Science Immunology

immunology.sciencemag.org/cgi/content/full/2/15/eaan8289/DC1

Supplementary Materials for

Dynamics of the human antibody repertoire after B cell depletion in systemic sclerosis

Charles F. A. de Bourcy, Cornelia L. Dekker, Mark M. Davis, Mark R. Nicolls, Stephen R. Quake*

*Corresponding author. Email: quake@stanford.edu

Published 29 September 2017, *Sci. Immunol.* **2**, eaan8289 (2017)
DOI: 10.1126/sciimmunol.aan8289

This PDF file includes:

Materials and Methods
Fig. S1. Comparison of healthy and SSc-PAH cohorts at baseline.
Fig. S2. Calculation of rates governing repertoire homeostasis.
Fig. S3. Assessment of repertoire elasticity under depletion.
Table S1. Data values displayed in Fig. 1E.
Table S2. Data values displayed in Fig. 1F.
Table S3. Data values displayed in Fig. 4A.
Table S4. Data values displayed in Fig. 4C.
Table S5. Data values displayed in Fig. 4D.
Table S6. Repertoire isotype percentages for each participant and time point (longitudinal experiment).
Table S7. IGH constant-region primer pool.
Table S8. IGH variable-region primer pool.
Table S9. Read counts before and after processing.
Table S10. Data values displayed in fig. S2A.
Table S11. Data values displayed in fig. S2B.
Table S12. Data values displayed in fig. S2D.
References (*44*–*46*)

Materials and Methods

IGH repertoire sequencing

The assay used in the present study proceeded in the following steps, which were a minor variation of published protocols (*30, 39, 40*): PBMC isolation from whole blood using a Ficoll gradient according to Stanford Human Immune Monitoring protocols (storage by freezing in 10% (vol/vol) DMSO/40% (vol/vol) FBS); total RNA extraction from thawed cells using the Qiagen AllPrep kit; RT on 250 ng total RNA using SuperScript III reverse transcriptase (Life Technologies) and IGH constant domain specific primers containing 8 or 12-nucleotide random barcodes synthesized by Integrated DNA Technologies (Table S10); second-strand synthesis using Phusion High-Fidelity DNA Polymerase (New England Biolabs, 98 °C for 4 min, 52 °C for 1 min, 72 °C for 5 min) and IGH variable domain specific primers containing 8 or 12-nucleotide random barcodes synthesized by Integrated DNA Technologies (Table S11); two rounds of cDNA purification with Ampure XP beads (Beckman Coulter) at a 0.8:1 ratio; PCR amplification using Platinum Taq DNA Polymerase High Fidelity (Life Technologies) and primers containing sample-multiplexing indexes combined with Illumina adapters; another round of purification with Ampure XP beads (0.7:1 ratio); sample pooling at equal volume; gel-purification using E-Gel EX Agarose Gels 2% (Invitrogen) and Freeze ‘N Squeeze DNA Gel Extraction Columns (Bio-Rad); 150-bp paired-end NextSeq sequencing (Illumina).

Data pre-processing

Data pre-processing was carried out in a manner analogous to a previous study (*30*). First, we carried out the following steps of the pRESTO (*41*) suite (version 0.4.8): raw read filtering with a threshold of 20 for the mean Phred quality; removal of PCR primer sequences from both ends in

two passes, dealing first with the random barcodes of length 8 nucleotides and then with the random barcodes of length 12 nucleotides, annotating each read with its barcode; consolidation of read groups with identical barcodes into a consensus sequence after alignment based on primer sequences; removal of read groups above a 0.1 error threshold against the consensus sequence; removal of read groups below a 70% agreement on constant-domain primer identity; merging of paired-end mates into a single sequence based on the read group consensus; removal of pairs below overlap significance threshold $p < 10^{-5}$ or above overlap error threshold 0.3; removal of sequences containing more than 10 ambiguous inner nucleotides,;assignment of isotypes and subisotypes based on constant-domain sequence match with removal of sequences above error threshold 0.2; trimming of the constant-domain sequence; assignment of molecular abundances (number of paired-end barcodes observed for the same IGH sequence); assignment of consensus counts (number of read pairs supporting an IGH sequence); removal of sequences with consensus count under 2.

Next: IMGT/HighV-QUEST (*42*) was used for VDJ gene assignment; non-functional sequences were removed; TIgGER (*33*) was used to determine V-genotype and correct V-allele calls (after pooling of time points for the longitudinal experiment, not applicable to the baseline comparison experiment); a priori indistinguishable alleles (based on the sequenced region) were collapsed; Change-O (version 0.3.3) (*43*) was used to infer a germline sequence for each observed sequence, masking the CDR3 region with ambiguous nucleotides; low-abundance sequences only 1 nucleotide away from a high-abundance sequence (abundance threshold: 20) with an identical combination of V-gene, J-gene and CDR3 length were removed as they might have resulted from RT errors; finally, sequences with identical combinations of V-gene, J-gene and CDR3 length were clustered into lineages using single-linkage clustering on the CDR3

nucleotide sequence with a Hamming distance cutoff of 0.1 times the CDR3 length (using the R package ‘stringdist’ (*44*)). Snakemake (*45*) was used to manage computational workflows.

UniFrac calculation

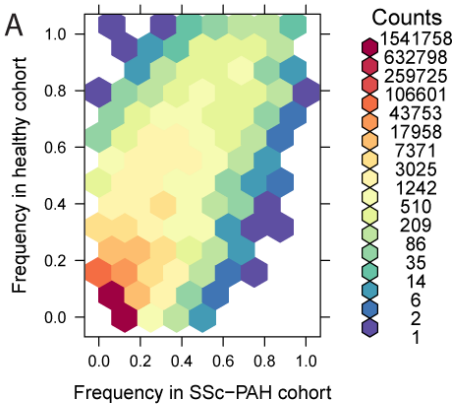
UniFrac (*29*) distances were computed as described in a previous antibody repertoire study (*30*), subsampling each repertoire/time-point to a depth of 10^4 distinct sequences if the total number of available sequences exceeded that number.

Mutation analysis

Reported nucleotide mutation counts were based on mismatches between each observed sequence and its inferred germline sequence. Reported AA changes for the V-segment were extracted from IMGT/HighV-QUEST output.

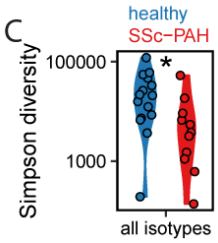
Plots

For Fig. 2C, abundance histograms were generated with a bin size of 20 and the lowest bin (which was the dominant one) was omitted from the plot to allow the tail of the distribution to be visualized. In Fig. 2D,F, points straddling the lower edge of the panel correspond to the value 0, which cannot be rendered on the logarithmic scale.



B

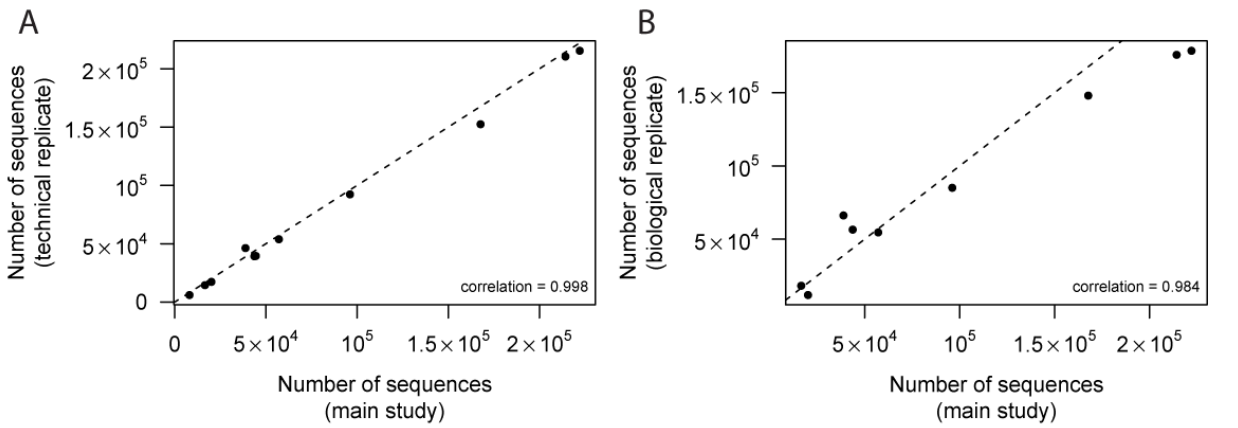
CDR3	p-value (Fisher)	p-value (Benjamini-Hochberg)	incidence in SSc-PAH cohort (N=11)	incidence in healthy cohort (N=15)
ARDRGSGWFDY	2.07×10^{-6}	1	1	15
ARGRDGYNSFDY	4.87×10^{-5}	1	0	12
ARGVGATGDY	4.87×10^{-5}	1	0	12
ARLQYYYYGMDV	1.06×10^{-4}	1	3	15
ARISGYYYYGMDV	1.57×10^{-4}	1	2	14
…	…	…	…	…
ARGGSSSPHFDY	2.01×10^{-3}	1	6	0
ARVPGYCSGGSCYDY	7.02×10^{-3}	1	5	0
ARDSGTAAFDY	7.02×10^{-3}	1	5	0
ARVRGVSYYFDY	7.02×10^{-3}	1	5	0
ARELYSSSWYNWFDP	7.02×10^{-3}	1	5	0
…	…	…	…	…



D

Segment	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
IGHV2-5	35.65864	-0.7430365	0.19214253	-3.867111	0.0001101322	0.004185025
IGHV4-30	7773.35680	0.2341236	0.09501339	2.464111	0.0137353509	0.260971666
IGHV1-24	10456.45858	0.3079372	0.13630655	2.259152	0.0238739475	0.302403335
IGHV3-23	5407.71243	0.1651125	0.11761566	1.403831	0.1603690906	0.909221620
IGHV3-30	4236.31313	0.2495112	0.20137958	1.239010	0.2153419626	0.909221620

Fig. S1. Comparison of healthy and SSc-PAH cohorts at baseline. (**A**) Two-dimensional histogram of AA CDR3 sequences indicating prevalence in the two cohorts. For each distinct AA CDR3 sequence, we determined which subjects contained that sequence or any of its one-mismatch derivatives in their repertoire. (SSc-PAH data from the cross-sectional experiment (described in Fig. 2), the baselines of the longitudinal experiment (described in Fig. 1,3,4,5,6), and a resequencing of the baseline libraries from the longitudinal experiment were considered.) The frequency of the sequence in a cohort is the fraction of subjects in the cohort that contained at least one relevant transcript. It is observed that most sequences are low-prevalence, and that prevalences in the two cohorts are generally similar with no obvious outliers. (**B**) The top part of the table shows the five AA CDR3 sequences with the overall lowest unadjusted p-values from Fisher's Exact Test. The bottom part shows five sequences with the lowest unadjusted p-values that were more prevalent in the SSc-PAH cohort than the healthy cohort. (**C**) Simpson diversity of the overall repertoire (1 divided by the sum of the squared fractional abundances of all sequences; shown on a logarithmic axis). Each point corresponds to one subject. (**D**) Differential gene expression analysis results output by DESeq2 (*46*), contrasting total repertoires for the SSc-PAH cohort relative the healthy cohort. "padj" is the Benjamini-Hochberg adjusted p-value. Only the 5 genes with the lowest padj-value are displayed; only IGHV2-5 was statistically significant.



C

	estimate (week^{-1})	std. error (week^{-1})	p-value
intercept(SR1) - intercept(SR6)	-8.24×10^{3}	3.90×10^{2}	2.98×10^{-5}
intercept(SR2) - intercept(SR6)	-7.77×10^{3}	5.65×10^{2}	1.62×10^{-4}
intercept(SR4) - intercept(SR6)	-6.82×10^{3}	6.42×10^{2}	4.44×10^{-4}
intercept(SR5) - intercept(SR6)	3.55×10^{4}	3.94×10^{3}	8.47×10^{-4}
intercept(SR6)	8.53×10^{3}	2.89×10^{2}	7.83×10^{-6}
slope(SR1) - slope(SR6)	5.96×10^{-3}	5.33×10^{-2}	0.916
slope(SR2) - slope(SR6)	-3.99×10^{-2}	7.65×10^{-2}	0.630
slope(SR4) - slope(SR6)	-4.30×10^{-2}	4.31×10^{-2}	0.375
slope(SR5) - slope(SR6)	-3.29×10^{-1}	3.41×10^{-2}	6.41×10^{-4}
slope(SR6)	-3.86×10^{-2}	2.01×10^{-3}	4.31×10^{-5}

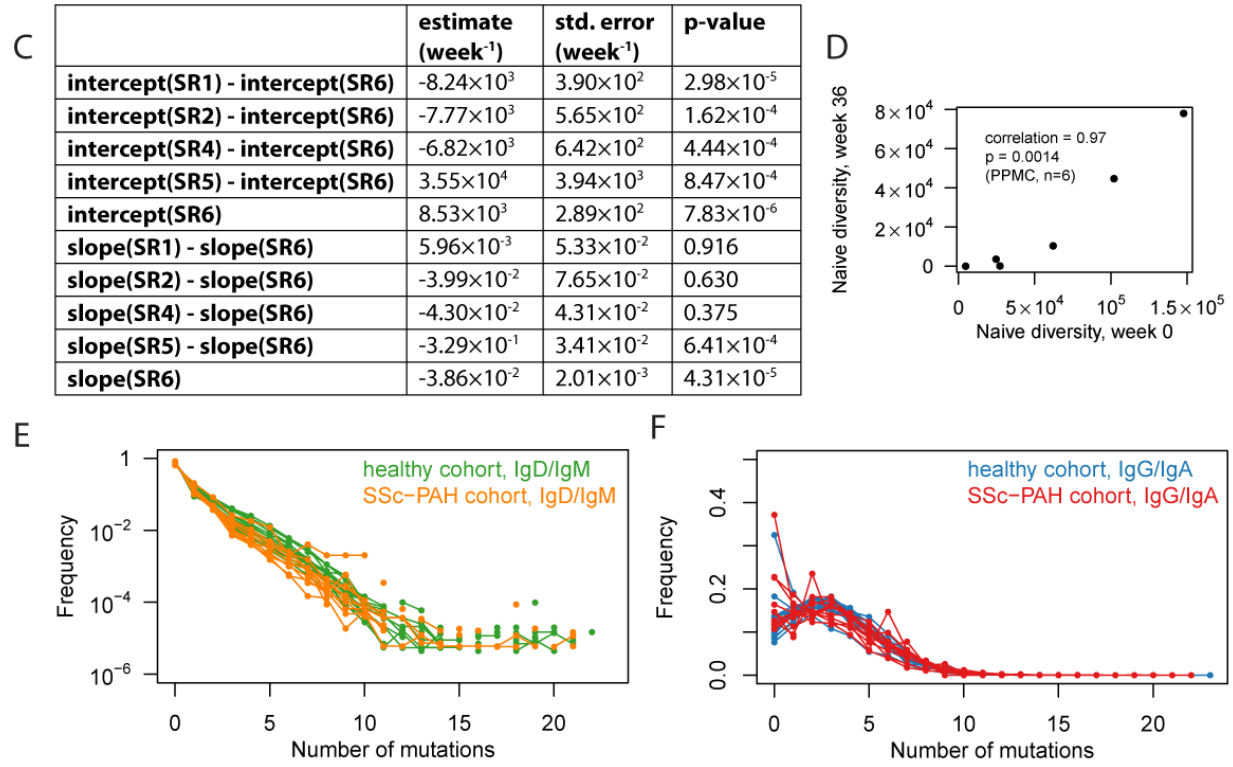


Fig. S2. Calculation of rates governing repertoire homeostasis. (**A**) Observed sequence diversity is consistent across technical replicates. Values displayed on the horizontal axis correspond to the SSc-PAH baseline data from the longitudinal experiment described in the main text; technical replicates were generated by resequencing the corresponding libraries at a comparable depth. Dotted line corresponds to equal values of the horizontal and vertical axes for comparison. Numeric data values are also provided in Table S10. (**B**) Observed sequence diversity is consistent across biological replicates. Values displayed on the horizontal axis correspond to the SSc-PAH baseline data from the longitudinal experiment described in the main text; biological replicates were generated by making new libraries from different RNA aliquots of the same samples and sequencing them at a comparable depth (same data as the SSc-PAH group of the cross-sectional experiment described in Fig. 2 of the main text). The same mass of input RNA (250 ng) was used to make all libraries; samples for which insufficient RNA was left

over from the main study were omitted (SR3 baseline, SP5 baseline). Dotted line corresponds to equal values on the horizontal and vertical axes for comparison. Numeric data values are also provided in Table S11. (**C**) Fit coefficients for a linear model of the form $y_i = a_6 + \sum_{p=1}^{5} a_{p6} I_p + s_6 x_i + \sum_{p=1}^{5} s_{p6} x_i I_p$ with standard errors and p-values for rejecting the null hypothesis that a coefficient is zero. Here (x_i, y_i) is an $\left(M_0, \frac{dM_0}{dt}\right)$ data point; I_p is an indicator value with value 1 if the sample (x_i, y_i) is from participant p and 0 otherwise; $a_{p6} = a_p - a_6$ and $s_{p6} = s_p - s_6$ where a_p and s_p are the intercept and slope in the linear model applicable to the $\left(M_0, \frac{dM_0}{dt}\right)$ data of participant "SR"p (p=1,2,4,5,6). M_0 trajectories were only modeled from the week 24 visit onwards to ensure onset of replenishment had already occurred; for participant SR1 it was necessary to begin later at week 36 for the trajectory to be monotonically increasing. Before conducting the fit, M_0 temporal trajectories were first lightly smoothed (each 2 successive values were averaged and assigned to the middle of the 2 corresponding time points); then dM_0/dt values were approximated as $\frac{\Delta M_{0.}}{\Delta time}$ from each pair of adjacent time points and the average M_0 of the same two time points was used as the corresponding M_0-value to give an $\left(M_0, \frac{dM_0}{dt}\right)$ pair. Participant SR3 is omitted because an insufficient number of time points after onset of depletion were available; SR6 was chosen as the reference for slope and intercept because SR6 was tracked over the longest period of time. The table spells out "intercept" and "slope" instead of a and s, e.g. intercept(SR5) instead of a_5 and slope(SR5) instead of s_5. Note that slope(SR1) – slope(SR6), slope(SR2) – slope(SR6) and slope(SR4) – slope(SR6) were all indistinguishable from zero (p-values above 0.1), suggesting that the slope of the linear model is similar for 4 out of the 5 modeled participants. (**D**) Number of non-mutated IgM or IgD sequences at baseline versus week 36. PPMC: Pearson Product Momentum Correlation. Numeric data values are also provided in Table S12. (**E**) Distribution of mutation loads in the IgD/IgM compartment at baseline. The distribution roughly corresponds to an exponential decay. (**F**) Distribution of mutation loads in the isotype-switched compartment at baseline. The distribution tends to be peaked at a non-zero value.

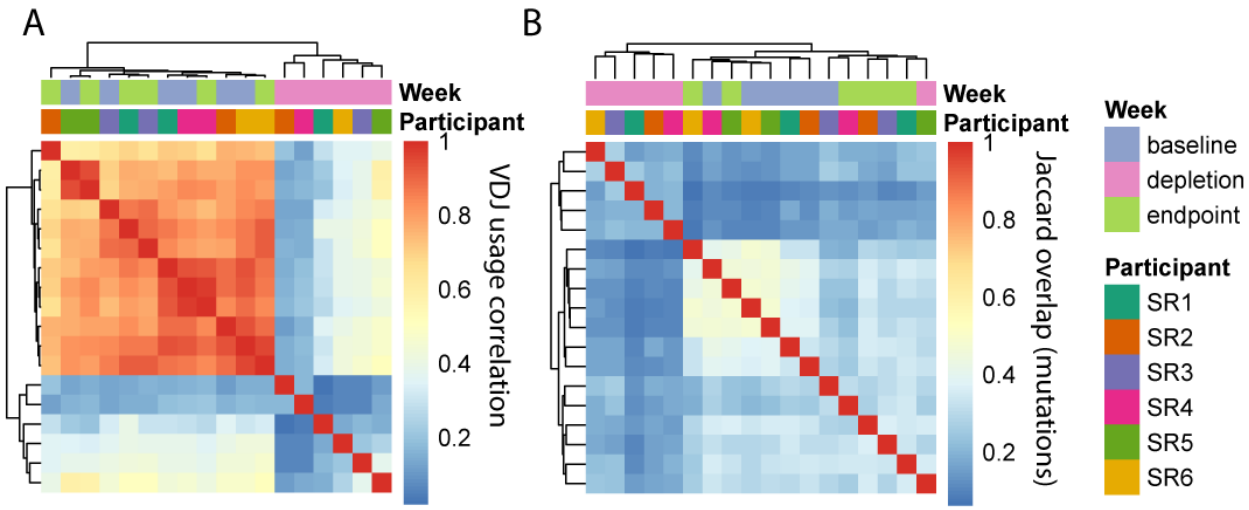


Fig. S3. Assessment of repertoire elasticity under depletion. (**A**) Comparison of VDJ usages (sequence-weighted) between participants and time points, displayed as a correlation heatmap. "Depletion" corresponds to the week 4 sample, "endpoint" to the last time point sampled. Rows and columns were ordered using complete-linkage hierarchical clustering with Euclidean distance. (**B**) Comparison of AA mutation sets between participants and time points, displayed as a heatmap of Jaccard overlap indices. Jaccard indices of the sets of position-labeled AA changes were computed separately for each V-gene and then averaged over all V-genes. "Depletion" corresponds to the week 4 sample, "endpoint" to the last time point sampled. Rows and columns were ordered using complete-linkage hierarchical clustering with Euclidean distance.

Table S1. Data values displayed in Fig. 1E.

Participant	Fold-change, sequence diversity	Fold-change, isotype-switched proportion
SP1	2.51×10^{0}	2.64
SP2	1.88×10^{0}	0.85
SP3	5.34×10^{-1}	1.26
SP4	5.84×10^{-1}	1.75
SP5	2.02×10^{0}	0.82
SR1	4.68×10^{-3}	11.34
SR2	6.60×10^{-3}	8.92
SR3	2.79×10^{-2}	11.04
SR4	3.26×10^{-3}	19.25
SR5	1.01×10^{-2}	33.02
SR6	4.03×10^{-3}	10.04

Table S2. Data values displayed in Fig. 1F.

Participant	UniFrac(week 0, week 36)	Study arm
SP1	0.862	B
SP2	0.834	B
SP3	0.828	B
SP4	0.800	B
SP5	0.827	B
SR1	0.963	A
SR2	0.901	A
SR3	0.973	A
SR4	0.913	A
SR5	0.860	A
SR6	0.861	A

Table S3. Data values displayed in Fig. 4A.

Participant	Week	M_0
SR1	49	1.57×10^3
SR1	61	4.39×10^3
SR1	74	5.75×10^3
SR1	87	6.51×10^3
SR6	36	5.89×10^4
SR6	49	1.24×10^5
SR6	62	1.65×10^5
SR6	80	1.91×10^5

Table S4. Data values displayed in Fig. 4C.

Naïve diversity, week 0	a (1/week)
2.46×10^4	5.26×10^2
2.72×10^4	3.24×10^2
6.21×10^4	1.18×10^3
1.02×10^5	8.70×10^3
1.48×10^5	6.08×10^3

Table S5. Data values displayed in Fig. 4D.

Naïve diversity, week 0	Time to onset of replenishment (weeks)
4.68×10^3	57
2.46×10^4	44
2.72×10^4	58
6.21×10^4	34
1.02×10^5	33
1.48×10^5	23

Table S6. Repertoire isotype percentages for each participant and time point (longitudinal experiment).

Participant	Week	% IgD	% IgM	% IgA1	% IgA2	% IgG1	% IgG2	% IgG3	% IgG4	% IgE
SP1	**-1**	9.02	87.97	1.3	0.52	0.94	0.18	0.08	0	0
SP1	**2**	10.93	80.48	3.53	0.34	3.86	0.38	0.47	0	0
SP1	**11**	9.87	82.17	4.06	0.78	2.17	0.7	0.23	0.02	0
SP1	**23**	9.28	79.97	4.21	0.56	4.91	0.78	0.29	0	0.01
SP1	**27**	8.33	83.69	4.22	0.62	2.29	0.59	0.26	0	0
SP1	**35**	8.53	84.64	4.23	0.44	1.67	0.27	0.21	0	0
SP2	**0**	14.38	77.56	2.36	0.7	4.35	0.42	0.23	0	0.01
SP2	**2**	13.55	76.99	3.28	1.02	3.69	1.05	0.42	0	0
SP2	**4**	10.9	78.78	4.15	1.49	2.97	1.35	0.37	0	0
SP2	**11**	10.15	86.5	1.33	0.35	1.26	0.26	0.14	0	0
SP2	**24**	7.4	88.31	2	0.39	1.24	0.42	0.24	0	0
SP2	**35**	11.43	84.93	1.24	0.54	1.39	0.35	0.11	0	0.01
SP2	**47**	7.85	84.67	1.85	0.26	4.35	0.82	0.21	0	0
SP3	**0**	9.38	88.78	0.99	0.2	0.22	0.14	0.24	0.04	0.01
SP3	**2**	9.94	87.12	1.17	0.59	0.45	0.29	0.39	0.05	0
SP3	**4**	10.58	87.23	0.83	0.43	0.37	0.2	0.3	0.05	0
SP3	**12**	7.26	90.32	0.73	0.34	0.46	0.46	0.37	0.07	0
SP3	**24**	6.7	91.55	0.72	0.38	0.26	0.17	0.17	0.05	0
SP3	**28**	6.85	91.09	0.79	0.36	0.32	0.24	0.34	0.01	0
SP3	**37**	9.45	88.49	0.84	0.47	0.32	0.17	0.22	0.04	0
SP3	**49**	12.97	84.82	1.24	0.31	0.26	0.16	0.23	0.02	0
SP3	**76**	7.49	91.34	0.42	0.19	0.22	0.12	0.21	0.01	0
SP4	**-1**	10.91	86.28	1.75	0.54	0.38	0.08	0.03	0.03	0
SP4	**2**	11.68	84.81	2.1	0.63	0.6	0.08	0.08	0.02	0
SP4	**3**	6.26	89.62	2.93	0.58	0.42	0.1	0.04	0.05	0
SP4	**13**	5.15	89.13	3.73	0.82	0.78	0.26	0.07	0.06	0
SP4	**23**	5.93	92.2	1.2	0.28	0.27	0.07	0.02	0.03	0
SP4	**38**	7.26	90.53	1.35	0.34	0.37	0.08	0.03	0.05	0
SP5	**0**	13.8	78.04	5.41	0.92	0.82	0.74	0.23	0	0.03
SP5	**2**	13.44	81.68	3.08	0.76	0.43	0.45	0.15	0	0
SP5	**4**	9.93	81.39	4.42	1.67	1.19	1.03	0.37	0	0.01
SP5	**11**	13.99	81.38	2.47	0.63	0.65	0.68	0.19	0	0.01
SP5	**36**	10.46	82.35	4.53	1.36	0.61	0.51	0.17	0	0.01
SP5	**48**	7.54	83.54	5.7	1.17	1.32	0.5	0.22	0	0.01
SR1	**0**	5.36	91.39	1.59	0.78	0.37	0.48	0.03	0	0
SR1	**4**	0.32	70.79	12.38	7.62	2.86	6.03	0	0	0
SR1	**13**	0	55.32	11.7	22.34	1.06	9.57	0	0	0
SR1	**23**	4.42	60.82	7.76	3.53	14.38	3.22	5.74	0.13	0
SR1	**37**	0.23	76.59	13.64	5	2.27	2.27	0	0	0
SR1	**49**	2.25	68.45	17.46	6.62	1.27	3.94	0	0	0
SR1	**62**	3.66	87.37	6.12	1.51	0.65	0.61	0.07	0	0
SR1	**74**	6.13	89.11	2.76	0.94	0.54	0.51	0	0	0
SR1	**87**	5.92	88.46	3.26	1.16	0.55	0.63	0.02	0	0
SR1	**100**	5.82	90.62	2	0.69	0.49	0.38	0	0	0
SR2	**-1**	6.2	85.14	4.88	0.98	1.85	0.81	0.14	0	0
SR2	**3**	0.34	16.55	70.95	4.39	4.05	3.38	0.34	0	0
SR2	**11**	0	28.57	30.41	4.15	17.05	19.82	0	0	0
SR2	**23**	1.21	8.57	78.04	3.02	6.19	1.7	1.26	0	0

SR2	**35**	4.96	39.83	30.86	7.41	5.71	10.79	0.42	0	0
SR2	**47**	6.72	88.05	3.33	0.73	0.43	0.53	0.2	0	0
SR2	**69**	5.36	67.02	21.71	2.91	1.96	0.78	0.26	0	0.01
SR3	**0**	10.85	84.27	2.07	0.37	1.16	1.15	0.1	0.02	0
SR3	**2**	0.81	66.67	11.38	9.76	3.25	5.69	2.44	0	0
SR3	**4**	1.59	36.87	32.1	5.57	6.63	15.38	1.86	0	0
SR3	**13**	1.28	52.56	28.21	2.56	3.85	11.54	0	0	0
SR3	**25**	0	0	0	0	100	0	0	0	0
SR3	**38**	1.2	47.2	18.4	7.2	12	13.6	0.4	0	0
SR3	**47**	2.95	69.99	14.03	3.37	1.82	7.29	0.14	0.42	0
SR3	**59**	9.6	89.4	0.54	0.12	0.22	0.11	0.01	0	0
SR4	**0**	11.06	85.68	0.83	0.37	1.73	0.19	0.13	0.01	0
SR4	**2**	5.56	37.3	12.7	31.75	10.32	2.38	0	0	0
SR4	**4**	0	23.14	18.09	23.4	19.15	8.78	7.45	0	0
SR4	**12**	1.2	50.2	18.33	22.31	4.38	3.59	0	0	0
SR4	**29**	2.09	45.86	17.63	19.72	8.52	5.93	0.19	0.06	0
SR4	**36**	9.74	77.63	3.03	1.75	6.92	0.79	0.14	0	0
SR4	**47**	12.42	83.03	2.35	0.95	0.62	0.46	0.09	0.08	0
SR4	**60**	10.9	86.16	1.11	0.27	1.14	0.17	0.19	0.05	0.01
SR5	**0**	13.43	83.85	1.71	0.34	0.33	0.28	0.05	0.01	0
SR5	**2**	0.15	3.73	35.48	26.29	12.18	20.34	1.72	0.11	0
SR5	**4**	0.14	2.31	43.31	35.22	1.76	16.23	0.99	0.05	0
SR5	**12**	0.05	17.83	48.6	21.16	2.19	9.89	0.24	0.05	0
SR5	**25**	13.49	82.85	2.46	0.67	0.04	0.47	0.01	0.01	0
SR5	**29**	7.67	90.14	1.26	0.49	0.08	0.34	0.02	0.01	0
SR5	**37**	7.68	87.82	2.57	1	0.2	0.68	0.04	0	0
SR5	**63**	11.34	86.12	1.19	0.6	0.19	0.53	0.03	0.01	0
SR6	**-1**	6.45	84.59	2.67	1.25	2.54	1.67	0.43	0.4	0
SR6	**3**	0.6	14.64	24.4	27.36	8.21	23.87	0.66	0.18	0.08
SR6	**5**	0.14	6.09	37.95	27.29	11.91	16.62	0	0	0
SR6	**11**	0.32	13.51	39.11	21.78	7.31	17.17	0.32	0.32	0.16
SR6	**23**	0.43	9.28	43.19	27.54	6.81	11.16	1.16	0.43	0
SR6	**36**	8.27	83.83	4.53	1.67	0.71	0.58	0.34	0.06	0
SR6	**48**	7.53	87.86	2.35	0.94	0.56	0.6	0.14	0.02	0
SR6	**62**	7.08	87.42	2.82	1.27	0.7	0.5	0.17	0.04	0
SR6	**76**	6.51	84.55	3.98	1.72	1.29	1.32	0.59	0.06	0
SR6	**106**	5.58	84.5	4.69	1.53	1.63	1.32	0.51	0.25	0

Table S7. IGH constant-region primer pool. Follows a previously published protocol (*40*). Ns correspond to random nucleotides, to be used as a barcode for recognizing unique molecules in the starting RNA material.

Sequence (5' to 3')	Proportion in pool	Intended isotype specificity
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGGGGAAGAAGCCCTGGAC	1	IgA
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGGGGAAGAAGCCCTGGAC	1	IgA
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGGGAAGTAGTCCTTGACCA	1	IgG
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGGGAAGTAGTCCTTGACCA	1	IgG
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNGAAGGAAGTCCTGTGCGAG	1	IgM
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGAAGGAAGTCCTGTGCGAG	1	IgM
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNAAGTAGCCCGTGGCCAGG	1	IgE
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNAAGTAGCCCGTGGCCAGG	1	IgE
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNTGGGTGGTACCCAGTTATCAA	1	IgD
TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNTGGGTGGTACCCAGTTATCAA	1	IgD

Table S8. IGH variable-region primer pool. Based on a previously published protocol (*39*). Ns correspond to random nucleotides, to be used as a barcode for recognizing unique molecules in the starting RNA material.

Sequence (5' to 3')	Proportion in pool	Intended isotype specificity
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNAGCCTACATGGAGCTGAGC	1	V1
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNAGCCTACATGGAGCTGAGC	1	V1
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNAGGTGGTCCTTACAATGACCAAC	1	V2
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNAGGTGGTCCTTACAATGACCAAC	1	V2
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNTCTGCAAATGAACAGCCTGA	1	V3
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNTCTGCAAATGAACAGCCTGA	1	V3
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNTGTTCAAATGAGCAGTCTGAGAG	0.2	V3
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNTGTTCAAATGAGCAGTCTGAGAG	0.2	V3
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTCTGCAAATGGGCAGCCTGA	0.2	V3
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNTCTGCAAATGGGCAGCCTGA	0.2	V3
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTTCTCCCTGAAGCTGAACTCTG	1	V4/V6
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNTTCTCCCTGAAGCTGAACTCTG	1	V4/V6
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGCCTACCTGCAGTGGAGCAG	1	V5
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNGCCTACCTGCAGTGGAGCAG	1	V5
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTTCTCCCTGCAGCTGAACTCTG	1	V6
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNTTCTCCCTGCAGCTGAACTCTG	1	V6
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGCATATCTGCAGATCAGCAGC	1	V7
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNGCATATCTGCAGATCAGCAGC	1	V7
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNCAGATCAGCAGCCTAAAGGC	1	V7
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNCAGATCAGCAGCCTAAAGGC	1	V7

Table S9. Read counts before and after processing.

Participant	Week	Number of raw read pairs	Number of unique sequences after processing	Number of unique molecules (UIDs) after processing
SP1	**-1**	7065459	20184	149318
SP1	**2**	10259251	38310	328056
SP1	**11**	12400086	50615	352586
SP1	**23**	8023023	35428	226146
SP1	**27**	11748379	44285	340467
SP1	**35**	15131779	67667	511813
SP2	**0**	5394332	16625	102781
SP2	**2**	10796111	32465	274736
SP2	**4**	9098808	24961	194046
SP2	**11**	11462855	37591	273145
SP2	**24**	3819380	18004	95821
SP2	**35**	4018362	76366	336313
SP2	**47**	7267439	19960	149433
SP3	**0**	8376136	57164	277110
SP3	**2**	4765950	25946	95534
SP3	**4**	9502224	52576	261804
SP3	**12**	1809970	8490	32928
SP3	**24**	4270901	45854	147929
SP3	**28**	7544912	46943	218022
SP3	**37**	7774177	47066	210665
SP3	**49**	11771748	60141	329905
SP3	**76**	14494701	82727	582603
SP4	**-1**	5198036	222120	894333
SP4	**2**	2738879	70069	194885
SP4	**3**	16076699	157419	887092
SP4	**13**	18247714	102062	664319
SP4	**23**	16184874	219237	1162252
SP4	**38**	16897337	206761	991418
SP5	**0**	9521534	44543	287543
SP5	**2**	6134873	39370	211714
SP5	**4**	4571618	53108	321076
SP5	**11**	5842453	126559	754563
SP5	**36**	11828086	40020	286402
SP5	**48**	14738789	41373	328690
SR1	**0**	8248237	43729	247616
SR1	**4**	510133	315	6473
SR1	**13**	480922	94	1172
SR1	**23**	775702	1585	13468
SR1	**37**	556725	440	9442
SR1	**49**	3772844	710	9668
SR1	**62**	15431354	10907	146175
SR1	**74**	5590711	9342	59257
SR1	**87**	2925310	8901	63794
SR1	**100**	6477611	14804	87665
SR2	**-1**	5730424	38860	183429
SR2	**3**	614084	296	6247
SR2	**11**	506043	217	4507

SR2	**23**	3600132	3876	79999
SR2	**35**	5574124	13194	237309
SR2	**47**	6492734	16379	129700
SR2	**69**	7168492	17081	174544
SR3	**0**	2192922	8163	58819
SR3	**2**	354302	123	2074
SR3	**4**	697234	377	2682
SR3	**13**	698250	78	946
SR3	**25**	237465	1	3
SR3	**38**	489761	250	2653
SR3	**47**	1327233	713	6604
SR3	**59**	5934739	55925	237436
SR4	**0**	2922129	96150	295844
SR4	**2**	236182	126	2530
SR4	**4**	1404466	376	8269
SR4	**12**	1321136	251	4303
SR4	**29**	3885951	6749	119298
SR4	**36**	6005556	17160	141965
SR4	**47**	6248519	30319	156838
SR4	**60**	2625105	27295	90390
SR5	**0**	3579726	214301	542300
SR5	**2**	1836159	4583	74243
SR5	**4**	764307	2212	25954
SR5	**12**	1600109	2103	37657
SR5	**25**	4940860	63766	243757
SR5	**29**	5600003	232972	644607
SR5	**37**	6210490	114687	379403
SR5	**63**	6627985	249875	739812
SR6	**-1**	14955089	167746	980834
SR6	**3**	3558040	3812	93155
SR6	**5**	1055709	722	14164
SR6	**11**	1251398	629	15796
SR6	**23**	1270782	690	14593
SR6	**36**	16021627	70516	522482
SR6	**48**	14857243	244729	1414672
SR6	**62**	15003782	240768	1393535
SR6	**76**	11419627	343512	1941926
SR6	**106**	20174849	343052	2799804

Table S10. Data values displayed in fig. S2A.

Number of sequences (main study)	Number of sequences (technical replicate)
8.16×10^{3}	6.15×10^{3}
1.66×10^{4}	1.46×10^{4}
2.02×10^{4}	1.74×10^{4}
3.89×10^{4}	4.64×10^{4}
4.37×10^{4}	3.94×10^{4}
4.45×10^{4}	3.96×10^{4}
5.72×10^{4}	5.40×10^{4}
9.62×10^{4}	9.24×10^{4}
1.68×10^{5}	1.53×10^{5}
2.14×10^{5}	2.11×10^{5}
2.22×10^{5}	2.15×10^{5}

Table S11. Data values displayed in fig. S2B.

Number of sequences (main study)	Number of sequences (biological replicate)
1.66×10^{4}	1.83×10^{4}
2.02×10^{4}	1.19×10^{4}
3.89×10^{4}	6.63×10^{4}
4.37×10^{4}	5.66×10^{4}
5.72×10^{4}	5.47×10^{4}
9.62×10^{4}	8.52×10^{4}
1.68×10^{5}	1.48×10^{5}
2.14×10^{5}	1.76×10^{5}
2.22×10^{5}	1.79×10^{5}

Table S12. Data values displayed in fig. S2D.

Naïve diversity, week 0	Naïve diversity, week 36
4.68×10^{3}	4.50×10^{1}
2.46×10^{4}	3.62×10^{3}
2.72×10^{4}	1.62×10^{2}
6.21×10^{4}	1.04×10^{4}
1.02×10^{5}	4.47×10^{4}
1.48×10^{5}	7.80×10^{4}