Supplementary Information

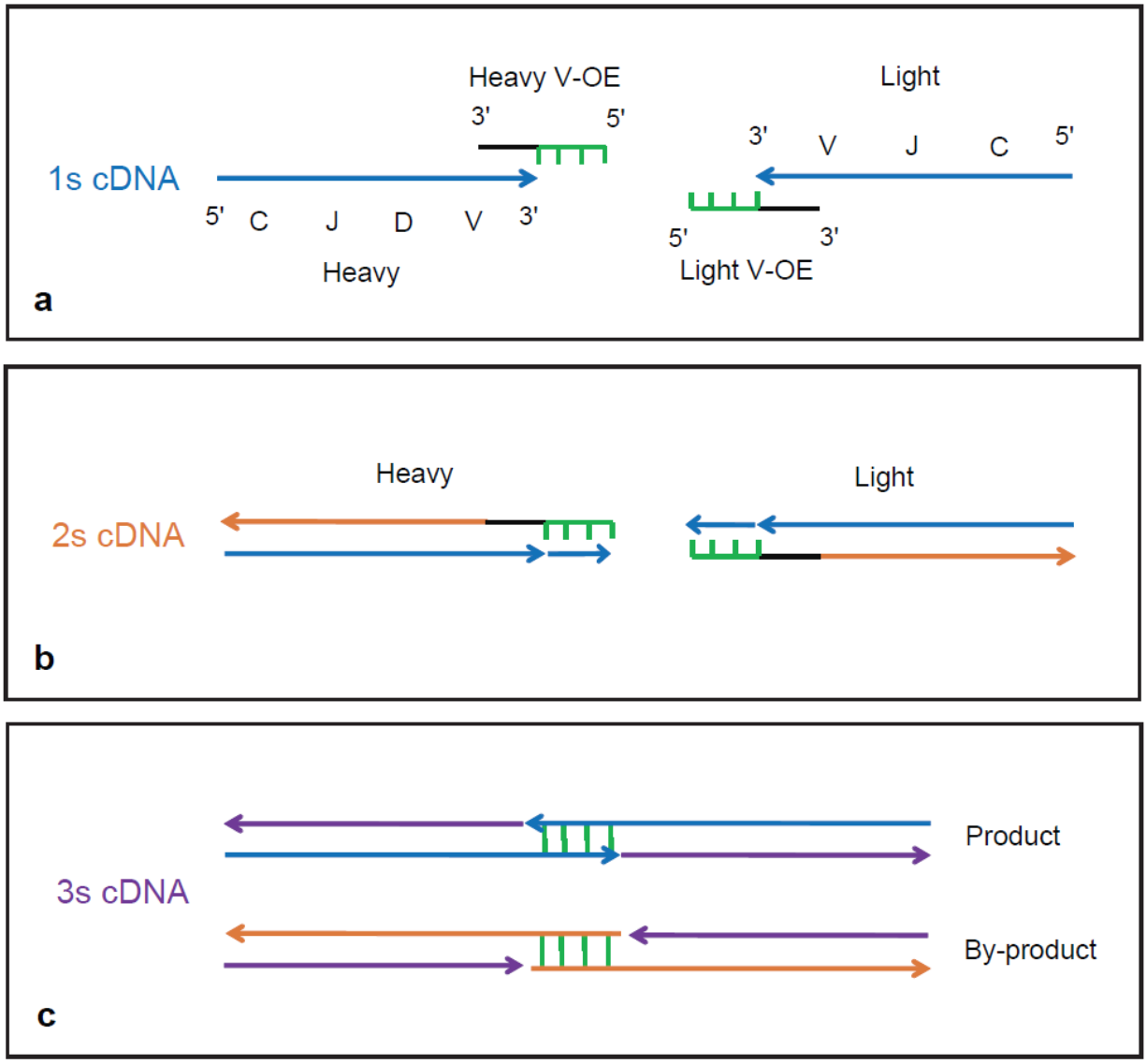
High-Throughput Sequencing of the Paired Human Immunoglobulin Heavy and Light Chain Repertoire

Brandon J. DeKosky[1], Gregory C. Ippolito[2], Ryan P. Deschner[1], Jason J. Lavinder[3], Yariv Wine[1], Brandon M. Rawlings[1], Navin Varadarajan[4], Claudia Giesecke[5], Thomas Dörner[5], Sarah F. Andrews[6], Patrick C. Wilson[6], Scott P. Hunicke-Smith[3], C. Grant Willson[1,7], Andrew D. Ellington[3,7], George Georgiou[1,2,3,8]

[1]Department of Chemical Engineering, University of Texas at Austin, Austin, Texas, United States of America
[2]Section of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, Texas, United States of America
[3]Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, United States of America
[4]Department of Chemical & Biomolecular Engineering, University of Houston, Houston, Texas, United States of America
[5]Department of Medicine, Rheumatology, and Clinical Immunology, Charité University Medicine, Berlin, and DRFZ Berlin, Germany
[6]Section of Rheumatology, Department of Medicine, University of Chicago, Chicago, Illinois, United States of America
[7]Department of Chemistry & Biochemistry, University of Texas at Austin, Austin, Texas, United States of America
[8]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 Biotechnology: doi:10.1038/nbt.2492



Supplementary Figure 1 An overview of the linkage (overlap extension) RT-PCR process. a) V-region primers (black) with a 5' complementary heavy/light overlap region (green) anneal to first strand cDNA. b) Second strand cDNA is formed by 5' to 3' extension; the overlap region is incorporated into all cDNA. c) After denaturation, heavy and light chains with first strand sense anneal to generate a complete 850 bp product through 5' to 3' extension. The CDR-H3 and CDR-L3 are located near the outside of the final linked construct to allow CDR3 analysis by 2x250 paired-end Illumina sequencing. Linkage RT-PCR primer sequences are given in Supplementary Table 5 (V-region primers denoted "fwd-OE" and constant region primers denoted "rev-OE").

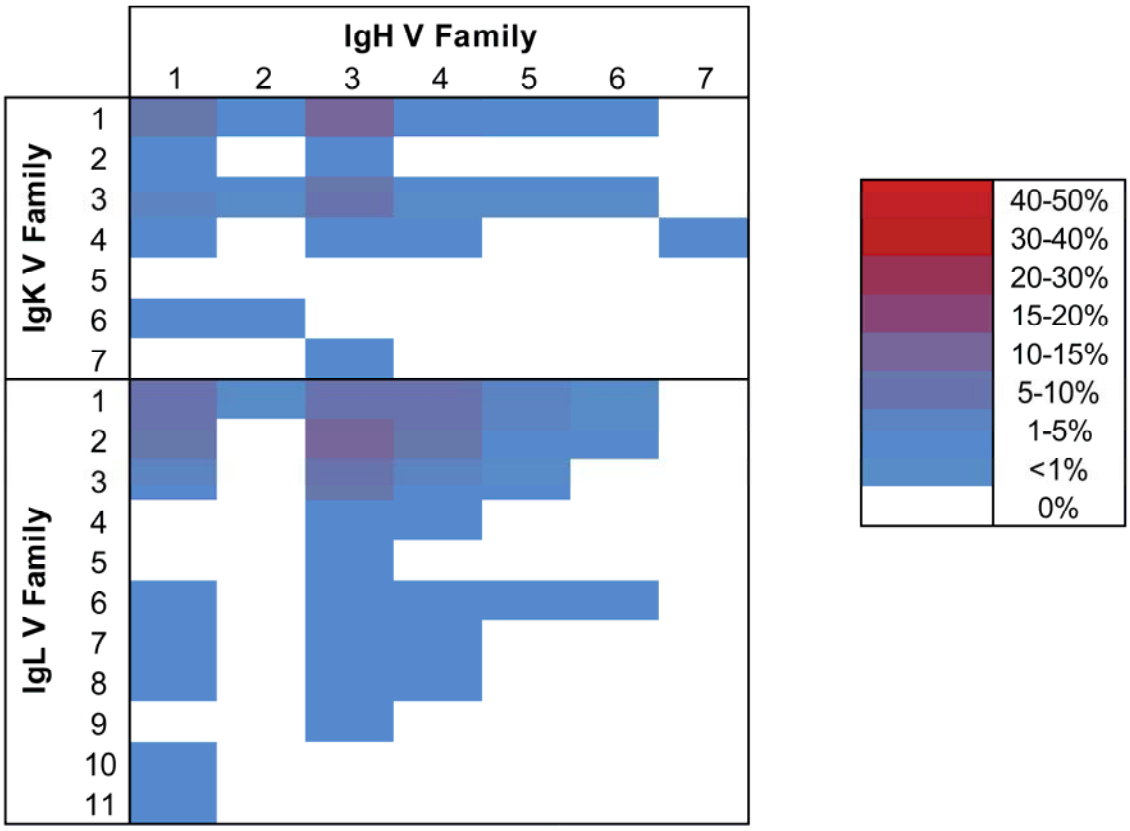
Nature Biotechnology: doi:10.1038/nbt.2492

Supplementary Table 1 Key statistics from several paired VH:VL repertoires. TD-tetanus toxoid/diphtheria toxoid, MSD-Merck Sharpe & Dohme

Immunization	n/a	Tetanus Toxoid (TD, MSD)	Influenza (2010-11 Fluvirin)
Cell Type	IgG^+ B lymphocytes	Day 7 post-TT boost TT^+ plasmablasts	Day 14 memory B cells
Fresh Cells vs. Freeze/Thaw	Fresh	Freeze/Thaw	Freeze/Thaw
Cell:Well Ratio	1:10	1:425	1:39
% cells as single cells	95.1%	99.9%	98.7%
Unique CDR-H3 Recovered	2,716	86	240
Control Cell Spike	IM-9	ARH-77	IM-9
Accuracy Ratio[1]	78:1	650:1	942:1

[1] For known spiked cells, (reads correct VL):(reads top incorrect VL)

Nature Biotechnology: doi:10.1038/nbt.2492



Supplementary Figure 2 A heat map of VH:VL pairings from IgG$^+$ class-switched peripheral B cells isolated from a healthy volunteer (n=2,248). The experiment presented here is a replicate of Fig. 2a using donated blood from a different individual.

Nature Biotechnology: doi:10.1038/nbt.2492

Supplementary Table 2 Key statistics for the IgG^+ VH:VL pairing experiment from a second volunteer (Supplementary Fig. 2).

Immunization	n/a
Cell Type	IgG^+ B lymphocytes
Fresh Cells vs. Freeze/Thaw	Fresh
Cell:Well Ratio	1:10
% cells as single cells	95.1%
Unique CDR-H3 Recovered	2,248
Control Cell Spike	IM-9
Accuracy Ratio[1]	125:1

[1] For known spiked cells, (reads correct VL):(reads top incorrect VL)

Nature Biotechnology: doi:10.1038/nbt.2492

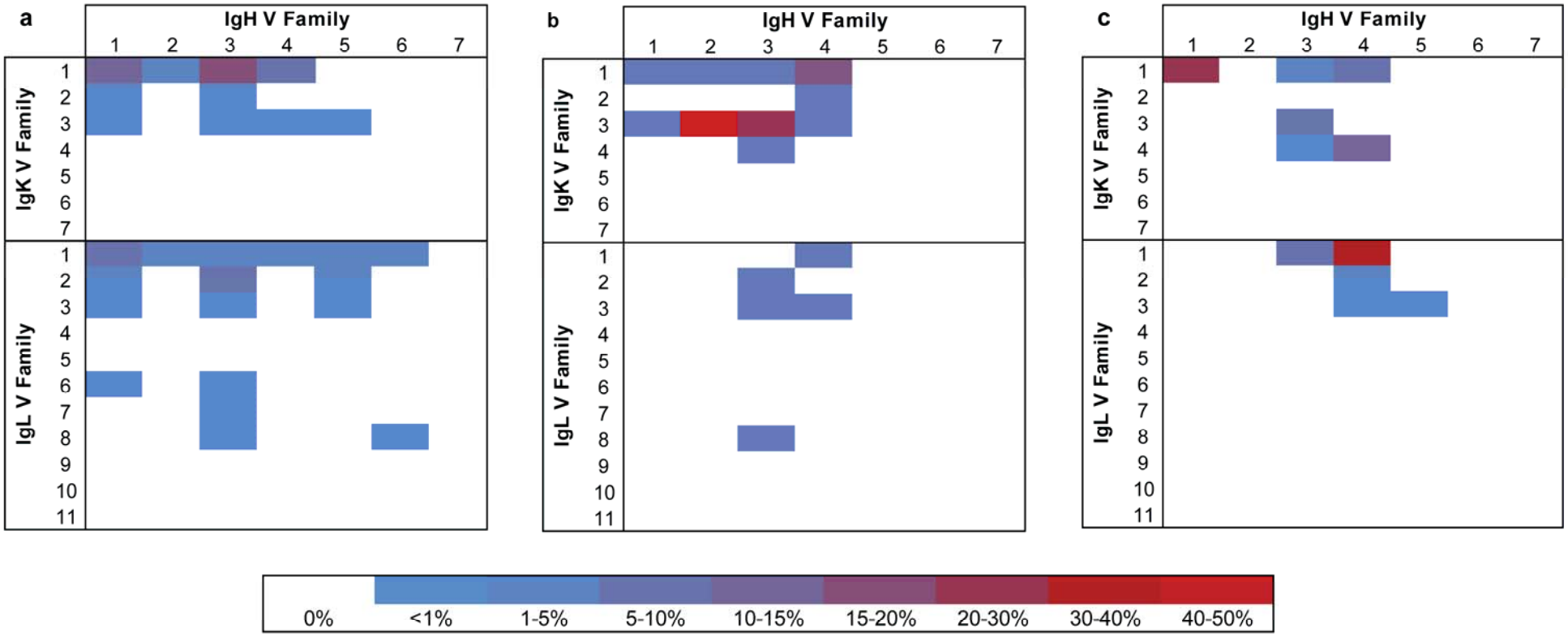
Supplementary Table 3 Analysis of overlapping heavy chain sequences and paired light chain sequences identified by both single cell RT-PCR and high-throughput VH:VL pairings in a memory B cell population isolated from an individual 14 days post-vaccination with the 2010-2011 trivalent FluVirin influenza vaccine.

Seq ID	Isotype	CDR-H3	Paired CDR-L3[1]	Source
2D02	IgM	gcgagaggcggaaatgggcgacccttgacaac	gcagcatgggatgacagcctgaatggttgggtg	Sanger scRT-PCR
2D02	IgM	gcgagaggcggaaatgggcgacccttgacaac	gcagcatgggatgacagcctgaatggttgggtg	MiSeq VH:VL
3D05	IgM	gcgagaaggtactttgactac	gnagcatgggatgacagcctgaatgtttggntg	Sanger scRT-PCR
3D05	IgM	gcgagaaggtactttgactac	gcagcatgggatgacagcctgaatgtttggctg	MiSeq VH:VL
1E02	IgG1	gcgcgacatggccctgcgggaaaaagcgcgtatggttttgatatc	cagtcctatgacagcggactgaatggttatgtggtc	Sanger scRT-PCR
1E02	IgG	gcgcgacatggccctgcgggaaaaagcgcgtatggttttgatatc	cagtcctatgacaacagactgaatggttatgtggtg	MiSeq VH:VL
3A01	IgG3	gcgagagtaatagcagctcgcgaccgccggatcactcctaactactaccgccctatggacgtc	caggtgtgggatagtagtagtgaccatcaggtg	Sanger scRT-PCR
3A01	IgG	gcgagagtaatagcagctcgcgaccgccggatcactcctaattactaccgccctatggacgtc	caggtgtgggacagtagtagtgatcatcaggtg	MiSeq VH:VL

[1] The 2D02 and 3D05 CDR-L3 sequences are highly similar but differ by two bases

Nature Biotechnology: doi:10.1038/nbt.2492

Nature Biotechnology: doi:10.1038/nbt.2492



Supplementary Figure 3 As Fig. 2b comprised the lowest sample size in Figure 2 (n=86 unique pairs, compared to Fig. 2a, n=2,716, and Fig. 2c, n=240) a simulation was performed to randomly select 86 VH:VL pairs from Fig. 2a and 2c and normalize all panels to 86 unique sequences. (a) healthy donor peripheral IgG^+ B cells, (b) day 7 tetanus-toxoid specific plasmablasts, and (c) day 14 post-influenza vaccination memory B cells. The simulation presented here facilitates comparison between panels a, b, and c.

Nature Biotechnology: doi:10.1038/nbt.2492

Supplementary Table 4 Statistical analysis of pairing accuracy

Experiment	*Resting IgG+ repertoire*	*Tetanus toxoid D7 plasmablasts*
Figure	2a	2b
Cell:Well Ratio	1:10	1:425
Fraction Single Cells[1]	0.951	0.999
Spiked Clone	IM-9	ARH-77
Spike %	4.0	7.5
Estimated # spiked cells	2,830	30
Estimated Spiked Cells as Single Cells[1]	2,691	30
Estimated Spiked Cells as 2-cells-per-well[1]	139	0
Paired Reads in Dataset	287,572	30,238
Correctly Paired Spike VH:VL Reads	14,805	871
Predicted Mispairing Rate[1]	4.9%	0.1%
Spike VH : Top Non-Spike VL Mispairing Rate	1.3%	0.15%
Top Non-Spike VH : Spike VL Mispairing Rate	7.8%	0.31%
Total Recovered VH:VL Pairs From Sample	2,716	86

[1] Calculated from the Poisson distribution (Supplementary Information, pg. 3)

Nature Biotechnology: doi:10.1038/nbt.2492

Supplementary Table 5 Overlap Extension (OE) RT-PCR primer mix

Conc. (nM)	Primer ID	Sequence
400	CHrev-AHX89	*CGCAGTAGCGGTAAACGGC*
400	CLrev-BRH06	*GCGGATAACAATTTCACACAGG*
40	hIgG-rev-OE-AHX89	*CGCAGTAGCGGTAAACGGC* AGGGYGCCAGGGGGAAGAC
40	hIgA-rev-OE-AHX89	*CGCAGTAGCGGTAAACGGC* CGGGAAGACCTTGGGGCTGG
40	hIgM-rev-OE-AHX89	*CGCAGTAGCGGTAAACGGC* CACAGGAGACGAGGGGGAAA
40	hIgKC-rev-OE-BRH06	*GCGGATAACAATTTCACACAGG* GATGAAGACAGATGGTGCAG
40	hIgLC-rev-OE-BRH06	*GCGGATAACAATTTCACACAGG* TCCTCAGAGGAGGGYGGGAA
40	hVH1-fwd-OE	TATTCCCATGGCGCGCCCAGGTCCAGCTKGTRCAGTCTGG
40	hVH157-fwd-OE	TATTCCCATGGCGCGCCCAGGTGCAGCTGGTGSARTCTGG
40	hVH2-fwd-OE	TATTCCCATGGCGCGCCCAGRTCACCTTGAAGGAGTCTG
40	hVH3-fwd-OE	TATTCCCATGGCGCGCCGAGGTGCAGCTGKTGGAGWCY
40	hVH4-fwd-OE	TATTCCCATGGCGCGCCCAGGTGCAGCTGCAGGAGTCSG
40	hVH4-DP63-fwd-OE	TATTCCCATGGCGCGCCCAGGTGCAGCTACAGCAGTGGG
40	hVH6-fwd-OE	TATTCCCATGGCGCGCCCAGGTACAGCTGCAGCAGTCA
40	hVH3N-fwd-OE	TATTCCCATGGCGCGCCTCAACACAACGGTTCCCAGTTA
40	hVK1-fwd-OE	GGCGCGCCATGGGAATAGCCGACATCCRGDTGACCCAGTCTCC
40	hVK2-fwd-OE	GGCGCGCCATGGGAATAGCCGATATTGTGMTGACBCAGWCTCC
40	hVK3-fwd-OE	GGCGCGCCATGGGAATAGCCGAAATTGTRWTGACRCAGTCTCC
40	hVK5-fwd-OE	GGCGCGCCATGGGAATAGCCGAAACGACACTCACGCAGTCTC
40	hVL1-fwd-OE	GGCGCGCCATGGGAATAGCCCAGTCTGTSBTGACGCAGCCGCC
40	hVL1459-fwd-OE	GGCGCGCCATGGGAATAGCCCAGCCTGTGCTGACTCARYC
40	hVL15910-fwd-OE	GGCGCGCCATGGGAATAGCCCAGCCWGKGCTGACTCAGCCMCC
40	hVL2-fwd-OE	GGCGCGCCATGGGAATAGCCCAGTCTGYYCTGAYTCAGCCT
40	hVL3-fwd-OE	GGCGCGCCATGGGAATAGCCTCCTATGWGCTGACWCAGCCAA
40	hVL-DPL16-fwd-OE	GGCGCGCCATGGGAATAGCCTCCTCTGAGCTGASTCAGGASCC
40	hVL3-38-fwd-OE	GGCGCGCCATGGGAATAGCCTCCTATGAGCTGAYRCAGCYACC
40	hVL6-fwd-OE	GGCGCGCCATGGGAATAGCCAATTTTATGCTGACTCAGCCCC
40	hVL78-fwd-OE	GGCGCGCCATGGGAATAGCCCAGDCTGTGGTGACYCAGGAGCC

Nature Biotechnology: doi:10.1038/nbt.2492

Supplementary Table 6 Nested PCR primers.

Conc. (nM)	Primer ID	Sequence
400	hIgG-all-rev-OEnested	ATGGGCCCTGSGATGGGCCCTTGGTGGARGC
400	hIgA-all-rev-OEnested	ATGGGCCCTGCTTGGGGCTGGTCGGGGATG
400	hIgM-rev-OEnested	ATGGGCCCTGGGTTGGGGCGGATGCACTCC
400	hIgKC-rev-OEnested	GTGCGGCCGCAGATGGTGCAGCCACAGTTC
400	hIgLC-rev-OEnested	GTGCGGCCGCGAGGGYGGGAACAGAGTGAC

Nature Biotechnology: doi:10.1038/nbt.2492

Supplementary Table 7 VH and VL Separate Amplification Primers

Conc. (nM)	Primer ID	Sequence
400	hIgG-all-rev-OEnested	ATGGGCCCTGSGATGGGCCCTTGGTGGARGC
400	hIgA-all-rev-OEnested	ATGGGCCCTGCTTGGGGCTGGTCGGGGATG
400	Linker-VHfwd-BC2	NNNNTGAAGGGGCTAGCTATTCCCATCGCGG
400	hIgKC-rev-OEnested	GTGCGGCCGCAGATGGTGCAGCCACAGTTC
400	hIgLC-rev-OEnested	GTGCGGCCGCGAGGGYGGGAACAGAGTGAC
400	Linker-VLfwd-BC2	NNNNTGAAGGGCGCCGCGATGGGAAT

Nature Biotechnology: doi:10.1038/nbt.2492