Supplementary Information

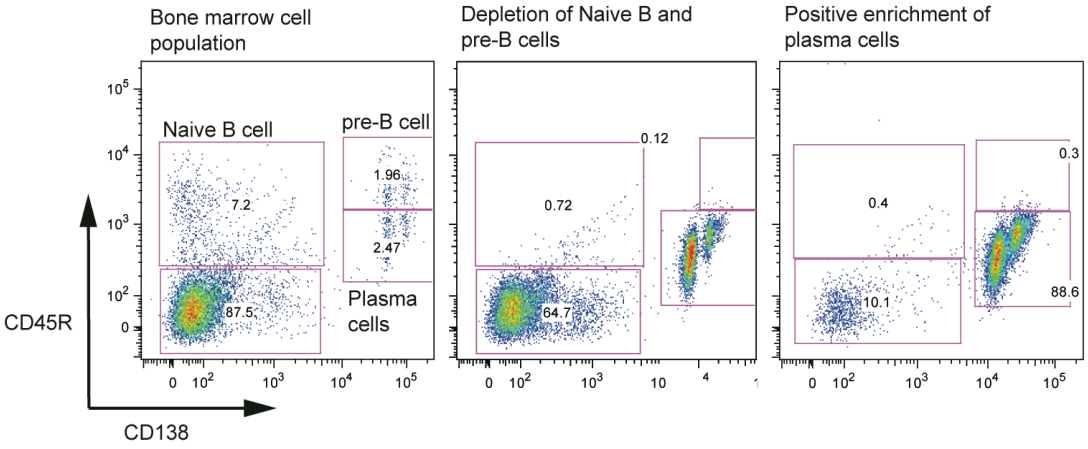
Isolation of monoclonal antibodies without screening by mining the variable gene repertoire of plasma cells

Sai T. Reddy[1,2], Xin Ge[1], Aleksandr E. Miklos[3,4], Randall A. Hughes[3,4], Seung Hyun Kang[1] , Kam Hon Hoi[2], Constantine Chrysostomou[1], Scott P. Hunicke-Smith[3], Brent L. Iverson[3,5], Philip W. Tucker[3,6], Andrew D. Ellington[3,4,5] & George Georgiou[1,2,3,6]

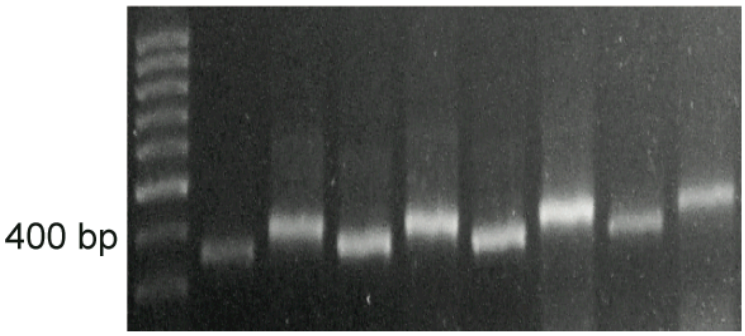
[1]Department of Chemical Engineering, [2]Department of Biomedical Engineering, [3]Institute for Cellular and Molecular Biology, [4]Applied Research Laboratories, [5]Department of Chemistry and Biochemistry, [6]Section of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, Texas, 78712, USA.

Correspondence should be addressed to G.G. (gg@che.utexas.edu).

Nature Biotechnology: doi: 10.1038/nbt.1673

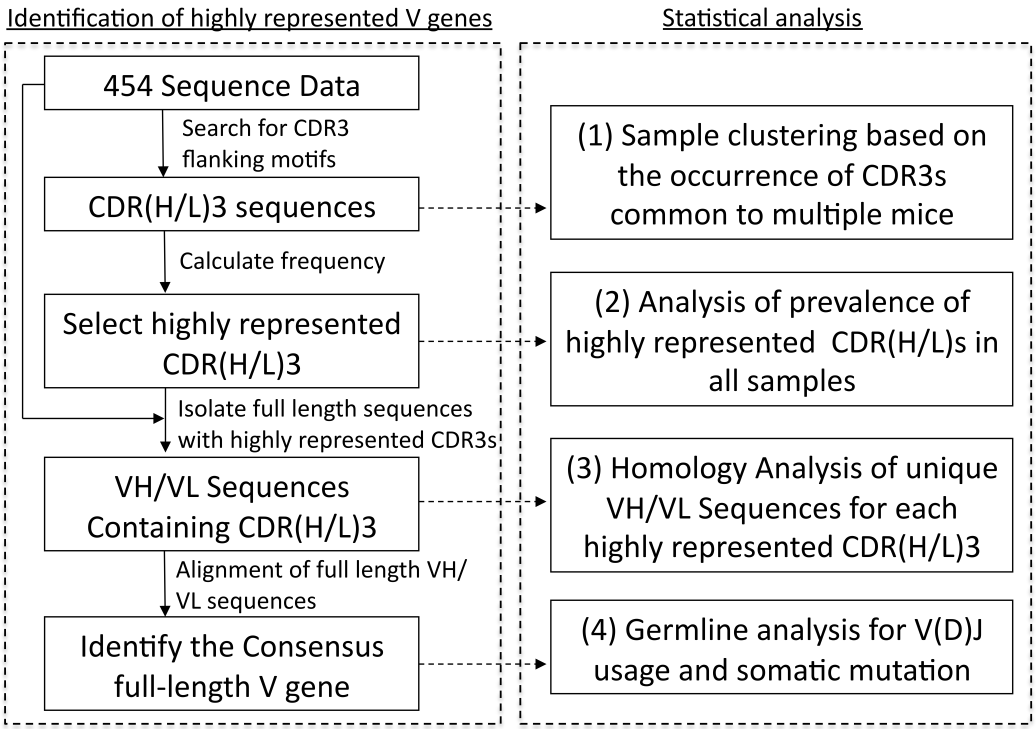


Supplementary Figure 1. Isolation of plasma cells from bone marrow. Left Panel: Flow cytometry plot of the total mouse bone marrow cell population labeled with anti-CD45R-APC and anti-CD138-PE antibodies. Middle panel: Bone marrow cells remaining following depletion of $CD45R^+$ cells. Right Panel. Cell population isolated following magnetic sorting with anti-$CD138^+$ conjugated magnetic beads.



Supplementary Figure 2. Variable light (V_L) and variable heavy (V_H) chain genes from bone marrow plasma cells. Agarose gel electrophoresis of V_L and V_H genes amplified by PCR from cDNA derived from bone marrow plasma cells of mice immunized with different antigens. From left: 1st lane is DNA Ladder; 2nd, 4th, 6th, 8th lanes are V_L (~370 bp) ; 3rd, 5th, 7th, 9th lanes are V_H (~400 bp).

Nature Biotechnology: doi: 10.1038/nbt.1673



Supplementary Figure 3. Flow chart of the bioinformatics pipeline for V gene analysis. First, CDR3s were identified by homology to conserved flanking amino acid sequences motifs. CDR3s found at the highest frequency (typically with frequency > 1%) were used to group the V gene sequences of interest. Homology analysis of V genes containing highly represented CDR3s was performed by multiple sequence alignment and calculation of pairwise identity. Finally, germline analysis of highly represented full-length V genes was performed to determine somatic mutations and V(D)J and V-J gene usage.

Supplementary Table 1. Highly conserved motifs at the N- and C- termini of CDR3s, and their occurrence frequency in the Kabat database.

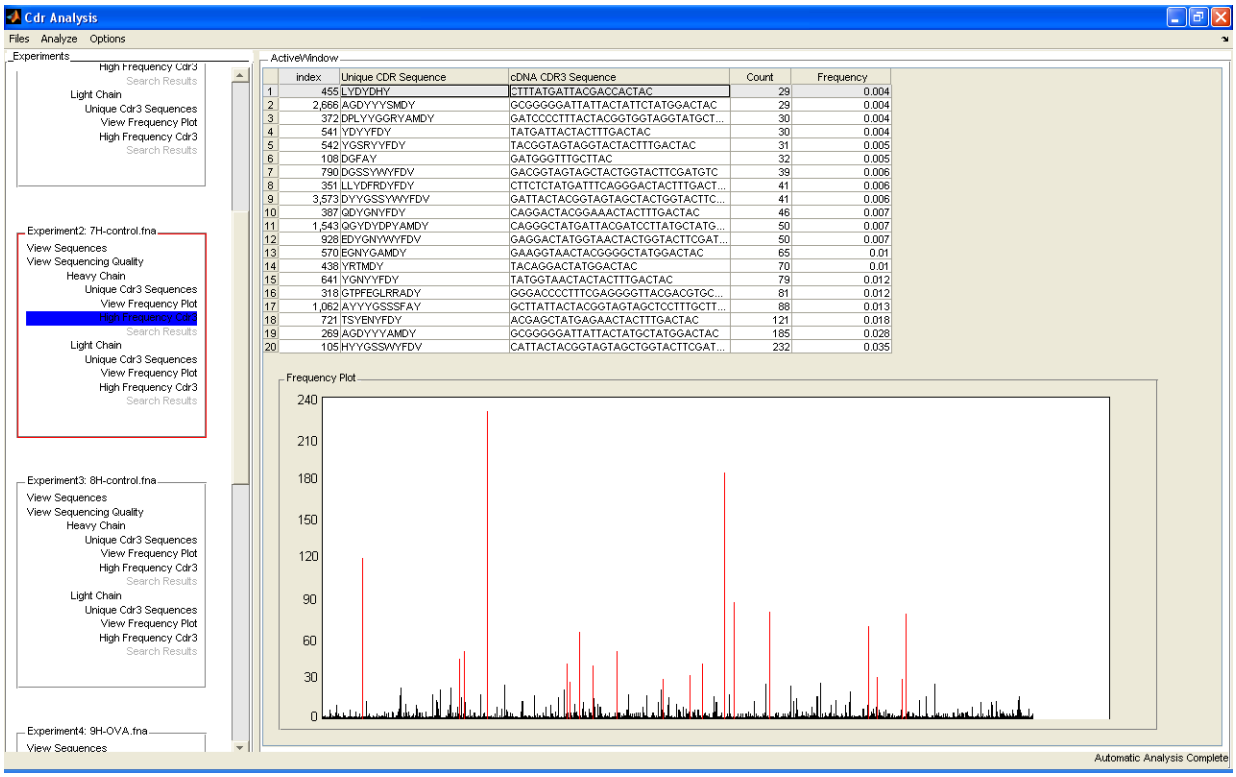
N-CDR3H	**Residue (Kabat #)**	**Asp (86)**	**Tyr/Phe (90)**	**Tyr/Phe (91)**	**Cys (92)**
	Frequency	98.3%	99.4%	98.8%	99.5%
C-CDR3H	**Residue (Kabat #)**	**Trp (103)**	**Gly (104)**	**Gly (106)**	**Ser/Thr (107)**
	Frequency	99.4%	98.6%	99.7%	98.7%
N-CDR3L	**Residue (Kabat #)**	**Asp (82)**	**Tyr (86)**	**Tyr/Phe (87)**	**Cys (88)**
	Frequency	98.5%	98.2%	97.7%	99.1%
C-CDR3L	**Residue (Kabat #)**	**Phe (98)**	**Gly (99)**	**Gly (101)**	**Thr (102)**
	Frequency	99.5%	99.4%	98.5%	98.1%

Supplementary Table 2. Number of CDR3 containing sequences identified from 454 GS-FLX DNA sequencing using CDR3 flanking consensus motifs.

Sample	454 GS-FLX Sequencing Size	Sequences Containing CDR-H3 motif			Sequences Containing CDR-L3 motif		
		Total Number	Number of Unique CDR-H3	Number of CDR-H3 as Single Copy	Total Number	Number of Unique CDR-L3	Number of CDR-L3 as Single Copy
Adjuvant-1	32066	6681	2706	1811	7112	1638	1053
Adjuvant-2	86720	16743	4640	2890	21241	3136	1888
Ova-1	63872	15350	4789	3010	13355	2251	1355
Ova-2	72257	15751	3821	2401	17200	2786	1700
C1s-1	43753	11595	2440	1443	13972	1706	1045
C1s-2	39961	9071	1799	999	14664	1477	847
Bright-1	36599	9453	2025	1178	12209	1383	632
Bright-2	39790	11769	2530	1210	10441	1422	578
Total	**415018**	**96413**	**24750**	**14942**	**110194**	**15799**	**9098**
Unique Sequences Across All Samples			**21271**			**8690**	

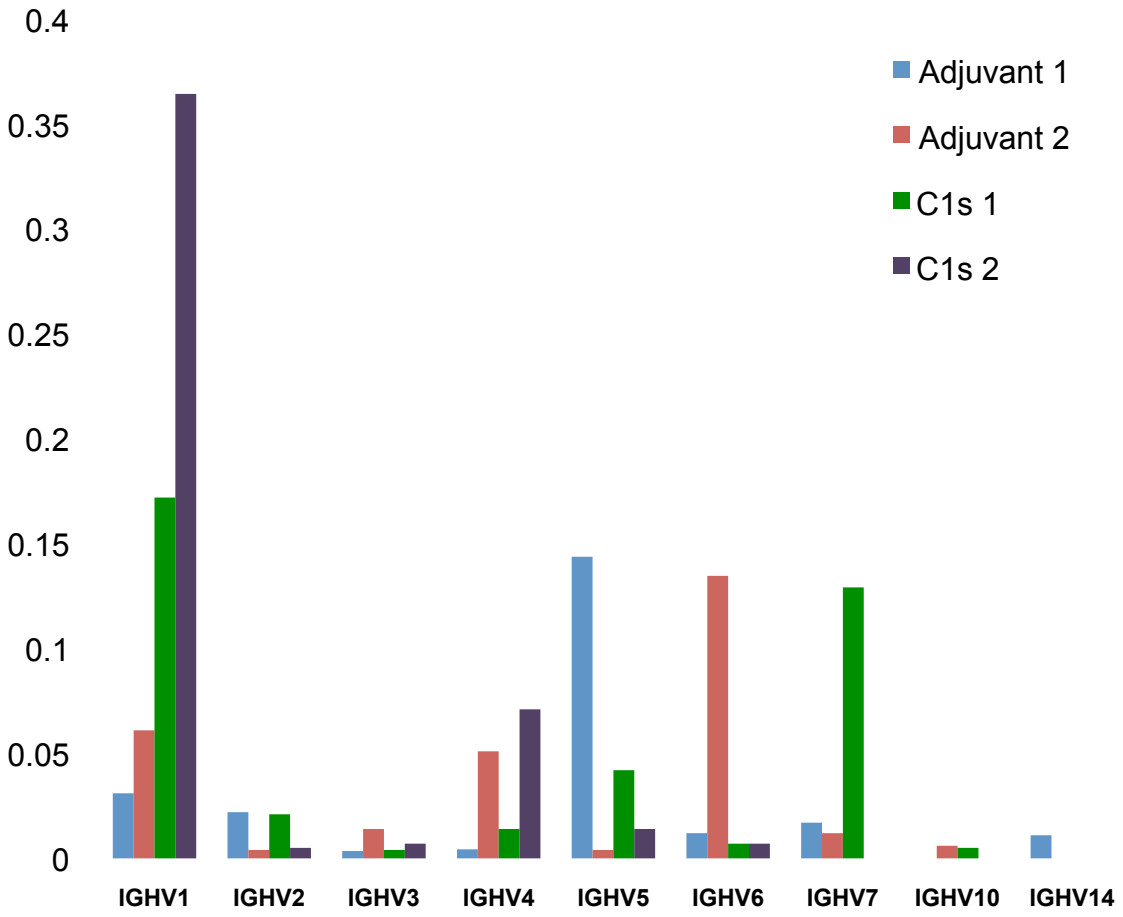
Supplementary Table 3. Comparison of 454 DNA sequencing results generated by two different facilities from two different cDNA library preparations (SeqWright and UT-Austin Genome Sequencing Center).

Sample	Adjuvant-1		OVA-2	
Facility	SeqWright	UT-Austin	SeqWright	UT-Austin
Total seq reads	7112	20850	17200	17291
CDR-L3 seq	%	%	%	%
WQGTHFPLT	7.02	10.9	5.32	6.80
QQYNSYPLT	5.28	4.97	4.05	3.70
QQYSSYPLT	4.91	3.88	3.46	2.82
QQSNSWYT	2.00	2.20	1.41	1.45
QNGHSFPLT	1.27	0.98	1.12	0.98
WQGTHFPRT	1.18	1.71	1.23	1.64
QHHYGTPPWT	1.52	1.48	0.77	0.87
QQHYSTPWT	1.42	1.35	2.01	2.01
HQWSSYPT	0.48	0.47	1.87	0.43
SQTTHVPPT	1.15	1.06	0.74	0.64
WQGTHFPQT	1.69	1.92	1.63	1.83
AQFYSYPLT	1.45	1.22	1.69	1.19
MQHLEYPYT	0.65	0.72	0.72	0.45
QQGQSYPWT	0.65	0.77	1.07	1.31
QQWNYPLIT	0.84	0.77	0.78	0.87
QQGNTLPWT	0.58	0.68	0.59	0.59



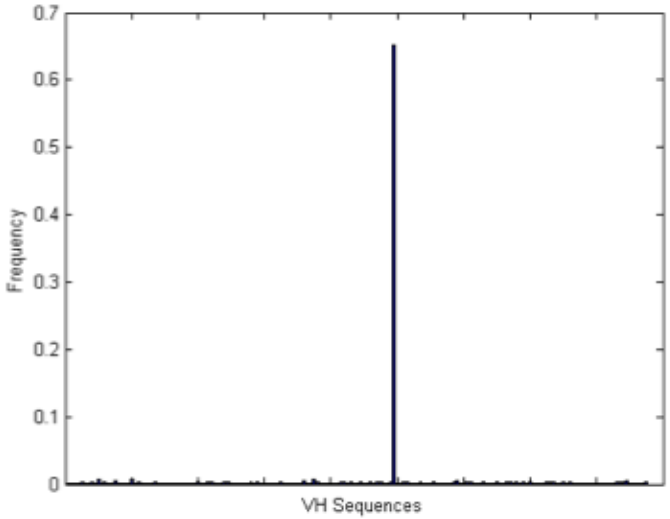
	index	Unique CDR Sequence	cDNA CDR3 Sequence	Count	Frequency
1	455	LYDYDHY	CTTTATGATTACGACCACTAC	29	0.004
2	2,666	AGDYYYYSMDY	GCGGGGGATTATTACTATTCTATGGACTAC	29	0.004
3	372	DPLYYGGRYAMDY	GATCCCCTTTACTACGGTGGTAGGTATGCT...	30	0.004
4	541	YDYYFDY	TATGATTACTACTTTGACTAC	30	0.004
5	542	YGSRYYFDY	TACGGTAGTAGGTACTACTTTGACTAC	31	0.005
6	108	DGFAY	GATGGGTTTGCTTAC	32	0.005
7	790	DGSSYWYFDV	GACGGTAGTAGCTACTGGTACTTCGATGTC	39	0.006
8	351	LLYDFRDYFDY	CTTCTCTATGATTTCAGGGACTACTTTGACT...	41	0.006
9	3,573	DYYGSSYWYFDV	GATTACTACGGTAGTAGCTACTGGTACTTC...	41	0.006
10	387	QDYGNYFDY	CAGGACTACGGAAACTACTTTGACTAC	46	0.007
11	1,543	QGYDYDPYAMDY	CAGGGCTATGATTACGATCCTTATGCTATG...	50	0.007
12	928	EDYGNYWYFDV	GAGGACTATGGTAACTACTGGTACTTCGAT...	50	0.007
13	570	EGNYGAMDY	GAAGGTAACTACGGGGCTATGGACTAC	65	0.01
14	438	YRTMDY	TACAGGACTATGGACTAC	70	0.01
15	641	YGNYYFDY	TATGGTAACTACTACTTTGACTAC	79	0.012
16	318	GTPFEGLRRADY	GGGACCCCTTTCGAGGGGTTACGACGTGC...	81	0.012
17	1,062	AYYYGSSSFAY	GCTTATTACTACGGTAGTAGCTCCTTTGCTT...	88	0.013
18	721	TSYENYFDY	ACGAGCTATGAGAACTACTTTGACTAC	121	0.018
19	269	AGDYYYAMDY	GCGGGGGATTATTACTATGCTATGGACTAC	185	0.028
20	105	HYYGSSWYFDV	CATTACTACGGTAGTAGCTGGTACTTCGAT...	232	0.035

Supplementary Figure 4. Example of graphical user interface (GUI) for V gene repertoire analysis. A GUI application was developed for organization and graphical representation of high-throughput sequencing results of V genes. The program imports and organizes data sets from different samples, identifies CDRH3 and CDRL3 using the described CDR flanking motifs, and extracts CDR3 frequency distributions.

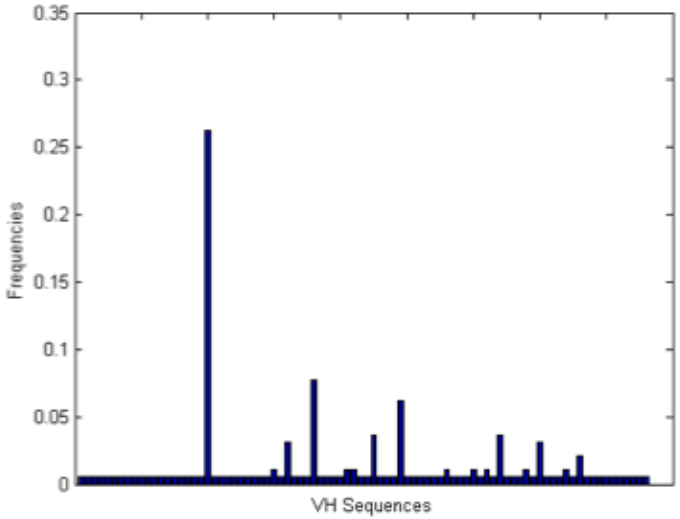


Supplementary Figure 5. V_H germline family representation in adjuvant and C1s immunized repertoires. Bar graph represents the frequency of V_H gene families among the top 30 V_H sequences in each repertoire (representing 24-47% of the total V_H repertoire). A clear skewing towards *IGHV1* is demonstrated in immunized mice.

A. C1s-1.1 V_H (CDRH3=GNYYYAMDY)



B. C1s-2.2 V_H (CDRH3=WLLLAY)



C. Top 3 full-length VH sequences containing C1s-2.2 CDR3

Rank	%	Sequence
1	26.3	EVKLVESGGGFVKPGGSLKLSCAASGFTFSTYGMSWVRQTPEKRLEWVASISAGGTTYYSDSVKGRFTISRDNARNILYLQMSSLRSEDTAMYYCARWLLLAYWGQGTLVT
2	7.7	D---------L--------I--------------------------------S--S---P--
3	6.2	----------L--S-----I--------------------------------S--S---P--

Supplementary Figure 6. Homology analysis of full-length V genes containing the same CDR3. (A) The dataset of full-length C1s-1.1 V_H genes (CDRH3 = GNYYYAMDY) is dominated by a single sequence (65%), as all the other sequences occur at a frequency <0.8%. (**B)** Frequency distribution of full-length C1s-2.2 V_H genes (CDRH3 = WLLLAY) shows multiple sequences with high frequency. **(C)** Top three full-length V_H sequences containing C1s-2.2 CDR3 and their respective frequencies (%).

Supplementary Table 4. Frequencies (%) of the most highly represented CDRL3 and CDRH3[a] sequences.

Clone	Percent	CDRL3	Clone	Percent	CDRH3
Ovalbumin			**Ovalbumin**		
1.1L	11.69	WQGTHFPLT	1.1H	7.17	TYGSSYYAMDY
1.2L	4.44	QQYNSYPLT	1.2H	1.13	TRLLWLYAMDY
1.3L	3.38	QQSNSWYT	1.3H	0.57	DVYDGYAMDY
1.4L	2.20	QHHYGTPPWT	1.4H	0.54	NPYAMDY
2.1L	5.32	WQGTHFPLT	2.1H	7.59	RTTVSRDWYFDV
2.2L	4.05	QQYNSYPLT	2.2H	3.22	YYYGSSAMDY
2.3L	3.46	QQYSSYPLT	2.3H	2.22	DGWYYFDY
2.4L	2.01	QQHYSTPWT	2.4H	2.11	EDDYDLFAY
2.5L	1.87	HQWSSYPT	2.5H	1.25	DTTVVEGDYFDY
C1s			**C1s**		
1.1L	12.95	WQGTHFPQT	1.1H	7.93	GNYYYAMDY
1.2L	6.94	QQWSSYPQLT	1.2H	2.64	DMISYWYFDV
1.3L	3.81	QNDHSYPLT	1.3H	1.67	EDYGNYWYFDV
1.4L	3.16	QQGQSYPFT	1.4H	1.17	EGYYYGSSYFDY
2.1L	17.10	FQGSHVPLT	2.1H-A	10.99	SDRYDGYFDY
2.2L	2.62	QQSNEDPWT	2.1H-B	9.93	SDRFDGYFDY
2.3L	2.20	WQGTHFPH	2.2H	3.30	WLLLAY
2.4L	1.64	QQHYSTPFT	2.3H	1.65	SDGYYYFDY
Bright			**Bright**		
1.1L	6.63	LQYASSPFT	1.1H	7.19	HDYGNYVDY
1.2L	4.73	WQGTHFPRT	1.2H	5.62	DGNYQEDYFDY
1.3L	4.51	QQNNEDPRT	1.3H	1.91	EGYAYDVDY
1.4L	3.59	QQRSSYPLT	1.4H	1.21	YDYGKDFDY
2.1L	7.25	WQGTHFPQT	2.1H	2.56	RGDGNYFFDY
2.2L	4.51	QQGQSYPWT	2.2H	2.27	GDEAWFAY
2.3L	3.12	LQYASSPYT	2.3H	2.04	EGDFDY
2.4L	2.59	FQGSHVPWT	2.4H	1.63	GGNYDYAMDY

[a]CDRH3 sequences present at high frequency in both immunized mice and adjuvant only mice were considered background and thus excluded from the list.

Supplementary Table 5. Germline identity and the number of amino acid somatic mutations (SM) of the most highly represented V_L and V_H[a] genes.

Clone	Germline V_L	#SM V_L
Ovalbumin		
1.1L	V1-135*01 - J5*01	2
1.2L	V6-23*01-J5*01	4
1.3L	N/A - low alignment score	N/A
1.4L	V12-44*01 - J1*01	1
2.1L	V1-135*01 - J5*01	2
2.2L	V6-15*01 - J5*01	1
2.3L	N/A - low alignment score	3
2.4L	V6-23*01 - J5*01	2
C1s		
1.1L	V1-135*01 - J1*01	2
1.2L	V4-55*01 - J5*01	3
1.3L	V8-28*01 - J5*01	1
1.4L	V15-103*01 - J1*01	2
2.1L	V1-117*01 - J5*01	4
2.2L	V3-5*01 - J1*01	1
2.3L	V1-135*01 - J2*01	4
2.4L	V6-25*01 - J4*01	0
Bright		
1.1L	V9-120*01 - J4*01	1
1.2L	V1-135*01 - J1*01	3
1.3L	V3-10*01 - J1*01	2
1.4L	V4-57*01 - J1*01	3
2.1L	V1-135*01 - J1*01	3
2.2L	V15-103*01 - J1*01	2
2.3L	V9-120*01 - J2*01	1
2.4L	V1-117*01 - J1*01	0

Clone	Germline V_H	#SM V_H
Ovalbumin		
1.1	V1-5*01 - J4*01 - D1-1*01	5
1.2	V6-6*02 - J4*01 - D2-2*01	2
1.3	V1S132*01 - J4*01 - D2-3*01	14
1.4	V5-6*01 - J4*01 - D6-1*01	7
2.1	V5-17*02 - J1*01 - D2-1*01	5
2.2	N/A - low alignment score	N/A
2.3	V5-17*02 - J2*01 - D2-3*01	5
2.4	V1-67*01 - J3*01 - D2-4*01	8
2.5	V3-2*02 - J2*01 - D1-1*01	3
C1s		
1.1	V1-5*01 - J4*01 - D2-1*01	2
1.2	V7-3*02 - J1*01 - D2-4*01	0
1.3	V2-9*02 - J1*02 -D2-1*01	3
1.4	V1-14*01 - J2*01 - D1-1*01	4
2.1 (A, B)	V1-4*01 - J2*01 - D2-14*01	3
2.2	V5-9-1*01 - J3*01 - D2-3*01	8
2.3	V1-5*01 - J2*01 - D2-3*01	3
2.4	V1-7*01 - J2*01 - D2-4*01	5
Bright		
1.1	V5-6*01 - J2*01 - D2-1*01	1
1.2	V5-6*01 - J2*01 - D2-1*01	2
1.3	V5-17*02 - J2*01 - D1-3*01	5
1.4	V5-17*02 - J2*01 - D1-1*01	6
2.1	V5-6*01 - J2*01 - D2-1*01	3
2.2	V5-6*01 - J3*01 - D4-1*01	4
2.3	V5-6-4*01 - J2*01 - D3-2*02	4
2.4	V14-3*02 - J4*01 - D1-1*02	9

[a]V_H sequences present at high frequency in both immunized mice and adjuvant only mice were considered background and thus excluded from the list.

Nature Biotechnology: doi: 10.1038/nbt.1673

Supplementary Table 6. Average somatic mutations in nucleotides (nt) and amino acids (AA) of the top 30 V_H sequences in each BM-PC repertoire.

Mouse	# of nt	# of AA
Adjuvant - 1	5.3	2.9
Adjuvant - 2	5.2	2.7
OVA - 1	5.5	3.3
OVA - 2	7.0	3.8
C1s - 1	5.0	2.6
C1s -2	6.7	3.9
Bright -1	4.9	2.8
Bright - 2	7.5	4.6

Supplementary Table 7. The most highly represented CDR3 groups and their full-length variable heavy (V_H) and variable light (V_L) gene frequencies and homologies.

Antigen	CDR3	CDR3 Freq (%)	1st V_H Freq (%)[a]	2nd V_H Freq (%)[a]	V_H Homology[b]
OVA-1.1	GSSYYAMDY	7.11	60.0	1.7	96.1
OVA-1.2	DYYGSSYWYFDV	1.10	47.1	5.8	89.9
OVA-1.3	DNWDWYFDV	0.57	49.0	4.0	95.0
OVA-1.4	LLWLYAMDY	0.54	54.7	4.7	97.3
OVA-2.1	RTTVSRDWYFDV	7.61	15.3	5.6	92.3
OVA-2.2	YYYGSSAMDY	3.23	26.0	10.8	96.0
OVA-2.3	DGWYYFDY	2.22	22.7	4.1	89.1
OVA-2.4	EDDYDLFAY	2.10	9.4	8.7	94.9
C1s-1.1	GNYYYAMDY	7.93	68.8	1.1	97.9
C1s-1.2	DDGYWYFDV	5.14	60.9	5.3	90.0
C1s-1.3	YYYGSSAMDY	4.37	58.5	3.7	94.5
C1s-1.4	DMISYWYFDV	2.64	70.9	1.1	90.0
C1s-2.1	SDRYDGYFDY	10.99	11.1	9.4	95.7
C1s-2.2	SDRFDGYFDY	9.93	12.5	4.2	94.7
C1s-2.3	WLLLAY	3.30	26.3	7.7	88.8
C1s-2.4	YGNYFDY	2.47	72.1	1.4	96.8
Bright-1.1	HDYGNYVDY	7.20	66.2	2.6	98.7
Bright-1.2	DGNYQEDYFDY	5.62	63.1	5.9	98.6
Bright-1.3	EGYAYDVDY	1.91	27.4	23.9	95.6
Bright-1.4	DDYDWYFDV	1.54	59.3	2.8	97.5
Bright-2.1	RGDGNYFFDY	2.57	16.1	14.0	95.0
Bright-2.2	GDEAWFAY	2.27	43.3	6.7	97.1
Bright-2.3	EGDFDY	2.03	14.9	8.1	95.3
Bright-2.4	YYYGSSYFDV	1.84	77.8	0.7	99.2

Antigen	CDRL3	CDR3 Freq (%)	1st V_L Freq (%)[a]	2nd V_L Freq (%)[a]	V_L Homology[b]
OVA-1.1	WQGTHFPLT	11.70	41.4	1.8	92.1
OVA-1.2	QQSNSWYT	4.40	54.5	2.4	94.0
OVA-1.3	QQYSSYPLT	3.38	46.2	1.9	93.9
OVA-1.4	QHHYGTPPWT	2.20	49.7	2.1	93.7
OVA-2.1	WQGTHFPLT	5.32	33.3	2.3	93.7
OVA-2.2	QQYSSYPLT	4.05	43.6	1.1	94.3
OVA-2.3	QQYNSYPLT	3.46	20.1	4.5	92.3
OVA-2.4	QQHYSTPWT	2.01	50.2	2.6	95.3

C1s-1.1	WQGTHFPQT	12.95	68.8	1.1	97.9
C1s-1.2	QQWSSYPQLT	6.94	60.9	5.3	90.0
C1s-1.3	QNDHSYPLT	3.81	58.5	3.7	94.5
C1s-1.4	QQGQSYPWT	3.16	70.8	1.1	98.5
C1s-2.1	FQGSHVPLT	17.10	5.7	4.7	90.4
C1s-2.2	QQSNEDPWT	2.62	65.7	2.8	97.4
C1s-2.3	WQGTHFPH	2.20	36.1	18.5	96.5
C1s-2.4	WQGTHFPT	2.15	39.2	15.6	96.9
Bright-1.1	LQYASSPFT	6.64	74.0	1.0	98.3
Bright-1.2	WQGTHFPRT	4.73	60.8	1.5	97.9
Bright-1.3	QQNNEDPRT	4.51	61.8	3.7	97.8
Bright-1.4	QQRSSYPLT	3.59	68.4	0.8	96.5
Bright-2.1	WQGTHFPQT	7.24	44.5	5.7	95.8
Bright-2.2	QQGQSYPWT	4.50	71.3	1.0	98.8
Bright-2.3	LQYASSPYT	3.12	70.7	2.0	98.6
Bright-2.4	FQGSHVPWT	2.58	47.3	3.8	95.0

[a]The frequencies of the top two V_H and V_L full-length sequences of a particular CDR3 group. [b]The V_H and V_L homologies were determined by calculating the pairwise identity by multiple sequence alignment of all V genes that shared the same CDR3.

Supplementary Table 8. Occurrence of the highest frequency CDRH3s from the bone marrow plasma cell repertoire of mice immunized with C1s, and their relative frequency in mice immunized with adjuvant or different antigens.

Antigen	**C1s - 1**	C1s -2	Br-1	Br-2	OVA-1	OVA-2	Adjuv- 1	Adjuv- 2
Total Seq Reads	**11,595**	9,071	9,453	11,769	15,350	15,751	6,681	16,743
CDRH3	**%**	%	%	%	%	%	%	%
GNYYYAMDY	**7.93**	0.01	0.01	0.01	0.00	0.00	0.04	0.01
DDGYWYFDV	**3.00**	0.00	0.12	0.00	0.01	0.16	0.00	0.28
YYYGSSAMDY	**2.68**	0.15	0.24	0.29	0.00	3.22	0.00	0.00
DMISYWYFDV	**2.64**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DGYDWYFDV	**2.26**	0.06	0.00	0.05	0.02	0.46	0.00	0.07
DDYDWYFDV	**1.97**	0.03	1.53	0.01	0.02	0.10	0.09	2.11
GSSYYAMDY	**1.67**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EDYGNYWYFDV	**1.67**	0.21	0.41	0.40	7.17	0.53	0.00	0.01
QGYDYDPYAMDY	**1.22**	0.00	0.00	0.00	0.00	0.00	0.75	0.00
EGYYYGSSYFDY	**1.17**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TGFDY	**1.11**	0.00	0.00	0.02	0.01	0.01	0.00	0.01
KGSTTATYFDY	**1.06**	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Antigen	C1s - 1	**C1s -2**	Br-1	Br-2	OVA-1	OVA-2	Adjuv-1	Adjuv-2
Total Seq Reads	11,595	**9,071**	9,453	11,769	15,350	15,751	6,681	16,743
CDRH3	%	**%**	%	%	%	%	%	%
SDRYDGYFDY	0.09	**10.99**	0.04	0.01	0.00	0.00	0.00	0.00
SDRFDGYFDY	0.05	**9.93**	0.05	0.03	0.00	0.00	0.00	0.00
WLLLAY	0.00	**3.30**	0.01	0.02	0.00	0.00	0.00	0.00
YGNYFDY	0.02	**2.39**	0.01	0.01	0.00	0.02	0.00	0.02
SDGYYYFDY	0.02	**1.65**	0.00	0.00	0.00	0.00	0.00	0.00
SGGNYDAMDY	0.00	**1.17**	0.02	0.00	0.00	0.00	0.00	0.00
YYDYDKAYYFDY	0.03	**1.15**	0.00	0.00	0.00	0.00	0.00	0.00

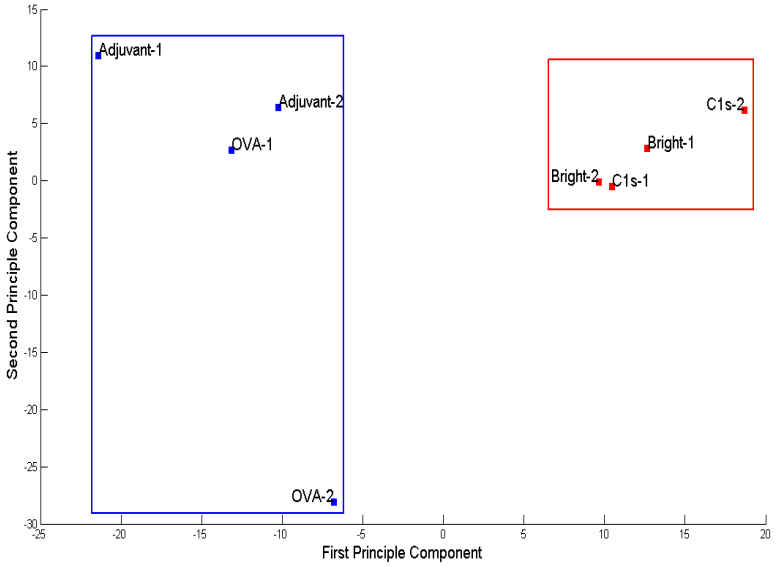
Nature Biotechnology: doi: 10.1038/nbt.1673

Supplementary Table 9. Occurrence of the highest frequency CDRL3s from the bone marrow plasma cell repertoire of mice immunized with C1s, and their relative frequency in mice immunized with adjuvant or different antigens.

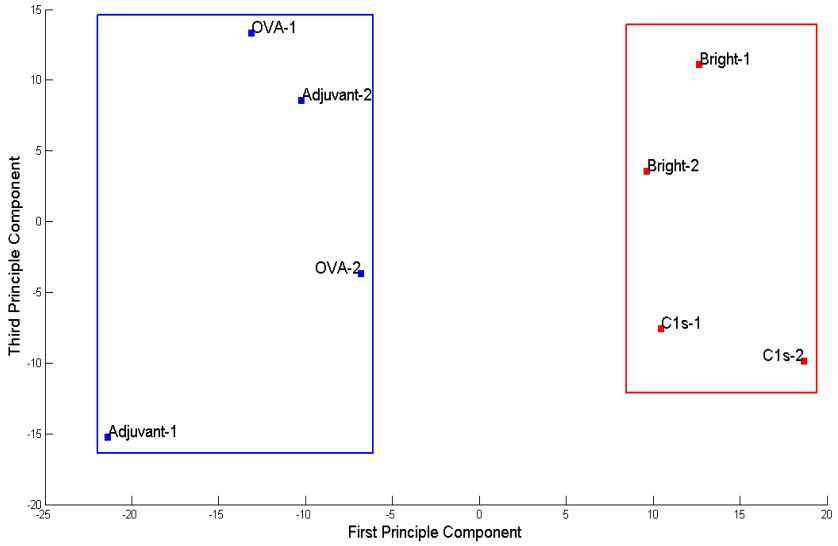
Antigen	**C1s - 1**	C1s -2	Br-1	Br-2	OVA-1	OVA-2	Adjuv-1	Adjuv-2
Total Seq Reads	**13,972**	14,664	12,228	10,452	13,355	17,200	7,112	21,241
CDRL3	**%**	%	%	%	%	%	%	%
WQGTHFPQT	**12.95**	1.56	2.33	7.25	1.66	1.63	1.69	1.54
QQWSSYPQLT	**6.94**	0.01	0.00	0.00	0.00	0.00	0.00	0.00
QNDHSYPLT	**3.81**	0.04	0.13	0.04	0.11	0.12	0.18	0.10
QQGQSYPWT	**3.16**	0.48	3.21	4.50	0.67	1.07	0.65	1.42
FQGSHVPWT	**2.76**	0.83	1.04	2.59	0.37	0.28	0.44	0.30
QQGQSYPFT	**2.64**	0.10	0.39	0.12	0.01	0.00	0.00	0.01
LQHGESPFT	**2.43**	1.55	2.59	2.41	0.00	0.02	0.01	0.01
QQGQSYPLT	**2.07**	1.05	2.97	0.11	0.25	0.26	0.17	0.58
QQSKEVPPT	**1.74**	0.53	0.08	0.11	0.07	0.04	0.13	0.07
QQHYSTPWT	**1.47**	1.41	0.36	0.82	0.88	2.01	1.42	1.13
QNDYSFT	**1.46**	0.04	0.01	0.05	0.00	0.00	0.00	0.00
WQGTHFPWT	**1.17**	0.68	1.55	0.99	0.54	0.83	0.93	0.95
WQGTHFPRT	**1.12**	1.15	4.73	0.94	1.27	1.23	1.18	2.70
FQGSHVPFT	**0.97**	0.14	0.47	0.47	0.01	0.00	0.00	0.00
WQGTHFPT	**0.95**	0.04	0.03	0.00	0.00	0.01	0.00	0.00

Antigen	C1s - 1	**C1s -2**	Br-1	Br-2	OVA-1	OVA-2	Adjuv-1	Adjuv-2
Total Seq Reads	13,972	**14,664**	12,228	10,452	13,355	17,200	7,112	21,241
CDRL3	%	**%**	%	%	%	%	%	%
FQGSHVPLT	0.19	**17.10**	2.29	0.23	0.16	0.23	0.24	0.19
QQSNEDPWT	0.23	**2.62**	0.50	1.00	0.28	0.32	0.24	0.23
WQGTHFPH	0.01	**2.20**	0.00	0.00	0.00	0.00	0.00	0.00
WQGTHFPT	0.00	**2.11**	0.00	0.07	0.02	0.02	0.04	0.00
QQHYSTPFT	0.19	**1.64**	0.17	0.15	0.00	0.00	0.00	0.01
WQGTHFPQT	12.95	**1.56**	2.33	7.25	1.66	1.63	1.69	1.54
LQHGESPFT	2.43	**1.55**	2.59	2.41	0.00	0.02	0.01	0.01
LQGSHVPLT	0.01	**1.50**	0.01	0.00	0.01	0.00	0.00	0.00
QQHYSTPWT	1.47	**1.41**	0.36	0.82	0.88	2.01	1.42	1.13
LQHGESPYT	0.44	**1.23**	0.84	0.66	0.25	0.44	0.76	0.61
SQSTHVPWT	0.89	**1.21**	1.68	1.30	0.50	0.51	0.32	0.28
WQGTHFPRT	1.12	**1.15**	4.73	0.94	1.27	1.23	1.18	2.70

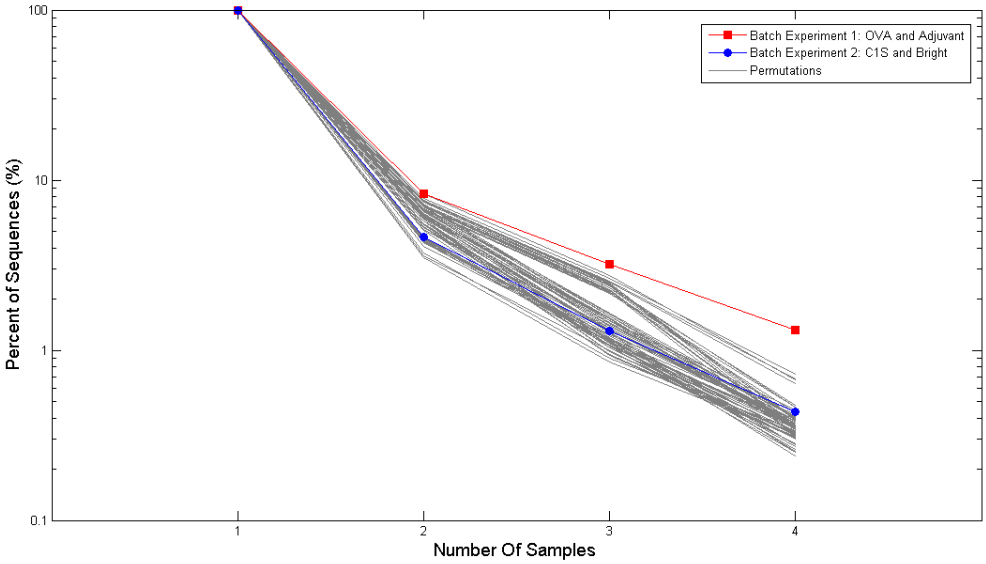
A. First Principle Component Vs. Second Principle Component



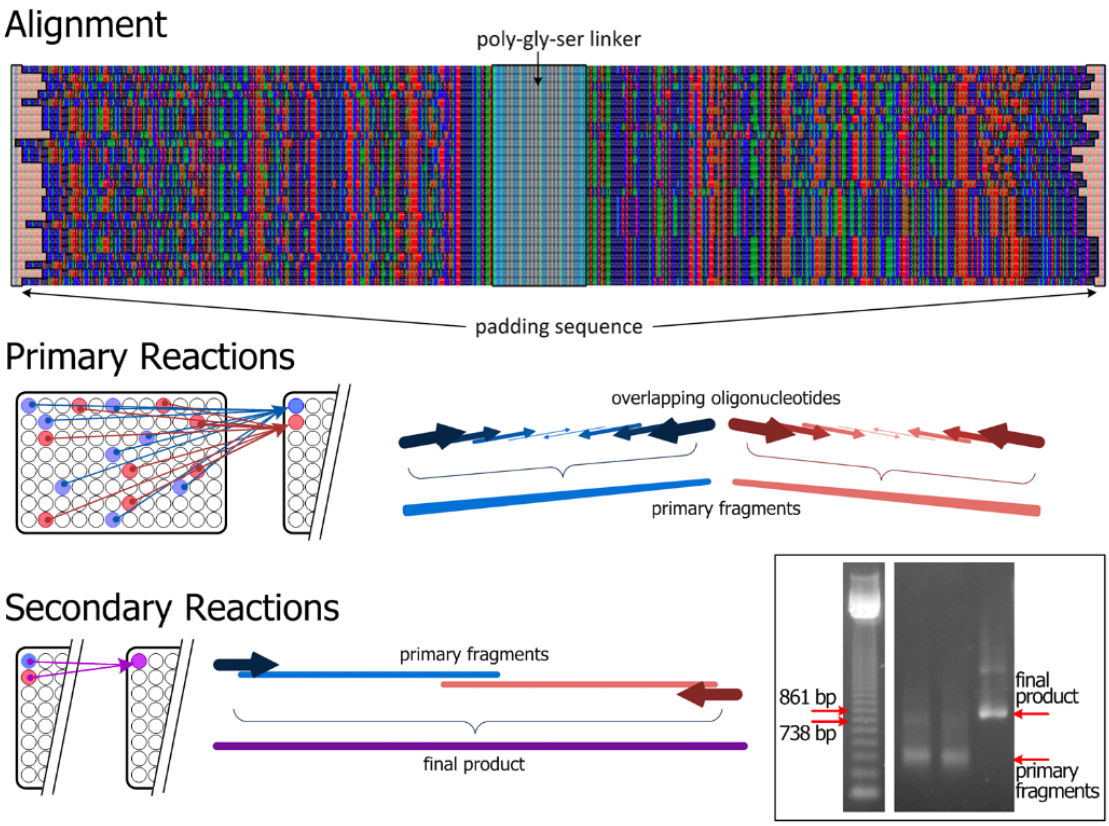
B. First Principle Component Vs. Third Principle Component



Supplementary Figure 7. Principal component analysis (PCA) of CDRH3 sequences from bone marrow plasma cell repertoires of different mouse groups. Ovalbumin (OVA) and Adjuvant only mice were one immunization group (derived from the same cage and same litter) while C1s and Bright mice were another immunization group. First principal component analysis identified two main cluster groups, blue and red, which represent different experiments (i.e. immunizations carried out on different dates).



Supplementary Figure 8. Percentage of CDRH3s distributed across subsets of four mouse populations. The percentage of common CDRH3 sequences between all combinations of four mice chosen from a set of eight. The percent similarity between mice immunized on the same day with either ovalbumin (OVA) or adjuvant is shown in red; percent similarity between mice immunized on the same day with either Bright or C1s is shown in blue; gray represents the similarity for every other possible combination.



Supplementary Figure 9. Construction of synthetic antibody genes. Highly represented V_H and V_L genes from bone marrow plasma cell repertoires were synthesized as single chain variable fragments (scFvs) by joining V genes with a poly-gly-ser linker. **Alignment**. The poly-gly-ser linker serves as the anchor point for aligning the set of desired genes. Appropriate restriction sites are added to facilitate cloning and the sequences are padded to uniform length to enable a single overlapping oligonucleotide assembly scheme for building all genes. **Primary assembly.** Primary fragments are generated from overlapping sets of oligonucleotides using inside-out nucleation PCR (scheme right). **Secondary assembly.** The second step of the assembly is a conventional overlap-extension PCR joining the primary fragments together to form the final product (scheme right). The primary PCR products are diluted with water by the liquid-handling robot and a portion of each diluted primary reaction is added to the secondary reaction (example left). **Gel image.** Agarose gel of typical scFv assembly products. First lane: DNA ladder; second and third lanes; primary products at ~400 bp; fourth lane: final product at ~810 bp.

Supplementary Table 10. Antibody single chain variable fragments (scFvs) identified by by combinatorial pairing of top four V_L and V_H genes.

C1s-1 V_L-V_H Pairing	ELISA Signal[a]
1L-1H	4.36
2L-1H	19.92
3L-2H	4.36
4L-3H	6.8
4L-4H	3.7
C1s-2	
1L-1HB	63.3
3L-1HB	4.65
2L-2H	3.1
3L-2H	4.22
3L-3H	8.8

[a]ELISA signal correlates to *E.coli* whole cell lysates expressing scFvs, measured as the OD_{450} signal of scFvs binding to antigen-coated wells relative to wells coated with unrelated antigen (bovine serum albumin).

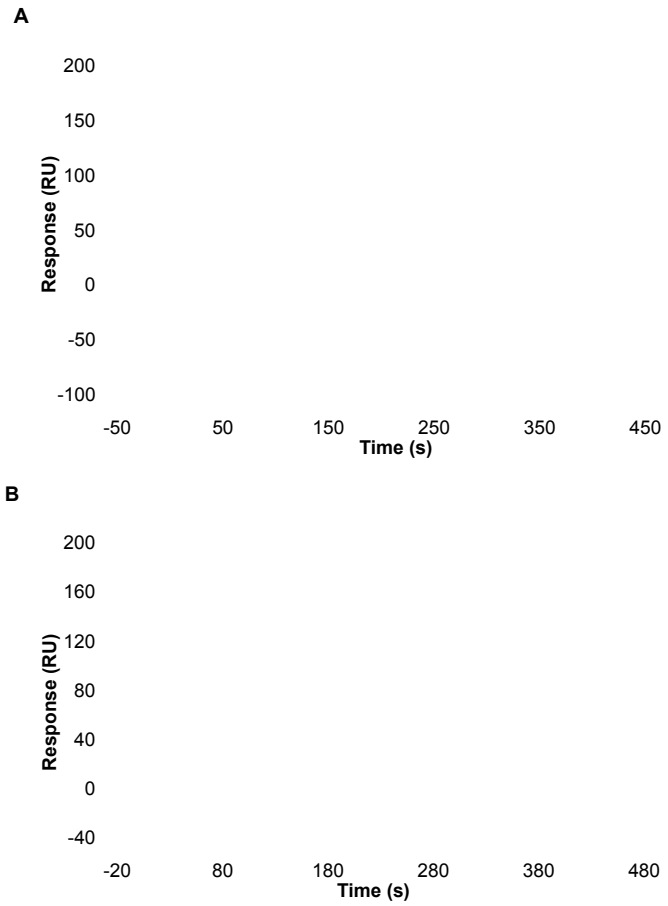
Supplementary Table 11. Serum IgG titers in mice immunized with different antigens.

Antigen	IgG Titer (serum dilution)
Ovalbumin - 1	1:5,000
Ovalbumin - 2	1:5,000
C1s - 1	1:10,000
C1s - 2	1:50,000
Bright - 1	1:10,000
Bright - 2	1:10,000

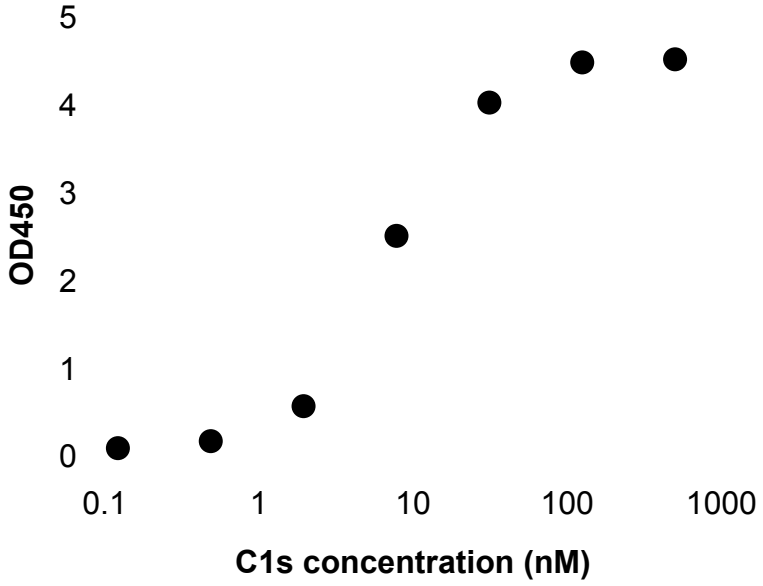
IgG titer was determined by maximum mouse serum dilution that gave an ELISA signal above background (binding with pre-immunized mouse serum).

Supplementary Table 12. Biophysical characterization of antibody single chain variable fragments (scFv) and IgGs derived by mining the bone marrow plasma cell repertoire of mouse C1s-2, which displayed the highest serum IgG titer.

V_L-V_H pair	2.1L-2.1HB (scFv)	2.1L-2.1HB (IgG)	2.3L-2.2H (IgG)
% V_L	17.10	17.10	2.20
% V_H	9.93	9.93	3.30
CDRL3	FQGSHVPLT	FQGSHVPLT	WQGTHFPH
CDRH3	SDRFDGYFDY	SDRFDGYFDY	WLLLAY
k_{on} (M^{-1} sec^{-1})	2.3×10^4	2.4×10^4	4.5×10^5
k_{off} (sec^{-1})	5.0×10^{-4}	1.2×10^{-3}	1.9×10^{-4}
K_D (nM)	20	50	0.43



Supplementary Figure 10. Kinetic binding analysis of purified anti-C1s IgGs by Surface Plasmon Resonance (Biacore). (A) Anti-Cls 2.1L-2.1HB was injected onto a chip with immobilized C1s at 25nM, 50nM, 100nM or 200nM and **(B)** As above for anti-C1s 2.3L-2.2H injected at 2.625nM, 5.25nM, 10.5nM or 21nM.



Supplementary Figure 11. Detection of C1s by sandwich ELISA using antibodies derived by mining bone marrow plasma cell repertoires. Anti-C1s scFv 2.1L-2.1HB was coated on the plate and used as the capture antibody and anti-C1s IgG 2.3L-2.2H was used as a detection antibody.

Supplementary Figure 12. Immunoprecipitation of C1s from human serum by using antibodies derived by mining bone marrow plasma cell repertoires. Anti-C1s IgG 2.3L-2.2H was used to capture C1s in human serum, following binding on Protein-A agarose beads. Western blot analysis was performed with anti-C1s scFv 2.1L-2.1HB as the primary antibody following by detection with anti-polyHis-HRP. Lane 1: 100 kDa and 70 kDa M.W. markers; lane 2: no capture antibody; lanes 3 and 4: capture with 1.5 ug/ml and 3 ug/ml 2.3L-2.2H antibody, respectively.

Supplementary Table 13. Primer mixes for PCR amplification of V_L and V_H genes.

V_L -Forward Primers	**Sequence**
YarivL-FOR1	AGC CGG CCA TGG CGG AYA TCC AGC TGA CTC AGC C
YarivL-FOR 2	AGC CGG CCA TGG CGG AYA TTG TTC TCW CCC AGT C
YarivL-FOR 3	AGC CGG CCA TGG CGG AYA TTG TGM TMA CTC AGT C
YarivL-FOR 4	AGC CGG CCA TGG CGG AYA TTG TGY TRA CAC AGT C
YarivL-FOR 5	AGC CGG CCA TGG CGG AYA TTG TRA TGA CMC AGT C
YarivL-FOR 6	AGC CGG CCA TGG CGG AYA TTM AGA TRA MCC AGT C
YarivL-FOR 7	AGC CGG CCA TGG CGG AYA TTC AGA TGA YDC AGT C
YarivL-FOR 8	AGC CGG CCA TGG CGG AYA TYC AGA TGA CAC AGA C
YarivL-FOR 9	AGC CGG CCA TGG CGG AYA TTG TTC TCA WCC AGT C
YarivL-FOR 10	AGC CGG CCA TGG CGG AYA TTG WGC TSA CCC AAT C
YarivL-FOR 11	AGC CGG CCA TGG CGG AYA TTS TRA TGA CCC ART C
YarivL-FOR 12	AGC CGG CCA TGG CGG AYR TTK TGA TGA CCC ARA C
YarivL-FOR 13	AGC CGG CCA TGG CGG AYA TTG TGA TGA CBC AGK C
YarivL-FOR 14	AGC CGG CCA TGG CGG AYA TTG TGA TAA CYC AGG A
YarivL-FOR 15	AGC CGG CCA TGG CGG AYA TTG TGA TGA CCC AGW T
YarivL-FOR 16	AGC CGG CCA TGG CGG AYA TTG TGA TGA CAC AAC C
YarivL-FOR 17	AGC CGG CCA TGG CGG AYA TTT TGC TGA CTC AGT C
YarivL-FOR 1 Lambda	AGC CGG CCA TGG CGG ARG CTG TTG TGA CTC AGG AAT C
V_L - Reverse Primers	
YarivL-REV 1	GAT GGT GCG GCC GCA GTA CGT TTG ATT TCC AGC TTG G
YarivL-REV 2	GAT GGT GCG GCC GCA GTA CGT TTT ATT TCC AGC TTG G
YarivL-REV 4	GAT GGT GCG GCC GCA GTA CGT TTT ATT TCC AAC TTT G
YarivL-REV 5	GAT GGT GCG GCC GCA GTA CGT TTC AGC TCC AGC TTG G
YarivL-REV 1 Lambda	GAT GGT GCG GCC GCA GTA CCT AGG ACA GTC AGT TTG G
YarivL-REV Lambda 2	GAT GGT GCG GCC GCA GTA CCT AGG ACA GTG ACC TTG G

V_H - Forward Primers	
YarivH-FOR 1	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA KGT RMA GCT TCA GGA GTC
YarivH-FOR 2	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA GGT BCA GCT BCA GCA GTC
YarivH-FOR 3	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GCA GGT GCA GCT GAA GSA STC
YarivH-FOR 4	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA GGT CCA RCT GCA ACA RTC
YarivH-FOR 5	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GCA GGT YCA GCT BCA GCA RTC
YarivH-FOR 6	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GCA GGT YCA RCT GCA GCA GTC
YarivH-FOR 7	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GCA GGT CCA CGT GAA GCA GTC
YarivH-FOR 8	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA GGT GAA SST GGT GGA ATC
YarivH-FOR 9	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA VGT GAW GYT GGT GGA GTC
YarivH-FOR 10	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA GGT GCA GSK GGT GGA GTC
YarivH-FOR 11	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA KGT GCA MCT GGT GGA GTC
YarivH-FOR 12	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA GGT GAA GCT GAT GGA RTC
YarivH-FOR 13	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA GGT GCA RCT TGT TGA GTC
YarivH-FOR 14	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA RGT RAA GCT TCT CGA GTC
YarivH-FOR 15	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA AGT GAA RST TGA GGA GTC
YarivH-FOR 16	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GCA GGT TAC TCT RAA AGW GTS TG

Nature Biotechnology: doi: 10.1038/nbt.1673

YarivH-FOR 17	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GCA GGT CCA ACT VCA GCA RCC
YarivH-FOR 18	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA TGT GAA CTT GGA AGT GTC
YarivH-FOR 19	GTT ATT GCT AGC GGC TCA GCC GGC AAT GGC GGA GGT GAA GGT CAT CGA GTC
V_H - Reverse Primers	
YarivH-REV 1	CCC TTG AAG CTT GCT GAG GAA ACG GTG ACC GTG GT
YarivH-REV 2	CCC TTG AAG CTT GCT GAG GAG ACT GTG AGA GTG GT
YarivH-REV 3	CCC TTG AAG CTT GCT GCA GAG ACA GTG ACC AGA GT
YarivH-REV 4	CCC TTG AAG CTT GCT GAG GAG ACG GTG ACT GAG GT