Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/3/e1501371/DC1

Supplementary Materials for

Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting

Tarik A. Khan, Simon Friedensohn, Arthur R. Gorter de Vries, Jakub Straszewski, Hans-Joachim Ruscheweyh, Sai T. Reddy

Published 11 March 2016, *Sci. Adv.* **2**, e1501371 (2016)
DOI: 10.1126/sciadv.1501371

The PDF file includes:

Supplementary Materials

Supplementary Materials and Methods

Design of Experiments (DoE) for library preparation optimization
Based on pilot experiments we determined four potential critical process parameters: 1) quantity of input cDNA copies, 2) number of cycles in the 1st-step multiplex-PCR, 3) quantity of DNA copies input into 2nd-step adapter-extension-PCR, and 4) number of sequencing reads analyzed. Therefore we designed a Design of Experiments central composite design with two center points (10 individual sample preparations, fig. S6) based on the three experimental factors: 1, 2, and 3 (above). We then used *in silico* subsampling of reads (following CLC Genomics Workbench pre-processing and quality-filtering of merged reads) used as an input into the MAF pipeline. The final product yield generated after the 2nd-step PCR was designed to be constant to ensure enough material for NGS, while not subjecting the sample to over-amplifcation (PCR chimeras are more likely when the template:primer ratio is high, fig. S4). Therefore the correlated ddPCR-qPCR data was used to guide the amount of input copies from the purified product of 1st-step multiplex-PCR into the 2nd-step adaptor-extension-PCR reaction. This analysis was performed on RNA extracted from hyperimmunized splenocytes from a single mouse with the new TAK multiplex-PCR primer set for mouse V_H (excluding TAK_612). 13 spike-in clones (excluding IGHV1S81*02 clones) were analyzed to illustrate the influence of sample preparation on clonal frequency accuracy. The dataset S16 produced a lower number of reads and only the lowest read sampling level was used for analysis. All three-way interactions and curvature effects were analyzed to determine significant effects using the software platform JMP 11 (figs. S7A, S8A, and S9A). Then only the significant factors were used to generate modeled data (figs. S7B-C, S8B-C, and S9B-C).

Nominal Logistic Regression Modeling to Predict Immune Status of Clones
Nominal logistic regression modeling analyzes multivariate data that is categorically labeled with different statuses (e.g. sample material from hyperimmunized or untreated mice). This process generates a model that predicts based on the individual data points (clonotypes) which category (hyperimmunized or untreated) best fits. On a global level the average degree of separation may be used to compare the similarity (and difference) of immune status. While on the clonotype level, the separation represents if a given clonotype was responsive to immunization or antigen exposure. The top 100 ranked clonotypes (based on frequency) of all data sets, n = 3 samples from three separate hyperimmunized mice (IM_1a, IM_2, and IM_3) and n = 3 samples from untreated mice (UM_1, UM_2, and UM_3), were analyzed based on three parameters: log (MAF bias corrected frequency), intraclonotype diversity index (IDI), and non-silent somatic hypermutations (SHM (ns)). The model factors consisted of all main effects, curvature of main effects, secondary interactions, and the tertiary interactions (Supplementary fig. S19B). The significant factors were found to be (in order of importance): log(frequency), IDI, SHM (ns), and log(frequency)*IDI. In contrast the uncorrected data (fig. S19A) did not find the main effect of log(frequency) to be significant, likely due to inaccurate frequencies. The overall model performance of the MAF corrected data was improved, represented by an area under the curve of the receiver operating characteristic (AUC of ROC) of 0.94, in contrast AUC was 0.78 for the uncorrected data (fig. S20A-B). The models were then reduced to the significant factors of the MAF corrected data (fig. S20C-D), which maintain similar AUC values (figs. S20C-D). The probabilities of the reduced modeled data were plotted with each parameter used in the model, as well as the non-modeled clonotype rank (by frequency) (fig. S21). The uncorrected data showed a clear correlation of probability with SHM (ns), while the MAF corrected data has the strongest correlation with IDI. The same reduced factors were used to perform three training-test models using the following data sets: model 1) training: (IM_2, IM_3, UM_2, UM_3) test: (IM_1a, UM_1); model 2) training: (IM_1a, IM_3, UM_1, UM_3) test: (IM_2, UM_2); model 3) training: (IM_1a, IM_2, UM_1, UM_2) test: (IM_3, UM_3). The results of the test data are presented in Figure 6B.

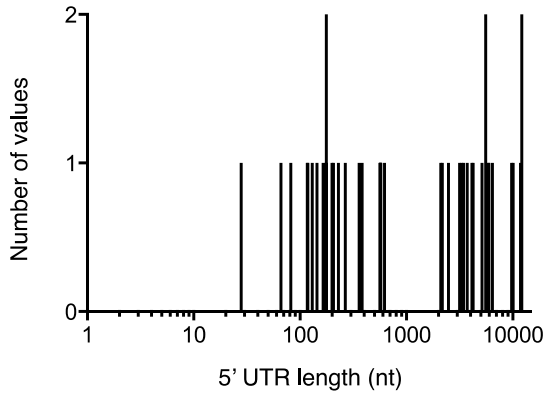


Fig S1. The 5′UTR lengths of mouse IGHV transcripts. The 5′UTR lengths of transcribed mouse IGHV genes were identified using the Immunogenetics Database (IMGT) website: **http://www.imgt.org/genedb/**. Several transcripts show very long 5′UTRs, for these genes it would not be possible to sequence a full-length VDJ region using a template-switching reaction [5' rapid amplification of cDNA ends (5' RACE)] (5′RACE) followed by PCR with a uniform primer. Full-length sequences that are returned by 5′RACE may be due to early termination products of 1st-strand synthesis, which would have their own inherent biases.

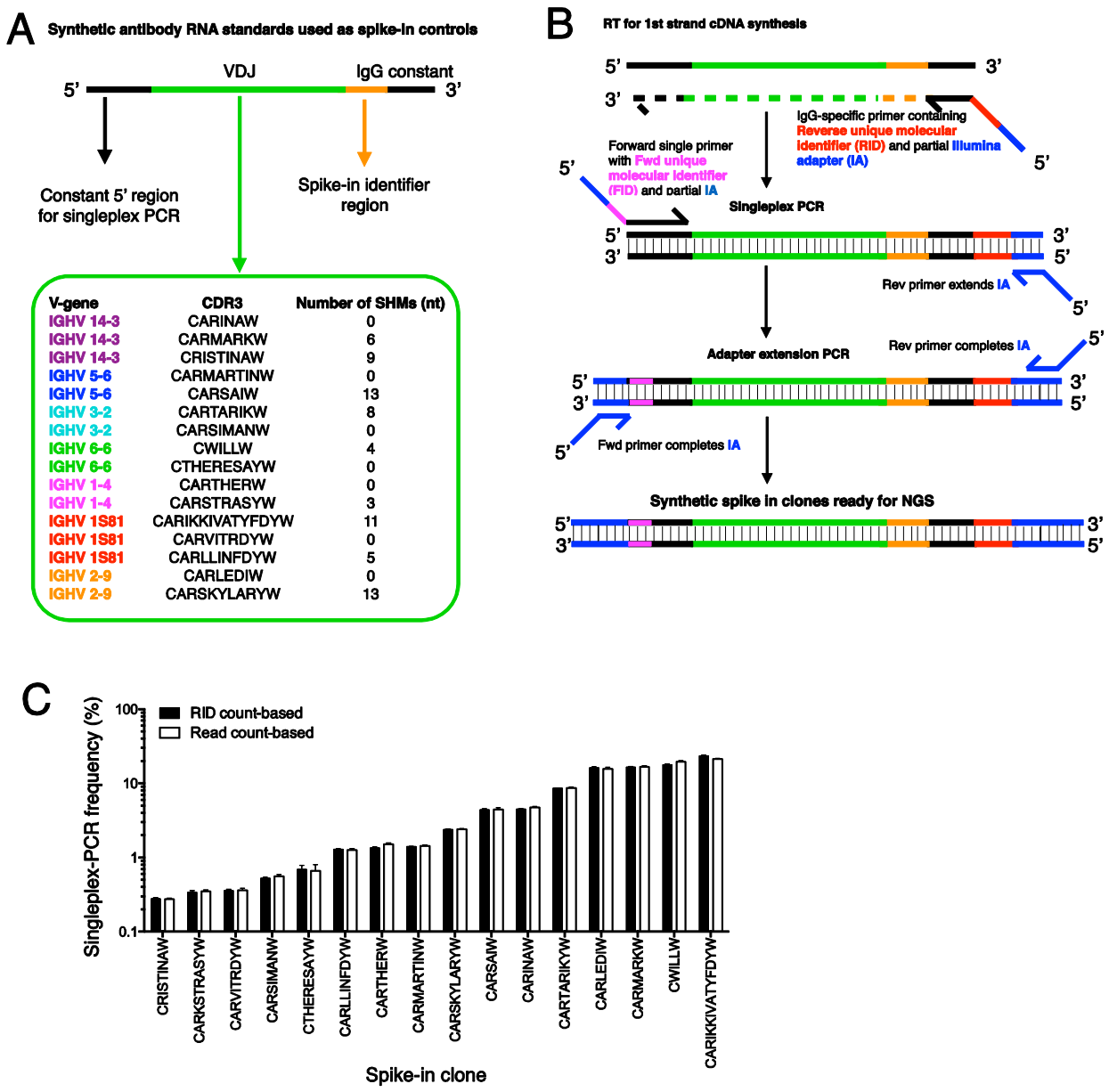


Fig. S2. Antibody synthetic spike-in genes. (**A**) All 16 synthetic spike-in clones were designed to have a 5′ constant region used for singleplex PCR controls, a spike-in identifier region for ddPCR and bioinformatic sorting (following Ig-seq), and a partial region corresponding to mouse IgG constant region. The 16 spike-in clones have 16 unique CDR3 amino acid (a.a.) sequences, 7 different V-genes, designed positions of somatic hypermutation (SHM). More details on spike-in design can be found in Materials and Methods. (**B**) Workflow for library preparation of spike-ins for control experiments. A reverse transcription (RT) step is performed with mouse IgG-specific primer (TAK_402) to generate first-strand cDNA with an RID and IA overhang. Singleplex PCR is performed with constant forward primer with an FID and IA overhang (TAK_472) and constant reverse primer specific for IA (TAK_423), followed by adapter-extension PCR with constant forward primer (TAK_424) and reverse primer with Illumina index sequence (e.g., TAK_531_IDX1). A complete list of primer sequences can be found in table S4. FID regions were used for error correction on RID regions (see Materials and Methods). (**C**) Spike-in frequencies are shown as means ± STD (based on RID-counts or Read-counts) obtained from replicate libraries (n = 5) generated by singleplex PCR from a master stock pool (see Materials and Methods); datasets used are J1-J5 (see table S1). For all further analysis singleplex PCR spike-in frequencies were RID-count-based mean values.

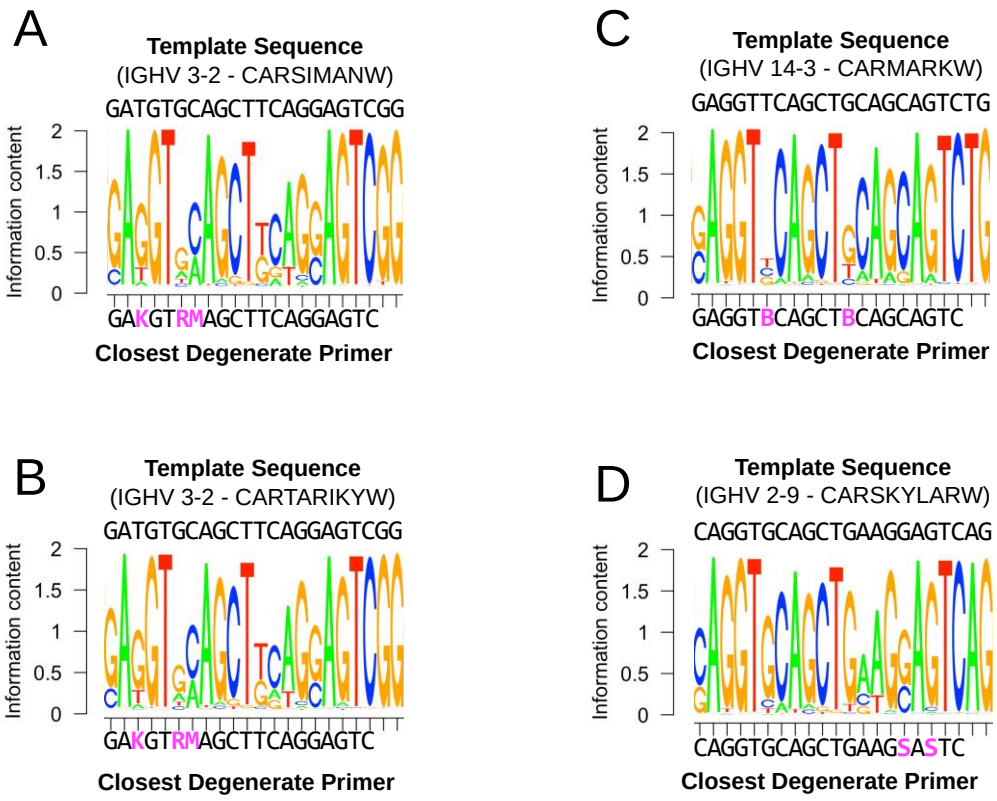


Fig. S3. Nucleotide sequence logos of the primer-binding regions of selected spike-in clones. A high diversity of nucleotides is observed in spike-in positions corresponding to both degenerate positions and non-degenerate positions in primers, suggesting the substantial mispriming is taking place during amplification. Mispriming appears to be systematic as different spike-ins sharing the same V-gene (e.g., IGHV 3-2) have similar nucleotide degeneracy in their primer binding regions. Ig-seq data are from multiplex PCR library preparation from mouse splenic cDNA with synthetic spike-ins (dataset Reddy-PS-1 was used, see **table S2**).

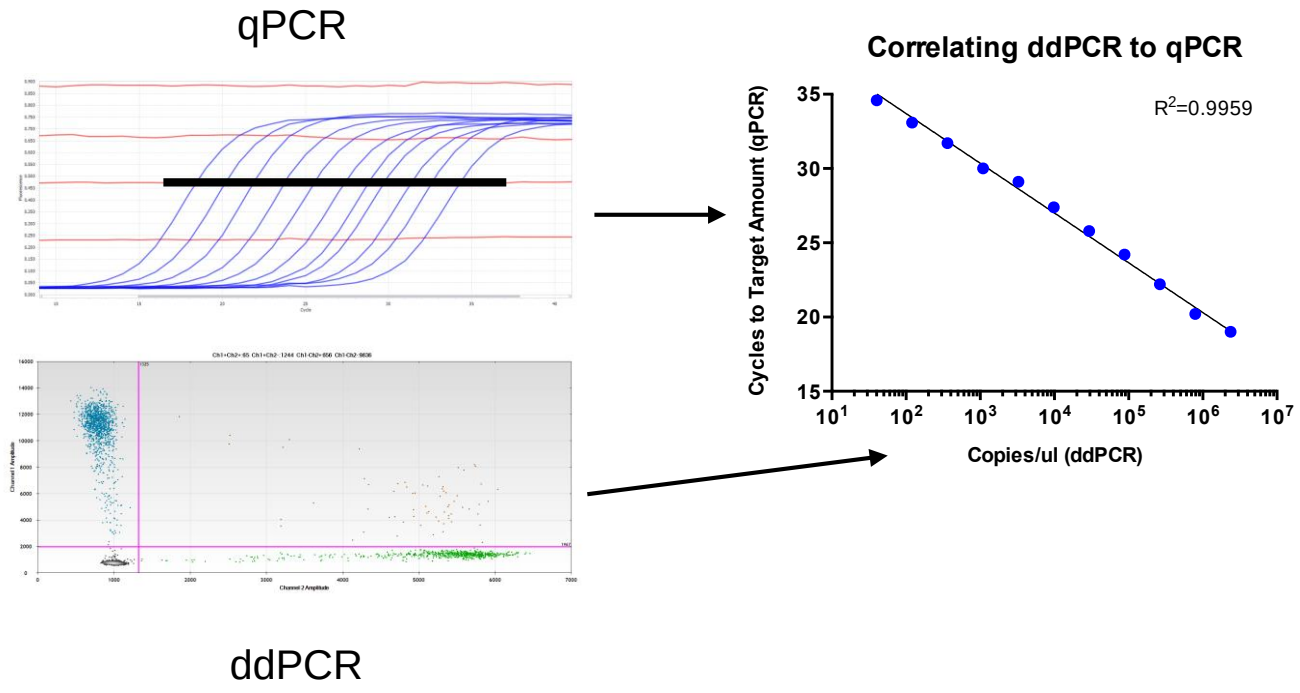


Fig. S4. Precise library quantification by linking qPCR to ddPCR. qPCR was used to visualize when reaction saturation occurs based on template amount and number of cycles, this data was then correlated to a ddPCR assay (using a synthetic minigene) which had a full natural antibody gene and Illumina adapters. This correlation allowed us to determine the proper number of cycles to use for the 2nd-step adapter-extension PCR reaction based on the desired number of ddPCR quantified copies of 1st-step multiplex PCR product. More details can be found in Materials and Methods.

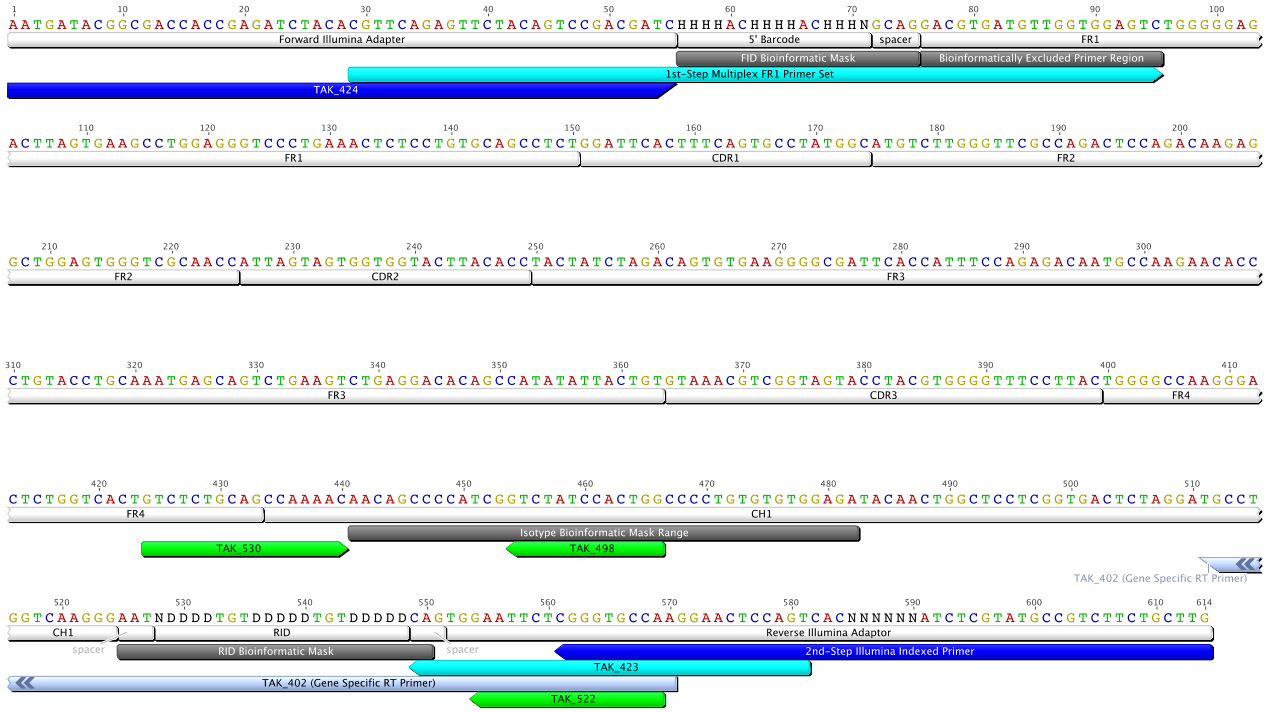


Fig. S5. Annotated example of biological sequence obtained from MAF library preparation.

Annotated colors correspond to the following:
White: Antibody sequence annotations
Dark grey: Bioinformatic annotations
Pale blue: RID tagged RT gene specific primer
Teal: 1st-step PCR primers
Dark blue: 2nd-step PCR primers
Green: ddPCR primers

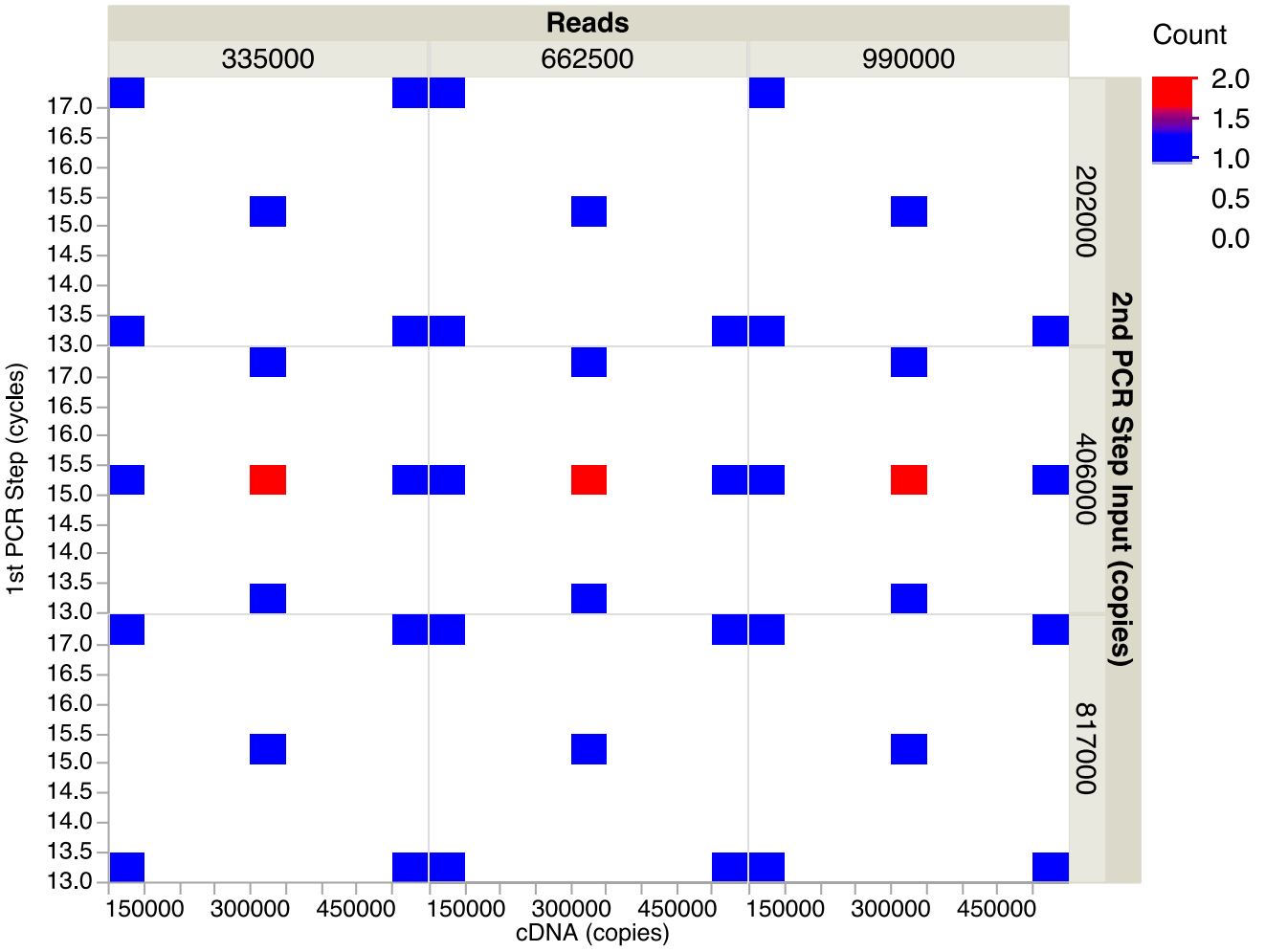


Fig. S6. Design of experiments (DoE) for library preparation optimization. A DoE, response surface central composite designed experiment was used to determine how several factors influenced library preparation and Ig-seq data. Factors shown are: 1) quantity of input cDNA copies, 2) number of cycles in the 1st-step multiplex PCR, 3) quantity of DNA copies input into 2nd-step adapter extension PCR, and 4) number of reads analyzed (based on *in silico* random sampling prior to bioinformatics pipeline analysis). For more information on DoE see Supplementary Materials and Methods.

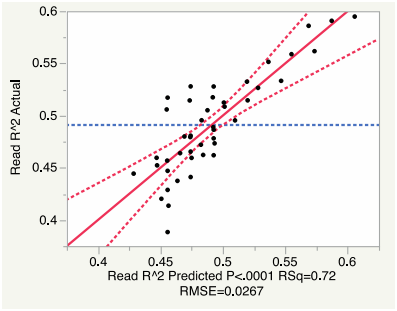
A

Uncorrected read-based frequency bias

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
1st PCR Step (cycles)	-0.026497	0.005656	-4.68	<.0001*
Reads	5.9099e-8	1.537e-8	3.84	0.0007*
cDNA (copies)*1st PCR Step (cycles)*Log(2nd PCR Step Input Copies)	0.0839914	0.02203	3.81	0.0007*
cDNA (copies)*1st PCR Step (cycles)	0.0185172	0.006549	2.83	0.0087*
cDNA (copies)*Log(2nd PCR Step Input Copies)	-0.058982	0.021885	-2.70	0.0120*
1st PCR Step (cycles)*1st PCR Step (cycles)	0.0226794	0.009577	2.37	0.0253*
cDNA (copies)*Reads	3.3554e-8	2.018e-8	1.66	0.1080
cDNA (copies)	-0.007616	0.005656	-1.35	0.1894
Log(2nd PCR Step Input Copies)	-0.024141	0.01886	-1.28	0.2114
Reads*Reads	-8.63e-14	7.82e-14	-1.10	0.2793
Reads*Log(2nd PCR Step Input Copies)	6.3873e-8	6.719e-8	0.95	0.3502
cDNA (copies)*cDNA(copies)	0.0066794	0.009577	0.70	0.4915
1st PCR Step (cycles)*Log(2nd PCR Step Input Copies)	0.0118846	0.021885	0.54	0.5916
cDNA (copies)*Reads*Log(2nd PCR Step Input Copies)	4.1037e-8	7.706e-8	0.53	0.5987
1st PCR Step (cycles)*Reads*Log(2nd PCR Step Input Copies)	-4.071e-8	7.706e-8	-0.53	0.6016
Log(2nd PCR Step Input Copies)*Log(2nd PCR Step Input Copies)	0.0365919	0.104025	0.35	0.7277
cDNA (copies)*1st PCR Step (cycles)*Reads	6.5167e-9	2.338e-8	0.28	0.7826
1st PCR Step (cycles)*Reads	4.266e-9	2.018e-8	0.21	0.8342

B



C

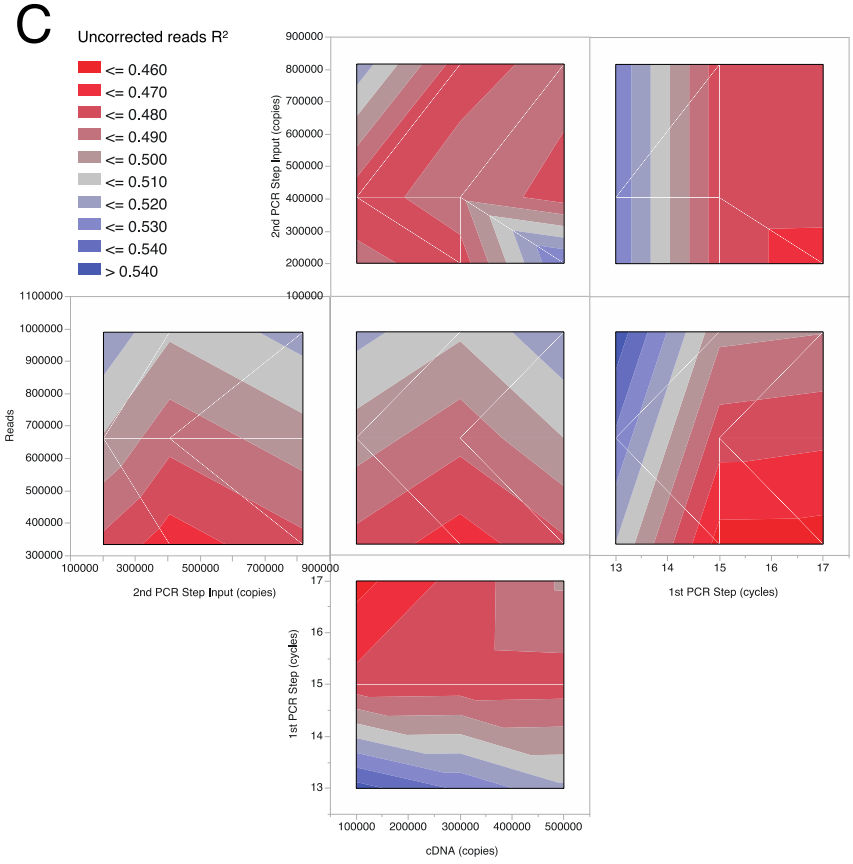


Fig. S7. Response surface methodology analysis of clonal frequency bias with uncorrected data. The R^2 values correlating spike-ins multiplex PCR vs. control singleplex PCR library preparation are based computing frequencies using uncorrected read counts. For more information on DoE see Supplementary Materials and Methods.

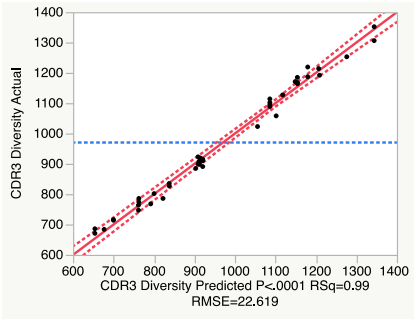
A

CDR3 diversity

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
cDNA (copies)	174.68603	4.69752	37.19	<.0001*
Reads	0.0003755	1.277e-5	29.41	<.0001*
cDNA (copies)*cDNA(copies)	-116.1517	7.95428	-14.60	<.0001*
cDNA (copies)*Reads	0.0001763	1.676e-5	10.52	<.0001*
Reads*Reads	-4.91e-10	6.49e-11	-7.57	<.0001*
Log(2nd PCR Step Input Copies)	104.51176	15.66333	6.67	<.0001*
Reads*Log(2nd PCR Step Input Copies)	0.0002479	5.58e-5	4.44	0.0001*
cDNA (copies)*Reads*Log(2nd PCR Step Input Copies)	0.0001792	0.000064	2.80	0.0093*
Log(2nd PCR Step Input Copies)*Log(2nd PCR Step Input Copies)	-172.1165	86.39518	-1.99	0.0565
cDNA (copies)*Log(2nd PCR Step Input Copies)	35.996971	18.17625	1.98	0.0579
1st PCR Step (cycles)*Log(2nd PCR Step Input Copies)	28.926654	18.17625	1.59	0.1231
cDNA (copies)*1st PCR Step (cycles)*Log(2nd PCR Step Input Copies)	22.995985	18.29675	1.26	0.2196
1st PCR Step (cycles)*1st PCR Step (cycles)	9.8482689	7.95428	1.24	0.2263
1st PCR Step (cycles)*Reads*Log(2nd PCR Step Input Copies)	5.464e-5	0.000064	0.85	0.4008
1st PCR Step (cycles)*Reads	-5.37e-6	1.676e-5	-0.32	0.7511
cDNA (copies)*1st PCR Step (cycles)	1.0581282	5.43928	0.19	0.8472
cDNA (copies)*1st PCR Step (cycles)*Reads	2.1229e-6	1.942e-5	0.11	0.9138
1st PCR Step (cycles)	-0.325376	4.69752	-0.07	0.9453

B



C

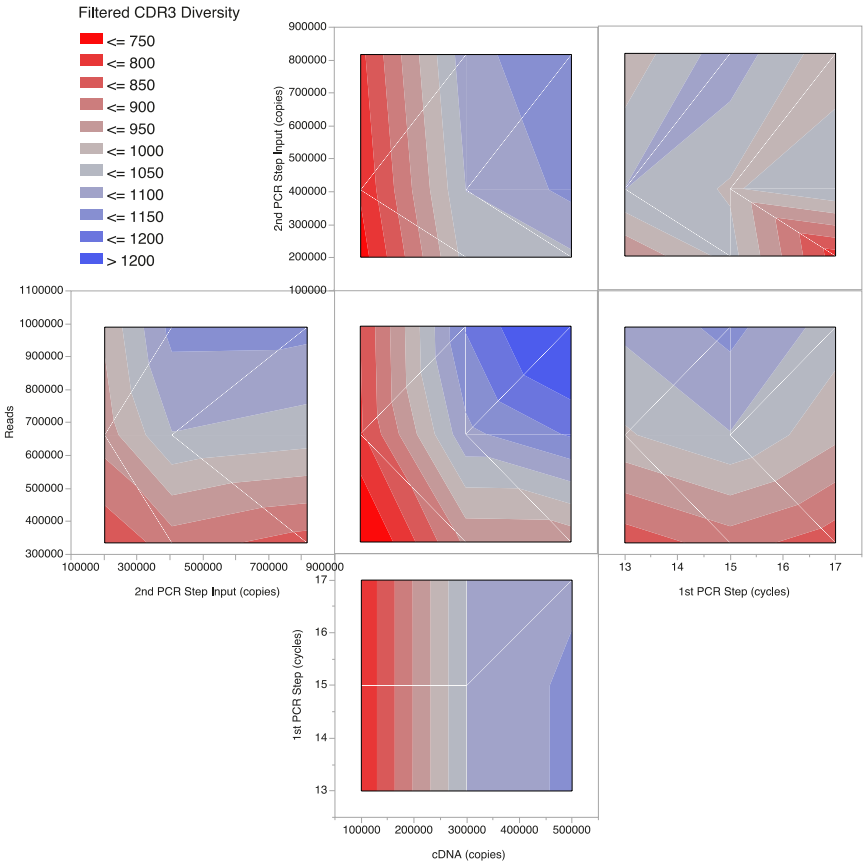


Fig. S8. Response surface methodology analysis of CDR3 diversity. The CDR3 diversity reported is based on MAF corrected data that removes flagged hotspot errors and clones that have less than three RIDs with three reads each. For more information on DoE see Supplementary Materials and Methods.

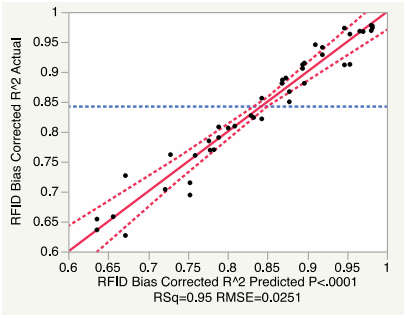
A

MAF-corrected frequency bias

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
cDNA (copies)	-0.098419	0.005195	-18.95	<.0001*
Reads	1.8304e-7	1.412e-8	12.96	<.0001*
cDNA (copies)*Reads	9.2977e-8	1.854e-8	5.02	<.0001*
Reads*Reads	-2.89e-13	7.18e-14	-4.03	0.0004*
cDNA (copies)*cDNA(copies)	0.0299659	0.008796	3.41	0.0021*
Log(2nd PCR Step Input Copies)*Log(2nd PCR Step Input Copies)	-0.307517	0.09554	-3.22	0.0033*
Reads*Log(2nd PCR Step Input Copies)	1.8172e-7	6.171e-8	2.94	0.0066*
1st PCR Step (cycles)*1st PCR Step (cycles)	0.0248492	0.008796	2.82	0.0088*
Log(2nd PCR Step Input Copies)	0.0411777	0.017321	2.38	0.0248*
cDNA (copies)*Reads*Log(2nd PCR Step Input Copies)	1.5545e-7	7.078e-8	2.20	0.0368*
1st PCR Step (cycles)	-0.009368	0.005195	-1.80	0.0825
1st PCR Step (cycles)*Log(2nd PCR Step Input Copies)	0.0330879	0.0201	1.65	0.1113
1st PCR Step (cycles)*Reads*Log(2nd PCR Step Input Copies)	-5.774e-8	7.078e-8	-0.82	0.4218
cDNA (copies)*1st PCR Step (cycles)*Reads	1.4583e-8	2.148e-8	0.68	0.5029
1st PCR Step (cycles)*Reads	8.0222e-9	1.854e-8	0.43	0.6686
cDNA (copies)*1st PCR Step (cycles)*Log(2nd PCR Step Input Copies)	-0.005436	0.020234	-0.27	0.7902
cDNA (copies)*1st PCR Step (cycles)	0.0006495	0.006015	0.11	0.9148
cDNA (copies)*Log(2nd PCR Step Input Copies)	-0.001596	0.0201	-0.08	0.9373

B



C

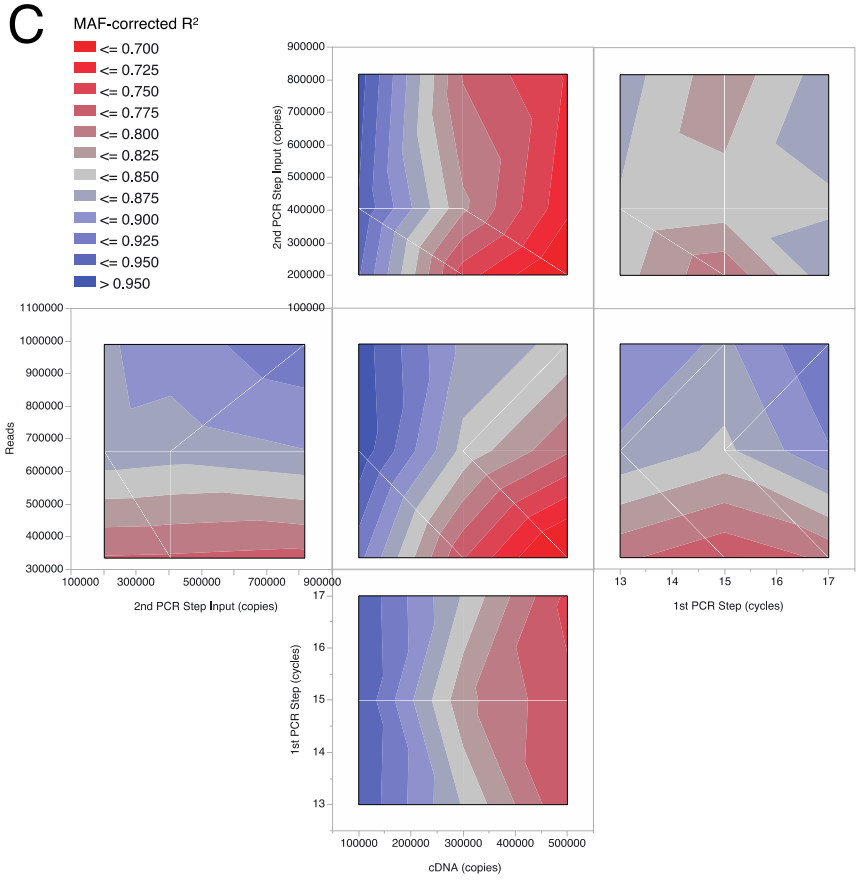


Fig. S9. Response surface methodology analysis of clonal frequency bias with MAF-corrected data. The R^2 values correlating spike-ins multiplex PCR vs. control singleplex PCR library preparation is based on MAF bias corrected frequencies. For more information on DoE see Supplementary Materials and Methods.

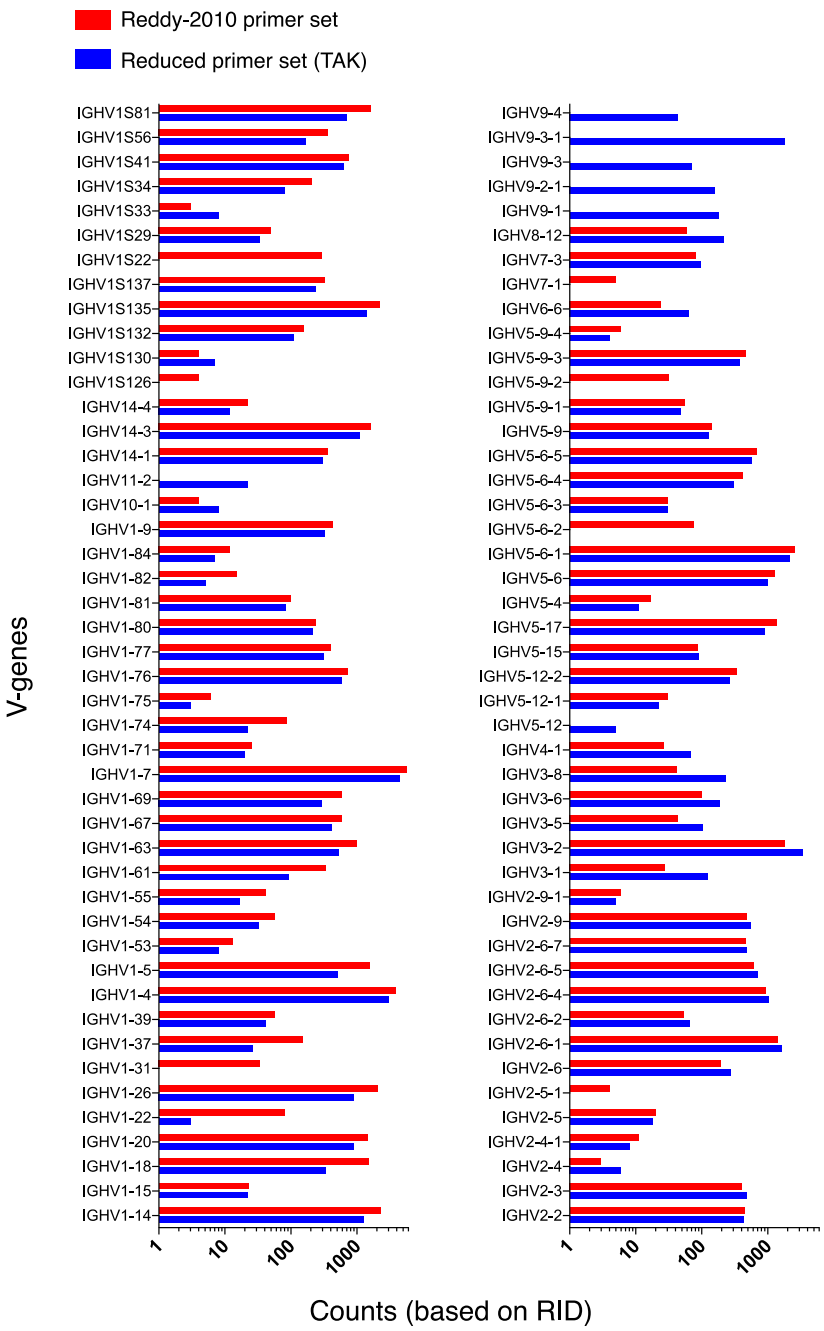


Fig. S10. Comparison of V-gene coverage using new reduced primer set (TAK) and previously published primer set (Reddy-2010). TAK primer set has a smaller number of primers but is still able to cover the majority of V-genes covered by Reddy-2010. Ig-seq was performed on samples from library preparation with the same starting material (immunized mouse splenic cDNA with ~10% spike-ins), but using TAK or Reddy-2010 primer sets for the multiplex PCR step. 650k preprocessed sequences for each data set were processed in the MAF bioinformatic pipeline. Datasets used are Reddy-PS-Compare and TAK-PS-Compare, see **table S2.** A complete list of primer sequences can be found in **table S5.**

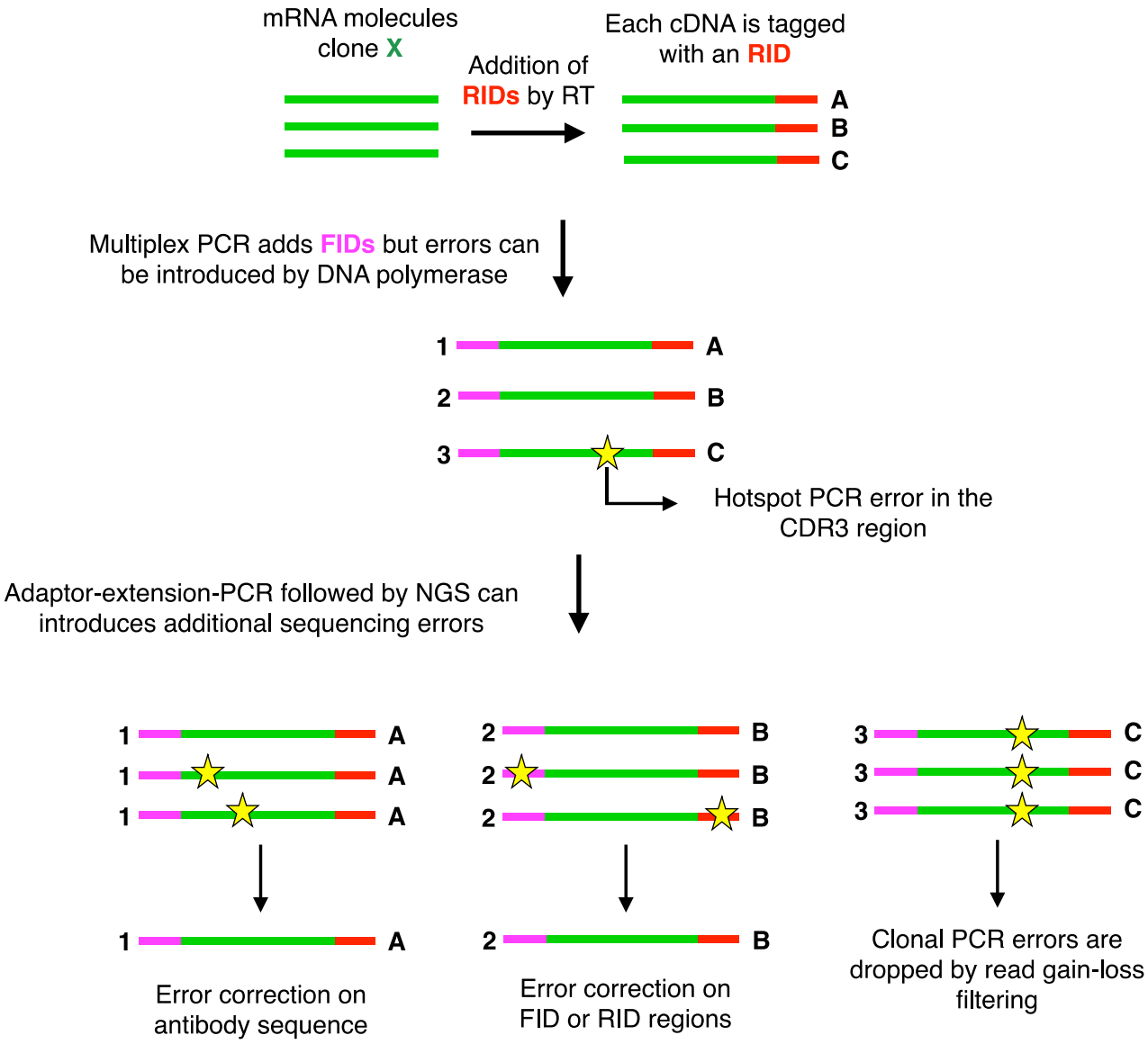


Fig. S11. Schematic of multistage error correction pipeline. For simplicity, Illumina adaptor regions are not shown. Further details of error correction methods can be found in Results and Materials and Methods.

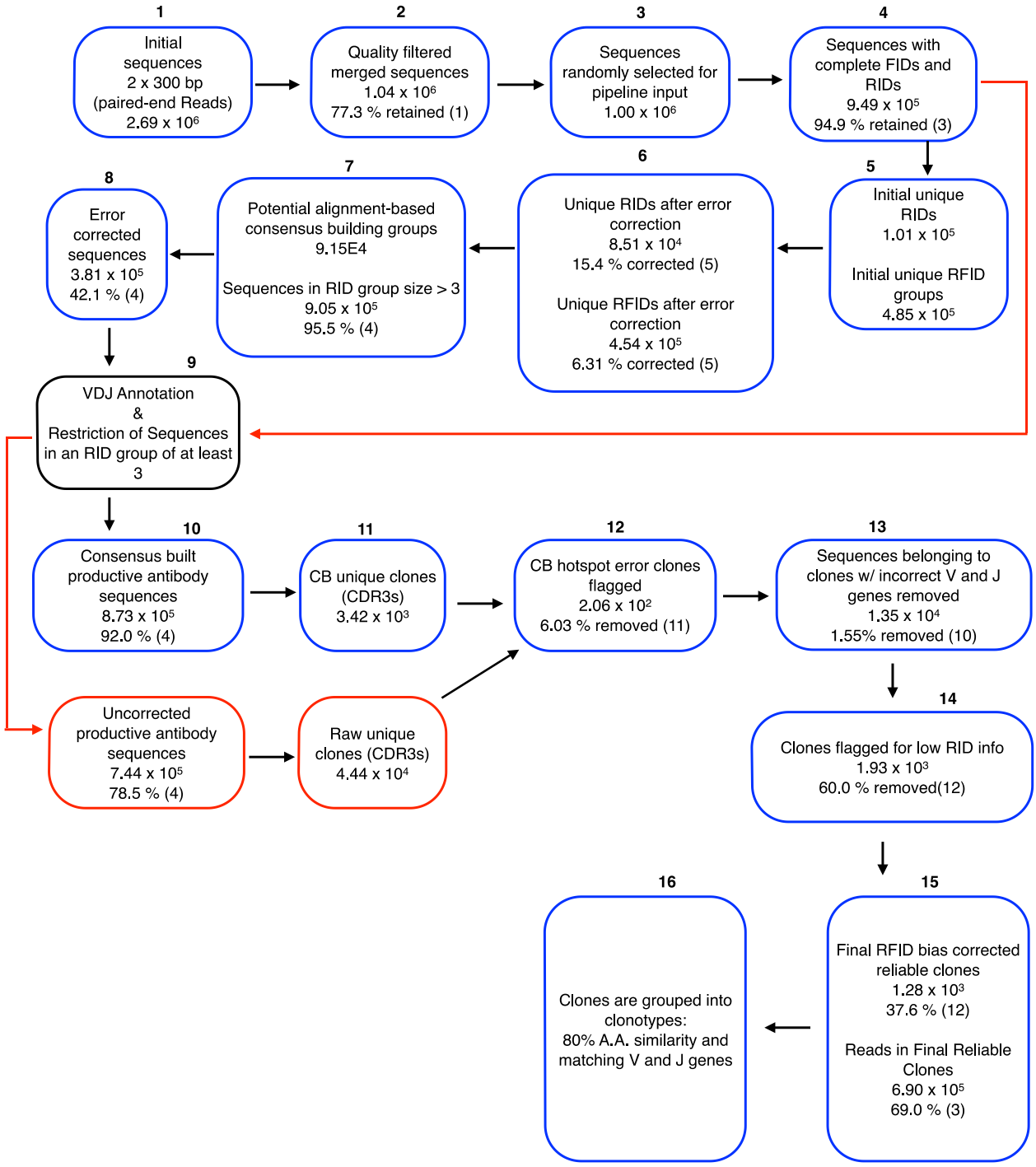


Fig. S12. Flow chart of multistage error correction pipeline. Ig-seq read numbers and statistics correspond to representative MAF library preparation from immunized mouse splenic cDNA with ~10% synthetic spike-ins (dataset used is IM_1a, see table S7).

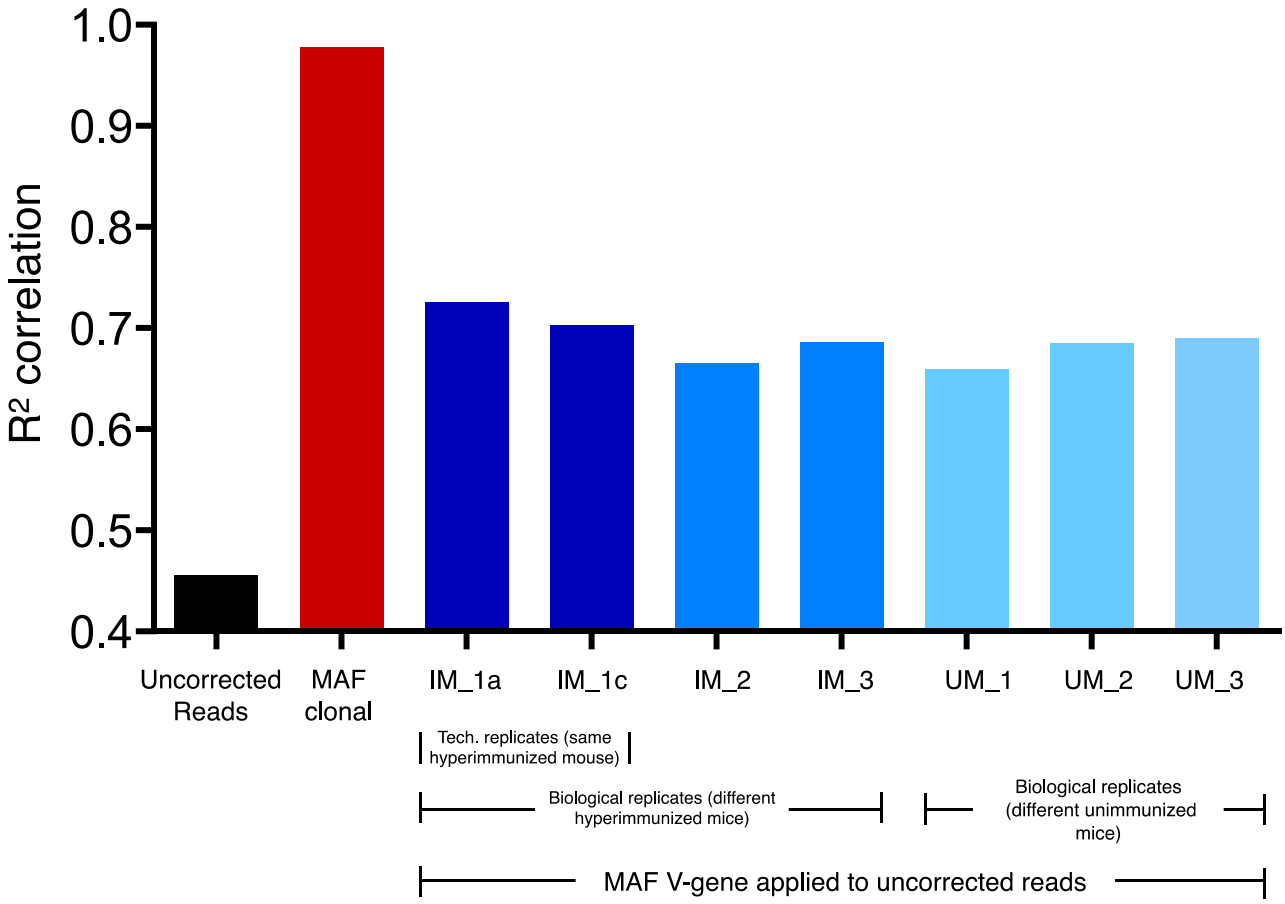


Fig. S14. Bias correction using MAF V-gene bias factor. R^2 correlation values of spike-in clonal frequencies with multiplex-PCR versus singleplex-PCR. Evaluation of the MAF V-gene bias factor (F3+) on uncorrected reads from different datasets (see Materials and Methods). Correction of uncorrected reads was performed by using the MAF V-gene bias factors from the same data set (IM_1a) and a technical replicate (IM_1b). Correction using MAF V-gene bias factors (from IM_1a) on biological replicates (separate hyperimmunized mice and untreated mice). Uncorrected reads and MAF clonal bias correction from dataset IM_1a are shown for reference. Ig-seq data are from library sample preparations of splenic cDNA with synthetic spike-ins from hyperimmunized mice (n = 3) and untreated mice (n = 3) (datasets used for hyperimmunized are IM_1a, IM_1b, IM_2, IM_3; datasets used for untreated mice are UM_1, UM_2, UM_3, see table S7). Singleplex spike-in frequencies are mean values obtained from replicate libraries (n = 5) generated by singleplex-PCR (see fig. S2 and table S1).

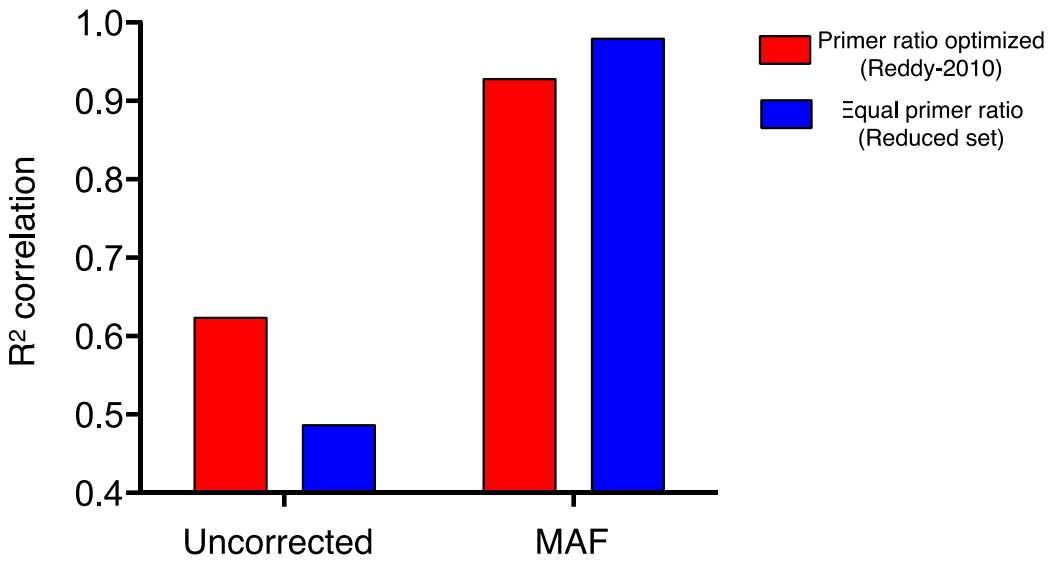


Fig. S15. Comparison of bias correction with new reduced primer set (TAK) and previously published primer set (Reddy-2010). R^2 correlation values of spike-in clonal frequencies with multiplex PCR versus singleplex PCR using different primer sets and different bias correction methods. MAF clonal bias correction with TAK primer set achieves the highest accuracy. Ig-seq was performed on samples from library preparation with the same starting material (immunized mouse splenic cDNA with ~10% spike-ins), but using primer set Reddy-2010 or primer set TAK for the multiplex-PCR step. 650k preprocessed sequences for each data set were processed in the MAF bioinformatic pipeline. Datasets used are Reddy-PS-Compare and TAK-PS-Compare, see table S2. A complete list of primer sequences can be found in table S5. Singleplex spike-in frequencies are mean values obtained from replicate libraries (n = 5) generated by singleplex-PCR (see fig. S2 and table S1).

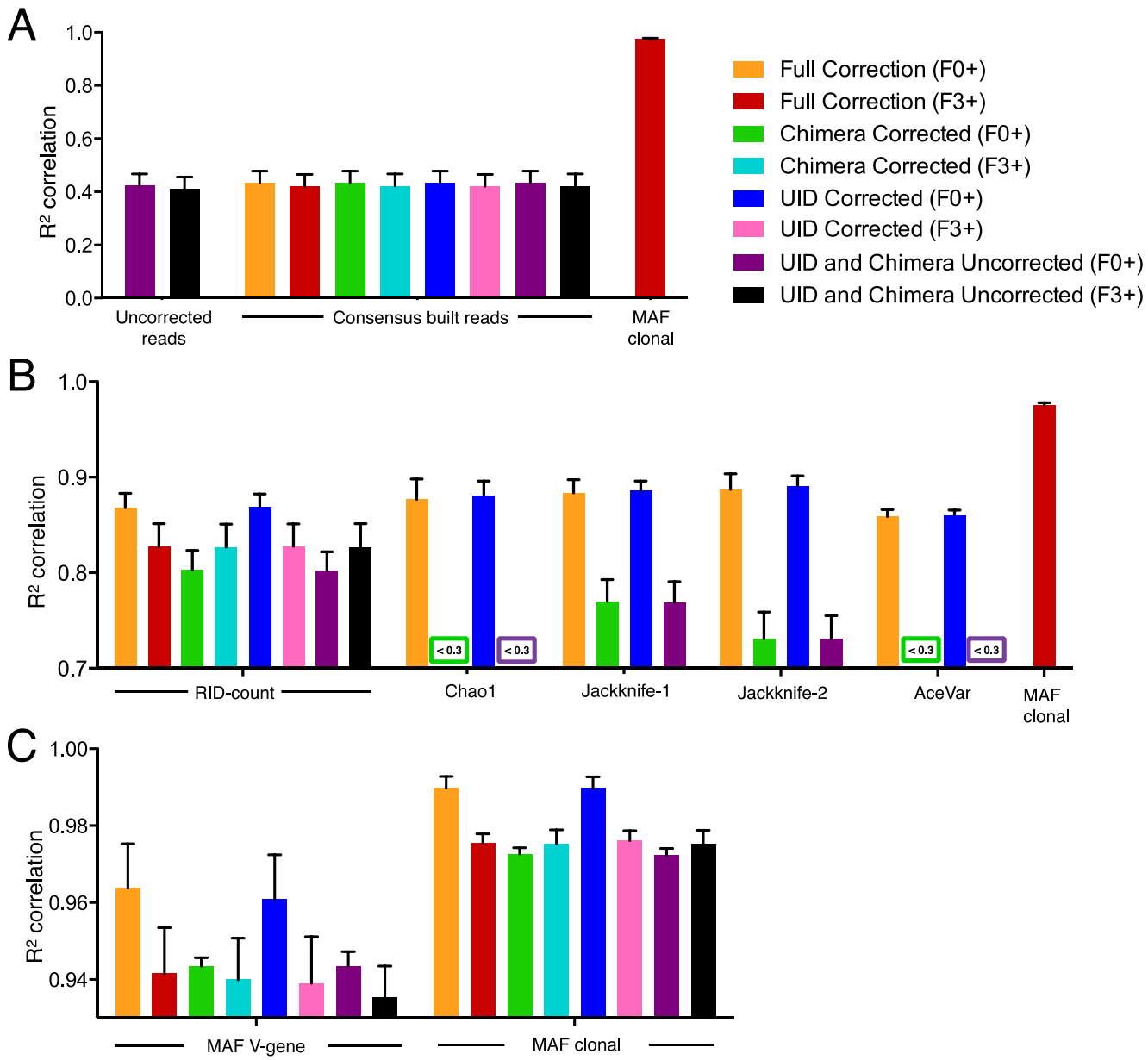


Fig. S13. Error correction effects on various bias correction methods. R^2 correlation values of spike-in clonal frequencies from multiplex-PCR versus singleplex-PCR using different versions of bias correction and different stages of error correction and filtering. Based on fig. S12, UID Corrected is stage 6; Chimera Corrected is Stage 13, and Full Correction is product of the full pipeline. Unrestricted (F0+) and restricted (F3+) refers to unfiltered productive sequences and filtering out sequences belonging to RID groupings with less than three reads, respectively. (**A**) Comparison of read counting based frequencies to MAF clonal. Raw reads are based on non-consensus built Ig-seq data. Consensus built reads correspond to clonal frequencies computed by reads counts after consensus RID building. (**B**) Comparison of RID counting methods to determine clonal frequencies. RID-count simply computes the RID count based on the total number of unique RIDs associated with a clone. The non-parametric species richness estimators take into account the total number of unique RIDs as well as the number RIDs with very few reads (e.g. a single read (Chao1 and 1st-order Jackknife) or a single and two reads (2nd-order Jackknife). Acevar was computed with k=10 and uses the traditional ACE if $CV_{rare} \leq 0.8$ or ACE-1 if $CV_{rare} > 0.8$. (**C**) MAF bias factor (FID / RID) was calculated with either median biological V-gene MAF bias factors (MAF V-gene) or individual clonal-based MAF bias factors (MAF Clonal) (see Materials and Methods). In all cases the MAF clonal bias correction achieves the highest accuracy. Also other methods of bias correction such as RID-counting or non-parametric species richness estimators benefit substantially by the MAF error correction pipeline. Ig-seq data are from replicate library sample preparations (n = 3) from mouse splenic cDNA with ~10% synthetic spike-ins (data are presented as means ± STD and are from replicate datasets IM_1a, _1b,_1c, see table S7). Singleplex spike-in frequencies are mean values obtained from replicate libraries (n = 5) generated by singleplex PCR (see fig. S2 and table S1).

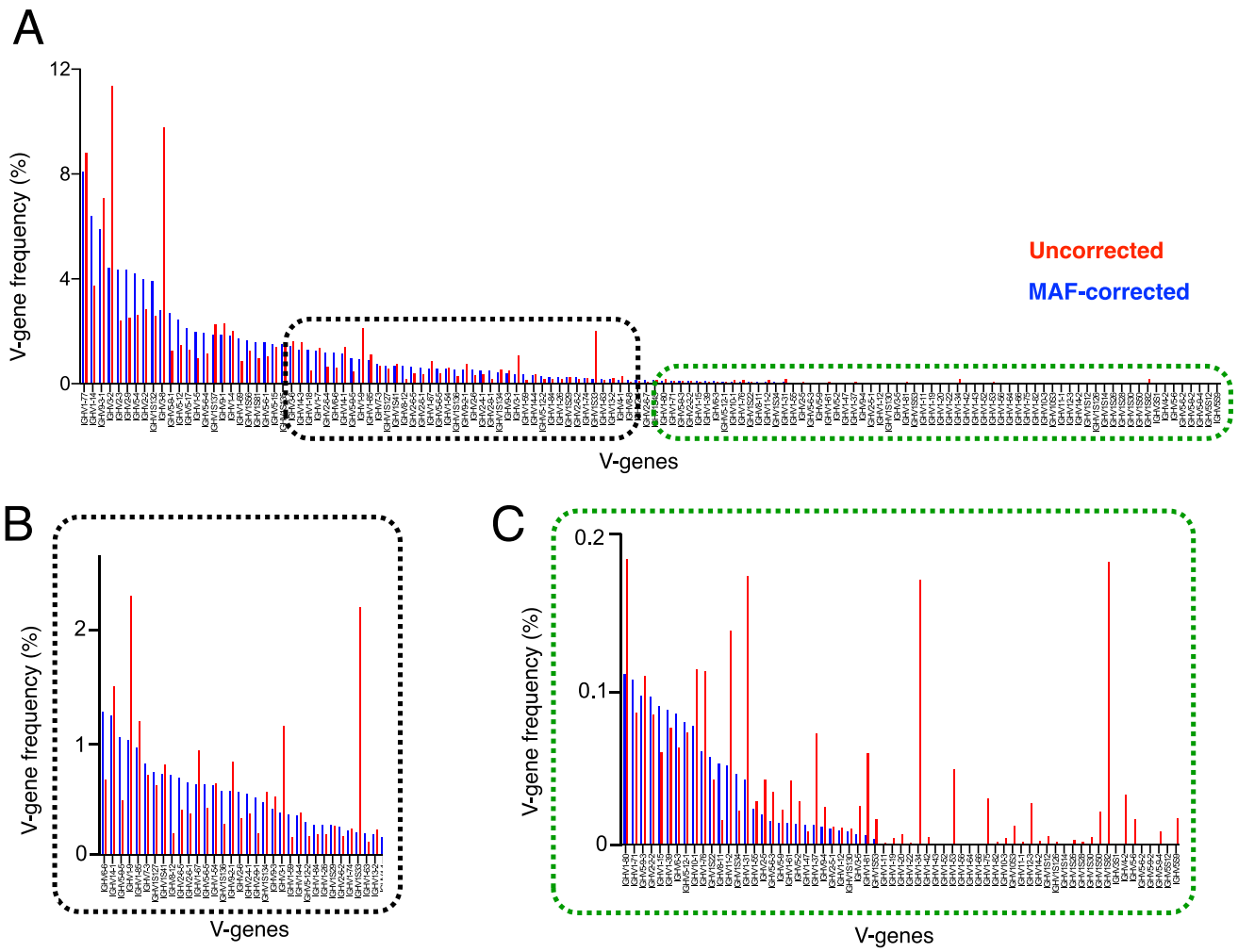


Fig. S16. Comparison of V-gene (germlines) before and after MAF correction. Ig-seq read numbers and statistics correspond to representative MAF library preparation from mouse splenic cDNA with synthetic spike-ins (dataset used is IM_1a, see table S7).

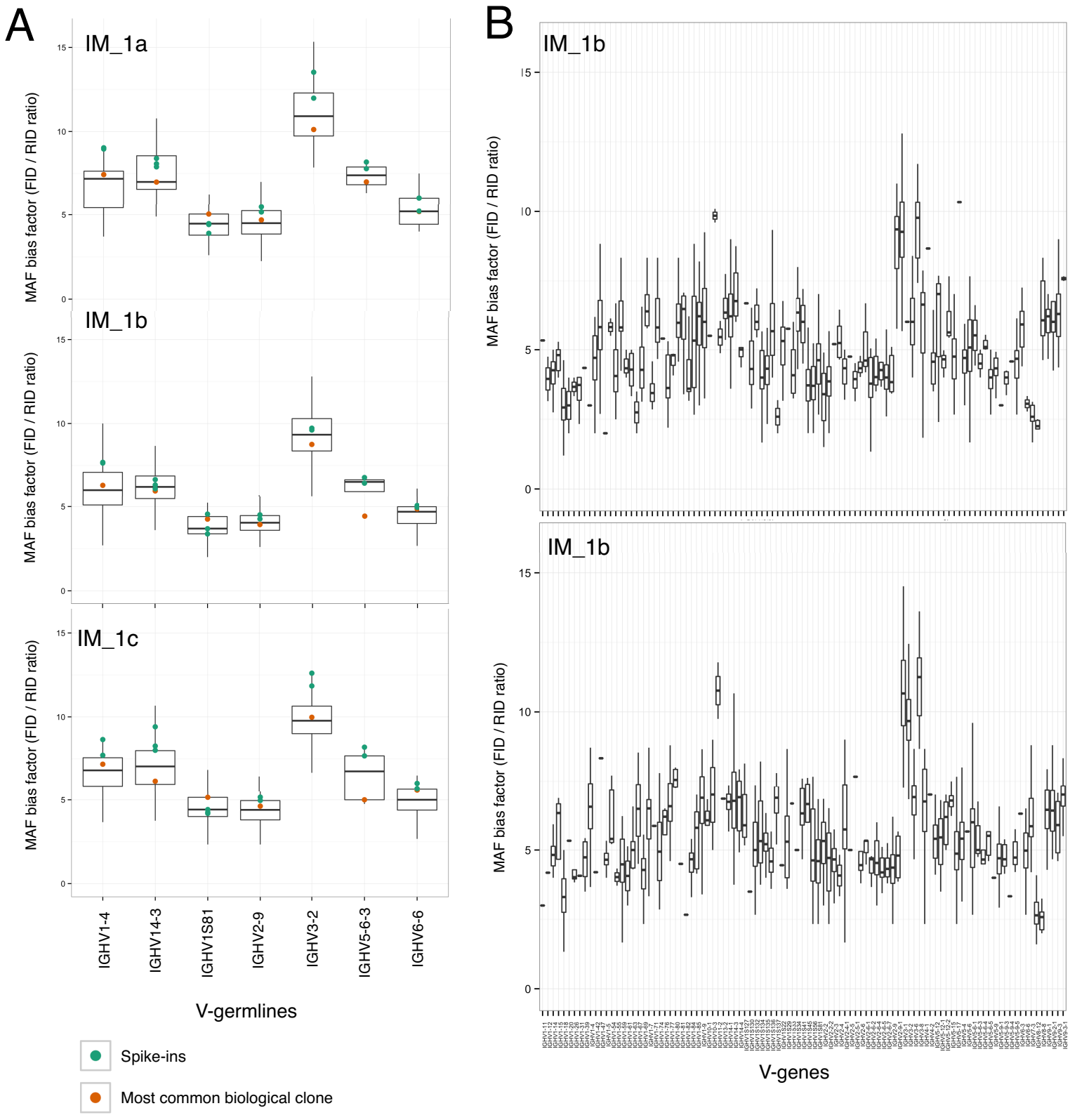
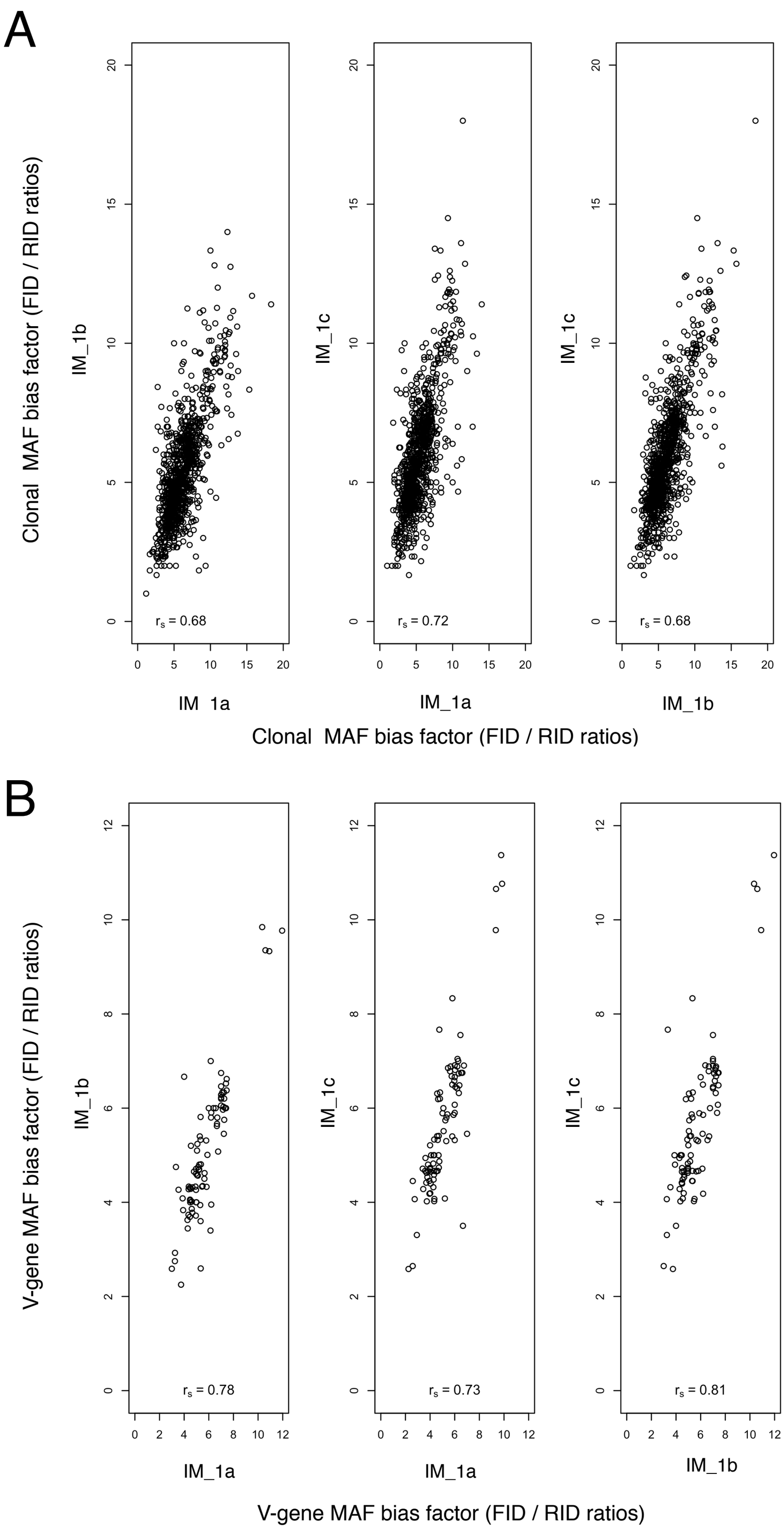

A
IM_1a
IM_1b
IM_1c
MAF bias factor (FID / RID ratio)
0
5
10
15
IGHV1-4
IGHV14-3
IGHV1S81
IGHV2-9
IGHV3-2
IGHV5-6-3
IGHV6-6
V-germlines
Spike-ins
Most common biological clone
B
IM_1b
IM_1b
MAF bias factor (FID / RID ratio)
V-genes

Fig. S17. The MAF bias factor across V-genes. (**A**) The MAF bias factor (FID / RID) of spike-in V-genes or all V-genes (**B**) shows grouping within and across replicate datasets. MAF bias factor across all V-genes for replicate IM_1a can be found in Fig. 5E. Ig-seq data are from replicate MAF library sample preparations (n = 3) from mouse splenic cDNA with synthetic spike-ins (datasets used were IM_1a, _1b, _1c, see table S7).



Supplementary Figure 18 Correlation of MAF bias correction factor across datasets.

(**A**) Clonal- (**B**) and V-gene-based MAF bias correction ratio (FID / RID) of MAF replicate library preparations.
Ig-seq data are from replicate MAF library sample preparations (n = 3) from mouse splenic cDNA with synthetic spike-ins (datasets used were IM_1a, _1b, _1c, see **Supplementary Table 7**).

Fig. S18. Correlation of MAF bias correction factor across data sets. (**A**) Clonal- (**B**) and V-gene-based MAF bias factor (FID / RID) of MAF replicate library preparations. Ig-seq data are from replicate MAF library sample preparations (n = 3) from mouse splenic cDNA with synthetic spike-ins (datasets used were IM_1a, _1b, _1c, see table S7).

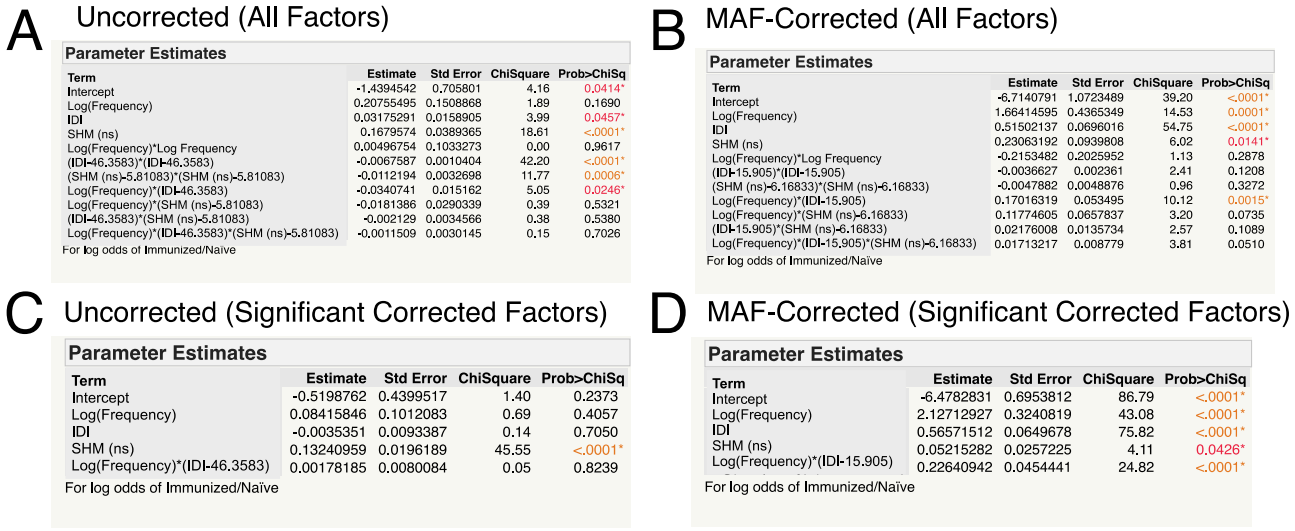


Fig. S19. Nominal logistic regression modeling based on Ig-seq clonotype measurements. The parameter estimates for nominal logistic regression models (trained with data sets: IM_1a, IM_2, IM_3, UM_1, UM_2, UM3) are shown for factors using uncorrected (**A, C**) and MAF corrected (**B, D**) data. First all factors were used in the models (**A, B**). The models were then reduced to the significant factors determined for the MAF corrected data (**C-D**). For more information, see **Supplementary Materials and Methods**.

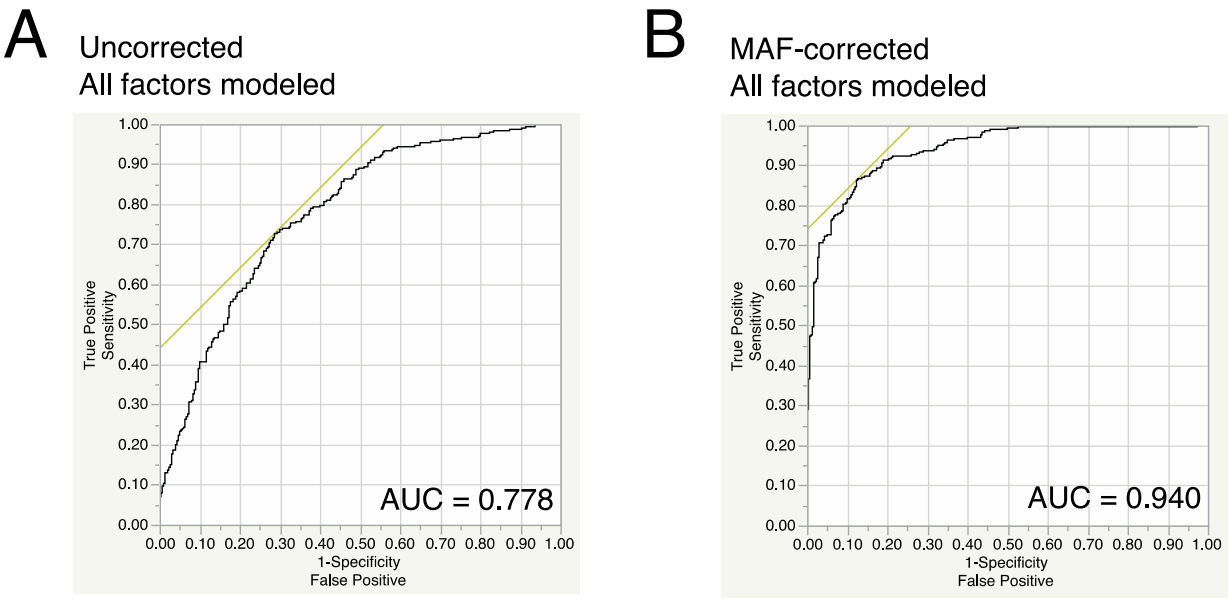


Fig. S20. Comparison of the sensitivity and specificity of the nominal logistic regression models. The receiver operating characteristics and area under the curve (AUC) for nominal logistic regression models (trained with data sets: IM_1a, IM_2, IM_3, UM_1, UM_2, UM3) are shown for factors using uncorrected (**A)** and MAF-corrected (**B)** data. For more information, see **Supplementary Materials and Methods**.

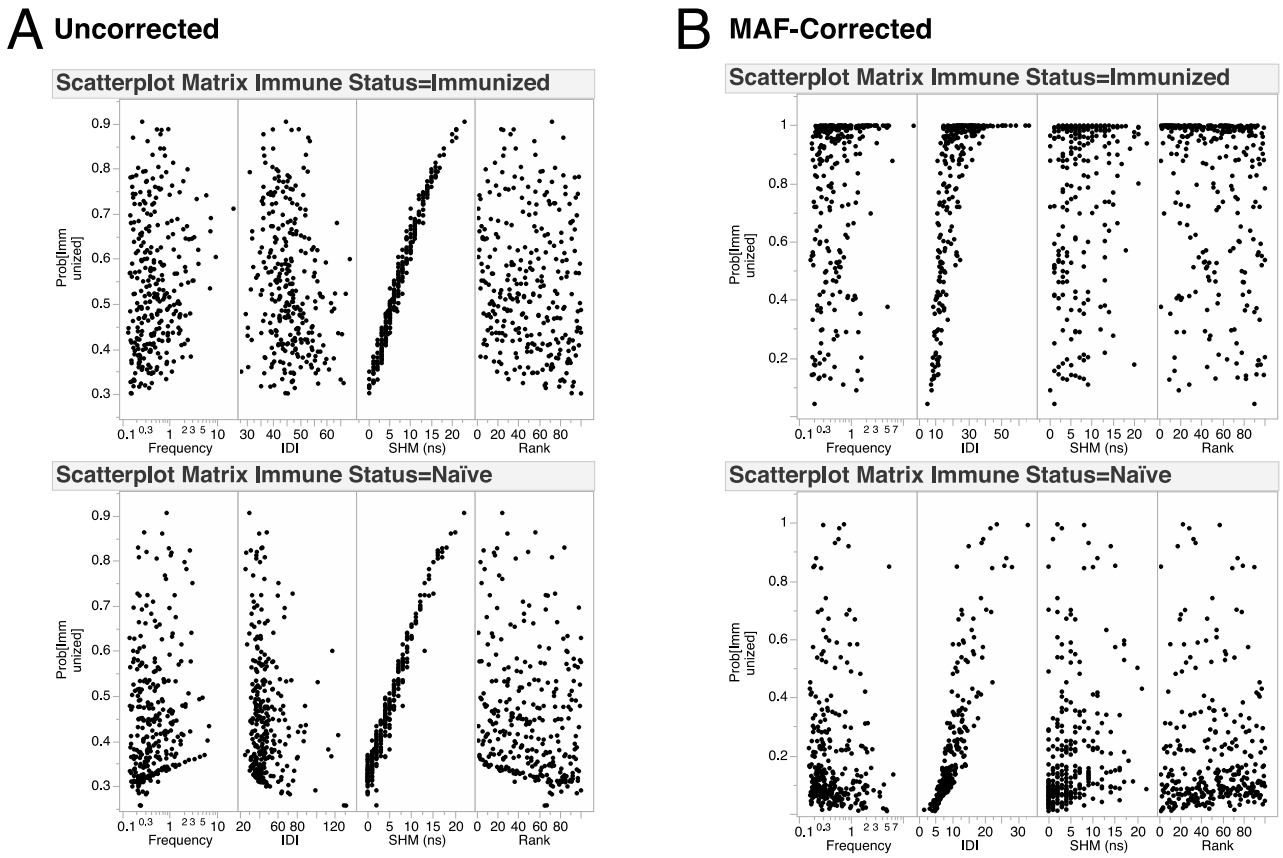


Fig. S21. Comparison of factor correlations with prediction probabilities of the nominal logistic regression models. The probabilities of clonotypes belonging to the hyperimmunized group from the nominal logistic regression models (trained with data sets: IM_1a, IM_2, IM_3, UM_1, UM_2, UM3) based on only MAF corrected significant factors are plotted with parameters used in the model (frequency, intraclonotype diversity index (IDI), and median non-silent somatic hypermutations (SHM (ns)), along with the non-model parameter rank based on clonotype frequency. Models fro both uncorrected (**A**) and MAF corrected (**B**) data are shown. The data sets for hyperimmunized mice (n=3) is shown on top and the data sets for the untreated mice (n=3) is shown on the bottom. For more information see Supplementary Materials and Methods.

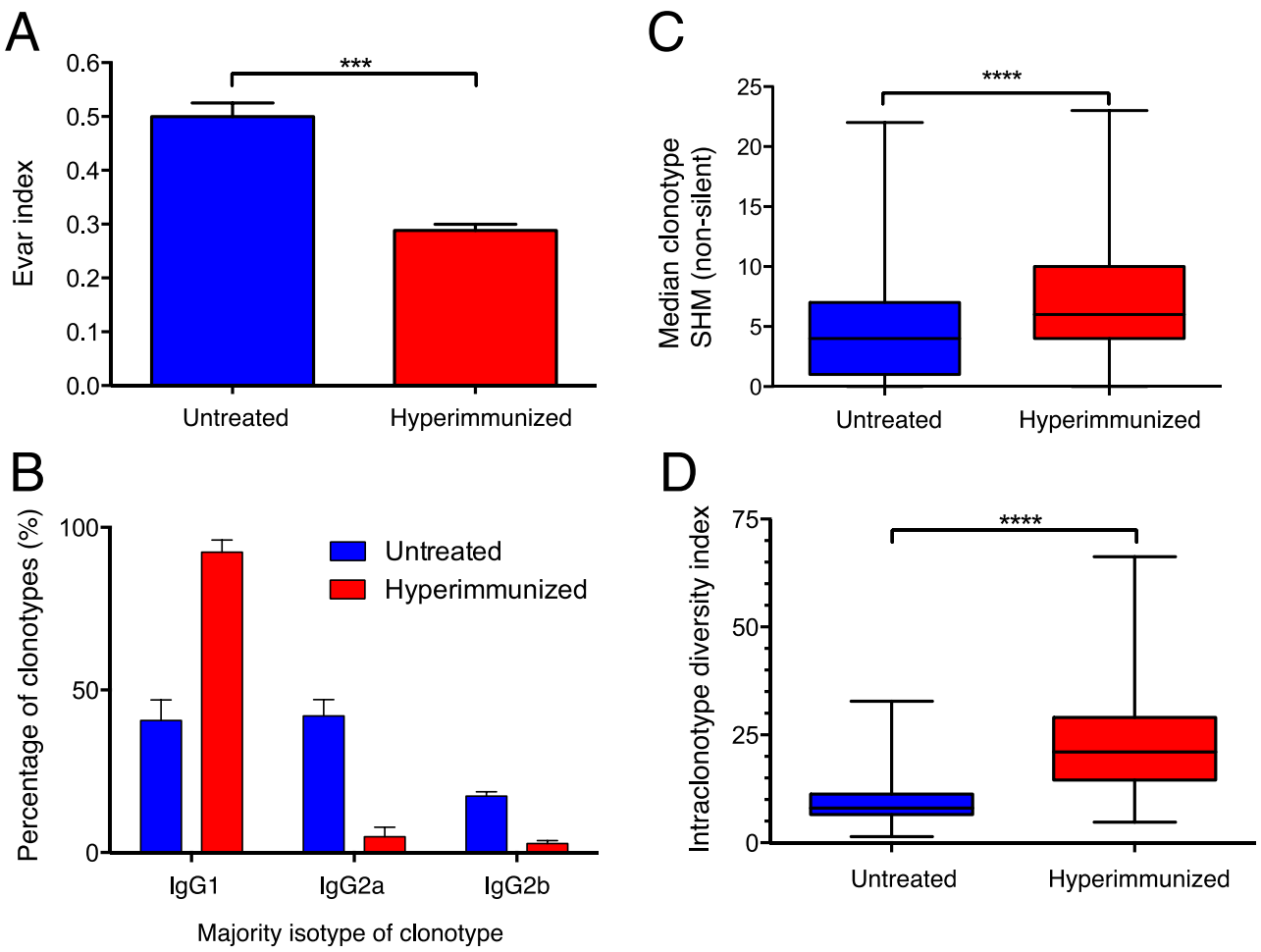


Fig. S22. Various immune profiling metrics from MAF-corrected Ig-seq data. Ig-seq data are from library sample preparations of splenic cDNA with synthetic spike-ins from hyperimmunized mice (n = 3) and untreated mice (n = 3) (datasets used for hyperimmunized are IM_1a, IM_2, IM_3; datasets used for untreated mice are UM_1, UM_2, UM_3, see table S7). Only clonotypes (>80% CDR3 a.a. similarity with matching CDR3 lengths, V and J genes) with the majority of reads being IgG 1, IgG2a, and IgG2b were used for analysis. The Evar index (**A**) was used (top 500 clonotype frequencies of each data set) to determine the extent of polarization. Values closer to 1 represent more evenly distributed clonotype frequencies within a data set, while values closer to 0 represent more skewed clonotype frequencies. The majority isotype of each clonotype (top 500 clonotype frequencies of each data set) was assigned to the clonotype frequency to determine the relative amounts of each isotype in the data sets (**B**). The median non-silent somatic hypermuations (SHM) for each of the top 100 clonotypes (based on frequency) for each data set were compared based on immune status (**C**). The intraclonotype diversity index for each of the top 100 clonotypes (based on frequency) for each data set were compared based on immune status (**D**). P values less than 0.001 and less than 0.0001 are represented by (***) and (****), respectively.

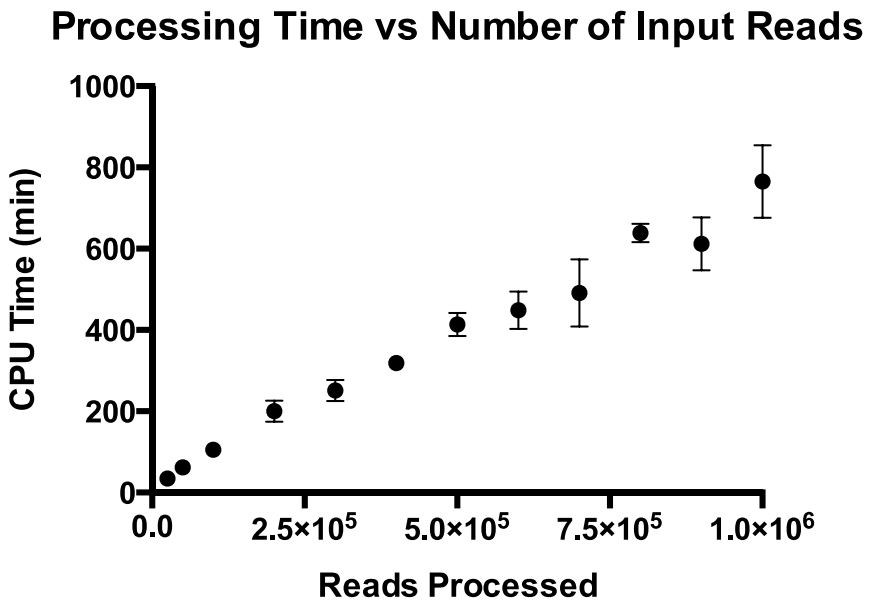


Fig. S23. Processing time of reads for MAF error and bias correction pipeline. The steps 2-6 of our MAF error and bias correction pipeline (see Materials and Methods) were integrated into a single workflow, processing time of this workflow scaled directly with the number of sequencing reads. Typically cluster computing enabled parallel processing of several data sets (e.g. 16-18 Ig-seq data sets of 1 x 10^6) at one time.

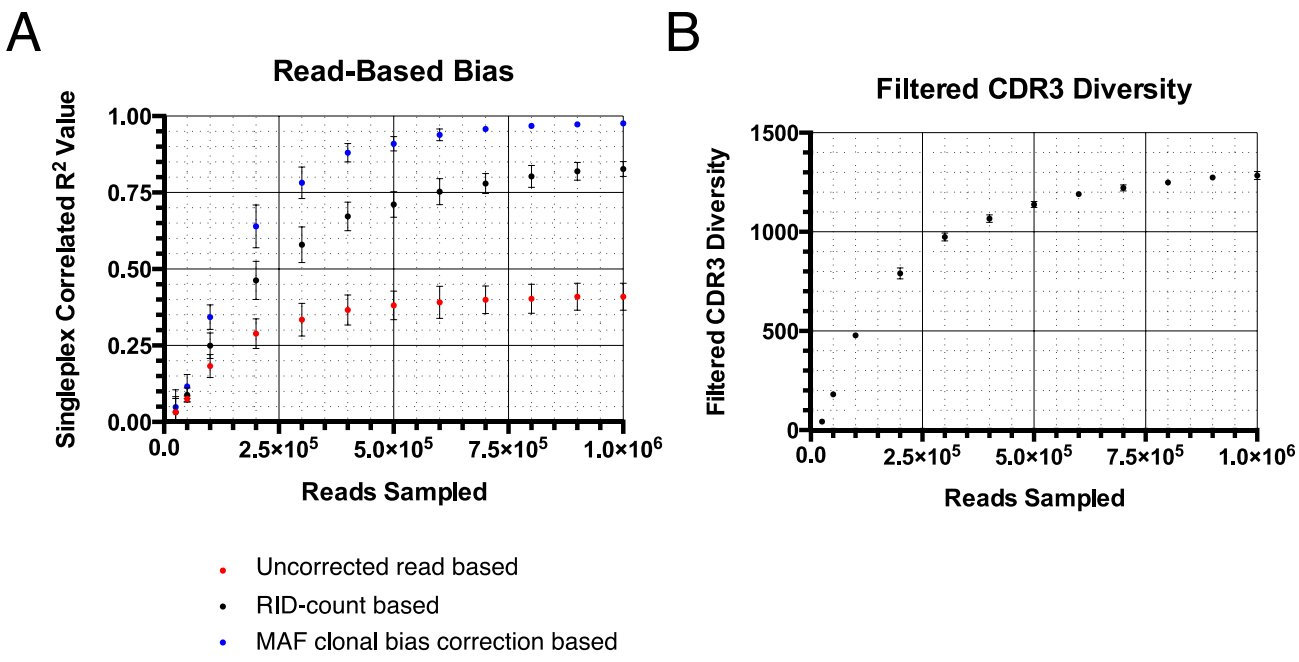


Fig. S24. Effect of the number of reads analyzed using final MAF sample preparation conditions. *In silico* random subsampling of CLC Genomics Workbench pre-processing and quality-filtering of merged reads was used to evaluate the effects on (**A**) clonal frequency bias and (**B**) filtered clonal diversity. MAF clonal bias corrected data achieved high accuracy (R^2 correlation of spike-in clonal frequencies to singleplex-PCR) with less reads compared to uncorrected or RID-count based data.

Dataset	Paired-End Reads	Post-CLC Merged Reads	Processed Reads	Productive Reads
J1	918000	373000	370000	330000
J2	1150000	477000	370000	365000
J3	1060000	423000	370000	364000
J4	1240000	510000	370000	365000
J5	1170000	458000	370000	354000

Table S1. Ig-seq read count statistics for spike-ins following replicate library preparation by singleplex PCR (see fig. S2, B and C). Information on bioinformatic preprocessing and annotation (productive reads) can be found in Materials and Methods.

Dataset	Paired-End Reads	Post-CLC Merged Reads	Processed Reads	Productive Reads
Reddy-PS-1	1560000	585000	400000	357000
Reddy-PS-2	2180000	827000	400000	356000
Reddy-PS-3	1090000	416000	400000	346000
Reddy-PS-Compare	1980000	749000	650000	522000
TAK-PS-Compare	1730000	672000	650000	583000

Table S2. Ig-seq read count statistics following MAF library preparation by multiplex PCR (See Fig. 2A). Datasets Reddy-PS-1, -2, -3 were from replicate MAF library preparation from the same starting material (immunized mouse splenic cDNA with ~10% spike-ins) using multiplex PCR step with primer set Reddy-2010 (described previously in Ref. 16). Datasets Reddy-PS-Compare and TAK-PS-Compare were from library preparation starting from the same starting material (immunized mouse splenic cDNA with ~10% spike-ins), but using primer set Reddy-2010 or newly designed reduced primer set TAK for the multiplex PCR step. A complete list of primer sequences can be found in table S5. Information on bioinformatic preprocessing and annotation (productive reads) can be found in Materials and Methods.

Annotation type	Comparison with IMGT HighV-Quest
V Gene Matches	0.998
J Gene Matches	0.999
CDR3 Matches	0.998
SHM Matches	0.99
Nonsilent SHM Matches	0.977

Table S3. A comparison of the VDJ annotation tool used in this study (modified from Laserson *et al*. (*12*)) with IMGT HighV-Quest. The same NGS data was used for both annotation tools, corresponding to 500,000 sequences (randomly selected) from dataset IM_2, see table S7. Additional information on bioinformatic preprocessing and VDJ annotation can be found in Materials and Methods.

Dataset	Paired-End Reads	Post-CLC Merged Reads	Processed Reads
S1	2400000	1060000	3 levels
S2	394000	1750000	3 levels
S3	2500000	1110000	3 levels
S4	3210000	1440000	3 levels
S5	2850000	1270000	3 levels
S6	2940000	1320000	3 levels
S7	3810000	1710000	3 levels
S8	3270000	1460000	3 levels
S9	2540000	1140000	3 levels
S10	3210000	1430000	3 levels
S11	2530000	1110000	3 levels
S12	2390000	1070000	3 levels
S13	2870000	1270000	3 levels
S14	2200000	992000	3 levels
S15	2430000	1090000	3 levels
S16	787000	336000	1 levels

Table. S4. Ig-seq read count statistics for DoE for library preparation optimization. Information on bioinformatic preprocessing and annotation (productive reads) can be found in Materials and Methods. For more information on DoE see Supplementary Materials and Methods.

Table S5. A complete list of primers and sequences used in this study.

MAF Primers

Primer Name	Sequence (5' - 3')	Function
TAK_402	TTGGCACCCGAGAATTCCACTGHHHHHACAHHHHHACAHHHHNATTCCCTTGACCAGGCA	Mouse IgG1,2a,2b,2c RT Primer (mixed at 95%)
TAK_403	TTGGCACCCGAGAATTCCACTGHHHHHACAHHHHHACAHHHHNATTCCTTTGACAAGGCATCC	Mouse IgG3 RT Primer (mixed at 5%)
TAK_472	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAGCGTCCAGTCACTCAGTTACT*C	Singleplex Spike-In 1st-Step Forward Primer
TAK_423	ACTGGAGTTCCTTGGCACCCGAGAATTCCACT*G	1st-Step PCR Reverse Primer
TAK_424	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGAT*C	2nd-Step PCR Forward primer
TAK_531_IDX_1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 1)
TAK_532_IDX_2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 2)
TAK_533_IDX_3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 3)
TAK_534_IDX_4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 4)
TAK_535_IDX_5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 5)
TAK_536_IDX_6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 6)
TAK_537_IDX_7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 7)
TAK_538_IDX_8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 8)
TAK_539_IDX_9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 9)
TAK_540_IDX_10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 10)
TAK_541_IDX_11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 11)
TAK_542_IDX_12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 12)
TAK_543_IDX_13	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 13)
TAK_544_IDX_14	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 14)
TAK_545_IDX_15	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 15)
TAK_546_IDX_16	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 16)
TAK_547_IDX_17	CAAGCAGAAGACGGCATACGAGATCTCTACGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 17)
TAK_548_IDX_18	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 18)
TAK_549_IDX_19	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 19)
TAK_550_IDX_20	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 20)
TAK_551_IDX_21	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 21)
TAK_552_IDX_22	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 22)
TAK_553_IDX_23	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 23)
TAK_554_IDX_24	CAAGCAGAAGACGGCATACGAGATGCTACCGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 24)
TAK_425_IDX_25	CAAGCAGAAGACGGCATACGAGATatcagtGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 25)
TAK_426_IDX_26	CAAGCAGAAGACGGCATACGAGATgctcatGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 26)
TAK_427_IDX_27	CAAGCAGAAGACGGCATACGAGATaggaatGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 27)
TAK_428_IDX_28	CAAGCAGAAGACGGCATACGAGATcttttgGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 28)
TAK_429_IDX_29	CAAGCAGAAGACGGCATACGAGATTAGTTGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 29)
TAK_430_IDX_30	CAAGCAGAAGACGGCATACGAGATCCGGTGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 30)
TAK_431_IDX_31	CAAGCAGAAGACGGCATACGAGATATCGTGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 31)
TAK_432_IDX_32	CAAGCAGAAGACGGCATACGAGATTGAGTGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 32)
TAK_433_IDX_33	CAAGCAGAAGACGGCATACGAGATCGCCTGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 33)
TAK_434_IDX_34	CAAGCAGAAGACGGCATACGAGATGCCATGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 34)
TAK_435_IDX_35	CAAGCAGAAGACGGCATACGAGATAAAATGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 35)
TAK_436_IDX_36	CAAGCAGAAGACGGCATACGAGATTGTTGGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 36)
TAK_437_IDX_37	CAAGCAGAAGACGGCATACGAGATATTCCGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 37)
TAK_438_IDX_38	CAAGCAGAAGACGGCATACGAGATAGCTAGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 38)
TAK_439_IDX_39	CAAGCAGAAGACGGCATACGAGATGTATAGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 39)
TAK_440_IDX_40	CAAGCAGAAGACGGCATACGAGATTCTGAGGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 40)
TAK_441_IDX_41	CAAGCAGAAGACGGCATACGAGATGTCGTCGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 41)
TAK_442_IDX_42	CAAGCAGAAGACGGCATACGAGATCGATTAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 42)
TAK_443_IDX_43	CAAGCAGAAGACGGCATACGAGATGCTGTAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 43)
TAK_444_IDX_44	CAAGCAGAAGACGGCATACGAGATATTATAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 44)
TAK_445_IDX_45	CAAGCAGAAGACGGCATACGAGATGAATGAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 45)
TAK_446_IDX_46	CAAGCAGAAGACGGCATACGAGATTCGGGAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 46)
TAK_447_IDX_47	CAAGCAGAAGACGGCATACGAGATCTTCGAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 47)
TAK_448_IDX_48	CAAGCAGAAGACGGCATACGAGATTGCCGAGTGACTGGAGTTCCTTGGCACCC*G	2nd-Step Reverse Primer (Illumina Index 48)

Table S5. A complete list of primers and sequences used in this study. (cont.)

Reddy-2010 multiplex-PCR primer set for mouse V_H
MAF-modified regions

Primer Name	Sequence (5' - 3')	Function	Ratio in primer set
TAK_404	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAKGTRMAGCTTCAGGAGT*C	FID-Multiplex Forward Primer	0.076
TAK_405	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAGGTBCAGCTBCAGCAGT*C	FID-Multiplex Forward Primer	0.076
TAK_406	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG CAGGTGCAGCTGAAGSAST*C	FID-Multiplex Forward Primer	0.057
TAK_407	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAGGTCCARCTGCAACART*C	FID-Multiplex Forward Primer	0.076
TAK_408	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG CAGGTYCAGCTBCAGCART*C	FID-Multiplex Forward Primer	0.132
TAK_409	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG CAGGTYCARCTGCAGCAGT*C	FID-Multiplex Forward Primer	0.038
TAK_410	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG CAGGTCCACGTGAAGCAGT*C	FID-Multiplex Forward Primer	0.019
TAK_411	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAGGTGAASSTGGTGGAAT*C	FID-Multiplex Forward Primer	0.038
TAK_412	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAVGTGAWGYTGGTGGAGT*C	FID-Multiplex Forward Primer	0.095
TAK_413	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAGGTGCAGSKGGTGGAGT*C	FID-Multiplex Forward Primer	0.038
TAK_414	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAKGTGCAMCTGGTGGAGT*C	FID-Multiplex Forward Primer	0.038
TAK_415	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAGGTGAAGCTGATGGART*C	FID-Multiplex Forward Primer	0.038
TAK_416	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAGGTGCARCTTGTTGAGT*C	FID-Multiplex Forward Primer	0.019
TAK_417	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GARGTRAAGCTTCTCGAGT*C	FID-Multiplex Forward Primer	0.038
TAK_418	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAAGTGAARSTTGAGGAGT*C	FID-Multiplex Forward Primer	0.038
TAK_419	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG CAGGTTACTCTRAAAGWGTST*G	FID-Multiplex Forward Primer	0.095
TAK_420	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG CAGGTCCAACTVCAGCARC*C	FID-Multiplex Forward Primer	0.066
TAK_421	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GATGTGAACTTGGAAGTGT*C	FID-Multiplex Forward Primer	0.013
TAK_422	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACHHHNGCAG GAGGTGAAGGTCATCGAGT*C	FID-Multiplex Forward Primer	0.013

Table S5. A complete list of primers and sequences used in this study. (cont.)

New reduced multiplex-PCR primer set (TAK) for mouse V_H
MAF-modified regions

Primer Name	Sequence (5' - 3')	Function	Ratio in primer set
TAK_562	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG GAGGTGAAGCTTCTCGAGT*C	FID-Multiplex Forward Primer	0.067
TAK_564	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG GAGGTGCAGCTTGTTGAGT*C	FID-Multiplex Forward Primer	0.067
TAK_568	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGATCCAGTTGGTGCAGT*C	FID-Multiplex Forward Primer	0.067
TAK_575	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG GAAGTGCAGCTGTTGGAGA*C	FID-Multiplex Forward Primer	0.067
TAK_582	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGGT/ideoxyI/CAGCTGCAGCAGY*C	FID-Multiplex Forward Primer	0.067
TAK_583	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGGTTM/ideoxyI/GCTGCAACAGT*C	FID-Multiplex Forward Primer	0.067
TAK_584	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGGTYCA/ideoxyI/CT/ideoxyI/CAGCAGT*C	FID-Multiplex Forward Primer	0.067
TAK_585	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGGTGCAGCTGAAGSAGT*C	FID-Multiplex Forward Primer	0.067
TAK_586	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG GAGGTGCAGCTTCAGGAGT*C	FID-Multiplex Forward Primer	0.067
TAK_587	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG GAAGTGAA/ideoxyI/CTTGAGGWGT*C	FID-Multiplex Forward Primer	0.067
TAK_588	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGGTTACTCTGAAAGAG*T	FID-Multiplex Forward Primer	0.067
TAK_589	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGAT/ideoxyI/CAGCTT/ideoxyI/AGGAGT*C	FID-Multiplex Forward Primer	0.067
TAK_590	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG GAGGTG/ideoxyI/AGCTGGTGGAGT*C	FID-Multiplex Forward Primer	0.067
TAK_591	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG GAGGTGCAGCTTGTAGAGA*C	FID-Multiplex Forward Primer	0.067
TAK_612	CGTTCAGAGTTCTACAGTCCGACGATCHHHHACHHHHACH HHNGCAG CAGGT/ideoxyI/CAGCTGCAGCAGc*C	FID-Multiplex Forward Primer	0.067

Table S5. A complete list of primers and sequences used in this study. (cont.)

ddPCR Primers

Primer Name	Sequence (5' - 3')	Function
TAK_530	GTC TC/ideoxyI/ /ideoxyI/CA GCC AAA AC	Forward J Gene Amplifying
TAK_522	TGGCACCCGAGAATTC	Reverse Illumina Adapter Amplyfing
TAK_498	/56-FAM/C+C+A GT+G GAT +A+GA +C/3IABkFQ/	Biological Isotype Probe
TAK_499	/5HEX/CG+T C+T+G ACT +AGA +ACT +C/3IABkFQ/	Spike-In Isotype Probe

			Somatic Hypermutaions (nt)			Unique CDR3s (a.a.)		Unique Intraclonal Variants (nt)			
Synthetic spike-in standard clones (CDR3 a.a. sequence)	**V Gene**	**Relative Spike-In Frequency (%)**	**Designed SHM**	**Read-Based Median**	**Median of RID Medians**	**Uncorrected**	**Full MAF Pipeline**	**Uncorrected**	**Full MAF Pipeline**	**Average correction (%)**	**Intraclonal diversity index (RID normalized)**
CARIKKIVATYFDYW	IGHV1S81	23.38	11	11 ± 0	11 ± 0	375 ± 9	1 ± 0	2407 ± 182	67 ± 5	97.24	7 ± 1
CWILLW	IGHV6-6	17.76	4	4 ± 0	4 ± 0	114 ± 20	1 ± 0	3364 ± 333	78 ± 2	97.7	6 ± 0
CARMARKW	IGHV14-3	16.55	6	6 ± 0	6 ± 0	263 ± 12	1 ± 0	5619 ± 250	100 ± 6	98.23	6 ± 0
CARLEDIW	IGHV2-9	16.17	0	0 ± 0	0 ± 0	188 ± 38	1 ± 0	3033 ± 473	75 ± 5	97.54	7 ± 0
CARTARIKYW	IGHV3-2	8.59	8	8 ± 0	8 ± 0	323 ± 18	1 ± 0	5674 ± 393	57 ± 6	99.01	6 ± 1
CARINAW	IGHV14-3	4.53	0	0 ± 0	0 ± 0	98 ± 10	1 ± 0	1791 ± 142	31 ± 6	98.3	7 ± 1
CARSAIW	IGHV5-6-3	4.42	13	13 ± 0	13 ± 0	90 ± 17	1 ± 0	1898 ± 143	30 ± 6	98.49	7 ± 1
CARSKYLARYW	IGHV2-9	2.39	13	13 ± 0	13 ± 0	83 ± 21	1 ± 0	588 ± 56	15 ± 4	97.57	10 ± 2
CARMARTINW	IGHV5-6-3	1.4	0	0 ± 0	0 ± 0	76 ± 12	1 ± 0	638 ± 112	10 ± 2	98.64	8 ± 0
CARTHERW	IGHV1-4	1.34	0	0 ± 0	0 ± 0	69 ± 9	1 ± 0	694 ± 64	7 ± 2	99.08	6 ± 2
CARLLINFDYW	IGHV1S81	1.28	5	5 ± 0	5 ± 0	43 ± 8	1 ± 0	178 ± 23	3 ± 1	98.73	6 ± 1
CTHERESAYW	IGHV6-6	0.69	0	0 ± 0	0 ± 0	33 ± 12	1 ± 0	233 ± 94	5 ± 3	98.34	9 ± 4
CARSIMANW	IGHV3-2	0.53	8	8 ± 0	8 ± 0	65 ± 6	1 ± 0	552 ± 63	6 ± 2	99.15	10 ± 2
CARVITRDYW	IGHV1S81	0.36	0	0 ± 0	0 ± 0	11 ± 1	1 ± 0	59 ± 6	2 ± 1	97.63	15 ± 8
CARKSTRASYW	IGHV1-4	0.34	3	3 ± 0	3 ± 0	42 ± 7	1 ± 0	210 ± 30	3 ± 2	99.19	8 ± 4
CRISTINAW	IGHV14-3	0.28	9	9 ± 0	9 ± 0	27 ± 5	1 ± 0	143 ± 18	2 ± 0	99.29	8 ± 1
										Average correction across all spike-ins (%)	
										98.38	
Pearson Coefficient (compared to frequency)						**0.819 (****)**	**0 (ns)**	**0.726 (**)**	**0.925 (****)**		**-0.449 (ns)**

Table S6. Error correction statistics for spike in clones. The intraclonal diversity index was determined for each clone by dividing the number of intraclonal variants by the clonal count (based on RID-count). Ig-seq data are from replicate MAF library sample preparations (n = 3) from mouse splenic cDNA with ~10% synthetic spike-ins (mean values are from replicate datasets IM_1a, _1b, 1c, see table S7). Relative spike-in frequencies are mean values obtained from replicate libraries (n = 5) generated by singleplex PCR (see fig. S2 and table S1).

Column #	1	2	3	4	5	6	7	8	9	10
	Dataset	Initial Paired-End Reads	Merged Reads After CLC	Percent Passing CLC	Processed Reads	Reads with FID and RID	% of Reads w/ FID and RID Found	Initial Unique FID/RID Combinations	Initial Unique RIDs	Corrected Unique FID/RID Combinations
Immunized	IM_1a	2.69E+06	1.04E+06	77.3	1.0E+06	9.49E+05	94.9	4.85E+05	1.01E+05	4.54E+05
	IM1_b	3.01E+06	1.17E+06	77.8	1.0E+06	1.00E+06	100.0	4.25E+05	1.01E+05	3.91E+05
	IM_1c	2.75E+06	1.01E+06	73.6	1.0E+06	9.16E+05	91.6	4.70E+05	1.00E+05	4.39E+05
	IM_2	2.77E+06	1.04E+06	74.8	1.0E+06	9.25E+05	92.5	4.23E+05	8.51E+04	3.85E+05
	IM_3	2.70E+06	1.01E+06	75.2	1.0E+06	9.01E+05	90.1	4.47E+05	9.86E+04	4.08E+05
Untreated	UM_1	7.18E+06	1.22E+06	34.0	1.0E+06	1.00E+06	100.0	6.00E+05	1.51E+05	5.60E+05
	UM_2	1.73E+07	1.27E+06	14.6	1.0E+06	1.00E+06	100.0	5.93E+05	1.27E+05	5.52E+05
	UM_3	7.01E+06	1.10E+06	31.4	1.0E+06	9.83E+05	98.3	5.46E+05	1.36E+05	5.08E+05

Column #	1	11	12	13	14	15	16	17	18	19	20
	Dataset	Corrected Unique RIDs	% of RFIDs Corrected	% of RIDs Corrected	Consensus Building Groups	% of "double tagged" sequences	% of Sequences in a Group size of at Least 3	% of Sequences Corrected	Raw Productive Reads	Raw % of reads	CB Productive Reads
Immunized	IM_1a	8.51E+04	6.3	15.4	9.15E+04	1.3	95.5	42.1	7.44E+05	78.49	8.73E+05
	IM1_b	8.16E+04	7.9	18.9	8.68E+04	1.2	96.5	43.2	7.88E+05	78.75	9.26E+05
	IM_1c	8.45E+04	6.6	15.8	9.08E+04	1.3	95.5	44.8	7.13E+05	77.77	8.43E+05
	IM_2	6.11E+04	9.1	28.3	6.52E+04	0.9	96.8	41.8	7.40E+05	79.91	8.81E+05
	IM_3	7.49E+04	8.6	24.1	8.00E+04	1.1	95.6	41.1	7.24E+05	80.38	8.39E+05
Untreated	UM_1	1.30E+05	6.5	14.2	1.39E+05	1.6	91.3	38.7	7.76E+05	77.64	9.11E+05
	UM_2	1.05E+05	6.9	17.3	1.19E+05	2.1	92.2	38.6	7.81E+05	78.11	9.10E+05
	UM_3	1.16E+05	6.8	15.1	1.23E+05	1.4	92.6	39.5	7.73E+05	78.60	9.07E+05

Column #	1	21	22	23	24	25	26	27	28	29	30
	Dataset	CB % of reads	Raw Unique Productive CDR3s	CB Unique Productive CDR3s	CB Productive CDR3s deemed hotspot error	% of CB Productive CDR3s deemed hotspot error	Additional F3+<3 Clones Removed	% F3<3 Clones removed post hotspot	Final Good Unique Filtered CDR3s	% final clones of initial CB clones	% of VDJ output reads in final filtered CDR3s
Immunized	IM_1a	92.03	4.44E+04	3.42E+03	2.06E+02	6.03	1.93E+03	60.00	1.28E+03	37.59	79.1
	IM1_b	92.64	4.67E+04	3.50E+03	2.14E+02	6.11	1.98E+03	60.27	1.31E+03	37.30	81.7
	IM_1c	91.97	4.49E+04	3.35E+03	2.17E+02	6.48	1.87E+03	59.62	1.26E+03	37.77	79.6
	IM_2	95.21	4.44E+04	3.03E+03	1.82E+02	6.01	1.85E+03	64.84	1.00E+03	33.05	79.3
	IM_3	93.20	4.37E+04	3.23E+03	2.02E+02	6.26	1.85E+03	61.04	1.18E+03	36.52	76.2
Untreated	UM_1	91.07	5.46E+04	7.44E+03	3.10E+02	4.17	5.18E+03	72.61	1.95E+03	26.25	84.1
	UM_2	91.04	5.88E+04	7.57E+03	3.05E+02	4.03	5.47E+03	75.29	1.79E+03	23.71	76.7
	UM_3	92.24	5.72E+04	8.05E+03	3.59E+02	4.46	5.70E+03	74.06	2.00E+03	24.78	81.3

Table S7. Expanded Ig-seq processing statistics. Datasets IM_1a, _1b, _1c correspond to replicate library sample preparations (n = 3) of a hyperimmunized mouse. Datasets IM_2, IM_3 correspond to 2 different hyperimmunized mice and UM_1, UM_2, UM_3 correspond to 3 different untreated mice. All datasets were from MAF library preparation of splenic cDNA with ~10% synthetic spike-ins.

Gene	Sequence	Function
TAK_1	CGTCCAGTCACTCAGTTACTCGAGGTTCAGCTGCAGCAGTCTGGGGCAGAGCTTGTGAAGCCAGGGGCCTCAGTCAAGTTGTCCTGCACAGCTTCTGGCTTCAACATAAAGACACCTATATGCACTGGGTGAAGCAGAGGCCTGAACAGGGCCTGGAGTGGATTGGAAGGATTGATCCTGCGAATGGTAATACTAAATATGACCCGAAGTTCCAGGGCAAGGCCACTATAACAGCAGACACATCCTCCAACACAGCCTACCTGCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTAGAATAAATGCGTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGATCCAATTGGATATC	Spike-In CARINAW
TAK_2	CGTCCAGTCACTCAGTTACTCGAGGTTCAGCTGCAGCAGTCTGGGGCAGAGCTTGTGAAGTCAGGGGCCACAGTCAAGTTGTCCTGCACAGCTTCTGGCCTCAACATAAAGACACCTATATGCACTGGCTGAAGCAGGGCCTGGAGTGGATTGGAAGGATTGATCCTGCGAATGGTAATACTAAATATGACCCGAAGTTCCAGGGCAAGGCCACTATAACAGCAGACACATCCTCCAACACAGCCTACCTGCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTAGAATGGCCAGAAAGTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGCAATTGGATATC	Spike-In CARMARKW
TAK_3	CGTCCAGTCACTCAGTTACTCGAGGTTCAGCTGCAGCAGTCTGGGGCAGAGCTTGTGAAGCCAGCGGCCTCACTCAAGTTGTCCTGCACAGCTTCTGGCTACAACATAAAGACATCTATATCCACTGGGTGAAGCAGAGGCCTGAAAAGGGCCTGGAGAGGATTCGAAGGACTGATCCTGCGAATGGTAATACTAAATATGACCCGAAGTTCCAGGGCAAGGCCACTATAACAGCAGACACATCCTCCAGCAGCCTGCAGCTCAGCAGCCTGACATCTGAGGACACTGCCGTCTATTACTGTGCTAGAATGCCAGAAGTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGCAATTGGATATC	Spike-In CRISTINAW
TAK_4	CGTCCAGTCACTCAGTTACTCGAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCTGAAACTCTCCTGTGCAGCCTCTGGATTCACTTTCAGTAGCTATGGCATGTCTTGGGTTCGCCAGACTCCAGACAAGAGGCTGGAGTTGGTCGCAACCATTAATAGTAATGGTGGTAGCACCTATTATCCAGACAGTGTGAAGGGCCGATTCACCATCTCCAGAGACAATGCCAAGAACACCCTGTACCTGCAAATGAGCAGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGAATGGCGAGAACAATAAACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGATCCAATTGGATATC	Spike-In CARMARTINW
TAK_5	CGTCCAGTCACTCAGTTACTCGAGGTGCAGCTGGTGGAGTCTGGGGGAGGCATAGTGCAGCCTGGAGGGTCCCTGAAACGCTCCTGTGCAGCCTCTAGATTCACTATCAGTAGCTATGGCATGTCTTCGGTTCGCCAGACTCCAGACAAGAGGCTGGAGCTGGTCGCAACCATTGATAGTAATGGTGGTAGCACCGATTATCCAGACAGTGTGAAGCGCCGATTCACCATCTCAGGAGACAATGCCAAGAACACCCTGTACCTGCAAATGCGCAGTCTGAAGTCTGAGGACACAGCCATGTATTACTGTGCAAGAAGCGCTATATGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGATCCAATTGGATATC	Spike-In CARSAIW
TAK_6	CGTCCAGTCACTCAGTTACTCGATGTGCAGCTTCAGGAGTCGGGACCTGGCCTGGTGAAACCTTCTCAGTCACTTTCTCTCACCTGCACTGTCACTGGCTACTCAATCACCAGTGGTTATGGCTGGAACTGGATCCGGCAGTTTCCAGGAAACAAACTGGAATGGATGGGCTACATAAGCTACAGTGGTAGCACTAACTACAACCCATCTCTCAAAAGTCGAATCTCTATCACTCGAGACACATCCAAGAACCAGTTCTTCCTGCAGTTGAATTCTGTGACTACTGAGGACACAGCCACATATTACTGTGCAAGAATCCATAATGGCAAAATACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGATCCAATTGGATATC	Spike-In CARTARIKYW
TAK_7	CGTCCAGTCACTCAGTTACTCGATGTGCAGCTTCAGGAGTCGGGACCTGGCCTGGTGAAACCTTCTCAGTCTCTGTCCCTCACCTGCACTGTCACTGGCTACTCAATCACCAGTGATTATGCCTGGAACTGGATCCGGCAGTTTCCAGGAAACAAACTGGAGTGGATGGGCTACATAAGCTACAGTGGTAGCACTAACTACAACCCATCTCTCAAAAGTCGAATCTCTATCACTCGAGACACATCCAAGAACCAGTTCTTCCTGCAGTTGAATTCTGTGACTACTGAGGACACAGCCACATATTACTGTGCAAGATCCATAATGGCAAACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGCAATTGGATATC	Spike-In CARSIMANW
TAK_8	CGTCCAGTCACTCAGTTACTCGAAGTGAAGCTTGAGGAGTCTGGAGGAGGCTTGGTGCAACCTGGAGGATCCATGAAACTCTCCTGTGTTGCCTCTGGATTCACTTTCAGTAACTACTGGATGAACTGGGTCCGCCAGTCTCCAGAGAAGGGGCTTGAGTGGGTTGCTGAAATTAGATTGAAATCTGATGATTATGTAACACATTATGCGGAGTCTGTGAAAGGGAGGTTCACCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCAAATGAACAACTTAAGAGCTGAAGACACTGGTATTTATTACTGTTGGATACTGCTATGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGCAATTGGATATC	Spike-In CWILLW
TAK_9	CGTCCAGTCACTCAGTTACTCGAAGTGAAGCTTGAGGAGTCTGGAGGAGGCTTGGTGCAACCTGGAGGATCCATGAAACTCTCCTGTGTTGCCTCTGGATTCACTTTCAGTAACTACTGGATGAACTGGGTCCGCCAGTCTCCAGAGAAGGGGCTTGAGTGGGTTGCTGAAATTAGATTGAAATCTAATAATTATGCAACACATTATGCGGAGTCTGTGAAAGGGAGGTTCACCATCTCAAGAGATGATTCCAAAAGTAGTGTCTACCTGCAAATGAACAACTTAAGAGCTGAAGACACTGGCATTTATTACTGTACCCATGAGCGTGAAAGTGCTTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGCAATTGGATATC	Spike-In CTHERESAYW
TAK_10	CGTCCAGTCACTCAGTTACTCCAGGTCCAGCTGCAGCAGTCTGGGGCTGAACTGGCAAGACCTGGTGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTTACTAGCTACACGATGCACTGGGTAAAACAGAGGCCTGGACAGGGTCTGGAATGGATTGGATACATTAATCCTAGCAGTGGTTATACTAAGTACAATCAGAAGTTCAAGGACAAGGCCACATTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAACTGAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGAACTCATGAGAGATGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGCAATTGGATATC	Spike-In CARTHERW
TAK_11	CGTCCAGTCACTCAGTTACTCCAGGTCCAGCTGCAGCAGTCTGGGGCTGAACTGGTAAAACCTGGGGCCTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTTACTAGCTACACGATGCACTGGGTAAAACAGAGGCCTGGACAGGGTCTGGAATGGATTGGATACATTAATCCTAGCAGTGGTTATACTAAGTACAATCAGAAGTTCAAGGACAAGGCCACATTGACTGCAGACAAATCCTCCAGCACAGCCTACATGCAACTGAGCAGCCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGAACACAAGAGCTTCTTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGCAATTGGATATC	Spike-In CARKSTRASYW
TAK_12	CGTCCAGTCACTCAGTTACTCCAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAAGGCTGGGGCTTCAGTGAAGATGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCTTGGACAAGGCCTTGAGTGGTTTGCAGAGACTAATCCTACCAACGGTCGTACTTACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCTACTCAGCGGCCCGACATTTGAGGACTCTGCGGTCTATTACTGTGCAAGAATTAAGAAGATAGTAGCTACCTACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACAATTGGATATC	Spike-In CARIKKIIVATYFDYW
TAK_13	CGTCCAGTCACTCAGTTACTCCAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAAGCCTGGGGCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGCACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGAGAGATTAATCCTAGCAACGGTCGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCCGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGAGTTCAAAAGAGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACAATTGGATATC	Spike-In CARVITRDYW
TAK_14	CGTCCAGTCACTCAGTTACTCCAGGTCCAACTGCAGCAGCCTGGGGCTGAACTGGTGAATCCTGGGGCTTCAGTGAACCTGTCCTGCAAGGCTTCTGGCTACACCTTGACCAGCTACTGGATGCATTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGTGAGATTAATCCTAGCAACGGTCGTACTAACTACAATGAGAAGTTCAAGAGCAAGGCCACACTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAACTCAGCAGCCCGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGACTACTAATTAACTTTGACTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACAATTGGATATC	Spike-In CARLLINFDYW
TAK_15	CGTCCAGTCACTCAGTTACTCCAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACTTGCACTGTCTCTGGGTTTTCATTAACCAGCTATGGTGTACACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGAGTGGCTGGGACTAATATGGGCTGGTGGAAGCACAAATTATAATTCGGCTCTCATGTCCAGACTGAGCATCAGCAAAGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCTGCAAACTGATGACACAGCCATGTACTACTGTGCCAGACTCGAGGATATATGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGATCCAATTGGATATC	Spike-In CARLEDIW
TAK_16	CGTCCAGTCACTCAGTTACTCCAGGTGCAGCTGAAGGAGTCAGGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACTTGCACCGTCTATGGGTATTCATTAACCAGGTATGGTGTACACTGGGTTCGCCAGCCTCCAGGTAAGGGTCTAGAGTGGCTGGGACTAATATGGGCTGGTGGAAGCACAAATTATAATTCGGCTCTCATGTCCAGACTAAGCATCAGCATAGACAACTCCAAGAGCCTAGTTTTCTTAATAATGAACAGTCTGCAAACTGATGACACAGCCTTGTACTACTGTGCCAGATCTAAGTACCTCGCTAGGTACTGGGGCCAAGGCACCACTCTCACAGTCTCCTCAGCCAAAACACTAGAAGTAGCGATAGATCTACGAGTTCTAGTCAGACGTCATACAACTGGCTCCTCGGTGACTCTAGGATGCCTGGTCAAGGGTTATGGATCTGGACATCACCATCATCATCACGGAGGATCCAATTGGATATC	Spike-In CARSLYLARYW
PCR Control	CGACACTGCTCGATCCGCTCGCACCATGTCGTTCAGAGTTCTACAGTCCGACGATCTATCACATCAACTCAGGCAGCAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTGAGGCCTGGGTCTTCAGTGAAGCTGTCCTGCAAGGCTTCTGGCTACACCTTCACCAGCTACTGGATGGGTGAAGCAGAGGCCTGGACAAGGCTTGAGTGGATTGGGTGAAGCACAGAGGCCTGGACAAGGCCTTGAATGGATTGGTAACATTTACCCTTCTGATAGTGAAACTCACTACAATCAAAAGTTCAAGGACAAGGCCACATTGACTGTAGACAAATCCTCCAGCACAGCCTACATGCAGCTCAGCAGCCTGACATCTGAGGACTCTGCGGTCTATTACTGTGCAAGGAATCTATGTGTAATGATGTAGTGACAGTGGAATTCTCGGGTGCCAAGGAACTCCAGTGTGAGGATCGACGAGAGCAGCGCGACTGG	Control Antibody Gene with Illumina Adaptors

Table S8. Synthetic genes used in this study. Synthetic antibody sequences TAK_1-16 were ordered from IDT in the pIDTBlue vector (in vitro transcription ready). PCR Control antibody gene with Illumina adaptors was ordered from IDT in the pIDTSmart (Amp) vector.