Supplementary Information

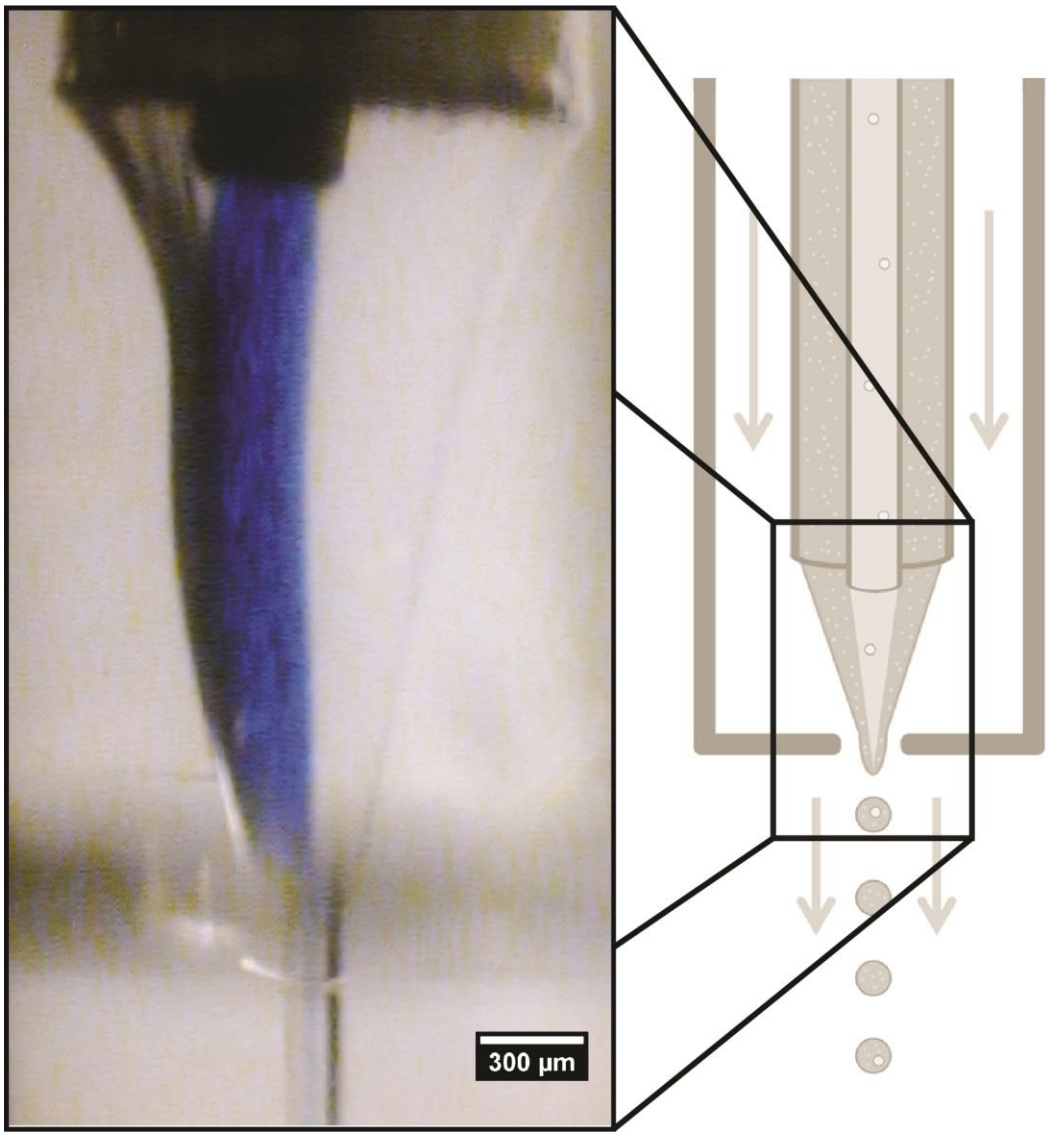
In-Depth Determination and Analysis of the Human Paired Heavy and Light Chain Antibody Repertoire

Brandon J. DeKosky[1], Takaaki Kojima[1,2], Alexa Rodin[1], Wissam Charab[1], Gregory C. Ippolito[3], Andrew D. Ellington[4], George Georgiou[1,3,5,6,*]

1. Department of Chemical Engineering, University of Texas at Austin, Austin, Texas, United States of America
2. Laboratory of Molecular Biotechnology, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan
3. Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas, United States of America
4. Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas, United States of America
5. Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA.
6. Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas, United States of America

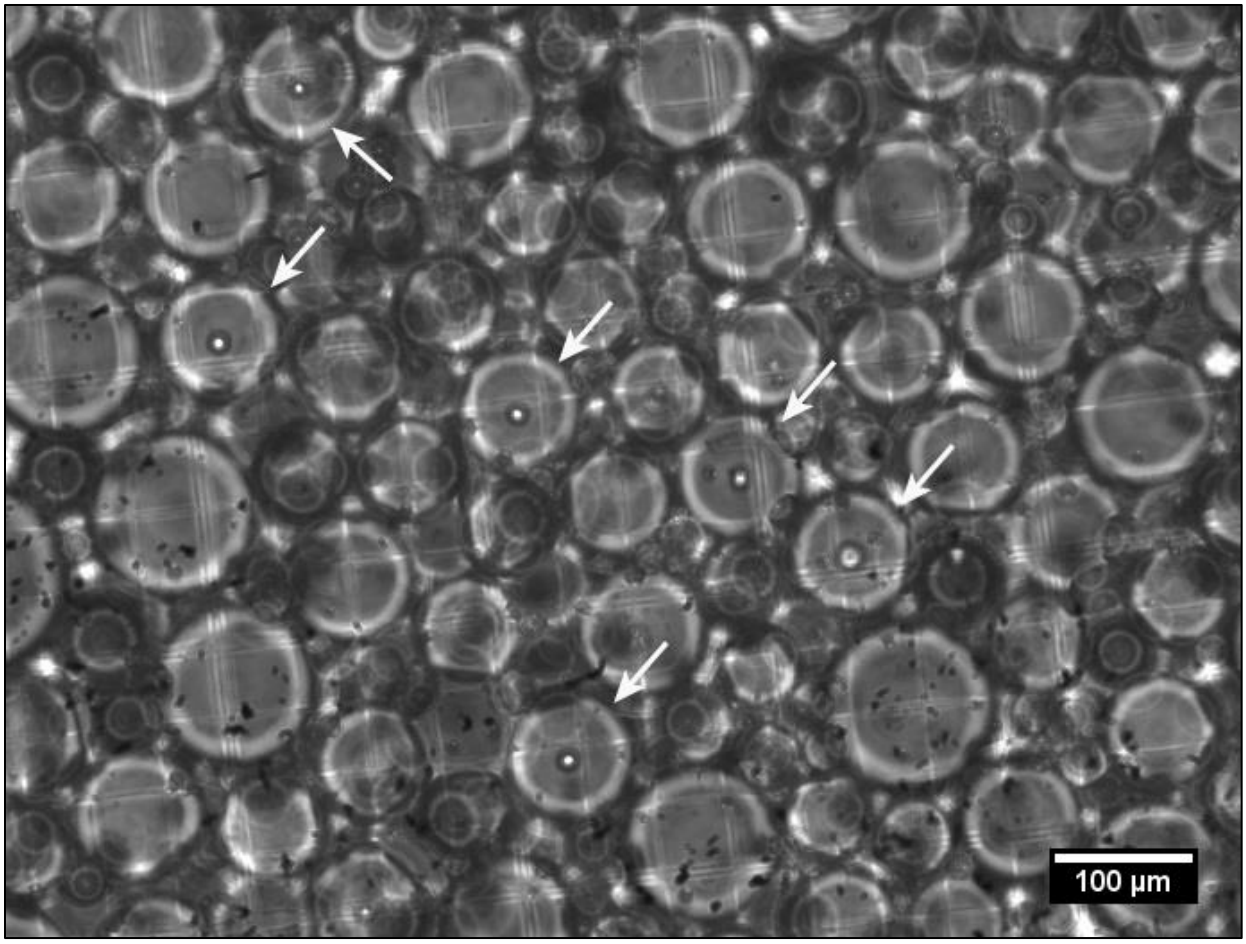
*Correspondence should be addressed to G.G. (gg@che.utexas.edu)

Nature Medicine: doi:10.1038/nm.3743



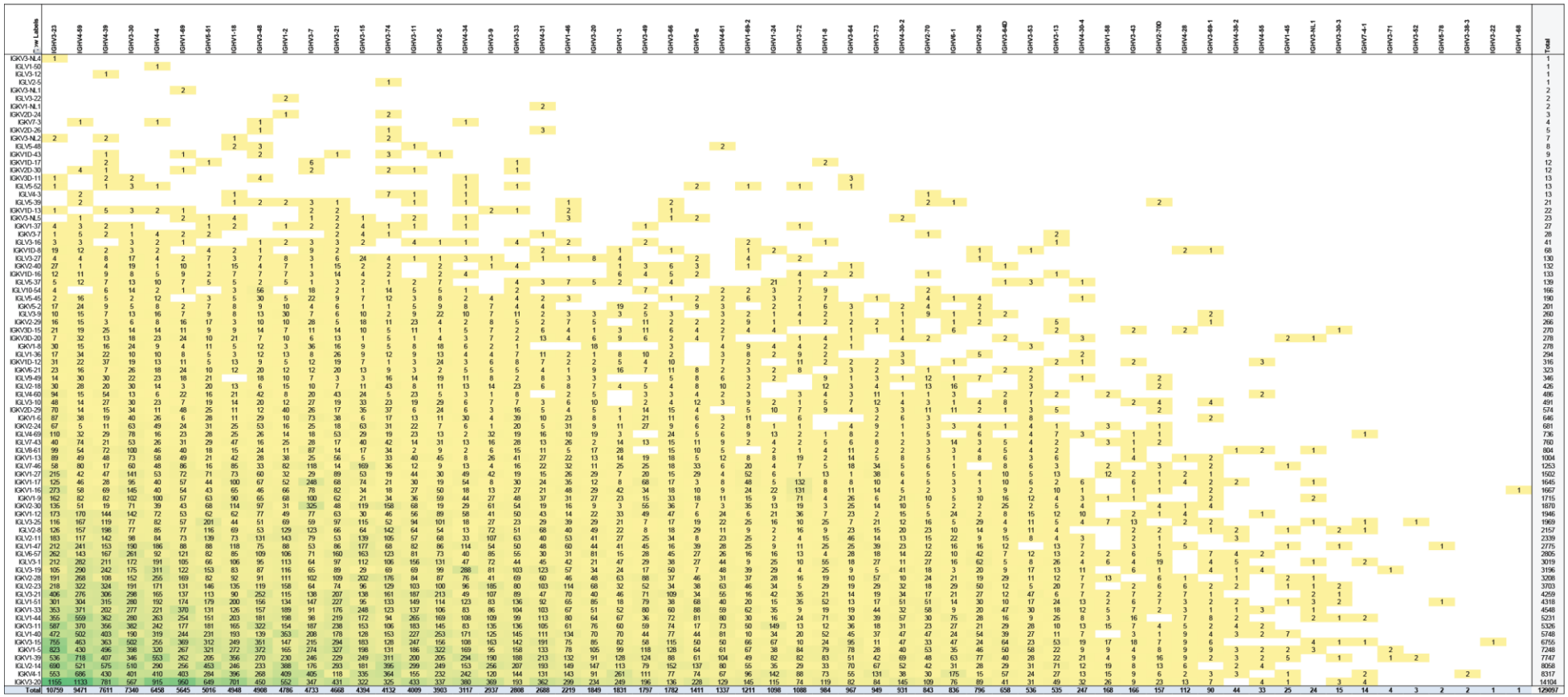
Supplementary Figure 1 A micrograph of the axisymmetric flow-focusing nozzle during emulsion generation (left), placed in context of the diagram from Figure 1a (right), where PBS/0.4% Trypan blue exits the inner needle and cell lysis buffer exits the outer needle.

Nature Medicine: doi:10.1038/nm.3743



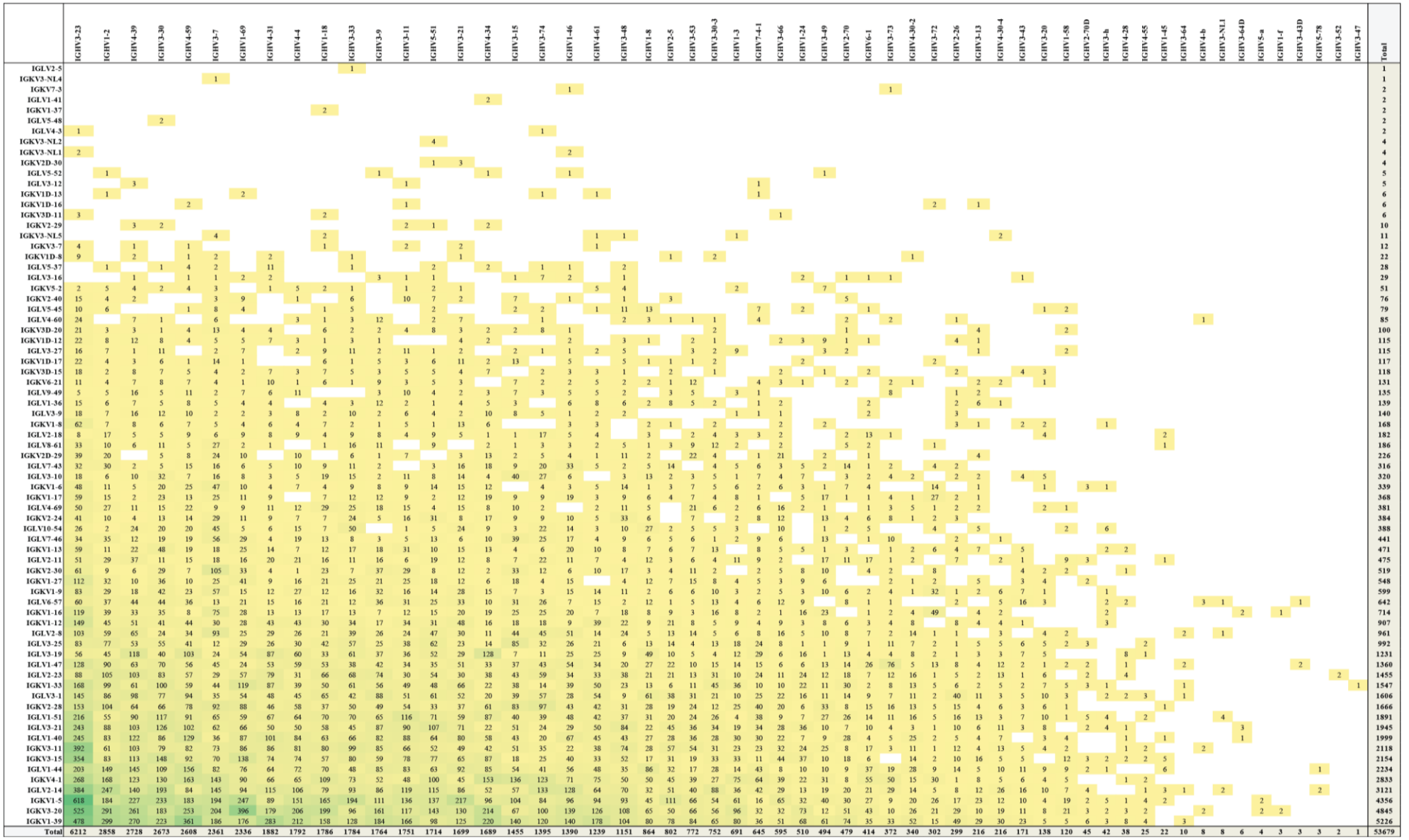
Supplementary Figure 2 MOPC-21 immortalized B cells encapsulated in emulsion droplets. The outer aqueous stream that normally contains cell lysis buffer (Fig. 1a, gray solution) was replaced with 0.4% Trypan blue in PBS to examine cell viability throughout the flow focusing and emulsification process. Emulsified cell viability was approximately 90% and cell viability did not differ substantially from non-emulsified controls.

Nature Medicine: doi:10.1038/nm.3743



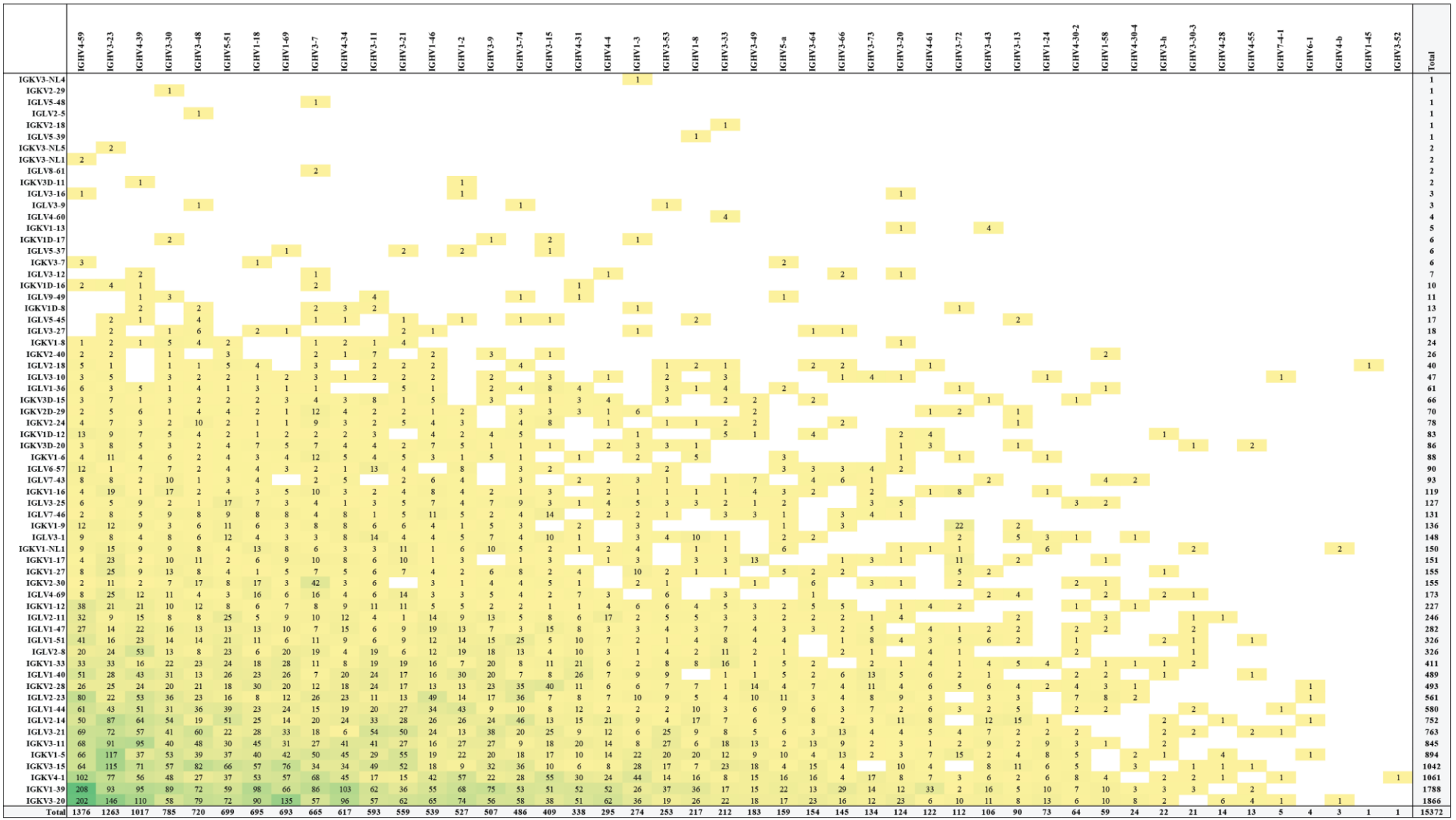
Supplementary Figure 3 Heat map of V-gene usage for 129,097 VH:VL clusters recovered from Donor 1. Sequences were collected using primers targeting the framework 1 region; raw data is available in the online supplement.

Nature Medicine: doi:10.1038/nm.3743



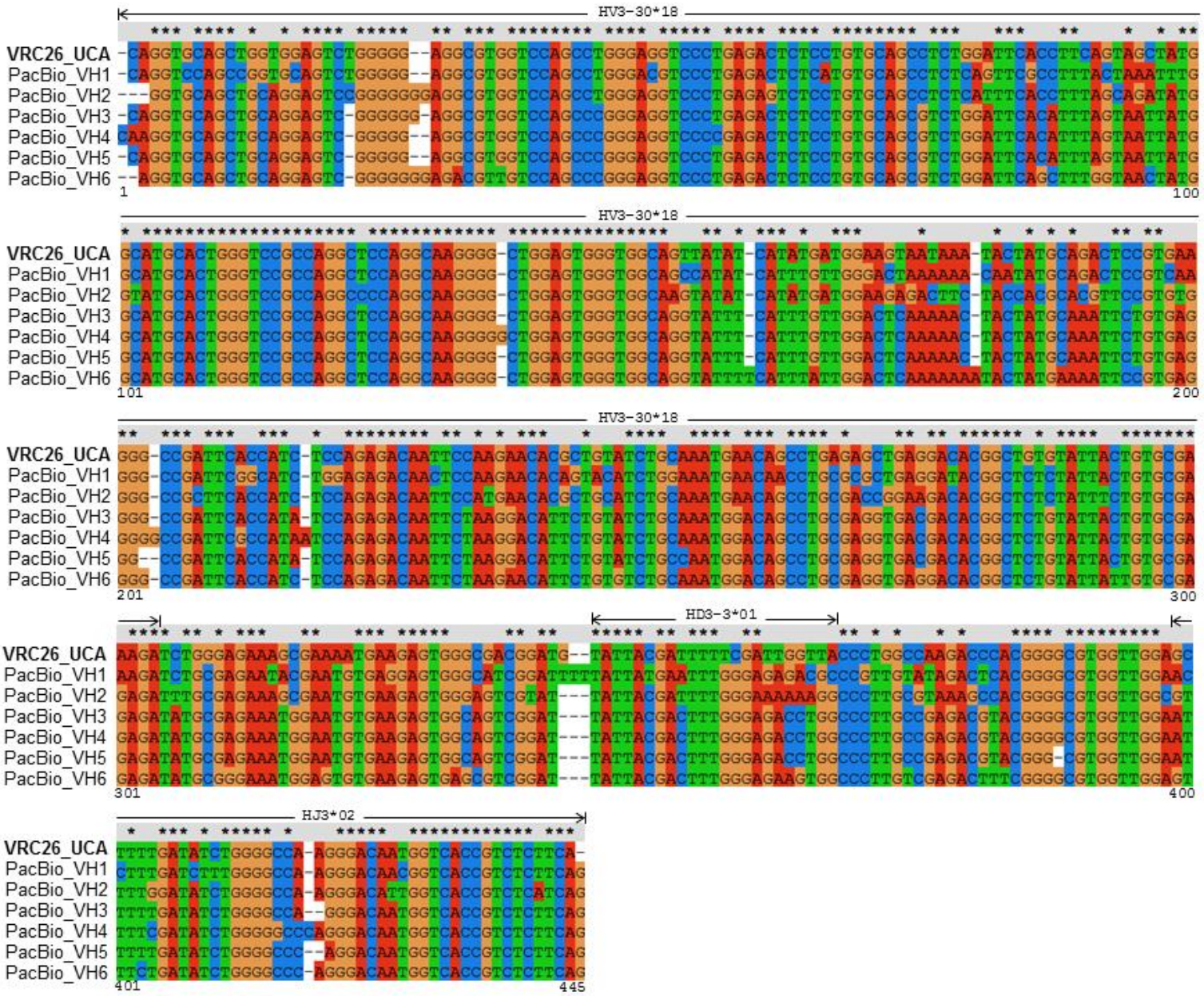
Supplementary Figure 4 Heat map of V-gene usage for 53,679 VH:VL clusters recovered from Donor 2. Sequences were collected using primers targeting the framework 1 region; raw data is available in the online supplement.

Nature Medicine: doi:10.1038/nm.3743



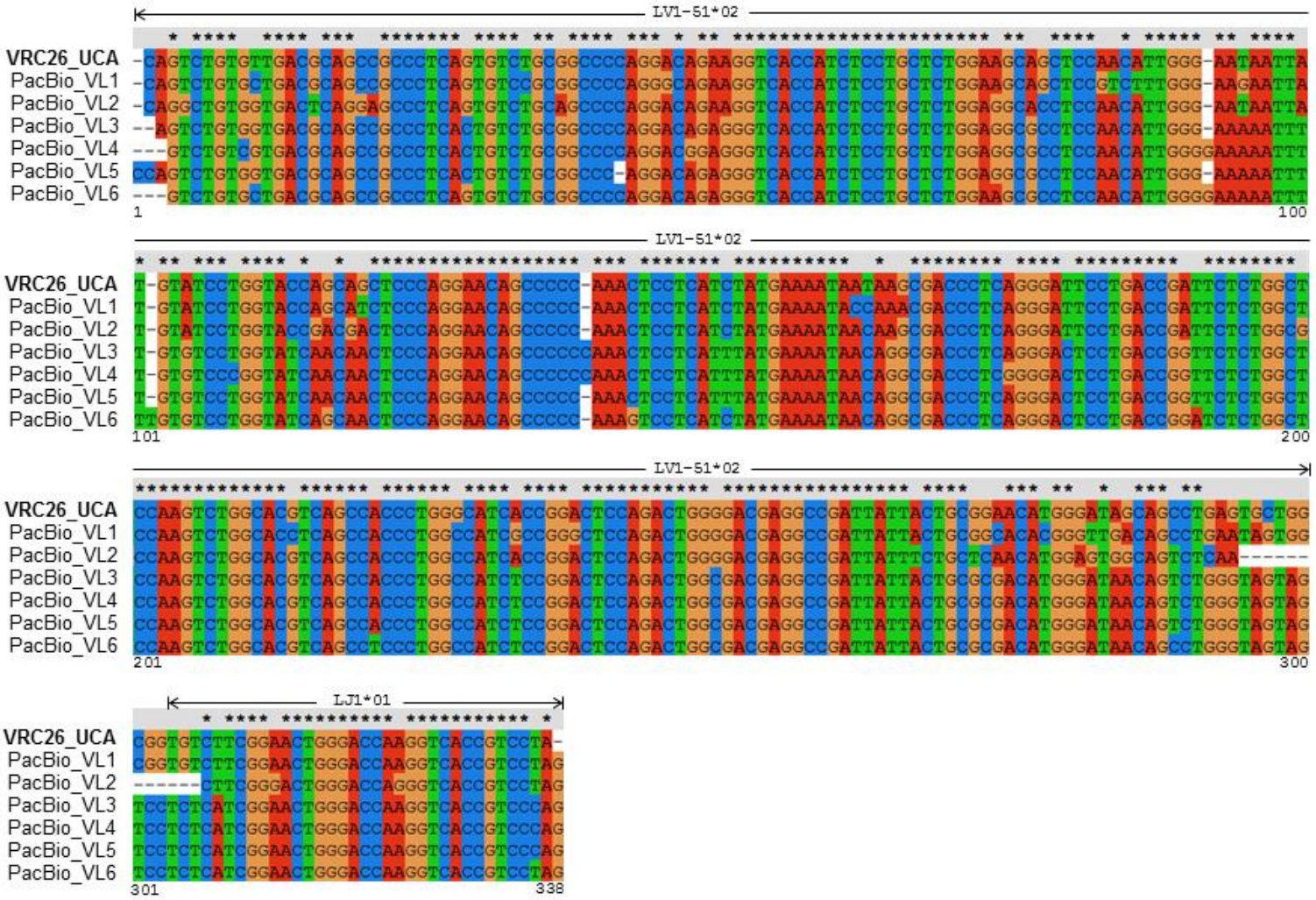
Supplementary Figure 5 Heat map of V-gene usage for 15,372 VH:VL clusters recovered from Donor 3. Sequences were collected using primers targeting the leader peptide region; raw data is available in the online supplement.

Nature Medicine: doi:10.1038/nm.3743



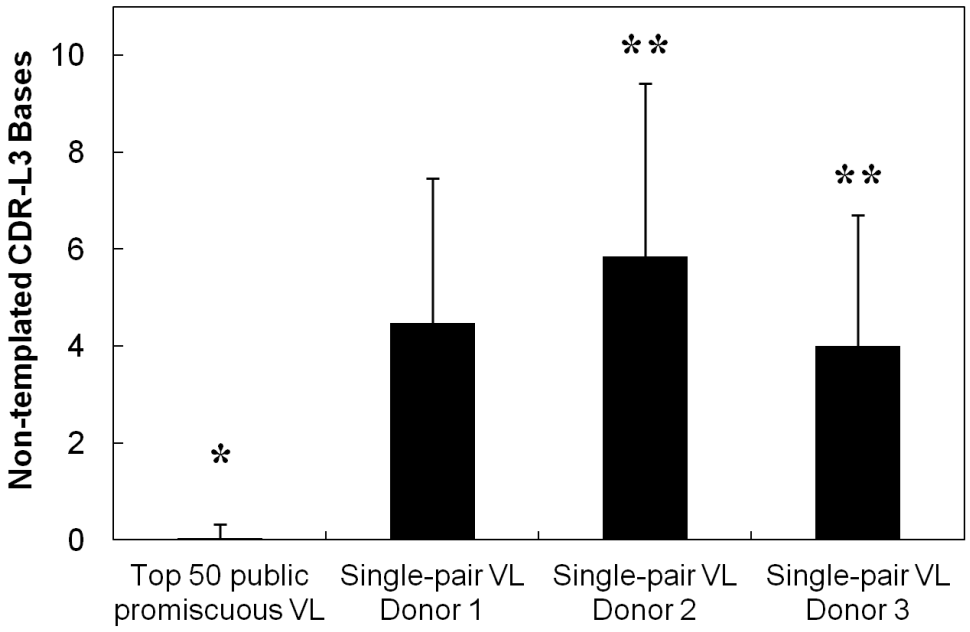
Supplementary Figure 6 VH alignment of the six VRC26 HIV broadly neutralizing antibody variants recovered by PacBio sequencing of complete ~850bp VH:VL amplicons. Sequences were recovered from CD27^{+} peripheral B cells of the CAP256 donor and aligned to the VRC26 VH unmutated common ancestor (UCA, Doria-Rose et al., *Nature* 2014). Corresponding light chain variants are shown in Supplementary Figure 7.

Nature Medicine: doi:10.1038/nm.3743



Supplementary Figure 7 VL alignment of the six VRC26 HIV broadly neutralizing antibody variants recovered by PacBio sequencing of complete ~850bp VH:VL amplicons. Sequences were recovered from $CD27^{+}$ peripheral B cells of the CAP256 donor and aligned to the VRC26 VL unmutated common ancestor (UCA, Doria-Rose et al., *Nature* 2014). Corresponding heavy chain variants are shown in Supplementary Figure 6.

Nature Medicine: doi:10.1038/nm.3743



Supplementary Figure 8 Comparison of the number of non-templated bases (sum of somatic mutations and non-templated insertions) in the top 50 public, promiscuous VL nucleotide junctions shared by Donors 1, 2, and 3 to 50 randomly selected VL junctions paired with only a single heavy chain in the Donor 1, Donor 2, or Donor 3 repertoires (mean±s.d.). Statistical significance noted where $p < 0.05$ (* $p < 10^{-10}$ compared to all other groups, ** $p = 0.0043$).

Nature Medicine: doi:10.1038/nm.3743

Supplementary Table 1 VH:VL pairing analysis of a mixture of HEK293 cells transfected with 11 different known antibodies. The maximum read count for each row and column is highlighted; 11/11 antibodies were identified and paired correctly in this control experiment. Read count variation was expected due to varying transfection & expression efficiency for the 22 distinct heavy and light chain plasmids, and antibody clones #10 and 11 exhibited notable VH-VL imbalance by total read counts. The signal:topVLnoise ratio (the relevant parameter for native pair assignment, see Supplementary Table 2) averaged 35:1 overall and 87:1 if noise from light chains 10 and 11 (which showed VH-VL imbalance, see total VH and VL reads) was excluded.

		Heavy Chain											
		1H	**2H**	**3H**	**4H**	**5H**	**6H**	**7H**	**8H**	**9H**	**10H**	**11H**	**Total**
Light Chain	**1L**	**1,842**	4	20	13	18	16	39	20	49	6	4	**2,031**
	2L	0	**4,916**	34	31	59	41	102	127	146	28	8	**5,492**
	3L	0	2	**6,251**	9	38	25	116	60	118	13	2	**6,634**
	4L	21	27	75	**14,592**	81	158	348	189	397	75	51	**16,014**
	5L	5	15	97	41	**16,204**	99	192	231	277	86	19	**17,266**
	6L	2	12	92	37	64	**16,427**	358	180	404	62	23	**17,661**
	7L	9	13	218	72	112	180	**21,315**	203	1,320	78	45	**23,565**
	8L	4	39	85	71	242	145	365	**32,393**	506	79	72	**34,001**
	9L	4	29	182	105	116	186	1,335	323	**35,391**	109	46	**37,826**
	10L	12	24	944	189	1,597	1,080	3,519	1,898	4,291	**8,535**	98	**22,187**
	11L	32	66	1,153	272	1,258	1,655	6,405	6,567	6,185	555	**14,126**	**38,274**
	Total	**1,931**	**5,147**	**9,151**	**15,432**	**19,789**	**20,012**	**34,094**	**42,191**	**49,084**	**9,626**	**14,494**	**220,951**

Nature Medicine: doi:10.1038/nm.3743

Supplementary Table 2 Accuracy statistics for human VH:VL paired analysis with an ARH-77 immortalized cell line control spike.

Estimated input human B cells	20,000
Estimated ARH-77 spiked cells	260
VH:VL Reads after CDR3 clustering	403,897
Recovered CDR-H3:CDR-L3 Clusters	1,751
Correct ARH-77 VH:VL Reads (Signal)	2,604
ARH-77 Top Incorrect VL Reads (topVLnoise)	27
ARH-77 2nd-Ranked Incorrect VL Reads	19
ARH-77 3rd-Ranked Incorrect VL Reads	16
ARH-77 Signal:topVLnoise Ratio*	96.4

* The key metric for VH:VL pair assignment (see main text)

Nature Medicine: doi:10.1038/nm.3743

Supplementary Table 3 Memory B cell counts before and after *in vitro* activation. Values must be considered rough estimates due to varying contributions of hemocytometer sampling, centrifugation/recovery cell loss, and cell death, stasis, and expansion over four days *in vitro*.

Sample	*FACS Count* **Fresh Bmems**	*Hemocytometer Count* **After 4d Activation**
Donor 1	1.8 million	1.6 million viable
Donor 2	1.1 million	1.3 million viable
Donor 3	347k	300k viable
ARH-77 spike experiment	87k	20k viable

Nature Medicine: doi:10.1038/nm.3743

Supplementary Table 4 Leader peptide overlap extension primers.

Conc (nM)	Primer ID	Primer Sequence
40	VH1_LP	tattcccatcgcggcgcACAGGTGCCCACTCCCAGGTGCAG
40	VH3_LP	tattcccatcgcggcgcAAGGTGTCCAGTGTGARGTGCAG
40	VH4/6_LP	tattcccatcgcggcgcCCCAGATGGGTCCTGTCCCAGGTGCAG
40	VH5_LP	tattcccatcgcggcgcCAAGGAGTCTGTTCCGAGGTGCAG
40	hVλ1for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAGGTCCTGGGCCCAGTCTGTGCTG
40	hVλ2for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAGGTCCTGGGCCCAGTCTGCCCTG
40	hVλ3for-2_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAYWCTGCACAGGCTCTGTGACCTCCTAT
40	hVλ4/5for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAGGTCTCTCTCSCAGCYTGTGCTG
40	hVλ6for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAGTTCTTGGGCCAATTTTATGCTG
40	hVλ7for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAGGTCCAATTCYCAGGCTGTGGTG
40	hVλ8for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAGAGTGGATTCTCAGACTGTGGTG
40	hVκ1/2for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAATGAGGSTCCCYGCTCAGCTGCTGG
40	hVκ3for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCACTCTTCCTCCTGCTACTCTGGCTCCCAG
40	hVκ4for_LP	gcgccgcgatgggaataNNNNNNNNNNNNNNNNTCTGCTCGAGTTCGGTCAATTTCTCTGTTGCTCTGGATCTCTG

Nature Medicine: doi:10.1038/nm.3743