V Gene Diversity Allows the B Cell Receptor to Accommodate Specific Antigen Shapes

V Gene Diversity Allows the B Cell Receptor to Accommodate Specific Antigen Shapes

Alex Lee Zhu

A Thesis Submitted to the Faculty of
Harvard Medical School
in Partial Fulfillment of the Requirements for the Degree of
Master of Medical Sciences in Immunology
Harvard University
Boston, Massachusetts
May, 2018

Thesis Advisor: Dr. Daniel Lingwood, PhD Alex Lee Zhu

V Gene Diversity Allows the B Cell Receptor to Accommodate Specific Antigen Shapes

Abstract

Forming complementarity to previously unseen foreign material is a key function of the antibody response. The antibody-ligand binding surface is created by a diversity of V gene encoded and non-encoded complementarity determining regions (CDRs) displayed by the germline B cell receptor (BCR). New antigen is thought to be chiefly engaged by the stochastically derived CDR3, calling into question the role of diversity in remaining germline encoded CDR1 and CDR2 regions. To functionally address this role, we generated humanized animals in which antibody response proceeds with normal CDR3 diversification and antibody light chain (LC) usage, yet is constrained to one or four human V_H genes. We find that while these animals accommodated IgG antibody responses to diverse protein antigens, single V_H gene mice failed to generate antibodies against bacterial exopolysaccharides—components of carbohydrate nanospheres present in bacterial biofilms which constitute thymus-independent type-2 antigens (TI-2). IgM responses to TI-2 antigens are mounted exclusively through BCR cross-linking; thus, production of IgM to TI-2 antigens could be a direct measure of antigen accommodation by the BCR. Single V_H gene mice could generate antibodies to TNP-ficoll, indicating that the TI-2 IgM response was not compromised; rather, it was exopolysaccharide geometry that could not be accommodated. By contrast, animals with four-V_H genes generated robust IgM responses to the exopolysaccharides, indicating that diversity in the V_H gene repertoire could overcome this geometric constraint. Our data thus demonstrates that V_H gene diversity functions to engage CDR3-independent antigen geometries and may have been evolutionarily retained for this purpose.

Table of Contents

List of Figures

Appendix Figures

Acknowledgements

I would like to thank

Professors Shiv Pillai and Michael Carroll
for blessing me with such an opportunity to grow

Diane Lam
for her kindheartedness and all the cookies

Selina Sarmiento
for ensuring that things run smoothly

my classmates
for some of my fondest memories

my colleagues at the Ragon
for countless insights

and
Daniel Lingwood
for the revelation that research is a joy

Chapter 1

Background

Antibodies recognize an enormous variety of unique molecular shapes to draw the distinction between self and nonself. Binding to three-dimensional epitopes occurs at the antigen binding site (ABS) on the bivalent Fab end of immunoglobulin which each consist of six complementarity determining regions (CDR). In order to perceive a vast diversity of non-self antigens, the immune system must generate an equally vast repertoire of antigen binding partners. Although our genomes are limited to merely an estimated 20,000 genes, our immune systems undergo a somatic recombination process known as V(D)J recombination (Fig. 1a) to produce billions of unique antibodies[1,2]. V(D)J recombination, a hallmark of adaptive immunity, is the phenomenon by which B lymphocyte cell receptors (BCR) are assembled. Different combinations of variable (V), diversity (D), and joining (J) gene segments are mixed and matched to produce a continuous genetic construct which, when expressed, yields a membrane-bound antigen receptor at the B cell surface. Productive recombination events and successful antigen recognition lead to proliferation signals that positively select for useful antibodies among the random reassortment products[3,4]. This microevolutionary process is the key to how our immune systems generate massive diversity from limited genetic material.

The diversity of vertebrate antibody repertoires can be attributed to two sources: combinatorial diversity and N-region junctional diversity (Fig. 1a). Combinatorial diversity involves the differential incorporation of germline encoded V genes that form CDRH1 & CDRH2. N-region junctional diversity regards the stochastic assembly process during which terminal deoxynucleotidyl transferase (TdT) randomly inserts N-nucleotides at the sites of V-D or D-J junctions[5]. This markedly contributes to antibody diversity and gives rise to the

hypervariable CDRH3, which occupies a central position in the antigen binding site (Fig. 1b) and plays a crucial role in epitope recognition[6,7]. The product of these two sources of diversity allows for the immune system to elicit an enormous variety of antigen binding partners.

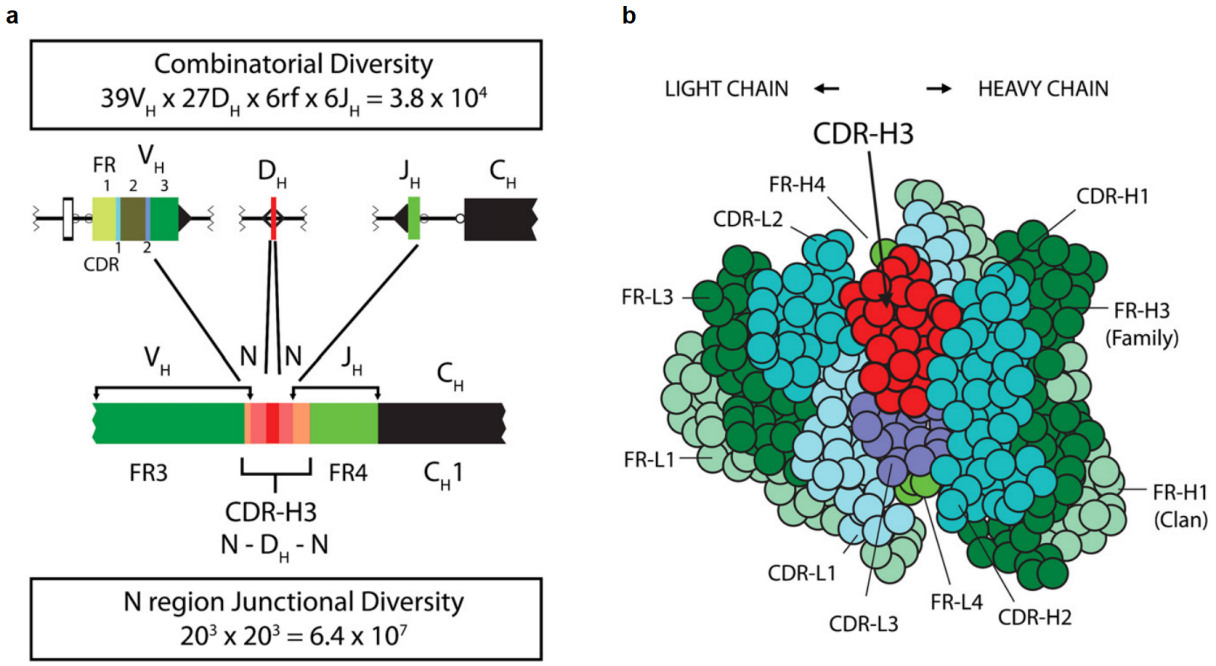

Figure 1 | Sources of diversity during antigen binding site assembly. a, Mixing and matching V, D, and J genes gives rise to combinatorial diversity, while stochastic TdT nucleotide insertions give rise to N-region junctional diversity. **b**, Bird's-eye view of an antigen binding site showing positions of CDRs. CDRH3 is shown in red, occupying a central position[7].

Historically, a largely unanswered question in the field remains as to which of these two sources of diversity dominantly drives antigen specificity. Determining the answer to this is particularly challenging given that both types of diversity contribute to the same antigen binding site. In 2000, a landmark paper published by Mark Davis's lab reported findings that offered insight into this question. The investigators were able to experimentally disentangle combinatorial diversity from N-region junctional diversity by exhibiting control over murine V genes[8]. They achieved this by genetically engineering a strain of mice restricted to a single V gene; by doing so, they exerted experimental control over combinatorial diversity so that

antibody responses in these mice would solely reflect the contributions of N-region junctional diversity to antigen recognition. Upon immunizing these mice with a variety of antigens and assessing the consequential antibody responses, elicitation of antibodies to most protein antigens were found. Upon making these observations, the investigators were led to conclude that CDRH3 hypervariability is the dominant driving force in the formation of diverse antigen binding partners.

If CDRH3 is in fact substantially adequate for the recognition of most antigens, then the question remains: why do we have about forty V genes? What is the role of V gene diversity? Moreover, why were V genes retained throughout our evolution[9-11]? For my thesis, I aimed to elucidate the functional role of V gene diversity. I approached this challenge by immunizing several strains of genetically modified mice with various antigens and assessing the ability of the B cell receptor to accommodate those antigens. By changing only the number of V genes in this experimental system, observed differences in immune responses were able to be attributed solely to V gene diversity.

The Lingwood Lab in collaboration with industry partners at Bristol-Myers Squibb have engineered multiple strains of humanized mice that possess human heavy chain and light chain loci[12,13]. Two strains of mice are restrained to single V genes: one expresses IGHV1-69*01 alone and another expresses only IGHV1-2*02. Another mouse line possesses a total of four V genes (IGHV5-5, IGHV3-30, IGHV1-69*09, IGHV4-34), which I shall refer to as quadruple V gene mice. All strains have unrestricted human CDRH3 diversity through the expression of fully humanized D and J genes. Early findings reported in the literature revealed that antibody functionality was compromised when human constant regions were engineered into mice expressing human antibodies[14]; thus, for optimal antibody functionality and expression, mouse

constant regions have been preserved in all of the strains involved in these experiments. Alongside the experimental groups, every immunization experiment was performed in parallel with wild type mice to compare antibody responses generated by the full repertoire of murine V genes to the same immunization regimens.

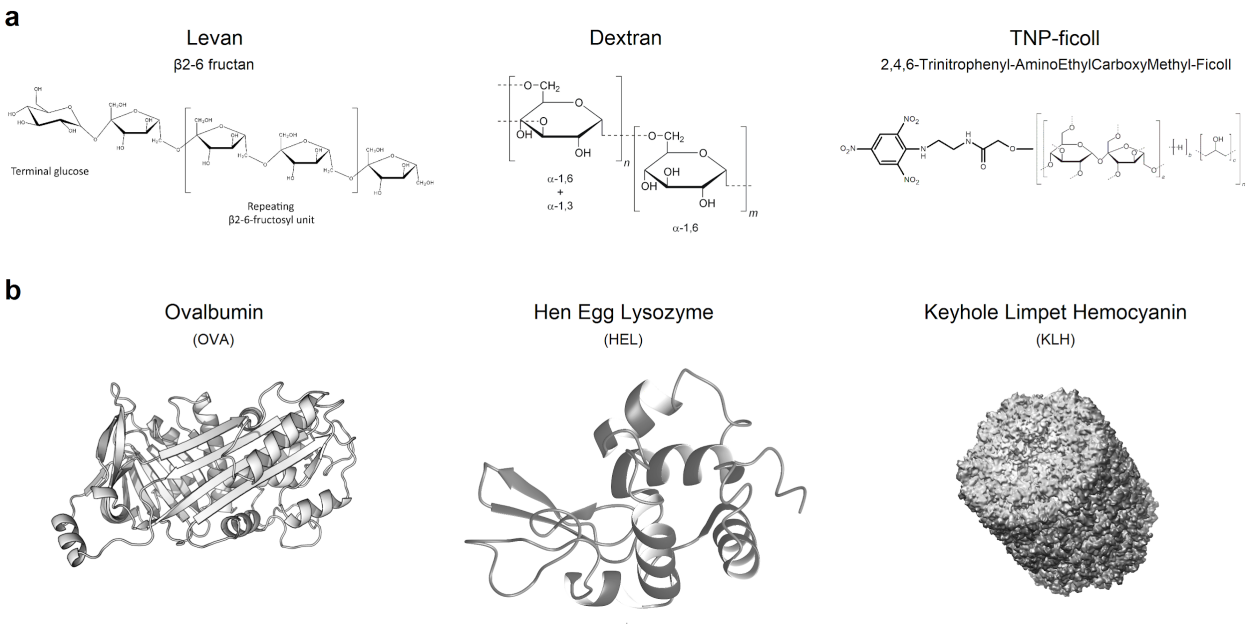

Figure 2 | Antigens selected for immunization. a, Chemical structures of thymus-independent type 2 exopolysaccharide antigens levan, dextran, and TNP-ficoll. **b**, Tertiary structures of protein antigens ovalbumin, hen egg lysozyme, and keyhole limpet hemocyanin.

In order to assess the contribution of V gene diversity to antigen accommodation, thymus-independent type 2 antigens (TI-2) dextran and levan were selected for immunization (Fig. 2a). These TI-2 antigens are well-characterized carbohydrate antigens that do not activate toll-like receptors or induce germinal center reactions[15-17]. IgM responses to TI-2 antigens are produced exclusively through BCR crosslinking; thus, the quantity of IgM produced to a TI-2 antigen can be used to directly assess the ability of the BCR to accommodate antigen. Additionally, the TI-2 antigen 2,4,6-Trinitrophenyl-AminoEthylCarboxyMethyl-FICOLL (TNP-ficoll) was used to assess B cell activation.

Responses to thymus-dependent (TD) antigens were assessed following immunization with protein antigens ovalbumin (OVA), hen egg lysozyme (HEL), and keyhole limpet hemocyanin (KLH) (Fig. 2b). These classical protein antigens were used as positive controls to ensure that normal IgG and IgM immune response mechanisms were fully functional after V gene restriction in these genetically recombineered humanized mouse strains[18-20]. Robust responses to protein immunogens confirmed that immunoglobulin production was not compromised in these mice.

TI-2 antigens dextran and levan are also notably carbohydrates—specifically, they are bacterial exopolysaccharides[21,22]. Bacteria synthesize and secrete these carbohydrates into their surrounding environments as a component of the biofilms they persist in[23,24]. This fact lends greater implications to the significance of my experiments by allowing for the framing of these findings through a broader evolutionary perspective; as I will show, my data makes for an increasingly convincing case that coevolutionary pressures stemming from our relationship with microbial commensals are responsible for shaping the immune repertoires that we are equipped with today.

Since a lack of diversity in the V gene repertoire may compromise humoral responses to bacterial antigens, I approached this phenomenon from another perspective by exploring the effects of V gene restriction on the homeostatic profiles of murine bacterial commensals. How does diversity of the V gene repertoire influence an organism's symbiotic relationship with their microbiota? Or more broadly, to what extent do genetics determine a host's homeostatic commensal profile[25]? By determining the gut commensal flora in V gene restricted mice as compared to wild type mice, a functional role of the V gene repertoire could be described in terms of the bacterial phyla that become established in the intestines of these mice. Not only does

the diversity of bacterial variety speak to this, but also the relative abundance of each bacterial phylum. By sequencing the 16S ribosomal subunit[26] extracted from the commensal bacteria isolated from the various V gene restricted mouse strains, the abundance and variety of gut microbiota could be compared to determine how V gene diversity influences commensal populations.

Chapter 2

Experimental Setup

Addressing the functional role of V gene diversity involved monitoring immune responses to various antigens and comparing the magnitudes of antibody responses to these immunizations between several different genotypes of V gene restricted mice. Immunogens were selected from the TI-2 carbohydrate antigen class due to the fact that IgM responses elicited to these molecules arise strictly through BCR crosslinking[17]. TI-2 antigens were selected for this desirable characteristic as to simplify the experimental setup by eliminating as many cellular, costimulatory, and antigen presentation related aspects of the humoral immune response irrelevant to my research question as possible. With this system in place, quantification of IgM production could then be used to directly assess the accommodative capacity of BCRs to TI-2 antigens.

Immunizations were performed after the mice had reached a minimum of six weeks of age. Prebleeds were drawn prior to injections to measure baseline preexisting antibodies to immunogens and to quantify net induced antibody production as a result of immunizations. Classical immunization regimens were used to monitor immune response development subsequent to injection. Antigen dosage and route of administration were determined by published reports of optimal results found in previous literature[8]. IgM responses to TI-2 carbohydrate antigens followed a two-week immunization regimen, consisting of bleeds drawn three days, seven days, and fourteen days post prime. Both IgG and IgM responses to protein antigens followed a classical five-week prime-boost immunization strategy. After the initial priming injection with adjuvant, the first bleed would be drawn two weeks later; a boost injection

identical to the prime would take place on the third week, followed by bleeds drawn one week and two weeks after boosting.

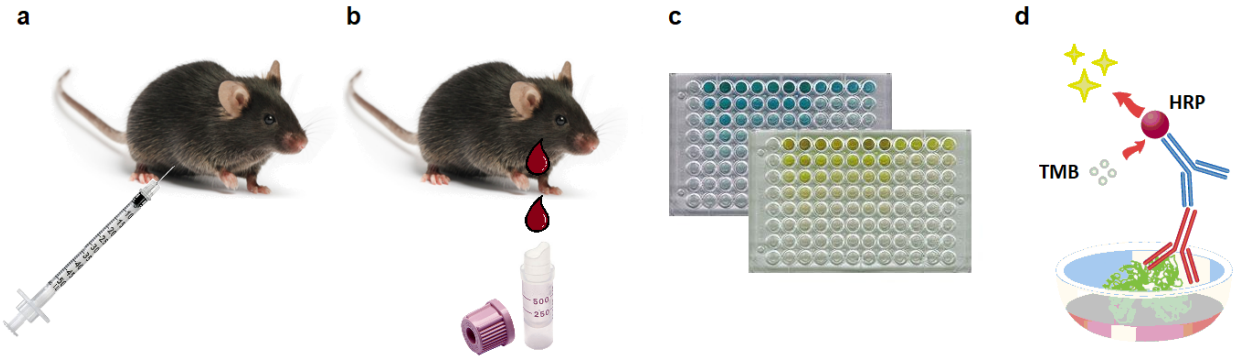

Figure 3 | Experimental overview. a, Injections were administered either intravenously or intraperitoneally. **b**, Bleeds were drawn according to separate regimens for carbohydrate and protein antigens. **c**, Sera from the blood of injected mice was serially diluted to assay for antigen-specific antibodies. **d**, Schematic diagram of an indirect ELISA. Antibodies in the sera bind to antigen-coated wells; secondary anti-mouse IgM or IgG antibodies conjugated to HRP then bind and catalyze the dye signal produced by developing reagent TMB.

Antibody responses to injected antigens were quantified by enzyme-linked immunosorbent assays (ELISA)[27]. Murine sera collected from bleed time points containing immunization-induced antibodies were serially diluted and exposed to antigen-coated 96-well plates. Quantification was determined by measuring the optical density (OD) of pigment-producing chemical reactions linked to secondary antibodies that bind the elicited murine antibodies (Fig. 3c, d). After subtracting out background levels present in the prebleed and background absorbance detected in blank wells, humoral responses were quantified on an OD scale of 0 to 4 to indicate the magnitude of antibody production, thereby signalling the BCRs ability to accommodate antigen.

BCR Sequencing:

In order to delve further into the characteristics of CDRH3 diversity in single V gene restricted mice, BCRs were sequenced using Illumina MiSeq to determine the distribution of CDRH3 lengths and amino acid compositions generated[28]. Following surgical excision of the spleen, lymphocytes were extracted and stained with a panel of antibodies in preparation for cell sorting. Fluorescence-activated cell sorting (FACS) was performed to bulk sort IgM positive B cells. These cells were then lysed to expose their transcriptomes; primers specific to the immunoglobulin locus were then used to amplify IgM templates containing the CDRH3s generated. After polymerase chain reaction (PCR) product purification and DNA quantification, dilutions to be loaded onto the MiSeq flow cell were prepared for optimal cluster density. Analysis of the read outputs will be performed in collaboration with Sam Kazer of Alex Shalek's lab.

16S MiSeq:

Another approach we took to understand the functional purpose of V gene diversity involved comparing the intestinal microbiota between the different V gene restricted mouse strains. To investigate the effect of V gene repertoires on gut commensal bacteria, I extracted the DNA of intestinal microbiota sampled from about a hundred V gene constricted mice. Beyond wild type controls, genotypes involved in this study included mice restricted to a single V gene and four V genes with mouse light chains, along with single V gene mice with both human heavy chains and human light chains. In order to characterize the gut commensal profiles in these animals, I amplified the 16S ribosomal subunit via PCR from the bacterial microbiomes. At the same time, barcoded primers labelled the amplicon product from each sample for multiplexed

sequencing[29]. After confirming the quality and purity of each sample by running gel electrophoresis, a library was prepared for Illumina MiSeq. Once obtained, sequence data was demultiplexed; samples were then identified within each respective bacterial phylum, allowing for the comparison of gut commensal profiles between the various V gene restricted mice.

Materials and Methods

Mice

Transgenic humanized mice of C57BL/6J background were bred in collaboration with expert mouse geneticists from Bristol-Myers Squibb to express the human heavy chain and light chain loci[30]. The mouse Fc region was retained for optimal antibody functionality. Two strains of mice were restricted to single V genes (either IGHV1-69*01 alone or IGHV1-2*02 alone), while quadruple V gene mice were restricted to four V genes (IGHV5-5, IGHV3-30, IGHV1-69*09, IGHV4-34)[31]. Counterparts to each of these three strains were also generated to express human heavy chains with retainment of mouse light chains. Mice were genotyped by extracting DNA from tail clips or ear clips, amplifying the V gene heavy chain locus using specific primers and PCR, then detecting the presence or absence bands found after running the DNA products through gel electrophoresis.

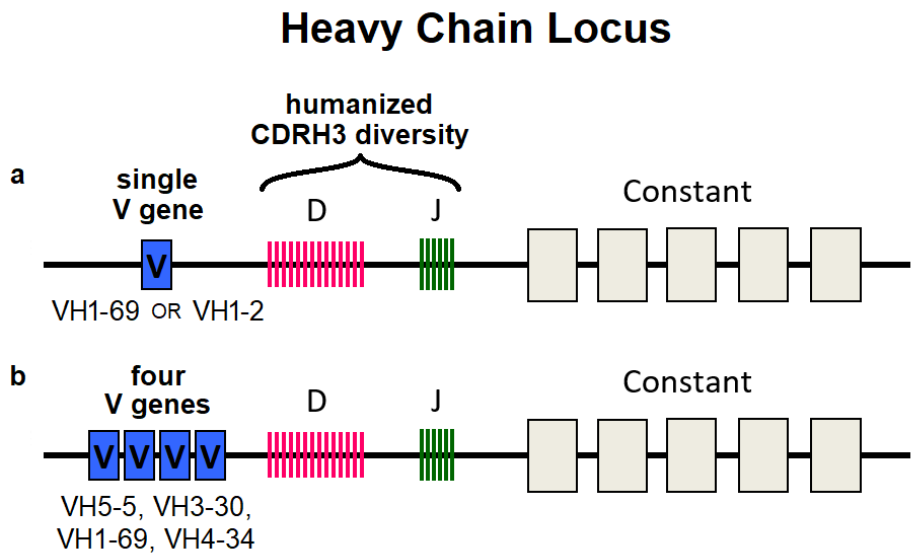

Figure 4 | Schematic diagram of the heavy chain locus HC1 in humanized V gene restricted mice. a, Single V gene restricted mice possess either VH1-69*01 alone or VH1-2*02. **b**, Quadruple V gene mice have been engineered to express four human V genes.

Reagents

TI-2 antigens used for immunizations included dextran, levan, and TNP-ficoll. Protein immunogens included OVA, HEL, and KLH. Sigma adjuvant was administered as a component alongside protein injections. Intravenous injections were performed retro-orbitally; during this procedure, mice were anesthetized with 5% isoflurane and closely monitored under a constant supply of oxygen[32].

ELISA reagents include 3,3',5,5'-tetramethylbenzidine (TMB) for HRP-conjugated secondary antibody development, along with quenching with 1 normal sulfuric acid solution. Horseradish peroxidase (HRP) conjugated secondary anti-mouse-IgM antibodies from goats or anti-mouse-IgG antibodies from sheep were used to detect antigen-specific immunoglobulins generated through immunization.

Immunizations

Mice of at least six weeks of age were used in immunization experiments. All injections were administered with a total volume of 100 μl dose in phosphate buffered saline (PBS). TI-2 immunogens (dextran, levan, and TNP-ficoll) were administered without adjuvant. 60 μg of dextran was administered intravenously via retro-orbital injection; 35 μg of levan was administered intraperitoneally; 35 μg of TNP-ficoll was administered intravenously. IgM responses to these immunogens were assessed by performing ELISA on sera drawn at 3 days, 7 days, and 14 days post prime[8].

10 μg of each proteins OVA, HEL, and KLH were delivered together as a cocktail with Sigma adjuvant[33] by intraperitoneal injection. A bleed was performed two weeks post prime, followed by a boost injection identical to the prime. Two more bleeds were performed one week

and two weeks post boost to allow for IgG class switching to occur. Background optical density and prebleeds were subtracted from post-immunization time points to obtain net induced antibody responses.

ELISA

Immunoglobulin responses were assayed by coating 96-well plates (Thermo Fisher Scientific-Nunclon Flat Bottom Polystyrol) with 400 nanograms of TI-2 antigen or 200 nanograms of protein antigen per well prior to adding serial dilutions of mouse serum[34]. After coating plates overnight at 4°C, blocking would be performed with 1% bovine serum albumin for an hour at room temperature to eliminate sites of potential nonspecific binding. Plate washes were performed using 1x PBS with 0.05% Tween 20 (PBST). Serum was obtained at each time point by performing sublingual bleeds of each mice and centrifuging the blood samples at 5000 RPM for ten minutes and collecting the topmost serum.

Serial dilutions of the serum in PBS of 1:20, 1:100, 1:500, 1:2500, 1:12500, 1:62500, and 1:312500 would be added to the blocked and washed plates, followed by a negative control well containing PBS to measure background levels of optical density. Antibodies in the serum would have an hour to bind to antigens in the well, then would be washed with PBST thrice. Secondary anti-mouse-IgM antibodies from goats or anti-mouse-IgG antibodies from sheep conjugated to HRP would be applied to all wells of each plate at 0.02% concentration. After an hour of binding and three more washes of PBST, 100 µl of TMB would be used for development for ten minutes in the dark. The reaction between HRP and TMB forms a blue signal dye that is quenched with 100 µl of 1 normal sulfuric acid, turning the signal yellow and halting any further reaction[27]. Lastly, optical density would be measured at 450 nm with a reference wavelength of 570 nm on a

TECAN Infinite M1000 Pro plate reader. To obtain net induced humoral responses, background measurements from the negative control wells were subtracted out along with any preexisting responses found in prebleeds. ELISA dilution curves were generated using GraphPad PRISM software.

IgM^+ B cell Sorting:

Following the surgical excision of spleens from euthanized mice, lymphocytes were dissociated by physically mashing the spleen in a tissue culture plate containing a media consisting of 2% fetal bovine serum (FBS) in 1x PBS. All lymphocytes were then transferred to eppendorf tubes and centrifuged at 1500 rpm for 5 minutes at 4°C. A visible pellet would form if cell counts surpassed one million. The cells were then washed by resuspension with 2% FBS and centrifuged again. Incubation with a 1% Ammonium-Chloride-Potassium (ACK) Lysing Buffer destroyed all red blood cells present in the sample, leaving the lymphocytes unaffected. Centrifugation once again collected the cells into a pellet and allowed for the aspiration of all undesired components in the supernatant. Following another wash step, the lymphocytes were stained with 1% solutions of the following antibodies:

Alexa Fluor 700	anti-mouse CD4 antibody
PE/Cy7	anti-mouse CD3 epsilon antibody
Brilliant Violet 421	anti-mouse CD19 antibody
Brilliant Violet 605	anti-mouse IgM antibody
PerCP-Cy5.5	anti-mouse IgG antibody

Cellular markers CD3 epsilon and CD4 were used to negatively select out T cell populations. B-lymphocyte antigen CD19 was used to gate for the desired B cells. To distinguish between class-switched B cells, IgG positive B cells were gated out, leaving IgM positive B cells (Fig. 5) to be

used for BCR sequencing. Incubation with these photoreactive antibodies occured in the dark for an hour at 4°C with gentle agitation.

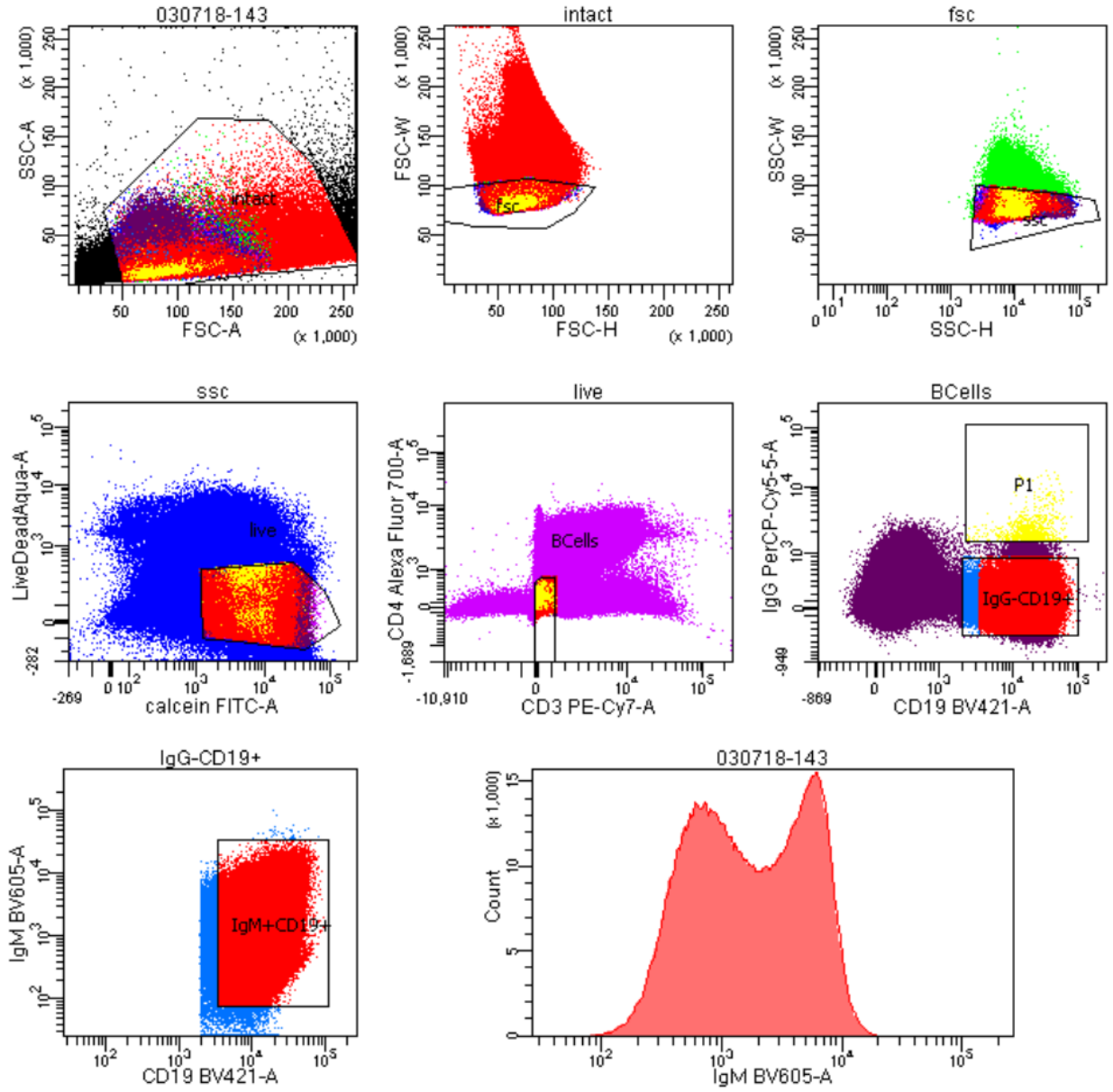

Figure 5 | Gating strategy used to sort for IgM$^+$ B cells from humanized mouse spleens. Among viable lymphocytes, T lymphocytes were excluded. From the B cells expressing CD19, IgG was negatively selected for. The desired IgM positive B cells were then collected.

In order to distinguish between alive, dead, and dying cells, the cell-permanent dye Calcein AM was used identify living B cells. To prepare live and dead cellular controls for

fluorescence-activated cell sorting (FACS), a Ramos cell line had been cultured in suspension. Two equal quantities of Ramos cells were aliquoted into separate eppendorf tubes—one half to be kept alive, and the other half to be killed. The 'live' cell aliquot was left untouched while the 'dead' aliquot was killed by heating the sample to 95°C for thirty minutes. Both alive and killed cells were then mixed together and stained with the following:

Calcein acetoxymethyl	Live cells
LIVE/DEAD™ Fixable Aqua Stain	Dead cells

A compensation panel was then prepared by individually combining each of the staining antibodies used with AbC flow cytometry compensation beads from Thermo Fisher Scientific for a total of seven compensation tubes[34]. Once antibody staining of the lymphocytes was completed, the cells underwent a wash step and were then stained with Calcein AM and Aqua. After the final wash, the cells were resuspended in 2% FBS media and filtered through fine mesh blue-capped FACS collection tubes. Stained and fully prepared, the cells were stored at 4°C in the absence of light until it was time to sort.

Following fluorescence compensation to eliminate spectral overlap, each sample of lymphocytes underwent high speed multiparameter cell sorting. Among the living cells, T cells were gated out, and CD19 was used to identify B cells[35]. As droplets containing individual lymphocytes passed through laser emission, IgM-expressing B cells were identified by the presence of BV605 fluorochrome and given a positive charge to deflect them into a separate collection tube. Once sorting was complete, each mouse spleen was found to yield over three million IgM positive B cells. A biological replicate of this B cell FACS run can be found in the appendix (Fig A.5). The obtained IgM^+ B cells were then lysed using a lysis buffer to expose their transcriptomes. Messenger RNA was converted into a more stable cDNA library for further downstream processing.

BCR Sequencing:

Following nucleic acid extraction, whole transcriptome amplification (WTA) was then performed through PCR. Primers were then designed to target and specifically amplify the heavy chain CDRs through another 45 cycles of PCR. The PCR product was then run through a DNA agarose gel containing 1% Sybr Safe™ adjacent to a lane of 100 base pair ladder. With the band determined to be at the correct size, a stringent gel extraction was performed to cleanly obtain the desired product. After confirming the size and concentration of the amplicon through TapeStation and and Qubit, a six picomolar library was prepared for DNA sequencing. Computational analysis of the reads then revealed the size distributions of CDRH3 generation along with amino acid composition to be compared alongside human BCR data. This comparison would then reveal to us how informative our humanized mouse model is of human antibody generation.

16S Ribosomal RNA MiSeq:

Stool samples were collected from immunologically naïve mice around six to ten weeks of age and frozen immediately. Beyond a wild type control group, samples were collected from heterozygous and homozygous mice of genotypes IGHV1-69*01, IGHV1-2*02, quadruple V gene mice, and fully humanized IGHV1-2*02 mice (possessing human heavy chain and human light chain). About eight to ten mice from each of these genotypic groups were sampled depending on availability. Using the QIAamp PowerFecal® DNA Kit from Qiagen[36], stool pellets were vigorously vortexed with beads for ten minutes to break up solids and lyse commensal microbes in each sample. A series of centrifugation and wash steps isolated the bacterial DNA

prior to filtering the sample through a spin column. The concentration of each DNA sample was then quantified using NanoDrop and stored at -20°C.

Comparisons of the commensal profiles between different genotypes of mice were based off of phylogenetic classification of bacteria according to hypervariable V4 region sequences of the 16S ribosomal subunit. In order to identify the variety of intestinal commensals found within each group of mice, the 16S V4 region was amplified using PCR. Multiplexed Illumina MiSeq was used to sequence the 16S regions; to keep track of sample origins, each one received uniquely barcoded reverse primers[29]. To ensure that the samples were free of contamination and that the PCR products were of the anticipated size, each DNA sample and its negative control were run on a SYBR Safe™ agarose gel. Size separation by electrophoresis confirmed the length of each amplicon. Samples found to have clean negative controls proceeded onward, while those containing any signs of contamination were eliminated for quality control.

A total of ninety individual samples were pooled together into six subpools, each containing fifteen samples; subpooling was performed to combine approximately equal quantities of DNA from each sample according to the concentrations determined by spectrophotometry. These six subpools of PCR products were then cleaned up using Qiagen QIAquick® PCR purification spin columns to remove all enzymes, primers, and other reagents, leaving only the DNA to be sequenced. A 'master' library containing all ninety samples was made by combining the six subpools, to be loaded onto the MiSeq flow cell.

Illumina MiSeq operates on the principle of sequencing by synthesis (SBS), which involves the imaging of fluorescent terminator dyes bound to nucleotides to determine the order of base calls[37-39]. After nucleotides are bound and imaged, the terminators are cleaved, allowing the next nucleotide to bind. Due to the nature of this sequencing strategy, the concentration of the

DNA library loaded into the MiSeq sequencer directly determines the cluster density of sequences on the flow cell. Cluster density is an important factor for the quality of base calls; overclustering prevents the camera from distinguishing bases, while underclustering fails to deliver a signal of adequate strength to the camera. In order to achieve an optimal cluster density of approximately 800 clusters per mm^2, the DNA concentration of the master library was meticulously quantified using NanoDrop, PicoQuant, TapeStation, Bioanalyzer and Qubit. Once an accurate concentration had been determined, a six picomolar dilution was prepared and loaded into the MiSeq cartridge with the appropriate additional primers. Finally, the sequences obtained were computationally analyzed by determining the phylogenetic variety of the bacteria found in each sample.

Results

Antigen-specific humoral responses were compared between the different genotypes of V gene restricted mice. Net induced responses were obtained by subtracting out background levels found in sera drawn prior to injection. Wild type mice served as a positive control, as they possess full murine V gene repertoires. In addition to TI-2 antigens, classic protein antigens were used to confirm that IgM and IgG responses were indeed intact, and that null responses were not false negatives resulting from crippled humoral immunity due to genetic engineering.

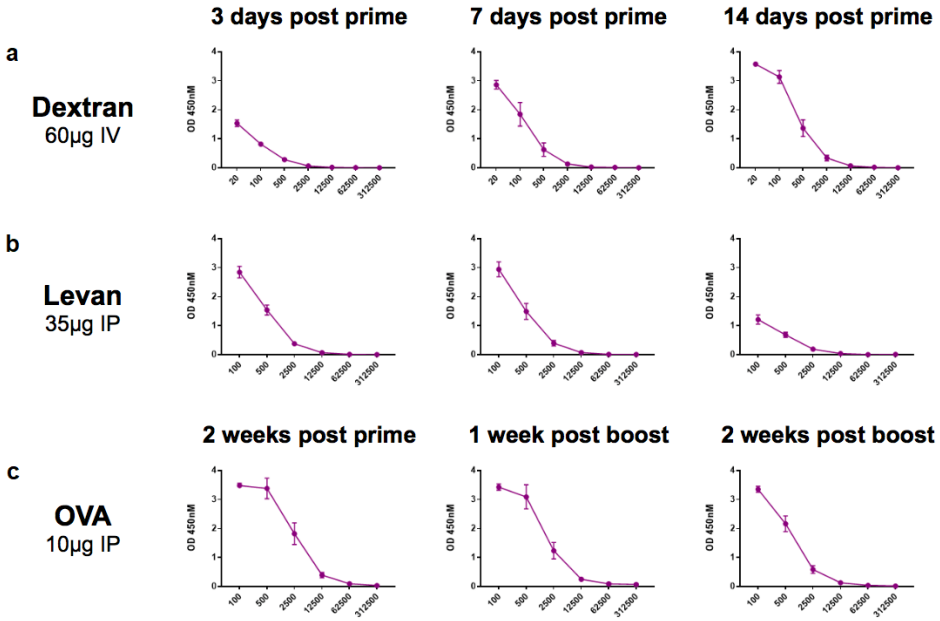

Figure 6 | Development of antigen-specific responses in wild type mice. a, IgM responses to dextran. **b**, IgM responses to levan. **c**, IgG responses to ovalbumin.

In wild type mice, robust responses to dextran and levan were observed; IgM responses to dextran increased over the course of two weeks (Fig. 6a), while responses to levan were found to wane slightly over time (Fig. 6b). This could be due to lower immunogenicity, or perhaps the lower administered dosage of levan compared to dextran or difference in route. Unsurprisingly, robust IgG responses were produced to proteins OVA (Fig. 6c), HEL, and KLH (Fig. A.6.1).

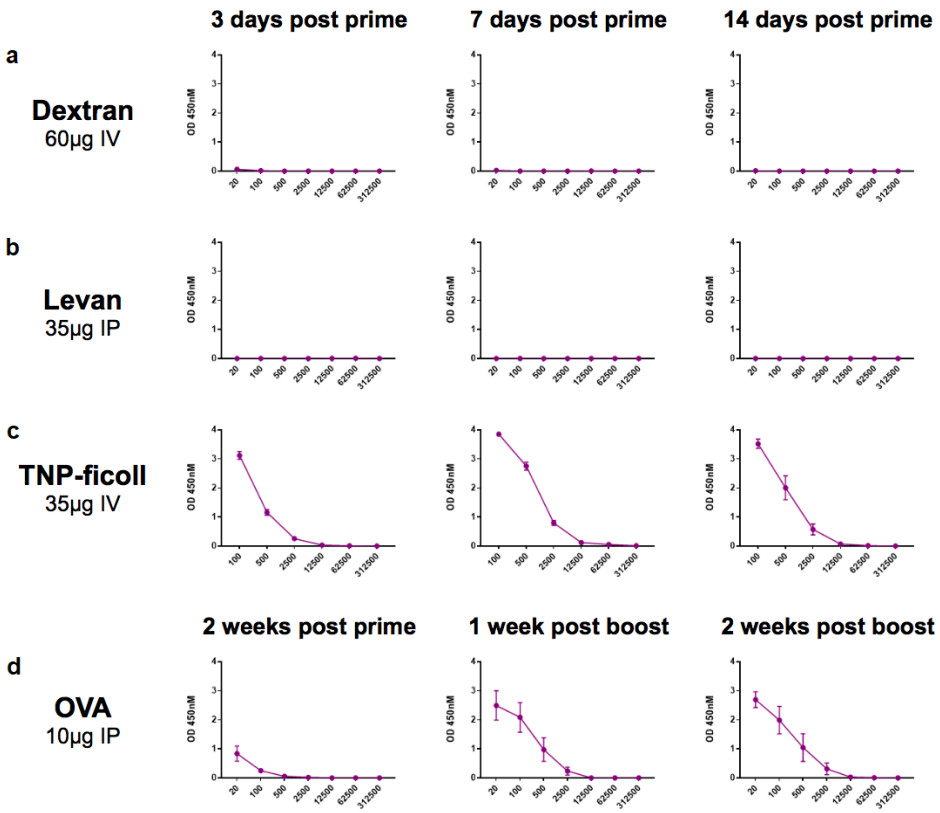

Figure 7 | Development of antigen-specific responses in IGHV1-69*01 restricted mice. **a**, IgM responses to dextran. **b**, IgM responses to levan. **c**, IgM responses to TNP-ficoll. **d**, IgG responses to ovalbumin.

Humoral responses observed in the single V gene restricted mice were generally of lesser magnitude compared to the responses of wild type mice. In the mice restricted to IGHV1-69*01, no responses were found to both dextran and levan (Fig. 7a, b), indicating that CDRH3 hypervariability was not sufficient to accommodate these TI-2 antigens. However, when immunized with another TI-2 antigen TNP-ficoll, a strong IgM response was observed (Fig. 7c), signifying that antigen recognition by BCR crosslinking occurred without issue. When injected with a cocktail of OVA (Fig. 7d), HEL, and KLH (Fig. A.7.1), boostable IgG responses were observed in this strain. Additionally, these mice were also observed to produce IgM responses to OVA, HEL, and KLH (Fig. A.7.2).

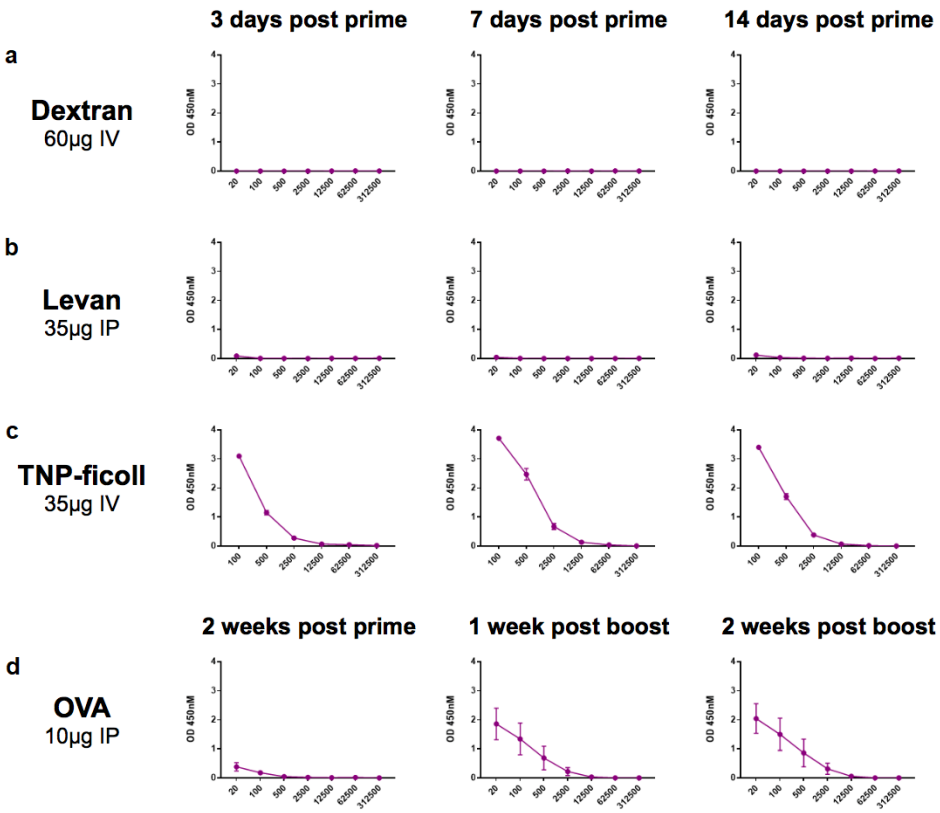

Figure 8 | Development of antigen-specific responses in IGHV1-2*02 restricted mice. **a**, IgM responses to dextran. **b**, IgM responses to levan. **c**, IgM responses to TNP-ficoll. **d**, IgG responses to ovalbumin.

To control for the possibility that the responses observed in IGHV1-69*01 mice may have been specific to that particular V gene, the same immunization regimens were performed in IGHV1-2*02 mice. Remarkably similar results to the responses observed in IGHV1-69*01 mice were found; the IGHV1-2*02 restricted mice were also unable to elicit antibody responses to dextran and levan (Fig. 8a, b). Despite this, IgM responses to TNP-ficoll were found in these mice (Fig. 8c), once again confirming that accommodation of TI-2 antigens through BCR crosslinking was operational. With these findings, nonresponsiveness to dextran and levan were concluded to be a result of single V gene restriction, a result not specific to IGHV1-69. IgG antibody responses were again found to be fully functional (Fig. 8d).

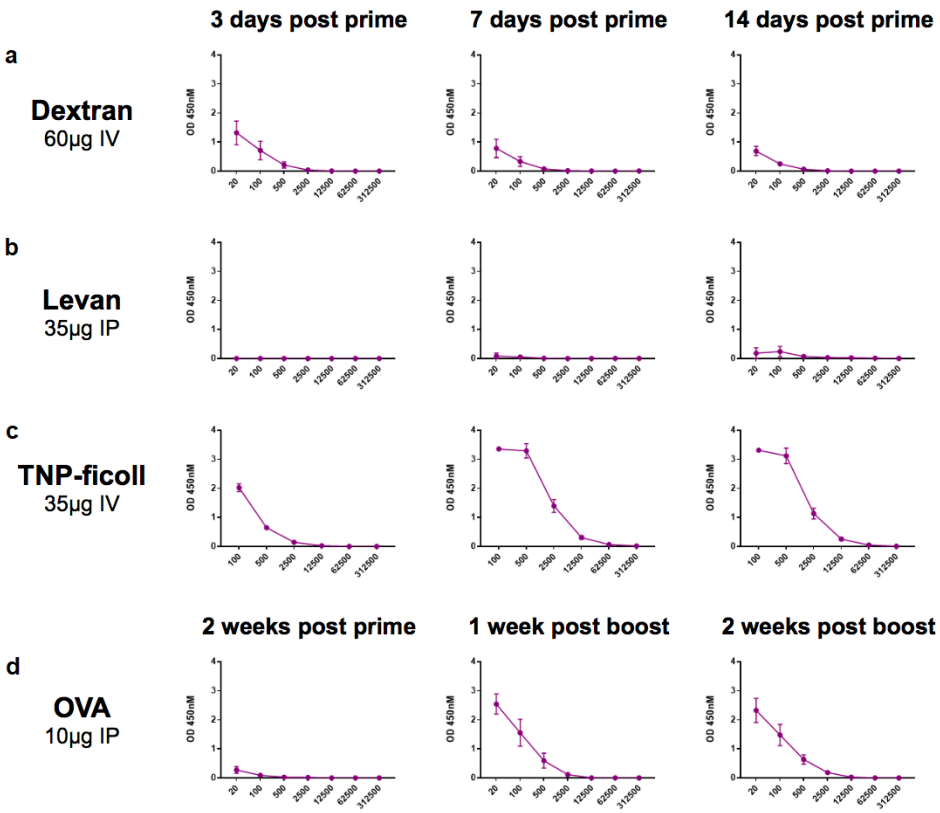

Figure 9 | Development of antigen-specific responses in quadruple V gene restricted mice. **a**, IgM responses to dextran. **b**, IgM responses to levan. **c**, IgM responses to TNP-ficoll. **d**, IgG responses to ovalbumin.

Quadruple V gene mice that underwent the same immunization regimens were found to respond moderately to dextran (Fig. 9a), a key response that was not observed in the single V gene mice. A miniscule response to levan was found fourteen days post prime (Fig. 9b). This finding indicates that, even with an increased V gene repertoire, there still do remain CDRH3-independent antigen geometries unable to be accommodated. Titration of diversity back into the V gene repertoire rescued the response to dextran, demonstrating that usage of a greater number of V genes can provide aid in the accommodation of particular antigen shapes. IgM responses to TNP-ficoll were found to be significantly greater (Fig. 9c) as compared to the single V gene mice, indicating that a larger V gene repertoire contributes not only to increases in the varietal

breadth of antigen recognition, but also to the robustness of responses to particular antigens. The appendix contains additional IgG and IgM responses to proteins (Fig. A6-A9).

BCR Sequencing Results

To confirm that our humanized mice produced humoral responses comparable to antibody generation in humans, IgM^+ B cells were isolated from the spleens of immunologically naïve mice over six weeks of age by FACS. The BCRs were then amplified and sequenced via 500 cycle MiSeq. Human peripheral blood mononuclear cells (PBMCs) were also isolated and BCR sequenced according to the same protocol. Over five million reads from each source were then used to determine the CDRH3 length distribution which averaged at 15 amino acids, along with the amino acid compositions at each Kabat position (Fig. 10).

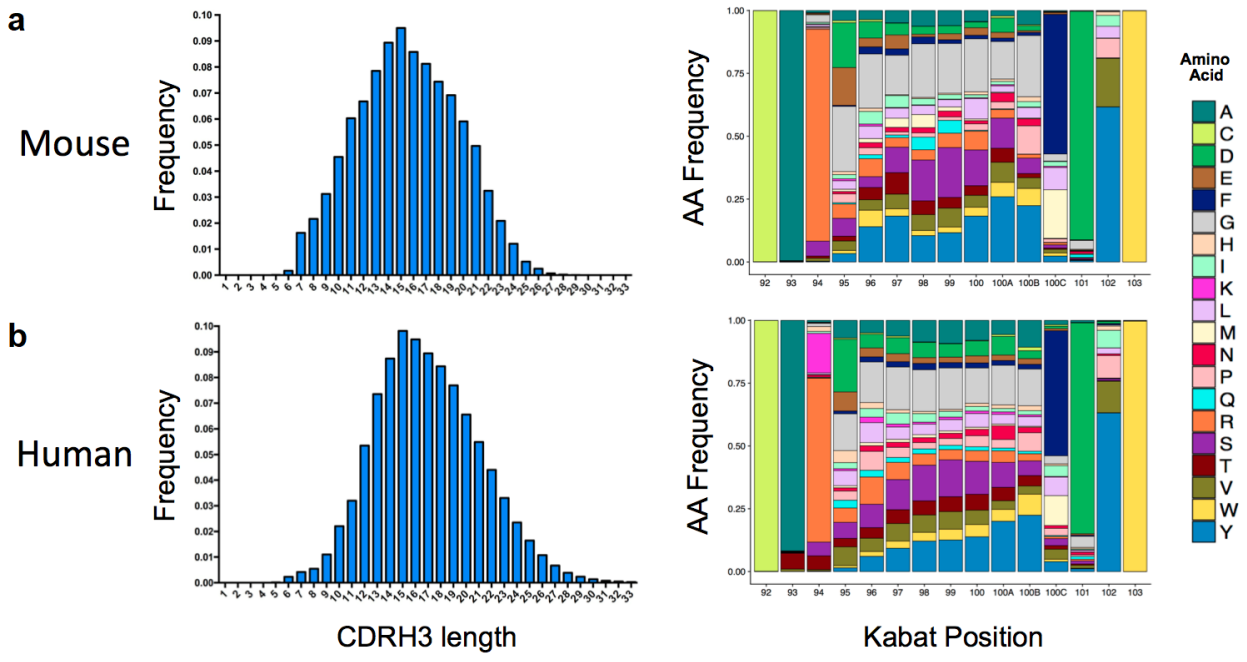

Figure 10 | CDRH3 lengths and amino acid composition observed in humanized mice as compared to humans. a, Distribution of the lengths of CDRH3s in number of amino acids generated by humanized mice along with the amino acids found at each Kabat position. **b,** Corresponding CDRH3 data found in humans.

16S MiSeq Results and Analysis

After several attempts, a six picomolar library dilution was determined to be optimal. yielding a cluster density of 735 k/mm^2 with a 89.4% pass filter rate, totalling to over fourteen million high-quality unpaired reads. Once the sequences were obtained, the fastq output file was demultiplexed, resulting in individual fastq files for each mouse. Using the trimmomatic script[40], all primer, linker, and adapter sequences (necessary for binding library sequences to the flow cell) were removed, leaving solely the 16S V4 region of interest. By running these sequences through the DADA2 script[41] in R, each 16S rRNA read was taxonomically assigned to bacterial phyla and visualized to reveal the variety and quantity of gut commensals found the V gene restricted mice (Fig. 11).

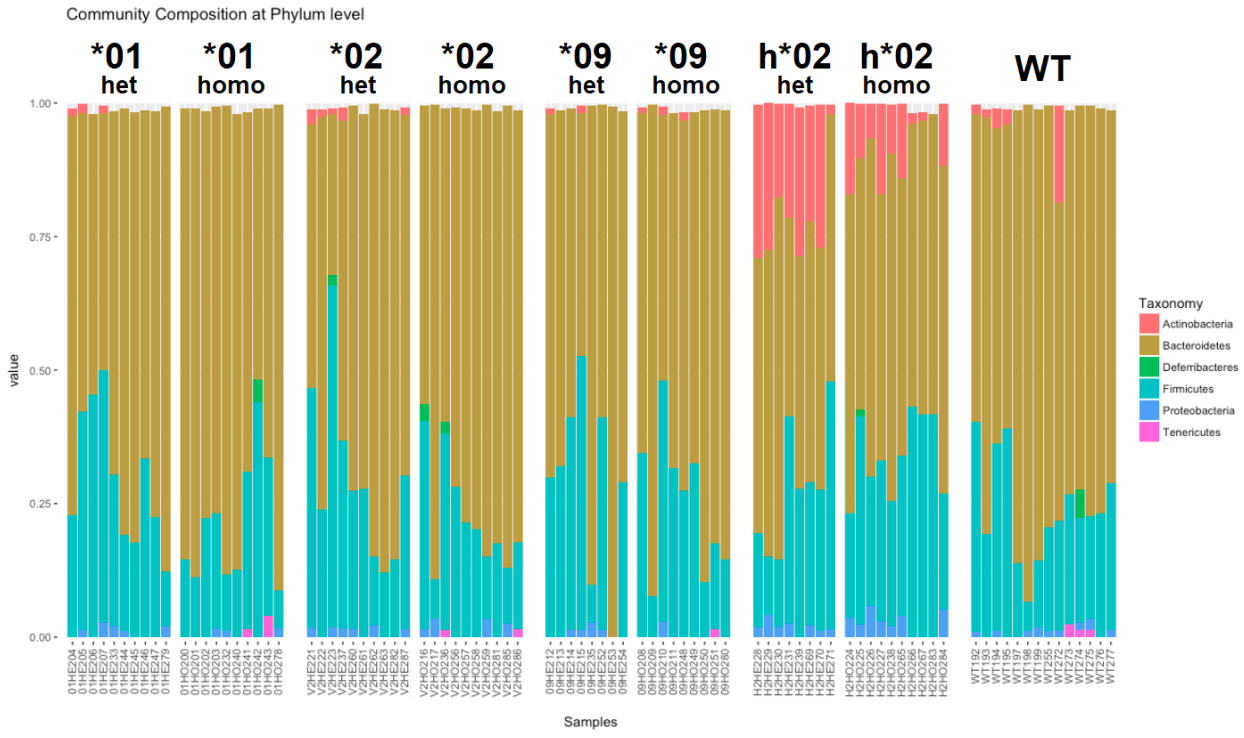

Figure 11 | Gut commensal profiles determined by 16S rRNA sequencing in V gene restricted mice. Actinobacteria prevalence found to be particularly high in mice with both human heavy chains and human light chains. Relative proportions of each bacterial phyla have been color-coded for comparison.

Differences between commensal profiles were also visualized by generating a Principal Coordinates Analysis (PCoA) plot. By doing so, diversity between the commensal bacteria of the various genotypes of mice could be easily compared by Bray-Curtis dissimilarity (Fig. 12). Single V gene mice (VH1-69*01 shown in red and VH1-2 mice shown in blue) and wild type (shown in purple) were found to be relatively similar. The profiles of four V gene mice (shown in yellow) were found to be loosely correlated with heterozygous VH1-69*01 commensal flora. The most striking difference stood out prominently in the commensal profiles of fully humanized VH1-2 mice, which congregated separately from the other genotypes with accordance to both

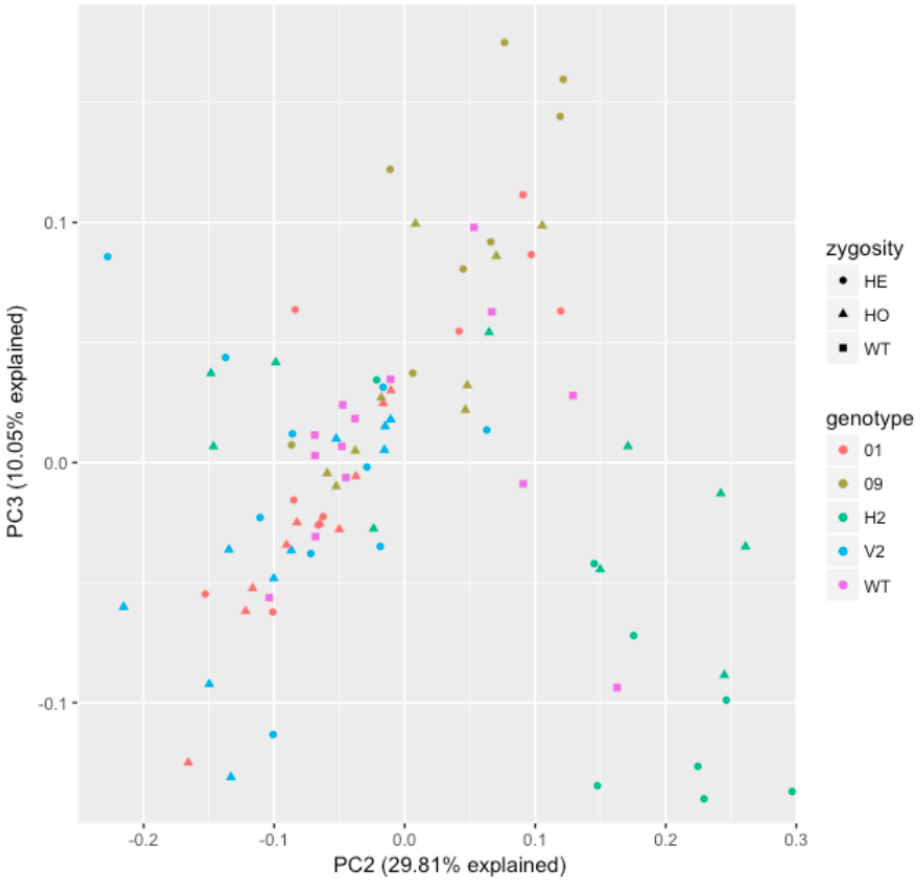

Figure 12 | PCoA of commensal flora between humanized mouse strains. Distances between points based off of compositional dissimilarity. Fully humanized murine commensal profiles shown in green symbols. WT commensals shown as purple squares. Four V gene profiles shown in yellow symbols. Single V gene profiles shown in red and blue symbols.

principal coordinates used for analysis. The fully humanized homozygous mice were found to be even more distinct than the heterozygous mice, owing to the comparatively larger proportions of actinobacteria found to inhabit their intestinal tracts.

Chapter 3

Discussion

The results of my experiments confirm the findings reported by Davis et al. concerning the fact that antigen-specific antibodies can be effectively generated in mice despite restriction of their V gene repertoires. The lack of IgM responses to TI-2 antigens dextran and levan immunizations in single V gene restricted mice was crucial in the identification of CDRH3-independent antigens. By observing responses to these CDRH3-independent antigens in quadruple V gene mice, the presence of an IgM response could then be solely attributed to the increase in V gene diversity. Taken together, the findings shown here strongly suggest that V gene diversity arose to fulfill a functional role in the accommodation of CDRH3-independent antigen geometries. Inherent biases in the usage of specific V genes in antibody responses to particular infectious agents[42,43] supports this contention; the accumulation of V genes over the evolutionary development of our immune systems may be very likely shaped by the pathogens we coevolved with.

Throughout the evolution of the vertebrate immune system, V gene accumulation has been speculated to have arose through duplication of preexisting V genes followed by mutation[44] to form other unique V genes that encode for CDRs helpful in the recognition of commonly encountered antigens. Although it may be the case that N-region junctional diversity can stochastically give rise to multitudes of various CDRH3 formations, random assembly alone is not adequate in the recognition of all foreign antigens as demonstrated by the dextran and levan immunization experiments. The retention and incorporation of germline-encoded CDRs during antibody production thereby provides one with evolutionary fitness through the innate-like coverage of specific antigen shapes. Over the course of millions of years, it is plausible that

closely coevolved microorganisms exerted selection pressure on our immune systems; ancestors who mutated to express V genes to accommodate those microorganisms may have gained an advantage to pass those genes onward.

To determine how informative our mice were of human immunology, numerous parameters have and will continue to be checked. We compared CDRH3 diversity generated (Fig. 10) and found a striking resemblance of human CDRs from B cells isolated from our mouse models. In conjunction with a high-resolution look into the somatic hypermutation occurring in these mouse strains, we also plan to take a closer look at the validity of these models in terms of cellular immunology by determining the ratio of T cells and B cells through flow cytometry. The proportion of these two immune cell types is an important figure that may influence the interpretation of both effector functions and humoral immunity observed in the murine models. Checking such parameters is paramount for quality science to proceed, so that we may defend our observations and hypotheses with confidence.

Regarding the impact of V gene diversity on intestinal commensal profiles in humanized mice, my findings reveal that particular V gene restriction or the possession of single versus quadruple V genes did not account for significant differences in gut commensal profiles. The size of the V gene repertoire therefore does not have a significant impact on diversity of microbial populations of the intestines. Surprisingly, the most prominent signal identified was the presence of actinobacteria found to be significantly more populous in fully humanized mice possessing human light chains in their antibodies. This leads us to hypothesize that the mouse light chain contribution to humoral immunity found in the partially humanized mice may prevent members of the actinobacteria phylum from inhabiting murine intestines. The specific mechanism which gives rise to this difference in commensal bacteria is not immediately obvious;

further exploratory studies have been planned for the near future to delve into this phenomenon. Ultimately, profiling intestinal commensals in humanized V gene restricted mice can be greatly informative about V gene usage following events of immunological intervention such as vaccination and its effects on microbiota.

Limitations

The practicality of this research project can be attributed primarily to the availability of the humanized mouse models engineered in collaboration with geneticists at Bristol-Myers Squibb. At the same time, the difficulty of recombineering V gene restricted mice for the purpose of determining the functional role of V gene diversity is in itself a limitation. Hypothetically, one could take a truly thorough look into this idea by immunizing a series of mice with steadily increasing numbers of V genes and observe improved accommodation to a variety of antigens the more V genes a strain possessed. While the availability of the quadruple V gene mice made this project possible, it was at the same time a limiting factor regarding not only what we could feasibly test for, but also the extent of the conclusions and claims we could draw from the antibody responses observed. This does however appear to be the most direct approach to experimentally address the function of retained human V genes over the evolutionary course of our immune systems[45].

An additional factor to consider is the specific V genes that the mice were restricted to. Naturally, different V genes do not contribute to antigen specificity in the same manner. From my results it was repeatedly clear that individual V genes IGHV1-69*01 and IGHV1-2*02 were unable to accommodate TI-2 antigens whatsoever. However, the possibility that another V gene in the quadruple V gene mice (either IGHV5-5, IGHV3-30, or IGHV4-34) might have encoded for a CDR that binds an epitope on dextran cannot be discounted. Because of this, it remains unclear whether *in vitro* immunological accommodative capacity to TI-2 antigens can be attributed solely to the increased V gene diversity or rather the incorporation of particular V genes during antibody generation. This distinction cannot be definitively drawn without a more

elaborate experimental setup to allow for higher resolution insight as to which factor truly enables TI-2 antigen accommodation in the antibody response.

In terms of identifying differences in commensal flora between the profiles of humanized mice, one clear limitation is the depth of taxonomy achieved. The preliminary study reported in this thesis assigned phylogeny according to the V4 hypervariable region of rRNA isolated from bacteria found in mouse stool. To achieve a higher-resolution look into the variety of microbes in the intestines of these mice, one would have to increase the depth of sequencing to parse out microbial diversity well beyond the phylum level. Such a feat would not only be far more costly, but also entail a more challenging computational analysis of the genetic sequences acquired to make accurate taxonomic assignments.

Future Research

In order to rigorously determine whether antibody responses to dextran in the quadruple V gene mice were due to antibody assembly using a particular V gene or increases in the V gene repertoire, one would require the quadruple V gene mice along with four other strains of mice each of which are individually restricted singly to the four V genes of the quadruple V gene strain. I foresee two conclusions to be drawn after immunizing these strains of mice with TI-2 antigens. In the first possible outcome, one may find small responses in one or more of the single V gene restricted mice. This would suggest that the quadruple V gene strain's accommodative capacity to dextran might possibly be an effect of the summation of those individual V genes, in which one of the four V gene is what drives humoral responses to antigen geometries like dextran. Such a result would not contradict the conclusions drawn because a diverse repertoire of V genes would still enable more robust antibody responses to CDRH3-independent antigens. Another possible outcome might be that all four of the single V gene restricted mice would elicit no responses to the TI-2 antigens; such a result would not only be surprising, but also further strengthen the case that diversity in the V gene repertoire in itself is what enables accommodation to CDRH3-independent antigens.

In the meantime, Bristol-Myers Squibb has just recombineered a new strain of humanized mice for continuation experiments. This particular strain has been recombineered to possess six human V genes (IGHV3-23, IGHV1-69*09, IGHV3-30, IGHV4-34, and two copies of IGHV5-51) with fully human D and J genes and murine constant regions[30]. Once humoral responses to dextran, levan, TNP-ficoll, and protein immunogens are quantified in this strain, we will be able to independently determine if the BCR accommodation of TI-2 antigens such as dextran in the quadruple V gene mice arose due to V gene diversity. Antibody responses to either levan or

dextran in these sextuple V gene mice will strongly support our contention that diversity in the V gene repertoire is the causal factor that enables BCR accommodation of CDRH3-independent antigens. In the unlikely event that no response to TI-2 antigens are found in the sextuple V gene mice, then we can attribute the accommodation of dextran to one or more of the three V genes expressed by the quadruple V gene mice (since one of them, IGHV1-69, was found to be non-responsive in the single V gene restricted strain). It is far more likely, however, that we would find a more robust IgM response to dextran in sextuple V gene mice as compared to the quadruple V gene mice. It may even be plausible that the sextuple mice might be able to accommodate and respond to levan, providing further support to reinforce the idea that V gene diversity arose to provide protection against antigen shapes that CDRH3 is unable to recognize[46,47].

Regarding V gene diversity and its effects on commensal bacteria, the next logical step is to implement rigorous controls to determine exactly which genetic factor it is that influences the successful or unsuccessful inhabitation of the murine intestinal tract by actinobacteria[48]. The mouse strains found to consistently contain actinobacteria in their commensal bacteria expressed both human heavy and light immunoglobulin chains, whereas mice expressing murine light chains with human heavy chains did not have significant populations of actinobacteria in their microbiota. Could it be that chimeric mouse light chain with human heavy chain antibodies prevent the establishment and growth of actinobacteria in the gut? If so, several follow up experiments may be performed to investigate this hypothesis. Observations can be made on the binding and neutralization of actinobacteria *in vitro* by secreted IgA antibodies harvested from partially humanized mice. Conversely, clearance of actinobacteria by partially humanized mice following gastrointestinal inoculation can be tested for; likewise, 16S sequencing of the

microbiota from both strains of mice following cohabitation can confirm this principle as well. These follow up studies would ultimately be aimed at understanding host genetic determinants of gut commensal flora, in addition to interventional events that may influence the plasticity of ones microbial profile.

Presented here in this thesis is a set of experiments aimed at addressing a fundamentally basic question left unanswered throughout the history of immunology. Diversity in the V gene repertoire of the vertebrate immune system likely evolved to overcome a weakness in adaptive immunity concerning overdependence of antigen recognition on the stochastically generated CDRH3. It is postulated that those who inherited larger and more diverse V gene repertoires likely gained an advantage in the ever-ongoing coevolutionary arms race between us and our ubiquitous commensal microbes. Fitness derived from a more consistent mode of recognizing pathogen-associated antigens through encoded V genes not only allowed for more rapid and reliable antibody responses to be produced, but also freed up precious biological resources to be allocated towards other vital life processes. It is astonishing to speculate how V gene diversity may have arose through such an extraordinary selection process, resulting in a more powerful adaptive immune system with evolutionary 'memory' directly encoded in the genome towards certain commonly encountered antigen shapes not reliably recognized solely through CDRH3 hypervariability.

Bibliography

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575-581 (1983).
2. Alt, F.W., Oltz, E.M., Young, F., Gorman, J., Taccioli, G. and Chen, J. VDJ recombination. *Immunology today* **13**, 306-314. (1992).
3. Meffre, E., Casellas, R. and Nussenzweig, M.C. Antibody regulation of B cell development. *Nature immunology*, **1**, 379. (2000).
4. Jung, D. and Alt, F.W. Unraveling V(D)J recombination: insights into gene regulation. *Cell* **116**, 299-311. (2004).
5. Benedict, C.L., Gilfillan, S., Thai, T.H. and Kearney, J.F. Terminal deoxynucleotidyl transferase and repertoire development. *Immunological reviews* **175**, 150-157. (2000).
6. Al-Lazikani, B., Lesk, A.M. and Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *Journal of molecular biology* **273**, 927-948. (1997).
7. Schroeder, H.W. and Cavacini, L. Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology* **125**, S41-S52. (2010).
8. Xu, J.L. and Davis, M.M. Diversity in the CDR3 region of V_H is sufficient for most antibody specificities. *Immunity* **13**, 37-45. (2000).
9. Janeway Jr, C.A. The immune system evolved to discriminate infectious nonself from noninfectious self. *Immunology Today* **13**, 11-16. (1992).
10. Bartl, S., Baltimore, D. and Weissman, I.L. Molecular evolution of the vertebrate immune system. *Proceedings of the National Academy of Sciences* **91**, 10769-10770. (1994).
11. Flajnik, M.F. and Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Reviews Genetics* **11**, 47. (2010).
12. Chen, J., Trounstine, M., Alt, F.W., Young, F., Kurahara, C., Loring, J.F. and Huszar, D., Immunoglobulin gene rearrangement in B cell deficient mice generated by targeted deletion of the JH locus. *International Immunology* **5**, 647-656. (1993).
13. Chen, J., Trounstine, M., Kurahara, C., Young, F., Kuo, C.C., Xu, Y., Loring, J.F., Alt, F.W. and Huszar, D. B cell development in mice that lack one or both immunoglobulin kappa light chain genes. *The EMBO Journal* **12**, 821-830. (1993).
14. Morrison, S.L., Johnson, M.J., Herzenberg, L.A. and Oi, V.T. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proceedings of the National Academy of Sciences* **81**, 6851-6855. (1984).
15. Stein, K.E. Thymus-independent and thymus-dependent responses to polysaccharide antigens. *Journal of Infectious Diseases* **165**, 49-52. (1992).
16. Mosier, D.E. and Subbarao, B. Thymus-independent antigens: complexity of B-lymphocyte activation revealed. *Immunology Today* **3**, 217-222. (1982).
17. Mond, J.J., Lees, A. and Snapper, C.M. T cell-independent antigens type 2. *Annual Review of Immunology* **13**, 655-692. (1995).

18. Adorini, L., Harvey, M.A., Miller, A. and Sercarz, E.E. Fine specificity of regulatory T cells. II. Suppressor and helper T cells are induced by different regions of hen egg-white lysozyme in a genetically nonresponder mouse strain. *Journal of Experimental Medicine* **150**, 293-306. (1979).
19. Babbitt, B.P., Allen, P.M., Matsueda, G., Haber, E. and Unanue, E.R. Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* **317**, 359. (1985).
20. Harris, J.R. and Markl, J. Keyhole limpet hemocyanin (KLH): a biomedical review. *Micron*, **30**, 597-623. (1999).
21. Hehre, E.J. and Neill, J.M., Formation of serologically reactive dextrans by streptococci from subacute bacterial endocarditis. *Journal of Experimental Medicine* **83**, 147-162. (1946).
22. Ebert, K.H. and Schenk, G. Mechanisms of biopolymer growth: the formation of dextran and levan. *Advances in Enzymology and Related Areas of Molecular Biology* **30**, 179-221. (1968).
23. Mukasa, H. and Slade, H.D. Mechanism of adherence of Streptococcus mutans to smooth surfaces I. Roles of insoluble dextran-levan synthetase enzymes and cell wall polysaccharide antigen in plaque formation. *Infection and Immunity* **8**, 555-562. (1973).
24. Gibbons, R.J. and Nygaard, M. Synthesis of insoluble dextran and its significance in the formation of gelatinous deposits by plaque-forming streptococci. *Archives of Oral Biology* **13**, 1249-IN31. (1968).
25. Carmody, R.N., Gerber, G.K., Luevano Jr, J.M., Gatti, D.M., Somes, L., Svenson, K.L. and Turnbaugh, P.J. Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host & Microbe* **17**, 72-84. (2015).
26. Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. and Schloss, P.D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology* **79**, 5112-5120. (2013).
27. Engvall, E. and Perlmann, P. Enzyme-linked immunosorbent assay, ELISA: III. Quantitation of specific antibodies by enzyme-labeled anti-immunoglobulin in antigen-coated tubes. *The Journal of Immunology* **109**, 129-135. (1972).
28. Robins, H. Immunosequencing: applications of immune repertoire deep sequencing. *Current Opinion in Immunology* **25**, 646-652. (2013).
29. Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M. and Gormley, N. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6**, 1621. (2012).
30. Lonberg, N., Taylor, L.D., Harding, F.A., Trounstine, M., Higgins, K.M., Schramm, S.R., Kuo, C.C., Mashayekh, R., Wymore, K., McCabe, J.G. and Munoz-O'Regan, D. Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* **368**, 856. (1994).

31. Pascual, V. and Capra, J.D. Human immunoglobulin heavy-chain variable region genes: organization, polymorphism, and expression. *Advances in Immunology* **49**, 1-74. Academic Press. (1991).
32. Constantinides, C., Mean, R. and Janssen, B.J. Effects of isoflurane anesthesia on the cardiovascular function of the C57BL/6 mouse. *ILAR Journal/National Research Council, Institute of Laboratory Animal Resources* **52**, 21. (2011).
33. Sigma-Aldrich. "Sigma Adjuvant System S6322." www.sigmaaldrich.com, sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/5/s6322dat.pdf
34. Höffkes, H.G., Schmidtke, G., Uppenkamp, M. and Schmücker, U. Multiparametric immunophenotyping of B cells in peripheral blood of healthy adults by flow cytometry. *Clinical and Diagnostic Laboratory Immunology* **3**, 30-36. (1996).
35. Ghia, P., ten Boekel, E., Rolink, A.G. and Melchers, F. B-cell development: a comparison between mouse and man. *Immunology Today* **19**, 480-485. (1998).
36. Hart, M.L., Meyer, A., Johnson, P.J. and Ericsson, A.C. Comparative evaluation of DNA extraction methods from feces of multiple host species for downstream next-generation sequencing. *PloS One* **10**, 0143334. (2015).
37. Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M. and Ravel, J. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6. (2014).
38. Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T. and Quince, C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research* **43**, 30-37. (2015).
39. Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. and Gu, Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341. (2012).
40. Bolger, A.M., Lohse, M. and Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120. (2014).
41. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581. (2016).
42. Van Braeckel-Budimir, N., Gras, S., Ladell, K., Josephs, T.M., Pewe, L., Urban, S.L., Miners, K.L., Farenc, C., Price, D.A., Rossjohn, J. and Harty, J.T. A T Cell Receptor Locus Harbors a Malaria-Specific Immune Response Gene. *Immunity* **47**, 835-847. (2017).
43. Dunand, C.J.H. and Wilson, P.C. Restricted, canonical, stereotyped and convergent immunoglobulin responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **370**, doi: 10.1098/rstb.2014.0238 (2015).
44. Cooper, M.D. and Alder, M.N. The evolution of adaptive immune systems. *Cell* **124**, 815-822. (2006).

45. Flajnik, M.F. and Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Reviews Genetics* **11**, 47. (2010).
46. Cohn, M., Langman, R. and Geckeler, W. Diversity 1980. *Immunology* **80**, 153-201. (1980).
47. Rajewsky, K., Forster, I. and Cumano, A. Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* **238**, 1088-1094. (1987).
48. Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K. and Kachman, S.D. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences* **107**, 18933-18938. (2010).

Appendices

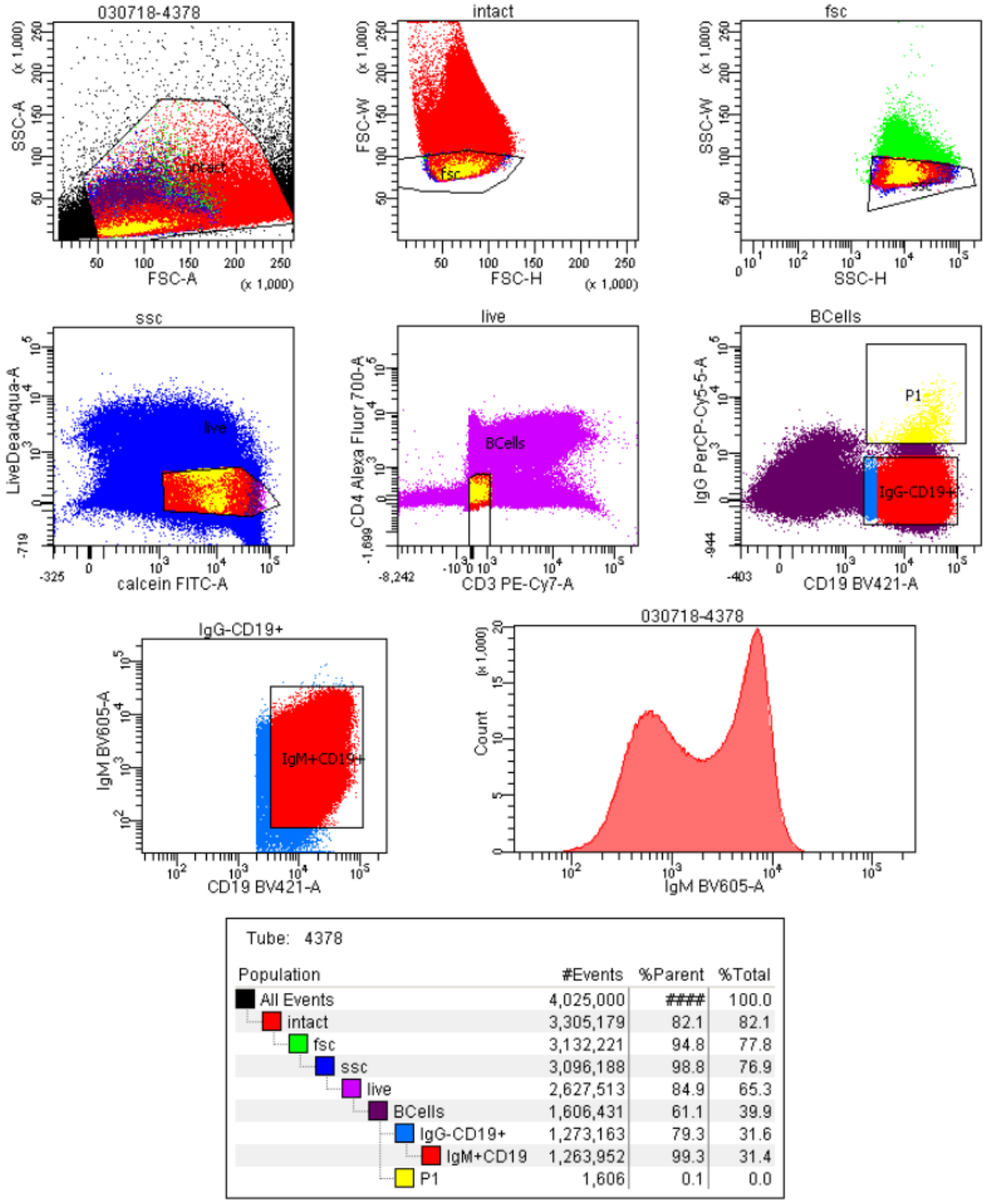

Figure A.5 | Biological replicate IgM$^+$ B cell sort performed on a fully humanized, six-week old, immunologically naïve IGHV1-2*02 mouse.

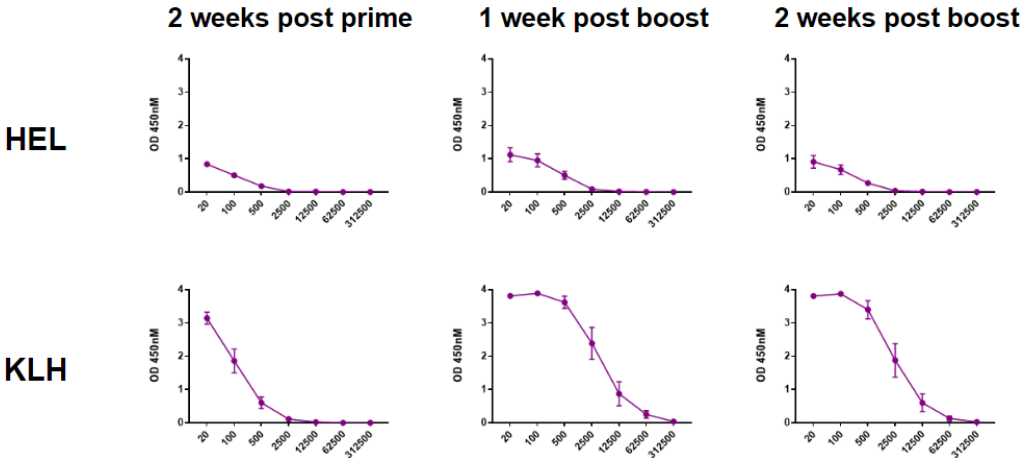

Figure A.6.1 | IgG responses to HEL and KLH in wild type mice determined by ELISA.

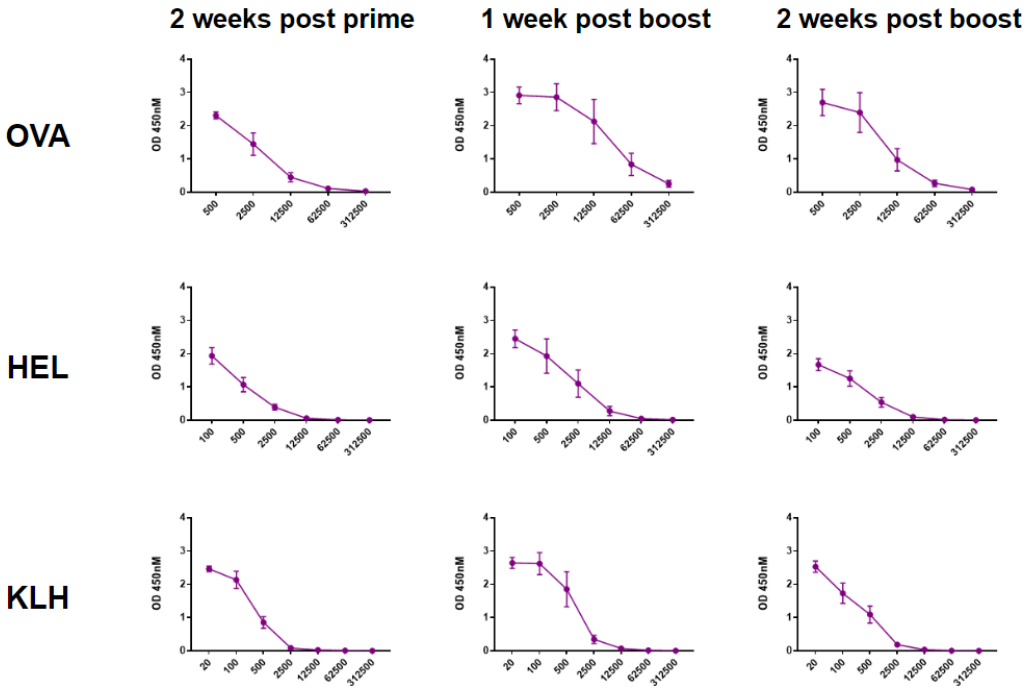

Figure A.6.2 | IgM responses to OVA, HEL, and KLH in wild type mice.

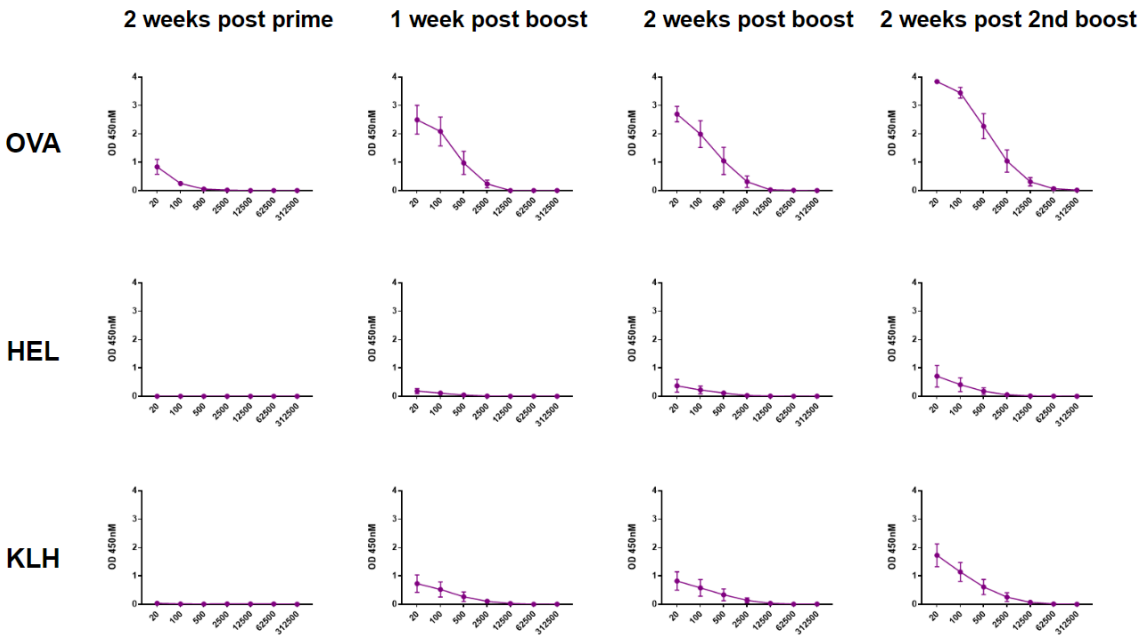

Figure A.7.1 | IgG responses to OVA, HEL, and KLH in IGHV1-69*01 single V gene mice. A second boost confirms that this genotype is in fact able to accommodate HEL and KLH.

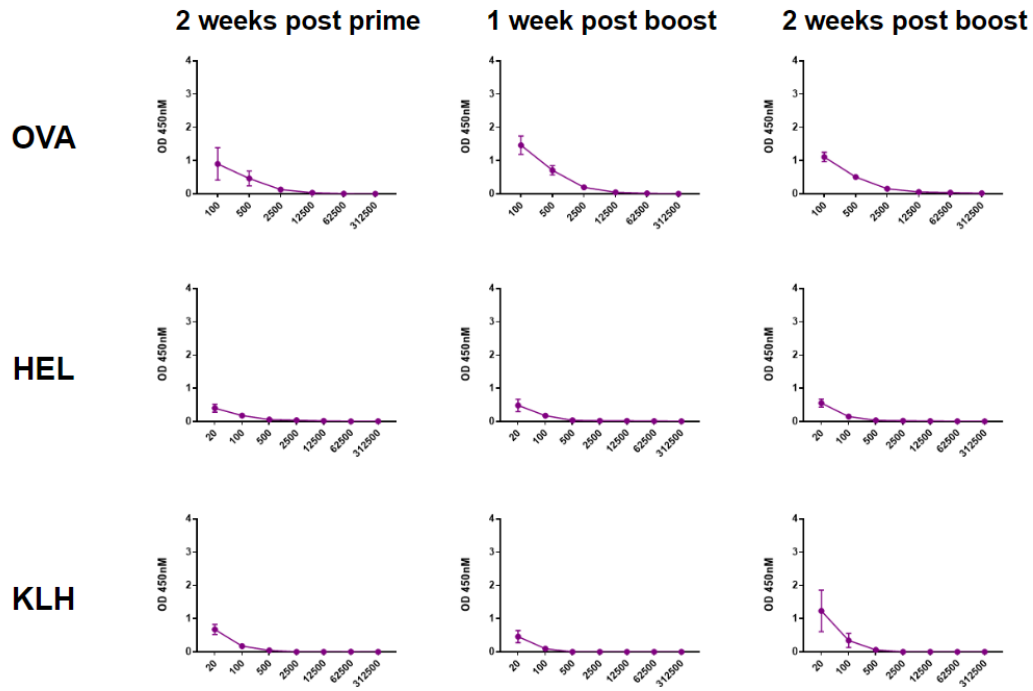

Figure A.7.2 | IgM responses to OVA, HEL, and KLH in IGHV1-69*01 single V gene mice.

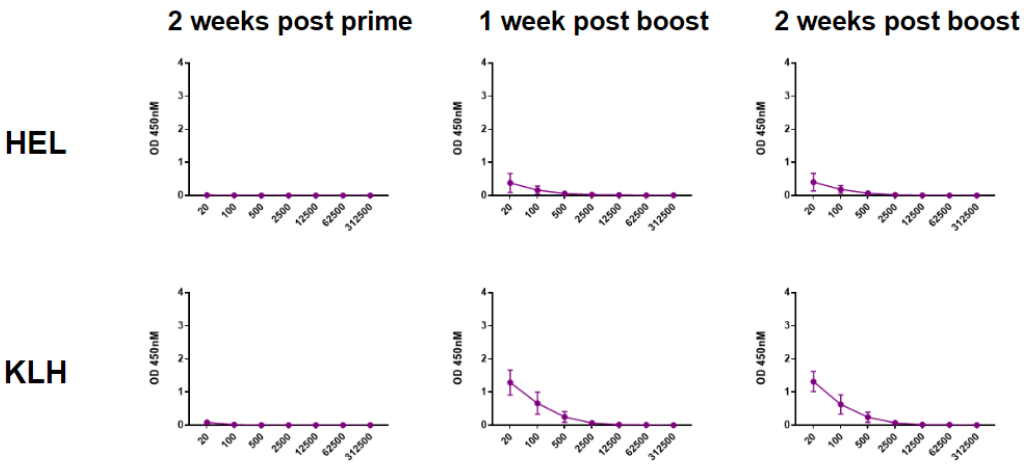

Figure A.8.1 | IgG responses to HEL and KLH in IGHV1-2*02 single V gene mice.

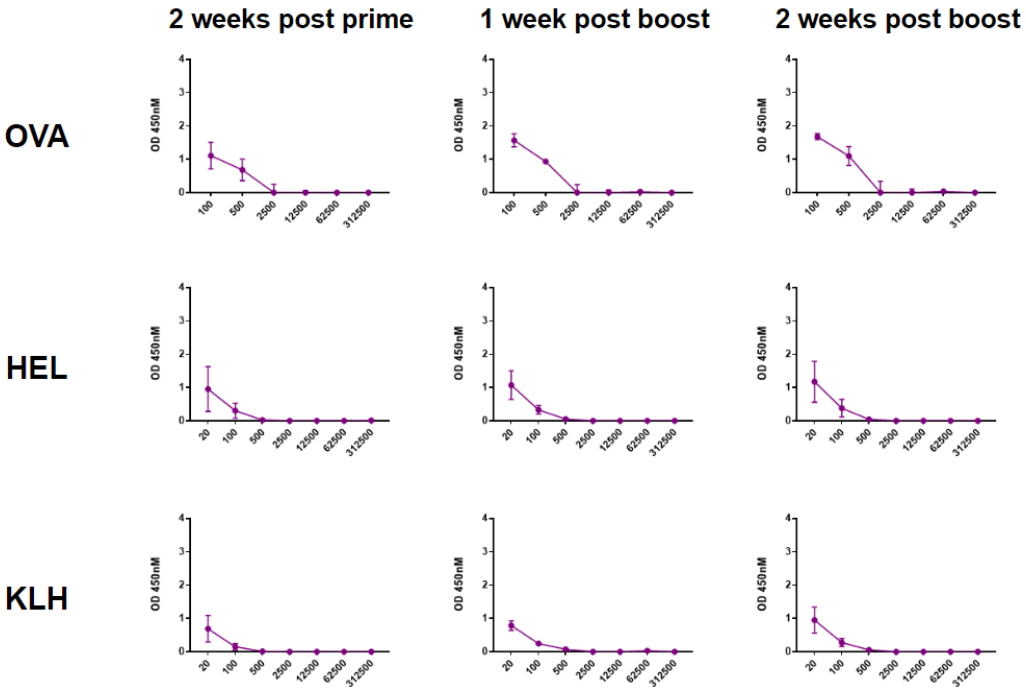

Figure A.8.2 | IgM responses to OVA, HEL, and KLH in IGHV1-2*02 single V gene mice.

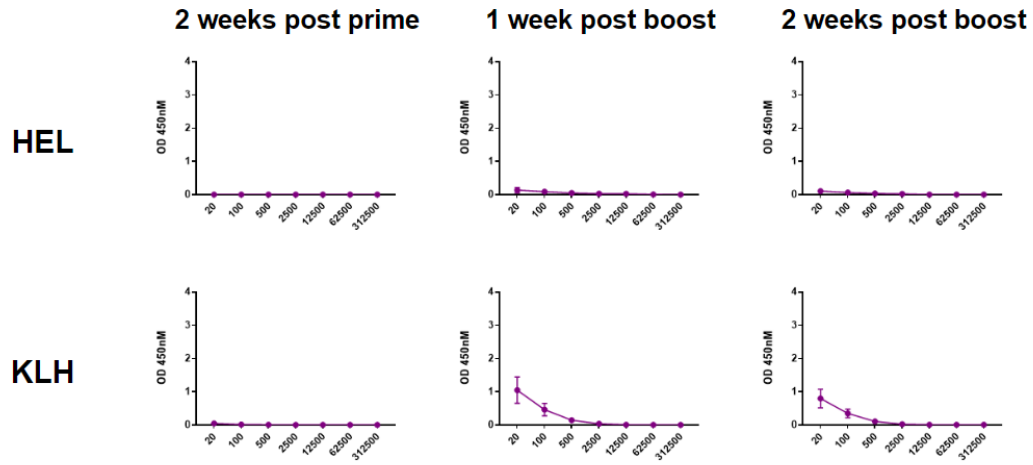

Figure A.9.1 | IgG responses to HEL and KLH in quadruple V gene mice.

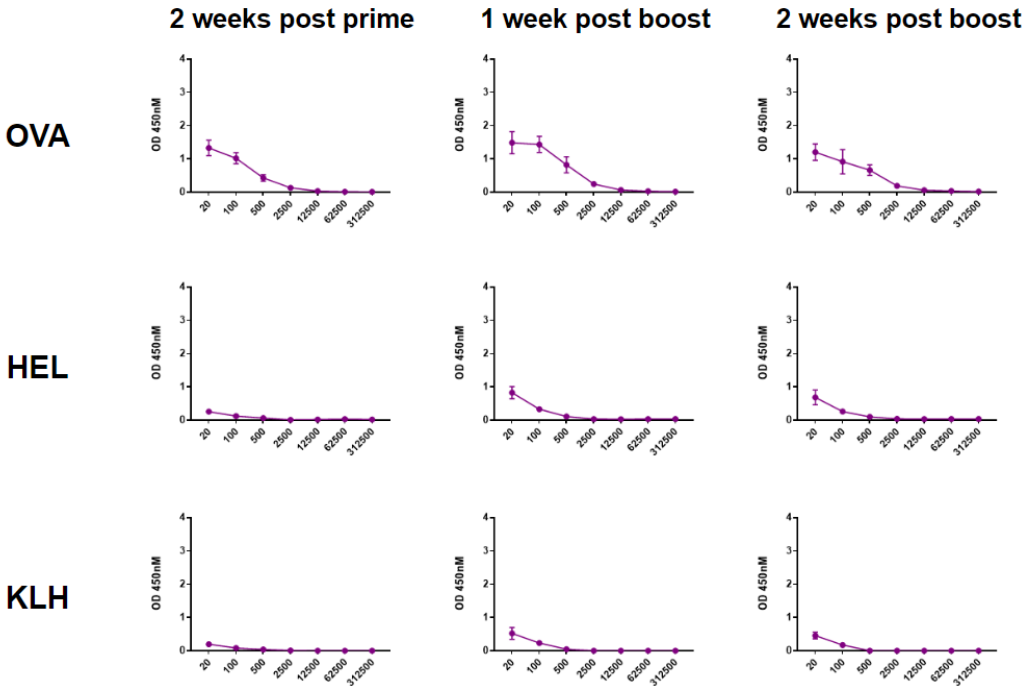

Figure A.9.2 | IgM responses to OVA, HEL, and KLH in quadruple V gene mice.

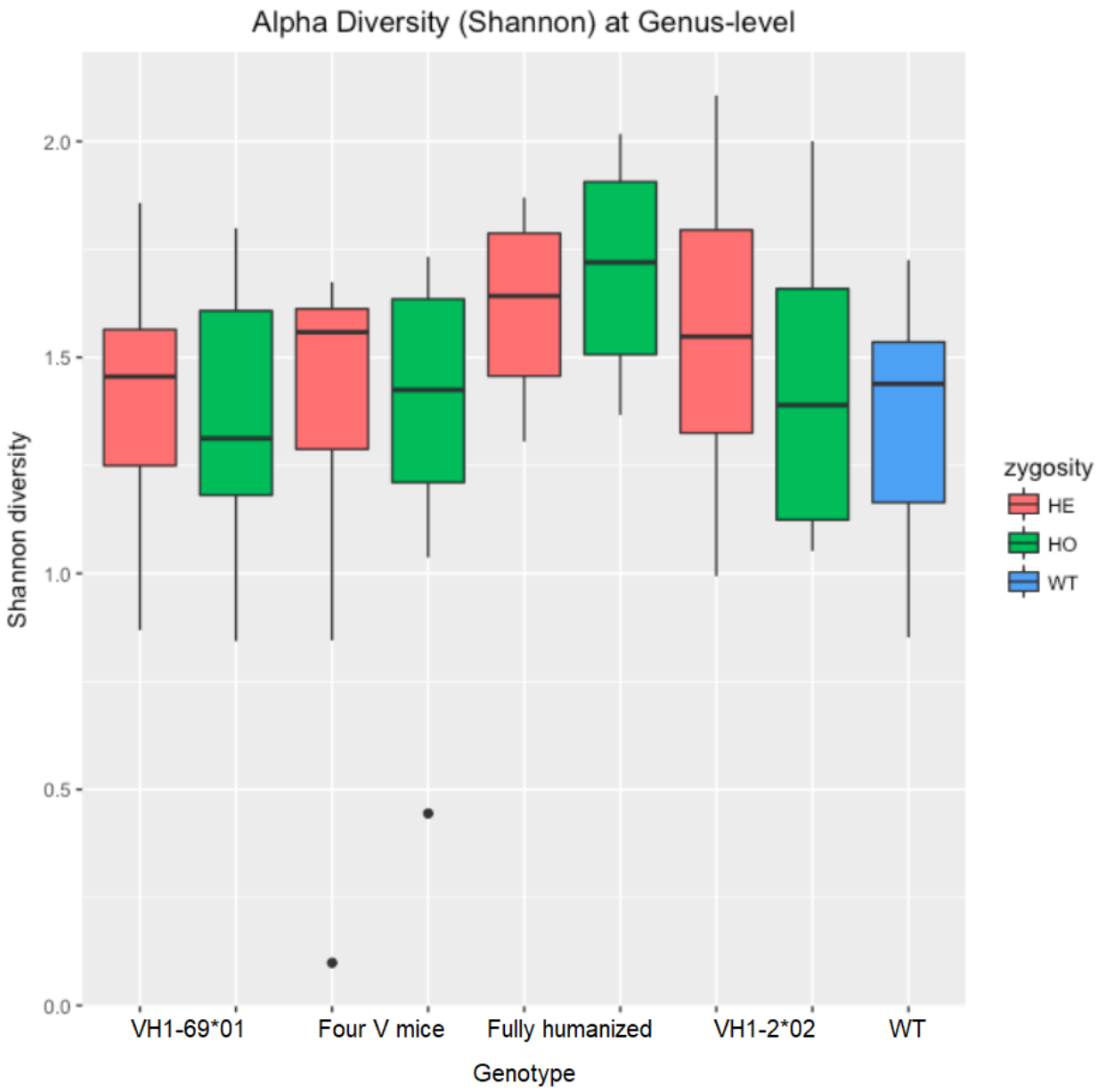

Figure A.12 | Alpha diversity indicating the richness of gut commensal diversity within each humanized mouse strain.