Supplemental materials for

Revealing how variations in antibody repertoires correlate with vaccine responses

Yana Safonova[1,2], Sung Bong Shin[3], Luke Kramer[4], James Reecy[4], Corey T. Watson[2], Timothy P.L. Smith[3], and Pavel A. Pevzner[1,*]

[1] Computer Science and Engineering Department, University of California San Diego, San Diego, CA, USA

[2] Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY, USA

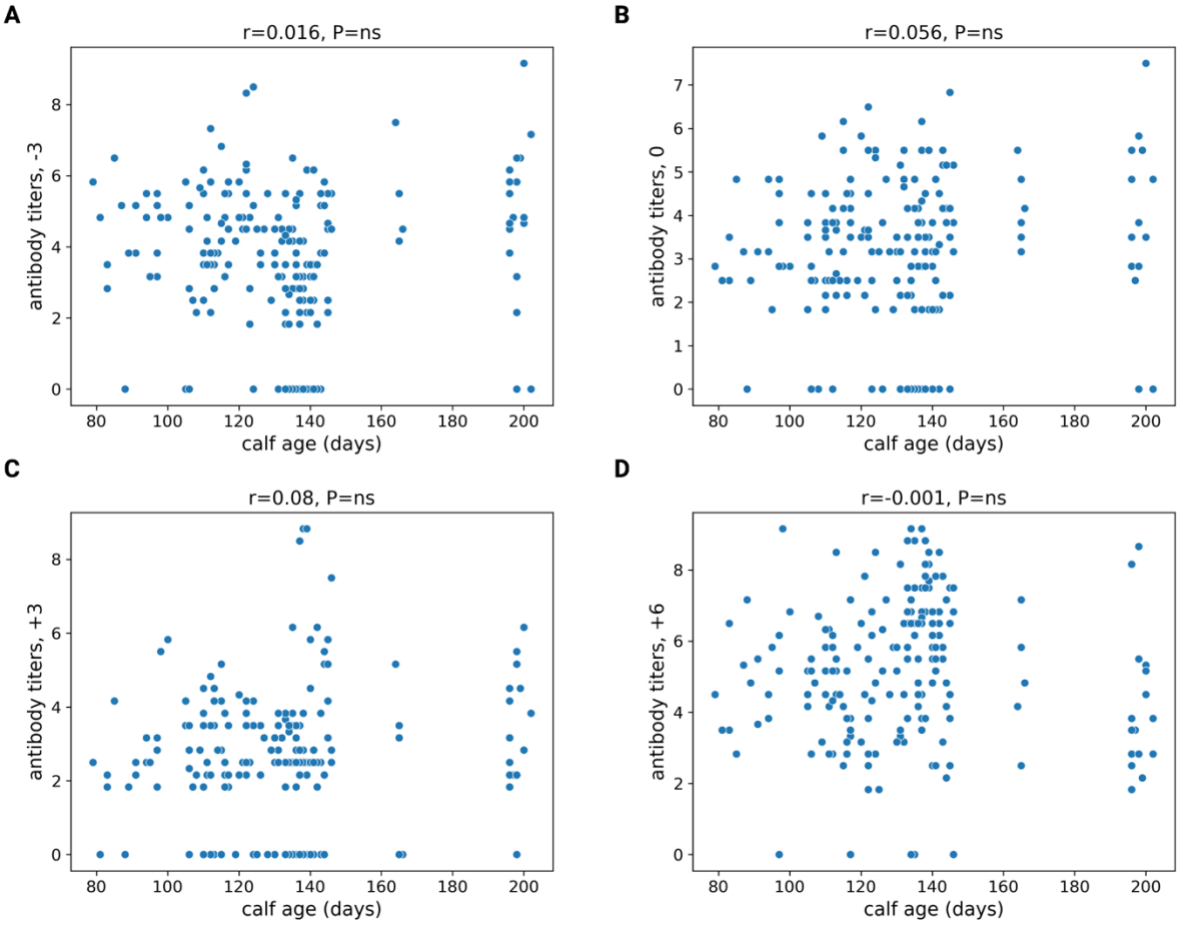
[3] U.S. Meat Animal Research Center, USDA-ARS, Clay Center, NE, USA

[4] Department of Animal Science, Iowa State University, Ames, IA, USA

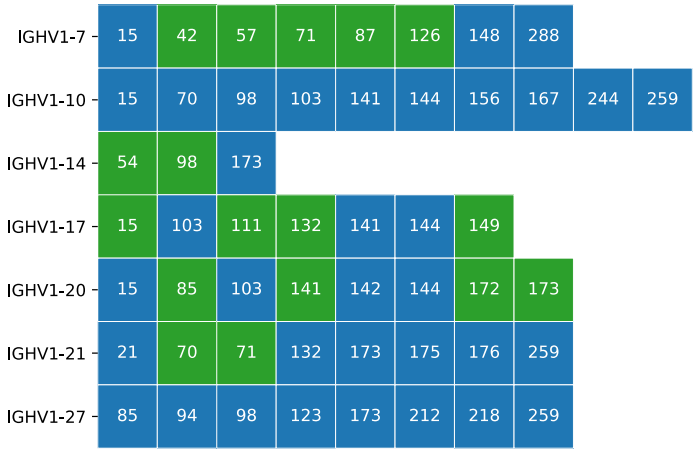
[*] Corresponding author: ppevzner@eng.ucsd.edu

The document includes Supplemental Figures S1–12, Supplemental Table S1, and Supplemental Methods.

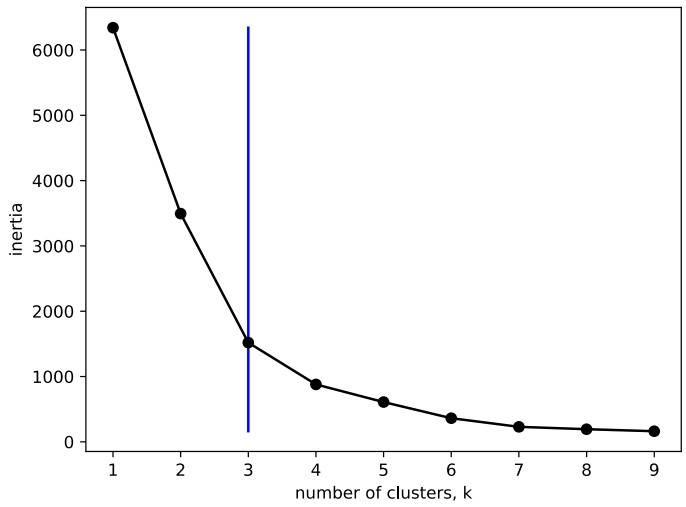
Supplemental Figures



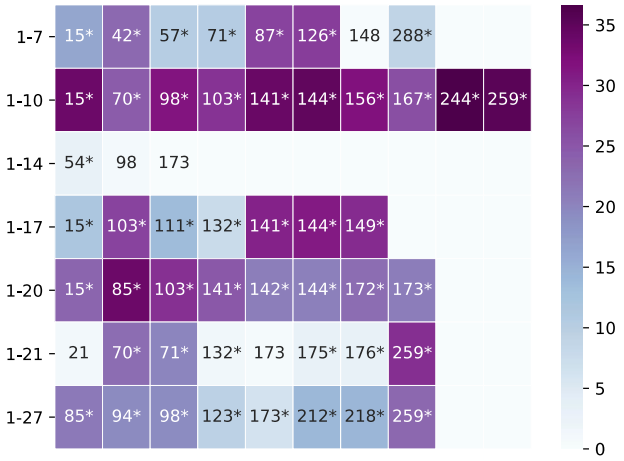
Supplemental Figure S1. Calf ages vs antibody titers at time points "–3" (*A*), "0" (*B*), "+3" (*C*), and "+6" (*D*).



Supplemental Figure S2. 52 identified GSVs. The number within a cell shows the nucleotide position of the GSV in the first IMGT allele of the gene. 17 green cells represent known germline variations. The remaining 35 blue cells represent either still unknown germline variations or frequent SHMs.



Supplemental Figure S3. Inertias of clusters computed via *k*-means clustering for the matrix shown in Figure 4A. The vertical blue line indicates the optimal number of clusters identified using the elbow method.



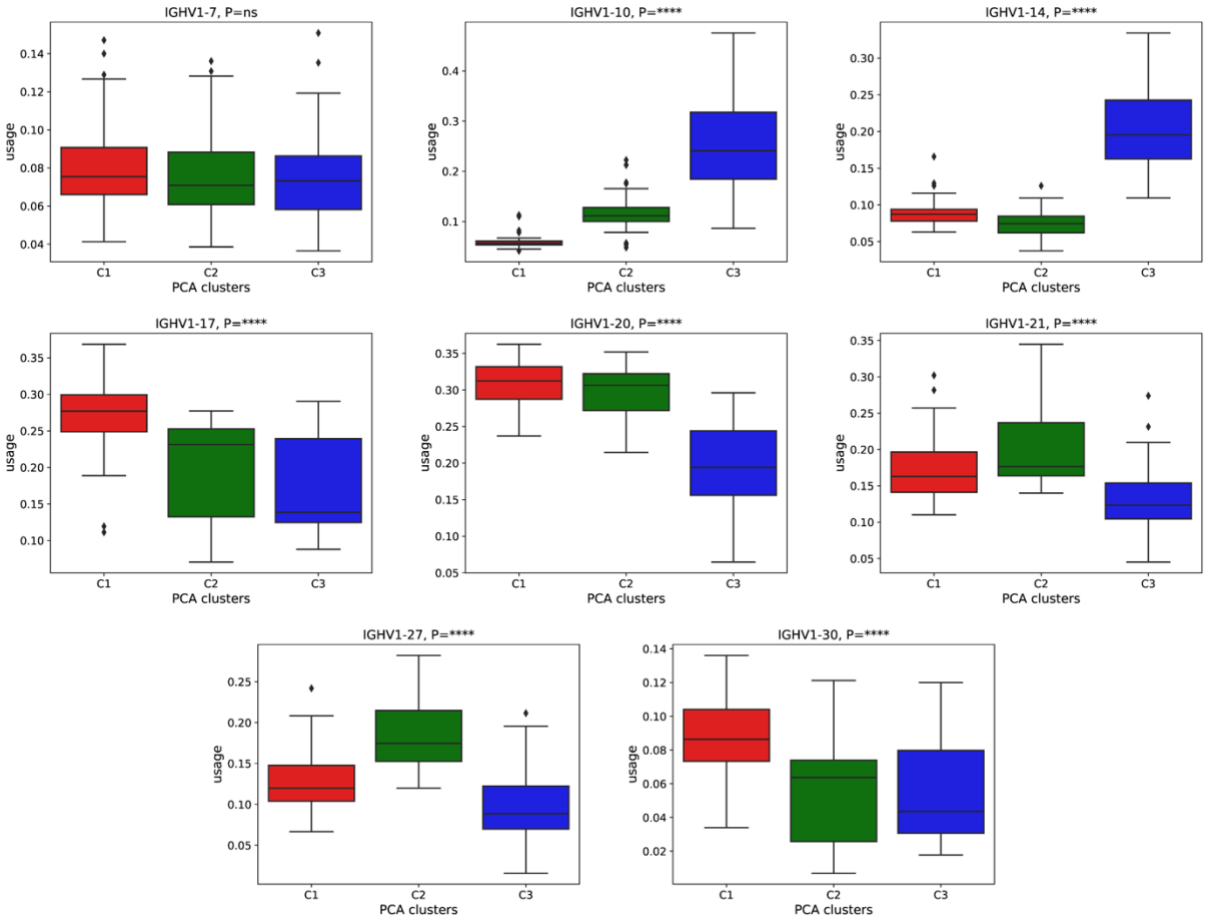
34

35 **Supplemental Figure S4. Associations of *R*-ratios of 52 GSVs of V genes with clusters C1–C3 shown in Figure**

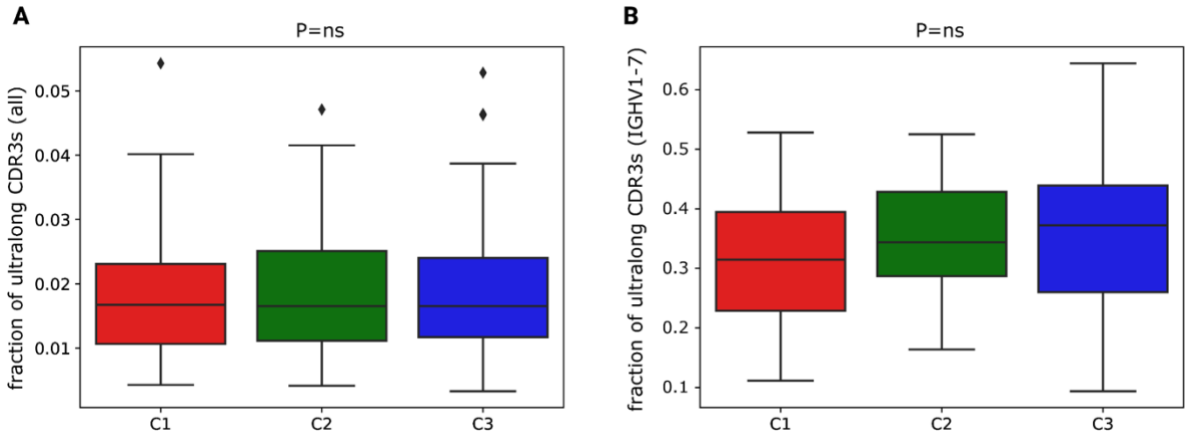
36 **4B.** Cells show association likelihoods (= $-\log_{10}$P) varying from 0 (white) to 36 (violet). Variations are labeled by

37 positions in V genes. Statistically significant associations (P<0.05) are additionally labeled with "*".

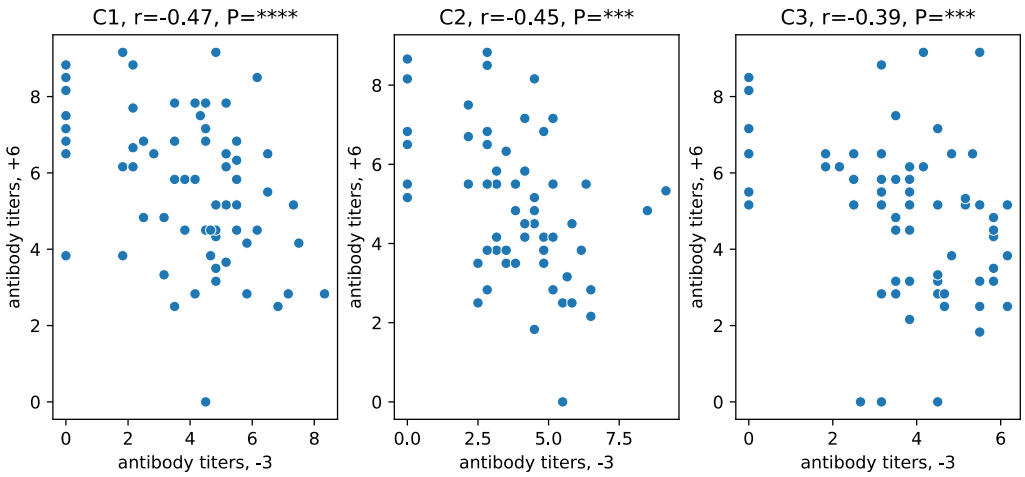
38



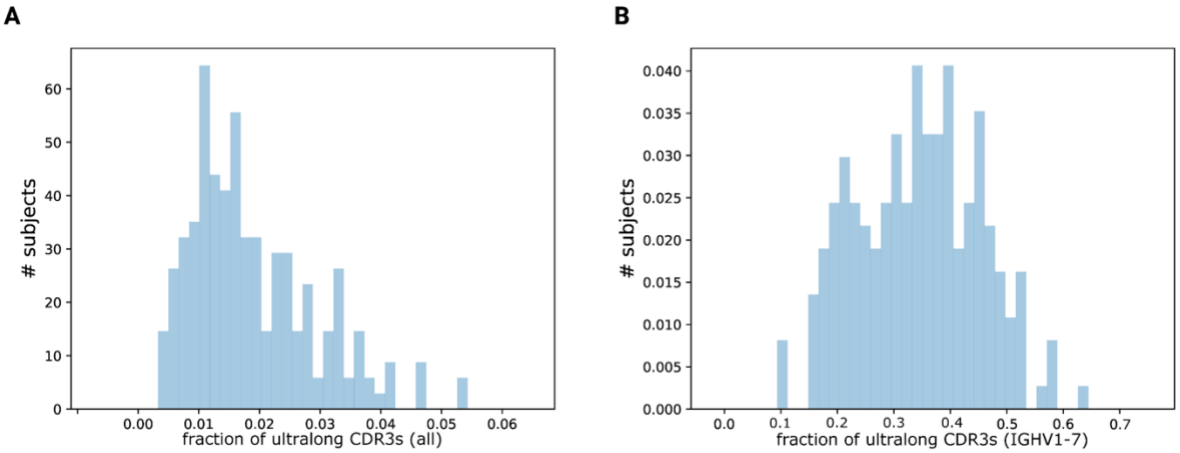
Supplemental Figure S5. Distribution of usages of eight V genes in three clusters identified using PCA shown in Figure 4. P-values were computed using the Kruskal-Wallis test.



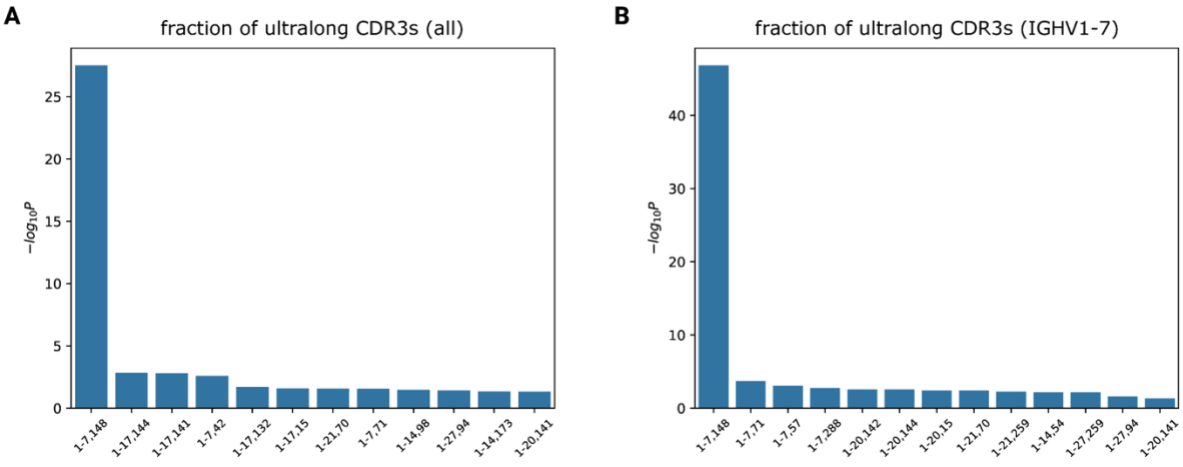
Supplemental Figure S6. Fraction of ultralong CDR3s in all CDR3s (*A*) and CDR3s derived from IGHV1-7 only (*B*) across clusters C1, C2, and C3. P-values were computed using the Kruskal-Wallis test.



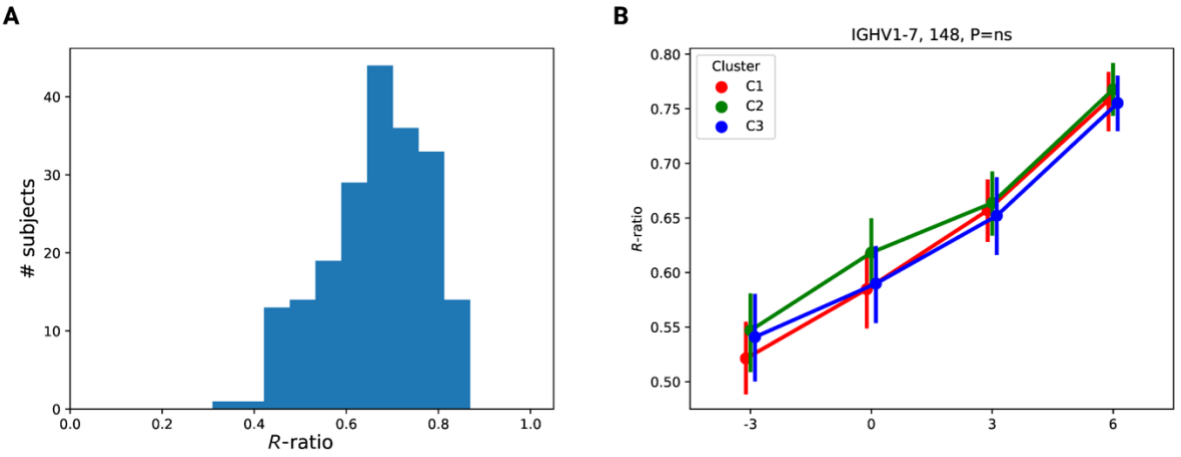
Supplemental Figure S7. Antibody titers at time point "–3" vs antibody titers at time point "+6" in clusters C1 (left), C2 (middle), and C3 (right). The Pearson correlations (*r*) and P-values (P) are shown the top of panels.



Supplemental Figure S8. Fractions of ultralong CDR3s. (*A*) The distribution of the fractions of ultralong antibodies among all antibodies in the combined datasets. (*B*) The distribution of the fraction of sequences with ultralong antibodies in all antibodies derived from IGHV1-7 in the combined datasets.



Supplemental Figure S9. Likelihoods of association P-values between fractions of ultralong CDR3s and GSVs. Likelihood is computed as the negative logarithm of the P-value to the base of 10. GSVs are labeled by the V gene and the position in it and shown in the descending order of likelihoods. Only GSVs with P-values below 0.05 are shown. (*A*) and (*B*) show fractions of ultralong CDR3s in all CDR3s and CDR3s derived from IGHV1-7 only, respectively.



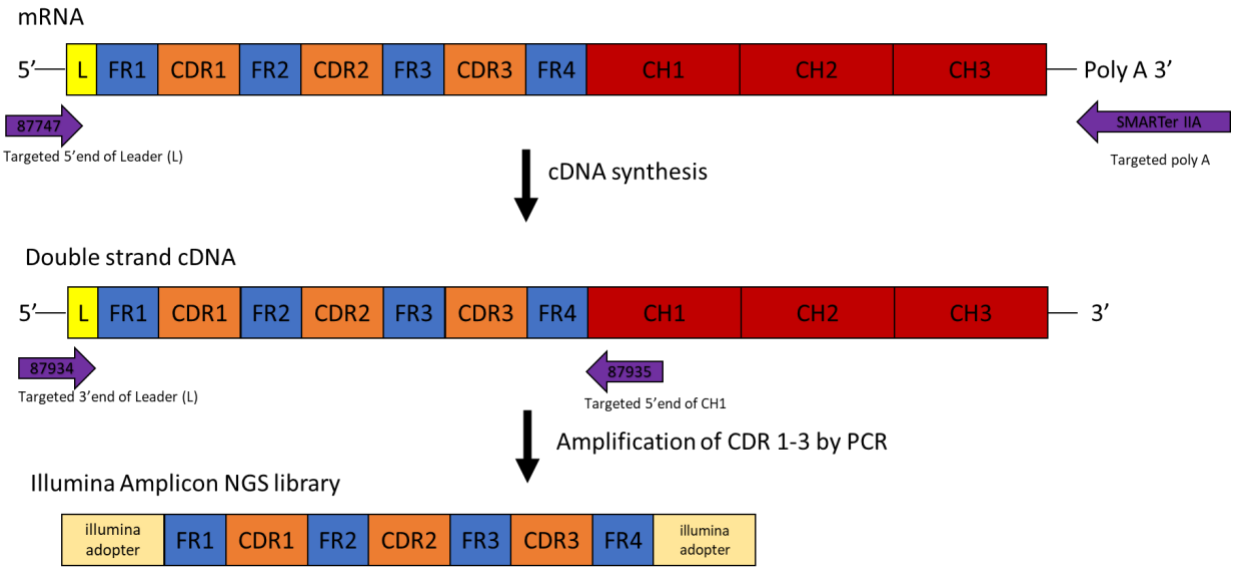
58

59 **Supplemental Figure S10. Characteristics of the GSV (IGHV1-7, 148).** (*A*) The distribution of *R*-ratios at position
60 148 in IGHV1-7. (*B*) The distribution of *R*-ratios at position 148 across clusters C1–C3. Vertical lines show 95%
61 confidence intervals.

```
                                    CDR1                    * CDR2
IGHV1-7    QVQLRESGPSLVKPSQTLSLTCTVSGFSLSDKAVGWVRQAPGKALEWLGGIDTGGSTGYNPGLKSRLSITKDNSKSQVSLSVSSVTTEDSATYYCTTVHQ
4K3D       EVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGNTGYNPGLKSRLSITKDNSKSQVSLSVSSVTTEDSATYYCTSVHQ
4K3E       EVQLRESGPSLVQPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGSTGYNPGLKSRLSITKDNSKSQVSLSVSSVTTEDSATYYCTTVHQ
5E99       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGTAGYNPGLKTRLSITKDNSKSQVSLTVSSVATEDSATYYCVTVYQ
5IHU       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGTAGYNPGLKTRLSITKDNSKSQVSLTVSSVATEDSATYYCVTVYQ
5IJV       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGNIDTGGITGYNPGLKSRLSITKDNSKNQVSLSVSSATAEDSATYYCTTVHQ
5ILT       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPEKALEWLGSIDTSGTTGYNSGLKSRLSIIKDNSKSQVSLSVSSVTTEDSATYYCTIVHQ
6E8V       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGSTGYNPGLKSRLSITKDNSESQVSLSVSSVTTEDSATYYCTTVHQ
6E9G       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGNAGYNPGLKSRLSITQDNSKSQVSLSVSTVTTEDSATYYCTTVQQ
6E9H       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGNAGYNPGLKSRLSITQDNSKSQVSLSVSTVTTEDSATYYCTTVHQ
6E9I       QVQLRESGPSLVKPSQTLSLTCTASGLSLSDKAVGWVRQAPGKALEWLGSIDTGGSAGYNPGLKSRVSITKDNSKSQVSLSVRGVTTEDSATYYCTTVHQ
6E9K       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGNAGYNPGLKSRLSITQDNSKSQVSLSVSTVTTEDSATYYCTTVHQ
6E9Q       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGNAGYNPGLKSRLSITQDNSKSQVSLSVSTVTTEDSATYYCTTVHQ
6E9U       QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGSIDTGGNAGYNPGLKSRLSITQDNSKSQVSLSVSTVTTEDSATYYCTTVHQ
           .          .         .  .          .      .  . ..   .  . .  ..  ..   . .....       .. .
```

62

63 **Supplemental Figure S11. Known bovine antibodies with ultralong CDR3s support variation at position 148 in**
64 **IGHV1-7.** Fragments of heavy chain sequences of 13 crystallized bovine antibodies with ultralong CDR3s
65 corresponding to IGHV1-7. Accession numbers of corresponding structures are shown on the left. The germline
66 sequence of IGHV1-7 is shown on the top of the alignment. Amino acid position 50 corresponding to the nucleotide
67 position 148 is marked with "*". Columns with at least one difference from the germline amino acid are marked with
68 "." on the bottom of the alignment. CDR1s and CDR2s (according to the IMGT notation) are shown by grey boxes.



Supplemental Figure S12. Illustration of the amplification of IGHG transcripts to generate Rep-Seq libraries. RNA extracted using poly(A)-enrichment was converted to first-strand cDNA using a poly(T) primer (SMARTer IIA primer included in kit). The second strand was generated using a primer specific to the leader region of the coding sequence (primer 87747, Table S3), creating IgG-enriched, double-strand cDNA. This cDNA was amplified by PCR using primers containing Illumina adapter sequence and flanking sequences at the 3' end of the VH leader (L) region (primer 87934) and the 5' end of the IgG heavy chain constant (CH1) region (primer 87935), respectively, which amplified the IgG cDNA across the heavy chain variable region (FR1–4, including CDR1–3).

Supplemental Tables

Gene	Average usage at "–3"	Average usage at "+6"	Usage fold
IGHV1-7	0.06	0.11	1.83
IGHV1-10	0.14	0.15	1.07
IGHV1-14	0.13	0.11	0.85
IGHV1-17	0.22	0.21	0.95
IGHV1-20	0.28	0.25	0.89
IGHV1-21	0.17	0.16	0.94
IGHV1-27	0.13	0.14	1.08
IGHV1-30	0.07	0.07	1.00

Supplemental Table S1. Average usages of eight V genes at time points "–3" and "+6". The column "Usage fold" shows the ratio of the average usages at "+6" to the average usage at "–3" for each V gene.

Supplemental Methods

Supplemental Method: The relations between pre- and post-vaccination immunity to the BRD vaccine in calves

The simplest vaccination scenario suggests the immunity does not exist before the vaccination and is gained after the vaccination. It results in low and high antibody titers before and after the vaccination, respectively. A more complex case might account for the impact of pre-existing immunity. If a subject has *optimal* pre-existing immunity (immunity that successfully responds to the vaccine), we except that antibody titers before and after the vaccination would be high. However, the BRD vaccination in 204 calves analyzed in this study represents an even more complex case. While the distributions of titers before and after the vaccination have similar mean values (Figure 2A), they anticorrelate: the higher titer before the vaccination, the lower titer after the vaccination, and vice versa (Figure 2B). This cannot be explained by optimal pre-existing immunity because titers before and after vaccinations would be correlated in that case. Thus, we suggest that pre-existing immunity is *suboptimal* and prevents gaining immunity from the vaccine in some calves.