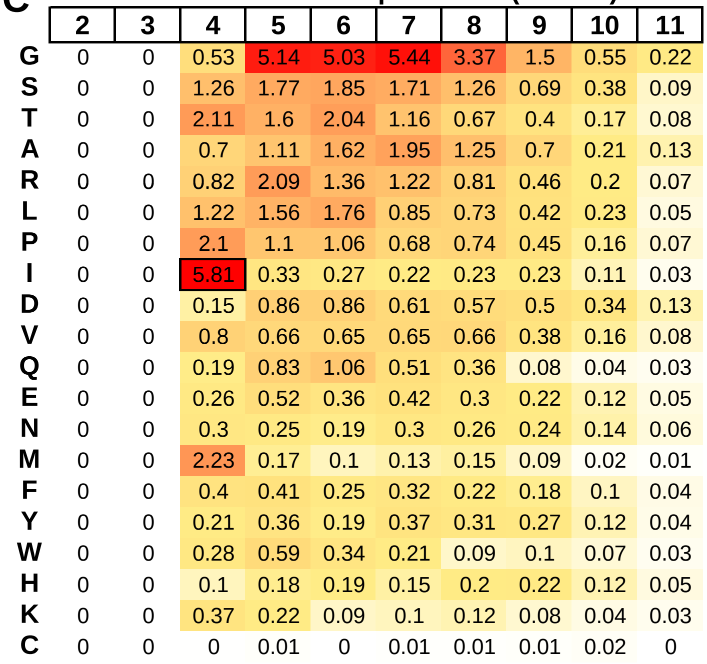
A Start at CDR3 position 2 (7.36%)

	2	3	4	5	6	7	8	9	10	11
G	1.04	2.22	3.26	4.87	4.17	3.37	1.41	0.46	0.26	0.07
S	6.52	3.15	1.11	1.26	1.54	0.98	0.54	0.31	0.19	0.02
T	6.19	1.2	1.39	1.11	1.02	0.44	0.31	0.17	0.06	0.02
A	0.31	0.93	0.78	1.3	1.19	0.81	0.76	0.26	0.09	0.07
R	0.76	2.46	1.57	1.54	0.67	0.78	0.28	0.19	0.07	0.04
L	0.09	1.3	1.43	1.04	0.65	0.35	0.22	0.24	0.11	0
P	0.06	0.43	1.54	0.69	0.52	0.5	0.48	0.26	0.06	0.07
I	0.63	0.19	0.83	0.13	0.24	0.26	0.17	0.09	0.04	0.04
D	0	0.57	1.15	0.72	0.65	0.5	0.39	0.24	0.15	0.02
V	0.13	0.17	0.54	0.59	0.72	0.35	0.41	0.09	0.02	0.04
Q	0.02	1.09	0.7	0.69	0.67	0.11	0.2	0.06	0.04	0
E	0	0.57	0.41	0.3	0.56	0.31	0.19	0.11	0.09	0.04
N	0.37	0.54	0.22	0.24	0.28	0.26	0.28	0.15	0.15	0
M	0	0.19	0.11	0.17	0.11	0.09	0.15	0.04	0.06	0
F	0.02	0.13	0.24	0.35	0.28	0.35	0.15	0.19	0.02	0
Y	0	0.13	0.2	0.19	0.28	0.33	0.07	0.15	0.02	0.04
W	0.02	0.2	0.28	0.13	0.15	0.11	0.11	0.06	0.04	0.02
H	0.02	0.26	0.07	0.11	0.2	0.13	0.15	0.07	0.07	0.06
K	0.04	0.48	0.31	0.13	0.07	0.02	0.04	0.06	0	0
C	0.19	0.09	0	0	0	0.02	0.04	0	0	0



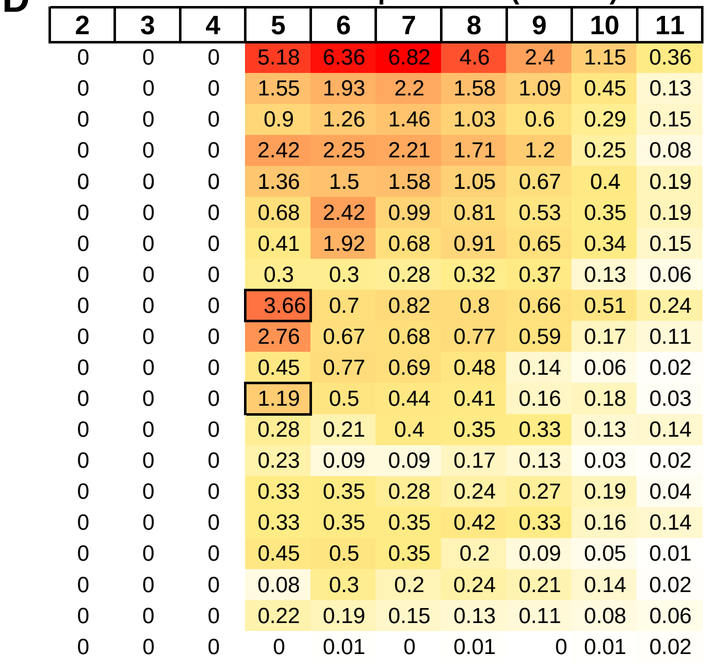
B Start at CDR3 position 3 (24.43%)

	2	3	4	5	6	7	8	9	10	11
G	0	0.88	1.76	5.1	5.34	4.53	2.2	1.09	0.4	0.13
S	0	5.6	1.28	1.33	1.27	1.36	0.91	0.4	0.16	0.05
T	0	3.58	1.24	1.82	1.36	0.8	0.59	0.24	0.07	0.04
A	0	0.33	0.83	1.16	1.42	1.25	0.98	0.47	0.17	0.08
R	0	3.37	1.53	1.45	1.07	1.03	0.59	0.22	0.13	0.04
L	0	0.15	1.93	1.22	1	0.55	0.56	0.36	0.17	0.06
P	0	0.07	3.42	0.82	0.86	0.52	0.56	0.37	0.13	0.02
I	0	0.61	0.72	0.23	0.17	0.24	0.22	0.14	0.1	0.02
D	0	0.07	0.85	0.96	0.64	0.67	0.51	0.36	0.2	0.04
V	0	0.08	0.52	0.52	0.6	0.65	0.47	0.26	0.08	0.05
Q	0	0.14	0.74	0.74	0.86	0.36	0.25	0.13	0.02	0.01
E	0	0.1	0.54	0.59	0.48	0.41	0.33	0.15	0.08	0.04
N	0	1.1	0.34	0.16	0.16	0.32	0.29	0.17	0.11	0.03
M	0	0.31	0.25	0.13	0.15	0.08	0.09	0.1	0.02	0.01
F	0	0.07	0.35	0.29	0.26	0.21	0.19	0.13	0.06	0.02
Y	0	0.02	0.26	0.22	0.24	0.28	0.22	0.2	0.05	0.03
W	0	0.05	0.46	0.28	0.28	0.13	0.13	0.04	0.04	0.01
H	0	0.06	0.31	0.13	0.16	0.2	0.15	0.2	0.05	0.02
K	0	1	0.32	0.2	0.07	0.08	0.11	0.07	0.05	0.04
C	0	0.05	0	0	0.01	0	0	0.01	0.01	0.01



C Start at CDR3 position 4 (50.77%)

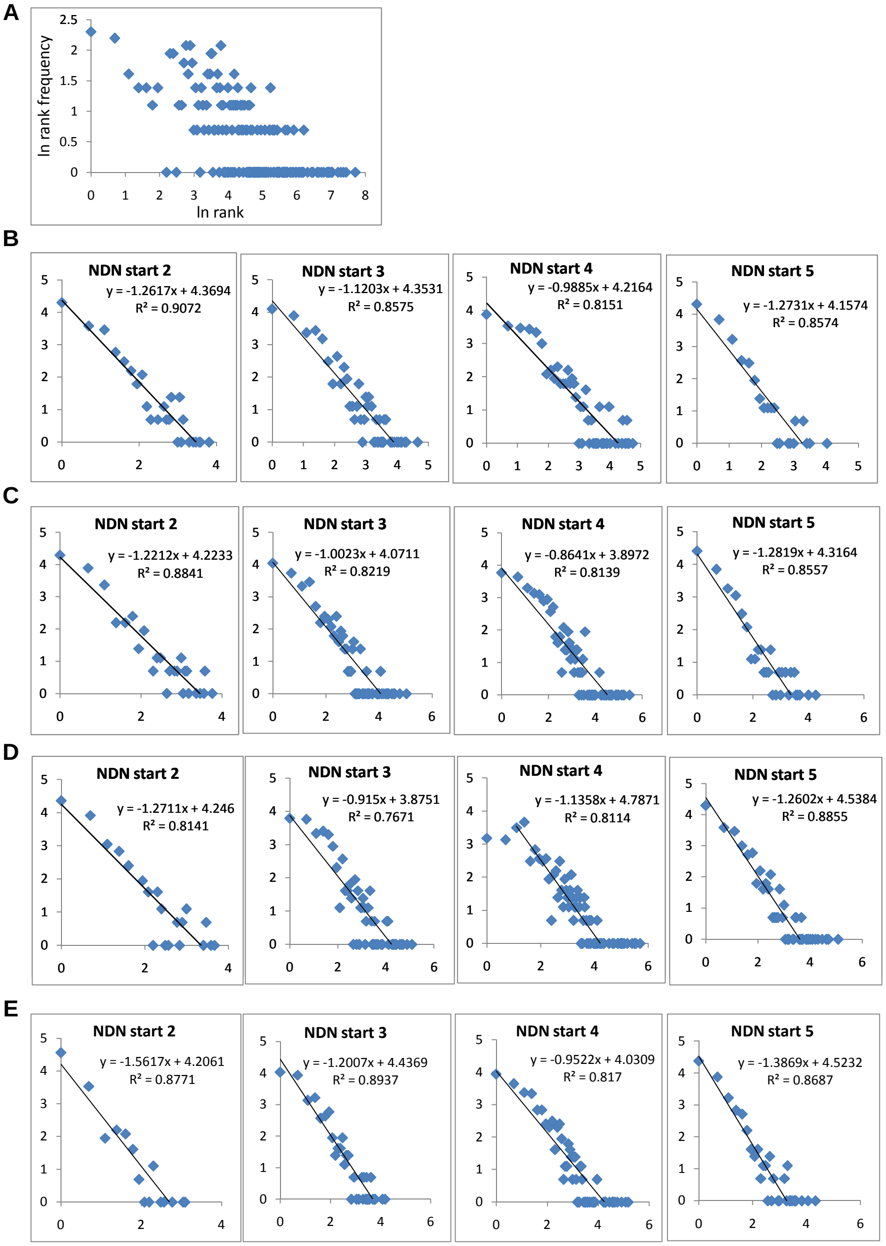
	2	3	4	5	6	7	8	9	10	11
G	0	0	0.53	5.14	5.03	5.44	3.37	1.5	0.55	0.22
S	0	0	1.26	1.77	1.85	1.71	1.26	0.69	0.38	0.09
T	0	0	2.11	1.6	2.04	1.16	0.67	0.4	0.17	0.08
A	0	0	0.7	1.11	1.62	1.95	1.25	0.7	0.21	0.13
R	0	0	0.82	2.09	1.36	1.22	0.81	0.46	0.2	0.07
L	0	0	1.22	1.56	1.76	0.85	0.73	0.42	0.23	0.05
P	0	0	2.1	1.1	1.06	0.68	0.74	0.45	0.16	0.07
I	0	0	5.81	0.33	0.27	0.22	0.23	0.23	0.11	0.03
D	0	0	0.15	0.86	0.86	0.61	0.57	0.5	0.34	0.13
V	0	0	0.8	0.66	0.65	0.65	0.66	0.38	0.16	0.08
Q	0	0	0.19	0.83	1.06	0.51	0.36	0.08	0.04	0.03
E	0	0	0.26	0.52	0.36	0.42	0.3	0.22	0.12	0.05
N	0	0	0.3	0.25	0.19	0.3	0.26	0.24	0.14	0.06
M	0	0	2.23	0.17	0.1	0.13	0.15	0.09	0.02	0.01
F	0	0	0.4	0.41	0.25	0.32	0.22	0.18	0.1	0.04
Y	0	0	0.21	0.36	0.19	0.37	0.31	0.27	0.12	0.04
W	0	0	0.28	0.59	0.34	0.21	0.09	0.1	0.07	0.03
H	0	0	0.1	0.18	0.19	0.15	0.2	0.22	0.12	0.05
K	0	0	0.37	0.22	0.09	0.1	0.12	0.08	0.04	0.03
C	0	0	0	0.01	0	0.01	0.01	0.01	0.02	0



D Start at CDR3 position 5 (17.44%)

	2	3	4	5	6	7	8	9	10	11
G	0	0	0	5.18	6.36	6.82	4.6	2.4	1.15	0.36
S	0	0	0	1.55	1.93	2.2	1.58	1.09	0.45	0.13
T	0	0	0	0.9	1.26	1.46	1.03	0.6	0.29	0.15
A	0	0	0	2.42	2.25	2.21	1.71	1.2	0.25	0.08
R	0	0	0	1.36	1.5	1.58	1.05	0.67	0.4	0.19
L	0	0	0	0.68	2.42	0.99	0.81	0.53	0.35	0.19
P	0	0	0	0.41	1.92	0.68	0.91	0.65	0.34	0.15
I	0	0	0	0.3	0.3	0.28	0.32	0.37	0.13	0.06
D	0	0	0	3.66	0.7	0.82	0.8	0.66	0.51	0.24
V	0	0	0	2.76	0.67	0.68	0.77	0.59	0.17	0.11
Q	0	0	0	0.45	0.77	0.69	0.48	0.14	0.06	0.02
E	0	0	0	1.19	0.5	0.44	0.41	0.16	0.18	0.03
N	0	0	0	0.28	0.21	0.4	0.35	0.33	0.13	0.14
M	0	0	0	0.23	0.09	0.09	0.17	0.13	0.03	0.02
F	0	0	0	0.33	0.35	0.28	0.24	0.27	0.19	0.04
Y	0	0	0	0.33	0.35	0.35	0.42	0.33	0.16	0.14
W	0	0	0	0.45	0.5	0.35	0.2	0.09	0.05	0.01
H	0	0	0	0.08	0.3	0.2	0.24	0.21	0.14	0.02
K	0	0	0	0.22	0.19	0.15	0.13	0.11	0.08	0.06
C	0	0	0	0	0.01	0	0.01	0	0.01	0.02

Supplemental Figure 4



0.9675