Supplementary Information

Analysis of recombinational switching at the antigenic variation locus of the Lyme spirochete using a novel PacBio sequencing pipeline

Theodore B. Verhey[1], Mildred Castellanos[1], George Chaconas[1,2,*]

[1]Department of Biochemistry and Molecular Biology, Snyder Institute for Chronic Diseases, Cumming School of Medicine, University of Calgary, Alberta, Canada

[2]Department of Microbiology, Immunology and Infectious Diseases, Snyder Institute for Chronic Diseases, Cumming School of Medicine, University of Calgary, Alberta, Canada

* Corresponding author
E-mail: chaconas@ucalgary.ca (GC). Phone 1-403-210-9692

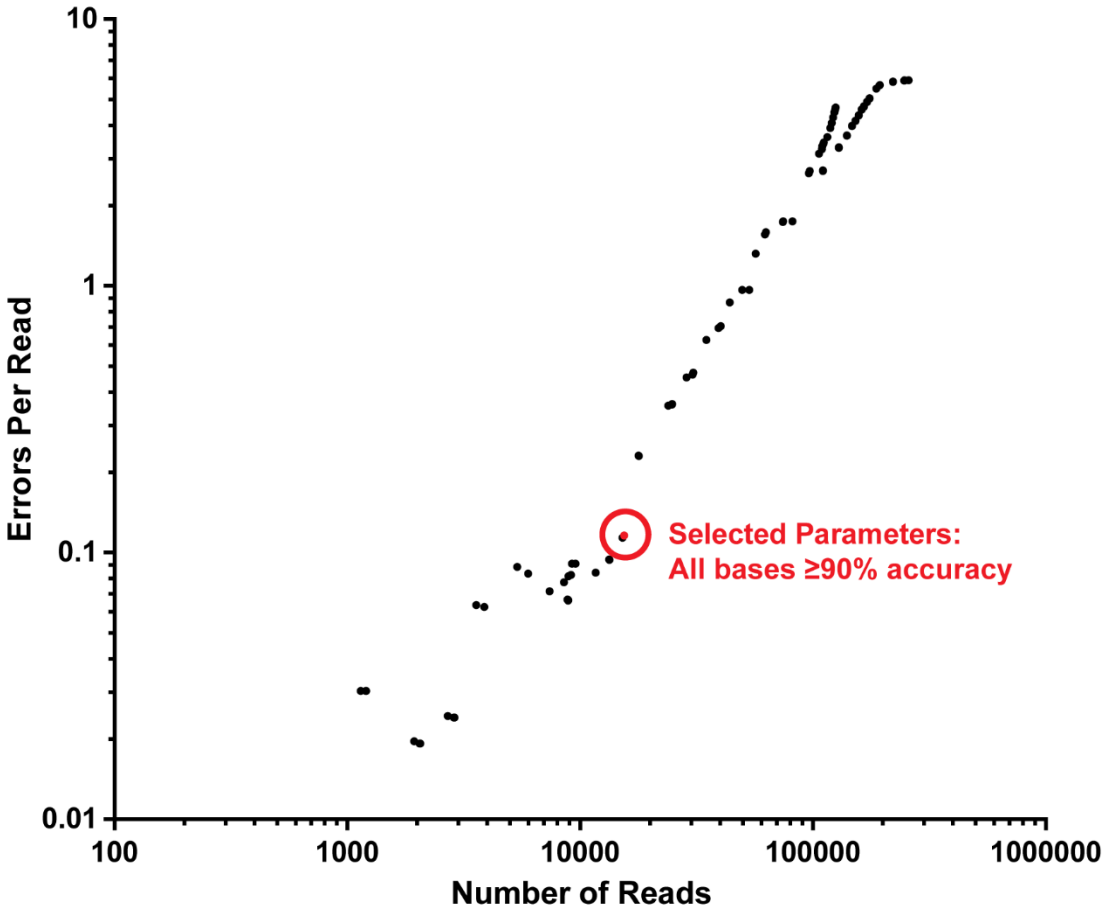

Fig S1. Optimizing sequence filters to balance dataset size and error rate. We tested 23,575 combinations of 3 different filtering parameters: number of subreads, average read quality, and minimum base quality. For each parameter set, we calculated the error rate in our week 0 controls and the resulting size of the entire dataset. By filtering our dataset for reads where every base exceeds 90% quality, we observe a 51-fold drop in errors.

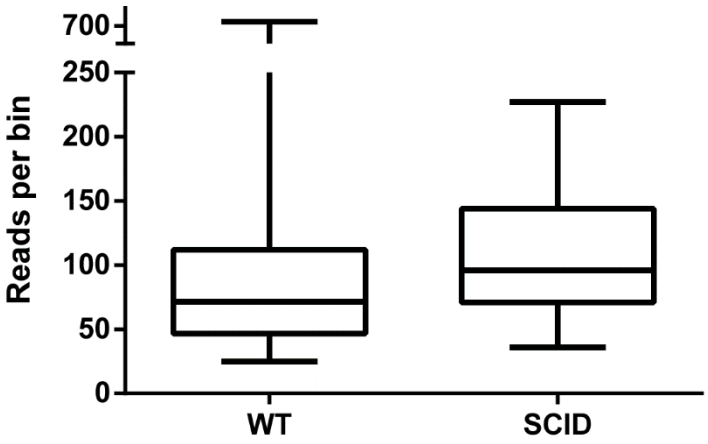

Fig S2. Number of reads per sample bin. Amplicons from wild-type mice and SCID mice were pooled and sequenced separately. The whiskers demarcate the maximum and minimum bin sizes.

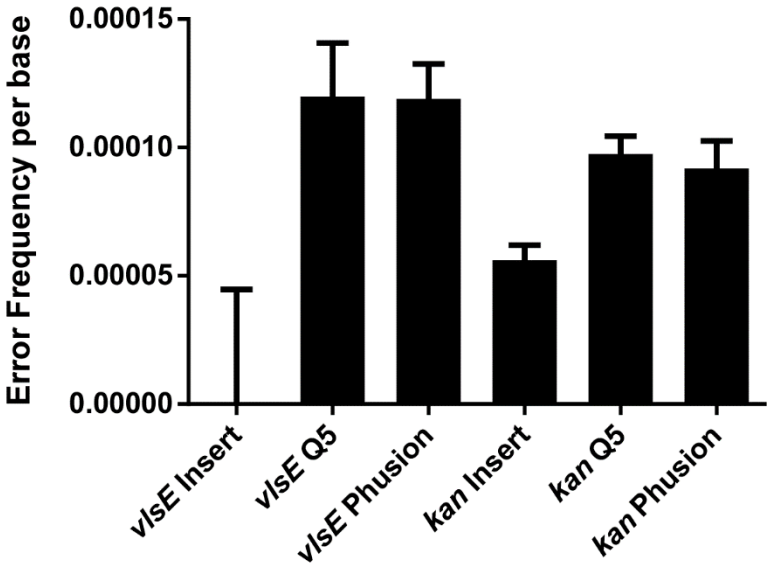

Fig S3. Error frequency in negative controls. We measured error frequency by sequencing the initial *vlsE* gene as well as the *kan* gene. For both the *kan* and *vlsE* controls, we sequenced PCR amplicons amplified using high-fidelity Q5 and Phusion polymerases, and a purified restriction digest fragment from a cloned insert as a control for PCR errors. The *vlsE* insert sequence had zero errors in the 30 vlsE sequences in that bin. For the purposes of generating an error bar, we overestimated the SEM of our *vlsE* sample by introducing a single error into one of the reads.

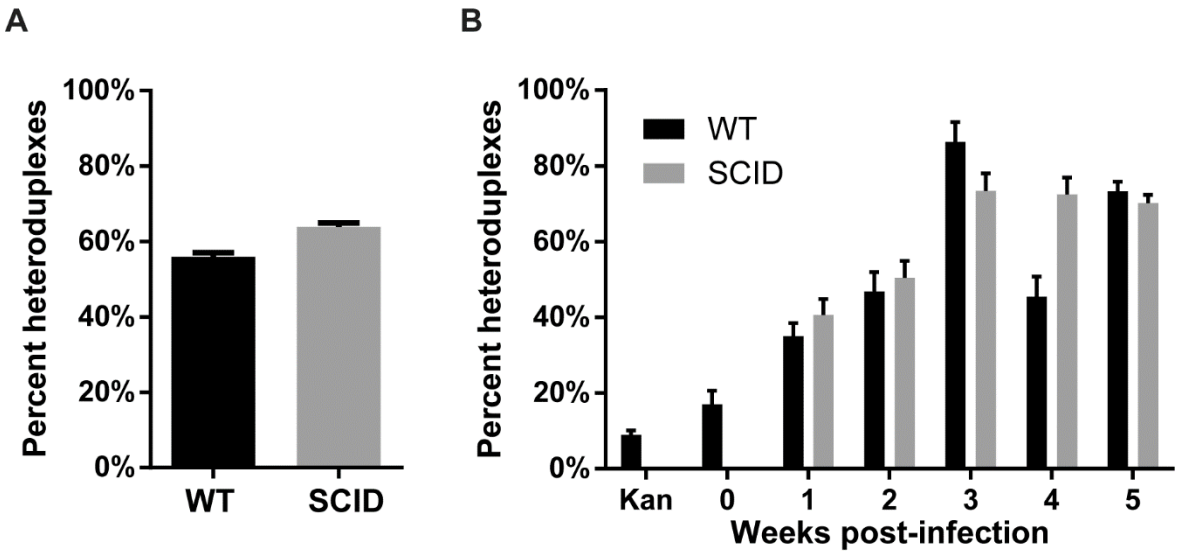

Fig S4. Heteroduplex analysis of sequence output. Using 2,112 *vlsE* molecules and 569 *kan* molecules where SSC reads existed for both strands, heteroduplexes were identified by counting those with divergent sequences. We measured the average heteroduplex frequency (± SEM) for **A)** the whole datasets, and **B)** over the time-course of the experiment.

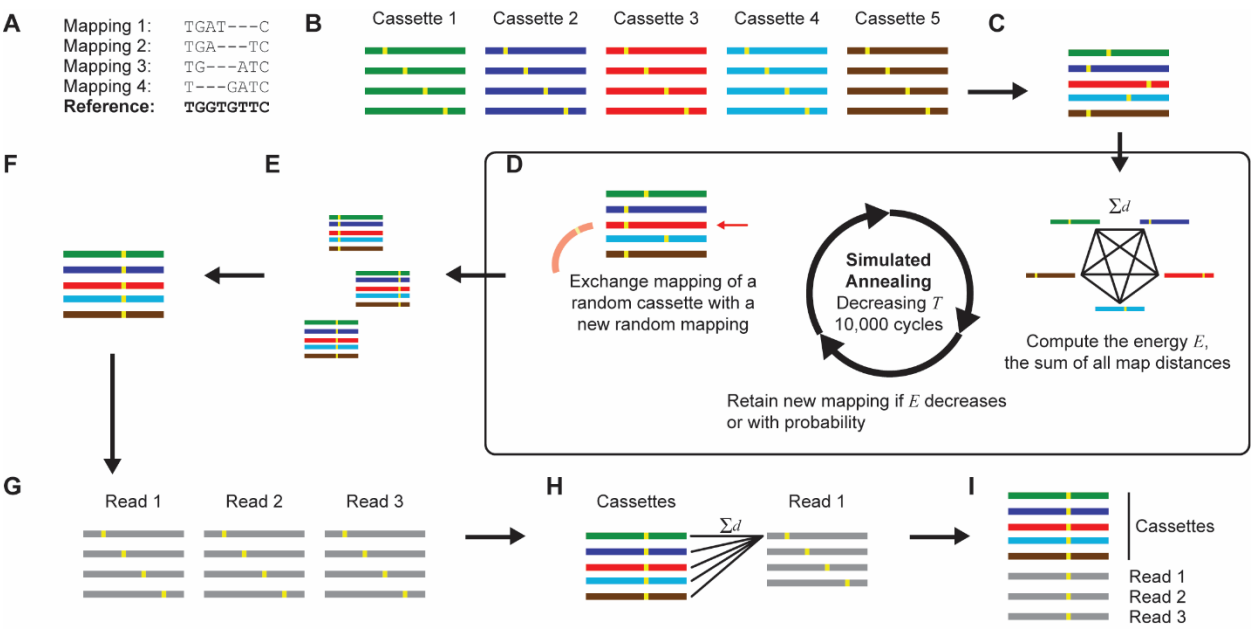

Fig S5. Schematic of alignment process for consistent mapping of ambiguous polymorphisms. A) A case study demonstrating that one sequence can have multiple equivalent alignments to a reference sequence. B) Each cassette sequence is mapped to the reference independently, and all possible equivalent alignments are retained. In this example, an ambiguous polymorphism (yellow) that is shared among all the cassettes can be mapped in 4 different positions. C) A mapping from each cassette is chosen at random to be the initial multiple alignment. D) Simulated annealing is used to iteratively improve the multiple alignment and find a globally optimal solution. Analogous to the annealing of a large population of similar DNA strands, SA minimizes the energy (sum of all pairwise mapping distances in the multiple alignment) by slowly decreasing the temperature (which affects the probability of accepting high-energy transitions) while swapping a random cassette's mapping for another random mapping for 10,000 iterations. E) Steps C and D are repeated to generate many often-equivalent solutions to the minimization problem. F) One solution is arbitrarily chosen to be the standard to which read mappings will be compared. G) Each sequenced read is mapped to the reference using the same pairwise aligner used to map the cassettes, and multiple equivalent mappings are retained. H) For each equivalent mapping, the similarity to the cassettes is measured. This similarity is the sum of distances between the mapping and each of the chosen cassette mappings. I) Only mappings that are closest to the cassettes are retained, and thus each read aligns as closely as possible to the cassettes.

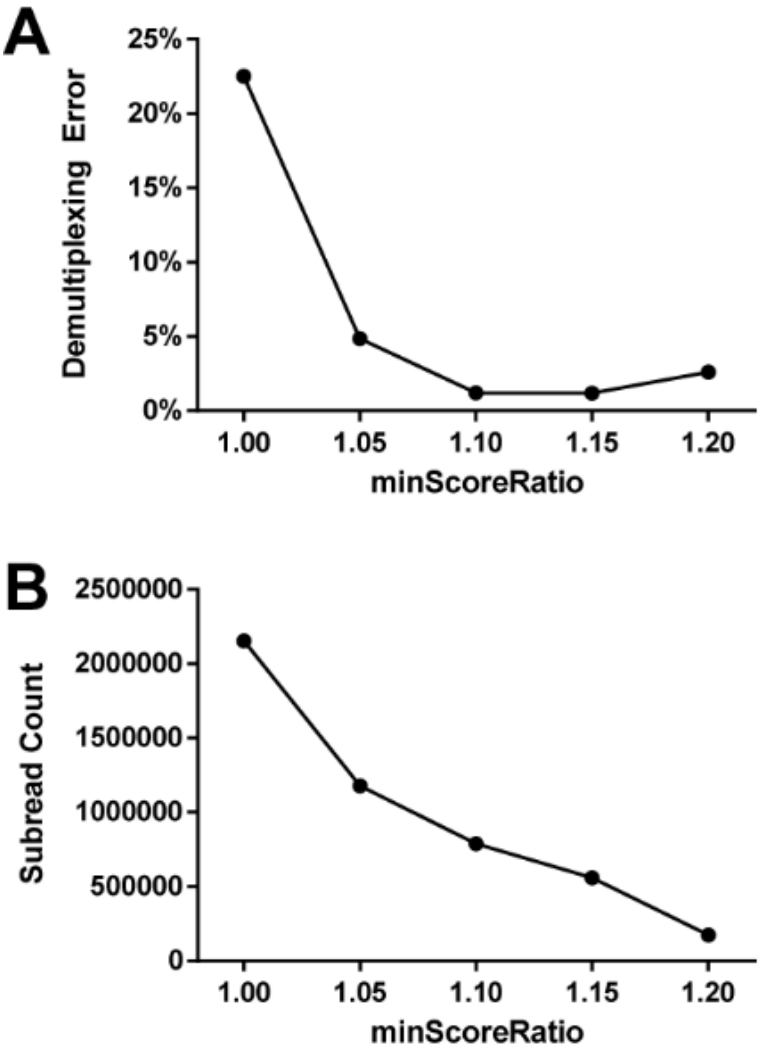

Fig S6. Effect of minScoreRatio parameters on demultiplexing error and sample size. The minScoreRatio parameter of pbssc is optimized to minimize demultiplexing error and maximize the dataset size (here shown as a total sequencing subread count). Pbssc-optimizer was used to assess the effect of different values of the minScoreRatio parameter in pbssc on **A)** the demultiplexing error rate and **B)** the size of the dataset (See **Materials and Methods** for more details).

Table S1. Inferred amino acid mutations most strongly contributing to diversifying and purifying selection by the acquired immune response.

A.A.	f_{WT}/f_{SCID}	Protein Mutations (% of all mutations at position, in WT)
G138	9.98	138delG (100%)
E137	9.79	136_137insE (100%)
K161	7.74	K161E (98%), K161T (0.8%), 161delK (0.8%)
G261	6.82	260_261insGK (85%), 260_261insGE (14%), 260_261insGKGN (0.8%), 260_261insVE (0.3%)
K157	6.60	157_158delKV (52%), 156_157insA (47%), 156_157insD (0.5%), K157E (0.5%), 156_158delAKV (0.1%)
V158	6.11	158delV (48%), V158A (48%), 157_158delKV (4.6%), 158_159delVA (0.04%)
A151	6.00	151delA (98%), 151_152delAD (1.7%), 150_151insD (0.1%), A151D (0.1%), A151N (0.1%), A151P (0.05%), A151G (0.05%)
G266	5.92	266delG (34%), 265_266insNEEN (18%), 265_266insEN (13%), 265_266insDAEN (12%), 265_266insD (11%), 265_266insNED (2.9%), 265_266insNAEN (2.4%), 265_266insNEN (2.3%), 265_266delGG (1.2%), 265_266insDEEN (1.0%), 265_266insDD (0.8%), 265_266insNAD (0.5%), 265_266insN (0.2%), 265_266insNEED (0.2%), 265_266insDA (0.2%), 265_266insG (0.1%), 265_266insDAD (0.1%), 265_266insNAED (0.07%), 265_267delGGA (0.05%), 265_266insDEN (0.05%), 265_266insE (0.05%), 265_266insNSEN (0.05%), 265_266insDE (0.03%), 265_266insDADN (0.02%), 265_266insDN (0.02%)
D160	5.89	159_160insA (100%), D160N (0.01%), 159_160insV (0.01%)
A156	5.67	156delA (65%), 155_156insK (18%), 155_156insA (16%), 154_156delDAA (0.4%), 155_157delAAK (0.06%)
E193	5.32	193_194delEN (48%), 192_193insN (48%), 192_193insGN (3.3%), 193delE (0.2%)
A153	4.96	153delA (56%), A153N (27%), 152_153insN (9.5%), 152_153insND (4.1%), 152_153delDA (1.4%), 153_154delAD (1.3%), 152_153insNA (0.1%), 152_153insA (0.1%), A153D (0.09%), 153_155delADA (0.08%), A153V (0.08%), A153G (0.04%), 152_153insD (0.04%), A153S (0.02%), 152_153insNN (0.01%) 152_154delDAD (0.008%)
G265	4.25	265delG (64%), G265A (16%), 264_265insAEN (16%), 265_266delGG (1.6%), 264_265insAD (0.2%), 264_265insE (0.2%), 264_265insADD (0.03%), 264_265insG (0.2%), 264_265insD (0.1%), 264_265delDG (0.09%), 265_267delGGA (0.07%), 264_265insN (0.07%), 264_265insA (0.04%), 264_265insADN (0.02%) 263_265delKDG (0.02%),
A155	3.93	155delA (53%), 154_155insN (17%), 154_155insA (14%), 154_155insNN (7.8%), 154_155delDA (5.3%), 154_155insD (1.2%), 154_156delDAA (0.4%), 153_155delADA (0.4%), 154_155insDA (0.2%), 154_155insDN (0.1%), 154_155insS (0.08%)
D262	3.83	262delD (60%), 261_262insNA (18%), 261_262insKGNA (14%), D261N (7.9%), 261_262insDA (2.6%), 261_262insNK (2.4%), 261_262insKGDA (2.1%), 261_262insNE (2.1%), 261_262insKGNE (2.0%), 261_262insKG (1.9%), 262_263delDK (1.9%), D262E (1.6%), 261_262insKGNG (1.5%), 261_262insEGNA (1.1%), 261_262insEG (1.1%), 261_262insNG (0.1%), 261_262insEGDA (0.09%), 261_262insK (0.06%)
A.A.	**f_{SCID}/f_{WT}**	**Protein Mutations (% of all mutations at position, in SCID)**
E124	9.41	121_123delVSE (95%), 123_124insA (3.6%), 124delE (0.7%), 124_125delEL (0.7%)
E121	7.26	E121A (24%), 120_121insG (24%), 120_121insGAG (22%), E121G (20%), 120_121insGAA (9.1%)
S123	5.55	121_123delVSE (74%), S122A (26%)
V122	3.88	121_123delVSE (53%), 121delV (44%), V121A (3.5%)
G311	3.78	G311E (93%), 311delG (5.1%), 311_313delGAA (1.4%)
S224	3.75	222_224delAVS (33%), 223_225delVSA (33%), 224_226delSAV (33%), 223_224insV (1.3%)
A225	3.35	223_225delVSA (32%), 224_226delSAV (32%), 225_227delAVS (32%), 225delA (1.3%), A225V (1.3%)
V223	3.21	222_224delAVS (48%), 223_225delVSA (48%), 223delV (1.9%), V223A (1.9%)
V226	3.21	224_226delSAV (48%), 225_227delAVS (48%), V226G (3.8%)
F269	3.17	268_269insDAD (52%), F269N (26%), 269delF (6.5%), 268_269insNGAE (2.6%), 268_269insNGAD (1.3%)
E182	2.85	E181K (93%), 181_182insS (5.4%), E182G (0.4%), [37 large deletions (0.8%)]
A312	1.71	312delA (36%), 311_312insV (27%), A312V (27%), A312E (7.8%), 311_313delGAA (2.2%)
A313	1.36	A313S (33%), A313T (31%), 313delA (26%), 312_313insV (4.3%), 311_313delGAA (2.1%), 313_314delAE (1.6%), A313V (1.6%)
V187	1.15	187delV (97%), V187A (2.5%), 186_187insV (0.2%), [28 large deletions (0.4%)]
S181	0.94	180_181insS (71%), 180_181insD (5.8%), S181T (5.8%), S181N (5.8%), [39 large deletions (11.7%)]

Table S2. PCR Primers

Name	Direction	Barcode ID	Target	Sequence (5' – 3', barcodeTARGET)
B2736	F	1	*vlsE*	tcagacgatgcgtcatGCGATATAAGTAGTACGACGGGGAAACCAG
B2737	F	2	*vlsE*	ctatacatgactctgcGCGATATAAGTAGTACGACGGGGAAACCAG
B2738	F	3	*vlsE*	tactagagtagcactcGCGATATAAGTAGTACGACGGGGAAACCAG
B2739	F	4	*vlsE*	tgtgtatcagtacatgGCGATATAAGTAGTACGACGGGGAAACCAG
B2740	F	5	*vlsE*	acacgcatgacacactGCGATATAAGTAGTACGACGGGGAAACCAG
B2741	F	6	*vlsE*	gatctctactatatgcGCGATATAAGTAGTACGACGGGGAAACCAG
B2742	F	7	*vlsE*	acagtctatactgctgGCGATATAAGTAGTACGACGGGGAAACCAG
B2743	F	8	*vlsE*	atgatgtgctacatctGCGATATAAGTAGTACGACGGGGAAACCAG
B2744	F	9	*vlsE*	ctgcgtgctctacgacGCGATATAAGTAGTACGACGGGGAAACCAG
B2745	R	10	*vlsE*	agtcatcgtatcgcgcCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2746	R	11	*vlsE*	cgatcagctgagcgcgCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2747	R	12	*vlsE*	tctgtagtgcgtgcgcCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2748	R	13	*vlsE*	gtcgcgacgtcagtgtCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2749	R	14	*vlsE*	tatacgtatatagacgCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2750	R	15	*vlsE*	agctctgagtctctatCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2751	R	16	*vlsE*	tctactctcgcatctaCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2752	R	17	*vlsE*	cacgatagtcgctatgCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2753	R	18	*vlsE*	atctagcgtagtgatgCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2762	R	19	*vlsE*	tgcatgcacagatgcgCAAGGCAGGAGGTGTTTCTTTACTAGCAGC
B2758	F	8	*kan*	atgatgtgctacatctGTAATACAAGGGGTGTTATGAG
B2759	F	9	*kan*	ctgcgtgctctacgacGTAATACAAGGGGTGTTATGAG
B2760	R	18	*kan*	atctagcgtagtgatgTCTGATTAGAAAAACTCATCG
B2761	R	11	*kan*	cgatcagctgagcgcgTCTGATTAGAAAAACTCATCG

Table S3. Sample barcoding for wild type mice and controls

Forward and reverse barcodes for each sample for the wild-type mice and mouse-independent controls. Asterisks (*) denote samples that could not be cultured.

	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19
F1	Week 1 Blood Clone 1 Mouse 1	Week 2 Ear Punch Clone 1 Mouse 1	Week 3 Ear Punch Clone 1 Mouse 1	Week 4 Ear Punch Clone 1 Mouse 1	Week 5 Bladder Clone 1 Mouse 1	Week 5 Ear Clone 1 Mouse 1	Week 5 Heart Clone 1 Mouse 1	Week 5 Joint Clone 1 Mouse 1		
F2	Week 1 Blood Clone 1 Mouse 2	Week 2 Ear Punch Clone 1 Mouse 2	Week 3 Ear Punch Clone 1 Mouse 2	Week 4 Ear Punch Clone 1 Mouse 2	Week 5 Bladder Clone 1 Mouse 2	Week 5 Ear Clone 1 Mouse 2	Week 5 Heart Clone 1 Mouse 2	Week 5 Joint Clone 1 Mouse 2	Other samples	*vlsE* Phusion PCR/ digest product
F3	Week 1 Blood Clone 1 Mouse 3	Week 2 Ear Punch Clone 1 Mouse 3	Week 3 Ear Punch Clone 1 Mouse 3	Week 4 Ear Punch Clone 1 Mouse 3	Week 5 Bladder Clone 1 Mouse 3	Week 5 Ear Clone 1 Mouse 3	Week 5 Heart Clone 1 Mouse 3	Week 5 Joint Clone 1 Mouse 3	Other samples	*vlsE* Q5 PCR
F4	Week 1 Blood Clone 2 Mouse 1	Week 2 Ear Punch Clone 2 Mouse 1	Week 3 Ear Punch Clone 2 Mouse 1	*	Week 5 Bladder Clone 2 Mouse 1	Week 5 Ear Clone 2 Mouse 1	Week 5 Heart Clone 2 Mouse 1	Week 5 Joint Clone 2 Mouse 1	Other samples	Other Samples
F5	Week 1 Blood Clone 2 Mouse 2	Week 2 Ear Punch Clone 2 Mouse 2	Week 3 Ear Punch Clone 2 Mouse 2	Week 4 Ear Punch Clone 2 Mouse 2	Week 5 Bladder Clone 2 Mouse 2	Week 5 Ear Clone 2 Mouse 2	Week 5 Heart Clone 2 Mouse 2	Week 5 Joint Clone 2 Mouse 2	Week 0 Clone 1	Other Samples
F6	Week 1 Blood Clone 2 Mouse 3	Week 2 Ear Punch Clone 2 Mouse 3	Week 3 Ear Punch Clone 2 Mouse 3	Week 4 Ear Punch Clone 2 Mouse 3	Week 5 Bladder Clone 2 Mouse 3	Week 5 Ear Clone 2 Mouse 3	Week 5 Heart Clone 2 Mouse 3	Week 5 Joint Clone 2 Mouse 3	Week 0 Clone 2	Other Samples
F7	Week 1 Blood Clone 3 Mouse 1	Week 2 Ear Punch Clone 3 Mouse 1	Week 3 Ear Punch Clone 3 Mouse 1	Week 4 Ear Punch Clone 3 Mouse 1	Week 5 Bladder Clone 3 Mouse 1	Week 5 Ear Clone 3 Mouse 1	Week 5 Heart Clone 3 Mouse 1	Week 5 Joint Clone 3 Mouse 1	Week 0 Clone 3	Other Samples
F8	Week 1 Blood Clone 3 Mouse 2	Week 2 Ear Punch Clone 3 Mouse 2	Week 3 Ear Punch Clone 3 Mouse 2	Week 4 Ear Punch Clone 3 Mouse 2	Week 5 Bladder Clone 3 Mouse 2	Week 5 Ear Clone 3 Mouse 2	Week 5 Heart Clone 3 Mouse 2	Week 5 Joint Clone 3 Mouse 2	*kan* Phusion PCR	
F9	Week 1 Blood Clone 3 Mouse 3	*Kan* Q5 PCR	Week 3 Ear Punch Clone 3 Mouse 3	Week 4 Ear Punch Clone 3 Mouse 3	Week 5 Bladder Clone 3 Mouse 3	Week 5 Ear Clone 3 Mouse 3	Week 5 Heart Clone 3 Mouse 3	Week 5 Joint Clone 3 Mouse 3	*kan* digest product	

Table S4. Sample barcoding for SCID mice and controls

Forward and reverse barcodes for each sample for the wild-type mice and mouse-independent controls. Asterisks (*) denote samples that could not be cultured.

	R10	R11	R12	R13	R14	R15	R16	R17	R18
F1	Week 1 Blood Clone 1 Mouse 1	Week 2 Ear Punch Clone 1 Mouse 1	Week 3 Ear Punch Clone 1 Mouse 1	Week 4 Ear Punch Clone 1 Mouse 1	Week 5 Ear Clone 1 Mouse 1	Week 5 Bladder Clone 1 Mouse 1	Week 5 Heart Clone 1 Mouse 1	Week 5 Joint Clone 1 Mouse 1	Week 5 Culture Mixture of Clones 1, 2, 3
F2	Week 1 Blood Clone 1 Mouse 2	Week 2 Ear Punch Clone 1 Mouse 2	Week 3 Ear Punch Clone 1 Mouse 2	Week 4 Ear Punch Clone 1 Mouse 2	Week 5 Ear Clone 1 Mouse 2	Week 5 Bladder Clone 1 Mouse 2	Week 5 Heart Clone 1 Mouse 2	*	
F3	Week 1 Blood Clone 1 Mouse 3	Week 2 Ear Punch Clone 1 Mouse 3	Week 3 Ear Punch Clone 1 Mouse 3	Week 4 Ear Punch Clone 1 Mouse 3	Week 5 Ear Clone 1 Mouse 3	Week 5 Bladder Clone 1 Mouse 3	Week 5 Heart Clone 1 Mouse 3	Week 5 Joint Clone 1 Mouse 3	
F4	Week 1 Blood Clone 2 Mouse 1	Week 2 Ear Punch Clone 2 Mouse 1	Week 3 Ear Punch Clone 2 Mouse 1	Week 4 Ear Punch Clone 2 Mouse 1	Week 5 Ear Clone 2 Mouse 1	Week 5 Bladder Clone 2 Mouse 1	Week 5 Heart Clone 2 Mouse 1	Week 5 Joint Clone 2 Mouse 1	
F5	Week 1 Blood Clone 2 Mouse 2	Week 2 Ear Punch Clone 2 Mouse 2	Week 3 Ear Punch Clone 2 Mouse 2	Week 4 Ear Punch Clone 2 Mouse 2	Week 5 Ear Clone 2 Mouse 2	Week 5 Bladder Clone 2 Mouse 2	Week 5 Heart Clone 2 Mouse 2	Week 5 Joint Clone 2 Mouse 2	
F6	Week 1 Blood Clone 2 Mouse 3	Week 2 Ear Punch Clone 2 Mouse 3	Week 3 Ear Punch Clone 2 Mouse 3	Week 4 Ear Punch Clone 2 Mouse 3	Week 5 Ear Clone 2 Mouse 3	Week 5 Bladder Clone 2 Mouse 3	Week 5 Heart Clone 2 Mouse 3	Week 5 Joint Clone 2 Mouse 3	
F7	Week 1 Blood Clone 3 Mouse 1	Week 2 Ear Punch Clone 3 Mouse 1	*	Week 4 Ear Punch Clone 3 Mouse 1	Week 5 Ear Clone 3 Mouse 1	Week 5 Bladder Clone 3 Mouse 1	Week 5 Heart Clone 3 Mouse 1	Week 5 Joint Clone 3 Mouse 1	
F8	Week 1 Blood Clone 3 Mouse 2	*	Week 3 Ear Punch Clone 3 Mouse 2	Week 4 Ear Punch Clone 3 Mouse 2	Week 5 Ear Clone 3 Mouse 2	Week 5 Bladder Clone 3 Mouse 2	Week 5 Heart Clone 3 Mouse 2	Week 5 Joint Clone 3 Mouse 2	
F9	Week 1 Blood Clone 3 Mouse 3	Week 2 Ear Punch Clone 3 Mouse 3	Week 3 Ear Punch Clone 3 Mouse 3	Week 4 Ear Punch Clone 3 Mouse 3	*	Week 5 Bladder Clone 3 Mouse 3	*	Week 5 Joint Clone 3 Mouse 3	