

Determination of B cell IgH repertoire changes after immunization and spaceflight modeling

by

Trisha Ann Rettig

B.S., Beloit College, 2008

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology  
College of Arts and Sciences

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

2018

## **Abstract**

Antibodies are an essential part of the immune system. Each B cell, a type of white blood cell, produces a unique antibody. This antibody molecule is comprised of two identical light chains and two identical heavy chains. Each chain has a variable region, which is responsible for antigen binding, and a constant region, which is responsible for effector function in the host. The variable region in the heavy chain is composed of three gene segments, the variable (V), diversity (D), and joining (J) gene segments. The light chain is composed of only V- and J-gene segments. Each immunoglobulin locus contains multiple versions of each gene segment, ranging from over 130 possible V gene segments in the heavy chain to four possible J-gene segments in both the heavy and kappa light chain. The recombination of gene segments occurs in the germline DNA and results in the formation of the unique antibody. The diversity and binding abilities of the antibodies are important for a proper and robust immunological response. Of importance to binding and specificity is the complementary determining region three (CDR3) which plays a major role in determining specificity and antibody-antigen binding. Due to its uniqueness, is used as a measure of diversity in the repertoire.

In this work, I used Illumina MiSeq 2x300nt high-throughput sequencing to assess the mouse splenic transcriptome. The work I present here shows the splenic immunoglobulin gene repertoire from unchallenged, unvaccinated conventionally housed mice, mice flown aboard the International Space Station (ISS), and mice challenged with tetanus toxoid (TT) and/or adjuvant (CpG) and subjected to skeletal unloading by antiorthostatic suspension (AOS). AOS is used to induce some of the physiological changes that parallel those that occur during space flight. The characterization of the repertoire includes analysis of V-, D-, and J-gene segment usage, constant region usage, V- and J-gene segment pairing, and CDR3 length and usage.

The work included validation of the methodology needed for tissue preparation and storage aboard the ISS, showing that the data obtained was similar to those used in standard ground-based methodologies (Chapter 2). I further validated our nonamplified sequencing methodology with comparisons to methods that use amplification as part of the process (Chapter 3). My work characterized the antibody repertoire of the conventionally housed C57BL/6J mouse (Chapter 4), an important mouse strain in the field of immunology, and demonstrated the homogeneity of gene segment usage in unchallenged animals. We also demonstrated that short duration (~21 days) space flight does not significantly alter the antibody repertoire (Chapter 5). The work culminates in an AOS study to assess changes to the B-cell immunoglobulin repertoire after vaccination with TT and/or CpG. The results show that changes to V-, D-, and J-gene segment usage occur after antigen challenge with AOS causing decreased class switching and frequency of plasma cells. Tetanus toxoid challenge decreased multiple gene segment usage and CpG administration increased isotype switching to the IgA constant region (Chapter 6).

Determination of B cell IgH repertoire changes after immunization and spaceflight modeling

by

Trisha Ann Rettig

B.S., Beloit College, 2008

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology  
College of Arts and Sciences

KANSAS STATE UNIVERSITY  
Manhattan, Kansas

2018

Approved by:

Major Professor  
Stephen K. Chapes



# **Copyright**

© Trisha Ann Rettig 2018.

## **Abstract**

Antibodies are an essential part of the immune system. Each B cell, a type of white blood cell, produces a unique antibody. This antibody molecule is comprised of two identical light chains and two identical heavy chains. Each chain has a variable region, which is responsible for antigen binding, and a constant region, which is responsible for effector function in the host. The variable region in the heavy chain is composed of three gene segments, the variable (V), diversity (D), and joining (J) gene segments. The light chain is composed of only V- and J-gene segments. Each immunoglobulin locus contains multiple versions of each gene segment, ranging from over 130 possible V gene segments in the heavy chain to four possible J-gene segments in both the heavy and kappa light chain. The recombination of gene segments occurs in the germline DNA and results in the formation of the unique antibody. The diversity and binding abilities of the antibodies are important for a proper and robust immunological response. Of importance to binding and specificity is the complementary determining region three (CDR3) which plays a major role in determining specificity and antibody-antigen binding. Due to its uniqueness, is used as a measure of diversity in the repertoire.

In this work, I used Illumina MiSeq 2x300nt high-throughput sequencing to assess the mouse splenic transcriptome. The work I present here shows the splenic immunoglobulin gene repertoire from unchallenged, unvaccinated conventionally housed mice, mice flown aboard the International Space Station (ISS), and mice challenged with tetanus toxoid (TT) and/or adjuvant (CpG) and subjected to skeletal unloading by antiorthostatic suspension (AOS). AOS is used to induce some of the physiological changes that parallel those that occur during space flight. The characterization of the repertoire includes analysis of V-, D-, and J-gene segment usage, constant region usage, V- and J-gene segment pairing, and CDR3 length and usage.

The work included validation of the methodology needed for tissue preparation and storage aboard the ISS, showing that the data obtained was similar to those used in standard ground-based methodologies (Chapter 2). I further validated our nonamplified sequencing methodology with comparisons to methods that use amplification as part of the process (Chapter 3). My work characterized the antibody repertoire of the conventionally housed C57BL/6J mouse (Chapter 4), an important mouse strain in the field of immunology, and demonstrated the homogeneity of gene segment usage in unchallenged animals. We also demonstrated that short duration (~21 days) space flight does not significantly alter the antibody repertoire (Chapter 5). The work culminates in an AOS study to assess changes to the B-cell immunoglobulin repertoire after vaccination with TT and/or CpG. The results show that changes to V-, D-, and J-gene segment usage occur after antigen challenge with AOS causing decreased class switching and frequency of plasma cells. Tetanus toxoid challenge decreased multiple gene segment usage and CpG administration increased isotype switching to the IgA constant region (Chapter 6).

# Table of Contents

List of Figures .....	xi
List of Tables .....	xiv
Acknowledgements .....	xv
Chapter 1 - Introduction.....	1
Basic Overview .....	1
B Cell Development and VDJ Recombination .....	1
Repertoire Development and Change .....	7
Repertoire Responses.....	11
Sequencing the Repertoire .....	18
Effects of Spaceflight on the Immune System.....	20
Study Objectives .....	26
Figures .....	28
Chapter 2 - Validation of methods to assess the immunoglobulin gene repertoire in tissues	
obtained from mice on the International Space Station .....	32
Abstract .....	32
Introduction.....	33
Materials and Methods.....	35
Results.....	40
Discussion .....	46
Acknowledgements.....	53
Figures and Tables .....	54
Chapter 3 - A comparison of unamplified and massively multiplexed PCR amplification for	
murine antibody repertoire sequencing .....	64
Abstract .....	64
Introduction.....	65
Materials and Methods.....	66
Results.....	69
Discussion .....	76
Figures and Tables .....	83

Chapter 4 - Characterization of the naïve murine antibody repertoire using unamplified high-throughput sequencing.....	93
Abstract.....	93
Introduction.....	94
Materials and methods.....	96
Results.....	100
Discussion.....	108
Acknowledgements.....	114
Figures and Tables.....	115
Chapter 5 - Effects of spaceflight on the immunoglobulin repertoire of unimmunized C57BL/6 mice.....	123
Abstract.....	123
Introduction.....	124
Materials and Methods.....	126
Results.....	131
Discussion.....	141
Figures and Tables.....	148
Chapter 6 - Effects of tetanus toxoid and skeletal unloading on the antibody repertoire of C57BL/6J mice.....	161
Abstract.....	161
Introduction.....	162
Materials and Methods.....	166
Results.....	170
Discussion.....	180
Figures and Tables.....	186
Chapter 7 - Conclusions.....	202
Figures.....	212
References.....	213
Appendix A - Supplemental figures.....	253
Chapter 3 Supplemental Figures.....	253
Chapter 4 Supplemental Figures.....	255

Chapter 5 Supplemental Figures .....	264
Chapter 6 Supplemental Tables .....	271
Appendix B - Standard operating procedures for repertoire analysis.....	281
Part I – Illumina Sequencing Import to CLC and Quality Cleaning .....	281
Part II – Mapping Cleaned Sequences to V Gene Segments and Locus for IMGT Submission .....	286
Part III – Submission of FASTAs to IMGT and retrieval of IMGT Results .....	290
Part IV – Determination of Constant Region (IgH Only).....	292
Part V – Reformatting of IMGT Data, Duplicate Removal, and Functionality Assignment .	296
Part VI – Characterization of Data (V, D, J, Constant, Functionality, CDR3 Length) .....	304
Sub Appendix A – Programs Required.....	311
Sub Appendix 2 - List of Resources .....	312
Sub Appendix 3 – Current Naming and Sorting Conventions for CLC .....	313
Sub Appendix 4 – Current Naming Conventions for IMGT Processing Folders.....	314
Sub Appendix 5 – Important IMGT Links .....	315
Sub Appendix 6 – Data Collected from IMGT Output .....	316
Sub Appendix 7 – Checklists for Data Processing .....	318
Sub Appendix 8 – Functional Heavy B6 Gene Segments (IMGT) .....	324
Sub Appendix 9 – Functional Heavy B6 V-Gene Segments in Chromosomal Order (NCBI Order).....	325
Sub Appendix 10 – Functional Kappa B6 Gene Segments in Chromosomal Order .....	326
Sub Appendix 11 – Functional Lambda B6 Gene Segments in Chromosomal Order .....	327
Appendix C - Copyright Releases.....	328
Chapter 3 Copyright Release .....	328
Chapter 4 Copyright Release .....	329
Chapter 5 Copyright Release .....	330

## List of Figures

Figure 1.1. Antibody Structure .....	28
Figure 1.2. VDJ Recombination .....	29
Figure 1.3. Class Switching .....	30
Figure 1.4. Repertoire Shifts .....	31
Figure 2.1. Bioinformatic analysis workflows.....	54
Figure 2.2. Decision-making matrix to remove duplicate sequence reads after IMGT processing .....	55
Figure 2.3. Top ten VH gene segments used among treatment groups. ....	56
Figure 2.4. Top ten Vk used among treatment groups.....	57
Figure 2.5. D, J, and heavy chain constant usage among treatment groups. ....	58
Figure 2.6. CDR3 AA sequence usage among treatment groups. ....	59
Figure 2.7. Correlation of V gene segments between genome and reference mapping.....	60
Figure 3.1. R <sup>2</sup> values of sequencing technical replicates.....	83
Figure 3.2. Percent of repertoire for high frequency V-gene segments among data sets. ....	84
Figure 3.3. Overlap of CDR3 sequence detection between technical replicates. ....	85
Figure 3.4. CDR3 sequence capture among Com1, Com2, and KSU data sets.....	86
Figure 3.5. High frequency CDR3s detected among the Com1, Com2, and KSU data sets. ....	87
Figure 4.1. V-gene segment usage among unimmunized mouse pools.....	115
Figure 4.2. V-gene segment usage among unimmunized mouse pools for IgH (A) and Igk (B) by chromosomal location .....	116
Figure 4.3. Percent abundance of IgH D- (A) and J- (B) gene segments, IgH constant regions (C) and Igk J-gene segments (D).....	117
Figure 4.4. Combinations of V-gene families with DJ-gene segments for IgH (A) and J-gene segments for Igk (C). ....	118
Figure 4.5. CDR3 length for IgH (A) and Igk (B).....	119
Figure 4.6. Top CDR3 AA sequences and overlap of unique CDR3 sequences within mouse pools.....	120

Figure 4.7. Comparison of CDR3 alignments in gene segment combinations (IGHV1-26, IGHD1-1, IGHJ1) coding for a predominantly short (A), median length (B), and long (C) H-CDR3 region and $\kappa$ -CDR3 (IGKV1-110, IGKJ-2) (D).	121
Figure 5.1. Expression of Top V-Gene Segments	148
Figure 5.2. Expression of Top-V $\kappa$ Gene Segments from Spleen and Liver	149
Figure 5.3. Expression of D and J-Gene segments and IgH Constant Region Usage	150
Figure 5.4. Gene Segment Combinations in Ground Control and Flight Animals	151
Figure 5.5. CDR3 Length in IgH and Igk sequences	152
Figure 5.6. Top CDR3 Usage and Overlap of CDR3 Between Treatment Animals	153
Figure 5.7. Nucleotide Alignment of CDR3 from top V-D-J Combination	154
Figure 5.8. Substitution Mutations by Ig Region	155
Figure 6.1. Impact of AOS on Serum Tetanus Toxoid-specific IgG.	186
Figure 6.2. Assessment of High Frequency V-Gene Segment Usage Among Treatment Groups.	187
Figure 6.3. Impact of AOS, Tetanus Toxoid and CpG on V, D, J and constant-gene usage. Immunoglobulin heavy chain D- (A), JH- (B), Constant Region (C) and J $\kappa$ -Gene Segment (D) Usage are presented by Treatment Group.	188
Figure 6.4. Ranking of High Frequency V-/D-/J-Gene Segment Combinations for Heavy (A) and V-/J-Kappa (B) Chain.	189
Figure 6.5. Circos Plots of V/J Combinations by Treatment Group of Immunoglobulin Heavy (A) and Kappa (B) Chain	190
Figure 6.6. Determination of CDR3 Amino Acid Length for Immunoglobulin Heavy (A) and Kappa (B) Chain.	191
Figure 6.7. Overlap of Unique CDR3 Amino Acid Sequences by Variable for Heavy (A) and Kappa (B) Chain.	192
Figure 6.8. Overlap of Unique CDR3 Amino Acid Sequences for TT and CpG Treatment Groups for Heavy (A) and Kappa (B) Chain.	193
Figure 6.9. Assessment of High Frequency VH-Gene Segment Usage in Class-Switched Antibodies.	194
Figure 6.10. Overlap of Unique CDR3 Amino Acid Sequences Overlap by Variable for Class-Switched Antibodies.	195



Figure 6.11. Unique CDR3 Amino Acid Sequence Overlap for TT and CpG Treatment Groups for Class-Switched Antibodies. ....	196
Figure 7.1. High frequency V-Gene segment usage between young C57BL/6 and aged C57BL/6Tac mice .....	212

## List of Tables

Table 2.1. Sequences used for heavy chain identification .....	61
Table 2.2. Sequencing and mapping results from the cells, tissue, and size selected treatment groups.....	62
Table 2.3. Comparison of mapping techniques in HiSeq datasets.....	63
Table 3.1. Comparison of total productive reads among data sets. ....	88
Table 3.2. Percent of reads assigned to non-C57BL/6 (B6) or pseudogene V-gene segments ....	89
Table 3.3. $R^2$ of V-gene segment usage to the KSU data set with read counts. ....	90
Table 3.4. Comparison of read count to unique CDR3 amino acid sequences.....	91
Table 3.5 Comparison of CDR3 frequencies in the whole repertoire and the unique repertoire. ....	92
Table 4.1. Sequencing and mapping statistics from mouse pools 1, 2, and 3.....	122
Table 5.1. Spleen Sequencing Read Counts in Ground (G) and Flight (F) Mice .....	156
Table 5.2. Comparison of Flight and Ground V-Gene Segment Usage .....	157
Table 5.3. V $\kappa$ Liver Sequencing Read Counts in Ground (G) and Flight (F) Mice .....	158
Table 5.4. V-J Linear Regression Analyses.....	159
Table 5.5. CDR3 Length by Isotype .....	160
Table 6.1. Average number ( $\pm$ SEM) of reads obtained per treatment group .....	197
Table 6.2. Impact of AOS, Tetanus Toxoid and CpG on lymphocyte phenotype by fold-change. ....	198
Table 6.3. Assessment of Coefficient of Determination ( $R^2$ ) of V-Gene Segments.....	199
Table 6.4. $R^2$ of V/J Pairing Correlation by Variable .....	200
Table 6.5. Shared Unique CDR3 Amino Acid Sequences.....	201

## **Acknowledgements**

I would like to thank my friends and family for supporting through my schooling to get me to where I am today. Without their support, this wouldn't be possible. Especially that of my husband who has walked this whole path with me, tolerating the ups and downs, and always willing to bring me dinner.

I would also like to thank my committee, Stephen K. Chapes, Sherry Fleming, Susan Brown, and Jodi McGill for their feedback and support.

None of this work would be possible without my lab mates. Bailey, Savannah, and Claire have poured their hearts, souls, and hours of work into finalizing all of this data, providing feedback, helping write papers, and always willing to help where it's needed.

I would also like to acknowledge our collaborators, Mike Pecaute and Nina Nishiyama who housed the AOS animals and have been ever helpful in manuscript preparations, data sharing, and troubleshooting.

# **Chapter 1 - Introduction**

## **Basic Overview**

B cells are a type of lymphocyte that produces a heterodimeric protein called an antibody. B cells were originally discovered originating from the bursa in the chicken, but in mammals are mostly derived from the bone marrow.<sup>1</sup> Ehrlich presented the side chain theory in the late 1890s laying the foundation for the theory of antibodies.<sup>2</sup> B cells play an important role in the immune system both by secreting antibodies and by functioning as antigen presenting cells to T cells. Each B cell produces a highly unique antibody which can be secreted or membrane-bound. The antibodies are composed of two identical light chains and two identical heavy chains (Figure 1.1). Antibodies and B cells have numerous functions within the immune system including the clearing of pathogens, clearing of infected or dead host cells, and assisting with T-cell activation.<sup>3</sup> Antibodies are formed with a pair of an identical heavy and light chains. These protein chains are the product of multigenic assembly of variable (V), diversity (D – heavy chain only), and joining (J) gene segments and a constant region gene (Figure 1.2).<sup>4</sup> Tonegawa and Hozumi's understanding of V(D)J recombination shattered the prevailing dogma at the time of "one gene, one protein".<sup>5</sup>

## **B Cell Development and VDJ Recombination**

B cells begin development as hematopoietic stem cells, progress to multipotent progenitors, diverge from the granulocyte/erythroid lineages, and eventually progress to lymphoid-primed multipotent progenitor cells.<sup>6-8</sup> From here, they develop into common lymphoid progenitor cells where they are more fully committed to being lymphocytes.<sup>8,9</sup> At this point, the cells will usually fully commit to the B cell lineage, progressing through the pro-B cell,

pre-B cell, immature and mature subsets. However, work by Graf et al. and Rodriguez-Fraticelli et al. show that commitments may not be final, and changes in development may occur if the proper signals are present.<sup>8,10,11</sup>

V(D)J rearrangement occurs during B cell development and successful rearrangements of heavy and light chains serve as checkpoints in the differentiation process. Rearrangement is dependent on two recombinases, RAG1 and RAG 2, which assemble and create a synaptic cleavage complex that creates nicks and breaks in the genomic DNA near recombination signal sequences (RSS). These are located on either side of the gene segments, to recombine the V-, D-, and J-gene segments.<sup>12,13</sup> RSS are either 12nt or 23 nucleotides long, forming a specific loop when the RSSs pair.<sup>13</sup> RAG1 is then able to cleave this pairing with the help of RAG2 by introducing a single stranded nick, cleaving the phosphate backbone and creating a double-stranded break.<sup>13</sup> Non-homologous end joining then repairs the DNA, creating a junction between the gene segments.<sup>12</sup> In addition to annealing the ends, additional nucleotides, called “non-templated (N) and palindromic (P)” additions are added between the gene segments using nucleotidyl transferase and DNA ligase, Artemis.<sup>14</sup> These nucleotides add additional variability to the sequence that is not dependent on genomic sequence. N nucleotides are more heavily weighted towards the addition of a G or C.<sup>14</sup>

Rearrangement begins in the early pro-B cell by the joining of the D- and J-gene segments (Figure 1.2).<sup>9</sup> Once the D- and J-gene segments are joined, the cell enters the late pro B-cell phase and combines the V-gene segment with the already recombined DJ-gene segments by the same mechanism.<sup>12</sup> The rearranged heavy chain is then paired with a surrogate light chain,

made of the VpreB and  $\lambda 5$  proteins to form the pre-B cell receptor (pre-BCR).<sup>12,15</sup> The pre-BCR is composed of the heavy chain, the surrogate light chain, Ig $\alpha$ , and Ig $\beta$ .<sup>12</sup> Failure of the heavy chain to pair with the surrogate light chains will prevent the cell from proceeding further in development.<sup>15</sup> After the cell has survived this selection, rearrangement begins with the light chain.<sup>15</sup>

Rearrangement begins as only one locus. Allelic exclusion prevents multiple V(D)J recombinations from occurring at the same time. Rearrangement in humans and mice always begins with a heavy chain locus, followed by the kappa chain and then the lambda chain. However, other species, such as pigs, rearrange the lambda chain first.<sup>16</sup> Either the maternal or paternal locus is selected for initial recombination, though there appears to be no clear reason why one is selected over the other. Feedback prevents the second locus from rearranging until the first has finished.<sup>17</sup> Stochastic models tend to focus on decreased efficiency preventing simultaneous rearrangements while determinant models suggest that the chromosomes are marked during development, which facilitates the rearrangement of one locus.<sup>17</sup> There are experiments supporting that chromosomes are marked early in development, before hematopoiesis begins, yet other experiments have shown that the choice doesn't occur until the lymphocyte has already begun to mature.<sup>17</sup> Regardless of chromosome use, allelic exclusion is modulated by multiple factors that prevent RAG from accessing RSS, thereby preventing rearrangement and even physically separating the V-gene segment DNA from the DJ-joined region of DNA.<sup>18</sup>

In the small pre-B cell phase, light-chain rearrangement begins using the same mechanics used in the heavy chain.<sup>12</sup> About half of all light-chain rearrangements are successful on the first rearrangement attempt, but rearrangements can be attempted at all four loci (two kappa, two lambda).<sup>15</sup> In mice, the kappa locus is used about five times more than the lambda locus for light chain generation.<sup>15</sup> After a successful light chain is paired with a heavy chain, the cell then expresses a complete B cell receptor (BCR) and are considered immature B cells.<sup>15</sup> There appears to be no pairing restrictions between heavy and light chains and the overall pairing of specific VH and VL gene segments is correlated with overall V-gene segment usage in the repertoire.<sup>19</sup> These BCRs are then tested against auto antigens and strong binding leads to apoptosis.<sup>15</sup>

Although Ig-gene rearrangement can introduce variability to the B cell repertoire, the process is inexact. Multiple problems can stop the Ig molecule from developing into a functional BCR. Examples include out of frame joining of the D- and J- gene segments, and further out of frame joining of the V-gene segment to the DJ segment.<sup>15</sup> Rearrangements can also introduce stop codons due to nuclease trimming of the gene segments or the addition of the N and P nucleotides.<sup>14</sup> In the C57BL/6 mouse, 7.6% of D-gene segments have 5' trimming, and 27.4% have 3' trimming.<sup>14</sup> P nucleotide additions of AAAAAA on the 5' end was the most common, and G, GT, and GTAG were the most common at the untrimmed 3' end.<sup>14</sup> In the case of the J gene segment, 12.3% of C57BL/6 mice had no exonuclease trimming, and 31.6% had P nucleotide additions ranging in size from one to eleven nucleotides.<sup>14</sup> Additionally, when N nucleotide additions were examined, the V-D junction had an average of four nucleotides, while the D-J junction averaged 2.9 additions.<sup>14</sup>

The combination of the V-, (D-), and J-gene segments comprises what is termed complementary-determining region 3 (CDR3).<sup>15</sup> Two other regions, designated CDR1 and CDR2, are coded for entirely in the V-gene segment, thus, reducing diversity possibilities compared to CDR3.<sup>15</sup> These CDR region designations are most important because they identify the parts of the antibody that bind to antigen and are also the most likely to undergo somatic hypermutation,<sup>20</sup> which will be discussed later.

The failure of a V-D-J rearrangement could induce the B cell to utilize the other allelic DNA strand or a more 5' V-gene segment maybe selected for rearrangement to form a productive heavy chain.<sup>15</sup> Approximately 15% of pro-B cells fail to make a proper rearrangement on either of the heavy chain loci.<sup>15</sup> Additionally, the heavy and light chains are bound through disulfide bonds around  $\beta$ -bulges associated with the V-gene segments and it is likely that some rearrangements fail because they do not generate the correct structure for light chain binding.<sup>15</sup> Once a B cell expresses surface heavy and light chains it has reached the “immature stage”, and the cells leave the bone marrow and traffick to the spleen. Trafficking to and within the spleen is controlled by multiple factors such as integrins (LFA-1 and  $\alpha 4\beta 1$ ),<sup>21</sup> receptors (CCR7 and CCR5),<sup>22,23</sup> chemokines (CXCL13),<sup>23</sup> and cytokines (such as IL-4).<sup>24</sup> These spleen-localized cells now enter “transitional phases” and proceed through multiple selection or “check” points.<sup>25,26</sup> After B cells leave the bone marrow, it is estimated that less than 20% of all immature B cells reach the spleen, possibly because their BCRs are autoreactive and they are removed during the migration.<sup>15</sup>



## **B1 B Cells**

B1 B cells are different than the more common B2 B cells that traffick from adult bone marrow to spleen as described above. B1a B cells are CD5<sup>+</sup> and are most common in young animals and humans.<sup>27</sup> B1b B cells are similar, but they lack the CD5 surface molecule.<sup>27</sup> In the mouse, B1 B cells comprise 30% of the spleen on day five, and about one to two percent of the spleen by week eight.<sup>27</sup> B1 B cells develop in the yolk sack, fetal liver, and neonatal bone marrow. The bone marrow loses the potential to develop new B1 B cells by about six weeks of age in mice.<sup>27,28</sup> In humans, B1 B cell precursors are detected at seven weeks in the embryo, and they populate the periphery between week 16 and 22 of gestation.<sup>29</sup> This is evidenced by data showing that, bone marrow transfers can restore the B2 population in three-month-old mice, but not the B1 population.<sup>27,30</sup> Gene segment usage is heavily restricted in the B1 population, which will be discussed in greater depth later, and most B1 cells lack nontemplated (N and P) base insertions in their joining of gene segments.<sup>27</sup> B1 B cells are commonly autoreactive and are not selected against using “traditional” B2 B cell methods because they don’t go through transitional selection processes in the spleen. Autoreactivity may even serve as a positive selection step.<sup>27</sup>

## **Constant Regions**

The constant region of the light chain is selected during VJ recombination, but the heavy chain can undergo class switch recombination (CSR), which is a different mechanism than V(D)J recombination.<sup>31-33</sup> The constant regions in the heavy chain serve different biological effector functions within the host. After VDJ recombination, B cells predominantly express the  $\mu$  constant region (IgM) and/or the  $\delta$  constant region (IgD). This is the only time that normal B cells express immunoglobulins with multiple constant regions in a single cell. Using alternative splicing mechanisms, the cell can express IgM, IgD, or IgM and IgD together.

Class switching generally occurs in response to antigen using activation-induced cytidine deaminase (AID) to deaminate a cystine in the switch region. This creates a double stranded break in the switch region immediately following VDJ recombination location and one before the newly-selected constant region (Figure 1.3).<sup>34</sup> Like in VDJ recombination, the DNA between the two breaks is spliced out.<sup>34</sup> Thus, once class switching occurs, a B cell can no longer use a constant region that has been removed from the genome, though other heavy chain genes that are 5' to the splice site are still available. The human heavy chain contains nine constant regions ( $\alpha 1$ ,  $\alpha 2$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma 1$ ,  $\gamma 2$ ,  $\gamma 3$ ,  $\gamma 4$ , and  $\mu$ ). Mice contain nine constant regions ( $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma 1$ ,  $\gamma 2a$ ,  $\gamma 2b$ ,  $\gamma 2c$ ,  $\gamma 3$ , and  $\mu$ ), though not all strains have all constant regions. CSR generally occurs after the B cell encounters antigen and is dependent on T cells and their cytokines.<sup>34-36</sup> Constant regions, even of the same type, for example the IgGs, can also affect the binding abilities of the variable region.<sup>37</sup>

## **Repertoire Development and Change**

There are multiple factors that affect the host's B-cell repertoire and give it variability. The first of these is dependent on the unique V-, D-, and J-gene segments encoded in the host's genome. The second factor is the semi-random recombination of those gene segments. The third factor is the random addition of the N- and P-nucleotides during V(D)J recombination, and the last is the somatic mutations that occur during the response to antigen and the process of affinity maturation.<sup>38</sup> This combination of events is responsible for the creation of unique antibody sequences. However, another element, the host's infection and exposure history, also plays a role in the ratio or number of B cells displaying a specific idotype (specific antigen-binding sites unique to an antibody). Recent work by Meng et al. has shown that the repertoire varies from sampling location to sampling location.<sup>39</sup>

## **Genetic and Spatial Control of Repertoire Development**

It appears that most gene segment usage is genetically controlled, rather than purely random.<sup>19,40</sup> During V(D)J recombination, gene segments are selected during recombination using RSS. Small changes to the sequences can affect recombination rates and create changes to the repertoire.<sup>41</sup> There are also spatial restrictions on recombination due to requirements of DNA looping for the acts of cleavage and blunt end rejoining. A position bias also been observed in DJ pairing in that more 3' D-gene segments pair more often with 5' J-gene segments, and vice versa.<sup>42</sup> Assessment by a few groups has found that V-gene segment usage is not random and there appears to be some skewing in the V-gene segments that are used.<sup>43-45</sup> Causes for this skewing appear to be largely genetic, and are largely attributed to the chromatin structure of the locus, which allows for availability of RSS for recombination.<sup>43,44</sup>

## **Clonal Expansion**

B cells express a B-cell receptor composed of membrane bound antibody (heavy and light chains) along with a number of molecules that are important for signal transduction: Ig $\alpha$ , Ig $\beta$ , and CD19, CD21, and CD81.<sup>3</sup> The heavy and light chains have short cytoplasmic tails and lack domains necessary for biochemical function and signal transduction. Therefore, B cells rely on Ig $\alpha$  and Ig $\beta$  to transduce signal.<sup>3</sup> Ig $\alpha$  and Ig $\beta$  are needed for B cell differentiation as well as antigen binding,<sup>3</sup> which stimulates expansion of that specific B cell, or clone. This is called “clonal expansion”. This expansion increases the total number of B cells expressing this BCR in the host and results in increased representation of that specific idotype in the repertoire as a whole (Figure 1.4A-C). These cells can also further differentiate into memory B cells or plasma cells. Plasma cells are responsible for most of the antibodies in the serum whereas memory B

cells do not spontaneously secrete antibody but serve as antigen presentation cells and can be further differentiated into antibody secreting cells.<sup>46</sup> After antigen clearance from the host, most of this effector cell population collapses due to the lack of signal but there is a persistence of higher affinity clones (memory) that may or may not have classes switched.<sup>47</sup> Plasma cells that do survive the population collapse can live in the bone marrow for up to a year in mice, even without the help of memory B cells (Hammarlund 2017, Manz 1997).<sup>48,49</sup> Memory B cells can live in the bone marrow for decades.<sup>50</sup>

### **Somatic Hypermutation**

Once a B cell is activated, it moves to germinal centers, locations within secondary lymphoid tissues, where T and B cells interact to initiate a response. This is where B cells undergo further selection and specialization to their cognate antigen, called somatic hypermutation (SHM). SHM occurs when AID targets specific hot spots in the antibody region of the genome, though off-target mutations are also detected.<sup>51</sup> These mutations target specific mutational hot spots DGYW (A/G/T, G, C/T, A/T) and its complement WRCH (A/T, A/G, C, A/C/T).<sup>52-55</sup> They also occur more often in the CDRs than in other regions of the antibody sequence, and mutations occur at a five-fold higher frequency than background.<sup>56-58</sup> AID targets the G/C pairing, deaminating the cystine to an uracil. This leads to a variety of outcomes, which can induce mutations. The first repair option is that, during replication, the incorrectly paired U/G becomes a T/A pair. The second option is that the incorrect uracil is removed and replaced by another random base and subsequent proper base pairing using short-patch base excision repair. The final option is that mismatch repair removes multiple bases around the mismatch and adds them back using error-prone polymerases, introducing more mutations.<sup>52,55</sup> There appears to be a bias in mice to repair these mutations with simple replication as opposed to using a DNA

error-repair pathway, which is commonly used in humans.<sup>54</sup> The mutations may be silent or may result in a missense mutation. Depending on the location of the mutation, a missense mutation may affect the structure of the antibody or the CDRs, which affect binding capabilities of the antibody. Insertions and deletions (indels), though rare, can also happen during somatic hypermutation, but must be in triplicate to keep the antibody sequence in frame.<sup>59,60</sup> These mutations and indels may increase affinity of the BCR to bind its antigen, or it may decrease affinity or cause the BCR to fold incorrectly all together. There is also evidence that V-gene segments tend to collect, or not collect, similar mutations across multiple hosts and that different V-gene segments acquire different mutations from each other, likely due to their DNA sequences and mutational hotspots.<sup>55</sup> Additionally, SHM in humans occurs more frequently in the IgG4 isotype compared to other IgGs.<sup>61</sup> This suggests that there is some downstream, or cis, regulation of the process.

During the host response, the subsequent rounds of SHM and the clearance of antigen lead to increased competition for antigen. The functional consequence of these events is that the B cells expressing the higher affinity BCRs are the B cells most likely to be triggered when they bind antigen. This engagement limits antigen binding to lower affinity BCRs. The B cells unable to bind antigen will suffer from “death by neglect”, leaving only the higher-affinity B cells to continue to replicate and continue to undergo SHM (Figure 1.4D-F). At the population level there is an overall increase in affinity for an antigen, though there appears to be a maximum affinity level. Poulsen *et al* found that antibody affinity for patients after a single TT vaccine was  $2.9 \times 10^{-9}$  (Kd) for TT antigens, which was at least five-fold higher than the average TT affinity from control patients immunized with other vaccines.<sup>62</sup> This group additionally showed that

binding on-rates did not improve after subsequent vaccinations, but off-rates did improve, demonstrating that these binding responses are selected through two different mechanisms.<sup>62</sup> However, maximum affinity levels vary from antigen to antigen.<sup>63</sup>

### **Overall Effects of Changes in the Repertoire**

Multiple factors affect the individual's unique B-cell repertoire. In the case of outcrossed populations, the V, D, and J gene segments present in the locus will affect gene segment recombination availability. RSS alternations will also play a role in outbred populations for initial recombination availability.<sup>64</sup> These two aspects affect the potential B cells available to respond to antigen. After the initial development of naive B cells, exposure to antigen continues to shape the repertoire. Repertoire changes can serve as a sort of "immunological history," showing what the host has encountered, and these encounters will leave a lasting mark on the repertoire.<sup>65</sup>

## **Repertoire Responses**

### **B1 B cells**

B1 B cells are predominantly encoded by V-gene segments from the VH11 or VH12 families in mice, often paired with JH1.<sup>27,66,67</sup> In humans these biases skew towards the V3 family, VH1-2, D7-27, JH2, and JH 3 and appear to be highly restricted.<sup>29,68</sup> They predominantly use the IgM isotype, but IgA and IgG3 natural antibodies are also detected.<sup>69</sup> B1 B cells also lack N-region insertions during V(D)J recombination in young animals, but in elderly animals, an increasing number of N-nucleotides can be detected in the peritoneal B1 cells.<sup>27,67</sup> The lower levels of N-nucleotides bias towards shorter CDR3s in the heavy chain.<sup>68</sup> B1 B cells develop more quickly and can populate the peripheral lymphoid tissues more quickly than their B2

cousins, potentially providing an advantage in young animals for clearance of apoptotic cells.<sup>68</sup> B1 B cells can bind a number of self-antigens, as well as foreign antigens, including phosphatidylcholine (PtC), a protein expressed on red blood cells.<sup>27,29,68</sup> In tracking these anti-PtC specific cells, it was determined that B1 B cells undergo expansion in response to their antigen, even in germ-free animals.<sup>27,67</sup> It has even been shown that during expansion, both class switching and somatic hypermutation can occur and is cumulative with age, even in germ free conditions.<sup>67</sup> Therefore, B1 B cells have the same molecular mechanisms available for these processes but are somehow regulated to suppress them. As age increases, the repertoire also becomes increasingly restricted and less random.<sup>67</sup> Additionally, as age increases, the B1 B cell response to antigens remains the same or increases, whereas B2 B cell antibody response decreases, biasing the repertoire towards a more monoclonal, and self-reactive repertoire.<sup>70</sup>

## **Vaccination**

The overall goal of vaccination is to create an antigen-specific memory response from B cells, protecting the patient from the expansion of pathogenic challenges (bacteria, viruses, etc.).<sup>71</sup> Thus, repertoire responses have been characterized for several antigens.<sup>72,73</sup> Some vaccine challenges reveal that specific epitopes have selection biases for specific V-gene families; such as the VH5 family selection in an HIV immunization in Macaques or the usage of VH5-51 and VH3-7 in humans following an influenza vaccination. Some antigens show no bias in the V-gene family they induce, such as tetanus toxoid (TT) vaccinations, outlined below.<sup>62,72,74,75</sup> Additionally, responses to vaccinations can vary strongly from subject-to-subject; ranging from more monoclonal responses to more polyclonal response.<sup>74</sup> Public clones, or those shared among multiple individuals, are rare, even in response to the same antigen, and, when examining the response to an influenza vaccine, the IgM repertoires were more similar among subjects than the

IgG-expressing cells (“antigen experienced”), though Greiff et al. recently demonstrated that there is more similarity among plasma cells in those responding to antigenic challenge.<sup>19,44,76</sup> Dose also seems to play a role in repertoire diversity. Higher doses induce a more diverse repertoire and may result in a broader repertoire response.<sup>77</sup> Other hallmarks of immune responses include shorter CDR3s and higher levels of mutations.<sup>40</sup>

Aging has been shown to play a large role in vaccine responses. Those over the age of 65 are less protected by the influenza vaccine, and with reduced responses to the pneumococcal vaccine.<sup>76,78,79</sup> De Bourcy *et al.* discovered that in half of elderly patients, their repertoire was comprised of a small number of memory cell, and their repertoires were less plastic.<sup>79</sup> The aged repertoire also had a smaller naïve repertoire, providing less diversity to allow for binding to a new antigenic challenge.<sup>78,79</sup> Additional skewing occurs in gene segment usage with VH3 family members being more common in young patients, and V4 more common in the elderly.<sup>80</sup> After vaccination with the pneumococcal polysaccharide vaccine, changes seen to CDR3 size and amino acid composition were less drastic in the elderly as opposed to young cohorts and the IgA and IgM responses were impaired.<sup>78</sup> While both groups showed cellular changes in the peripheral blood at day seven, the younger cohort returned to their pre-vaccination normal values, and the elderly group did not.<sup>78</sup>

### *Tetanus Toxoid*

The tetanus toxoid vaccine has long been used to study the immune response<sup>81-83</sup> and serum antibody level responses continue to be studied.<sup>84-86</sup> In addition to being a commonly used vaccine, it is considered one of the most immunogenic vaccines.<sup>62</sup> It has also been used as a conjugate protein for vaccine development.<sup>87</sup>



The tetanus toxin is produced by *Clostridium tetani* and is a 150 kDa protein which is cleaved to form a heavy and light chain.<sup>88</sup> The light chain is a metalloprotease and is responsible for the disruption of neurotransmitters.<sup>88</sup> The toxin works by blocking the release of neurotransmitters from the prejunctional membranes of nerve cells causing spastic paralysis by preventing formations of the SNARE complex.<sup>88,89</sup> The heavy chain is also cleaved into two parts, the N-terminal is used to help translocate the light chain, and the C-terminal binds to sensitive cells.<sup>88</sup> For use in vaccination, the toxin is denatured with formaldehyde making it a toxoid.<sup>90</sup>

Previous studies of TT vaccination examined both the overall levels of anti-TT Abs in the serum<sup>90</sup> and the makeup and binding of those Abs.<sup>91</sup> The half-life of anti-TT Abs is 11 years.<sup>85</sup> While children show high levels of anti-TT Abs in the serum,<sup>86</sup> a decrease in anti-TT Abs in the serum occurs in those over 40; likely due to a decrease in vaccinations.<sup>90</sup>

Studies based on the gene-segment usage of anti-TT Abs have determined that Abs generated after challenge are widely varied in both heavy and light chains. This is likely due to the large size of the TT.<sup>71,82,92-95</sup> The ability of anti-TT Abs that do not bind to the C-terminus of the heavy chain to neutralize the toxin is currently unknown, though multiple anti-TT Abs have been shown to synergize to increase neutralization.<sup>82,96</sup> Paulsen et al. and Lavinder et al. suggest that the anti-TT Ab repertoire consists of between 50-100 different original clones and that these clones encompass gene segments throughout the entire repertoire.<sup>62,96-98</sup> However, nine months after boosting, only three anti-TT Ab clones were still detected, which shows that the repertoire returns to a “normal” state post vaccination.<sup>99</sup> When the isotypes of the antibodies were

examined, the response was predominantly IgG, though IgA and IgM have also been detected.<sup>100</sup> Meijer et al. demonstrated that many of the heavy chains of anti-TT Abs can pair with multiple light chains and still retain their anti-TT specificity.<sup>91</sup> These data suggest that the heavy chain dominates the antigen-binding process.

Two human subjects and their responses to TT were followed through multiple challenges. New clones were detected after the second booster, suggesting that new anti-TT Abs can be added after repeat exposure as naive cells continue to be introduced in the host.<sup>62</sup> Similar results have been seen in other vaccination studies and may provide a way to deal with antigenic shifts in the pathogen.<sup>62</sup> While the possibility of not detecting these clones in initial sequences is possible, it is highly unlikely due to the selection of TT specific antibodies and single cell sequencing.<sup>62</sup> They also studied the affinity of the anti-TT antibodies and determined that peak affinity was obtained after the first booster and peak mutations levels were achieved after a single vaccination.<sup>62</sup> Mutation rates for anti-TT plasma cells and memory B cells were around 10%, though anti-TT Abs that could bind the heavy chain C-terminus contained higher mutation rates.<sup>71</sup> CDR3 of the heavy chain was determined to have the largest role in binding to TT, and CDR1 of the light chain also played a role.<sup>89</sup>

In examining murine specific responses to TT, four heavy chain CDR3s were detected (CARGGNYAYW, CARQDRYGFALDYW, CARLGYDGVALDYW, CARDFQYGNFYFDYW) and five kappa chain CDR3s (CQQGSSIPRLTF, CQQGNTLPWTF, CQQGNTFPWTF, CQQGSSIPRLTF, CQQGSIPRLTF).<sup>98</sup> The light chains were able to pair with multiple heavy chains.<sup>98</sup> Gene families in the heavy chain included V1, V2, V3, and V5

with D-gene families D1 and D2, combined with all four J-gene segments.<sup>98</sup> The kappa chain used V-gene families V1, V4, V5, V6, V8, V9, V10, and V12 and all four J-gene segments.<sup>98</sup> Over 20 binding epitopes exist on the tetanus toxoid molecule, and binding to nine of these is known to cause neutralization and are located on the heavy chain of the toxin.<sup>82</sup>

## **Adjuvants**

Adjuvants are added to vaccines to help increase the immune response to antigenic challenge.<sup>101,102</sup> Aluminum salts were first used as adjuvants in 1926, followed by water-in-oil emulsions in 1937.<sup>102</sup> They can decrease the dosage required and increase the speed of the immune response.<sup>102</sup> They are also implicated in diversifying the antibody response to vaccination.<sup>101</sup> There are multiple classes of adjuvants, nonimmunostimulatory, which provide increased display of the adjuvant to the immune system, and immunostimulatory adjuvants which are directly targeted to specific parts of the immune system to increase response, and adjuvants that do both.<sup>102</sup> Immunostimulatory molecules often target Toll Like Receptors, which can bias the immune response towards specific T-cell responses, while others, such as emulsions increase immune cell recruitment to the area and antigen display.<sup>102</sup> In our studies, we use CpG, an adjuvant that binds to Toll Like Receptor 9 (TLR9). TLR9 is an intracellular, nucleic acid sensing receptor that signals through the MYD88 pathway, leading to activation of transcription factors such as NF- $\kappa$ B.<sup>103</sup> This leads to the expression of pro-inflammatory cytokines, increased production of antibodies, a T<sub>H</sub>1 based response, isotype switching, the development of marginal zone B cells, and memory B cell differentiation.<sup>102-106</sup> CpG has also been shown to play a role in promoting antigen presentation by dendritic cells, which also leads to an increased antibody response.<sup>107</sup> It has also been shown that immunizations with CpG demonstrate affinity

maturation in the germinal centers up to four months after immunization, further increasing antigen affinity past the initial vaccination.<sup>108</sup>

## **B Cell Cancers**

While infections are one of the most common ways we think about the changes to the repertoire, B cell cancers are also associated with overall changes. Like proper immune responses, B cell cancers also have associations with specific gene-segment usages and mutation developments. There are a variety of malignancies that develop from antibody producing cells at different time points in development.<sup>109</sup> Some of these cancers have strong associations with specific gene segments such as chronic lymphocytic leukemia (CLL) (VH3-21, VH1-69 and VL2-14), and multiple myeloma (VH3-74 and VH3-7).<sup>80,109</sup> Myeloma also has a strong negative association with gene segments. For example, a very common V-gene segment in humans, V4-34, has not been detected in cancerous cells, even though the gene segment is commonly used in the repertoire.<sup>110</sup> Myelomas appear to be very heterogenous in their gene segment usage and do not inhibit normal plasma cells.<sup>109,111</sup>

Myelomas also continue to undergo somatic hypermutation on both the heavy and light chain after development, suggesting that chronic activation or antigen selection may continue to affect the malignancy.<sup>109,110,112</sup> Many of these mutations are radical, rather than similar amino acid replacements in the framework region.<sup>111</sup> Follicular lymphomas also appear to be heavily mutated while only half of CLL patients exhibit mutated sequences.<sup>80,113</sup> Gastric MALT lymphomas are highly linked to *Helicobacter pylori* infections.<sup>80</sup> There is even evidence that gene segment usage by CLL is regionally diverse, suggesting antigen-driven selection.<sup>114</sup> In addition

to antigen selection, acute lymphoblastic leukemias continue to diversify by maintaining the same DJ joining region but rearranging with more 5' V-gene segments.<sup>115</sup>

The most useful part of repertoire sequencing for cancer however comes from minimal residual disease testing (MRD). Proof of concept for MRD has been established for multiple B cell cancers.<sup>116-118</sup> MRD works by using PCR to test for known B cell cancer antibody sequences in the patient.<sup>116</sup> MRD negativity is correlated with a decreased risk of relapse, allowing doctors to make more informed decisions regarding treatment.<sup>116</sup> Mutation detection in some cancers are also useful as prognostic indicators.<sup>118</sup>

## **Sequencing the Repertoire**

### **Experimental Design**

As costs of sequencing have decreased over time, the ability and interest in sequencing whole B cell and immunoglobulin repertoires has increased. With this comes additional experimental design considerations to carefully weigh advantages and disadvantages of specific methods and technologies.<sup>119</sup> Starting material, such as DNA or RNA, must be chosen, as well as the target of the sequencing. DNA sequencing allows for a direct count of the number of CDR3s, whereas RNA provides a more representative sample of transcripts being made and reflects the “active” repertoire.<sup>120</sup> Additional considerations include the importance of the various structures of the antibody. Documentation of CDR3 sequence is very important for repertoire analysis, but information about which V-gene segment(s) or constant region(s) are being used also provide clues about the repertoire and the host response (Greiff 2015).<sup>121</sup> Depth of sequencing is also an important factor to consider. While 454 sequencing provided long read lengths (>400 bp), more-recent Illumina methodology has a higher read number (captures more sequences). This greater

sequencing depth is offset by a higher error frequency, however, which is a unique issue for antibody repertoires since mutations are part of the B cell differentiation process and determining sequencing errors vs. somatic mutations is problematic.<sup>122,123</sup> Questions concerning the detection of a specific CDR3 with a known sequence may benefit from amplification-based DNA sequencing with cheaper Sanger methods, while characterization of an entire repertoire, such as the work presented here, requires the use of a high-throughput technology. When planning an experiment however, it is important to remember that a single sequencing is a snapshot in time and fails to adequately monitor the constant changes in the repertoire.<sup>124</sup> Additionally, peripheral blood, while often the only source of B cells for humans, fails to provide an adequate measure of the total immune repertoire contained in various immunological organs and niches.<sup>39,125</sup>

## **Amplification**

To our knowledge, all repertoire studies for B and T cells, except those outlined here, have been the result of an amplification process.<sup>44,123,126-129</sup> These datasets contain both bias and PCR artifacts, leaving their results difficult to interpret or compare from study to study.<sup>123</sup> Thus, the avoidance of biases from amplification is one of the highest priorities in repertoire sequencing.<sup>130</sup>

PCR errors are accumulated through the amplification process which can falsely inflate members of the repertoire, or they can add suspected mutations that do not exist<sup>123,124,131,132</sup> and may not be distinguishable from low level mutations that do exist in the repertoire.<sup>132,133</sup> PCR biases can be introduced because of primer binding properties, CG content, mispriming, non-specific binding, and errors during replication.<sup>130,134-136</sup> A specific issue in targeting antibody gene segments is primer-annealing efficiencies, since the gene segments that make up the murine

IgH locus are similar, though not identical.<sup>135</sup> The biases inherent in the multiplex PCR can lead to false repertoire skewing, gene frequency inaccuracies, and a less comprehensive view of the repertoire.<sup>118,122,124</sup> The development of these multiplex primer, is highly challenging,<sup>120</sup> although Carlson et al. argue that amplification methodologies can capture the entire repertoire, there are concerns.<sup>126</sup>

The issues with amplification are well known, specifically, the inability to amplify antibodies produced by hybridomas.<sup>137,138</sup> “Universal” primers for the human antibody repertoire do exist, but some questions remain if they cover the entire repertoire.<sup>139</sup> The difficulty in developing a “universal” or even highly comprehensive primer set for the mouse is likely due to their highly varied leader sequences, V-gene segments, and framework regions. Primer design would have to rely on massively multiplexed reactions and/or degenerate primers. Additionally, the most commercially viable amplification methods would need to amplify across multiple common strains adding additional levels of complexity. Methods to overcome the biases detected in amplification have been developed, such as the use of 5’ RACE<sup>119,133,140</sup> and the use of molecular barcodes or identifiers.<sup>141,142</sup> However, these methods are expensive and have their own draw backs. Replication of the entire repertoire using 5’RACE would still require the use of multiple constant region primers, leading to the same multiplexing issues. Barcoding can have errors and chimeric reads making repertoires difficult to reconstruct.<sup>143</sup>

## **Effects of Spaceflight on the Immune System**

### **Microbes in Space**

Microorganisms were found throughout the International Space Station (ISS) and Mir.<sup>144-</sup>

<sup>147</sup> These organisms are predominated by members of the human mucosal microbiome, however

over 35 different bacteria species and 30 fungal species have been identified on the surfaces and in the air aboard the ISS.<sup>147</sup> These species include opportunistic pathogens such as *Staphylococcus aureus*.<sup>147</sup> The crew are likely the largest reservoirs on ISS, and changes to their microbiomes, both with reduced number of species and transfers between crew members, can be observed in the confinement period prior to and after launch.<sup>148-150</sup>

Space flight and simulated microgravity have been shown to increase virulence in *Salmonella typhimurium* displaying both genetic changes and increased virulence with infection when assayed post flight.<sup>151-153</sup> *S. typhimurium* also showed a reduced sensitivity to acidic pH when grown in a microgravity HARV model.<sup>151</sup> However, increased virulence is not universal because *Streptococcus pneumoniae* did not show increased virulence when tested after space flight.<sup>154</sup> Other bacterial species such as *Staphylococcus aureus* and *Escherichia coli* showed an increase in the minimum inhibitory concentrations of antibiotics on short term flights, but long-term flights (four months) showed that *E. coli*, *Bacillus subtilis*, and *Pseudomonas ariginosa* generally became resistant to the 12 tested antibiotics, but *S. aureus* occasionally became more resistant.<sup>149,154</sup> Additionally, many bacterial strains also grew faster in their planktonic state in microgravity and have increased biomass.<sup>154,155</sup> While agar-based studies didn't show the same increased final cell count, they did show an increased number of cells attached as biofilms.<sup>153,154</sup> Therefore, the space environment may alter bacteria and change their virulence and/or pathogenicity while there are concomitant changes to host defenses. Microbial infections were not uncommon aboard Mir, ranging from eye infections to respiratory infections (Taylor 2005).<sup>149</sup> During the Apollo missions, over half of the astronauts reported infections during or



immediately after their missions.<sup>156,157</sup> Other changes to the immune system are evidenced in delayed wound healing and slowed type IV hypersensitivity reactions.<sup>158-161</sup>

## **Ground Models**

Multiple, significant physiological changes occur during spaceflight. Due to the costs, space, and time limitations, it was useful to use ground-based simulations to dissect the effects of spaceflight on host physiology. While the most studied responses tend to be fluid shifts and stress (cortisol) which reflect the new exposure to unloading, other environmental factors such as barometric pressure and exposure to radiation also change in space.<sup>162</sup> Mice housed to control barometric pressure to simulating that aboard ISS showed increased susceptibility to mengovirus infections compared to mice kept at normal atmospheric pressure.<sup>156</sup>

Some immunological changes that occur during spaceflight have been consistently observed through multiple studies, such as neutrophilia (post flight), and decreased numbers of CD4 and CD8 T cells. Other host responses are changed inconsistently. For example, the number of monocytes, CD4/CD8 ratios, and B and NK cell percentages may or may not be altered.<sup>163,164</sup> Therefore, there is clearly an environmental impact that can vary from individual-to-individual and from flight-to-flight. Additionally, astronauts tend to reduce their dietary intake while in space affecting metabolism.<sup>156,162,163</sup> Due to these differences, ground models offer some ability to control some of the multiple variables that influence the host physiology during space flight. Unfortunately, since there is an impact on so many physiological systems, it is unlikely that a single perfect ground-based model exists. It is important for scientists to know the limitations of their models and not over-interpret their data.

To model some of the effects of space flight, one ground-based model that is routinely used is antiorthostatic suspension (AOS) (also called hind limb unloading or skeletal unloading). This technique was originally developed by Morey and was further refined.<sup>164,165</sup> This technique unloads the hind limbs of mice or rats through various means to prevent weight bearing on the rear limbs while maintaining a 30-to-50% weight load on the front limbs.<sup>165</sup> Animals are still able to move at will using a grid on the floor, which reduces stress levels.<sup>165</sup> Many of the effects seen in spaceflight are modeled in AOS animals such as lowered bone mass in unloaded limbs, fluid shifts, muscle atrophy, increased levels of corticosterone, and altered cellular responses.<sup>166-</sup>

170

Bed rest with a 6° head down tilt is another common physiological microgravity model used with human subjects and can model some of the impacts seen in space flight such as the decrease in lymphocyte proliferation and decreased IL-6 production. However, in other cases, the studies disagree.<sup>163,171</sup> For example, cortisol levels increased in spaceflight, but decreased in bed rest.<sup>163</sup> Cortisol measurements from bed rest are inconsistent, and some argue that cortisol levels do not play a role in muscle remodeling in spaceflight.<sup>172</sup> Indeed, some artificially elevated corticosterone levels to look at the impact on host responses and they failed to see an impact.<sup>173</sup>

### **Immune System and Host Physiology**

General issues associated with spaceflight and the immune system include immune dysregulation, thymus involution (caused by increased glucocorticoids), decreased lymphocyte development, alterations in cytokine production, and changes to leukocyte distribution.<sup>174-176</sup> Human progenitors cultured during spaceflight, showed decreased erythropoiesis and

myelopoiesis after only 13 days of spaceflight and AOS rats showed decreased ability to respond to colony stimulating factors.<sup>156,177</sup>

Post-flight counts of peripheral blood showed decreased total lymphocyte number and poor response from T cells to mitogen stimulation after space shuttle flights and in AOS animals.<sup>165,178-183</sup> Mice in 2G hypergravity also showed a decreased response to ConA.<sup>184</sup> AOS rats also had a decreased production of superoxide and showed decreased levels of killing phagocytosed bacteria in macrophages.<sup>156,185</sup> Flight also suppressed the differentiation of these cells.<sup>161</sup>

Humans, AOS rats, and AOS mice showed decreased production of multiple cytokines including IFN- $\alpha$ , IFN- $\beta$ , CSF-1, IL-1, IL-2, TNF- $\alpha$ , M-CSF, IFN- $\gamma$  and IL-6.<sup>156,157,171,174,175,180,186</sup> A human Antarctica model, another ground based model for high-stress environments, showed an increase in pro-inflammatory cytokines such as IFN- $\gamma$  and IL-12 and a reaction in anti-inflammatory cytokines such as IL-10 and IL-1ra.<sup>156</sup> Elevation in astronauts was seen in IL-8, CCL2, CCLA, and CXCL5 during flights.<sup>171</sup> Administration of IL-2 or M-CSF to compensate for their reduced secretion appears to reverse some of the immune effects of AOS.<sup>180,186</sup>

## **B Cell Responses**

Prolonged exposure to space flight and microgravity increases bone resorption and decreases bone formation.<sup>177,187</sup> AOS caused both bone remodeling and an immediate decrease in B cell development.<sup>177</sup> B and T cells develop in specific niches in the bone marrow, and changes to bone structure are common after prolonged exposure to microgravity.<sup>177</sup> The B cell progenitor population (CLP and beyond) remained low throughout the 21 days of suspension.<sup>177</sup> However,

Armstrong *et al.* did not see a change in B cell hematopoiesis.<sup>186</sup> Gainger *et al.* reported that while normalized spleen mass stayed the same for AOS mice, the total number of cells was reduced by 25% and saw decreases in CD19<sup>+</sup> B cells with an increase in CD3<sup>+</sup> T cells, altering the B to T cell ratio.<sup>176</sup> Gridley *et al.* also detected decreases in T cells, but the decreases were even between CD4<sup>+</sup> and CD8<sup>+</sup> cells, not altering the ratios.<sup>175</sup> Ortega *et al.* detected a change in the number of differentiated cells in the bone marrow, with a shift towards more differentiation, rather than less mature cells.<sup>188</sup> AOS mice and space flight rats have lymphoid organ dependent inhibition in proliferation of lymphocytes, suggesting that lymphocyte depression is not a whole-body suppression, but rather tissue specific.<sup>161,165,189</sup> There is also some environmental impact on bone changes that are dependent on housing.<sup>190</sup> Rats housed individually lost more bone mass than group-housed rats.<sup>190</sup>

Many of the changes exhibited by AOS animals are like those seen in aging.<sup>177</sup> However, in AOS rats, though involution of the thymus was detected, no change in total antibody production was detected.<sup>156,191</sup> However, mice infected with *P. aeruginosa* showed a delayed protective antibody response.<sup>171</sup> AOS mice also showed a decreased resistance to infection with *Encephalomyocarditis* virus D and increased infections by *Klebsiella pneumoniae* and *Pasturella pneumotropica*.<sup>161,192,193</sup>

There is some indication that space flight affects the immunoglobulin gene rearrangement process. The expression of the transcription factor Ikaros, responsible for expression of the heavy chain, changed when *Pleurodeles waltl* embryos, a type of newt, were subjected to microgravity.<sup>177</sup> Spaceflight also increased their levels of IgY (like IgA in mammals), and

changes were detected in V-gene segment usage.<sup>177,191</sup> Total antibody levels in astronauts remains unchanged after flight, but shifts in isotypes were seen.<sup>177,194</sup> Isotype shifts away from the class switched IgG1, IgG2a, and IgG2b were seen in mice exposed to constant low dose radiation.<sup>191</sup>

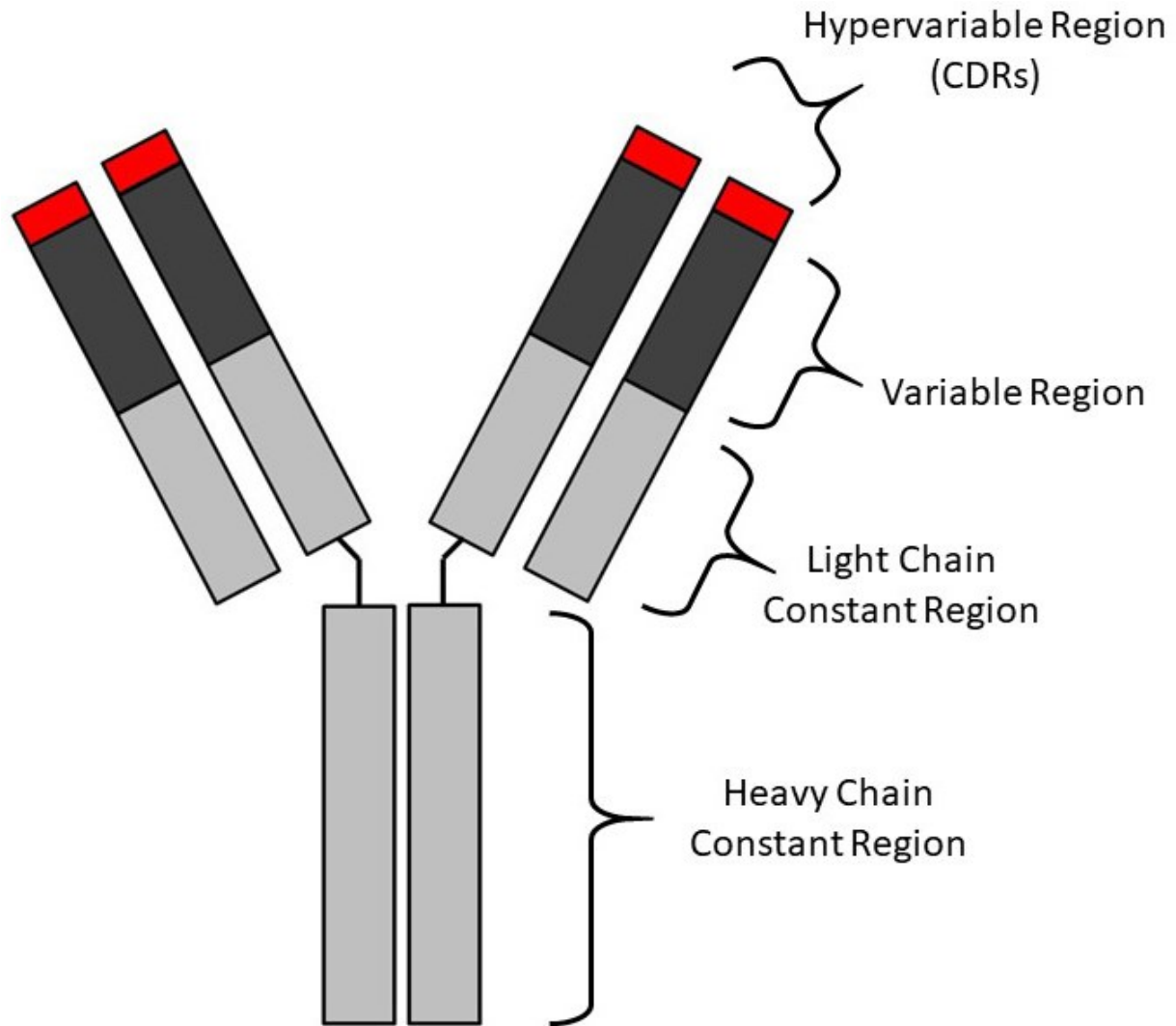
## Study Objectives

The goal of these studies is to characterize the antibody response to vaccination in animals exposed to microgravity. Specifically, we use bioinformatic analysis of RNA sequencing to determine the effects of space flight and immunization on the B-cell repertoire. To perform this work, we first developed and validated a workflow that would allow us to collect RNA from whole spleen and process it for our specific application. Due to the limitations of the International Space Station, we would not be able to process single cell suspensions, as most repertoire studies do. Samples would come back as whole frozen tissues preserved in RNALater. We confirmed that whole tissue sequencing did not result in significant differences from the cell sequencing.<sup>45</sup> We also validated our data processing workflow by comparing our data to that of ground-based research and found similar results.<sup>45</sup> As part of the validation process, we were able to characterize the repertoire of the normal mouse. These data help provide a more uniform background to determine the changes from “normal” for our continuing studies<sup>195</sup> on mice flown to the ISS<sup>58</sup> and mice subjected to AOS (this thesis). This work demonstrated the uniformity of the normal B cell/immunoglobulin repertoire and complements work done by others.<sup>14,44,56</sup> Additionally, to further supplement our data collection, we examined the impact of amplification on repertoire analysis. We found that using massively multiplexed PCR yielded deeper sequencing (more CDR3s and sequences detected) but failed to detect the full breadth of V-gene segments that unamplified sequencing does (Rettig, this thesis). The thesis culminates with our

studies analyzing the effects of hindlimb unloading on tetanus toxoid challenge with and without the B-cell adjuvant CpG.

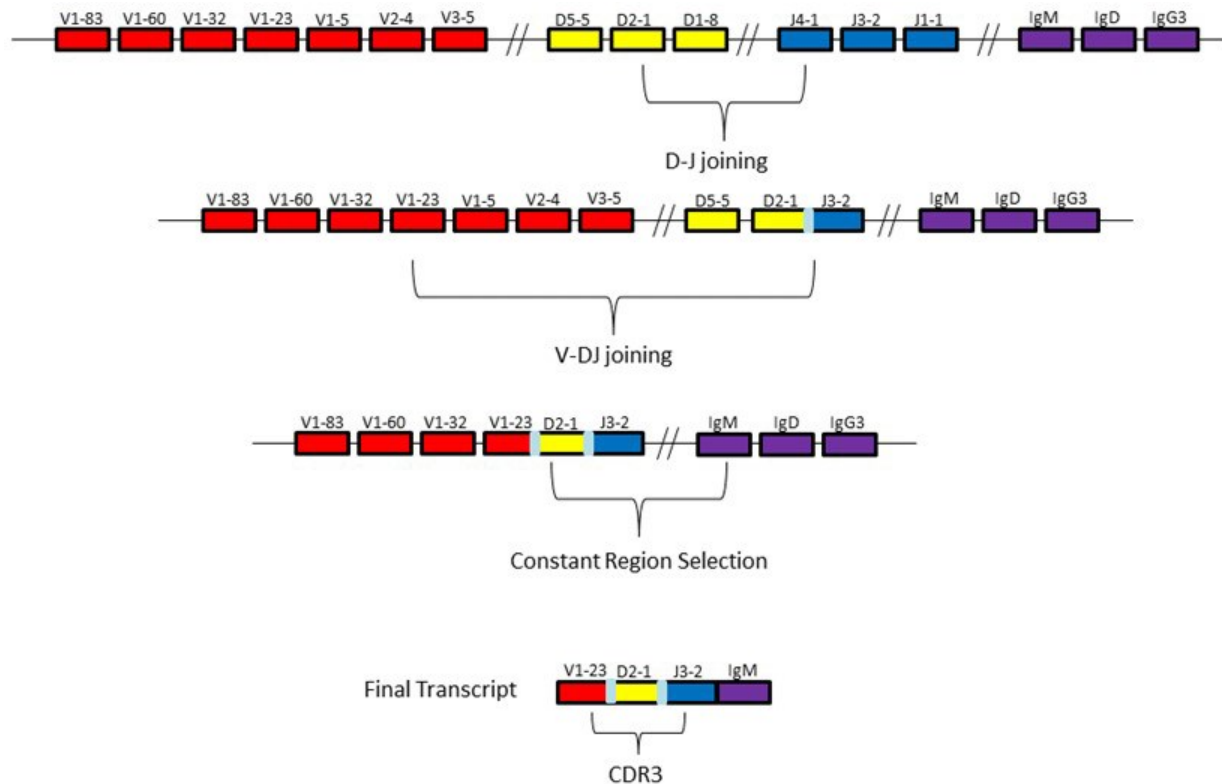
## Figures

**Figure 1.1. Antibody Structure**



Antibodies are composed of two identical heavy chains and two identical light chains. Each chain contains a constant region and a variable region. A hypervariable region contains complementarity determining regions (CDRs) and is responsible for antigen binding.

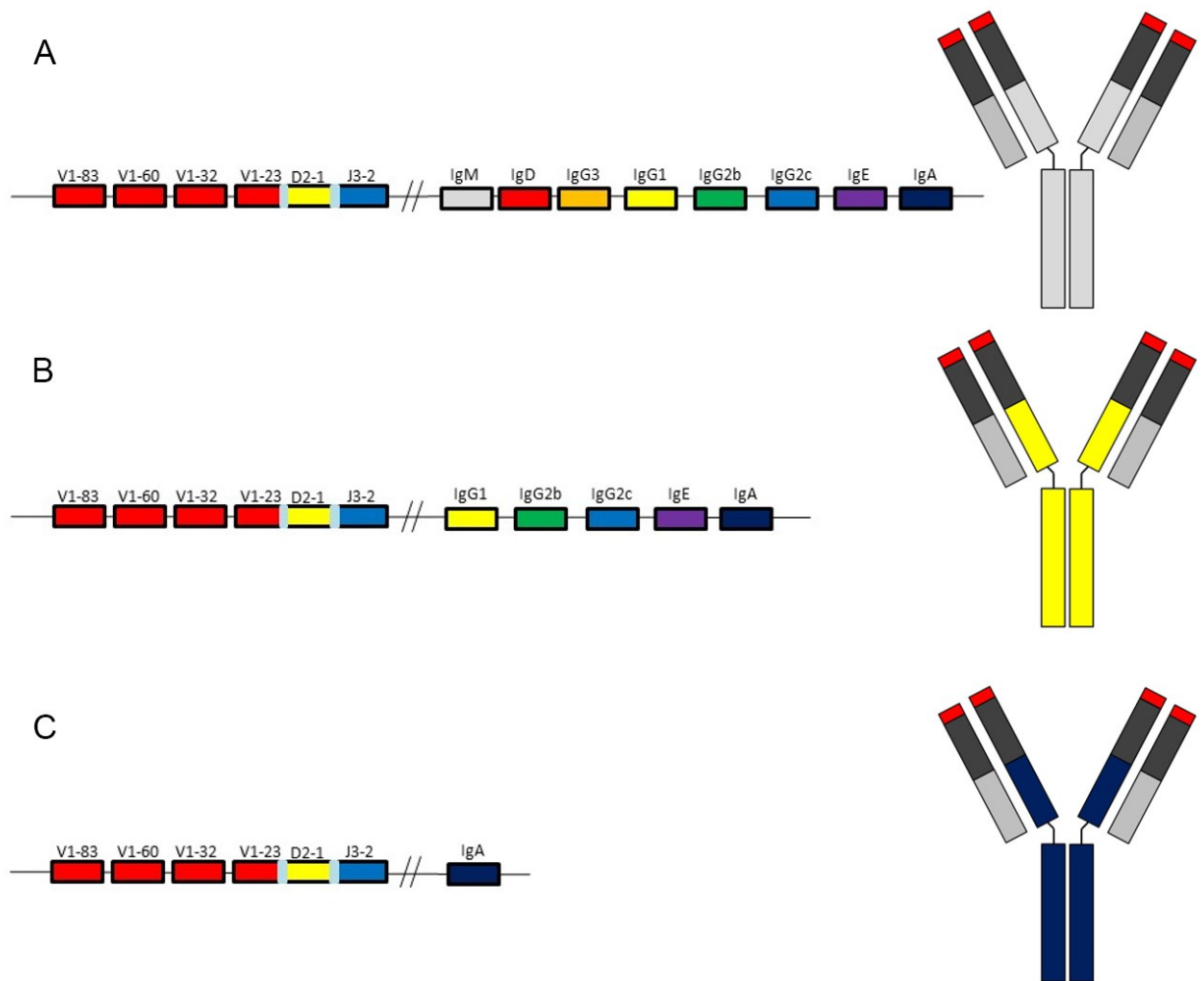
**Figure 1.2. VDJ Recombination**



VDJ recombination occurs in the germline, first by initiating recombination of a semi-random D- and J-gene segment. The intervening DNA is spliced out. Recombination then proceeds with a semi-random V-gene segment joining to the DJ rearrangement. N and P nucleotides are added between each gene segment recombination (light blue bars) which add additional variability. A constant region is then selected (IgM or IgD for initial recombination).

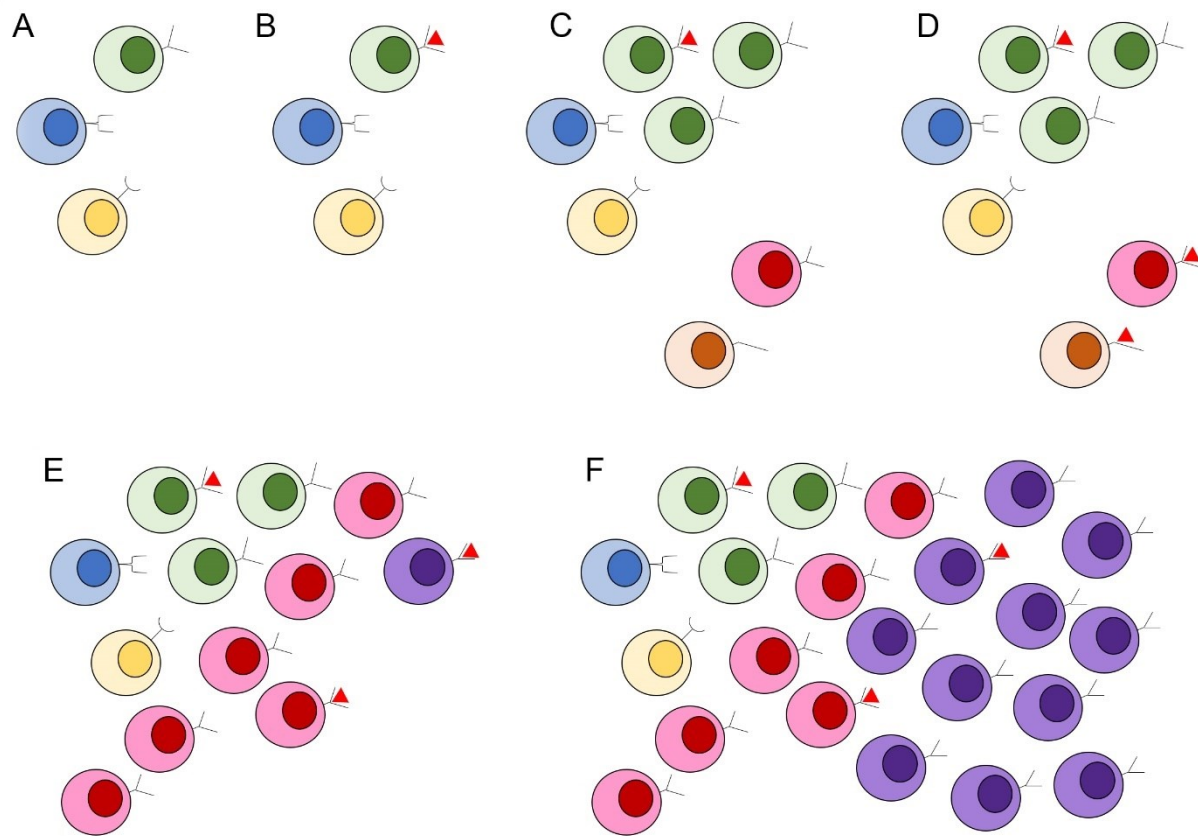


**Figure 1.3. Class Switching**



(A) Initial rearrangement results in an IgM or IgD isotype antibody, which is regulated by alternative splicing. (B) Class switching occurs by selecting a new constant region in response to cytokine messages and results in selection of a new isotype. The intervening DNA is removed from the germline. (C) An antibody can continue to class switch to a remaining constant region if proper signals are received.

**Figure 1.4. Repertoire Shifts**



(A) The naïve antibody repertoire contains a variety of different BCRs created from different heavy and light chain rearrangements. (B) Antigen (red triangle) presents and binds to one specific BCR triggering amplification of that B cell, which may undergo mutations due to SHM (C). Binding of that antigen to the newly mutated BCRs (D) will lead to apoptosis in the mutated BCRs that fail to bind with high affinity (orange cell) and (E) amplification in those that bind with higher affinity (red cell). This cell will also undergo mutation (F) and higher affinity BCRs will continue to multiply leading to changes in the repertoire.

## **Chapter 2 - Validation of methods to assess the immunoglobulin gene repertoire in tissues obtained from mice on the International Space Station**

### **Abstract**

Spaceflight is known to affect immune cell populations. In particular, splenic B cell numbers decrease during spaceflight and in ground-based physiological models. Although antibody isotype changes have been assessed during and after space flight, an extensive characterization of the impact of spaceflight on antibody composition has not been conducted in mice. Next Generation Sequencing and bioinformatic tools are now available to assess antibody repertoires. We can now identify immunoglobulin gene- segment usage, junctional regions, and modifications that contribute to specificity and diversity. Due to limitations on the International Space Station, alternate sample collection and storage methods must be employed. Our group compared Illumina MiSeq sequencing data from multiple sample preparation methods in normal C57Bl/6J mice to validate that sample preparation and storage would not bias the outcome of antibody repertoire characterization. In this report, we also compared sequencing techniques and a bioinformatic workflow on the data output when we assessed the IgH and Igk variable gene usage. Our bioinformatic workflow has been optimized for Illumina HiSeq and MiSeq datasets, and is specifically designed to reduce bias, capture the most information from Ig sequences, and produce a data set that provides other data mining options.

**Citation:** Rettig, T. A.\*, Ward, C.\*, Pecaut, M. J. & Chapes, S. K. Validation of Methods to Assess the Immunoglobulin Gene Repertoire in Tissues Obtained from Mice on the International Space Station. *Gravitational and space research: publication of the American Society for Gravitational and Space Research* 5, 2-23 (2017). \*Co-First Authors

## Introduction

For B cell development and specificity, there are a large number of heavy chain (IgH) and kappa light chain (Igκ) gene segments that are used to produce the immunoglobulin (Ig) receptor population repertoire.<sup>196</sup> This antibody repertoire is quite large and the possible specificities can theoretically exceed the number of actual antibody molecules in the host.<sup>133</sup> In antibodies, the antigen binding region is formed by six complementarity determining regions (CDRs) that loop out from the V region backbone formed by two beta-pleated sheets.<sup>197,198</sup> The germline V gene segment repertoire is necessary for host responses to pathogens and CDR1 and CDR2 are completely encoded for by variable (V region) gene segments.<sup>199</sup> Therefore, knowing which V gene segments are utilized is fundamental to understanding B-cell specificity and the development of effective immune responses. The CDR3 of both the heavy and light chains are highly variable due to their unique generation. During the creation of each Ig sequence, partially random splicing between V, D (heavy chain), and J gene segments occurs and random base insertions occur, called “n-nucleotide” additions. One hypothesis is that V gene segments have been maintained in the genome because of their importance in binding specific pathogens and provide essential host defense functions. However, CDR3 may be important because it is highly variable. Its role as a highly diverse and variable region is what provides the essential key to determining antigen specificity.<sup>200</sup>

The spaceflight environment can impact many parameters critical to the host immune response. In multiple species spaceflight affects the total body, thymus and spleen mass,<sup>189,201-211</sup> circulating corticosterone,<sup>209,210,212-219</sup> mitogen-induced proliferation, cytokine production and reactivity,<sup>179,189,192,210,213,215,220-232</sup> and lymphocyte subpopulation distribution.<sup>211,226,232-236</sup> Clearly

there are broad physiological impacts that affect many systems. The broader implications of this have been seen in space and ground-based models as detailed by Sonnenfeld and Crucian.<sup>237,238</sup>

B cells are among the immune components that are affected by spaceflight. The number of B cells in the spleen was reduced in mice flown on the space shuttle flight, STS-118.<sup>175</sup> The percentage of B cells in the bone marrow and spleen was also reduced in mice subjected to hindlimb unloading.<sup>176,177</sup> When rats were injected with sheep red blood cells 8 days prior to an 18.5-day COSMOS flight there were lower IgG concentrations compared to both immunized and non-immunized ground controls after landing.<sup>213</sup> IgM production was virtually eliminated in lymphocytes stimulated in vitro with pokeweed mitogen (PWM) on the International Space Station (ISS).<sup>239</sup> In this study, cells were activated on Earth, frozen down, and then put back into suspension in space where IgM secretion was significantly lower than similarly treated ground-based controls.<sup>239</sup> In a study of long-term ISS crewmember in-flight and post flight plasma samples, no significant changes to adaptive immunity or cytokine profiles were detected.<sup>171</sup> However, an assessment of peripheral blood revealed changes in the distribution of B cells and a reduction in T cell function following mitogen stimulation.<sup>238</sup>

Our group is interested in the impact of spaceflight on B cell immunoglobulin gene usage. Next Generation Sequencing (NGS) now allows us to analyze the repertoire of Ig gene segments that are used in the assembly of immunoglobulins that are transcribed by B cells and that are present in the host. NGS allows the determination of V, D and J gene usage, CDR3 assembly and the assessment of mutations that occur in response to immune challenge. In space, there exist certain limitations associated with tissue collection and storage methods traditionally

used on the ground. In preparation for sending mice to the ISS, our group needed to validate our procedures and our ability to obtain high-quality RNA that could be used for NGS for the assessment of Ig gene usage. Our group also sought to validate the usage of RNA extracted from whole tissue collected and stored with these limitations in mind. It was also necessary to develop a workflow that would facilitate the analysis of large amounts of data that would be generated during this project. In this manuscript, we describe the development of our workflow and validation of mouse NGS processing protocols for use with space flight experiments.

## **Materials and Methods**

### **Sample Preparation and RNA Extraction**

Spleens were removed from four 11-week-old, specific pathogen-free, female C57BL/6J mice housed in the Division of Biology vivarium at Kansas State University. One-half of each spleen was homogenized with a 70  $\mu$ M sieve to generate a single-cell suspension designated, “**cells**”. Spleen cells were pelleted at 350 x g and were resuspended in 5 mL of ice-cold ACK lysing buffer (155mM  $\text{NH}_4\text{Cl}$ , 10mM  $\text{KHCO}_3$ , 0.1mM EDTA) to lyse red blood cells. After five minutes, 10 mL of ice-cold isotonic medium was added to the suspension and cells were again pelleted at 350 x g and the supernatant was removed. The pellet was resuspended in Trizol LS (Ambion, Carlsbad, CA, USA) for RNA extraction. The remaining spleen half was immediately placed into Trizol LS for RNA extraction, and designated “**tissue**”. RNA extraction was performed according to the manufacturer’s instructions. RNAsin (40 units) was added to each RNA aliquot and stored in  $-80^\circ\text{C}$ . RNA quality was assessed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

## MiSeq Sequencing

One microgram of high-quality total RNA was used for RNA sequencing (RNA-seq) library construction using the TruSeq RNA Sample Preparation kit v2 (Illumina, San Diego, CA) with the following modification: one minute fragmentation time was applied to allow for longer RNA fragments. The obtained RNA-seq libraries were analyzed with the 2100 Bioanalyzer. The **tissue** sample described previously was then subjected to size selection and designated “**size selected**”. Selection of sequences 375-900 nucleotides (nt) in length (275-800 nt sequences plus 50 nt sequencing adaptors on each end of the cDNA) was performed using the Pippin Prep system (Sage Science, Beverly, MA, USA). All libraries were then quantified with qPCR according to Illumina recommendations. The sequencing was performed at the Kansas State University Integrated Genomics Facility on the MiSeq personal sequencing system (Illumina) using the 600 cycles MiSeq reagent v3 kit (Illumina) according to Illumina instructions, resulting in 2 x 300 nt reads.

## MiSeq Reference Mapping

The bioinformatics workflow used in our study is outlined in Figure 2.1. FASTQ sequencing files were imported into CLC Genomics Workbench v9.5.1 (CLC bio, Aarhus, Denmark) (<https://www.qiagenbioinformatics.com/>). Data were cleaned in the CLC program to remove low quality and short sequences. Due to Illumina sequencing artifacts, the first 12 nt were removed from each sequence. Sequences were quality cleaned by retaining the longest region of the sequences with at least 97% of the sequence with Phred scores over 20. Sequences with fewer than 40 nt were removed. Reads remained with paired-end sequences, designated “**paired**”, and, in cases of overlapping sequence pairs, reads were merged, designated as “**merged**” (Figure 2.1A, light blue line). Sequences were merged using a match score of +1,

mismatch cost of -2, a gap cost of -3, and a minimum score of 10. Cleaned paired (Figure 2.1A, purple lines) and merged (Figure 2.1A, red lines) sequences were mapped to specific C57BL/6 V gene sequences obtained from NCBI for the immunoglobulin heavy (IgH) and light (Igκ) chains. The paired and merged sequences were mapped using a match score of +1 and a mismatch score of -2. Additional putative antibody sequences were obtained by mapping sequences to the IgH and Igκ loci and the whole genome using the same scores. Mapped MiSeq sequences were combined and submitted to ImMunoGeneTics's (IMGT) High-V Quest (Figure 2.1A, green lines) (Alamyar et al., 2012). Functionally productive heavy chain sequences were imported into CLC and constant regions were determined using a motif search for the first 20 nt in each constant region that are provided in Table 2.1 (dashed line). Motifs were reassociated with their original sequences in Microsoft Excel for complete antibody (V[D]JC) identification. While two IgG subclass motifs were used to identify IgGs, they share partial sequence homology, therefore all IgG subclasses were combined resulting in an overarching IgG isotype. Kappa chain sequences were processed directly (Figure 2.1A, dotted line).

To collect a higher number of putative Ig sequences, our data handling workflow used multiple mapping processes which could result in the same sequence being submitted to IMGT multiple times. Failure to remove these duplicated sequences would lead to incorrect frequency assessments. Figure 2.2 outlines the procedure for duplicate sequencing read removal. Sequences were identified by their original MiSeq identification numbers for sorting. To retain the sequence with the most information and most accurate mapping, sequences were assessed based on functionality, constant region identification, V region score, and strand. Only one sequence per MiSeq identification number was retained and used for subsequent data compilation. Data from



the remaining productive and unknown functionality antibody sequences were compiled for V, D (IgH only), and J gene segment usage, CDR3 length, and CDR3 amino acid (AA) composition.

### **HiSeq Reference Mapping**

The MiSeq workflow described above was modified to analyze mouse liver transcriptomic data from the Rodent Research 1 (RR1) NASA validation flight provided by the NASA GeneLab program (<https://genelab-data.ndc.nasa.gov/genelab/>, Accession Numbers: GLDS-47, GLDS-48). These data include sequences from the livers of ground control and flight mice from two separate cohorts, CASIS (GLDS-47) and NASA (GLDS-48), that were generated using Illumina HiSeq (1 x 50 nt reads). Raw sequencing reads were imported into CLC and quality cleaned as described with the exception of short read removal as quality cleaned reads were below the threshold utilized in the original workflow (Figure 2.1B). Reads were then mapped to the V $\kappa$  references identified above. Total V $\kappa$  mapping counts were collected and analyzed in Excel.

### **MiSeq and HiSeq Genome Mapping**

FASTQ files were imported into CLC and quality cleaned as described previously (Figure 2.1C). MiSeq reads were merged as described previously (Figure 2.1C, blue arrow). Paired and merged MiSeq and unpaired HiSeq reads (Figure 2.1C, purple arrow) were mapped using the RNA-Seq tool in CLC to the current mouse reference genome (GRCm38). A match score of +1, a mismatch score of -2, and insertion and deletion scores of -3 were used to map reads to the genome. Due to the short read lengths of the HiSeq data, V-gene segment usage was compiled directly after the RNA-Seq analysis (Figure 2.1C, green arrow). For MiSeq data, reads

were collected, submitted to IMGT, duplicates removed and usage statistics compiled as described above (Figure 2.1C, dashed box).

### **NASA RNA Preparation and Sequencing**

Tissues from two sets of mice were analyzed. The first set of spleens and livers were removed from five 35-week-old female C57BL/6Tac mice aboard the ISS 21-22 days after launch (CASIS Flight, SpaceX-4). Five 35-week old female mice housed in the ISS Environmental Simulator were processed similarly with a four-day delay (CASIS Ground Controls).<sup>240</sup> Spleens and livers were placed in RNAlater (LifeTechnologies, Carlsbad, CA) for at least 24 hours at 4°C and then stored at -80°C while aboard the ISS, during transport, and upon return to Earth. The second set of tissues were isolated from seven 21-week-old female C57BL/6J mice that were euthanized aboard the ISS 37 days post-launch (NASA Flight, SpaceX-4). Carcasses were immediately frozen (-80°C) and after arriving on Earth, were dissected. Livers were preserved in RNAlater for at least 24 hours at 4°C and then frozen at -80°C. RNA was extracted from the tissues using Trizol (LifeTechnologies, Carlsbad, CA) according to the manufacturer's instructions. The resultant RNA was processed through an RNeasy mini column (QIAGEN, Hilden, Germany) as per manufacturer's instructions. RNA quality was assessed on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and stored at -80°C. RNA isolated from liver tissue was sequenced on Illumina HiSeq with single reads of 50nt (1x 50nt). Additionally, Illumina MiSeq sequencing of splenic RNA isolated from three CASIS ground control animals was performed as described earlier.

## Results

### Size Selection Yields Highest Number of Antibody Sequences

Most studies of Ig gene-segment use frequency have used single-cell suspensions or sorted cells to isolate specific cell population.<sup>14,199,241-243</sup> The advantage of these preparations is the exclusion of extraneous tissues and cells, which can enhance recovery of Ig sequences.

Our goal was to assess Ig gene segment usage in mice housed on the ISS. Due to limitations of animal and tissue handling on the ISS, only whole tissues would be available for analysis. To determine if we could obtain a sufficient number of Ig sequences from alternative sample preparations, we examined the total Ig sequence results from three different RNA preparation and sequencing treatment groups. The first two treatment groups comprised of RNA prepared from cells or whole tissue. A third treatment group was the same RNA from the tissue, which was subsequently size selected at 275-800 nt and sequenced independently. We selected this range of lengths because the IgH VDJ recombination sequences generally require 350-450 nt to gather information on V/D/J/C usage. Size selection also allows us to eliminate the inherent Illumina bias for short reads, while maintaining total transcriptome integrity for later data mining purposes.

Total sequence numbers generated were similar among the treatment groups with the cells treatment group resulting in 23.9 million reads, tissue treatment group with 25.9 million, and size selected treatment group with 21.5 million reads, as shown in Table 2.2. After cleaning, 18.7 million reads remained in the cells treatment group, 20.3 million in the tissue treatment group, and 12 million in the size selected group (Table 2.2). Mapping was performed as described in the materials and methods for both the IgH and Igκ loci and VH- and Vκ-gene

segments. Locus mapping returned higher levels of probable Ig sequences than V-gene segment specific mapping. V-gene segment mapping of sequences to both locus and gene segment references from quality cleaned reads was the lowest in the cells treatment group (1.56% IgH and 1.51% Igκ) and highest in the size selected treatment group (3.1% IgH and 2.69% Igκ).

After submission to IMGT and cleaning, antibody sequences were identified as potentially functional or of unknown functionality by IMGT. Unknown functionality sequences were comprised of partial sequences lacking CDR3 information to determine functionality. The total number of antibody sequences obtained for the IgH and Igκ was lowest in the cells treatment group and highest in the size-selected treatment group, (Table 2.2) mirroring the V-gene segment mapping results. Both productive and unknown IgH and Igκ antibody sequence counts were also lowest in the cells treatment group and highest in the size-selected treatment group. More unknown than productive sequences being identified in both IgH and Igκ. The size-selected treatment group produced both the highest number of productive antibody sequences and the highest total number of identified Ig sequences among treatment groups.

### **Comparison of Immunoglobulin Gene Segment Usage Among Treatment Groups**

We compared IgH and Igκ gene segment frequency using multiple metrics across all three treatment groups; the first of which is V gene segment usage. To assess the frequency of each VH- and Vκ- gene segment in normal mouse spleen, the total frequency of each V-gene segment was tabulated in Figure 2.3 and Figure 2.4 as a percentage of the total repertoire for our cells, tissue, and size selected treatment groups. VH-gene segment usage was similar among all three treatment groups (Figure 2.3A). V-gene segment V1-80 was detected most frequently, followed by V1-18, and V1-26 gene segments. The gene segment V1-80 was ranked either first

or third as a percentage of the total repertoire in all three treatment groups (Figure 2.3B). The gene segment V1-18 ranged between the first and third most frequently used gene segment. V1-26 was the second to fifth most frequently used VH-gene segment. While gene frequency detection rankings among cells, tissue and size selected treatment groups were not identical, there was high similarity in overall VH-gene segment detection and in repertoire usage. Correlations between treatment groups produced an  $R^2$  of at least 0.7562 ( $p < 0.0001$ , data not shown) between the cells and size selected treatment groups. The  $R^2$  between cells and tissue treatment group was higher ( $R^2 = 0.8149$ ,  $p < 0.0001$ , data not shown). Tissue and size selected treatment groups had the highest correlation ( $R^2 = 0.9645$ ,  $p < 0.0001$  data not shown).

Kappa chain V-gene segment usage was also compared among the different treatment groups. Figure 2.4 shows the percent of repertoire for the top ten most abundant  $V_k$  gene-segments of each treatment group. There was significant overlap in the top  $V_k$  of each treatment group when assessed as either percent of repertoire detected (Figure 2.4A) or when ranked from highest to lowest frequency (Figure 2.4B). Greater similarities in  $V_k$  existed between tissue and tissue size selected treatments. Correlations between  $V_k$  in treatment groups produced an  $R^2$  of at least 0.8335 ( $p < 0.0001$ , data not shown), with tissue and size selected treatment groups having the highest correlation ( $R^2 = 0.9894$ ,  $p < 0.0001$ , data not shown).

The frequency of D- and JH-gene segment use in normal mice was also assessed. Figure 2.5 shows that the cells, tissue and size selected D- and JH-gene-segment usage frequency was similar. D1-1 was the most frequently discovered D-gene segment in all three treatment groups, comprising almost 30% of the repertoire (Figure 2.5A). Due to the short length of the D gene, it

was often difficult to properly determine which D gene was used in an antibody. When a D gene was identified, but was attributed to a non-strain-specific D gene, they were labeled “undetermined”. These D-gene segments were also very common, occurring between 26%-28% of the time, in all three treatment groups. Gene segments D2-4, D4-1, D2-3, D2-5 and “no” D-gene (labeled NONE) were the next most frequent assignments, ranging from about six percent to eight percent of D-gene frequency.

We found that JH-gene segment usage was the same for the cells, tissue and size selected treatment groups (Figure 2.5B). JH2 was the most frequently used J-gene segment followed by JH1, JH4, and JH3, respectively. Gene segment usage in kappa chains was also similar in all three treatment groups, with Jκ1 as the most frequently used, followed by Jκ5, Jκ2 and Jκ4 respectively (Figure 2.5C). When less than six nucleotides from the J gene segment were identified, they were marked as <6 nt.

Five heavy chains, IgA, IgD, IgE, IgG (all subfamilies), and IgM are part of the normal mouse Ig repertoire. Almost 80% of the repertoire used the IgM constant region (Figure 2.5D). IgD, IgA, and IgG were detected at frequencies between three percent and twelve percent of the total repertoire and were evenly distributed among cells, tissue and size selected treatment groups. IgE was not detected in any of the treatment groups.

### **CDR3 AA Sequence Determination**

CDR3 is highly variable and it may be critical in determining antigen specificity.<sup>200</sup> Therefore, we assessed individual CDR3 frequency from each treatment group. The top five most common CDR3s from each treatment group for the heavy chain were compiled and shown

in Figure 2.6A, resulting in a total of ten unique CDR3s among treatment groups. The tissue and size selected treatment groups contained all of the most common CDR3s, however, the cells treatment group lacked the CARGIYYGSYFDYW sequence, which ranked as the second most common in the tissue data set (Figure 2.6B). We detected one hundred sixty-four CDR3 AA sequences in all three data sets at least once (Figure 2.6C). The tissue and size selected treatment data sets shared 607 CDR3 sequences. Thirty-eight and 82 CDR3 AA sequences were shared between cells and tissue and cells and size selected treatment groups, respectively. Each treatment group data set also contained a large number of unique CDR3 reads.

We found overlap among the top five kappa chain CDR3 of each treatment group (Figure 2.6D). All top CDR3 sequence were found among in all three datasets and CDR3 sequences appeared in at least the top 78 CDR3 sequences of the other treatment groups out of 2814 unique CDR3 that were identified (Figure 2.6E). Figure 2.6F shows the diversity of kappa chain CDR3 sequences. The total number of CDR3 amino acid sequences that were unique to each treatment groups was 441, 811, and 848 in cells, tissue, and tissue size selected treatment groups, respectively. Three-hundred and eighty-one unique kappa chain CDR3 sequences are shared among all three treatment groups.

### **Application of MiSeq Workflow to RR1 HiSeq Data**

The MiSeq workflow was adapted to process the liver Illumina HiSeq data from the RR1 validation flight. Due to short read length (38 nt), only V-gene segment usage was assessed. Due to low IgH read numbers, only V $\kappa$  information is presented. The V $\kappa$  percent of repertoire from each HiSeq RR1 mouse cohort (CASIS Ground, CASIS Flight, NASA Ground, and NASA Flight) was compared to the V $\kappa$  percent abundance of the cell, tissue, and size selected HiSeq

datasets. Table 2.3 shows poor correlation between reference mapped RR1 HiSeq cohorts and the MiSeq datasets described in the previous section. As all mice represented in this comparison are C57BL/6 mice, though ages and experimental conditions varied, we were concerned that the lack of concurrence in V $\kappa$  gene-segment usage of the RR1 mice and those used in the workflow discussed above may reflect the bioinformatic treatment of the data. To test this hypothesis, we modified the workflow to map sequencing reads to the entire *Mus musculus* genome rather than mapping reads to V $\kappa$  reference sequences, as genome mapping is commonly employed in transcriptomics analysis. This bioinformatic treatment yielded a higher correlation with the MiSeq datasets. Table 2.3 shows that the distribution of V $\kappa$  percent abundance was dependent on the mapping technique used for the HiSeq datasets.

### **Reference/Locus Mapping is Comparable to Whole Genome RNA-Seq Methods**

Because the data obtained from the HiSeq data set using reference mapping techniques were less correlative to V $\kappa$  gene use to our previously obtained MiSeq data for normal mice we were concerned that our initial bioinformatics techniques for MiSeq data may not be appropriate. To validate the accuracy of our bioinformatic treatments of the sequencing data that were submitted to IMGT, two different mapping methods were compared using Illumina MiSeq data from CASIS ground control animals. The reference mapping approach, used previously, mapped sequences to both the IgH V-gene segments (251 segments) and the entire IgH locus (2.8Mb) obtained from NCBI (NC\_000078.6, 113258768 to 116009954) or Ig $\kappa$  V-gene segments (164 segments) and the entire Ig $\kappa$  locus (3.2Mb) obtained from NCBI (NC\_000072.6, 67555636 to 70726754). Therefore, we used the whole genome mapping outlined above to compare to our reference mapping. Results were obtained by using the RNA-Seq analysis tool in CLC to map reads to the entire genome with the IgH and Ig $\kappa$  loci selected for submission to IMGT. The



IMGT output was processed similarly for both (genome *vs.* reference) mapping strategies. The median frequency of all VH- and V $\kappa$ -gene segments was compared if it was detected in at least one of the three animals and the data were compiled for both reference- and genome-mapping options. V-gene segments not detected in a treatment group were assigned a “zero” frequency. Assessment of the median frequencies of the two methods by linear regression in Figure 2.7 revealed that the frequency data for VH-gene segment usage was very similar regardless of the mapping technique ( $R^2 = 0.9973$ ,  $p = <0.0001$ ) (Figure 2.7A). There was also a strong correlation of V $\kappa$  usage between the two methods ( $R^2=0.9991$ ,  $p<0.0001$ ) (Figure 2.7B). Comparisons of D-gene segment, J-gene segment, constant region frequency and CDR3 lengths were also highly correlated using both techniques (data not shown). Therefore, we are confident that our reference mapping bioinformatics strategy is providing an accurate picture of V-gene usage.

## Discussion

Spaceflight presents unique difficulties in the collection, preparation and preservation of cells and tissues. Normal preparation methods such as the creation of single-cell suspensions are difficult and normal tissue preservations methods such as the use of liquid nitrogen for flash-freezing are unavailable, leading to the preservation of tissue in RNAlater followed by long term storage at  $-80^{\circ}\text{C}$ . In an effort to determine the acceptability of whole tissue preparations compared to more traditional single cell suspensions, we examined the differences in Ig sequences obtained from both treatment groups. We were concerned that tissue isolation methods may introduce artifacts into the data since many studies specifically focus on single cell suspensions; often sorted, to isolate B cells specifically.<sup>67,242-244</sup> While animals utilized in this study were not specifically challenged, the presence of plasma cells and plasmacytes, which

produce several orders of magnitude more immunoglobulin transcripts, cannot be excluded and represents a weakness of our workflow for the analysis of naïve repertoires. Similarly, underrepresentation of subpopulations with limited stability may lead to other biases.<sup>25,26</sup>

Our data indicate that comparable results were obtained from both the tissue and the cells treatment groups as there were strong correlations in V-gene usage. Due to the combinatorial nature of CDR3, the level of shared sequence identity was encouraging among the three treatment groups. It should be noted, however, that even within the top H-CDR3, a higher degree of similarity was found in the tissue and size selected treatment groups than was found in the cells treatment group. While the high level of similarity between the tissue and size selected cohort is to be expected given that the size selected treatment was a subset of the tissue RNA, the deviation in the cells treatment group could possibly result from an unbalanced sampling of B cell subsets depending on the portion of the spleen that was utilized during sample preparation, which would be exacerbated by clonal expansion. For instance, marginal zone B cells have been shown to have a preference for short H-CDR3 amino acid motifs that lack n-nucleotide additions as compared to follicular B cells.<sup>245,246</sup>

In an effort to reduce Illumina bias towards short reads seen in the cells and tissue treatment groups, ten months later, we sequenced the same tissue total RNA using size selection. The extended storage time after initial sequencing and additional freeze/thaw cycles are likely the cause of reduced numbers of post-cleaning reads due to RNA degradation. Nevertheless, the size-selected data set still provided the highest number of productive and unknown antibodies. Subsequent preparations have verified that size selection is helpful in providing the highest

number of antibody sequences (data not shown). Therefore, we have chosen to include size selection in our protocol for NGS assessment of Ig V-gene usage.

The antibody repertoires from numerous species have been analyzed using a variety of amplification, sequencing, and analysis techniques.<sup>134,244</sup> We chose to assess Ig gene usage without using amplification. Although many studies use amplification in order to obtain a higher number of reads, this may lead to bias into the repertoire.<sup>133,134</sup> Bias may be introduced due to primers or to errors created during the PCR reaction.<sup>134,199</sup> The large number of primers needed to amplify all the V genes in mice also presents some obstacles. Some have used 5' RACE with primers based on the constant region.<sup>134,137</sup> However, in order to amplify the entire repertoire, multiple 5' RACE primers are required, which still increases the chances for primer bias and increases costs significantly. Our goal was to examine the breadth of the antibody repertoire by gathering information about V, D, J, constant region usage and CDR3 composition. The detected V-gene and CDR3 sequences appear to parallel the repertoires reported using more focused amplification methods. Therefore, we have a methodology for future studies that will examine the immune response to vaccination.

During the course of our studies, we had the opportunity to work with both HiSeq and MiSeq data. While Illumina sequencing (HiSeq and MiSeq) produces a higher volume of sequence reads, they are shorter and more prone to errors than sequencing with 454 or Sanger methods.<sup>134</sup> However, Illumina sequencing has improved over time and is arguably now the NGS of choice. Our sequencing with Illumina MiSeq allowed us to obtain reads of up to 560 nt when forward and reverse sequencing ends were paired. This provides enough sequencing length to

capture information from the V-gene segment to the constant region of both the heavy and light chains. We also found that as the sequences became longer, there was a drop off in sequencing quality, which has been previously reported.<sup>247</sup>

Our workflow for Ig sequence isolation selected for sequences with the most information. This required the merging of overlapping read pairs to provide sequences long enough to identify the V, D, J, and constant regions. In order to collect the highest number of possible Ig sequences, we used multiple mapping techniques to both the V-gene segment and the locus in an effort to collect every possible Ig sequence. Preliminary workflow attempts found that each mapping technique isolated some unique sequences and that locus mapping resulted in a high number of “false positive” sequences. Subsequent sequence removal in Excel selected for antibodies containing the most data retaining productive antibodies with constant regions and high V-gene scores, a measure of the length and accuracy of match to the germline V gene segment. By utilizing multiple mapping methods and subsequent selection for the sequence with the most information, we are able to obtain a relatively large number of antibody sequences without introducing primer bias or PCR-induced sequencing errors.

To the best of our knowledge, this is the first data set of tissue based, non-amplified MiSeq analysis of the antibody repertoire. While our results are not a direct technique match to other published data sets, our normal mouse V-gene usage is consistent with the findings from other laboratories. For example, Collins used 5' RACE from the constant region followed by sequencing using 454 on a splenic cell suspension.<sup>14</sup> Of the top ten VH-gene segments identified by Collins, we identified five of the same VH genes within our top ten most frequently detected.

All except the V1-59 gene segment were among the highest contributors to our repertoire.<sup>14</sup> In addition, JH-gene segment frequencies were also relatively uniform with the J2 gene segment as the most frequently used.<sup>14</sup> Yang performed sequencing on cell sorted B cells isolated from the spleen followed by amplification with primers specific for many, but not all, of the V heavy chains of mice and constant region primers to amplify V, D, J and part of the constant region.<sup>67</sup> Their PCR products were then sequenced on the Illumina platform and aligned to known VH-gene segments.<sup>67</sup> They identified V1-26 as their most common V-gene segment, which ranked between the second and fifth most common V gene in our data sets. V-gene segments V1-82, V1-64, and V1-55 were also identified as common V-gene segments in their analysis, all of which were frequently detected in our data.<sup>67</sup> Kaplinski also examined sorted spleen cells, amplified with PCR and sequenced on MiSeq with 2 x 150 nt reads.<sup>242</sup> Sequencing results were analyzed through idAB for identification.<sup>242</sup> In contrast to our data analysis, Kaplinski examined only V gene segments found in productive antibodies, where we compiled all V gene segments identified in productive and unknown functionality sequences.<sup>242</sup> Of the most common VH-gene segments provided, four, V1-80, V1-26, V1-53, and V1-82 were identified in our top ten grouping.<sup>242</sup> Our data sets also isolated D1-1 as the most common D-gene segment. Kramer *et al.* examined sorted splenic follicular B cells, using IgM restricted PCR and sequenced using the Sanger method.<sup>243</sup> As we discovered, Kramer *et al.*'s most common VH family was V1, followed by V2 and V5 at relatively equal levels (Kramer et al., 2016). In contrast, we found that the V6 gene-segment family was detected at a higher level than found in the Kramer analysis.<sup>243</sup> The J4 gene segment was also used more than detected in our data set.<sup>243</sup> We both identified D1 and D2 as the most common D gene-segment families.

We also compared our data to Ig $\kappa$  gene family usage. Aoki-Ota assessed skewed V $\kappa$  gene segment usage and V-J gene segment usage in unimmunized C57BL/6 mice using primer amplified total RNA of B cells from spleen, bone marrow and lymph node using 454 pyrosequencing.<sup>241</sup> Their sequencing data was analyzed using the NCBI basic local alignment tool with reference sequences for V $\kappa$  and J $\kappa$  obtained from the IMGT data base. The top seven V gene segments identified in their study were also found to be among the most abundant V $\kappa$  gene segments in all of our treatment groups. Additionally, V-J pairing of their top gene segments paralleled our data. Lu examined the effects of primer bias and mouse to mouse variation in V $\kappa$ - and J $\kappa$ -gene segments and CDR3 regions using primer amplified total RNA isolates of Balb/c splenic B cells on the 454 pyrosequencing platform.<sup>199</sup> As with our study, sequencing reads were submitted to the IMGT HighV-quest tool, however, only functionally productive immunoglobulins were used in their analysis.<sup>199</sup> In spite of the strain differences between our studies, of the V $\kappa$ -gene segments representing over one-percent of the antibody repertoire reported by Lu *et al.*, at least 80% of those gene segments appearing at a frequency 0.5% or higher in our analyses.

Although there was not 100% agreement among our study and the others, there was a high degree of consistency. Variations in data may result from sequencing and tissue isolation techniques and natural variation among animals, including mouse strain. In addition, since we did not amplify for IgV gene segments, we likely may have missed rarer B cell clones. Primer biases in other studies may have also contributed to some of the differences. Nevertheless, it is clear that our approach provided an unbiased, representative sample of actively transcribing B cells.

Our group utilized liver Rodent Research One sequencing data sequenced on the Illumina HiSeq platform (1 x 50 nt) that was available from the NASA Genelab project. The sequencing length was the largest limitation of these data. Our MiSeq data were sequenced in both directions at a length of 300 nt. Some paired-end reads also contained overlaps, allowing us to merge these sequences and provide reads up to around 560 nt. HiSeq sequencing reads were not of sufficient length to obtain CDR3 composition from the IMGT HighV-Quest tool, limiting the analyses that could be used to assess the antibody repertoire. Therefore, the applicability of publicly available datasets to independent research questions is dependent upon the sequencing method used to acquire the data. For Ig gene repertoire studies, we recommend the use of sequencing methods that result in longer reads, though short reads may be useful for other research questions.

Initial comparisons to assess similarity of the different mouse cohorts showed a lack of correlation between the HiSeq RR1 data to the normal mouse MiSeq V $\kappa$  usage. We thought that part of the discrepancy may be from problems with the short HiSeq sequences, specifically when forced to align to V-gene segments when multiple matches are excluded. Mapping short HiSeq reads to the entire mouse genome remedied the inconsistencies observed between RR1 and normal mouse MiSeq data, likely due to the limitations of the RNA-Seq analysis employed. This demonstrates that the bioinformatic treatment of the data can impact results. We found that mapping longer MiSeq sequencing reads from RNA isolated from mice within the CASIS ground RR1 cohort to both the whole mouse genome and V $\kappa$  reference sequences yielded a strong correlation. This validates the applicability of the MiSeq workflow described in this work on additional MiSeq datasets and reinforces that sequence read length must be taken into account when selecting bioinformatics methods.

In conclusion, our goals for this project were to examine the breadth of the antibody repertoire gathering information about V, D, J, and constant region usage and CDR3 composition and to lay the foundation for future studies that will examine the immune response to vaccination during space flight. We have determined that whole tissue preparations as will be available from the ISS will yield similar results when examining the antibody repertoire. We also determined that performing a size selection to isolate likely antibody sequences provided the highest number of Ig reads. A novel workflow using multiple mapping methods to characterize NGS data for Ig repertoire data was developed and genome and reference mapping methods were validated through the use of publicly available datasets. This novel workflow can be used for future studies on the antibody repertoire regardless of whether they are ISS- or ground-based.

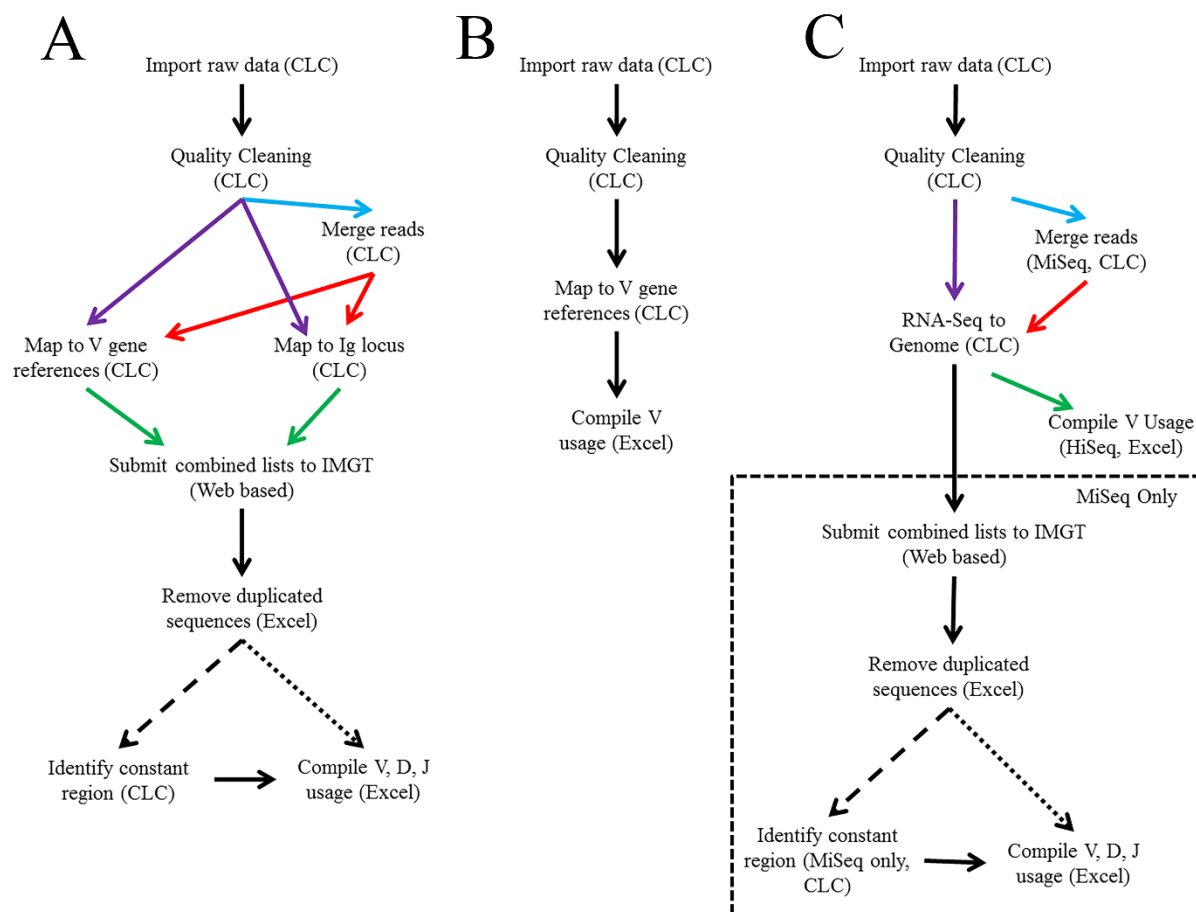
### **Acknowledgements**

This work was supported by NASA grants NNX13AN34G and NNX15AB45G, NIH grant GM103418, the Molecular Biology Core supported by the College of Veterinary Medicine at Kansas State University, and the Kansas State University Johnson Cancer Center. GeneLab data are from the NASA GeneLab Data Repository (<http://genelab.nasa.gov/data>). We thank Ms. Melissa Gulley for her help in the lab and Mr. Ricky J. Rettig for his help and expertise in Microsoft Excel. We also thank Drs. Ruth Globus and Sungshin Choi at NASA Ames for their help with RR1 data and Dr. Alina Akhunova, Director of the Kansas State University Integrated Genomics Facility, for her help, dedication and sequencing expertise.



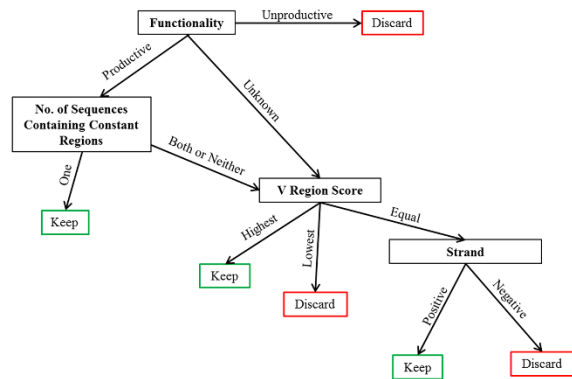
## Figures and Tables

**Figure 2.1. Bioinformatic analysis workflows.**



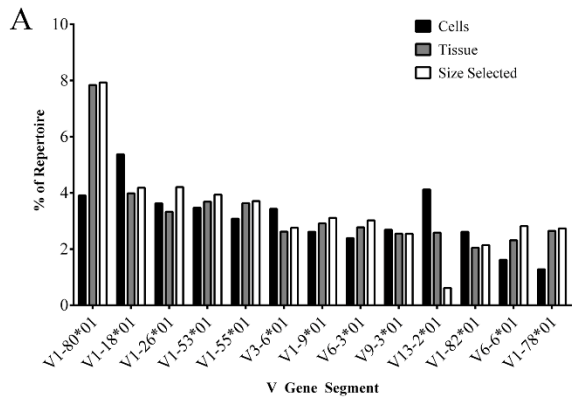
(A) Workflow for MiSeq reference mapping strategy using CLC Genomics Workbench software, the ImMunoGeneTics (IMGT) data base and Excel. (B) Workflow for HiSeq reference mapping strategy. (C) Workflow for MiSeq and HiSeq referenced mapping strategy.

**Figure 2.2. Decision-making matrix to remove duplicate sequence reads after IMGT processing**



Mapped sequences that were identified using Illumina sequence identification tags and sequences identified multiple times were removed as outlined.

**Figure 2.3. Top ten VH gene segments used among treatment groups.**

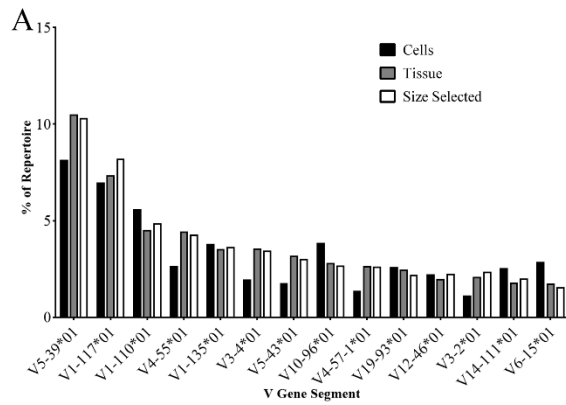


**B**

	Cells	Tissue	Size Selected
IGHV1-80*01	3	1	1
IGHV1-18*01	1	2	3
IGHV1-26*01	4	5	2
IGHV1-53*01	5	3	4
IGHV1-55*01	7	4	5
IGHV1-9*01	9	6	6
IGHV3-6*01	6	9	9
IGHV6-3*01	11	7	7
IGHV9-3*01	8	11	11
IGHV1-82*01	9	15	13
IGHV6-6*01	19	13	8
IGHV1-78*01	25	8	10
IGHV13-2*01	2	10	51

(A) The top ten VH gene segments for each treatment group are presented as a percent of repertoire with corresponding percent of repertoire in other treatment groups listed. (B) Top ten VH gene segments are listed by rank order (most frequent to least frequent). Dark red indicates higher rank moving to white, of lower rank. VH-gene segments with identical ranks are displayed as ties.

**Figure 2.4. Top ten V $\kappa$  used among treatment groups.**

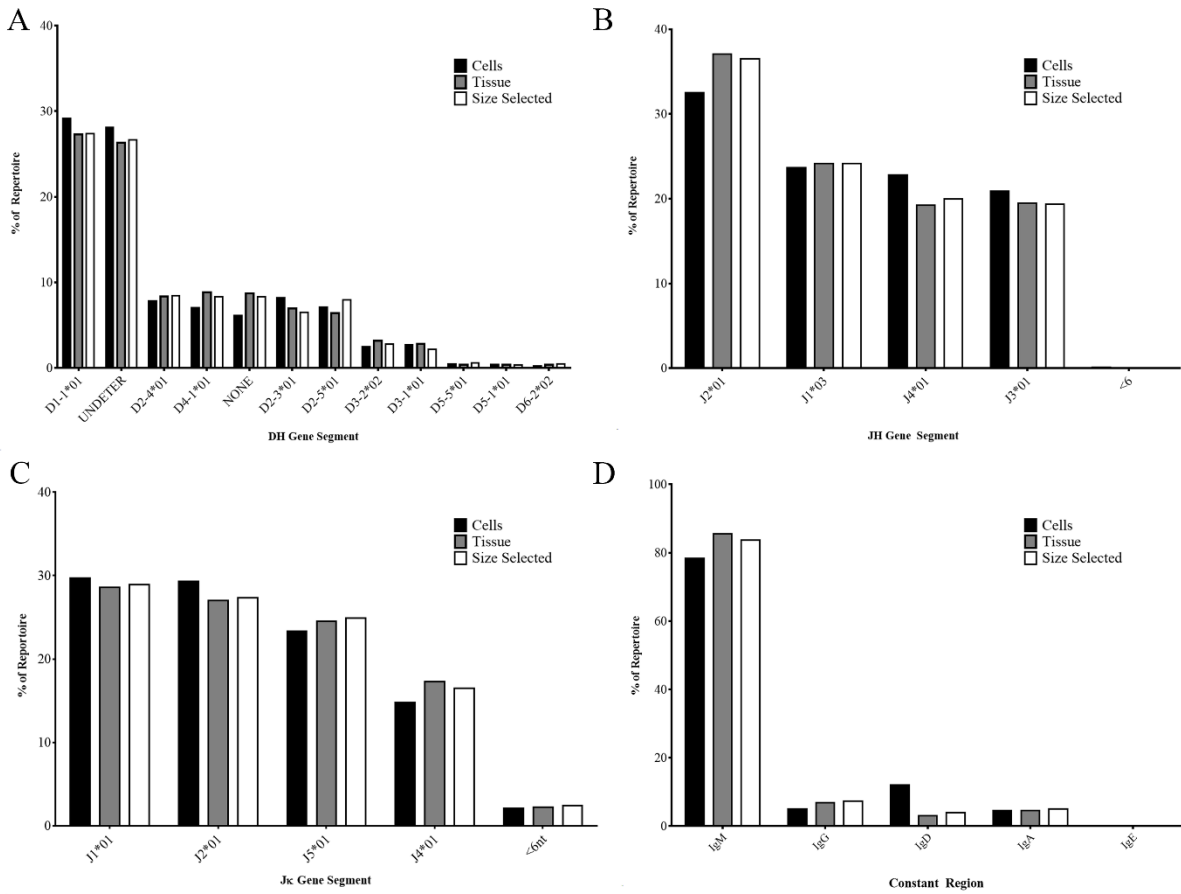


B

	Cells	Tissue	Size Selected
IGKV5-39*01	1	1	1
IGKV1-117*01	2	2	2
IGKV1-110*01	3	3	3
IGKV4-55*01	7	4	4
IGKV1-135*01	5	6	5
IGKV3-4*01	13	5	6
IGKV5-43*01	18	7	7
IGKV10-96*01	4	8	8
IGKV4-57-1*01	27	9	9
IGKV19-93*01	8	10	13
IGKV12-46*01	10	15	12
IGKV3-2*01	32	14	10
IGKV14-111*01	9	17	16
IGKV6-15*01	6	19	23

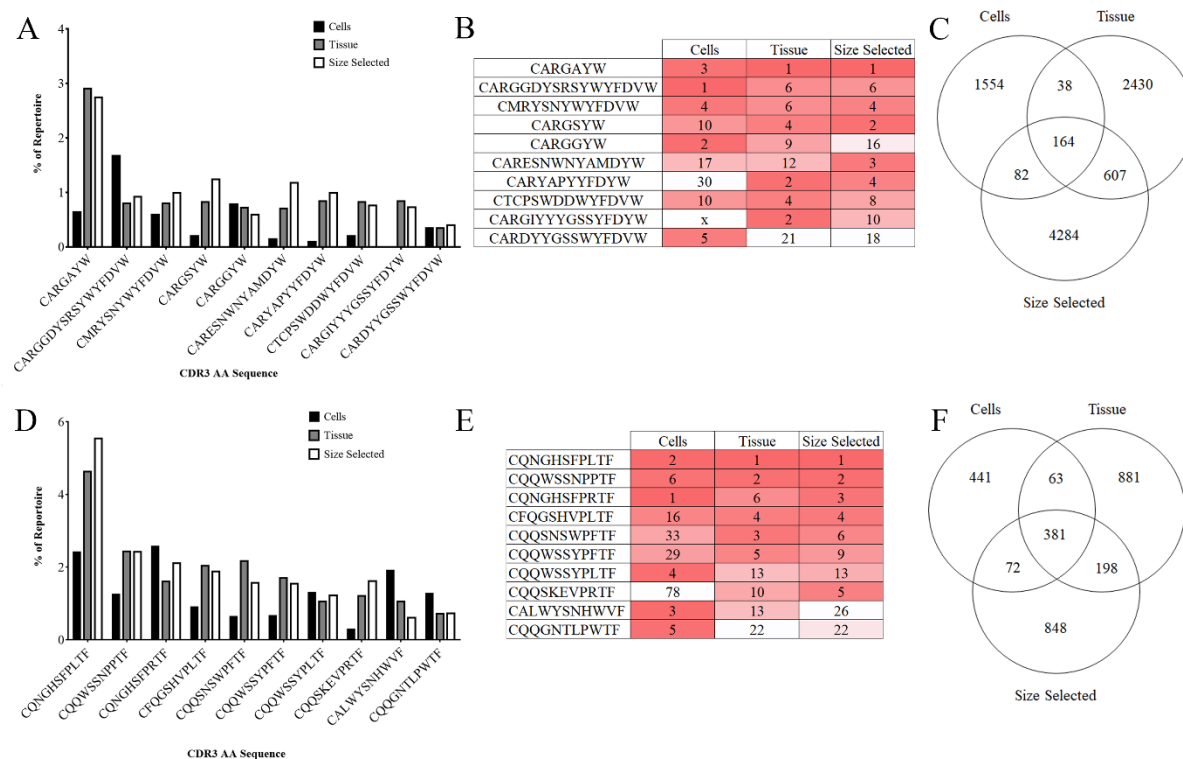
(A) The top ten V $\kappa$  gene segments for each treatment group are presented as a percent of repertoire with corresponding percent of repertoire in other treatment groups listed. (B) The top ten V $\kappa$  gene segments are listed by rank order (most frequent to least frequent). Dark red indicates higher rank moving to white, lower rank. VH-gene segments with identical ranks are displayed as ties.

**Figure 2.5. D, J, and heavy chain constant usage among treatment groups.**



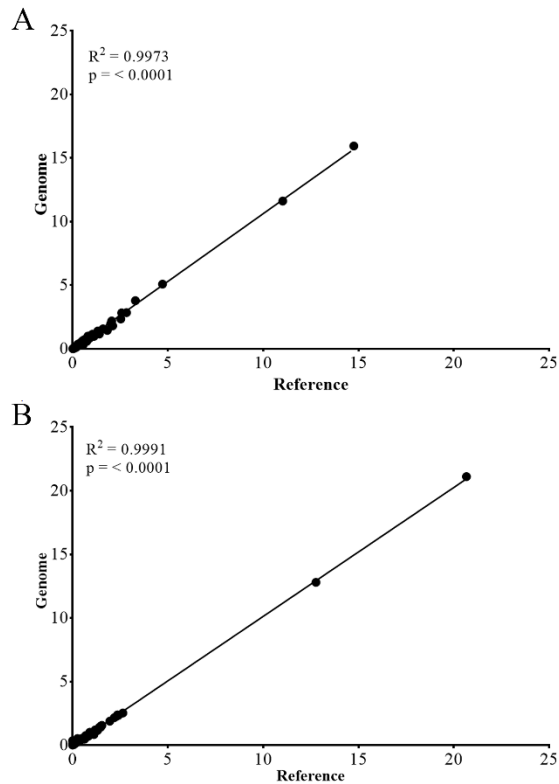
(A) DH gene segment usage by percent of repertoire. (B) JH gene segment usage by percent of repertoire. (C) Jk gene segment usage by percent of repertoire. (D) Heavy chain constant region usage by percent of repertoire.

**Figure 2.6. CDR3 AA sequence usage among treatment groups.**



(A) Top five most common heavy chain CDR3 AA sequence usage is presented as percent of repertoire. (B) Top five most common heavy chain CDR3 AA sequence usage is presented by rank. Dark red indicates higher rank moving to white, of lower rank. An x denotes that the AA sequence was not found. (C) Unique heavy chain CDR3 AA sequences identified within and among treatment groups. (D) Top five most common kappa chain CDR3 AA sequence usage is presented as percent of repertoire. (E) Top five most common heavy chain CDR3 AA sequence usage is presented by rank. (F) Unique heavy chain CDR3 AA sequences identified within and among treatment groups.

**Figure 2.7. Correlation of V gene segments between genome and reference mapping.**



(A) Linear regression of median VH gene segment usage from genome and reference mappings.  $R^2=0.9973$ ,  $p<0.0001$ . (B) Linear Regression of median Vκ gene segment usage from genome and reference mapping.  $R^2=0.9991$ ,  $p<0.0001$ .

**Table 2.1. Sequences used for heavy chain identification**

Constant Region	Motif Sequence
IgA	GAGTCTGCGAGAAATCCCAC
IgD	GTAATGAAAAGGGACCTGAC
IgE	TCTATCAGGAACCCTCAGCT
IgG1/2b/2c	GCCAAAACAACAGCCCCATC
IgG3	AACAACAGCCCCATCGGTCT
IgM	TCAGTCCTTCCCAAATGTCT

Motifs used to determine the constant region of heavy chain Ig sequences.



**Table 2.2. Sequencing and mapping results from the cells, tissue, and size selected treatment groups.**

	Cells	Tissue	Size Selected
Total Reads	23.9 M	25.9 M	21.5 M
Post Cleaning	18.7 M	20.3 M	12M
IgH Mapped	278318	313194	327015
VH Mapped	12851	26559	42375
Igκ Mapped	261037	273562	264938
Vκ Mapped	20776	35719	64540
Heavy Chain Productive	2036	4991	8939
Heavy Chain Unknown	6139	11374	14047
Light Chain Productive	3439	6799	11595
Light Chain Unknown	6894	10462	12393

**Table 2.3. Comparison of mapping techniques in HiSeq datasets.**

		Reference <sup>2</sup>	Genome <sup>3</sup>	Compared <sup>4</sup>
Cohort	N	R <sup>2</sup> (p-value)	R <sup>2</sup> (p-value)	R <sup>2</sup> (p-value)
CASIS G	3	0.030 (.0637)	0.101 (0.0015)	0.011 (.027)
CASIS F	3	0.001 (.776)	0.216 (<.0001)	0.042 (.074)
NASA G	7	<0.001 (.854)	0.379 (<.0001)	0.013 (.262)
NASA F	7	0.004 (.521)	0.277 (<.0001)	0.006 (.476)

Mapping techniques were compared by assessing the correlation of Vκ usage between multiple HiSeq and Miseq datasets. HiSeq datasets included sequencing data from CASIS and NASA ground (G) or flight (F) RR1 mice. The comparison groups are as follows:

<sup>2</sup>Vκ gene segment usage from reference-mapped HiSeq data versus Vκ gene segment usage of MiSeq sample preparation datasets.

<sup>3</sup>Vκ gene segment usage from genome-mapped HiSeq data versus Vκ gene segment usage of MiSeq sample preparation datasets.

<sup>4</sup>Vκ gene segment usage from reference-mapped HiSeq Data versus Vκ gene segment usage of genome-mapped HiSeq data

## **Chapter 3 - A comparison of unamplified and massively multiplexed PCR amplification for murine antibody repertoire sequencing**

### **Abstract**

Sequencing of the antibody repertoire has steadily become cheaper and easier for researchers. These sequencing methods usually rely on some form of amplification, often a massively multiplexed PCR prior to sequencing. In an effort to eliminate potential biases and create a data set that could be used for other studies, our lab compared unamplified sequencing results from the splenic heavy-chain repertoire in the mouse to those processed through two commercial applications. We also compared the use of mRNA vs total RNA, reverse transcriptase, and primer usage for cDNA synthesis and submission. We determined that the use of mRNA for cDNA synthesis resulted in higher read counts and reverse transcriptase or primer usage had no statistical effects on read count. Although most of the amplified data sets contained more antibody reads than the unamplified data set, we detected more unique V-gene segments in the unamplified data set. We also determined that although the unique CDR3 detection was much lower in the unamplified data set, it detected most of the high frequency CDR3s. We have shown that unamplified profiling of the antibody repertoire is possible, detects more V-gene segments, and detects high frequency clones in the repertoire.

## Introduction

Antibodies are proteins produced by B cells that play a vital role in adaptive immunity. These molecules are complex, and their diverse specificities are defined at several levels. At the protein level, these molecules are heterodimers joined with disulfide bonds between heavy and light chains. At the genetic level, the antibody specificity is influenced by semi-randomly recombining a variable (V), diversity (D), and joining (J) gene segments that are encoded in the genome. This V-D-J, or in the case of the light chain, V-J rearrangement, also influences antibody specificity by random and semi-random base insertions or deletions during the recombination process. In the end, the collection of antibodies produced inside an organism, the antibody repertoire, is a fingerprint of what antigens an organism has been exposed to and a possible measure of the level of immunocompetence. The sequencing of the antibody repertoire has become easier, cheaper, and faster in recent years. Antibody repertoires have been sequenced in numerous species, in response to vaccinations,<sup>40,74,78</sup> and infections,<sup>248-250</sup> and have been employed in cancer detection; providing valuable feedback regarding the immune system's response to challenges and for early cancer detection in patients.<sup>251,252</sup>

Our laboratory studies the effects of microgravity on the immune system.<sup>162,165,175,180,185,188</sup> Specifically, we were interested in understanding the impact of space flight on immunoglobulin gene usage and changes in the antibody repertoire in response to vaccination. In an effort to supplement our unamplified MiSeq data sets, we explored the use of commercial processes that use technologies to amplify sequences to increase the depth of coverage of specifically targeted immunoglobulin gene transcripts. These data sets are created using massively multiplexed PCR reactions that are subsequently sequenced on the Illumina

platform. Multiplex amplification strategies have been used to explore the T cell<sup>123,126</sup> and B cell repertoires<sup>44,127-129</sup> and both biases and PCR artifacts have been detected in the data sets.<sup>123</sup> Causes of bias or errors generated during a multiplex PCR amplified library preparation can range from primer/target binding property differences, CG content, errors in transcription, secondary target structure interference, multiplexing competition artifacts, as well as the inability to design primers that can specifically amplify all V-gene segments in the genome.<sup>123,126,130,131,137</sup>

We started this project with the hypothesis that our unamplified data set would provide comparable results to those seen in the commercially amplified data sets. However, we are unaware of any immunoglobulin repertoire studies that have been done to compare data obtained using amplification techniques compared to the repertoire in a total RNASeq library. Concurrent with performing these analyses, we discovered that the required sample preparation for commercial sequencing also varied. Therefore, we found it necessary to examine the impact of sample preparation as part of our effort. This manuscript describes our results comparing a data set generated using unamplified total RNA (TRNA) to commercially amplified data sets. We examined the role of commercial amplification and cDNA generation methods as well as the impact of the starting material on sequence output.

## **Materials and Methods**

### **KSU RNA Preparation**

RNA was prepared as outlined in Rettig et al 2017.<sup>45,58</sup> Briefly, RNA was extracted from the spleens of four nine-week old female C57Bl/6J mice. TRNA was submitted to the Kansas State University Integrated Genomics Facility for sequencing and cDNA was prepared using

standard Illumina protocols. cDNA made from the TRNA, was size selected to 275-800bp length and sequenced on the Illumina MiSeq at 2x300bp using Illumina's protocol.

### **Commercial Sample preparations**

mRNA was extracted from the TRNA isolated from the sample used for the KSU data set using the PolyATtract mRNA isolation system (Promega, Fitchburg, WI) following manufacturer's instructions. RT-PCR for samples amplified with Avian Myeloblastosis Virus (AMV) reverse transcriptase based sample preparation was performed using the Access RT-PCR System (Promega, Fitchburg, WI) following the manufacturer's instructions. RT-PCR for Moloney Murine Leukemia Virus (MMLV) reverse transcriptase based sample preparation was performed using the SuperScript III First-Strand Synthesis System (Invitrogen, Carlsbad, CA) following manufacturer's instructions. Starting material for RT-PCR was either TRNA from the KSU data set or the purified mRNA. RT-PCR products were submitted to Adaptive Biotechnologies (Seattle, WA) (Com1) on dry ice following company protocols. TRNA and mRNA (unamplified) were submitted to iRepertoire, Inc. (Huntsville, AL) (Com2) on dry ice following company protocols.

### **Bioinformatic Analysis**

KSU sequencing results were analyzed as described previously.<sup>45,58</sup> Briefly, sequencing results were quality controlled, antibody-specific sequences were isolated, and submitted to ImMunoGeneTics (IMGT)<sup>253</sup> for analysis. Individual sequences are assigned a unique ID by the sequencing machinery during Illumina sequencing. These sequence IDs were used to identify unique sequences. The sequence containing the most high quality information was retained for

further analysis as outlined in Rettig et al.<sup>45</sup> No further filtering of reads was performed. Both commercial sequencers provided their own bioinformatic analyses of the sequencing results. The raw sequencing results from Com1 and Com2 were also submitted to IMGT for analysis and subjected to the bioinformatics pipeline standardly employed for KSU data sets. IMGT's nomenclature and classifications are used throughout this paper, including ORF and putative gene segments not yet fully mapped to the genome. The IMGT lists contain all of the gene segments presented in the NCBI data base. IMGT's High-V Quest occasionally assigned multiple potential V-gene segments to a single sequence, likely due to incomplete capture of the entire V-gene sequence or high homology between gene segments. In all IMGT processed data, sequences that contained two possible V-gene segment possibilities were assigned a weighted value of 0.5 per sequence, as opposed to one for full matches. Sequences with V-gene segments that were assigned more than two potential matches were excluded from analysis. Initial results were tabulated using the company's proprietary bioinformatic results. However, to determine the role of bioinformatic handling of the data, some of Com1 and Com2 data were subjected to the standard KSU bioinformatic workflow analysis and CDR3 analysis<sup>45</sup> using IMGT's results.

## **Statistical Tests**

All statistical analyses were carried out using GraphPad Prism (Version 6.0). Paired T-tests were performed using the raw read counts. Coefficient of determinations ( $R^2$ ) were performed by comparing the percent of repertoire between animals. Percent of repertoire is determined by dividing the read count for a specific V-gene segment by the total number of reads detected and multiplying by 100.

## Results

Most studies examining immunoglobulin repertoires use amplification to increase the depth of sequencing, but amplification comes with some drawbacks. We wanted to assess the comparability of amplified and non-amplified data from identical samples. In preparation to do this comparison, we found that different commercial amplification methodologies required different types of sample preparation. For example, sample submission for the Com1 data sets required a cDNA sample. The Com1 process amplified the resulting cDNA using proprietary primers and sequencing on the Illumina platform. After an initial submission showed a low correlation between the Com1 sequencing and the KSU data set (data not shown), we hypothesized that cDNA preparation plays a role in determining the amplified repertoire. To test this hypothesis, we assessed the role of starting material (mRNA or TRNA), reverse transcriptase (AMV vs MMLV), and primer templates (oligo-dT or random hexamer) on the sequenced B cell immunoglobulin repertoire. Com2 submissions required the submission of raw RNA, rather than cDNA.

### Assessment of Transcriptional Read Counts

The amplified data sets from Com1 returned between 7,084 and 1,263,003 sequences, dependent on the preparation method. We found that mRNA starting material yielded more total transcriptional reads than TRNA ( $P=.013$ , 2 tailed matched T test; Table 3.1). Generally, the AMV reverse transcriptase and random hexamer primers also tended to yield higher numbers of transcripts. The use of AMV and random hexamer primers resulted in more total productive reads in three out of four of the comparisons directly comparing primers, however, the overall differences were not statistically different ( $P>0.05$ , 2 tailed matched T test; Table 3.1). In the



Com2 data set, we found a moderate number of reads, about one-half of those detected in the highest Com1 numbers. These compare to 11,200 sequence reads generated in the KSU data set using a total MiSeq approach.

### **Determination of Sequencing Reproducibility**

To assess the repeatability of the amplified Com1 and Com2 data sets, we examined the correlation of V-gene segment usage. In the C57Bl/6 mouse, the V-gene segment is the most varied in the heavy chain (IgH locus) comprising a total of 133 functional V-gene segments and alleles compared to 19 for the D-gene segment and four for the J-gene (2017, 2017). The nucleotide sequences in the V-gene segments are also highly varied and require a complex multiplex PCR to amplify. Correlations were assessed using the data provided by the commercial vendor's proprietary bioinformatics.

Non-strain specific V-gene segment assignments accounted for between 1.06% and 2.16% of the sequencing results from Com 1 and 4.4% and 3.1% for Com2 (Table 3.2). Although there were differences in the immunoglobulin gene transcripts that were detected, there was a high correlation in the V-gene sequences detected among the different technical replicates in the Com1 data ( $R^2$  range from 0.7159 to 0.9939, all  $p < 0.0001$ ) (Figure 3.1). The  $R^2$  between technical replicates in Com2 was 0.9644 ( $p < 0.0001$ ). Therefore, although total transcripts generated varied with sample preparation, the V-gene segments that were detected were consistently detected using two different commercial approaches.

## Impact of Amplification on V-gene Segment Detection

The unamplified KSU approach produced a data set where a total of 131 V-gene segments were detected while the Com1 data sets contained between 86 and 107 V-gene segments. The Com2 data sets each contained 72 detectable V-gene segments. We compared results from all V-gene segments detected in the Com1 data set to the unamplified KSU data set.  $R^2$  were moderate ranging from 0.4648 to 0.6154 (all  $p < 0.0001$ ) (Table 3.3). The Com2 data sets also showed moderate  $R^2$  of 0.5615 for mRNA and 0.5574 for tRNA (all  $p < 0.0001$ ). To determine the differences in V-gene detection and the causes of these lower correlations, we compared the results from the most commonly detected V-gene segments in the KSU data sets to their frequencies in the Com1 and Com2 data sets. The protocol for Illumina sequencing uses mRNA selection, SuperScriptIII reverse transcriptase, and random hexamer primers. To best compare results, we used the Com1 mRNA-MMLV-Hex data set and the Com2 mRNA data set using the top 36 V-gene segments in the KSU data set. These V-gene segments comprise over one percent of the detected repertoire and are considered highly expressed. In the Com1 data set, of these 36 highly expressed V-gene segments, nine (V2-06, V1-26, V1-18, V1-50, V4-01, V1S95, V1S92, V1S108, V1-52) were detected at two-fold lower frequency than in the KSU data set (Figure 3.2). These nine V-gene segments were either absent or were near zero percent of the repertoire. Of these same top 36 V-gene segments in the Com1 data set, three (V6-3, V5-17, V11-2) were detected at two-fold greater than the KSU data set. The  $R^2$  for these top 36 V-gene segments to the KSU data set was 0.2339  $P = 0.0028$ ).

We also compared the Com2 mRNA data set to the same 36 V-gene segments from the unamplified KSU data. Thirteen of the 36 V-gene sequences (V2-6, V9-3, V4-1, V1S95, V1S92,

V1-81, V8-8, V2-2, V5-17, V1S108, V11-2, V1-52, V9-1) were detected at a twofold or lower level than at the KSU data set (Figure 3.2). Of those 13, eight (V2-6, V4-1, V1S95, V1S92, V1-81, V1S108, V11-2, V1-52) were at or near zero percent of the repertoire (Figure 3.2). Six other V-gene segments (V1-53, V3-6, V1-64, V1-76, V1-69, V10-1) were detected at twofold higher levels than those found in the KSU data set (Figure 3.2). The correlation of these top 36 V-gene segments was also poor at 0.2464 ( $P=0.0021$ ).

### **Direct Comparisons of Amplified and Unamplified Data Sets**

The comparisons in V-gene use were made using the bioinformatics provided by the commercial venture. To standardize the data handling to remove bioinformatic reasons for the differences in data, we processed the sequencing results from the Com1 mRNA-MMLV-Hex and Com2 mRNA data sets using the KSU bioinformatics work flow.<sup>45</sup> Briefly, the raw commercial sequencing results were submitted to IMGT for identification and were further analyzed using the KSU pipeline.

The KSU bioinformatic treatment of the Com1 data set correlated moderately with the commercially provided bioinformatics ( $R^2=0.5054$ ,  $p<0.0001$ ). After processing the Com1 data with the KSU bioinformatics pipeline, the  $R^2$  to the KSU data set decreased from 0.5728 (Table 3.3) with the original bioinformatics to 0.4582 ( $p<0.0001$ ) with the adjusted bioinformatics. However, 14 V-gene segments were detected in the Com1 data set using the KSU bioinformatics workflow that were not originally detected using the commercially provided bioinformatics (Figure A.1). When we processed the Com2 data using the KSU bioinformatic pipeline, the Com2 data set was highly correlated with the original commercially provided bioinformatics

treatment ( $R^2=0.9997$ ,  $p<0.0001$ ). When we compared Com2 data set processed with the KSU bioinformatics pipeline to the KSU RNASeq data set, the data still only had an  $R^2= 0.5620$  ( $p<0.0001$ ). The KSU bioinformatics workflow detected an additional 26 V-gene segments that were not detected by the commercial bioinformatics (Sup Fig 1).

When we reanalyzed the bioinformatics data from Com1 and Com2 using the KSU pipeline, we detected gene segments that were not detected in the original commercially provided bioinformatics. However, the inclusion of these gene segments, did not improve the comparability between the amplified data sets and the KSU RNASeq data. In the Com1 data set, some undetected gene segments (V1-18, V1-50, and V1-34) were not detected in the original bioinformatics but were detected at high levels ( $>1\%$ ) in the KSU/IMGT processed data (Sup Fig 1). The additional V-gene segments detected in the Com2 data set were found in less than  $<0.05\%$  of the repertoire (Sup Fig 1). These changes were not sufficient to improve  $R^2$  values.

### **Impact of Amplification on the Reproducibility of CDR3 Detection**

The comparison of two different amplified data sets (Com1 and Com2) to the unamplified KSU RNASeq data revealed that some V-gene segments were not detected in the amplified data sets. This is a concern since that precludes a complete picture of the V-gene repertoire. Nevertheless, amplified sequencing of the antibody repertoire is thought to provide an advantage in that the depth of coverage is increased over unamplified data sets due to the number of reads generated. To determine how extensive the discrepancy is between amplified and unamplified data we assessed the read depth (number of reads generated) and resampling efficiency of CDR3 (number of unique CDR3s resampled between replicates) using technical

replicates of samples sequenced with the various sequencing techniques. As expected, amplified data sets had both higher total read numbers and unique CDR3 numbers (Table 3.4).

Resampling/reproducibility has been assumed to improve with the depth of coverage. We had the unique opportunity to compare sequencing data from the same biological material subjected to multiple sequencing methodologies. We have also had the opportunity to do technical replicates on multiple samples subjected to RNASeq or amplification procedures. This allowed us the ability to look at CDR3 sampling reproducibility and to determine if amplification provided any advantage in CDR3 reproducibility. For the KSU unamplified data set, two C57BL/6J mouse spleen RNA samples (#32 and #39) were sequenced independently two times each and the CDR3s sampled were compared. In the KSU data set, 32-1 shared 28% of its total unique reads with 32-2. (Figure 3.3). 32-2 shared 24% of its reads with 32-1 (Figure 3.3). KSU data set, 39, showed 25% overlap of their total unique reads between each sampling (Figure 3.3). For the Com2 data, since there was such a strong correlation between the sequence output between mRNA and tRNA samples of C57BL/6J spleen samples ( $R^2 = 0.9644$ ), we considered these technical replicates. The mRNA data set shared 24% of its sequences with the tRNA data set while the tRNA data set shared 30% of its sequences with the mRNA data set. We also examined the resampling efficiency in the Com1 data set using the spleen mRNA-MMLV data sets that were reverse transcribed with two different primers, random hexamer and oligo-dT. Although this was not a perfect technical replicate, there was an  $R^2$  of almost 0.94 in V-gene segments detected (Figure 3.1). Therefore, we felt these served as “incipient” technical replicates. The random hexamer data set shared 20% of the detected CDR3s and the oligo dT

data set shared 32% of its CDR3 sequences. Therefore, regardless of data set, there was an average of 26-27% sampling overlap regardless of whether amplification was performed or not.

We also assessed the overlap in the detected CDR3s between the Com1 data set (mRNA-MMLV-Hex), Com2 data set (mRNA), and the KSU original data set to determine the extent of the overlap of CDR3 detection using the different methods. From the 295,116 CDR3 unique sequences that were detected, 2662 of those sequences were shared among all three data sets (Figure 3.4). The amplified data sets from Com1 and Com2 shared the most CDR3 sequences between them with 34,141 sequences found in both data sets (Figure 3.4). The KSU data set shared 59% of its detected CDR3 sequences with the Com1 and Com2 data sets (Figure 3.4). These data are consistent with the lower depth of sequencing of the unamplified data set compared to the Com1 and Com2 data sets where 19-32% overlap occurred in detected CDR3 sequences.

### **Detection of High Frequency CDR3s**

To gauge whether the highest frequency CDR3s can be detected by the different techniques, we assessed the 25 highest frequency CDR3s from each sequencing method. This resulted in a total of 48 unique CDR3s from the three different methods (Figure 3.5). The KSU data set detected all but one (CARGYFDVW) of these 48 sequences, the Com1 data set failed to detect four sequences (CARGTYW, CTWDEGNYW, CARGIYW, CARGSYW) and the Com2 data set detected all 48 sequences (Figure 3.5). The CDR3s that were not detected in the Com1 data set, did use V-gene segments that were detected in the data set. These data show that

although the depth of sequencing of the KSU data set was about 10% of the amplified data sets, the data set still captured most of the highest frequency CDR3's.

### **CDR3 Frequency Assessment**

While the high frequency CDR3s were shared among at least two sequencing runs, most of the CDR3 sequences detected were unique to each sequencing run (Figure 3.4). To determine the frequency of unique CDR3 sequences, we compared the most frequent, least frequent, and average percent of the measured CDR3 repertoire (Table 3.5). The highest frequency CDR3 was detected at 4.16%, 0.22%, 2.26%, and of the repertoire for the KSU, Com1, and Com2 data sets, respectively (Table 3.5). The lowest frequency CDR3s, representing only a single detected read, were 0.0088%, 0.0004%, and 0.0002%, of the repertoire for the KSU, Com1, and Com2 data sets, respectively (Table 3.5). The average detection level for the KSU data set was 0.015%, 0.0006% for the Com1 data set, and 0.0007% for the Com 2 data set (Table 3.5).

We also examined the frequency of CDR3s that were unique to each data set. Overall, the maximum and the average frequencies of the data sets were reduced (Table 3.5). This demonstrates that the unique reads in each data set were most likely transcripts from low frequency B cells. Moreover, these data suggest that even without amplification, the KSU data set was detecting the most prevalent CDR3s and many low-frequency sequences.

## **Discussion**

We have demonstrated that Illumina sequencing of total RNA from mouse spleen is able to capture a representative sample of the splenic B cell repertoire. This snapshot of the repertoire,

while producing less reads than amplified data sets, detects more V-gene segments than data sets that used two different amplification strategies and captures most of the high frequency CDR3s found in the amplified data sets. While amplified data sets provide more CDR3 depth of coverage, the unamplified data sets produced from an RNASeq allow for further data mining, eliminate as much primer bias as possible and provides an accurate representation of the repertoire.

Sequencing requirements of the B cell receptor are more challenging than those of the T cell receptor. There are no consensus sequences to reference.<sup>123</sup> Therefore, avoiding bias is one of the main priorities for antibody repertoire sequencing.<sup>130</sup> PCR errors are accumulated through the amplification process which can falsely inflate the repertoire or they can add suspected mutations that do not exist<sup>123,124,131,132</sup> and they may not be distinguishable from low level mutations that actually do exist in the repertoire.<sup>132,133</sup> PCR biases can be introduced because of primer binding properties, CG content, mispriming, non-specific binding, and errors during replication.<sup>130,134-136</sup> A specific issue in targeting antibody gene segments is primer annealing efficiencies since the gene segments that make up the murine IgH locus are similar, though not identical.<sup>135</sup> The biases inherent in the multiplex PCR can lead to false repertoire skewing, gene frequency inaccuracies, and a less comprehensive view of the repertoire.<sup>118,122,124</sup> The development of these multiplex primers, is highly challenging.<sup>120</sup>

Our results demonstrate some of the issues of assessing B-cell repertoires using massively multiplexed PCR reactions. While reproducibility for technical replicates was moderate to high, the correlations of these data sets to the unamplified KSU data sets were low to



moderate, even when the same bioinformatics processing was employed. Of significant concern is the Com1 set failed to detect 26 V-gene segments, and the Com2 set failed to detect 59 V-gene segments that were detected by the unamplified KSU RNASeq data set. The absence of these genes draws into question how one compares the various data sets with correlation coefficients that are below 0.3. Although Carlson et al. argues that amplification methodologies can capture the entire repertoire, there are concerns.<sup>126</sup> Even when we only looked at the V-gene families detected at the highest frequency there were omissions. Of the 36 V-genes that we categorized as “high frequency” (>1% of the repertoire), the Com1 data set failed to detect nine V genes, and the Com2 data set failed to detect 13 V genes. These results suggest that those methods are providing a false snapshot of the repertoire by missing those high frequency V-gene segments. Interestingly, three of these gene segments that we detect that are missing from the amplified data (V1S95, V1S92, V1S108) are identified by IMGT as genomic DNA, but not known to be rearranged.<sup>254</sup> Therefore, the inclusion of these V-gene segments in our analysis is debatable. We have detected them in this and other works,<sup>45,58,195</sup> at frequencies above one-percent in many samples and feel justified in their inclusion in our analysis. The Illumina sequencing protocols also rely on oligo-dT selection, showing that these transcripts have poly-A tails. Therefore, they appear to be functional messages. Even if these three genes are excluded from the analysis, the Com1 assessment failed to detect 33 V-gene sequences, and Com2 failed to detect 10. Our laboratory has previously examined the recapture of V-gene segments in multiple sequencings in our laboratory.<sup>45</sup> We sequenced two technical replicates of the same RNA and obtained a  $R^2$  of 0.9645 ( $P < 0.0001$ ) for V-gene segment usage.<sup>45</sup> This demonstrates that our workflow and sequencing methodology show high levels of repeatability.

The failure of primers to capture specific V-genes is not a new discovery, as primers failed to adequately sequence hybridomas previously.<sup>137,138</sup> “Universal” primers for the human antibody repertoire do exist, but some questions remain if they cover the entire repertoire.<sup>139</sup> The difficulty in developing a “universal” or even highly comprehensive primer set for the mouse is likely due to their highly varied leader sequences, V-gene segments, and framework regions. Primer design would have to rely on massively multiplexed reactions and/or degenerate primers. Additionally, the most commercially viable amplification methods would need to amplify across multiple common strains adding additional levels of complexity. Indeed, in our attempts to design “universal” primers, we found a minimum of 11 primer sets would be needed to detect V-genes associated for each isotype. Even then, there were still issues with matching PCR conditions and efficiency. Methods to overcome the biases detected in amplification have been developed, such as the use of 5’ RACE<sup>119,133,140</sup> and using molecular barcodes or identifiers.<sup>141,142</sup> However, these methods are expensive and have their own drawbacks. Replication of the entire repertoire using 5’RACE would still require the use of multiple constant region primers, leading to the same multiplexing issues. Barcoding can have errors and chimeric reads making repertoires difficult to reconstruct.<sup>143</sup> This latter issue is not a problem with our RNASeq data.

While bias exists in the massively multiplexed amplification process, there may be some sequencing errors in the unamplified KSU data set as well. While not specifically amplified for antibody sequences, random hexamers and oligo-dT capture beads are used prior to sequencing to generate the library<sup>255</sup> and some biases have been observed in random hexamer binding.<sup>256</sup> The use of oligo-dTs can result in enhancement of the 3’ end of transcripts.<sup>257</sup> We do not think this is particularly problematic since Illumina sequence methodology aims to reduce bias in their

library preparations by combining the random hexamer and oligo-dT capture. Additionally, while all libraries were sequenced on the Illumina platform, 3.19% of high-quality Illumina reads contain false base calls, which are impossible to differentiate using normal bioinformatics methods.<sup>123</sup> Over representation of specific dinucleotides can also be detected in sequencing which are not related to primer usage.<sup>258</sup> Therefore, although we hope to reduce bias and omission by doing RNASeq, we still have some technical issues that keep the data set from being a perfect reflection of the repertoire. Multiple technical replications help reduce the impact of this problem.

One analysis that was not pursued in this current investigation was the identification of clonally related sequences. We were specifically interested in the functional antibody repertoire present in the spleen and focused our analysis at the transcript level. We acknowledge that the overrepresentation of some sequences may be likely within our data sets since we don't use barcoding or clonality analysis to collapse similar mRNA sequences. Nevertheless, the overrepresentation of a specific sequence by an overly productive cell is also representative of cellular activation of transcription, and likely, functional antibody protein in the body.<sup>259,260</sup>

The lack of amplification also results in varying sequence lengths in our data set. We selected 40nt as our minimum cutoff to provide us enough information to detect V-, D-, and J-gene segments. While some short sequences were included in the data analysis, our overall average sequence length was 287nt, with productive sequences averaging 331nt and unknown sequences averaging 270 nt (data not shown). Overall, less than 0.5% of the sequences analyzed

where less than 100nt long (data not shown). Therefore, we do not think that sequence length is an issue in this study.

Although there may be issues with antibody repertoires generated using massively multiplexed PCR reactions, there are some advantages that may overshadow the disadvantages. For example, there is increased sequencing depth (e.g. 1,260,000 reads vs. 11,200 complete reads), more low frequency CDR3 detection (20-fold more unique CDR3 sequence than the unamplified KSU data set), and sequencing costs can be lower than an Illumina MiSeq run (by integrating multiplexing/barcoding).

Although amplification provides more detail in the CDR3 repertoire, if one is interested in the B cell clones that are most prevalent, then RNASeq does not appear to be at a disadvantage. CDR3 resampling was similar (20-32%) regardless of method. Additionally, when examining high frequency CDR3s, the unamplified data set only failed to detect a single unique CDR3 sequence, while the Com1 data set failed to detect four.

When preparing for antibody repertoire sequencing, multiple factors must be considered within the framework of the specific biological questions being asked. This includes needed coverage, cost, and starting material.<sup>119</sup> Additionally, it is important to consider that all repertoire sequencing is merely a snapshot of a constantly shifting image.<sup>124</sup> We will also never be able to fully capture the full diversity of the B cell immunoglobulin repertoire, which is estimated to range from  $10^6$ - $10^7$  possible unique rearrangements and mutations<sup>119,124</sup> to as much as  $10^{13}$ .<sup>44</sup> The failure of the KSU data set to detect rare clones compared to the amplified data sets is likely due

to this; but even the amplified data sets only sampled a fraction of the total repertoire. Therefore, one must decide how “deep” is adequate for the question being addressed.

During this investigation, we also had to address the issue that starting material may influence the quality of one’s sequencing. mRNA as a starting template increased reads up to two orders of magnitude. Furthermore, random hexamer primers and MMLV reverse transcriptase generally yielded higher read count results. Interestingly, the use of mRNA, with MMLV reverse transcriptase and hexanucleotide primers is most technically like that used in Illumina sequencing. However, additional data will be needed to confirm our observations since we did not pursue this aspect of the study in detail and the replicate number did not allow for robust statistical comparisons.

In conclusion, we have demonstrated that sequencing of unamplified splenic RNA provides a realistic snapshot of the total splenic B cell repertoire. We also have demonstrated that a good understanding of the bioinformatics work flow and reporting of the methodology is critical and cannot be understated. We understand that there are cellular biases and transcript stability differences within B cell subpopulations.<sup>25,26</sup> However, for the purpose of assessing a whole tissue B cell repertoire, unamplified RNASeq can provide a glimpse of the most prevalent B cell clones. The unamplified approach could just as well be applied to specific cell populations when the application required it. Moreover, an unamplified data set may detect V-gene segments that amplified data sets miss.

## Figures and Tables

Figure 3.1.  $R^2$  values of sequencing technical replicates.

		Com 1							Com2
		mRNA AMV Hex	mRNA MMLV dT	mRNA MMLV Hex	Total AMV dT	Total AMV Hex	Total MMLV dT	Total MMLV Hex	mRNA
Com 1	mRNA AMV dT	0.9939	0.9244	0.9582	0.8941	0.9396	0.9640	0.9065	
	mRNA AMV Hex		0.9183	0.9607	0.8960	0.9443	0.9639	0.9122	
	mRNA MMLV dT			0.9377	0.8615	0.8230	0.9313	0.8186	
	mRNA MMLV Hex				0.8158	0.8923	0.9673	0.9427	
	Total AMV dT					0.8677	0.8386	0.7166	
	Total AMV Hex						0.9136	0.8749	
	Total MMLV dT							0.9301	
Com2	TRNA								0.9970

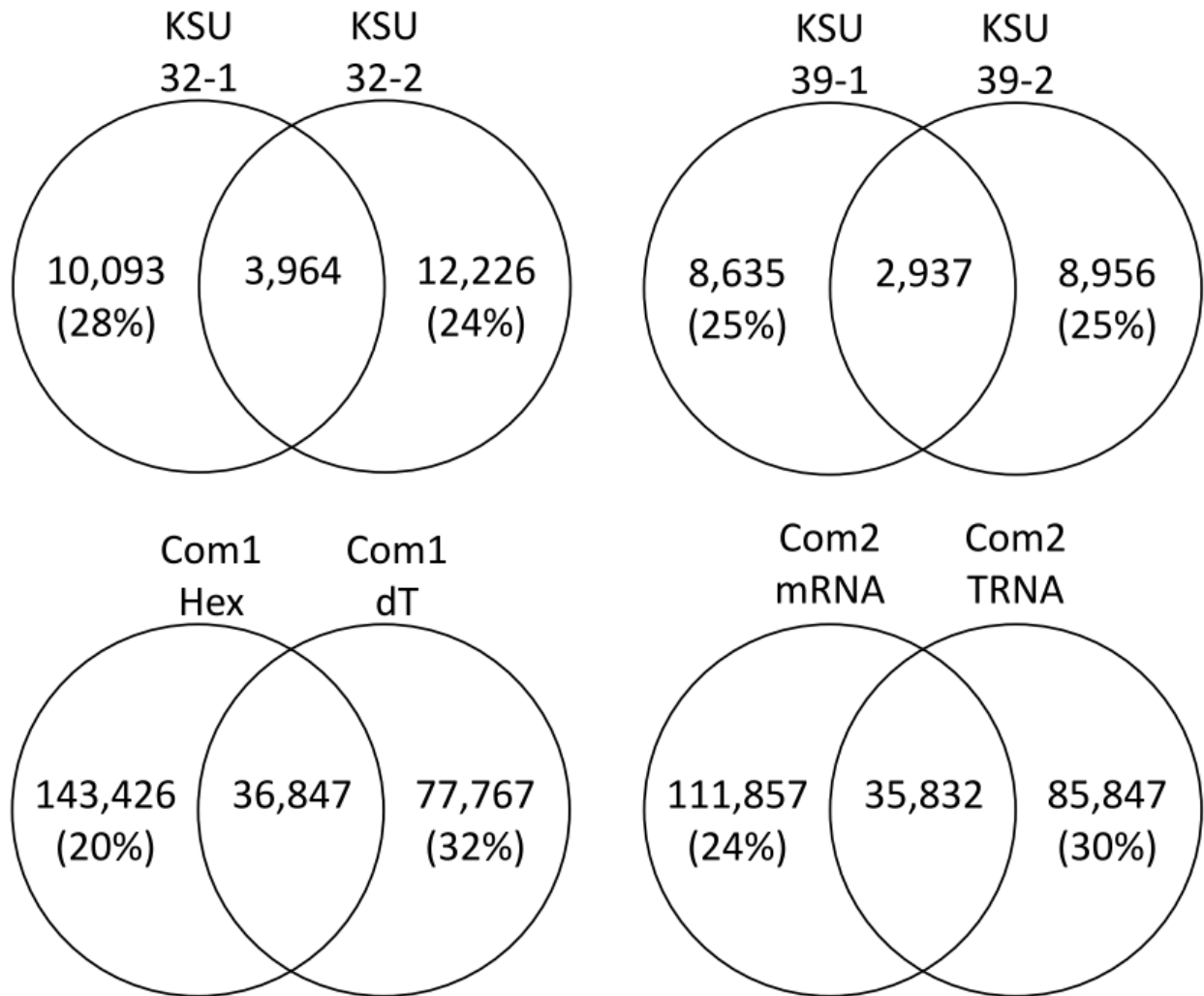
The percent of repertoire for each V-gene segment detected was compared between technical replicates. The highest  $R^2$  are dark red, the lowest  $R^2$  are blue.

**Figure 3.2. Percent of repertoire for high frequency V-gene segments among data sets.**

	KSU	Com1	Com2
IGHV1-80	6.2	6.1	6.3
IGHV2-6	3.7	1.0	
IGHV2-6-8	3.7	4.5	1.9
IGHV1-26	3.4	0.0	6.2
IGHV1-9	3.0	4.9	5.0
IGHV1-53	2.8	4.0	6.4
IGHV9-3	2.7	2.8	0.9
IGHV1-18	2.6		2.1
IGHV1-55	2.4	2.0	4.2
IGHV1-50	2.4		2.0
IGHV6-3	2.4	5.7	1.7
IGHV3-6	2.0	3.1	4.9
IGHV1-64	1.9	2.8	4.6
IGHV1-76	1.9	3.1	4.0
IGHV6-6	1.9	3.3	1.7
IGHV4-1	1.8	0.0	
IGHV1-82	1.7	2.0	1.7
IGHV1S95	1.7		
IGHV1S92	1.6		
IGHV1-81	1.6	2.1	
IGHV1-72	1.5	2.3	2.3
IGHV8-8	1.5	2.5	0.7
IGHV2-2	1.5	2.0	0.7
IGHV5-17	1.5	3.0	0.6
IGHV1-22	1.5	1.9	2.4
IGHV1-69	1.5	1.0	3.4
IGHV1S108	1.5		
IGHV7-3	1.4	1.1	2.6
IGHV10-1	1.4	2.2	4.2
IGHV10-3	1.3	1.5	1.7
IGHV2-3	1.2	2.0	2.3
IGHV14-4	1.2	1.2	1.8
IGHV11-2	1.2	2.5	0.1
IGHV1-52	1.1		
IGHV9-1	1.1	0.9	0.4
IGHV1-78	1.0	1.1	1.2

Percent of repertoire for the KSU, Com1 (mRNA-MMLV-hex), and Com2 (mRNA) are displayed. The highest value percent of repertoire is dark red while the lowest are white. Black boxes represent no detected reads (true zero). Rounded zeros are represented as 0.0.

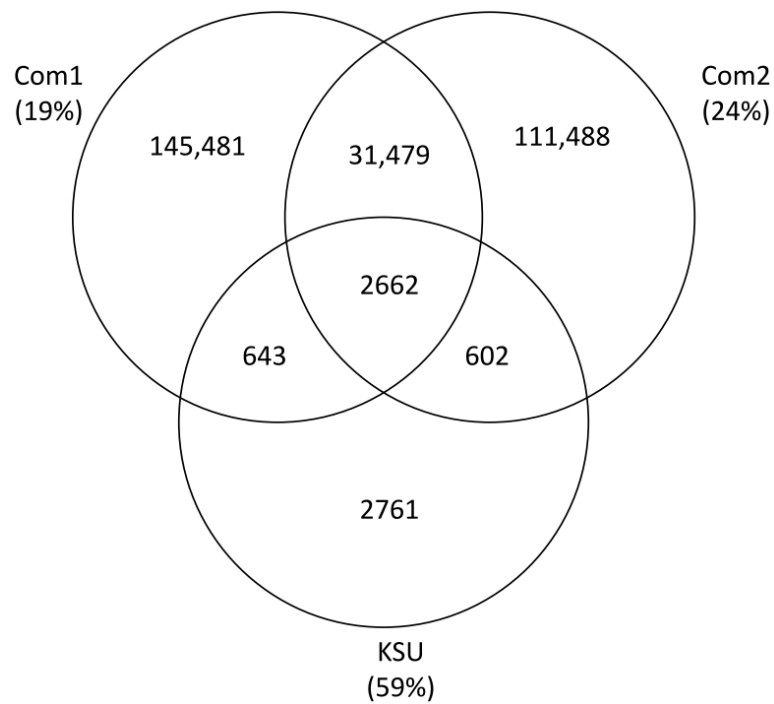
**Figure 3.3. Overlap of CDR3 sequence detection between technical replicates.**



CDR3 amino acid sequences were compared between technical replicates. Sequences unique to one data set are displayed in the outer circles. Sequences shared between data sets are in the overlap. Percent of shared CDR3 sequences is displayed in parentheses in the outer circles. (A) KSU data sets 32-1 and 32-2. (B) KSU data sets 39-1 and 39-2. (C) Com1 data sets mRNA-MMLV-Hex and mRNA-MMLV-dT. (D) Com2 data sets mRNA and TRNA.



**Figure 3.4. CDR3 sequence capture among Com1, Com2, and KSU data sets.**



CDR3 amino acid sequences were compared among the Com1 mRNA-MMLV-Hex, Com2 mRNA, and the KSU data sets. Percent of the repertoire shared with at least one other data set is listed in parentheses.

**Figure 3.5. High frequency CDR3s detected among the Com1, Com2, and KSU data sets.**

	KSU	Com1	Com2
CASVYDGYAFAYW	4.2	0.1	1.3
CARGAYW	3.0	0.2	2.3
CAREYDGYPYAMDYW	1.5	0.1	1.3
CARRWLHYAMDYW	0.9	0.1	1.1
CARGGYW	0.8	0.0	0.6
CARDYYGSSWYFDVW	0.7	0.1	0.9
CARTIYYGSSWFAYW	0.7	0.1	0.2
CAREELPHYFDYW	0.5	0.0	0.0
CARGGYDGYPHYAMDYW	0.5	0.0	0.1
CARGYYFDYW	0.5	0.0	0.4
CAKDYYGSSWYFDVW	0.4	0.1	0.5
CMRYSNYWYFDVW	0.4	0.0	0.0
CAREEGYYDWFAYW	0.4	0.0	0.0
CMRYSSYWYFDVW	0.4	0.0	0.0
CMRYGNYWYFDVW	0.3	0.0	0.0
CARRAHYYGSSYYFDYW	0.3	0.0	0.1
CASYSNYDYW	0.3	0.0	0.3
CARGTYW	0.3		0.4
CARWRVYYGNWYFDVW	0.3	0.0	0.0
CARDNWDWYFDVW	0.3	0.0	0.1
CARLYYYGSHWYFDVW	0.3	0.0	0.1
CARRDSNYLDYW	0.3	0.0	0.2
CASNAYYSNYVTPHFDYW	0.2	0.0	0.3
CARWYYGSSYEGYFDVW	0.2	0.0	0.6

	KSU	Com1	Com2
CTWDEGNYW	0.2		0.3
CASWDFAYW	0.2	0.1	0.2
CARSGGNWYFDVW	0.1	0.0	0.2
CASWEFAYW	0.2	0.0	0.1
CTRFAYW	0.2	0.0	0.1
CARGDYW	0.2	0.0	0.2
CARDDPFAYW	0.2	0.0	0.1
CARGGYDYDWYFDVW	0.2	0.0	0.0
CARGGFAYW	0.1	0.0	0.0
CARSHYSAWFAYW	0.1	0.0	0.1
CARGFAYW	0.0	0.0	0.0
CARGYFDVW		0.0	0.0
CATYGSPFAYW	0.1	0.0	0.0
CARGGYSDYDWYFDVW	0.2	0.0	0.0
CARDYGSSYDSYW	0.1	0.0	0.1
CARSQNYWFAYW	0.1	0.0	0.1
CAKAPIYYDYDGVFYYAMDYW	0.1	0.0	0.4
CARSPDGYAMDYW	0.2	0.0	0.3
CARSGAYYRDYYAMDYW	0.2	0.0	0.3
CARDGSHAMDYW	0.1	0.0	0.3
CARGIYW	0.2		0.2
CARQLYAMDYW	0.1	0.0	0.2
CARGSYW	0.2		0.2
CARRLDYW	0.0	0.0	0.2

The top 25 CDR3s from each data set (48 total) were compiled and percent of repertoire compared. Black boxes represent no detected reads (true zero). Rounded zeros are represented as 0.0.

**Table 3.1. Comparison of total productive reads among data sets.**

	KSU <sup>a</sup>	Com 1 <sup>a</sup>								Com 2 <sup>a</sup>	
		mRNA <sup>b</sup>				TRNA <sup>b</sup>				mRNA <sup>b</sup>	TRNA <sup>b</sup>
		AMV <sup>c</sup>		MMLV <sup>c</sup>		AMV <sup>c</sup>		MMLV <sup>c</sup>			
		dT <sup>d</sup>	Hex <sup>d</sup>	dT <sup>d</sup>	Hex <sup>d</sup>	dT <sup>d</sup>	Hex <sup>d</sup>	dT <sup>d</sup>	Hex <sup>d</sup>		
Total Productive Reads	11,200 <sup>e</sup>	553,521	1,263,003	883,532	1,035,461	7,975	6,867	208,979	220,772	637,214	766,075

a – Sequencing technique (Com1 and Com2 are amplified data sets)

b – starting material (mRNA – Messenger RNA, TRNA – Total RNA)

c – reverse transcriptase (AMV - Avian Myeloblastosis Virus, MMLV - Moloney Murine Leukemia Virus)

d – primer (dt – Oligo dT, Hex – Random hexamer)

e – an additional 27,896 reads were used for V-gene segment usage assessment. These sequences were not long enough for CDR3 detection.

**Table 3.2. Percent of reads assigned to non-C57BL/6 (B6) or pseudogene V-gene segments**

	Com 1 <sup>a</sup>								Com 2 <sup>a</sup>	
	mRNA <sup>b</sup>				TRNA <sup>b</sup>				mRNA <sup>b</sup>	TRNA <sup>b</sup>
	AMV <sup>c</sup>		MMLV <sup>c</sup>		AMV <sup>c</sup>		MMLV <sup>c</sup>			
	dT <sup>d</sup>	Hex <sup>d</sup>	dT <sup>d</sup>	Hex <sup>d</sup>	dT <sup>d</sup>	Hex <sup>d</sup>	dT <sup>d</sup>	Hex <sup>d</sup>		
% Non-B6 V-Gene segments	1.63	1.64	1.83	2.00	1.44	1.06	1.99	2.16	4.4	3.1

a – Sequencing technique (Com1 and Com2 are amplified data sets)

b – starting material (mRNA – Messenger RNA, TRNA – Total RNA)

c – reverse transcriptase (AMV - Avian Myeloblastosis Virus, MMLV - Moloney Murine Leukemia Virus)

d – primer (dt – Oligo dT, Hex – Random hexamer)

**Table 3.3. R<sup>2</sup> of V-gene segment usage to the KSU data set with read counts.**

	Com 1a								Com 2a	
	mRNAb				TRNAb				mRNAb	TRNAb
	AMVc		MMLVc		AMVc		MMLVc			
	dTd	Hexd	dTd	Hexd	dTd	Hexd	dTd	Hexd		
R <sup>2</sup> to KSU Dataset	0.5694	0.5733	0.4648	0.5728	0.4709	0.5710	0.5573	0.6154	0.5615	0.5574
Read Count	553,521	1,263,003	883,532	1,035,461	7,975	6,867	208,979	220,772	637,214	766,075

a – Sequencing technique (Com1 and Com2 are amplified data sets)

b – starting material (mRNA – Messenger RNA, TRNA – Total RNA)

c – reverse transcriptase (AMV - Avian Myeloblastosis Virus, MMLV - Moloney Murine Leukemia Virus)

d – primer (dt – Oligo dT, Hex – Random hexamer)

**Table 3.4. Comparison of read count to unique CDR3 amino acid sequences**

	KSU	mRNA-MMLV-Hex (Com1)	mRNA (Com2)
Read Count	11,200	1,035,461	637,214
Unique CDR3 Sequences	6668	180,266	146,231

**Table 3.5 Comparison of CDR3 frequencies in the whole repertoire and the unique repertoire.**

	Whole Repertoire			Unique Repertoire		
	KSU	Com1	Com2	KSU	Com1	Com2
Minimum	0.008758	0.000446	0.000221	0.008758	0.000446	0.000221
Maximum	4.1655	0.216969	2.259680	0.035032	0.005804	0.026044
Average	0.014997	0.000555	0.00684	0.008929	0.000478	0.00290

## Chapter 4 - Characterization of the naïve murine antibody repertoire using unamplified high-throughput sequencing

### Abstract

Antibody specificity and diversity are generated through the enzymatic splicing of genomic gene segments within each B cell. Antibodies are heterodimers of heavy- and light-chains encoded on separate loci. We studied the antibody repertoire from pooled, splenic tissue of unimmunized, adult female C57BL/6J mice, using high-throughput sequencing (HTS) without amplification of antibody transcripts. We recovered over 90,000 heavy-chain and over 135,000 light-chain immunoglobulin sequences. Individual V-, D-, and J-gene segment usage was uniform among the three mouse pools, particularly in highly abundant gene segments, with low frequency V-gene segments not being detected in all pools. Despite the similar usage of individual gene segments, the repertoire of individual B-cell CDR3 amino acid sequences in each mouse pool was highly varied, affirming the combinatorial diversity in the B-cell pool that has been previously demonstrated. There also was some skewing in the V-gene segments that were detected depending on chromosomal location. This study presents a unique, non-primer biased glimpse of the conventionally housed, unimmunized antibody repertoire of the C57BL6/J mouse.

**Citation:** Rettig, T. A.,\* Ward, C.,\* Bye, B. A., Pecaut, M. J. & Chapes, S. K. Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. *PLoS One* **13**, e0190982, doi:10.1371/journal.pone.0190982 (2018). \*Co-First Authors.



## Introduction

B cells are an important part of the adaptive immune system, arising from hematopoietic stem cell precursors. These cells express surface immunoglobulin (Ig) receptors and secrete these same proteins as antibodies into the serum after differentiation into plasma cells.<sup>8,261</sup>

As B cells develop, they rearrange Variable- (V), Diversity- (D), and Joining- (J) gene segments, which combine with a constant region to form the antibody structure.<sup>5,262</sup> Antibodies consist of heterodimers of heavy and light chains.<sup>262</sup> The heavy chain is formed from V-, D-, and J-gene segments combined with a constant region,<sup>263</sup> while light chains lack a D-gene segment.<sup>5,264</sup>

There are three complementarity determining regions (CDR). CDR1 and CDR2 are encoded in the V-gene segment. CDR3 consists of a combination of V-, (D-, heavy-chain), and J-gene segments.<sup>265</sup> Of the CDRs, CDR3 contributes the most to binding specificity. Antibodies are further characterized by the constant region, or isotype, which is influenced by the stage of B-cell development and antigen specificity.<sup>266</sup>

The total collection of antibody specificities present within an individual is known as the antibody repertoire. Diversity of the antibody repertoire results from four main components: the initial germ line (inherited), diversity from recombination of that germline, the imprecisions during V(D)J recombination, and somatic mutations.<sup>67,121,267</sup> The antibody repertoire has been examined in many studies by high-throughput sequencing (HTS) and fully mapped in the zebrafish.<sup>268</sup>

Repertoires can serve as a fingerprint or snapshot of the current immune-system status and these types of data have been used to explore the development of host defense to infectious disease,<sup>73,78,248,269-271</sup> cancer,<sup>115,128,272,273</sup> autoimmune disease,<sup>274,275</sup> and early disease detection.<sup>65</sup> With the development of HTS, we are now able to detect the differences between or among B-cell repertoires such as B2 (adaptive antibodies) and B1 (natural antibodies) B cells<sup>67</sup> or memory and naïve repertoires.<sup>199,276</sup> HTS has accelerated the characterization of the widely differing human Ig haplotypes,<sup>277-281</sup> and strain-specific gene segment usage in mice.<sup>14</sup>

Our long-term goals are to investigate the repertoire of B cells in mice in space and how it changes in response to antigen challenge. More specifically, our lab is interested antibody repertoire dynamics within the context of spaceflight. Due to the cost of these experiments, creating datasets that can be mined by our lab or others is important. The antibody repertoire is traditionally assessed through the amplification of Ig sequences that have been isolated from sorted B cell populations.<sup>133</sup> While these practices increase the likely hood of recovering rare Ig sequences and allow for the dissection of the antibody repertoire by B-cell populations, cell sorting may not be possible within the design of certain experiments. During the development of methodology to assess Ig-gene usage by mice subjected to space flight we performed multiple HTS runs to validate sample preparation, bioinformatic methodology, and reproducibility.<sup>45</sup> We also wanted some background data about the Ig repertoire in the normal B6 mouse population. Knowing that there can be significant mouse-to-mouse variability in the Ig repertoire<sup>44,58</sup> and to minimize the impact any one mouse might have in the validation data set, we chose to pool multiple mice specifically for these validation experiments.<sup>45</sup> We now present these data on the splenic repertoire of conventionally housed, unimmunized, unchallenged, adult C57BL/6J mice.

## **Materials and methods**

### **RNA extraction and sequencing**

Tissue extraction and sequencing were performed as described previously (Rettig 2017).<sup>45</sup> Briefly, spleens were collected from three independent pools of four, specific pathogen-free (based on the RADIL Advantage Basic profile), female, C57BL/6J mice nine-to-eleven weeks old. Animals were euthanized with isoflurane overdose followed by cervical dislocation. Briefly, mice were exposed to 400  $\mu$ L isoflurane in a gauze-pad enclosed in a histopathology cassette in a 450 mL chamber as was described by Huerkamp et al.<sup>282</sup> Animals used in pool one were raised in the Laboratory Animal Care Services (LACS) facility (breeder stock renewed less than 2 years previous) in the Division of Biology at Kansas State University. Mice from pools two and three were received directly from Jackson Laboratories and were housed in the LACS facility. Mice were fed LabDiet 5001 and had access to water and food ad libitum. Mice were maintained on a 12/12 light/dark cycle. Mice for pools two and three were allowed to acclimate in the vivarium for 22-31 days prior to sacrifice. Animal procedures were approved by the Institutional Animal Care and Use Committee at Kansas State University. After euthanasia, spleen tissue was processed immediately for RNA extraction with Trizol LS according to the manufacturer's instructions. Pool one contained RNA from one-half of the spleen tissue while pools two and three contained RNA extracted from complete spleens. Equal concentrations of total splenic RNA (RIN>8) were pooled and mixed from each of the four mice, resulting in three final pools. At least one microgram of RNA from each pool was submitted for size selection (275-800 nucleotides) and sequencing on Illumina MiSeq at 2x300 nucleotides as described previously (Rettig 2017). Sequencing was performed at the Kansas State University Integrated

Genomics Facility using the standard Illumina sequencing protocol, including oligo-dT-bead selection of mRNA and reverse transcription to cDNA. A reduced fragmentation time (one minute) was used to yield longer transcripts.<sup>45</sup> To avoid potential primer bias and maintain a dataset that could be further mined, we did not amplify Ig sequences. The authors note that a subset of mouse pool one data was used in a methods paper presented by our group. The context and focus of that previous manuscript does not overlap with the current work.<sup>45</sup>

## **Bioinformatics**

Sequence selection, mapping, and final processing was performed as outlined previously.<sup>45</sup> Briefly, sequences were imported into CLC Genomics Workbench v9.5.1 (<https://www.qiagenbioinformatics.com/>) and cleaned to obtain high quality reads using a quality score of 97% of the sequencing containing a Phred score over 20. Both paired and merged (overlapping pairs) were mapped to V-gene segments and the loci using a match score of +1 and a mismatch score of -2 to identify potential antibody sequences. The sequences were collected and submitted to ImMunoGenTics's (IMGT) High-V Quest for identification.<sup>283</sup> Due to the chances of collecting the same read multiple times through the mapping and identification process, only one sequence per sequence ID was analyzed (procedure outlined in Rettig et al).<sup>45</sup> Productive and unknown functionality sequences were identified via IMGT and used for subsequent analyses. Productive antibody sequences were defined as in frame and did not contain a stop codon, however binding abilities were not assessed. Unknown sequences did not contain enough sequencing information to determine functionality. Gene segments were identified using IMGT's nomenclature and using IMGT's list of gene segments, including functional and open reading frame-defined gene segments. We implemented two procedural

changes to further define the repertoire that are different from the Rettig et al. paper.<sup>45</sup> In this manuscript, when calculating the percent abundance in the repertoire, we also include V-gene segments where one or two possible V-gene segments were detected. When one single V-gene segment was detected in a sequence, it was assigned a value of one. When two potential V-gene segments were detected each gene segment was assigned a value of 0.5. The totals were then tabulated as described in Rettig et al.<sup>45</sup> Additionally, CDR3 sequences that did not fit the C-xx-W motif for IgH (heavy-chain) were reclassified as unknown functionality, unless a class-switched isotype (IgA, IgD, IgE) was detected. CDR3 sequences for Igκ (kappa-chain) that did not fit the C-xx-F motif were classified as unknown functionality. Sequencing reads containing hyperlengthy (greater than 2x the average, or 18 amino acids (AAs)) κ-CDR3 that fit the C-xx-F motif were also removed from analysis as we believed they were falsely identified through bioinformatic or sequencing errors.

Initial nucleotide alignments were created with MAFFT<sup>284</sup> using portions of the germline and CDR3 nucleotide sequences provided by IMGT. Sequences were sorted by identity, compared to the germline and the sequence order was then adjusted to group similarly-aligned sequences. Nucleotide sequences of identical length were then isolated from the full alignment and aligned with each other while retaining all previously-inserted gaps.

### **V(D)J pairing frequency**

Pairing frequency was assessed in productive sequencing reads from both IgH and Igκ datasets. All pairing of V-gene segments was only assessed from productive IgH and Igκ sequencing reads, referred to as VH and Vκ, respectively. Sequences identifying more than one

possible V-gene segment were excluded from this analysis. For both IgH and Igk, J- and D-gene segments designated as undetermined (U) were either not reported by IMGT, contained less than six nucleotides, or multiple gene segments were assigned to a single sequencing read. Total counts from VJ pairings for heavy- and light-chains were tabulated and Circos graphs were generated using Circos Online.<sup>285</sup>

### **Statistical analysis and graphics**

Linear Regressions were performed by comparing the percent of repertoire of V-gene segment or V(D)J combinations from pools 1 vs 2, 2 vs 3, and 1 vs 3 using the linear regression analysis tool in GraphPad (Version 6.0). Chi-square analysis of V-gene segment usage was performed on raw sequencing read counts using R version 3.4.2 (<https://www.r-project.org/>). All productive VH- and Vκ-gene segments and open reading frames listed on IMGT for the B6 mouse were analyzed and gene segments not found in our datasets were assigned a read count of zero. The IMGT productive list includes all gene segments detected in the NCBI annotation of the mouse genome. The IMGT chromosomal locations were also used for any chromosomal analysis, though not all gene segments have a defined chromosomal location. Gene segments that were not defined by location in IMGT were excluded from any chromosomal analysis. These analyses were performed on each mouse pool separately by comparing the observed raw-read count values of V-gene segments to an expected theoretical number of reads which was based on the null hypothesis that all V-gene segments will have the same number of raw reads. This value was determined by dividing the total number of antibody reads observed in a mouse pool by the number of possible gene segments. The analysis of gene-segment usage by chromosomal location was performed by dividing gene segments into four quadrants based on nucleotide

position. A quadrant was defined as one-fourth of the entire locus, as determined by number of nucleotides in the locus. Gene segment location was defined as the first nucleotide of the gene segment. A Chi-square analysis was performed on each mouse pool by summing the raw-read counts of all gene segments containing a 5' nucleotide position within each quadrant and comparing the observed total reads within a quadrant to an expected theoretical number of reads for each quadrant which was based on the null hypothesis that the number of raw reads is not statistically different between defined quadrants.

Percent of repertoire values were determined by dividing sequencing reads corresponding to each gene segment, constant region, or CDR3 length by the total number of gene segments, constant regions, or CDR identified in each mouse pool for normalized comparison between pools. Percent of repertoire for these variables were displayed as bar graphs, generated in GraphPad v6.0, or as heat maps, generated in Microsoft Excel. In addition to visualization of gene segment combinations through Circos graphs, the percent of repertoire for gene segment pairings was also visualized using the bubble chart tool in Microsoft Excel.

## **Results**

### **VH- and V $\kappa$ -gene segment usage**

We obtained between 8,714 and 11,200 IgH individual productive reads, and between 14,271 and 28,756 individual IgH reads of unknown functionality as identified by IMGT HighV-Quest (Table 4.1). Between 11,968 and 18,643 individual Ig $\kappa$  productive, and 12,602 and 39,410 individual Ig $\kappa$  reads of unknown functionality were identified by IMGT HighV-Quest (Table 4.1). Overall, we identified 132 VH- and 109 V $\kappa$ -gene segments within the repertoires of our

mouse pools (Figure A.2). As a general trend, the three pools resulted in similar frequencies and similar ranks for V-gene usage among groups. The ten most common V-gene segments from each pool were compiled, resulting in a total use of 14 VH-gene segments and 17 V $\kappa$ -gene segments (Figure 4.1).

The most common VH-gene segment in pools one and two was V1-80 (Figure 4.1A, Figure A.2A). V1-80 was the seventh most common gene segment used in pool three. V6-3 was the most common VH-gene segment in pool three, but ranked seventh and sixth in pools one and two, respectively (Figure 4.1B). V1-26 was the next most common VH-gene segment, ranking second in pools one and two, and third in pool three. Among the top ten most common VH-gene segments, most gene segments ranked between first and 17th within their pools, however, three outliers were found within these groups. V1-50 ranked 23rd in pool one, but it was ninth and second in pools two and three, respectively. V1-78 was ranked 31st in pool two, but ranked tenth and 17th in pools one and three, respectively. V1-18 was ranked 32nd in pool three but third and fourth in pools one and two, respectively. VH-gene usage between pools was well correlated (1 vs 2  $R^2 = 0.8427$ , 2 vs 3  $R^2 = 0.7054$ , 1 vs 3  $R^2 = 0.5842$ , all  $p < 0.0001$ ).

Seventeen V $\kappa$ -gene segments were among the top 10 most abundant V $\kappa$  of the repertoire in at least one of the three mouse pools, with five V $\kappa$ -gene segments appearing in the top ten of all three mouse pools (Figure 4.1, Figure A.2B). Pool one appeared enriched for V5-39, comprising 10.33% of the repertoire as compared to 1.03% in pool two and 2.84% in pool three (Figure 4.1C). Excluding this difference, V1-110, V1-117, V10-96, and V4-55 were the four most abundant gene segments in all three mouse pools. The lowest ranking of the most abundant



V $\kappa$  in any of the mouse pools were V5-39, ranking 32nd in pool two, while ranking first and seventh in pools one and three, respectively and V10-94, ranking 32nd in pool one, but fifth in pool two and 13th in pool three. (Figure 4.1D). All top V $\kappa$  that showed any variation in abundance among pools were still within the top 32 V $\kappa$ -gene segments. V $\kappa$  usage between pools was correlated, although not as highly as VH (1 vs 2  $R^2$ = 0.3684, 2 vs 3  $R^2$ =0.6700, 1 vs 3  $R^2$ =0.6414, all  $p$ =<0.0001).

With 113 productive VH and 93 productive V $\kappa$  described in IMGT by chromosomal location, each gene segment would be expected to appear as part of the repertoire roughly 0.84% and 1.06% of the time for VH and V $\kappa$ , respectively, if gene segment usage was random (Figure 4.2). When we assessed usage frequency, there was a non-random distribution of both VH- (Figure 4.2A) and V $\kappa$ - (Figure 4.2B) gene segments (Chi-square analyses; all pools  $p$ <0.0001).

We also assessed gene segment usage by chromosomal location in the IgH and Ig $\kappa$  loci as gene-segment spacing was not evenly distributed along the chromosome. Based on 5' nucleotide position, Q1-4 contained 22, 23, 33, and 33 VH-gene segments or 18, 21, 24, and 30 V $\kappa$ -gene segments, respectively. Since V-gene segment usage appears skewed, we tested whether the total expression of V-gene segments within a quadrant defined by nucleotide position was randomly distributed. We found that the expression within each quadrant was not randomly distributed, for both IgH and Ig $\kappa$ , suggesting that V-gene expression may be influenced by chromosomal location (all pools,  $p$ <0.0001).

### **DH-, JH-gene segment and IgH constant region usage**

We identified ten different D-gene segments used in our repertoires (Figure 4.3A). We also added one additional category for our analyses, termed “undetermined”. This label was applied to D-gene segments that were assigned by IMGT to non-C57BL/6J genes and antibody sequences containing a V- and J-gene segment, but not containing an identifiable D-gene segment. Due to the very short length of D-gene segments combined with alterations during recombination, D-gene segments were bioinformatically difficult to identify.

For all three groups, the D1-1 gene segment was the most common segment identified comprising 26-28% of the repertoire. Undetermined D-gene segments, however, made up a large part of the D-gene segment repertoire, comprising 31-33% of the repertoire. D2-3, D2-4, D4-1, and D2-5 were found in similar frequencies ranging from six to fifteen percent of the data set. D3-2, D3-1, D6-3, D5-5, D5-2, and D6-4 were found at low levels in all data sets; comprising under three percent of the total repertoire.

Four JH-gene segments were identified, with JH2 being the most common among all three groups (Figure 4.3B). The remaining J-gene segments, JH4, JH3, and JH1, were found at similar levels among groups totaling between 19% and 27% of the repertoire.

IgM was overwhelmingly the most commonly identified constant region making up between 78% and 84% of the total repertoire (Figure 4.3C). IgG was the next most common between seven and eleven percent of the total repertoire. IgA and IgD were rare totaling between

two and six percent of the repertoire. IgE was only detected in pool three at less than one percent. It was not detected in pools one and two.

### **J $\kappa$ -gene segment usage**

A total of four J $\kappa$ -gene segments were identified, with similar distribution of J $\kappa$ -gene segments between mouse pools (Fig 3-10D). Due to the even distribution of the three most abundant J $\kappa$ , the ranking of each gene segment varied slightly among the three mouse pools. Within each mouse pool, there was a small portion of J $\kappa$  that contained too few nucleotides to be assigned to a specific gene segment (0.03-0.38%).

### **IgH- and Ig $\kappa$ - gene segment combinations**

VH, DH, and JH family combination frequency was examined. Some preferential bias for specific gene segments seemed to exist (Figure 4.4A). For example, the JH4/DH2 combination appeared at a high frequency with VH1 (4.5% of repertoire), but not with any other VH gene family to the same degree. IgH gene segment recombination frequency correlated with gene segment abundance. VH1, which contains over half of all possible V-gene segments, also was the most commonly used VH family, which is seen as the dominant band in the Circos plot (Figure 4.4B).

The pairing of V $\kappa$  families to individual J $\kappa$  was also assessed (Figure 4.4C-D). Overall, the pairing of V $\kappa$  families with J $\kappa$  appeared random, however, certain V $\kappa$  families preferentially paired with specific J $\kappa$ -gene segments. For example, V4 paired less efficiently with J1, while V3 paired more efficiently with J1 (Figure 4.4C). Unlike VH, no single V $\kappa$  family was exceedingly dominant. Although V4 was the most represented gene family, its expression level was close to

that of the second next most prominent gene families, which varied by mouse pool as shown in the Circos plot (Figure 4.4D).

The percent of repertoire that each VJ-gene segment combination comprised within each mouse pool was compared by linear regression. Mouse pools showed modest correlation levels of VJ-gene segment recombination frequency in IgH (1 vs 2  $R^2 = 0.6055$ , 2 vs 3  $R^2 = 0.4419$ , 1 vs 3  $R^2 = 0.4399$ , all  $p < 0.0001$ ) and Igk (1 vs 2  $R^2 = 0.2340$ , 2 vs 3  $R^2 = 0.3598$ , 1 vs 3  $R^2 = 0.4607$ , all  $p < 0.0001$ ), with some enrichment for certain combinations within each mouse pool. V-(D)-J combinations were within bubble charts generate a visual comparison of pairing (Figure 4.4A and Figure 4.4C).

### **IgH and Igk CDR3**

The average IgH CDR3 (H-CRD3) length of all three data sets was 11 amino acid (AAs) long (Figure 4.5A). The lengths of the H-CDR3s followed a normal distribution except all three groups were enriched for five AA H-CDR3s. H-CDR3 AA length ranged from one to twenty-three amino acids in length with 11 AAs being the average for all three pools. Igk chain CDR3 length was conserved at nine AAs, comprising 87-90% of the repertoire (Figure 4.5B). While nine AAs was the most frequent  $\kappa$ -CDR3 length, one  $\kappa$ -CDR3 with a length of seven AAs was observed in the top CDR3 sequences. The distribution of CDR3 lengths was even among pools for both IgH and Igk (Figure A.3). Four  $\kappa$ -CDR3 sequences that fit the conserved kappa chain C-xx-F motif identified within the mouse pools (data not shown), while no such hyperlengthy H-CDR3 sequences fitting the C-xx-W motif were identified.

Interestingly, many of these H-CDR3s, though found in all three pools, were not necessarily common H-CDR3s. Only one H-CDR3, CARGAYW, was found among the top ten most common H-CDR3s of each pool. One additional H-CDR3, CARDYYGSSWYFDVW, was found in the top five of pools two and three. Of the 75 total H-CDR3s that appeared in all three pools, frequencies varied drastically, from being the most common to only being detected once (Figure 4.6A). Of the top five most common H-CDR3s in each data set, only three, CARGAYW, CARGGYW, and CMRYSNYWYFDVW occurred in all three data sets (Figure 4.6B, Figure A.4). CARGSYW occurred in pools one and two, CARRWLHYAMDYW in pools two and three, and CARYAPYYFDYW in pools one and three. The remaining most common H-CDR3s occurred in only one pool. A heatmap of all 75 shared H-CDR3 is shown in Figure A.4.

There were 3,217 total unique  $\kappa$ -CDR3 amino acid sequences identified among all three mouse pools (Figure 4.6C). While there were 2,345 individual  $\kappa$ -CDR3 amino acid sequences that were unique to the individual mouse pools, there were also 475  $\kappa$ -CDR3 shared among all three pools. None of these 475 shared  $\kappa$ -CDR3 were found within the top five  $\kappa$ -CDR3s of all three mouse pools (Figure 4-18D). There were 14 total  $\kappa$ -CDR3 sequences that were found within the top five  $\kappa$ -CDR3 for each mouse pool (Figure 4.6D, Figure A.4). There were no CDR3 sequences that were found within the top five  $\kappa$ -CDR3 in all three mouse pools.

### **Comparison of alignments of CDR3s**

To assess the heterogeneity in B-cell idiotypes created by the differential splicing of Ig genes, we compared B cells that used the same V-, D- and J-genes. Two gene combinations

containing complete V-, D-, and J-segments common to top 35 gene combinations found in all three mouse pools were selected for this analysis (Figure 4.7A-C, Figure A.5).

One heavy-chain VDJ-gene combination displaying a CDR3 region of variable length was selected from the 15 most common gene combinations among the three mouse pools and aligned to its germline sequence. From the full alignment, one short (four to eight AAs, Figure 4.7A), one medium (11 AAs, Figure 4.7B), and one long (14 AAs, Figure 4.7C) selection of nucleotide sequences were isolated and compared. Although the three groups were encoded by the same V-, D-, and J-gene segments, gene segment representation across each sample was variable. Most variability occurred in or around the D-gene segment, which could be due to splicing, N- and P-nucleotide additions, and deletions during somatic recombination. D-gene usage also appeared to be a factor determining CDR3 length. This was evidenced by increasing D-gene representation across CDR3 selections of decreasing length compared to the relative conservation of the V- and J-gene segments, though J-gene conservation seemed to decrease among extremely short CDR3s. Overall, the V-gene segment appeared to remain the most uniform.

While Igk contains mostly CDR3 sequences that are nine amino acids in length, many highly abundant VJ-gene segment combinations (such as V110 and J2) contained CDR3s of multiple lengths. Unlike IgH, the alignments of VJ pairings were relatively uniform among Igk as compared to the germline sequence in CDR3 sequences that were eight, nine and ten amino acids in length (Figure 4.7D).

## Discussion

To our knowledge, these data are the first unamplified sampling of the normal mouse antibody repertoire that has been described. Others have looked at Ig-gene segment usage with other strategies<sup>44,196,199,242,286</sup> but we wanted to determine if a straight-forward RNA-Seq approach would provide us with a reasonable assessment of B cell Ig-segment use without the limitations that amplification methods introduce (Rettig et al., this thesis).

To minimize potential single animal aberrations and repertoire skewing, we pooled splenic tissue of four unimmunized mice in three biological replicates. This approach was successful since we saw less variation with pooled samples compared to data sets that are made up of single mice.<sup>58</sup> Grieff et al. demonstrated that CDR3 and VDJ composition in pooled mouse samples were less polarized than that of an individually sequenced mouse subjected to antigen challenge.<sup>244</sup> Therefore, our data are consistent with that study. We also did the technical replication of material in pool one and found that there was good reproducibility ( $R^2 = 0.7562$ ).<sup>45</sup> Therefore, the data in the individual normal mouse pools and a compiled summary of those data are a strong reflection of the normal mouse repertoire.

The most common VH-gene segment was V1-80, which was the most common in pools one and two. V6-3 was the most common in pool three. In selecting the ten most common VH-gene segments from each pool, we identified 14 different VH-gene segments, with heavy overlap among pools. All VH-gene segments isolated comprised between 1% to 8% of the repertoire. This VH V-gene variation is consistent with previous observations<sup>65</sup> but the pools ameliorated the extreme variations that was reported by that group. We also saw that V $\kappa$ -gene segment usage

was comparable among mouse pools, with 17 gene segments comprising between 1% and 13% of the repertoire. Although we have found that gene segment use among all three pooled sample groups was similar, there were some differences. V5-39 was observed at a high frequency in pool one (10%) as compared to pools two (1%) and three (3%). This skewing could be the result of a mouse within pool one responding to a specific antigen that other mice did not respond to or more likely, represents the natural variability of mice.<sup>44</sup> Nevertheless, even pooling samples maintains the randomness of antibody gene selection and rearrangement and a pool of four individuals still has relative uniqueness.

Some have suggested that V-gene segment usage may be skewed.<sup>241</sup> Chi-square analyses of VH- and V $\kappa$ -gene segment use in our data set would support this contention since several VH and V $\kappa$ -gene segments were used more frequently than expected. Even though we have analyzed three independent biological samples made of pools of four mice, we recognize that an even larger data set will be needed to conclusively settle this discussion. Studies on humans have revealed similar V(D)J usage in spite of them being outbred populations<sup>276</sup> which also supports that there is some inherent selection in V-gene selection regardless of genetics. Additional studies looking at epigenetic changes or other transcriptional regulatory elements such as the characterization performed by Choi et al.<sup>287</sup> might also help understand mechanisms of V-gene segment selection.

For the D-gene segment, three usage levels were detected. D1-1 was the most used gene segment in all three pools; comprising around 26% of the total repertoire. Over 36% of D-gene segments could not be identified, likely due to the short length of the D-gene segment.



Nevertheless, we do see different populations of antibodies even when they do share similar D-genes segments. Some have large D-gene segments where others have little recognizable sequence.

JH-gene segment usage was relatively uniform. J2 was the most common among all three pools, with over 32% use in the repertoire. J1, J3, and J4 were evenly represented among all three pools totaling between 19-27% of the repertoire. Gene segments with less than six nucleotides were unable to be identified and occur at less than 0.1% of the heavy-chain repertoire. J $\kappa$ -gene segment usage is somewhat evenly distributed among J1, J2, and J5 comprising between 25-32% of the repertoire, in agreement with the findings of Aoki-ota et al.,<sup>241</sup> Lu et al. found a slightly different J $\kappa$  expression profile, possibly reflecting strain specific usage of J $\kappa$ .<sup>199</sup> As paralleled in the heavy-chain data, gene segments with less than six nucleotides were rare; occurring in less than 0.4% of the total repertoire.

Constant region usage in the heavy-chain was heavily dominated by IgM, which reflects the “naïve” status of our mice. Although IgM comprised over 78% of the total identified constant regions we did see the expression of IgG, IgA and IgD. IgE was rare, being detected in only pool three. However, when compared to serum data, even in naïve mice, there was a high level of circulating IgGs, which was not reflected in spleen tissue sequencing, which instead shows very high levels of IgM.<sup>288</sup> This could be due to a large B-cell population in the spleen that is not secreting antibody at high levels into the bloodstream.<sup>44</sup>

We looked at the common H-CDR3 sequences among the three mouse pools. There was little overlap; with only 75 H-CDR3s detected in all three pools and between 92 and 163 common H-CDR3s when we just looked at two pools. We detected between 4.6k to 6.2k unique sequences found only in each respective pool. While we sampled a small fraction of CDR3s present in the total antibody repertoire, Lu et al. and Greiff et al. used primer amplification to enrich for IgH transcripts and still found high CDR3 variability among individuals.<sup>44,199</sup> Similarly, in a study comparing monozygotic twins, Glanville et al., also demonstrated that CDR3 profiles between the individuals were quite diverse despite similar gene family usage between the twins.<sup>289</sup> Therefore, the pooling methodology that we employed did not significantly diminish the detection of the unique CDR3 repertoires that individuals have.

When we examined the  $\kappa$ -CDR3 usage among the three biological samples, there was a higher proportion of common  $\kappa$ -CDR3s. Unique  $\kappa$ -CDR3s within each pool ranged from 688 to 832 CDR3 identified within all three pools and between 108 and 178 CDR3 identified in only two pools. One explanation for light-chain CDR3 length homogeneity may be selection due to light-chain editing that occurs during B-cell maturation and the need to be able to interchange the light chains.

The small numbers of overlapping CDR3 sequences among our three pooled samples suggests that significant variation in the idiotypes could develop, even within an inbred population of mice and reinforces the idea of unique generation of B-cell diversity in inbred and outbred populations.<sup>276,289</sup> Moreover, we were curious if the size of the total pool of B cells could be estimated from our data. Using a model of capture-recapture methodology,<sup>290</sup> the Chapman

estimator,<sup>291</sup> and the number of common heavy-chain CDR3 sequences seen in each of our samplings, we estimated our B-cell pool to range from 1.5-14 x 10<sup>6</sup> cells. If we assume that there are 5 x 10<sup>6</sup> B cells in a nine to eleven-week-old female C57BL/6J mouse spleen, and our mouse pools were made up of four spleens, this estimate of the possible splenic B-cell pool is reasonably accurate, especially if we take into account some of the CDR3 sequences were detected multiple times (multiple B cells with the same IgH).

While CDR3 is commonly used to describe the antibody repertoire, many studies have reported the combinations of the V(D)J.<sup>120,128,244,292,293</sup> Compiling CDR3 nucleotide alignments allowed us to visualize the significance of individual gene segment involvement with the CDR3 in the context of specific V(D)J combinations. Sequencing outside of CDR3 also reveals biologically relevant information about antigen binding and allows for further characterization and potential lineage determination. Our sequencing technique enabled the identification of V(D)J-gene segments in addition to constant region, providing insight into pairing of V-gene families with (D)J-gene segments.

The information about the unchallenged Ig-gene repertoire also has other uses. It provides a comparative foundation when looking at host response to antigen<sup>294</sup> and has been used to isolated therapeutic antibodies. Antibodies for influenza in a mouse model and were a valuable tool in the detection of antigen specific responses.<sup>133,270</sup>

While a lack of amplification may extricate primer bias, we knew that it would come at the cost of potentially excluding rare B-cell clones. In humans, for example, a single clone may

only comprise 0.1% to 0.3% of the repertoire.<sup>295</sup> We have explored the differences between samples that have and have not been amplified and found a moderate correlation ( $R^2 = 0.5815$ ,  $0.5855$ ,  $p < 0.0001$ , Rettig et al., this thesis). While some of the differences arose from expected depth-of-sequencing issues, we unexpectedly found that discrepancies also resulted from gene segments being detected in the unamplified data set not detected in the amplified data sets (Rettig et al., this thesis). It is also important to note that the number of immunoglobulin reads detected in our RNASeq library equaled or exceeded those in other HTS studies (Boyd et al. 2009). Therefore, we are aware of the tradeoffs and benefits of the HTS strategy we have used.

Another issue which may affect the data presented stems from the use of whole spleen tissue rather than isolated B-cell populations.<sup>133</sup> Although our approach was necessary to accommodate requirements of a separate investigation,<sup>45</sup> we are aware that the inclusion of extraneous cells as result of using whole tissue could reduce the recovery of rare B-cell clones. In addition, some bias might be introduced because of cell subpopulation stability and frequency in whole spleen tissue.<sup>25,26</sup> Nevertheless, in spite of the limitations of our methodology, it appears that the repertoire we detected correlated with mouse studies that have used selection and amplification methods of various kinds. For example, Collins et al. detected five of the same VH genes that we detected among our highest 10 used VH-gene segments. JH2 was also the most frequently detected in both of our studies.<sup>14</sup> Yang and Kaplinski detected V-gene segment use that paralleled our findings with V1-26 identified by them as the most frequently used.<sup>67,242</sup>

Few studies have explored the light chain repertoire; however, more characterization will be possible with increasing use of single cell amplification.<sup>94,296,297</sup> While strain specificity has

been reported,<sup>14,44</sup> many V $\kappa$ -gene segments that were represented over one-percent of the time in unimmunized BALB/c mice were also identified in our study.<sup>199</sup> Aoki-Ota et al. also noted V $\kappa$ -gene segment skewing in their assessment of unimmunized C57BL/6 mice.<sup>241</sup> These similarities also suggest that the lack of amplification did not dramatically affect our assessment of the B-cell repertoire, and the differences seen are likely due to mouse-to-mouse variation that still manifests in our pooled samples.

In conclusion, we have presented an unamplified view of the conventionally housed, unimmunized, antibody repertoire. It appears that an RNASeq approach without amplification can provide an accurate assessment of V-gene use as well as a snapshot of the CDR3 present in the population; and helps validate this as a reasonable scientific approach. In addition, we lay the foundation for future work in our lab to characterize the unamplified whole tissue repertoire of the immunized C57BL/6 mouse.

### **Data Availability**

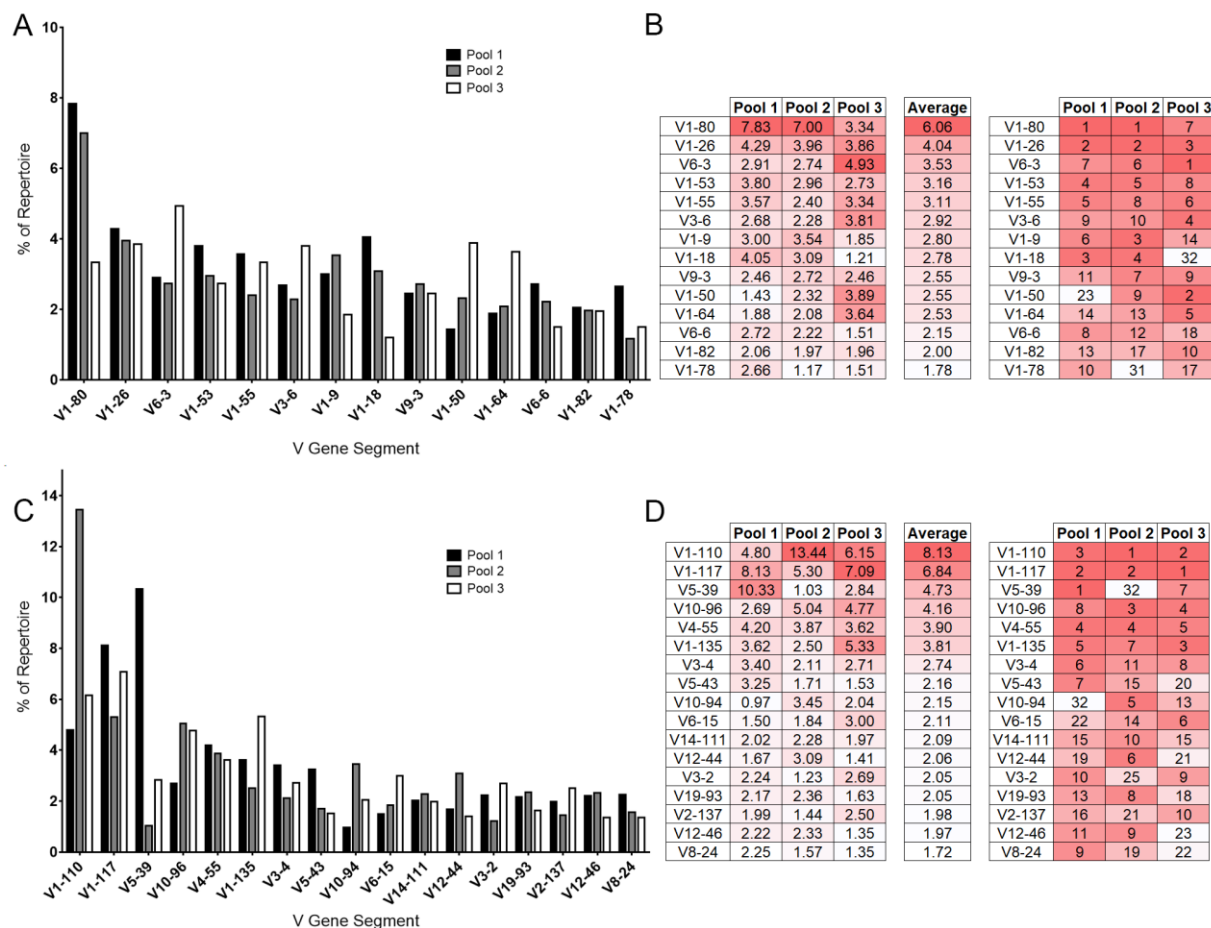
Raw FASTQ files are available through NASA's GeneLab (<https://genelab-data.ndc.nasa.gov/genelab/>) via accession number GLDS-141.

### **Acknowledgements**

We thank Ms. Melissa Gulley for her help in the lab and Dr. Alina Akhunova, Director of the Kansas State University Integrated Genomics Facility, for her help, dedication and expertise. We thank Mr. Michael Soriano, NBL Program Office, Office of Science, U.S. Department of Energy, Argonne, IL for his help with the cell population estimates in the spleen.

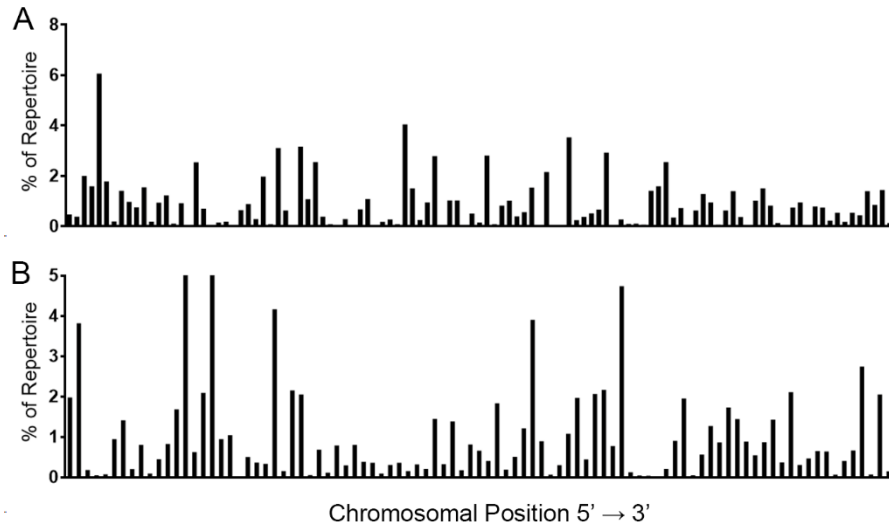
## Figures and Tables

**Figure 4.1. V-gene segment usage among unimmunized mouse pools.**



Sequencing reads mapped to each individual gene segment were divided by the total sequencing reads of all identified gene segments from each mouse pool for a normalized comparison between pools. (A) The VH representing the ten most abundant gene segments from each mouse pool are displayed. (B) The rankings of each gene segment contained within the top 10 most abundant VH from at least one of the mouse pools are compared. The most abundant gene segment is ranked as 1. Dark red indicates higher rank moving to white, of lower rank. Similarly, the top 10 abundant Vk are displayed (C-D).

**Figure 4.2. V-gene segment usage among unimmunized mouse pools for IgH (A) and Igk (B) by chromosomal location**



Gene segments are shown in order of chromosomal position (5' to 3'). The average value from three mouse pools for each CDR3 length is shown. Distribution was assessed via Chi-square analysis in R (version 3.4.2) (all pools,  $p < 0.0001$ ).

**Figure 4.3. Percent abundance of IgH D- (A) and J- (B) gene segments, IgH constant regions (C) and Igk J-gene segments (D).**

**A**

	Pool 1	Pool 2	Pool 3	Average		Pool 1	Pool 2	Pool 3
Undeter	33.18	31.24	31.15	31.86	Undeter	1	1	1
D1-1	27.57	25.89	26.01	26.49	D1-1	2	2	2
D2-3	6.47	15.47	9.62	10.52	D2-3	6	3	4
D2-4	8.42	7.66	11.57	9.22	D2-4	4	5	3
D4-1	8.43	7.79	8.28	8.17	D4-1	3	4	5
D2-5	7.99	5.51	6.77	6.75	D2-5	5	6	6
D3-2	2.77	2.23	2.69	2.57	D3-2	7	8	7
D3-1	2.19	2.80	1.92	2.30	D3-1	8	7	8
D6-3	2.12	0.88	1.46	1.49	D6-3	9	9	9
D5-5	0.62	0.18	0.28	0.36	D5-5	10	11	10
D5-2	0.17	0.34	0.21	0.24	D5-2	11	10	11
D6-4	0.06	0	0.04	0.03	D6-4	12	12	12

**B**

	Pool 1	Pool 2	Pool 3	Average		Pool 1	Pool 2	Pool 3
JH2	36.47	31.95	37.97	35.46	JH2	1	1	1
JH3	19.47	27.27	19.82	22.19	JH3	4	2	3
JH4	19.77	22.13	22.70	21.54	JH4	3	3	2
JH1	24.25	18.62	19.44	20.77	JH1	2	4	4
Undeter	0.03	0.04	0.07	0.05	Undeter	5	5	5

**C**

	Pool 1	Pool 2	Pool 3	Average		Pool 1	Pool 2	Pool 3
IgM	83.63	82.23	78.12	81.33	IgM	1	1	1
IgG	7.38	10.94	10.62	9.64	IgG	2	2	2
IgD	3.81	5.02	4.98	4.60	IgD	4	3	4
IgA	5.17	1.82	6.27	4.42	IgA	3	4	3
IgE	0	0	0.02	0.01	IgE	5	5	5

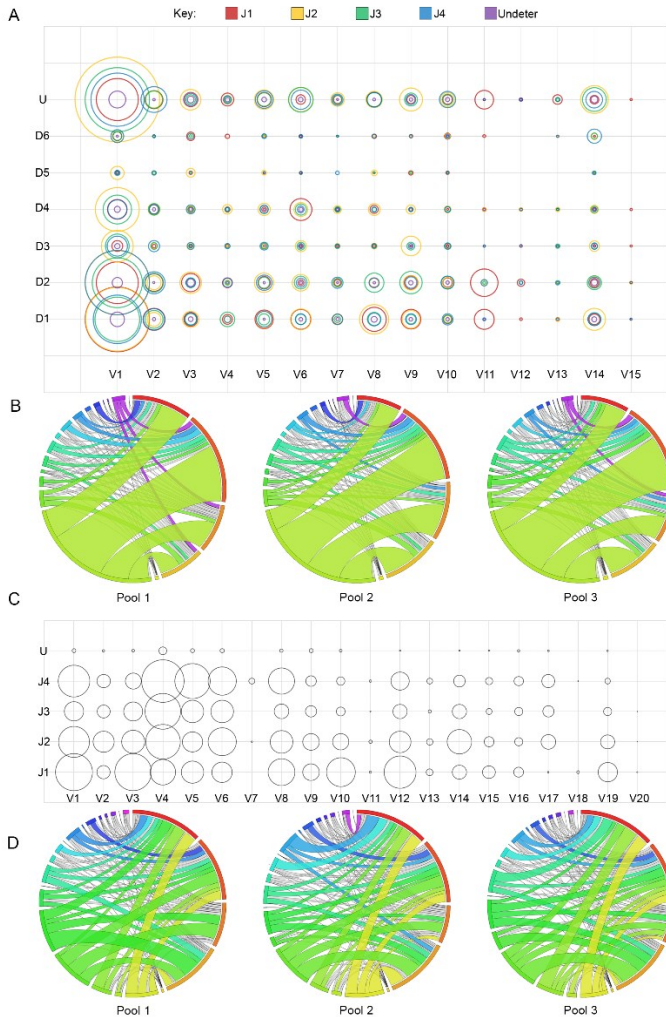
**D**

	Pool 1	Pool 2	Pool 3	Average		Pool 1	Pool 2	Pool 3
Jk1	28.92	30.69	31.68	30.43	Jk1	1	1	1
Jk2	27.00	27.08	25.85	26.65	Jk2	2	2	3
Jk5	25.29	26.22	26.94	26.15	Jk5	3	3	2
Jk4	18.76	15.63	15.18	16.52	Jk4	4	4	4
Undeter	0.03	0.38	0.35	0.25	Undeter	5	5	5

Sequencing reads corresponding to each gene segment or constant region were divided by the total number of gene segments or constant regions identified in each mouse pool for normalized comparison between pools (left side). The most abundant gene segment is ranked as one (right side). Dark red indicates higher rank moving to blue (A) or white (B-D), of lower rank. Sequencing reads designated undetermined (undeter) where portions of a D- or J-gene segment were identified but unable to be assigned to a specific C57BL/6J D- or J-gene segment.



**Figure 4.4. Combinations of V-gene families with DJ-gene segments for IgH (A) and J-gene segments for Igκ (C).**

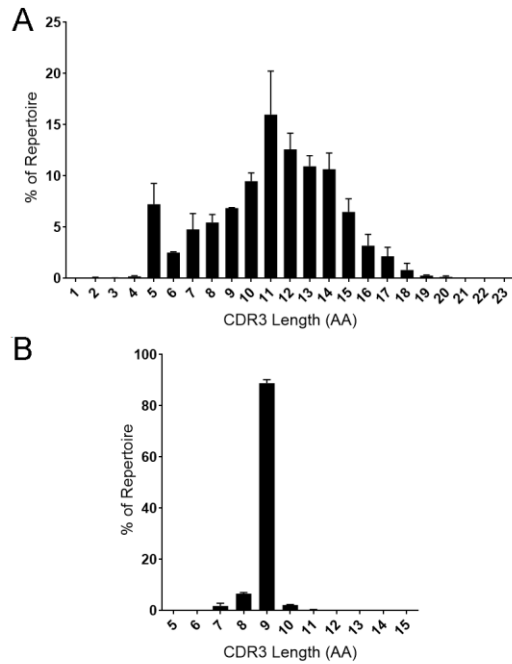


Increasing pairing frequency of V(D)J is represented by larger circles. Sequencing reads in which more than one C57BL/6 J-gene segment was attributed or too few nucleotides were present in the J-gene segment for designation by IMGT have been classified as undetermined (U). Pairing frequency is also represented by Circos graphs for IgH (B) and Igκ (D). Circos Plot Labels (starting at 12:00 position and the largest arc and continuing clockwise with occasional color references)

B – J1 (red), J2, J3, J4, U (yellow), V1, V2, V3, V4, V5, V6, V7 (Teal), V8, V9, V10, V11, V12, V13, V14 (purple), V15

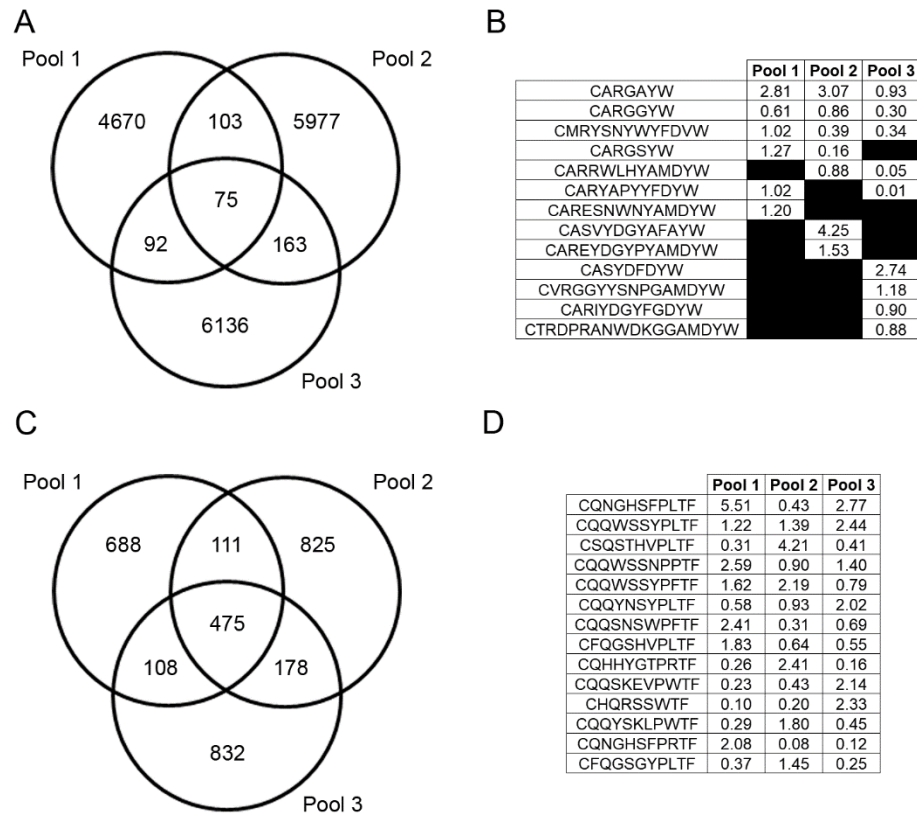
D – J1 (red), J2, J4, J5, U, V1 (yellow), V2, V3, V4, V5, V6, V7 (black sliver), V8, V9, V10, V11, V12, V13, V14, V15 (royal blue), V16, V17, V18, V19, V20 (black sliver, if present)

**Figure 4.5. CDR3 length for IgH (A) and Igκ (B).**



The average percent of repertoire of each CDR3 amino acid length from three mouse pools is displayed.

**Figure 4.6. Top CDR3 AA sequences and overlap of unique CDR3 sequences within mouse pools.**



A Venn diagram displays the overlap of the number of unique CDR3 amino acid sequences among mouse pools for IgH (A) and Igk (C). The percent of repertoire for the top five CDR3 amino acid from each mouse pool are shown for IgH (B) and Igk (D).



**Table 4.1. Sequencing and mapping statistics from mouse pools 1, 2, and 3.**

	Pool 1	Pool 2	Pool 3
<b>Total Reads</b>	25.1 M <sup>a</sup>	31.4 M	32.7 M
<b>Post Cleaning</b>	12.0 M	30.9 M	32.0 M
<b>Productive IgH</b>	8,714	11,200	10,224
<b>Unknown IgH</b>	14,271	27,896	18,756
<b>Productive Igκ</b>	11,968	18,643	16,293
<b>Unknown Igκ</b>	12,602	39,410	36,624

<sup>a</sup>M: million

## **Chapter 5 - Effects of spaceflight on the immunoglobulin repertoire of unimmunized C57BL/6 mice**

### **Abstract**

Spaceflight has been shown to suppress the adaptive immune response, altering the distribution and function of lymphocyte populations. B lymphocytes express highly specific and highly diversified receptors, known as immunoglobulins (Ig), that directly bind and neutralize pathogens. Ig diversity is achieved through the enzymatic splicing of gene segments within the genomic DNA of each B cell in a host. The collection of Ig specificities within a host, or Ig repertoire, has been increasingly characterized in both basic research and clinical settings using high-throughput sequencing technology (HTS). We utilized HTS to test the hypothesis that spaceflight affects the B-cell repertoire. To test this hypothesis, we characterized the impact of spaceflight on the unimmunized Ig repertoire of C57BL/6 mice that were flown aboard the International Space Station (ISS) during the Rodent Research One validation flight in comparison to ground controls. Individual gene segment usage was similar between ground control and flight animals, however, gene segment combinations and the junctions in which gene segments combine was varied among animals within and between treatment groups. We also found that spontaneous somatic mutations in the IgH and Igk gene loci were not increased. These data suggest that space flight did not affect the B cell repertoire of mice flown and housed on the ISS over a short period of time.

**Citation:** Ward, C.,\* Rettig, T. A.,\* Hlavacek, S., Bye, B. A., Pecaut, M. J., Chapes, S. K. Effects of spaceflight on the immunoglobulin repertoire of unimmunized C57BL/6 mice. *Life Science in Space Research* **16**, 63-75. doi: <https://doi.org/10.1016/j.lssr.2017.11.003> (2018)

\*Co-First Authors.

## Introduction

Spaceflight presents a unique set of challenges to the immune system. For example, spaceflight alters T- and B-lymphocyte functions, including recall responses in astronauts aboard the space shuttle and cytokine responses after missions to the international space station (ISS).<sup>171,179,215,222,228,233,298,299</sup> In addition to functional changes, lymphocyte subpopulations are altered. CD8<sup>+</sup> T-cell numbers were increased during flight while other T-cell subsets were decreased.<sup>299</sup> Similar changes in phenotype also occur in animal and tissue culture systems during spaceflight or ground-based spaceflight analogs such as anti-orthostatic suspension (AOS).<sup>164,165,237,300,301</sup>

Lymphocyte subpopulations change in response to spaceflight<sup>175,224,226,235,236,302-304</sup> and AOS.<sup>176,305</sup> Splenic T- and B-lymphocyte counts were decreased in mice flown on the 13-day mission of the Space Shuttle Endeavor (STS-118) compared to ground controls.<sup>175</sup> In the AOS model, Wei et al. found a reduced number of both T and B lymphocytes in the thymus and spleen of hindlimb unloaded Balb/c mice compared to normal controls.<sup>305</sup> Reductions in the mass of lymphoid organs has also been observed.<sup>176,201,205,208,210,211,302,306,307</sup> Spaceflight altered the phenotype of immune cells in the bone marrow, the lymphoid organ in which hematopoiesis occurs,<sup>188</sup> and AOS reduced the number of bone marrow B-cell progenitors.<sup>177</sup>

While many studies have characterized T-cell response to spaceflight,<sup>175,189,205,213,232,302,308-317</sup> fewer studies have characterized the impact of spaceflight on B-cell populations. The characterization of B-cell receptors, known as immunoglobulins (Igs), is of particular interest due to the (IgH) and light chains, which are encoded on separate loci.<sup>262</sup> The

heavy chain locus encodes multiple Variable- (V), Diversity- (D) and Joining- (J) gene segments, while the functionally equivalent  $\kappa$  (Ig $\kappa$ ) and  $\lambda$  (Ig $\lambda$ ) light chain loci contain only V- and J-gene segments (Early et al., 1980, Sakano et al., 1979). During early B-cell development in the bone marrow, B cells undergo recombination of heavy and light chain Ig loci, in which only one of each V(D)J-gene segment is selected for Ig use.<sup>5,262</sup> Random and palindromic nucleotide insertion at splice sites adds to Ig diversity.<sup>318-320</sup>

In the Ig molecule, complementarity determining regions (CDR) confer binding specificity. CDR1 and CDR2 are encoded entirely within the V-gene segment, while CDR3 contains a portion of the 3' end of the V-gene segment, the entire D-gene segment, and a portion of the 5' end of the J gene segment.<sup>200,262,265</sup> As a result of somatic recombination, B cells collectively express individual Igs that theoretically can bind virtually any pathogen.

An individual's Ig repertoire can be characterized using high-throughput sequencing (HTS) using either genomic DNA or messenger RNA sequences isolated from B-cell populations.<sup>45,133,321</sup> B cells will clonally expand after antigen-Ig receptor engagement, resulting in a higher portion of target-specific Ig receptors within the B-cell population. There have been a number of HTS-based Ig repertoire studies in human disease, ranging from infectious disease,<sup>77,78,125,271</sup> autoimmunity,<sup>274,275,322</sup> and cancer.<sup>109,116,273,323,324</sup> Greiff, et al. developed a profiling framework using the Ig repertoire as an indicator of an individual's immunological status.<sup>65</sup>



Some have explored the impact of spaceflight on Ig repertoires. *In vitro* challenge of human B cells during spaceflight resulted in lower concentrations of secreted Ig (Fitzgerald et al., 2009). There was no significant difference in pre- and post-flight Ig levels in peripheral blood of astronauts who flew aboard the ISS.<sup>194,217,298</sup> These samples, however, were not taken after challenge with a specific antigen. Rats immunized intraperitoneally with sheep red blood cells prior to spaceflight produced significantly less serum IgG compared to immunized ground control animals.<sup>213</sup>

Although some have explored Ig gene segment changes in the context of spaceflight or model analogs,<sup>191,325-327</sup> little has been done to characterize the impact of spaceflight on the Ig repertoire in mice. Given that changes in B cells and Ig concentrations occur during spaceflight conditions, we tested the hypothesis that spaceflight alters the Ig repertoire of mice flown on the ISS. We examined individual Ig gene segment usage, gene segment combinations, CDR3 composition, and frame work and CDR mutations in 35-week-old, unimmunized, female C57BL/6Tac mice flown aboard the ISS using high throughput sequencing.

## **Materials and Methods**

### **Tissue Samples**

RNA samples were provided by the NASA Ames Research Center. RNA was extracted from the spleen and liver of 35-week-old female C57BL/6Tac mice that were either housed in the ISS environmental simulator (ground control, n=5), or flown aboard the ISS via SpaceX-4 (n=5). Tissues from flight animals were collected on board the ISS 21-22 days post-launch in flight animals while tissues from ground control animals were processed similarly on a four-day

delay. Upon collection, spleens and livers were stored at 4°C in RNAlater (LifeTechnologies, Carlsbad, CA) for at least 24 hours and then stored at -80°C. RNA extraction was performed according to manufacturer's instructions with the RNeasy mini column (QIAGEN, Hilden, Germany) and stored at -80°C. This was a secondary science experiment and the dissection and timing of the experiment were dictated by the primary validation experiment. Animal care and experimental procedures were approved by the Institutional Animal Care and Use Committee at the NASA Ames Research Center.

### **Illumina MiSeq Sequencing**

RNA samples were subjected to Illumina MiSeq sequencing at the Kansas State University Integrated Genomics Facility as previously outlined in Rettig et al (Rettig et al., 2017). Briefly, sequencing was performed using the standard MiSeq protocol which includes oligo-dT-bead selection and reverse transcription of mRNA to cDNA. Ig-specific primer amplification was not utilized. Additionally, fragmentation was limited to one minute to reduce fragmentation and maintain longer reads. Illumina MiSeq with paired reads of 300 base pairs was performed on size selected (275-800 nt) total RNA isolated from the liver and spleen of three ground control and three flight animals based on highest RIN values (Spleen RIN:5.9-8.9, Liver RIN: 6.2-7.8) (Ground animals: G1, G2, G3; Flight animals: F1, F2, F3). Illumina MiSeq data from both spleen and liver are available by NASA GeneLab (<https://genelab.nasa.gov>, GLDS-ID Pending).

## **Bioinformatic Workflow**

Illumina MiSeq sequencing reads were processed as described previously (Rettig et al., 2017). Briefly, FASTQ files were imported into CLC Genomics Workbench v9.5.1 (<https://www.qiagenbioinformatics.com/>) and were quality trimmed and filtered to remove sequences less than 40 nt in length to prevent false V-gene segment assignment. Paired-end sequences and overlapping-paired (merged) sequences were mapped to both V-gene segment references obtained from the ImMunoGeneTics (IMGT) database (251 IgH segments, 135 Igκ segments), and to entire IgH and Igκ loci obtained from NCBI (NC\_000078.6, 113258768 to 116009954, and NC\_000072.6, 67555636 to 70726754, respectively). Mapped sequencing reads were submitted to the IMGT HighV-Quest tool for characterization of functionality and junctional analysis. Only one sequence per sequencing cluster was retained for further analysis as outlined in Rettig et al.<sup>45</sup> Briefly, per sequencing organization ID, the read with the most information is saved for further analysis. Antibody sequences can differ by as little as a single nucleotide and without a unique barcoding step during amplification, removing similar, but not identical sequences, could limit the breadth of the repertoire sampled. Due to this, no further filtering was performed on sequences. A motif search was performed in CLC on IgH sequences that were identified by IMGT as productive to determine their respective constant regions.

## **Gene Segment Usage**

Sequencing reads were analyzed using the IMGT HighV-Quest tool.<sup>283</sup> V-gene segment usage was characterized as either productive or unknown functionality, where a read was considered productive if it was in frame and did not contain a premature stop codon as defined by IMGT. Sequences that did not fit the C-xx-W motif in non-class switched H-CDR3 or C-xx-F

in  $\kappa$ -CDR3 were assigned an unknown functionality. D- and J-gene segment, and constant region usage was assessed in productive reads only. All gene segments considered functional by IMGT (includes open reading frames and gene segments without full NCBI mapping) were included in this analysis. Reads assigned to multiple C57BL/6 V-gene segments were tabulated using a weighted distribution. Reads containing only one possible V-gene segment were assigned a count of one. Reads containing two possible V-gene segments were assigned a count of 0.5 for each potential V-gene segment. Reads containing more than two potential V-gene segments were excluded from V-gene analysis. Multiple V-gene segment assignments likely resulted from reads containing less than a full V-gene segment as a result of random hexamer priming. Reads assigned to a single non-C57BL/6 D/J-gene segments or multiple C57BL/6 D/J-gene segments were reclassified as undetermined and kept for analysis. J-gene segments in which less than six nucleotides were identified were also classified as undetermined. Percent of repertoire was determined for each individual animal by dividing the number of sequencing reads for each gene segment by the total number of sequencing reads mapped to all gene segments.

### **Gene Segment Combination & CDR3 Analysis**

Reads assigned to non-C57BL/6 V-gene segments or multiple C57BL/6 V-gene segments were removed from our V(D)J combination analyses. V(D)J combination analyses were performed on productive sequencing reads and visualized through the use of bubble charts (Microsoft Excel) and/or circos graphs from Circos Online.<sup>285</sup> Percent repertoire was used to detail individual bubble charts and the average of percent repertoire was used when combining mice from each treatment group for V(D)J bubble chart analysis. CDR3 amino acid sequence

was presented as percent of repertoire as described above and by a highest to lowest ranking of abundance.

CDR3 nucleotide alignments were created using the MAFFT multiple sequence alignment program.<sup>284</sup> A V-D-J-gene segment combination was selected from the top ten percent most-represented gene segments across all individuals for both heavy and light chain. From each individual group, unique nucleotide sequences were isolated and aligned to their respective germline sequences provided by IMGT. Individual alignments were then stacked within their treatment groups and germline gaps were adjusted for consistency across treatment groups.

### **Complementarity Determining and Framework Region Mutation Analysis**

Nucleotide substitution mutation data for complementarity determining and framework regions for IgH and Igk were obtained from the IMGT HighV-Quest tool. Any mutations involving degenerate bases were removed. Nucleotide range (in base pairs), number of reads containing at least one mutation, total number of substitution mutations, and number of mutations per base pair position were determined for each region. Comparative values for each combination of region, Ig location, and treatment group were determined by calculating the average of values contained in each combination's respective replicates (n=3).

### **Statistical Analysis and Representation of Data**

All statistical analyses were performed in GraphPad (version 6.0). Dot plots and bar graphs were generated in GraphPad using mean values and standard deviation. Heat maps of gene segment usage were generated in Microsoft Excel.

## Results

### V-Gene Segment Usage

B cells originate in the bone marrow from hematopoietic precursors, traffic through the periphery and enter the spleen where they are further selected and mature.<sup>328</sup> To view a snapshot of the impact that spaceflight has on the splenic Ig repertoire of unimmunized mice, we sequenced total splenic RNA isolated from three ground control animals and three animals flown aboard the ISS. We assessed the composition of individual IgH and Igκ sequences. In spleen, between 104,135 and 149,675 IgH, and between 103,841 and 175,406 Igκ sequencing reads of productive or unknown functionality were detected in ground animals, while between 66,909 and 181,703 IgH, and between 81,889 and 107,928 Igκ were detected in flight animals (Table 5.1).

The V-gene segment contributes to the combinatorial diversity of the Ig repertoire in part due to the large number of possible V-gene segments that could be selected within an individual B cell. Among the six study animals 133 VH- and 108 Vκ-gene segments were detected. Overall, the frequency of highly abundant V-gene segments and less frequently identified V-gene segments were similar between treatment groups (Figure A.6). Despite a general similarity, a pairwise comparison of animals within treatment groups showed low-to-moderate levels of VH-gene segment correlation (Ground  $R^2$ : 0.356-0.695, p-values:<0.001; Flight  $R^2$ : 0.101-0.360, p-values:0.0001-<0.0001) that demonstrates that there is animal-to-animal variation (Table 5.2). A stronger correlation was seen in Vκ-gene segments (Ground  $R^2$ : 0.660-0.738, p-values <0.0001; Flight  $R^2$ : 0.465-0.606, p-values <0.0001) (Table 5.2). When comparing the average abundance of V-gene segments from ground and flight animals an  $R^2$  of 0.592 was observed in VH ( $p$ =<0.0001) and 0.810 was observed in Vκ ( $p$ =<0.0001) (Table 5.2).

When comparing VH-gene usage among all animals, nine gene segments represented over five percent of the repertoire in at least one animal (Figure 5.1A). No one gene segment was found at over the five percent level in all six animals. V1-53 was found in over five percent in five animals, V9-3 was over five percent in four animals, V1-26 and V3-6 in two animals, and the remaining (V1-78, V6-3, V5-4, V1-15, V1-19) were found at high levels in only one animal.

Within V $\kappa$ , six gene segments represented over five percent of the repertoire in at least one animal (Figure 5.1B). The most abundant gene segment, V5-39 comprised over 16 percent of V $\kappa$  usage in all six animals. No other V $\kappa$  represented over five percent of the repertoire in all six animals. One gene segment, V3-4, was found at over five percent of the repertoire in four animals (G1, G2, G3, F3). The remaining four gene segments (V2-137, V1-110, V4-61, and V6-25) were found at greater than five percent in only one animal. There was no statistical difference in top VH- or V $\kappa$ -gene segment usage between ground and flight animals (student's t-test,  $p=0.6478-0.9609$ ).

We also attempted to assess the antibody repertoire in the liver because of its role in fetal B-cell development. Only 592 to 1,429 Ig $\kappa$  sequencing reads were detected in ground control animals and 425 to 543 Ig $\kappa$  sequencing reads were detected in flight animals (Table 5.3). We did not characterize the heavy chain in the liver due to the low number of IgH sequencing reads that we detected. We assessed V $\kappa$ -gene segments that represented over five percent of the repertoire in the spleen or liver and found 11 gene segments. Only one gene segment, V5-39, was found in the top five across both tissues (Figure 5.2). V3-4 was found in the top five for all animals and

tissues except flight mouse two's liver sample, where it was ranked tenth. Overall, the average usage of these top V $\kappa$ -gene segments showed low to modest correlation between liver and spleen in ground animals ( $R^2 = 0.4081$ ,  $p < 0.0001$ ) and flight animals ( $R^2 = 0.2727$ ,  $p < 0.0001$ ). Analysis of statistical differences between individual V $\kappa$ -gene segments representing over five percent of the repertoire was not undertaken due to low read counts in the liver datasets.

### **D- and J-Gene Segment & Constant Region Usage**

Heavy chain Ig diversity is also achieved by using, modifying and splicing of D- and J-gene segments. To determine if space flight affected these processes we also assessed D- and J-gene usage. The most commonly detected D-gene segment in both ground control and flight animals was D1-1, comprising between 30.6% to 46.8% of the repertoire (Figure 5.3A, Figure A.7A). D2-4, D2-3, D4-1, and D2-5 were detected at similar levels among ground control and flight animals between 3.56% and 11.39% of the repertoire. D3-1, D3-2, D6-3, D5-5, and D5-2, and D6-4 were detected the least often with levels between 3.6% and  $>0.003\%$  of the repertoire. Because of extensive modification of D-gene segments during IgH rearrangement, D-gene segments were unable to be determined for between 24.4% and 36.6% of the repertoire for all animals. There was no statistical difference in D-gene segment usage between ground and flight animals (student's t-test,  $p = 0.1542-0.9840$ ). D-gene segment usage was highly correlated between ground and flight animals (linear regression,  $R^2 = 0.9935$ ,  $p < 0.0001$ ).

Additional Ig variability is gained from the inclusion of different J-gene segments. Within IgH, the distribution of J-gene segment usage was less uniform than D-gene segment usage in both ground and flight animals. There was no consensus on the most abundantly



expressed gene segment as each of the four JH-gene segments was the most abundant segment in at least one ground or flight animal (Figure 5.3B, Figure A.7B). While there was no consensus use of a particular J-gene segment, usage of any J-gene segment was between 13% and 40%, showing that usage is relatively uniform. Student's t-tests were performed to determine whether differences in individual gene segment usage were significant. No significant differences were found between ground and flight animals (student's t-test,  $p=0.2060-0.8662$ ). As the most abundant J $\kappa$ -gene segment, J5, comprises between 32.6% and 47.9% of the repertoire, and is the most abundant gene segment all study animals (Figure 5.3C, Figure A.7C). Interestingly, J1 ranked second most abundant in all ground animals while it ranked third in all flight animals. There were no statistical differences in individual JH- or J $\kappa$ -gene segment usage between ground control and flight animals (Student's t-test,  $p=0.0977-0.9262$ ). Linear regression revealed correlation between JH usage of ground and flight animals for Ig $\kappa$  ( $R^2=0.9928$ ,  $p=0.0076$ ) and IgH ( $R^2=0.8147$ ,  $p=0.0360$ ).

Ig isotype composition can provide insight into the developmental stage of B cells. Because animals in this experiment were specific-pathogen free, it is unsurprising that IgM predominated with between 62.38% and 82.81% of the repertoire (Figure 5.3D, Figure A.7D). IgG was the second most prominent isotype which trended towards a higher percentage of the repertoire in flight animals although the difference was not statistically significant (Student's t-test,  $p=0.2150$ ). Except for a relatively high expression of IgA in flight mouse two, IgA was detected in between 1.43% and 4.58% of the repertoire, IgD and IgE were detected less than one percent in all animals. There were no statistical differences in Ig isotype frequency between

ground and flight animals (Student's t-test,  $p=0.1075-0.8277$ ). There was a high correlation between ground and flight constant region usage (linear regression,  $R^2=0.9734$ ,  $p=0.0019$ ).

### **V(D)J Combinations**

V(D)J family combinations were examined as another way to determine if recombination of Ig gene segments was affected by spaceflight. We visualized V(D)J-gene segment combinations using both bubble charts and circos plots for both IgH and Igκ. IgH showed more variation between ground control and flight animals compared to Igκ when looking at the most common gene family combinations. Circos plots allowed us to examine top V-gene families, J-gene segments, and V/J pairing frequency in both IgH and Igκ.

For ease of display, we first grouped together all V-gene segments into their respective family. D- and J-gene segments remained as individuals. V1 was the most common IgH gene family used in all mice comprising, on average 51% of the V-gene family use in ground animals and, on average, 57% of the repertoire in flight animals (Figure 5.4A-B, Figure A.8A-F). In ground-treatment animals, V9 was the second most common gene family in ground mouse one and ground mouse two, while V2 was the second most common gene family in ground mouse three (Figure 5.4A, Figure A.8A-C). In flight animals, the second most common V-gene family used was unique among the three animals (flight one: V3, flight two: V2, and flight three: V9) (Figure 5.4B, Figure A.8D-F). The third most common V-gene family was also unique among the ground treatment animals being V2, V3, and V9 for ground mouse one, ground mouse two, and ground mouse three respectively (Figure 5.4A, Figure A.8A-C). V5 was the third most

common family in flight mouse one, V9 in flight mouse two, and V2 in flight mouse three (Figure 5.4B, Figure A.8D-F).

We found that the most common V/D/J combinations were correlated with the most frequently used gene families or segments within the repertoire. When averaging among repertoires, the most common IgH combination in ground-treatment animals was V9/D1/J1 (8.66%), though this combination was only in the top five most frequent combinations in ground mouse one and ground mouse two. The most common average combination in flight-treatment animals was V1/D1/J4 (8.35%), though this combination was only detected in the five most common combinations for one animal (Figure 5.4A-B, Figure A.9A-B). The V1/D1/J2 combination was shared among the top five gene family combinations in all mice; representing 5.94% of the repertoire in ground-treatment animals and 7.73% in the flight-treatment animals (Figure A.9A-B). A notable difference between ground and flight treatment groups was the usage of the V9-gene family. This family represented the top average VH-gene family used in the ground-treatment animals as well as the top combination in ground mouse one and ground mouse two (Figure 5.4A, Figure A.9), but only appeared once as the top gene family combination used for flight-treatment animals (Figure 5.4B, Figure A.9B).

We also examined the top five V/J pairing frequencies for IgH (Figure 5.4A-B, Figure A.8A-F, Figure A.9C-D). For ground animals, there were six unique pairings represented. The V1 family was used for four of the six unique V/J pairings compiled. V9 and V2 were also used. Of the six unique pairings, four were shared among all three mice (V1/J1, V1/J2, V1/J3, and V1/J4). One was shared among two mice (V9/J1), and one (V2/J4) was found only in ground

mouse three's five most common combinations (Figure 5.4A, Figure A.8A-C, Figure A.8C). For flight animals, seven unique pairings were found in the five most common pairs. V1 was again the overwhelmingly most common V family with four of the seven unique pairings including it. V3, V2, and V9 were all used by a single animal. Of the seven unique pairs, four were shared among all three mice (V1/J1, V1/J2, V1/J3, and V1/J4). These pairings were also among the most common in the ground-treatment group. Three (V3/J3, V2/J4, V9/J1) were found in a single mouse's five most common pairings (Figure 5.4B, Figure A.8D-F, Figure A.9D).

We undertook similar analysis for Igk sequences (Figure 5.4C-D, Figure A.8G-L, Figure A.9E-F). The most common V $\kappa$ -gene family expressed in all mice was V5, representing over one-fifth of the repertoire. The second most common V $\kappa$ -gene family in ground animals was V3 while in flight animals it was V6 or V4 (Figure 5.4C-D, Figure A.8G-L, Figure A.9E-F). The third most represented V $\kappa$ -gene families was V6 for ground mouse one, and V4 for ground mouse two, and V2 for mouse three (Figure 5.4C, Figure A.7G-I). In flight animals, the third most common V $\kappa$ -gene family was V4 for flight mouse one, and V3 for flight mouse two and three (Figure 5.4D, Figure A.9J-L).

Unlike IgH, Igk expressed more variety in V/J pairing variety. In Igk there were five pairings in ground animals (V2/J4, V3/J2, V3/J5, V5/J1, V6/J2) that were not shared with the top five flight-animal pairings (Figure 5.4C-D, Figure A.8G-L, Figure A.9E-F). There were also four pairings (V1/J2, V4/J2, V6/J1, V6/J4) found in flight animals not found in the top five ground-animal pairings (Figure 5.4C-D, Figure A.8G-L, Figure A.9E-F).

The correlation of the average percent of repertoire for V/J combinations between ground and flight animals was higher in Ig $\kappa$ , with a  $R^2$  of 0.8795 ( $p < 0.0001$ ), whereas IgH had a  $R^2$  of 0.3296 ( $p < 0.0001$ ) (linear regression). V/J combination usage among animals within each treatment group also showed stronger correlation within Ig $\kappa$  than IgH in both ground and flight treatment groups (Table 5.4).

### CDR3

CDR3 is important for conferring diversity in Ig specificity. Therefore, we assessed whether spaceflight had an impact on several properties of CDR3 because changes in CDR3 could affect host ability to respond to antigen. We found that CDR3 length in IgH was highly varied. The length ranged from one amino acid to 35 (Figure 5.5A). The average CDR3 lengths for all the ground control animals was  $12 \pm 0$ . The average CDR3 lengths for flight animals was also  $12 \pm 1$ . The CDR3 lengths were not normally distributed in the flight or ground animals (flight  $p < 0.0001$ , ground  $p = 0.0128$ ) with the majority of CDR3 lengths falling between 11 and 14 AAs. We also examined the heavy-chain CDR3 length by isotype (Table 5.5) and by treatment group. We found no significant difference by treatment group or by isotype (two-way ANOVA, interaction  $p = 0.8141$ , isotype  $p = 0.4589$ , treatment  $p = 0.6225$ ).

Kappa-chain CDR3 length was conserved at nine amino acids with 90.3 to 94.2 percent of all light chains having CDR3 rearrangements that were nine amino acids in length (Figure 5.5B). Only a small percentage of light chains had eight amino acids (2.8 to 6.9%), or ten amino acid long (0.9-3.9%) CDR3s. Ground mouse one (0.9%), was an exception and was enriched for 11 amino acid CDR3s (2.3%) compared to other animals (0.2-0.7%). CDR3 lengths over 18

amino acids were not displayed (20, 25, 27, 29, 32, 35, 36, 37, 39). CDR3 rearrangements of these lengths were often only detected in one animal and expression of these lengths did not exceed 0.01% of the repertoire. Additionally, these rearrangements may have been identified in error as many of these rearrangements contain intervening phenylalanine residues within the conserved kappa-chain CDR3 C-XX-F motif.

There was little overlap among IgH CDR3s regardless of treatment (Figure 5.6A, C). Of the top five CDR3s found in each animal, we identified 26 unique CDR3 AA rearrangements. Of those 26 CDR3s, four were found in all six animals. Two additional rearrangements were identified in three animals. One of those rearrangements, CASHGSSYLAWFAYW, was found in only flight animals and not found in any ground animals. Six CDR3s were found in two animals, and the remainder were found in a single animal. The vast majority of CDR3 rearrangements detected were unique to each animal (6,661-9,270 rearrangements), though there was a small amount of overlap among animals (103-163 CDR3 rearrangements). For flight animals, 78 CDR3 rearrangements were found in all three animals and 70 were found in all three ground animals. Of the rearrangements found in all three animals per treatment group, only 20 rearrangements were detected in all six animals.

There was considerable overlap in the top five Igκ CDR3 rearrangements of each animal (Figure 5.6B, D). Seventeen total top CDR3 rearrangements were identified and all CDR3 rearrangements were identified in every animal. One CDR3 was the most abundant rearrangements in five animals (CQNGHSFPLTF), still ranking third in the remaining animal (F1). Ground control and flight animals shared three rearrangements that ranked within the top

20 CDR3 in all animals. Overall, between 1,590-2,933 unique CDR3 were detected in ground control animals, and between 1,461-1,818 unique CDR3 were detected in all flight animals (Figure 5.6). Of the 725 and 579 CDR3 shared in all three ground control and flight animals, respectively, 446 were shared between all six ground and flight animals.

A CDR3 rearrangement nucleotide sequence alignment of one of the top V-D-J-gene segment combinations demonstrates significant variability among mice such that any variability between ground and flight treatment groups cannot be determined with confidence (Figure 5.7). Additional data sets are needed in order to assess the effect of spaceflight on CDR3 formation.

### **Mutations in Complementarity Determining and Framework Regions**

We also examined mutation frequencies in CDR and framework (FW) regions for each animal because mutations can affect Ig specificity. Mutation frequencies were normalized by animal and Ig region (FW1-3, CDR1-3) by dividing the percent of total substitution mutations (total mutations/total productive reads) by respective region length (Figure 5.8). There were no significant differences between the mutation frequencies of ground control and flight animals for any of the Ig regions in both IgH (Student's t-test,  $0.1916 < p < 0.9978$ ) and Igk (Student's t-test,  $0.3175 < p < 0.9865$ ). When comparing the substitution mutation frequency across Ig regions, more substitution mutations occurred in CDR3 compared to other regions in IgH and Igk (ANOVA; all  $p < 0.05$ ).

## Discussion

Spaceflight and ground-based analog models induce phenotypic and functional changes in T- and B- lymphocyte populations. Spaceflight also affects bone marrow, the site of B-cell differentiation and development. Therefore, we wanted to know whether the stress and physiological changes associated with spaceflight would affect the normal development of the highly-diversified and highly-specific antigen receptors on B-lymphocytes. If so, the ability to respond to pathogens might be affected. We characterized the antibody repertoire of C57BL/6Tac mice flown aboard the ISS and ground control animals using HTS and RNA-Seq.

HTS studies of antibody repertoires typically employ polymerase chain reaction amplification of Ig specific sequences from sorted B-cell populations. We assessed the B cell repertoire in whole spleen tissue because of the limitations of the primary science, a verification flight of mouse housing hardware. This precluded the sorting of cell populations. We previously showed that similar data could be generated using whole spleen tissue compared to whole spleen cell suspensions.<sup>45</sup> Although we do not account for B-cell subpopulations,<sup>67,242</sup> we do measure the total splenic Ig repertoire. Additionally, since we did not use specific amplification of Ig sequences the depth of sequencing was not as high as some have accomplished looking at Ig gene usage.<sup>67,242,329</sup> We have compared of amplified and unamplified data sets by our lab show reasonable correlations of the data, with more V-gene segments being detected in the unamplified data sets (Rettig 2017, this thesis). Amplification with multiplex Ig specific primers may introduce amplification bias as primers may bind with varying efficiency to V-gene segments, although there have been recent advances in experimental approach to address amplification bias.<sup>321</sup> We also found that the type of RNA-Seq analysis we are using in the



assessment of younger C57Bl/6J mouse Ig gene usage, correlated well with that of studies using amplification.<sup>45</sup> Therefore, we feel we have a reasonable snapshot of B-cell receptors present in the spleens of 35-week-old female mice. In addition, the sample preparation allows additional data mining of valuable mouse samples.

In both humans and mice, early B-cell development occurs in the fetal liver prior to postnatal development in the bone marrow. We attempted to determine the Ig repertoire within the adult liver of the ground and flight animals in comparison to the splenic Ig repertoire. Unfortunately, few Ig sequences were recovered in liver samples suggesting few B cells are actually resident or circulating in the liver under normal, steady-state conditions. The liver kappa chain V-gene repertoire did correlate some with the usage in the spleen and probably reflects circulating B-cell Ig expression, but we did not confirm that. We focused our efforts on the spleen data.

While previous analyses by our lab used splenic mRNA pooled from four animals,<sup>45,195</sup> the current study assessed individual mice and exhibited significantly more mouse-to-mouse variation than one might expect in inbred mice; even within treatment groups and compared to pooled mouse samples. Overall, V-gene segment usage correlated when analyzed using pairwise linear regression of animals within ground and flight treatment groups and there did not appear to be an impact of spaceflight on B-gene segment use. It is possible that differences in gene segment usage in ground and flight animals would be observed upon immunization. Studies on the effects of spaceflight in an immunized amphibian model showed altered VH-gene family and Ig class usage.<sup>191,325</sup>

We performed a number of analyses to determine if the Ig gene rearrangement process was affected by spaceflight. Averaged V-gene segment usage between the ground control and flight animals was moderately to highly correlated (VH:  $R^2=0.5922$ ,  $p<0.0001$ ; V $\kappa$ :  $R^2=0.831$ ,  $p<0.0001$ ). Many of the most abundant V-gene sequences were shared in flight and ground animals and there was no statistically significant difference in usage of individual top V-gene segments between ground control and flight animals (Student's t-test,  $p=0.06478-0.9609$ ). Similarly, no differences in D-, JH- or J $\kappa$ -gene segment usage and IgH constant region usage was seen between ground and flight animals.

We also examined whether spaceflight would affect the V/J-gene segment combinations that normally occur in specific pathogen-free mice. These too, were not different between ground control and flight animals for IgH and Ig $\kappa$ . V/J-gene segment combinations were moderately to highly correlated (IgH:  $R^2=0.3296$ , Ig $\kappa$ :  $R^2=0.8795$ , both  $p<0.0001$ ). An assessment of the impact of hypergravity on the similarly assembled T cell receptor repertoire of neonatal mice showed low correlation of individual Beta chain V- and J-gene segment recombination frequencies between control animals and animals subjected to centrifugation. 85% of gene V/J-gene segment combinations were not shared among the two treatment groups<sup>330</sup> and the differences could be attributed to changes in somatic recombination machinery under altered gravity conditions.<sup>330,331</sup>

The combinatorial diversity of Ig was shown through the assessment of overall CDR3 sequence overlap among animals, as a large number of sequences were unique to only one animal. Little overlap was observed in the top 5 H-CDR3 rearrangements within all six ground

and flight animals which totaled to 26 CDR3 rearrangements. Thirteen of these CDR3 rearrangements were only identified in one animal. More overlap was observed in the top 5  $\kappa$ -CDR3 rearrangements within all six ground and flight animals which totaled 17 CDR3 rearrangements, which were identified in all animals.

The mice studied in this investigation were not challenged and were housed under specific-pathogen free conditions. Mutational frequency in Ig is normally associated with antigen stimulation.<sup>332</sup> Therefore, we did not expect that these mice were undergoing high amounts of somatic mutations. The majority of the mutations detected occurred in CDR3 in both ground and flight mice. It is possible that we will see differences in mutation frequency between ground control and flight animals after experimental immune challenge. The frequency of somatic hypermutations in *P. waltoni* immunized in space was slightly lower than animals immunized on earth.<sup>326</sup> An experiment with antigen challenge of the Ig repertoire will be necessary to test this hypothesis.

The animals used in this experiment were older (35 weeks). The expression of genes necessary for Ig recombination do go down as mice age.<sup>333</sup> Therefore, it may be possible that we do not see differences between the treatment groups because the perturbations of spaceflight are not enough to disrupt the reduced B cell differentiations occurring in 35-week-old mice under normal steady-state conditions. Additionally, the half-life of B cells in secondary lymphoid organs is between four and seven weeks.<sup>334</sup> Given the brief time that the mice were subjected to space flight (three weeks) it is likely that the exposure was not long enough to affect V/D/J

recombination in the repertoire. Additional experiments, especially with young mice, will be needed to test this hypothesis.

Animals in this experiment were older and supplied from a different vendor (Taconic) than the 9-11-week-old C57BL/6J mice from Jackson Laboratories used in our previous experiments.<sup>45,195</sup> To assess whether differences existed in the Ig repertoire, between the older C57BL/6Tac mice and the younger C57BL/6J mice, we compared gene segment usage between the two mouse cohorts. Because no differences in top V-gene segments, (D)J-gene segments, and constant region usage were detected between RR1 ground control and flight animals, all six animals were pooled and compared to the C57BL/6J cohort. We selected the 25 most abundant V-gene segments from both RR1 and C57BL/6J cohorts, resulting in 33 top IgHV and 34 top IgKV. We found that seven of 35 IgHV and eight of 32 IGKV were expressed at significantly different levels within the repertoire between the two cohorts (Student's t-test). These differences are largely driven by gene segments that are highly expressed in the RR1 cohort such as IGHV1-80, which represented  $2.04 \pm 0.26\%$  of the repertoire in RR1 animals and  $6.6 \pm 2.40\%$  of the repertoire in the C57BL/6J cohort (Student's t-test,  $p=0.0032$ ). This is even more pronounced in IGKV5-39, which represented  $21.41 \pm 7.50\%$  of the repertoire in RR1 animals and only  $4.73 \pm 4.93\%$  of the repertoire in the C57BL/6J cohort (Student's t-test,  $p=0.0109$ ).

We also performed a linear regression of top V-gene segment usage, which showed poor correlation between cohorts in both IgH ( $R^2 = 0.1568$ ,  $p=0.0226$ ) and Igk ( $R^2 = 0.1681$ ,  $p=0.0160$ ). D-gene segment usage was only significantly different IGHD1-1, which represented  $39.7 \pm 5.23\%$  of the repertoire in RR1 animals and  $26.49 \pm 0.94\%$  of the repertoire in the

C57BL/6J cohort (Student's t-test,  $p=0.0040$ ). We found that J-gene segment usage varied between the two cohorts in three out of five IG $\kappa$ J and no differences in IGHV were detected (student's t-test). IgH constant region usage was significantly different for IgD (RR1:  $0.63 \pm 0.22\%$ , C57BL6/J:  $4.60 \pm 0.69\%$ ; Student's t-test,  $p=0.0074$ ) and IgG (RR1:  $23.83 \pm 6.19\%$ , C57BL6/J:  $9.65 \pm 6.19\%$  ; Student's t-test,  $p=0.0074$ ).

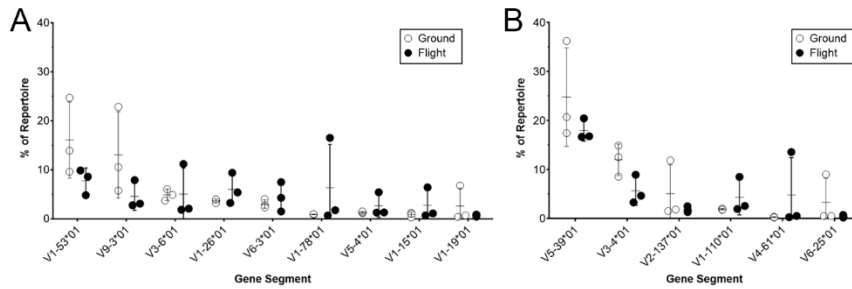
Although we cannot determine whether these differences are attributed to differences in vendor or differences in age, it is likely that repertoire differences are driven by a more mature Ig repertoire within the RR1 animals, as a higher percentage of IgH sequences demonstrate class switching. Both cohorts were unimmunized and maintained under specific pathogen-free conditions.

In conclusion, we have been able to successfully characterize immunoglobulin gene segment usage and junctional diversity within the antibody repertoire of unimmunized C57BL/6Tac mice flown aboard the ISS. Individual gene segment usage remained similar among animals within and among treatment groups, with the most abundant gene segments being conserved across all animals. Gene segment combinations and CDR3 sequences were highly varied, demonstrating the combinatorial diversity of the antibody repertoire, but that variation reflects the dynamics of individualized selection of Ig molecules and not any impact of spaceflight. A larger sample size would help solidify this conclusion, but these data provide preliminary suggestions that the recombinatorial processes that lead to the diverse Ig repertoires in mice are not affected by a short trip to and stay on the ISS. These data do not preclude that differences in the Ig repertoires of ground and flight animals will not be seen during active

immunization or in younger, possibly more recombinatorily active mice. Current studies in our lab aim to characterize antibody repertoire dynamics upon antigen challenge using a murine anti-orthostatic suspension model and during a future space flight.

## Figures and Tables

**Figure 5.1. Expression of Top V-Gene Segments**



(A) VH- and (B) Vκ-gene segment usage for gene segments representing over five-percent of the repertoire in at least one animal within ground and flight treatment groups. No significant difference in individual gene segment usage was detected between ground control and flight treatment groups (Student's *t*-test,  $0.1534 < p\text{-value} < 0.9609$ ). Significant differences between gene segment usage of combined ground and flight animals were found. In IgH, V1-53 was more abundant than many of the top V-gene segments (V1-78, V5-4, V1-15, V1-19); (Student's *t*-test, all  $p < 0.05$ ). In Igk, V5-39 was more abundant than many of the top V-gene segments (V3-4, V2-137, V1-110, V4-6, V6-2); (Student's *t*-test, all  $p < 0.05$ ).

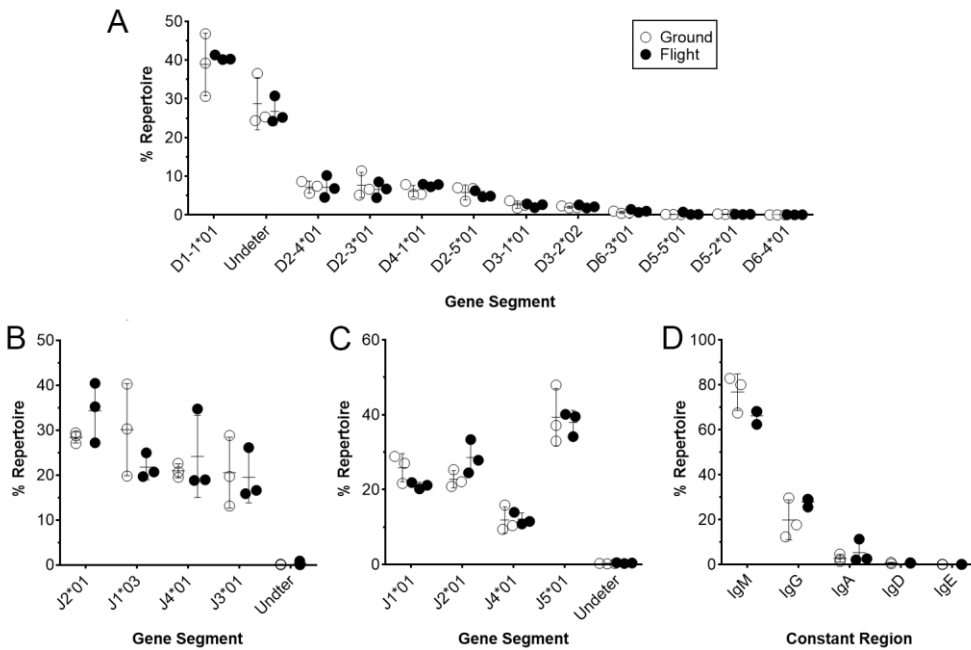
**Figure 5.2. Expression of Top-V $\kappa$  Gene Segments from Spleen and Liver**

	LG1	LG2	LG3	LF1	LF2	LF3		SG1	SG2	SG3	SF1	SF2	SF3
V5-39*01	4	1	4	1	2	2	V5-39*01	1	1	1	1	1	1
V3-4*01	5	2	2	5	10	5	V3-4*01	2	2	2	5	5	2
V1-110*01	12	5	10	14	5	8	V1-110*01	9	10	11	14	9	3
V2-137*01	16	9	3	11	20	13	V2-137*01	16	9	3	8	14	19
V3-2*01	2	3	31	40	4	15	V3-2*01	19	8	21	20	8	14
V6-20*01	3	14	44	20	36	15	V6-20*01	13	37	32	15	29	23
V4-91*01	52	30	58	34	36	3	V4-91*01	47	40	33	32	18	7
V4-61*01	47	20	76	62	3	19	V4-61*01	60	48	57	55	2	37
V4-81*01	8	4	1	1	1	1	V4-81*01	82	81	81	82	83	78
V6-25*01	6	90	58	94	70	76	V6-25*01	3	41	38	28	59	54
V6-14*01	1	78	76	76	87	76	V6-14*01	50	71	63	73	57	69

V $\kappa$ -gene segment usage for gene segments representing over five-percent of the repertoire in at least one animal from the liver or spleen of ground or flight animals are presented by rank. Liver ground (LG) and liver flight (LF) rankings are shown to the left and spleen ground (SG) and spleen flight (SF) rankings are shown to the right. V-gene segments are listed most frequent to least frequent. Dark red indicates higher rank moving to blue, lower percent rank.

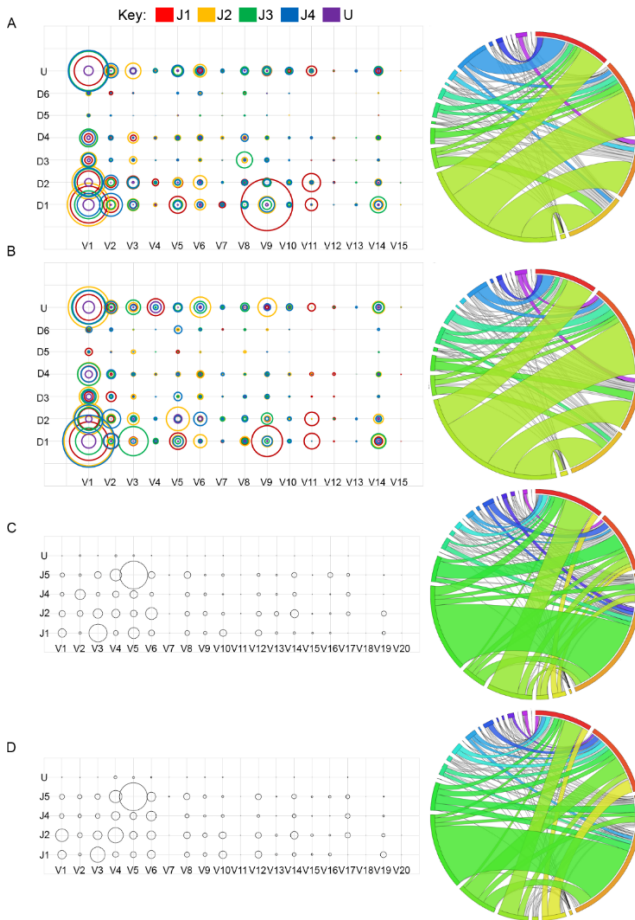


**Figure 5.3. Expression of D and J-Gene segments and IgH Constant Region Usage**



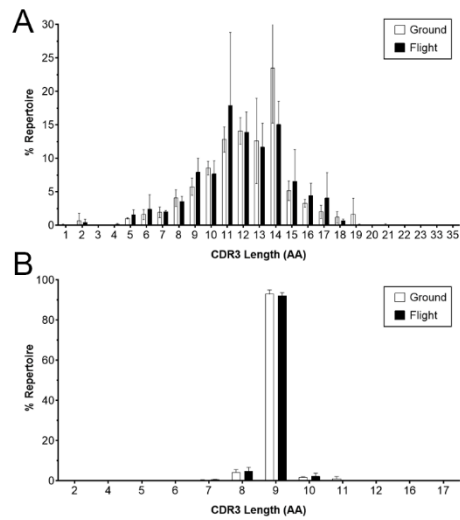
(A) D-gene segment, (B) JH-gene segment, (C) Jκ-gene segment, and (D) IgH constant region usage in animals within ground and flight treatment groups are presented as percent of repertoire. No significant difference in individual gene segment usage was detected between ground control and flight treatment groups (Student's *t*-test, D:  $0.1542 < p\text{-value} < 0.9840$ , JH:  $0.2060 < p\text{-value} < 0.8662$ , IgH Constant Region:  $0.1075 < p\text{-value} < 0.8277$ , Jκ:  $0.0977 < p\text{-value} < 0.9262$ ). D1-1 was used at a significantly higher rate than all other D-gene segments (Student's *t*-test, all significant  $p < 0.05$ ). No significant differences between JH-gene segment usage were found (Student's *t*-test, all  $p < 0.05$ ). In Igκ, significant differences in expression were found between all gene segments except between J1 and J2 (Student's *t*-test, all significant  $p < 0.05$ ).

**Figure 5.4. Gene Segment Combinations in Ground Control and Flight Animals**



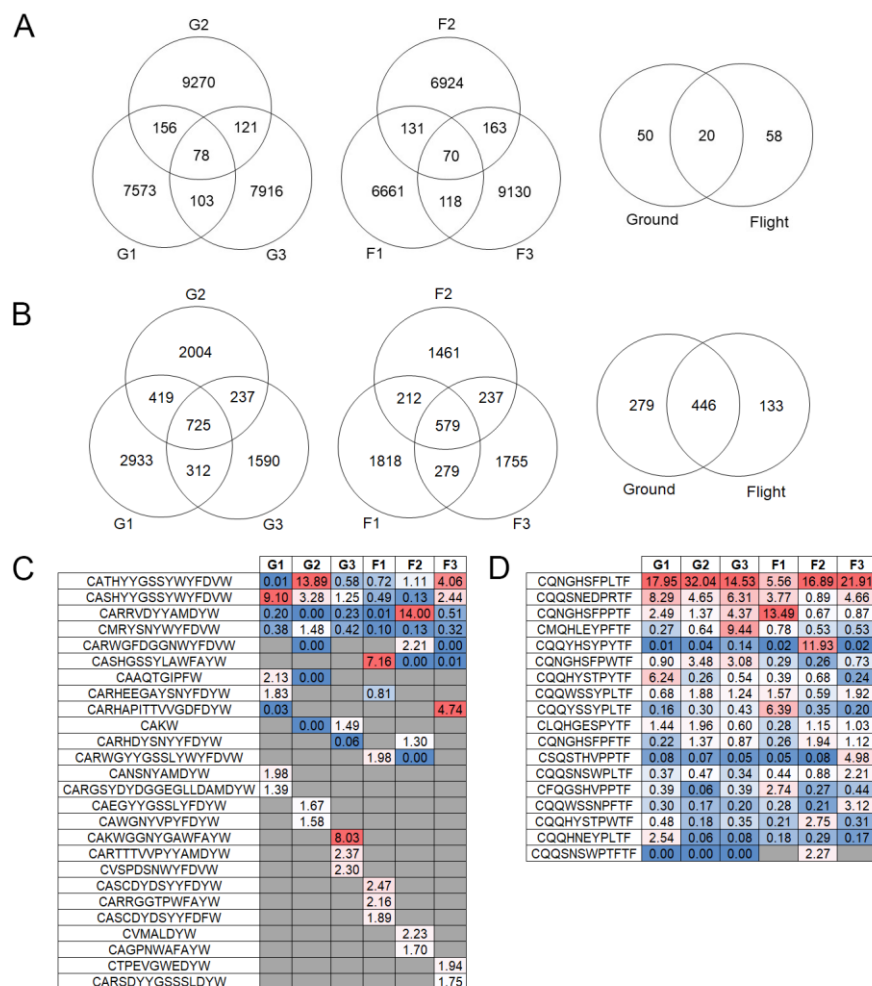
(A, B) Average IgH V/D/J combinations (bubble chart) and the V/J combinations (Circos plot) for ground treatment (A) animals and (B) flight animals. For bubble charts, V-gene family is represented along the x-axis, the D-gene segment is represented along the y-axis, and the J-gene segment is represented by a specific color. The size of the bubble corresponds to the average percent repertoire of the specific gene combination. Circos plots are read clockwise starting at the 12 o'clock position with J1 (red), J2, J3, J4, U, V1 (lime green), V2, V3, V4, V5, V6, V7, V8 (light blue), V9, V10, V11, V12, V13, V14, and V15 (sliver, no color). (C,D) Average Igk V/J combinations for ground treatment (C) animals and (D) flight animals. For bubble charts, V-gene family is represented along the x-axis and J-gene segment is represented along the x axis. The size of the bubble corresponds to the average percent repertoire of the specific gene combination. Circos plots are read clockwise starting at the 12 o'clock position with J1 (red), J2, J4, J5, U, V1 (yellow), V2, V3, V4, V5, V6, V7 (sliver, no color), V8, V9, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19 (light purple).

**Figure 5.5. CDR3 Length in IgH and Igκ sequences**



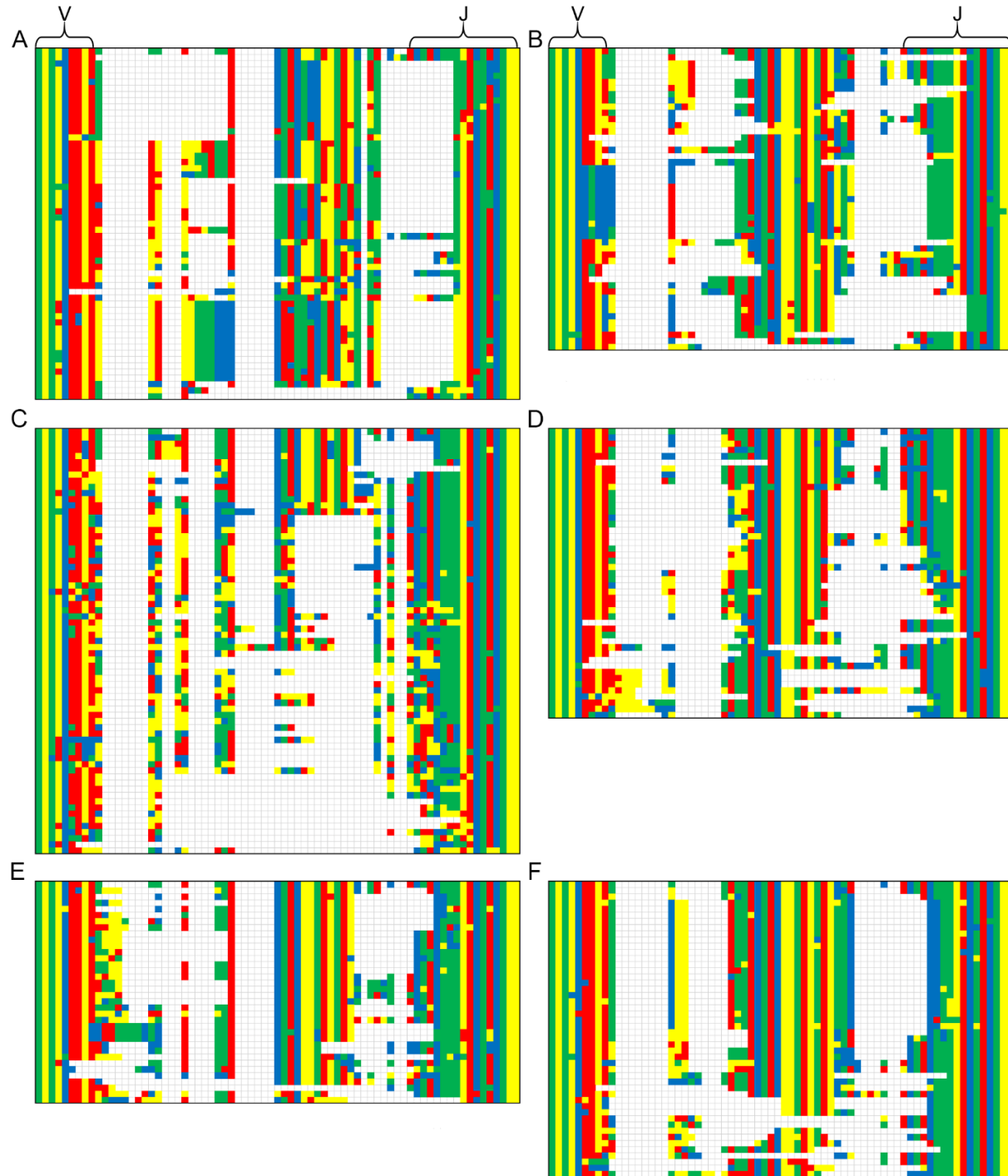
(A) IgH and (B) Igκ CDR3 amino acid length of ground control and flight animals as mean-average with standard deviation.

**Figure 5.6. Top CDR3 Usage and Overlap of CDR3 Between Treatment Animals**



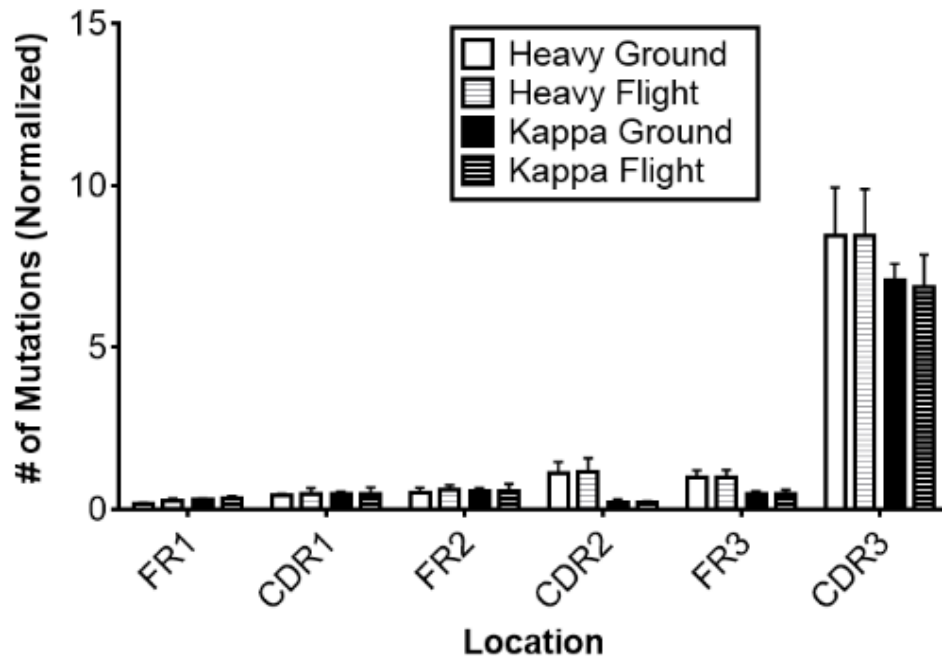
Venn diagrams show the overlap of unique CDR3 sequences among and between ground control (G1-G3) and flight treatment (F1-F3) groups in (A) IgH and (B) Igκ. CDR3 sequences were ranked within the top 5 most abundant rearrangements of any ground control or flight animals in (C) IgH and (D) Igκ. Dark red indicates higher percent of repertoire moving to blue, which represents lower percent of repertoire.

**Figure 5.7. Nucleotide Alignment of CDR3 from top V-D-J Combination**



Nucleotide alignment of heavy-chain gene segment V1-26\*01/D1-1\*01/J1\*03 across individuals in ground (G1 - A, G2 - C, G3 - E) and flight (F1 - B, F2 - D, F3 - F) treatment groups. Brackets in the germline region of the first individual in each treatment group delineate V- and J-gene regions. These bracketed regions remain the same across all individuals in the treatment group.

**Figure 5.8. Substitution Mutations by Ig Region**



Total number of substitution mutations in FRs 1-3 and CDRs 1-3 were observed for both IgH and Igκ chains. Abundance was first normalized by region length and then by total number of cleaned, productive reads in each respective data set and multiplied by 100 to attain percent abundance.

**Table 5.1. Spleen Sequencing Read Counts in Ground (G) and Flight (F) Mice**

	G1	G2	G3	F1	F2	F3
<b>Raw Reads<sup>a,b</sup></b>	51.4 M	45.2 M	43.5 M	40.5 M	48.4 M	55.8 M
<b>Cleaned<sup>a,c</sup></b>	13.2 M	31.4 M	30.9 M	14.6 M	13.0 M	14.1 M
<b>IgH IMGT<sup>d,e</sup></b>	124,102	104,135	149,675	85,802	66,909	181,703
<b>Igκ IMGT<sup>d,e</sup></b>	175,406	140,963	103,841	107,851	81,889	107,928

<sup>a</sup>M=Million

<sup>b</sup>Raw reads reflect unfiltered FASTQ files imported from the Illumina MiSeq personal sequencing system.

<sup>c</sup>Cleaned reads were quality trimmed to remove the first 12 base pairs, reads with a Phred score under 20, and sequences less than 40 nt in length.

<sup>d</sup>Mapped Ig Sequencing reads of productive or unknown functionality were obtained from the IMGT HighV-Quest tool.

<sup>e</sup>There is no statistical significance by student's t-test in the number of reads mapped to IgH ( $p=0.7215$ ) and Igκ ( $p=0.1424$ ).

**Table 5.2. Comparison of Flight and Ground V-Gene Segment Usage**

Comparison	VH R <sup>2</sup>	V $\kappa$ R <sup>2</sup>
<b>G1 v G2</b>	0.620	0.660
<b>G2 v G3</b>	0.356	0.738
<b>G1 v G3</b>	0.695	0.678
<b>F1 v F2</b>	0.101	0.465
<b>F2 v F3</b>	0.225	0.516
<b>F1 v F3</b>	0.360	0.606
<b>Mean G v F</b>	0.592	0.810

Pairwise linear regressions of VH- and V $\kappa$ -gene segment usage were performed among ground control (G) and flight (F) animals. A linear regression was performed on the mean-average V gene segment of ground and flight treatment groups.

<sup>a</sup>All comparison groups were correlated ( $p \leq 0.0001$ )



**Table 5.3. Vκ Liver Sequencing Read Counts in Ground (G) and Flight (F) Mice**

	G1	G2	G3	F1	F2	F3
<b>Raw Reads<sup>a,b</sup></b>	43.9 M	42.9 M	38.4 M	49.2 M	48.3 M	39.7 M
<b>Cleaned<sup>a,c</sup></b>	18.8 M	35 M	31.5 M	24.7 M	18.2 M	19.4 M
<b>Igκ IMGT<sup>d,e</sup></b>	1429	1267	592	543	425	438

<sup>a</sup>M=Million

<sup>b</sup>Raw reads reflect unfiltered FASTQ files imported from the Illumina MiSeq personal sequencing system.

<sup>c</sup>Cleaned reads were quality trimmed to remove the first 12 base pairs, reads with a Phred score under 20, and sequences less than 40 nt in length.

<sup>d</sup>Mapped Ig Sequencing reads of productive or unknown functionality were obtained from the IMGT HighV-Quest tool.

**Table 5.4. V-J Linear Regression Analyses**

Comparison	IgH R <sup>2</sup>	Igκ R <sup>2</sup>
G1 vs G2 <sup>a</sup>	0.534	0.750
G2 vs G3 <sup>a</sup>	0.124	0.7209
G1 vs G3 <sup>a</sup>	0.256	0.657
F1 vs F2 <sup>b</sup>	0.010	0.539
F2 vs F3 <sup>a</sup>	0.067	0.598
F1 vs F3 <sup>a</sup>	0.103	0.738
G vs F AVG <sup>a</sup>	0.330	0.880

Pairwise linear regression analyses of IgH and Igκ V/J-gene segment combination abundances were performed among grounds control (G) and flight (F) animals. A linear regression was performed on the mean-average V/J-gene segment combination abundances of ground and flight treatment groups.

<sup>a</sup>Comparison group was correlated ( $p < 0.0001$ )

<sup>b</sup>Comparison group was correlated for IgH ( $p = 0.035$ ) and Igκ ( $p < 0.0001$ )

**Table 5.5. CDR3 Length by Isotype**

Isotype	CDR3 Amino Acid Length					
	G1	G2	G3	F1	F2	F3
IgA	10	13	12	13	11	12
IgD	11	12	12	12	12	12
IgE	14	11	12	14	11	15
IgG	12	12	13	13	11	12
IgM	12	12	12	12	11	12

The mean-average CDR3 length of ground control (G) and flight (F) animals is displayed by isotype. Assessment by two-way ANOVA revealed no significant differences in CDR3 length by treatment group ( $p=0.6225$ ), isotype ( $p=0.4589$ ), or the interaction of the two variables

## **Chapter 6 - Effects of tetanus toxoid and skeletal unloading on the antibody repertoire of C57BL/6J mice**

### **Abstract**

Spaceflight is known to affect multiple aspects of the immune system, though while there are many studies on the subject, many are inconclusive for both humans and mouse models. One important part of the adaptive immune system is antibodies, which are vital for clearing infectious agents from the host. Each B cell produces a unique antibody composed of two identical heavy chains paired with two identical light chains. These chains are comprised of variable (V), diversity (D, heavy chain only), and joining (J) gene segments combined with a constant region. The gene segments are permanently rearranged in the germline and during rearrangement variations in gene segment selection and the joining of those segments, diversity is generated in the antibody repertoire. We assess the antibody repertoire response to antiorthostatic suspension (AOS), a physiological model of spaceflight, tetanus toxoid (TT), and CpG ODN (CpG).

We detected subtle changes in the repertoire after challenge with TT and CpG in both the heavy and light chain. We detected changes to V-, D-, and J-gene segments with AOS, TT, and CpG treatment. We also detected changes to V- and J-gene segment pairing for both heavy and light chain. There were lower levels of IgG detected in AOS animals than in loaded animals and increased levels of IgA in CpG treated animals. We also assessed class-switched heavy chain antibody sequences and detected changes in V-gene segment usage and identified potentially TT-specific CDR3 AA sequences. Our results demonstrate that small changes were detectable in the total splenic antibody repertoire of mice subjected to AOS and challenged with TT and CpG.

## Introduction

The spaceflight environment has dramatic influence on various immune functions. Studies in animals and humans have shown that spaceflight affects the total body, thymus and spleen mass,<sup>189,201-204,206,208,210,211,335-337</sup> circulating corticosterone,<sup>210,212-215,217-219,337,338</sup> mitogen-induced proliferation, cytokine production and reactivity,<sup>179,189,210,213,215,220,221,224,225,228-232,339-344</sup> and lymphocyte subpopulation distributions.<sup>211,233-236,343,345,346</sup> While many of these changes are dramatic, the spaceflight studies conducted to date are inconclusive with respect to how those changes affect the ability of the host to resist an actual *in vivo* immune challenge. However, given that spaceflight clearly disrupts immune cell population distributions and function *in vivo* and *ex vivo*, it seems likely that the spaceflight environment is detrimental to an immune response essential for the host to remain healthy during flight.

The ability of the host to resist infection is often related to the repertoire of activated B cells and their secreted antibodies. The general selection process for antigen-specific idiotypes has been known for some time.<sup>347-349</sup> The variable region of the immunoglobulin heavy chain is encoded by variable (V), diversity (D), and joining (J) gene segments in the immunoglobulin locus on mouse chromosome 12 or human chromosome 14 with separate loci for immunoglobulin light chains.<sup>350,351</sup>

To date, there have been very few studies characterizing the impact of the spaceflight environment on antibody responses. There are some indications that humoral responses are down-regulated in flight. When immunized 8 days prior to flight with sheep red blood cells, rats returning from an 18.5 day COSMOS flight had lower IgG concentrations compared to both

immunized and non-immunized ground controls.<sup>213</sup> More recently, IgM production was virtually eliminated in lymphocytes cultured and activated with pokeweed mitogen (PWM) on board the International Space Station (ISS) when compared to similarly activated ground controls. Furthermore, when B cells were activated in culture prior to flight, stored frozen, and then resuspended and assayed in space, their IgM production was slower than similarly treated ground-based controls.<sup>239</sup> In contrast, in another type of assay system, there were no significant differences in immunoglobulin levels, regardless of isotype class, after the Skylab missions,<sup>352</sup> nor after short-duration (10-11 day) space shuttle missions.<sup>194,232</sup> These data suggest individual and assay-dependent impacts. There have also been at least three independent studies which indicate that there may be an enhanced antibody response after flight. After Apollo missions 9-11, there were increases found in circulating IgG and/or IgA levels after landing which may be linked to an on-board infection.<sup>233</sup> After a 16-day space shuttle flight, total un-stimulated plasma IgE levels were elevated compared to preflight values.<sup>353</sup> After a 5-month stay aboard the ISS, splenic transcription levels of IgY (the newt counterpart of IgA) generated in response to food antigens was three times higher than similarly treated ground control values.<sup>325</sup> Despite the inconsistencies in these studies, the clear conclusion is that the humoral immune system is responsive to the space environment.

Tetanus toxin is secreted by the gram-positive bacterium, *Clostridium tetani*. The toxin interferes with neurotransmitter release causing systemic neuromuscular dysfunction. The impact is seizures and disruption of the autonomic nervous system leading to death. The tetanus toxoid (TT) is a formaldehyde-inactivated form of the toxin that has been approved for use as a vaccine in humans by the Food and Drug Administration. In the United States, TT is usually given as

part of a standard vaccination schedule during the first year of life and is often given in conjunction with vaccinations for diphtheria and pertussis with booster shots given to adults every 10 years.<sup>354</sup>

The antibody response to an antigen in a vaccine can sometimes be less than optimal. To overcome this, immunologists have developed adjuvants. Although the mechanisms of action are not fully understood, adjuvants generally prolong exposure and/or augment the innate response during the initial exposure to the antigen, thereby increasing phagocytosis and antigen presentation.<sup>355</sup> Adjuvants can also enhance responsiveness by activating signal transduction through toll-like receptors although there are still some details to be determined.<sup>356</sup> The adjuvants may improve both cell-mediated and humoral responses and the development of immunological memory.

Adjuvants come in many forms, ranging from components of inactivated bacteria and viruses to oil emulsions, and aluminum salts.<sup>355</sup> Although there are an increasing number of options available for adjuvants, we focused on a synthetic oligodeoxynucleotide (ODN) containing unmethylated CpG motifs (CpG). CpG is a relatively stable dinucleotide sequence that is found naturally and more frequently in viral and bacterial DNA than in vertebrates.<sup>357</sup> CpG motifs have been shown to appear with 20-fold greater frequency in bacterial DNA compared to mammalian DNA.<sup>358</sup> This sequence is recognized and differentiated from similar (methylated) sequences in vertebrate DNA by the innate immune system.<sup>359,360</sup> CpG ODN has already been shown to be an effective immune-stimulator in a few pre-clinical studies, either alone or in combination with other therapies including vaccines, chemotherapy and radiotherapy.<sup>361,362</sup>

CpG acts through the TLR9 pathway.<sup>361</sup> TLR9 is typically found within the endolysosomes of B cells and plasmacytoid dendritic cells.<sup>358,363,364</sup> However, the binding of CpG to TLR9 triggers a cascade of responses that includes the activation of innate populations (e.g. macrophages and neutrophils), the upregulation of surface receptors critical to antigen presentation (e.g. MHC-II and B7), the differentiation and proliferation of lymphocyte populations (e.g. B and T cells), and the release of various cytokines (e.g. IFN- $\gamma$ , IL-6 & -12, GM-CSF, and TNF- $\alpha$ ).<sup>365,366</sup> Typically, this drives Th1-dominated immune responses,<sup>367</sup> ultimately enhancing the expansion of antigen-specific B-and T-cell idiotypes, augmenting the antibody response, and improving the development of immunological memory.<sup>361</sup> CpG has been shown to improve TT-specific IgG production when given with inoculation, even above levels generated when using alum, the only adjuvant actually FDA-approved for use in humans.<sup>365</sup> CpG has also been shown to be effective in inducing memory B cells to proliferate, mature, and begin secreting antibody after a previous exposure to TT.<sup>368</sup>

In this investigation, we used antiorthostatic suspension (AOS) to induce some of the physiological changes that are associated with spaceflight. This technique was originally developed by Morey-Holton, et al.<sup>369</sup> (Morey-Holton and Wronski 1981) but has been used successfully in mice, as well.<sup>165</sup> We hypothesize that exposure to AOS will lead to a diminished capacity to generate antigen-specific B cells. Specifically, we hypothesize that AOS will lead to a diminished capacity of B cells to rearrange antibody genes which will affect both overall levels of all immunoglobulin (Ig) types, as well as antigen-specific IgG. We will detail TT-specific Ig



levels, memory generation, V-, D-, and J- gene segment use, V/(D)/J combinations, and CDR3 characteristics and use.

## **Materials and Methods**

### ***Suspension, Immunization, and TT specific IgG Analysis***

Ten-week-old female C57BL/6J mice were unloaded using antiorthostatic tail suspension (AOS). Ten animals were included in each treatment group. Two independent suspensions were done (40 mice per experiment). Animals from each of eight treatment groups were randomly sacrificed using 100% CO<sub>2</sub> followed by cervical dislocation on four different collection dates; two consecutive collection days per experiment, 20 mice per day. Animals were subjected to AOS for two weeks and then immunized with intraperitoneal injection of saline, 5 LF/mL tetanus toxoid (TT), 0.4 mg/mL CpG ODN 1826 (CpG), or both TT and CpG. Treatment group designation for this presentation will be by the convention (AOS, TT, CpG) with a – signifying that the animal did not receive treatment and + signifying the animal did receive treatment. For example, an animal receiving no treatments will be represented as (---) and an animal receiving all treatments as (+++). A standardized collection assembly-line was used to euthanize the animals and collect the tissue to minimize experimental variation. Blood, spleen and bone marrow were collected first as primary science for this experiment with other tissues collected thereafter. Blood was collected via cardiac puncture. Blood was allowed to clot, then centrifuged at 10,000g for five minutes at 4°C, collected and frozen at -80°C. Tissues including spleen were collected and snap frozen in LN<sub>2</sub>. Samples were stored in -80°C prior to analysis. Serum TT-specific IgG was measured from the second set of 40 mice by ELISA per manufacturer's instructions (Alpha Diagnostic International, San Antonio, TX).

### ***RNA Extraction and Sequencing***

Tissue extraction was performed as described previously.<sup>45,195</sup> Briefly, spleen tissue was processed with Trizol according to the manufacturer's instructions. Thirty-two spleen samples were sequenced, two from each AOS trial for a total of four animals per treatment group. Samples were selected for sequencing based on highest RIN scores. Total RNA was sequenced on the Illumina MiSeq 2x300 nucleotide platform at the Kansas State University Integrated Genomics Facility using a reduced fragmentation (one minute) protocol to result in longer sequences. Standard Illumina sequencing protocols include the use of oligo-dT selection for mRNA sequences and reverse transcription using random hexamer primers.

### ***Bioinformatics and Analysis***

Bioinformatic analysis was performed as previously outlined.<sup>45,195</sup> Briefly, Illumina sequencing results were imported into CLC Genomics Workbench v10.2 (<https://www.qiagenbioinformatics.com/>) and cleaned to assure high quality reads using a Phred score of over 20 for 97% of the sequence. Paired and merged (overlapping pairs) sequences were mapped to reference V-gene segments and their respective loci to collect potential antibody sequences. These sequences were collected and submitted to ImMunoGeneTic's (IMGT) High-V Quest for bioinformatic analysis. One sequence per Illumina sequence ID was analyzed as outlined in previous work.<sup>45,58,195</sup> Functionality was identified by IMGT, using their definitions, but binding ability or specificity was not assessed. Functional sequences are in frame and do not have a stop codon while unknown functionality sequences do not contain enough information to determine functionality, a result of short sequences.

Gene segment abundance was assigned as in Rettig et al.<sup>195</sup> with V-gene segments matching only one possibility being assigned a value of one, and partial matches of two gene segments a 0.5 value. Gene segments with more than one match were not included. CDR3 motifs were also analyzed as in Rettig et al.<sup>195</sup> with a C-xx-W motif or class switching (heavy chain) or a C-xx-F motif (kappa chain) required for a functional assignment. V(D)J pairing was also assessed as outlined in Rettig et al.<sup>195</sup> by identifying functionally productive sequences with a single identified V-gene segment. D- and J-gene segments that were not reported by IMGT, contained less than six nucleotides, or identified multiple sequences were reported as undetermined. Total counts were used to generate Circos graphs using Circos Online.<sup>285</sup>

For memory marker expression, sequencing results were mapped to the NCBI mouse reference genome<sup>370</sup> and Transcripts Per Million (TPM) were generated for each animal using CLC's RNA-Seq tool. TPMs were then analyzed using Three-way ANOVA. Fold-change was calculated by dividing the average TPM per gene by the average TPM in the negative control (---) group.

### ***Isolation of TT-Specific Heavy Chain CDR3s (H-CDR3s)***

Eight-week-old conventionally housed, specific pathogen free female C57BL/6 mice were challenged with 5 LF/mL TT and 0.4 mg/mL CpG three times. Challenges occurred between two and three weeks after the previous vaccination. Two weeks after the final challenge, animals were euthanized with isoflurane overdose followed by cervical dislocation (Huerkamp 2000) and spleen and bone marrow were collected. The spleen was homogenized through a 70µm sieve to create a single cell suspension. The cell suspensions from two mice were pooled for subsequent processing. Magnetic beads (Dynabeads M-270 Carboxylic Acid, ThermoFisher,

Waltham, MA) were conjugated with TT (26uL of 5LF/mL TT), per manufacturer's instructions and added to the cell suspension. Between  $12.5 \times 10^7$  and  $21.6 \times 10^7$  cells were incubated at  $1 \times 10^7$  cells/mL of isolation buffer for 20 minutes with the TT-coated magnetic beads per manufacturer's instructions. After the final purification wash step, Trizol was then added to the cell-bead conjugates and RNA was extracted per the manufacturer's instructions. RNA was prepared for sequencing as outlined above and analyzed using our standard bioinformatic workflow outlined above. Potential TT-specific H-CDR3s were collected by identifying sequences that occurred in all three mouse pools and occurred in the top 100 of at least one pool. The average H-CDR3 percent of repertoire of the potential TT-specific H-CDR3s was compiled using previous results from normal mice that our laboratory outlined in Rettig et al. 2018.<sup>195</sup> The average of the potential TT-specific H-CDR3s was compared to the average of the normal mouse H-CDR3 repertoire. H-CDR3 sequences meeting the criteria describe above and that occurred at a two-fold or higher rate than detected in the normal mouse data were considered TT-specific H-CDR3s.

### ***Statistical Analysis***

Three way ANOVAs were analyzed by creating a linear model in R to analyzed the three variables, AOS, TT, and CpG and their interactions in R. Post-hoc comparisons were analyzed using a Tukey HSD function in R. Difference of least square means was analyzed by first calculating variance for each gene segment (V, D, and J). Variation was calculated as  $(\% \text{ of repertoire per animal} - \text{Average } \% \text{ of repertoire})^2$ . Whole mouse variation was determined by summing that animal's variance for all gene segments. Variances per animal where then compared using Least Squared Means analysis in SAS v9.4. Linear regressions and correlation

coefficients were calculated in Graphpad Prism v6.0. Graphs were generated in Graphpad Prism v6.0. Circos graphs were generated using Circos Online.<sup>285</sup>

## **Results**

### ***Read Counts***

We obtained an average initial transcript read count between 38.4 and 44.5 million reads per animal (Table 6.1). After cleaning, an average of 24.2 to 28.2 million reads remained for analysis (Table 6.1). The average number of total assessed reads for the heavy chain was between about 83,000 and 109,000 reads and about 65,000 and 85,000 reads for kappa chain (Table 6.1). There were no significant differences among initial, cleaned, or final read counts for any treatment for main effects or interactions (Three-way ANOVA,  $p>0.05$ ).

### ***TT-Specific IgG Response***

TT-specific IgG was measured by ELISA in the serum to confirm seroconversion (Figure 6.1). Significant main effects were detected for TT and CpG (Three-way ANOVA,  $P=0.03$ ) and significant interactions were detected for TTxCpG (Three-way ANOVA,  $P=0.02$ ). AOS status did not affect the TT-specific IgG levels in either run.

### ***Memory Markers***

To analyze mobilization of B cells and memory cell formation, we assessed the level of several phenotypic molecules thought to be regulated by B cell activation and/or the formation of a memory response. Specifically, we did RNASeq (Table 6.2), on markers of B cells (B220), plasma cells (CD138), activated B cells (CD80), naïve B cells (CD19), memory B cells (CD27), antigen-primed, resting B cells (CD44), and stimulated B cells (CD62-L). Only transcriptional changes with a greater than two-fold change and P-value of  $<0.05$  were considered significant. Using this approach, there were no significant changes in response to any treatment group.

### ***VH Gene Usage***

To understand whether AOS affected the normal process of immunoglobulin gene rearrangement or if there were changes to the B-cell repertoire, we examined which V gene segments were used in over 10% of the repertoire in at least one of the treatment groups (Figure 6.2A). Eleven V-gene segments (V1-26, V3-6, V1-80, V1-53, V6-3, V8-8, V8-12, V1-42, V5-4, V9-4, V1-66) were detected at over 10% in at least one treatment group (Figure 6.2A). Of these V-gene segments, only treatment with CpG had a significant effect on V-gene segment usage. Treatment with CpG decreased the usage of the V1-80 gene segment (Three-way ANOVA,  $p=0.02$ ). When we examined all V-gene segments, there were significant treatment effects on multiple V-gene segments. Unloaded mice showed increased levels of V10-3 and V4-1, but decreased levels of V1-63 and V1-76 (Three-way ANOVA,  $p<0.05$ ; Appendix A.10). Administration of TT decreased expression of V1-31, V-74, V1-76, and V1-85 (Three-way ANOVA,  $p<0.05$ ). Treatment with CpG decreased expression of V10-3 and V1-36, but increased expression of V11-2, V1-63 and V1-80 (Three-way ANOVA,  $p<0.05$ ; Appendix A.10). Significant interactions were observed between AOSxTT for V11-1, AOSxCpG for V1-12, V15-2, and V1-63, and TTxCpG with V10-3, V1-12, V1-64, and V1-69 (Three-way ANOVA,  $p<0.05$ ; Appendix A.10). A significant interaction effect on V14-1 was also detected for AOSxTTxCpG treatment group. To determine if AOS had some impact on V gene use, we also compared V gene segment usage between the loaded PBS only control animals (Negative Control) and the unloaded (AOS) TT+CpG animals. Only one V-gene segment, V1-76, showed a significant change, a decrease in the AOS treatment group (Three-way ANOVA,  $p=0.03$ ; Appendix A.10). There was no significant difference in the variation of V-gene segment usage in the heavy chain by treatment group (Difference of Least Square Means,  $P>0.05$ ).

We also used linear regression to assess how well various treatment groups compared. The coefficient of determination ( $R^2$ ) (Table 6.3, all  $p < 0.0001$ ) ranged from 0.5563 to 0.8858 and showed moderate to high levels of correlation by variable (Table 6.3). The average correlation for the heavy chain in the AOS treatment group was 0.7284, TT was 0.7103, and CpG was 0.7013.

### ***V $\kappa$ Gene Usage***

Complete antibodies require light chains in humans and mice. Therefore, we also examined the V $\kappa$ -gene segments that comprised over 10% of the repertoire for the kappa chain (Figure 6.2B). Six V-gene segments (V5-39, V4-55, V8-30, V3-4, V6-23, V4-71) were used in over 10% of the repertoire in at least one treatment group (Figure 6.2B). Of these V-gene segments, V5-39 showed increased usage when CpG was administered while V4-55 usage decreased (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). V4-55 showed decreased usage with AOS+CpG treatment, TT+CpG treatment, and AOS+TT+CpG treatment. V5-39 showed increased usage in TT+CpG animals when compared to TT+NoCpG animals (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). V8-30 showed increased usage in NoTT+NoCpG animals vs NoTT+CpG animals and in HLU+NoTT+CpG animals when compared to all other treatment groups (Three-way ANOVA,  $p < 0.05$ ; Appendix A-11).

When examining all V-gene segments, significant treatment effects were observed. Unloaded mice showed an increased usage of V1-132 and V4-86 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). TT treatment decreased the usage of V12-46 and V4-70 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). Treatment with CpG increased usage of V5-39, V5-48, and V8-19, but

decreased usage of V4-55 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). Significant interactions were observed with TTxCpG for V4-57, V4-62, V6-14, V5-48, and V8-30 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). There was no significant difference in the variation of V-gene segment usage in the kappa chain by treatment group (Difference of Least Square Means,  $P > 0.05$ ).

The  $R^2$  for V $\kappa$  usage was assessed as was done for the heavy chain (Table 6.3, all  $p = < 0.0001$ ).  $R^2$  ranged from 0.1849 to 0.8164 and generally showed moderate to high levels of correlation by variable, though it appeared that the V gene correlation between the control and the use of CpG in the presence of TT (-+- vs -++) was drastically lower at 0.1849 (Table 6.3). The average correlation for AOS in the kappa chain was 0.6297, TT was 0.5947, and CpG was 0.4586.

#### ***D, JH, J $\kappa$ and IgH Gene Usage***

We also assessed D- and J-gene segment usage in the repertoire. Regardless of treatment, D1-1 was the most commonly used D-gene segment (Figure 6.3A). This was followed by D2-4, D2-3, D4-1, and D2-5 (Figure 6.3A). Undetermined D-gene segments comprised a large amount of the repertoire, as seen previously (Figure 6.3A).<sup>196</sup> Administration of TT caused a decrease in usage of the D5-5 gene segment. Significant interactions were detected for AOSxTT for D3-1 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.10), but no other significant differences were detected for main effects or interactions. JH-gene segment usage was similar for all gene segments, but administration of TT decreased the level of JH2 usage (Figure 6.3B), (Three-way ANOVA,  $p < 0.05$ ; Appendix A.10). Significant interactions were detected for JH2 with AOSxTT and AOSxCpG for JH3 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.10). Isotype switching in



response to antigen challenge suggests an active host immune response. Constant region usage was predominated by IgM regardless of treatment group (Figure 6.3C), but AOS treatment animals had increased IgM usage and decreased IgG (Three-way ANOVA,  $p < 0.05$ ; Appendix A.10). Animals treated with CpG showed higher levels of IgA (Three-way ANOVA,  $p < 0.05$ ; Appendix A.10).

J $\kappa$  showed similar usage levels of J $\kappa$ 1, J $\kappa$ 2, and J $\kappa$ 5, but lower levels of J $\kappa$ 4 (Figure 6.3D). Treatment with TT significantly decreased the number of undetermined J $\kappa$ -gene segments, but no other significant main effects or interactions were detected (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). There was no significant variation in D, JH, or J $\kappa$  usage among all treatment groups (Difference of Least Square Means,  $P > 0.05$ ).

### ***Gene segment combinations***

Recombination and combinatorial diversity are hallmarks of the immunoglobulin molecule. Changes in how gene segments are combined could impact the Ig repertoire and the ability of a host to respond to antigen. To assess the impact of AOS, TT and CpG on V/(D)/J assembly, we compiled the top five most common V/(D)/J combinations in each treatment group. Using this approach, there were 29 unique V/D/J combinations detected for the heavy chain among our eight treatment groups and 22 for the kappa chain (Figure 6.4). The most common V/D/J combination found in all eight groups was V1-26/D1-1/J1. It ranged from being the first to the eighth most common combination among our eight treatment groups. Nevertheless, other combinations such as V1-80/U/J3 and V1-55/D1/J1 were also present in all eight treatment groups (Figure 6.4A). Some combinations however, such as V5-9-1/D2/J4, were more restricted and were only detected in three treatment groups (---, +--, +-+) and only at a high

level in one (+--) (Figure 6.4A). Two other combinations (V1-47/D2-5/J3, V1-43/U/J4) were found in seven of eight treatment groups (Figure 6.4A).

Due to the high variation in D-gene segment usage among our mice, we also examined how well V/J pairings correlated between treatment groups (Table 6.4). Correlations were lower than the V-gene segment correlations, ranging from 0.6140 to 0.1414 (Table 6.4). AOS significantly increased the frequency of V1-12/J2, V1-22/J1, V1-55/U, V1-77/J4, VH3-8/J2, and V8-4/J2 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.12). AOS significantly decreased the frequency of VH1-76/J1, VH1-76/J4, VH2-3/J4, and VH9-2/J4 combinations (Three-way ANOVA,  $p < 0.05$ ; Appendix A.12). TT treatment significantly increased the frequency of V3-8/J2 and V6-3/U pairings but decreased the frequency of VH4-1/J1, VH1-76/J1, V1-85/U, and V2-3/J4 pairs (Three-way ANOVA,  $p < 0.05$ ; Appendix A.12). CpG treatment significantly increased the frequency of V10-3/J2, V1-53/J2, V1-53/J3, V1-74/J2, V5-17/J1 and V9-2/J4 pairings (Three-way ANOVA,  $p < 0.05$ ; Appendix A.12) but it decreased V11-2/J1, V1-26/U, V15-2/J1, V1-53/J4, and V5-15/J1 pairing frequency (Three-way ANOVA,  $p < 0.05$ ; Appendix A.12). We also present the V/J combinations by family to provide a more generalized view of V/J recombination (Circos graphs presented in Figure 6.5A).

We also assessed V/J combinations for the kappa chain (Figure 6.4B). The most common V/J combination found in all eight treatment groups was V5-39/J5; ranging from the first to the fifteenth most common combination. Many other combinations (V10-96/J1, V1-117/J1, V1-110/J1, V4-55/J4, and V4-55/J5) were also commonly found ranking between first and fifteenth in usage their respective treatment groups (Figure 6.4B). Unlike the heavy chain, all light chain

gene-segment combinations were found in all eight treatment groups, but light chain gene-segment combinations such as V1-117/J4, V8-30/J1, V-80/J4, V8-30/J2, and V1-122/J4 were found commonly only in one treatment group (Figure 6.4B).

V/J pairing frequencies in the kappa chain were generally lower than heavy chain pairing frequencies (Table 6.4) ranging from 0.6440 to 0.1439. AOS significantly increased the frequency of pairings for V10-95/U, V4-51/U, V4-63/J5, V6-14/U, V4-58/J4, and V4-86/U but decreased frequencies for V14-111/J2, V19-93/U, and V4-91/J5 (Three-way ANOVA,  $p < 0.05$ ; Appendix A.13). TT treatment significantly increased frequency of V4-73/U, V4-86/J4, V6-25/J1, V6-29/J4, V8-18/J4, and V4-58/J4 pairings (Three-way ANOVA,  $p < 0.05$ ; Appendix A.13). It also significantly decreased the frequencies of V12-46/J5, V14-100/U, V2-109/J2, V3-7/J1, V7-33/U, and V8-21/U pairs (Three-way ANOVA,  $p < 0.05$ ; Appendix A.13). CpG treatment significantly increased V11-125/U, V3-12/U, V4-53/J5, V4-58/J1, V4-73/U, V5-39/J1, V5-48/J1, V6-17/J5, V8-21/J2, V8-21/U, V8-24/J4, V9-123/J5, and V8-16/J1 pairing frequencies (Three-way ANOVA,  $p < 0.05$ ; Appendix A.13). We represent the V/J combination by family in Figure 6.5B.

### ***CDR3 Length***

CDR3 length has been shown to change in response to vaccine challenge (Poulsen, Meijer et al. 2007). The average heavy chain CDR3 length was 11 AAs for treatment groups --+, +--, and +-+ and 12 AAs for treatment groups ---, -+-, -++ , +-+ for (Figure 6.6A). The average CDR3 length for all treatment groups in kappa chain was 9AAs (Figure 6.6B). In the heavy chain, AOS treatment decreased the number of 14 AA CDR3s and CpG decreased very short 3AA CDR3s (Three-way ANOVA,  $p < 0.05$ ; Appendix A.10). Also, in the heavy chain, AOSxTT

and TTxCpG interactions affected 9 AA CDR3s (Three-way ANOVA,  $p < 0.05$ ; Appendix A.10). In the kappa chain, TT treatment decreased the number of 7 AA CDR3s (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11). Interactions were detected for HLUxTT with 10 AA CDR3s and with 8AA CDR3s for HLUxTTxCpG (Three-way ANOVA,  $p < 0.05$ ; Appendix A.11).

### ***Assessment of Changes in CDR3***

Previous studies (Greiff, Menzel et al. 2017, Rettig, Ward et al. 2017, Rettig, Ward et al. 2018, Ward, Rettig et al. 2018) have shown that CDR3 repertoires are highly unique. The data from this experiment are consistent with those observations (Table 6.5). Over 91% of the heavy chain CDR3 AA (H-CDR3) repertoire and 65% of the kappa chain CDR3 AA repertoire were found in only a single treatment group (Table 6.5). Only 560 (0.16%) of the H-CDR3s, and 3.68% of  $\kappa$ -CR3s were shared among all eight treatment groups (Table 6.5).

We also assessed the overlap of CDR3s between treatment groups. Both the loaded (non-AOS) and the AOS-treated mice shared 11% of their heavy chain repertoire (20,649 CDR3 AA sequences) (Figure 6.7A). The non-TT and the TT treated animals also shared 11% of their heavy chain repertoire (20,472 CDR3 AA sequences) (Figure 6.7A). The non-CpG and CpG treated animals only shared 4% of their heavy chain repertoires (8,043 CDR3 AA sequences) (Figure 6.7A). The kappa chain shared more CDR3s between the treatment groups (Figure 6.7B) than seen in the heavy chain. The non-AOS and AOS mice shared 43% and 44% of their repertoires respectively (7,625 CDR3 AA sequences) (Figure 6.7B). The non-TT and TT treated animals also shared 43% and 44% of their repertoires respectively (7,648 CDR3 AA sequences) (Figure 6.7B). The non-CpG and CpG animals shared 45% and 42% of their repertoires, respectively (7,558 CDR3 AA sequences) (Figure 6.7B).

Due to the lower overlap of CpG animals seen in the heavy chain, we assessed the overlap of the TTxCpG interaction in both heavy (Figure 6.8A) and kappa chain (Figure 6.8B). In the heavy chain, between 15% and 17% of the repertoire was shared among treatment groups (Figure 6.8A). In the kappa chain, between 29 and 30% of the repertoire was shared (Figure 6.8B). There appears to be no difference in CDR3 sharing among the TT and CpG treatment groups.

### **Assessment of Class Switched (IgG/IgA) Sequences**

In effort to narrow down the TT-specific response, we focused our analysis on class switched sequences, hypothesizing that these sequences were most likely to be TT-specific. As only the heavy chain can perform class switching, these analyses are limited to only heavy chain sequences with an identifiable constant region. We analyzed V-gene segments comprising over five percent of the repertoire in at least one treatment group (Figure 6.9). We detected 11 VH-gene segments. Of these 11 gene segments, six (V1-80, V1-26, V3-6, V8-8, V9-4, and V-16) were detected in our whole-repertoire analysis (Figure 6.3A, Figure 6.9). Statistical analysis revealed that of these 11 gene segments, significant changes were only seen in V4-1, which was used at higher levels in AOS animals than loaded animals (Three-way ANOVA,  $P=0.02$ ). When looking at all V-gene segments, AOS treatment increased usage of the V14-3, V1-81, and V4-1 gene segments, and decreased usage in V-10-1 and V2-2 (Three-way ANOVA,  $P<0.05$ ). TT challenge increased the usage of V2-5 and decreased usage of V1-85 and V5-9 (Three-way ANOVA,  $P<0.05$ ). Treatment with CpG increases usage of V1-81 and V5-9, but decreased usage of V1-84 (Three-way ANOVA,  $P<0.05$ ). Significant interactions were seen for AOS x TT for V1-23, V1-5, V1-9, V2-2, and V8-5 (Three-way ANOVA,  $P<0.05$ ). Interactions were also seen

for AOSxCpG in V15- and V2-3 (Three-way ANOVA,  $P < 0.05$ ). TTxCpG interactions were found for V1-64 and V8-11 and AOSxTTxCpG were found for V1-12, V1-55, V1-72, and V5-15 (Three-way ANOVA,  $P < 0.05$ ).

We also assessed CDR3 overlap within the class switched antibodies, as we did previously with the whole repertoire (Figure 6.10). While the AOS and TT treatment groups showed decreased levels of sharing, down to 3-4% from 11%, the CpG treatment group showed similar levels of sharing remaining at 3-4% (Figure 6.7A, Figure 6.10). We also analyzed the CDR3 overlap of the animals with and without TT and CpG treatment with class switched CDR3s (Figure 6.11). Interestingly, we see high levels of sharing among all treatment groups (42-53%) compare to analysis of the whole repertoire (15-17%) (Figure 6.8A, Figure 6.11).

We also analyzed likely TT-specific CDR3 sequences that were selected from B cells collected from hyperimmunized animals. To be considered likely TT-specific, a CDR3 had to be detected in all three animal pools and occur in the top 100 most frequent CDR3s in at least one pool. We identified 19 potential CDR3s and analyzed their frequency in our data set. Only one CDR3, CARPNWDRYFDVW, showed significant changes, increasing frequency in animals treated with TT, and decreasing in frequency in animals treated with CpG (Three-way ANOVA,  $P = 0.04$ ). However, this CDR3 was only detected in six of our 16 total TT treated animals, so the biological significance of this finding maybe low. We also detected a significant interaction affects for the CDR3 CARGYYW in AOSxCpG animals.

## Discussion

Space flight is known to have multiple, but individually disparate impacts, on the host immune response. However, little is known about the overall impact on the ability of a host to survive a pathogen challenge. To begin to address this, we used an experimental design to include an experimental challenge with a common vaccine antigen, TT, and skeletal unloading using AOS. Serum TT-specific IgG antibody concentrations in conjunction with the adjuvant CpG indicated a robust IgG response to antigen even in the context of AOS. Therefore, there was a successful immunization as measured 14 days after challenge. To parallel this, we characterized the antibody transcriptome response in the spleen. There were no significant impacts on total transcript reads and final antibody sequence reads (heavy chain and the kappa chain) which suggests that treatment did not significantly affect our ability to assess the repertoire and that there was no treatment group bias in our transcriptome. The number of transcripts that we were able to assess actually improved from our previous attempts to assess the repertoire of normal, conventionally-housed mice<sup>195</sup> or normal mice that were flown to the international space station.<sup>58</sup> We were able to successfully assess memory markers, the V-, D-, and J-gene segment repertoires, gene segment combinations, and CDR3 length and overlap in the current paper.

We identified eleven VH-gene segments that occurred at over 10% of the repertoire in at least one treatment group. Of these eleven gene segments, five (V1-26, V3-6, V1-80, V1-53, and V6-3) were detected at high levels in the normal conventionally-house mice from our previous study,<sup>195</sup> but none were detected at above 8% of the repertoire. These data are consistent with other studies of the immunoglobulin gene repertoire. Of our eleven gene segments, seven were

also detected at high levels by Collins et al.<sup>14</sup> Treatment with CpG significantly decreased levels of V1-80, likely due to increased expression of other V-gene segments, V10-3 and V1-36. Interestingly, treatment with tetanus toxoid didn't increase expression of any VH-gene segment but did decrease expression of five V-gene segments, suggesting they are not used in the TT response. This pattern is consistent with the impact of an adjuvant that stimulates all the B cells vs. a smaller number of clones that respond in an antigen-specific response. Although it is not clear what the cells are responding to when just given CpG, the mice were not in a germ-free environment. Therefore, the number of environmental antigens that mice<sup>371</sup> or humans<sup>372</sup> can respond to under normal living conditions is considerable.

There were some subtle changes in V-gene usage with AOS. Unloaded animals used less V1-63 and V1-76, but more V10-3 and V4-1. When we examined the homology of the gene segments whose expression went up, there was 69% homology between V10-3 and V4-1. There was 90% homology between V1-63 and V1-76. There was only 47% homology between the genes that increased in frequency and those which decreased, suggesting that there was some selection of V-gene usage in response to AOS. Consistent with this, there was greater change in V-gene use in situations with multiple treatments. Significant interactions for AOSxTT, AOSxCpG, TTxCpG, and AOSxTTxCpG were also detected for multiple VH-gene segments. These statistically-different subtle changes to the repertoire suggests that complex perturbations can impact the B-cell repertoire. Unfortunately, it is still not clear if there is a functional impact on the ability of the mice to survive a virulent pathogen. Given that there was also no statistically significant variation among treatment groups for over all VH-gene segment usage, it is unlikely



that the host response would be modified significantly. AOS, TT, or CpG showed moderate to high levels of correlation in their V-gene use ( $R^2 = 0.5563$  to  $0.8858$ ).

We found six highly expressed V $\kappa$ -gene segments in at least one treatment group. Three of these gene segments (V5-39, V4-55, and V3-4) were also found at high levels in the other normal mouse groups that we have examined in the past in the spleen and liver with V5-39 reaching over 10% of the repertoire.<sup>58,195</sup> V5-39, V4-58, and V8-19 also increased with the administration of CpG, while V4-55 was decreased. Unloaded mice showed increases in V1-132 and V4-86. TT animals also did not show specific V $\kappa$ -gene segment elevations, but rather decreases in V12-46 and V4-70, suggesting that these V-gene segments are not important to the TT response. Significant interactions were observed for TTxCpG, but no others and there was no significant variation among treatment groups. Correlations (TT vs. no TT, CpG vs. no CpG, AOS vs. no AOS) were lower than those found in VH-gene segment usage ranging from 0.1849 to 0.8164 which is consistent with the lower importance the light chain plays in determining specificity.<sup>296</sup>

As found previously with mice flown to the ISS,<sup>58</sup> there were few differences in D-, JH-, and J $\kappa$ -gene segment usage in response to AOS. D-gene usage was consistent with normal C57BL/6J mouse usage with the most common D-gene segments (D1-1, D2-4, D2-5, D4-1, and D2-5).<sup>195</sup> TT treatment did increase the use of D5-5, a low frequency gene segment, and significant interactions were detected for AOSxTT for D3-1. TT administration also increased JH2 usage and interactions were detected for AOSxTT and AOSxCpG for JH2. These data also support the hypothesis that complex physiological changes can alter the B cell repertoire.

There was a discrepancy between the impact of AOS on mouse IgG protein in serum and splenic transcript levels in our experiment. Although mice subjected to AOS had lower numbers of IgG transcripts and higher numbers of IgM transcripts, the level of circulating TT-specific IgG was not changed. This difference most likely reflects the kinetics of the cellular response to antigen and the longer lasting IgG protein levels compared to IgG transcripts. This would be consistent our methodology and the kinetics of a B cell response. Two weeks after immunization would have been past the peak in transcriptional activation in response to antigen. Suspended animals also showed a decreased amount of CD138, a plasma cell marker,<sup>373</sup> suggesting that AOS was shutting the activated plasma cells down sooner.

We looked at expressed CDR3s as a surrogate to B cell idiotypes present in the repertoire. The length of the CDR3s in heavy and light chain also suggest that there are subtle alterations in the B cell repertoire; especially with multiple physiological challenges. Overall, there was little difference in the average CDR3 length was similar among treatment groups (11 or 12AAs) for the heavy chain, and 9AAs for kappa chain which was similar to our previous normal mouse data.<sup>58,195</sup> AOS decreased 14AA length H-CDR3s and CpG decreased 3AA H-CDR3s. Interactions were observed for 9AA sequences with AOSxTT and TTxCpG. Kappa chain showed treatment with TT decreasing the frequency of 9AA  $\kappa$ -CDR3s and interactions for 10AAs with AOSxTT and 8AAs for HLUxTTxCpG.

Another way to assess the impact of individual treatments was to assess the number of CDR3s shared between treatment and control groups (Figure 6.7). That is, changes in the CDR3

overlap could reflect that a treatment is affecting the repertoire. The overlap in CDR3 usage between normal mouse pools in previous experiments averaged about 5%.<sup>195</sup> CDR3 overlap among normal mice in this experiment ranged 6-8% normal, untreated mice (---) (data not shown). The increased overlap could be due to increased sample size increasing the possibility of detecting low frequency CDR3s in multiple treatment groups. The current data sets contained up to 40 times more unique CDR3s than our previous data sets. Mice subjected to AOS or TT had slightly more CDR3 overlap compared to normal mice. This might suggest that the treatments narrowed the repertoire of B cells present in the population. It also supports the hypothesis that the treatments were narrowing the B cell repertoire, a phenomenon that was seen previously in antigen-stimulated plasma cells.<sup>44</sup> CpG treatment also had a strong impact which is consistent the mechanism by which CpG works; a non-specific activation of B cells through the TLR9 receptor and expansion of activated B cells. Interestingly, V-gene segment overlap went down with CpG treatment. If the CpG-treated mice were responding to a discrete environmental antigen, this could explain the diminished overlap of the V-gene repertoire after CpG. In contrast, there were no changes in shared CDR3s in response to any of the treatments light chain V-genes. This reaffirms the importance of the heavy chain in determining antibody specificity.

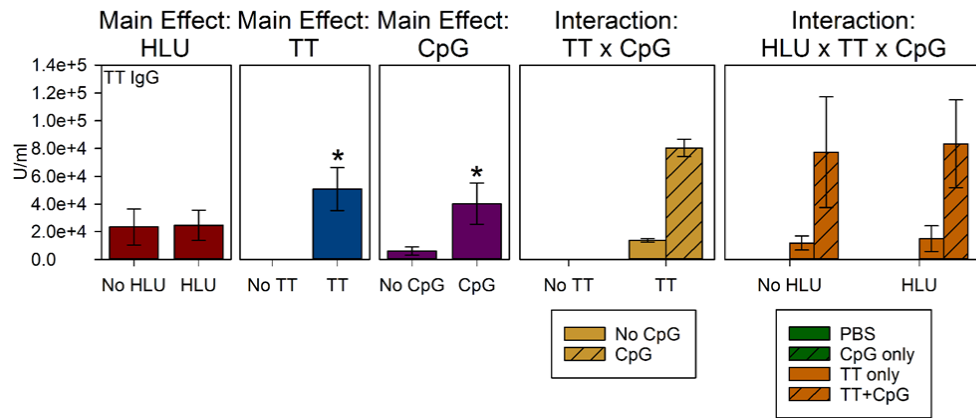
Due to the lower percentage of shared CDR3s and the lower V-gene segment usage  $R^2$  in the CpG- treated animals, we also assessed the compounding effects of TT and CpG treatment, without the influence of AOS. There was no difference in CDR3 sharing of the unique H-CDR3s (15-17%) or the  $\kappa$ -CDR3s (29%-30%) when including both the TT and CpG variable. While there were no differences in CDR3 sharing in the TT+CpG treatment animals, the combined results show that CpG's affect V-gene segment usage and CDR3 usage.

Previous studies in mouse and human have shown that the TT response generates many divergent idiotypes.<sup>71,82,92-95</sup> The anti-TT H-CDR3 repertoire is estimated between 50 and 100 unique CDR3s.<sup>62,96</sup> We attempted to isolate TT-specific CDR3s by analyzing the class-switched repertoire, hypothesizing that these sequences were more likely to be antigen specific. We did find shifts in the VH-gene segment usage of class switched compared to non-class switched repertoire, though none of the major VH-gene segments showed statically significant differences with TT administration. We also looked at CDR3 sharing in only class switched sequences and found decreased sharing among AOS and TT treated animals when examining only a single variable, however when we examined both TT and CpG treatment, we found increased sharing compared to the whole repertoire, even among untreated animals. We also detected as single CDR3, CARPNWDRYFDVW, that was found at higher levels in TT animals, but lower in CpG treated animals. However, due to the low prevalence in the repertoire (6/16 TT treated animals), there is some question about the biological significance of the change.

Overall, our data suggest that AOS, TT and CpG treatments had subtle impacts on the B cell repertoires of the treated mice. Changes were exacerbated by multiple treatments but still were subtle. Since animals were only vaccinated a single time we are only assessing the early host response. We realize that ground-based studies would benefit from looking at antigen-specific cell populations. However, this experiment was done preliminary to an experiment to be done on the ISS where cell isolation and selection methods will not be possible. Later follow-up studies looking at the primary and secondary response during spaceflight have been planned and may show a larger impact.

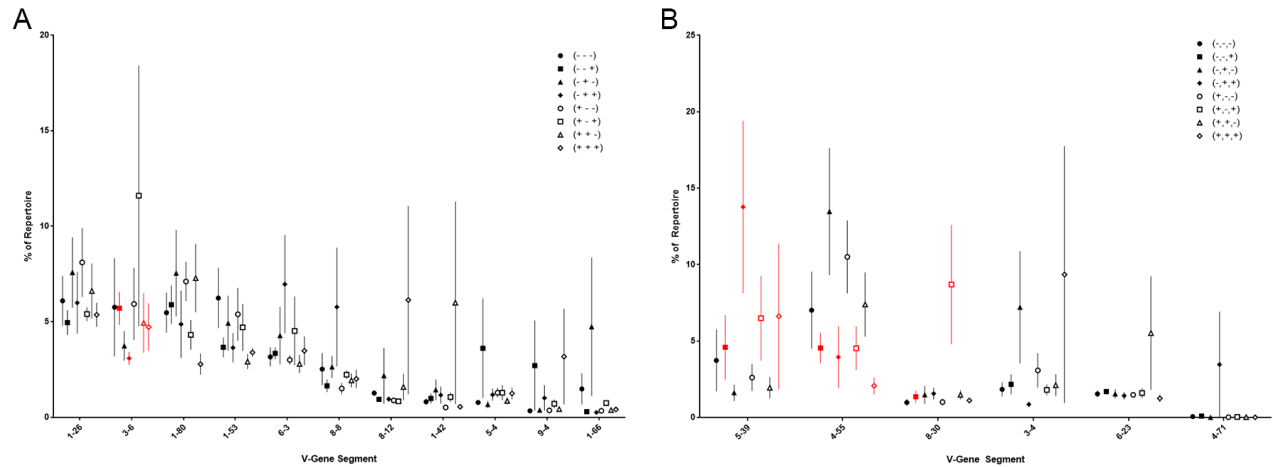
## Figures and Tables

**Figure 6.1. Impact of AOS on Serum Tetanus Toxoid-specific IgG.**



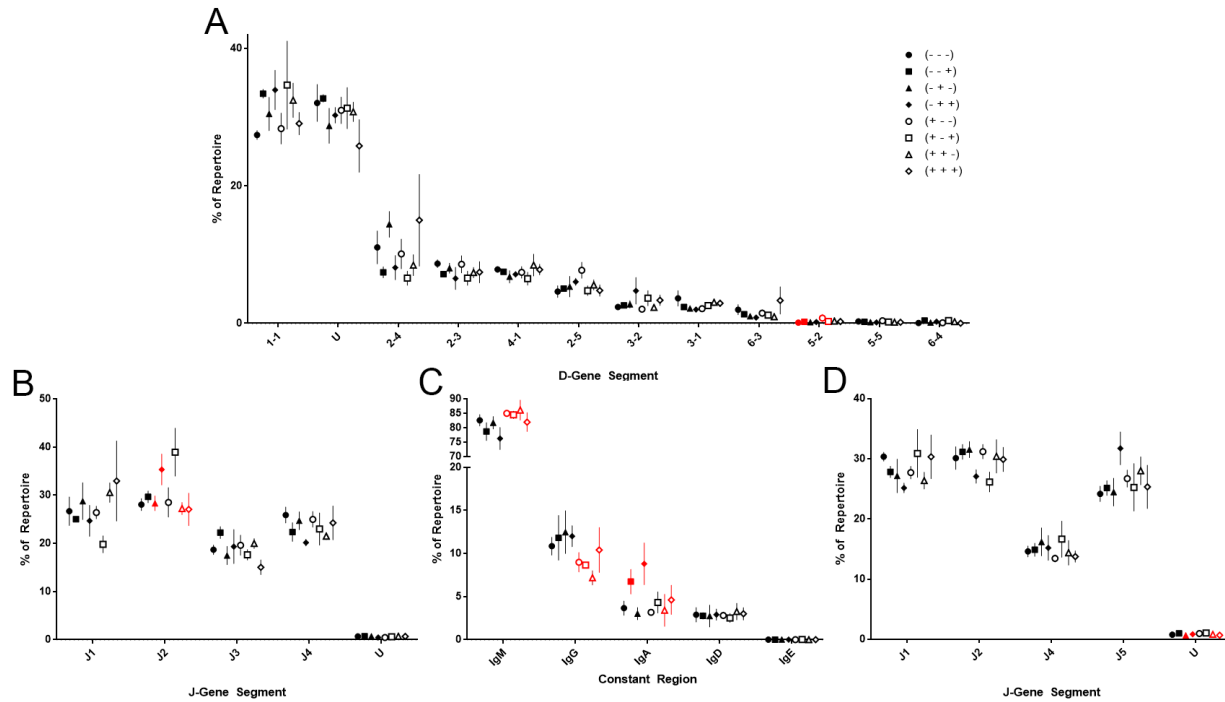
Values represent mean  $\pm$  SEM. N= 3-5 mice/group. Three-Way ANOVA was done as described in the methods; \*= P<0.05 for a Main Effect.

**Figure 6.2. Assessment of High Frequency V-Gene Segment Usage Among Treatment Groups.**



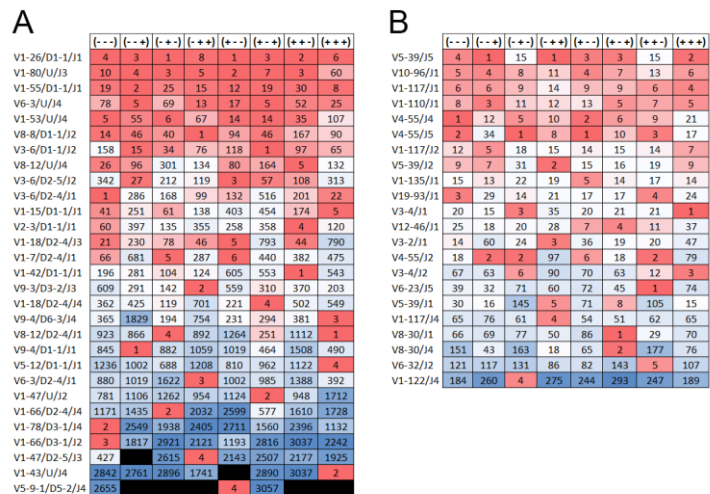
V-gene segments comprising over 5% (average) of the total repertoire in at least one of the treatment groups are presented mean  $\pm$  SEM (n=4 mice per treatment group). Red coloring indicates statically significant ( $P<0.05$ ) differences between similarly treated AOS vs. control animals. (A) Eleven VH-gene segments were detected at over 5% of the repertoire. (B) Six V $\kappa$ -gene segments were detected at over 5% of the repertoire.

**Figure 6.3. Impact of AOS, Tetanus Toxoid and CpG on V, D, J and constant-gene usage. Immunoglobulin heavy chain D- (A), JH- (B), Constant Region (C) and Jκ-Gene Segment (D) Usage are presented by Treatment Group.**



All gene segments are represented, displayed by average and SEM. Red coloring indicates statically significant ( $P < 0.05$ ) differences between similarly treated AOS vs. control animals. Gene segments that are unable to be identified by IMGT because of sequence ambiguities are considered undetermined (U).

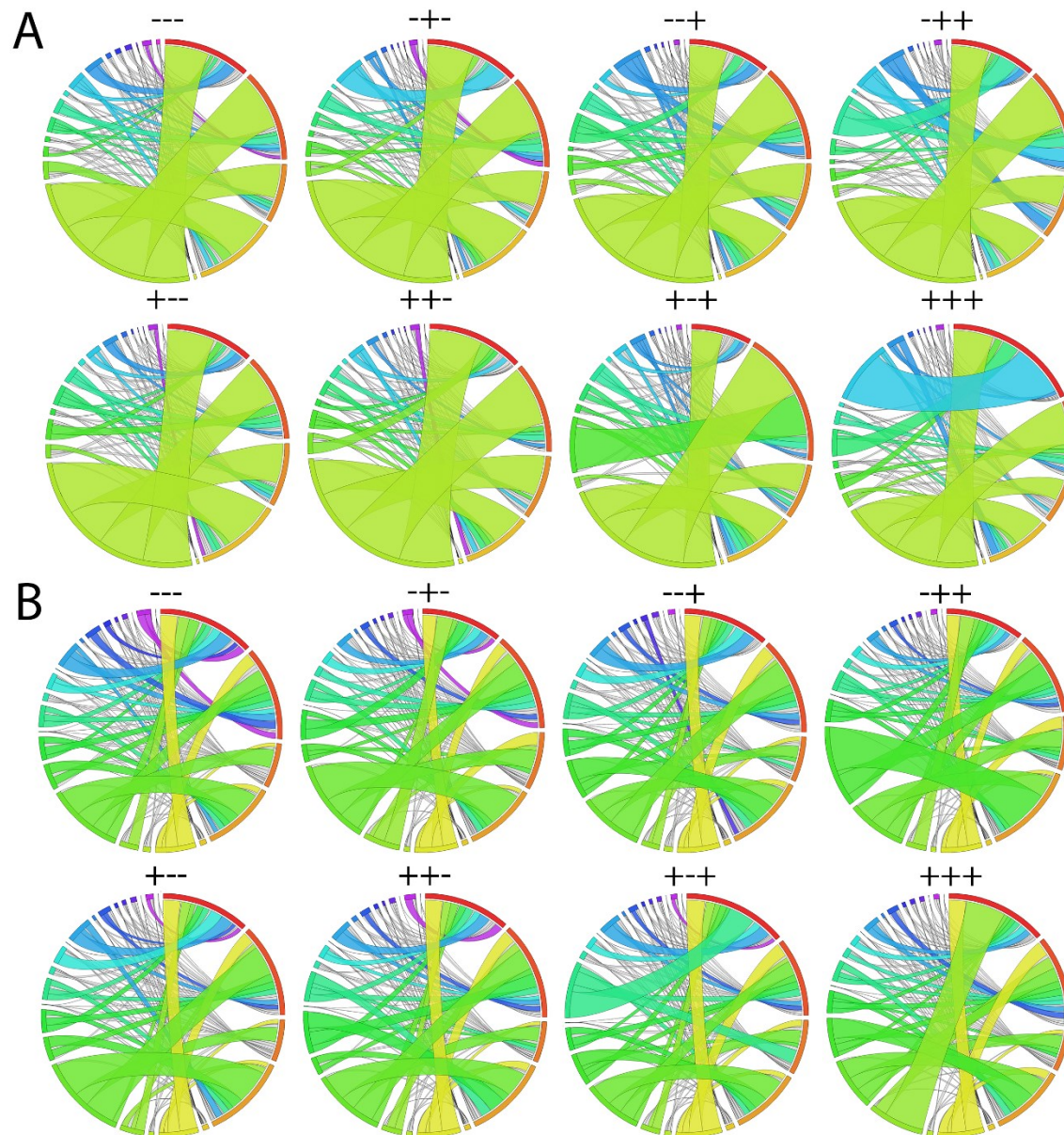
**Figure 6.4. Ranking of High Frequency V-/D-/J-Gene Segment Combinations for Heavy (A) and V-/J-Kappa (B) Chain.**



The top five most common V(D)J combinations (average) per treatment are shown by rank. Heat map frequencies range from dark red reflecting the more common (higher ranking) to blue reflecting the less common (lower ranking). Black signifies that the combination was not found.

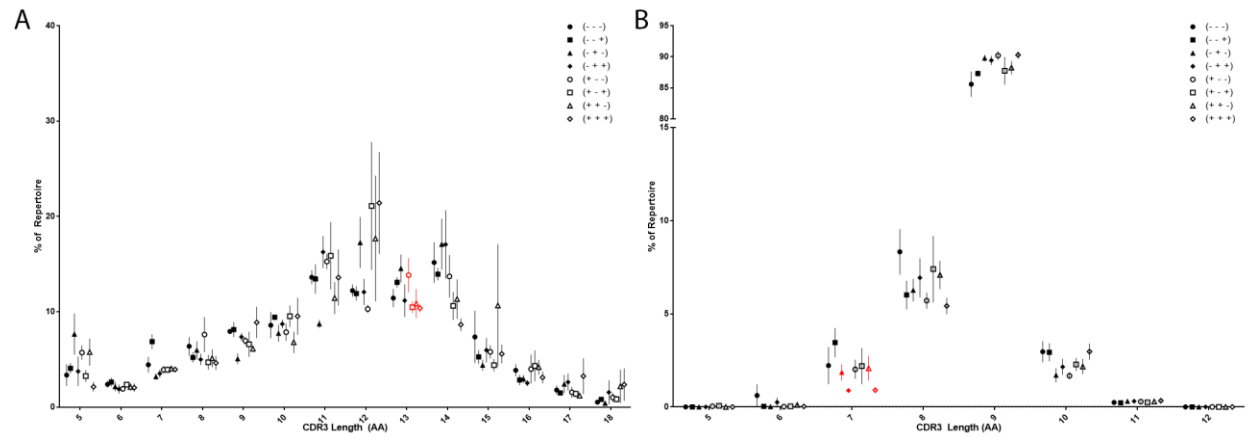


**Figure 6.5. Circos Plots of V/J Combinations by Treatment Group of Immunoglobulin Heavy (A) and Kappa (B) Chain**



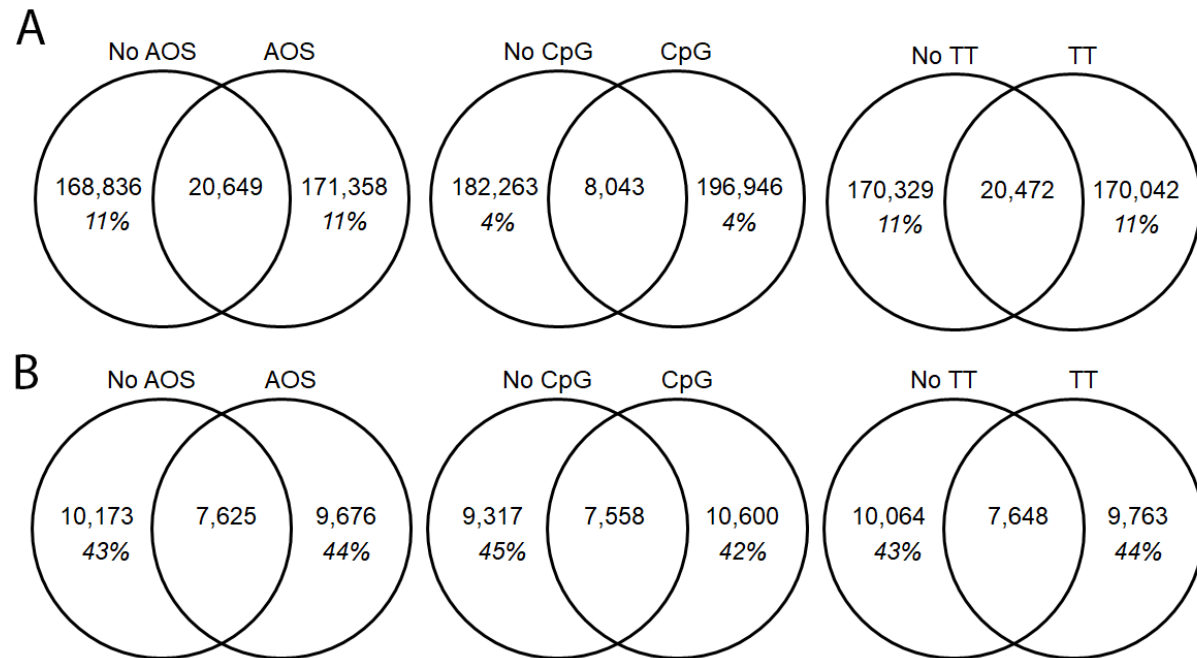
Each band represents a V/J pairing with more common V/J pairings being colored and having wider bands. Circos plot labels (starting at 12:00 position proceeding clockwise with occasional color references for reader orientation) (A) J1 (red), J2, J3, J4, U (yellow), V1, V2, V3, V4, V5, V6, V7 (Teal), V8, V9, V10, V11, V12, V13, V14 (purple), V15 (B) J1 (red), J2, J4, J5, U, V1 (yellow), V2, V3, V4, V5, V6, V7 (black sliver), V8, V9, V10, V11, V12, V13, V14, V15 (royal blue), V16, V17, V18, V19, V20 (black sliver, if present).

**Figure 6.6. Determination of CDR3 Amino Acid Length for Immunoglobulin Heavy (A) and Kappa (B) Chain.**



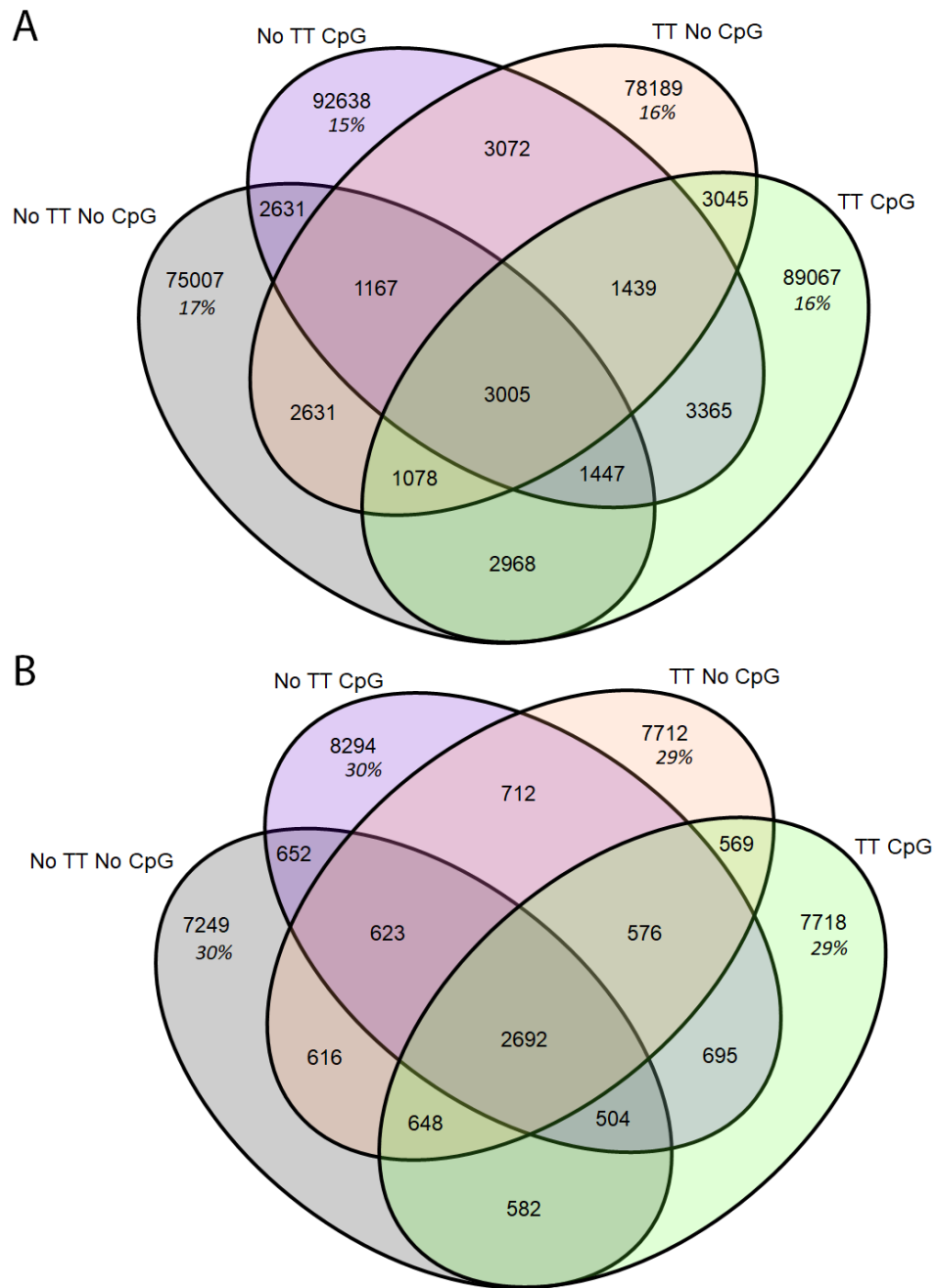
Selected CDR3 AA Length (based on highest usage) are displayed by mean  $\pm$  SEM (n=4 mice per treatment group). Red color indicates statistically significant ( $P < 0.05$ ) differences between similarly treated AOS vs. control animals.

**Figure 6.7. Overlap of Unique CDR3 Amino Acid Sequences by Variable for Heavy (A) and Kappa (B) Chain.**



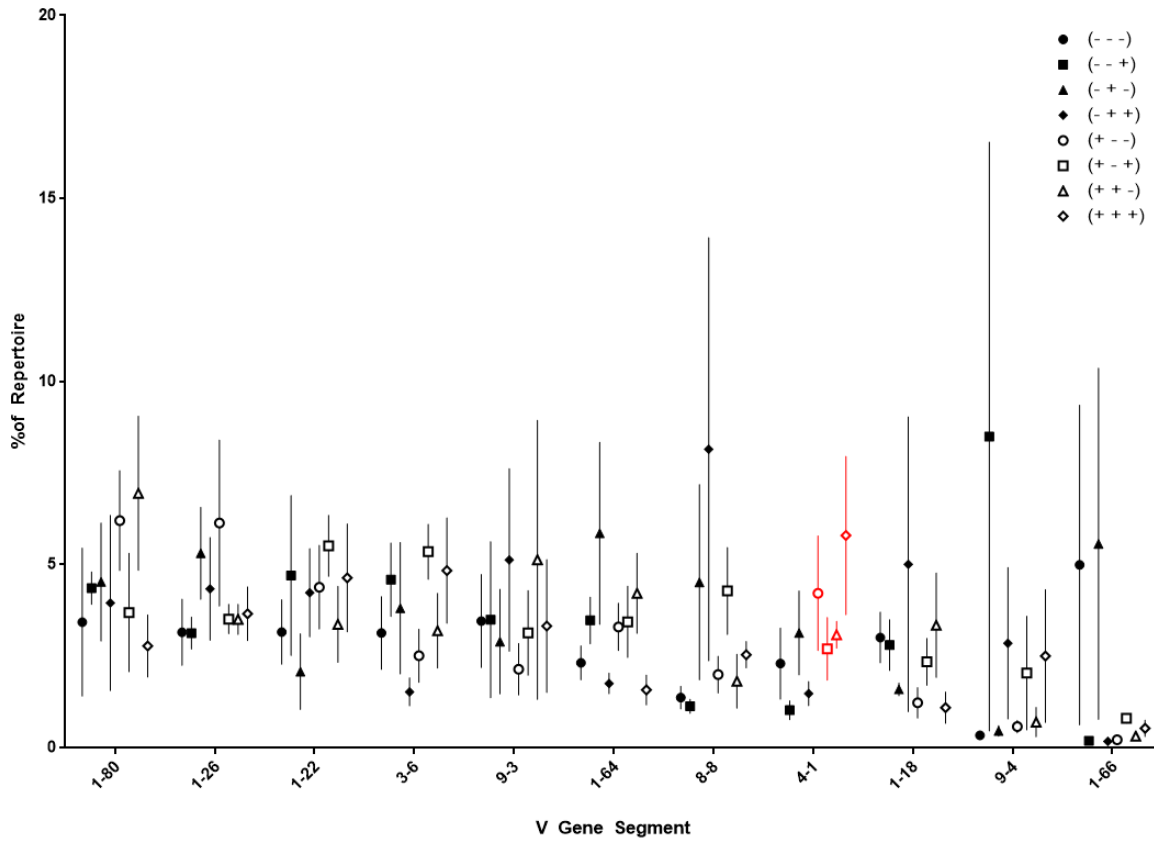
The overlap of unique CDR3 sequences is displayed in Venn Diagrams. The total number of unique CDR3s is displayed by variable group with the percent of the shared repertoire shown in italics below the number of unique amino acid sequences.

**Figure 6.8. Overlap of Unique CDR3 Amino Acid Sequences for TT and CpG Treatment Groups for Heavy (A) and Kappa (B) Chain.**



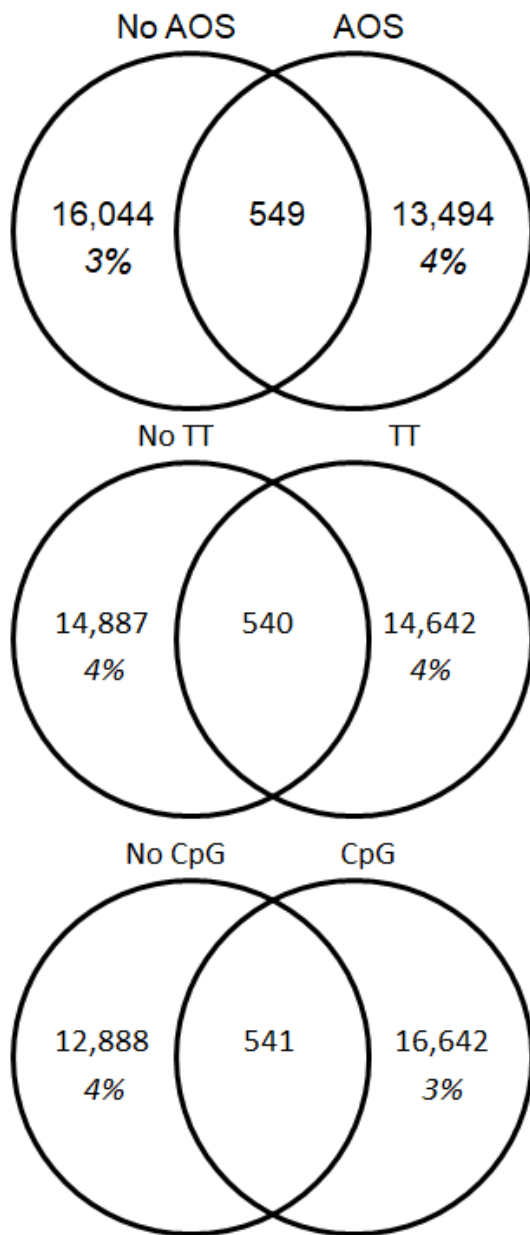
The overlap of unique CDR3 sequences is displayed in Venn Diagrams. The total number of unique CDR3 sequences in each treatment group, regardless of AOS status, with the percent of the shared repertoire shown in italics below the number of unique amino acid sequences.

**Figure 6.9. Assessment of High Frequency VH-Gene Segment Usage in Class-Switched Antibodies.**



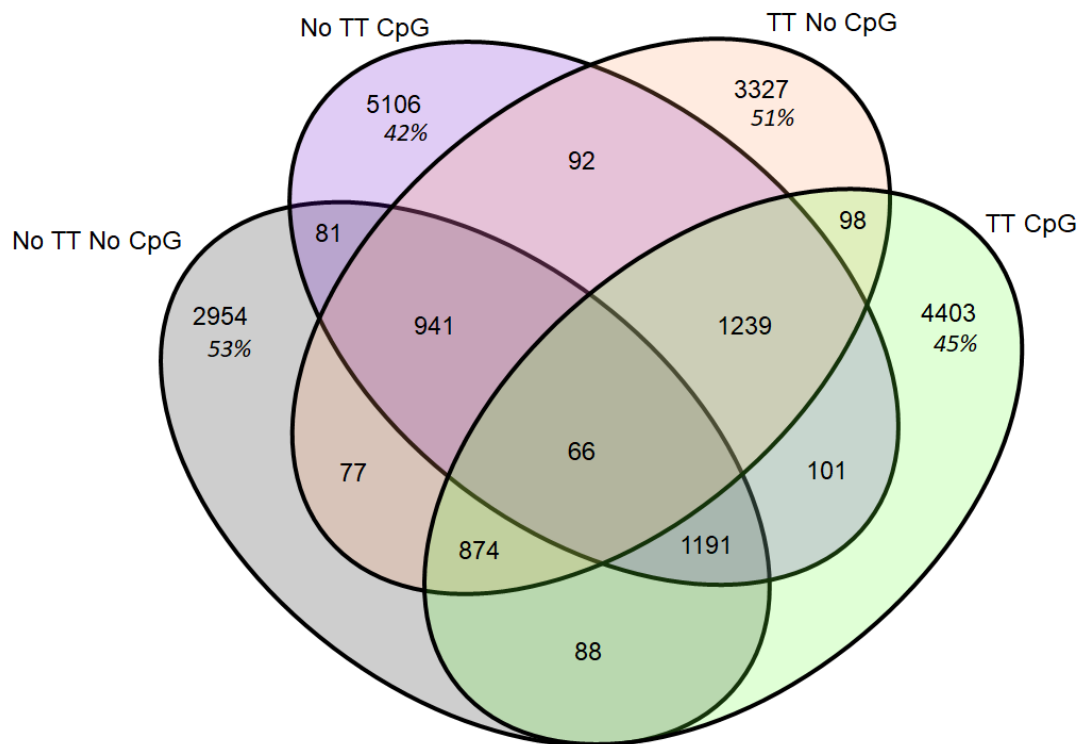
VH-gene segments found in over 5% (average) of the class switched (IgG/IgA) repertoire are displayed by mean  $\pm$  SEM (n=4 mice per treatment group). Red coloring indicates statically significant ( $P<0.05$ ) differences between similarly treated AOS vs. control animals.

**Figure 6.10. Overlap of Unique CDR3 Amino Acid Sequences Overlap by Variable for Class-Switched Antibodies.**



The overlap of unique CDR3 sequences is displayed in Venn Diagrams. The total number of unique CDR3 amino acid sequences is displayed for each variable group with the percent of the shared repertoire shown in italics below the number of unique amino acid sequences.

**Figure 6.11. Unique CDR3 Amino Acid Sequence Overlap for TT and CpG Treatment Groups for Class-Switched Antibodies.**



The overlap of unique CDR3 amino acid sequences is displayed in Venn Diagrams. The total number of unique CDR3 amino acid sequences in each treatment group, regardless of AOS status, with the percent of the shared repertoire shown in italics below the number of unique amino acid sequences.

**Table 6.1. Average number ( $\pm$  SEM) of reads obtained per treatment group**

	No AOS				AOS			
	No TT		TT		No TT		TT	
	No CpG	CpG	No CpG	CpG	No CpG	CpG	No CpG	CpG
Avg Initial Reads (M) <sup>a</sup>	44.5 $\pm 1.6$	42.6 $\pm 0.7$	42.2 $\pm 3.1$	38.4 $\pm 0.6$	42.9 $\pm 3.4$	39.3 $\pm 1.1$	39.1 $\pm 1.1$	38.6 $\pm 1.0$
Avg Cleaned Reads (M) <sup>b</sup>	26.2 $\pm 3.8$	26.1 $\pm 3.3$	24.8 $\pm 2.8$	26.5 $\pm 2.9$	24.3 $\pm 1.9$	28.2 $\pm 2.3$	27.3 $\pm 3.0$	27.9 $\pm 4.0$
Avg IgH Productive <sup>c</sup>	38,111 $\pm 2,816$	42,863 $\pm 9,371$	47,824 $\pm 10,912$	57,220 $\pm 15,734$	39,759 $\pm 9,865$	56,958 $\pm 12,905$	53,450 $\pm 16,204$	49,469 $\pm 15,726$
Avg IgH Unknown <sup>c</sup>	44,687 $\pm 1,694$	45,998 $\pm 6,591$	54,503 $\pm 10,072$	48,999 $\pm 12,974$	42,896 $\pm 5,799$	52,150 $\pm 9,586$	47,944 $\pm 9,847$	38,941 $\pm 10,437$
Avg IgH Total Reads <sup>d</sup>	82,798 $\pm 3,316$	88,861 $\pm 15,822$	10,2327 $\pm 20,148$	10,6219 $\pm 28,565$	82,655 $\pm 31,080$	10,9107 $\pm 22,721$	10,1394 $\pm 26,008$	88,410 $\pm 26,047$
Avg IgK Productive <sup>c</sup>	37,695 $\pm 3,178$	37,563 $\pm 6,783$	51,674 $\pm 13,051$	53,621 $\pm 15,019$	38,748 $\pm 7,956$	49,146 $\pm 8,989$	37,894 $\pm 9,357$	44,435 $\pm 10,659$
Avg IgK Unknown <sup>c</sup>	27,630 $\pm 1,887$	27,918 $\pm 3,260$	33,042 $\pm 7,584$	29,314 $\pm 7,441$	27,617 $\pm 1,692$	30,816 $\pm 3,428$	24,148 $\pm 5,528$	26,910 $\pm 7,380$
Avg IgK Total Reads <sup>d</sup>	65,325 $\pm 4,746$	65,481 $\pm 9,896$	84,716 $\pm 20,303$	82,934 $\pm 22,269$	66,365 $\pm 9,533$	79,962 $\pm 12,147$	62,042 $\pm 14,805$	71,345 $\pm 17,720$

(M) = Million

a – Number of sequences obtained from Illumina MiSeq sequencing

b – Number of sequences after quality control steps

c – Number of sequences identified by ImMunoGenetTics as productive of unknown sequences

d – Total of productive and unknown sequences



**Table 6.2. Impact of AOS, Tetanus Toxoid and CpG on lymphocyte phenotype by fold-change.**

	No AOS				AOS			
	No TT		TT		No TT		TT	
	No CpG	CpG	No CpG	CpG	No CpG	CpG	No CpG	CpG
B220	1.00	0.93	0.95	1.01	0.93	1.07	1.15	1.09
CD62-L	1.00	0.95	0.91	0.98	0.82	0.99	1.04	1.11
Cd44	1.00	0.92	0.81	0.94	0.78	1.00	1.04	0.92
Cd27	1.00	0.79	1.00	0.88	0.85	0.91	1.27	0.96
Cd19	1.00	0.99	1.02	1.06	0.99	1.08	1.07	1.16
CD138	1.00	0.89	1.23	0.98	0.83	1.00	0.77	0.82
Cd80	1.00	0.97	0.89	0.95	0.86	1.07	1.05	0.85

All P>0.05

**Table 6.3.** Assessment of Coefficient of Determination ( $R^2$ ) of V-Gene Segments.

	VH <sup>a</sup>	V $\kappa$ <sup>b</sup>
<b>AOS</b>		
(---) vs (+--)	0.8858	0.8011
(--+ ) vs (+-+)	0.7510	0.6067
(-+-) vs (++-)	0.7026	0.6504
(-++) vs (+++)	0.5745	0.4606
<b>TT</b>		
(---) vs (-+-)	0.7752	0.6461
(--+ ) vs (-++)	0.7086	0.5572
(+-+) vs (+++)	0.5793	0.4255
(+-- ) vs (++-)	0.778	0.7501
<b>CpG</b>		
(---) vs (--+)	0.8338	0.8164
(-+-) vs (-++)	0.6594	0.1849
(+-- ) vs (+-+)	0.7557	0.4507
(++-) vs (+++)	0.5563	0.3824
<b>AOS on TT+CpG</b>		
(---) vs (-++)	0.7027	0.4415
(---) vs (+++)	0.6500	0.4297

a -  $R^2$  values were determined by variable to measure variations between treatment groups for immunoglobulin heavy variable gene usage

b -  $R^2$  values were determined by variable to measure variations between treatment groups for immunoglobulin kappa variable gene usage

Treatment groups are represented by absence (-) or presence (+) of a treatment by (AOS, TT, CpG). An untreated animal is represented by (---) while an animal receiving only TT and CpG is represented as (-++)

**Table 6.4.** R<sup>2</sup> of V/J Pairing Correlation by Variable

	VH-JH <sup>a</sup>	V <sub>κ</sub> -J <sub>κ</sub> <sup>b</sup>
<b>AOS</b>		
(---) vs (+--)	0.6140	0.5905
(--+ ) vs (+-+)	0.3881	0.5896
(-+-) vs (++-)	0.4247	0.4669
(-++) vs (+++)	0.1510	0.4203
<b>TT</b>		
(---) vs (-+-)	0.4649	0.4879
(--+ ) vs (-++)	0.3857	0.5361
(+-+) vs (+++)	0.1412	0.4278
(+-- ) vs (++-)	0.3534	0.4969
<b>CpG</b>		
(---) vs (--+)	0.5533	0.6440
(-+-) vs (-++)	0.3685	0.1439
(+-- ) vs (+-+)	0.3791	0.3480
(++-) vs (+++)	0.1414	0.3133
<b>AOS on TT+CpG</b>		
(---) vs (-++)	0.3536	0.3893
(---) vs (+++)	0.3025	0.4242

a - R<sup>2</sup> values were determined by variable to measure variations between treatment groups for immunoglobulin heavy variable gene usage

b - R<sup>2</sup> values were determined by variable to measure variations between treatment groups for immunoglobulin kappa variable gene usage

Treatment groups are represented by absence (-) or presence (+) of a treatment by (AOS, TT, CpG). An untreated animal is represented by (---) while an animal receiving only TT and CpG is represented as (-++)

**Table 6.5. Shared Unique CDR3 Amino Acid Sequences**

Shared <sup>a</sup>	Count <sup>b</sup>	Percent <sup>c</sup>
<b>Heavy</b>		
1	331642	91.91
2	18568	5.15
3	5194	1.44
4	2268	0.63
5	1318	0.37
6	777	0.22
7	516	0.14
8	560	0.16
<b>Kappa</b>		
1	18095	65.86
2	3494	12.72
3	1758	6.40
4	1211	4.41
5	814	2.96
6	630	2.29
7	460	1.67
8	1012	3.68

a – Number of treatment groups containing a unique CDR3 sequence

b – Number of unique CDR3 sequences in each shared pool

c – Percent of unique samples shared in each shared pool

## Chapter 7 - Conclusions

The spaceflight model serves as a model of both a high-stress environment and as a model for non-weight bearing humans, such as those on bedrest or confined to wheelchairs. High-stress environments, regardless of weight bearing have been shown to depress the immune system. Additionally, as humans age and become less mobile, effects from aging and confinement will impact the immune system. Understanding changes to the immune system will help physicians be better able to tackle treatments for patients.

Our works have shown that we have developed a successful workflow for the analysis of the murine antibody repertoire from animals who have flown aboard the ISS and housed on the ground. This workflow (Chapter 2) was created and validated because most repertoire analyses have used some method of cell sorting to isolate B cells and then used amplification methods to select for specific Ig loci for the repertoire analysis. For the experiments on the ISS, single cell suspension preparation is not currently possible. Therefore, validation and ground-based studies were based on whole tissue sequencing. In addition, we did not use amplification for our data sets. The original decision to avoid amplification was to avoid sequencing biases, the difficulty in designing and implementing primer sets that would represent the entire Ig repertoire (see the following paragraph) and generate a data set that would allow for additional data mining, which is available for public use. We did explore some available methods to generate focused, amplified datasets and found drastic differences in results which were surprising and unacceptable of inclusion in our works (Chapter 3).

I explored the possibility of using primer amplification of immunoglobulin genes. Originally, I attempted to make my own set of degenerate primers. Using a phylogenetic tree, I developed “groups” of similar V-gene segments and attempted to design degenerate primers that would amplify these V-gene segments. We discovered that a high number of primers would be required (11 groups were created via analysis, each with at least one degenerate primer set). I successfully amplify IgM transcripts from whole spleen tissue with one set of degenerate primers but the amplification conditions among all the different primer sets were drastically different. Therefore, amplification efficiencies, template fidelity and reproducibility dissuaded us from using this method. During this process, we explored commercial sequencing options with two companies who used V- and J-gene segment or V-gene segment and constant region primers to amplify antibody gene sequences. Although the commercial methods used the Illumina platforms, as we do, there were different results which showed that V-gene segment usage was very different between identical KSU unamplified samples the amplified samples. After analysis of the amplified raw data through our bioinformatic pipeline, these extreme differences were still present so we hypothesize that these differences were caused from amplification bias. The failure of the amplified datasets to include a number of V-gene segments detected in at high frequency the KSU data was a concern. Therefore, we chose not to use the amplification approach. To our knowledge, we are the only group exploring the B cell immunoglobulin gene repertoire using an unamplified data set.

The main method of combating amplification bias is to use 5' RACE, a technique which uses a primer specific to the 3' end of the transcript that adds a poly A sequence to the 5' end after sequence duplication. The poly-A sequence is included in subsequent rounds of

amplification to amplify 5' variable transcripts. However, even when using the in 5' RACE technique, studies to date have focused only on a single isotype (usually IgG). For a complete picture of the Ig repertoire, multiple primers would be needed. This is an expensive technique and still has limitations of 3' primer bias and optimization requirements. Other methodologies include the addition of molecular barcodes to mRNA sequences to amplify 5' variable transcripts. It also allows for cluster generation so sequencing errors can be removed from the antibody sequences. This method prevents many of the issues associated with straight amplification, but there are still amplification pitfalls. Chapter 3 of this thesis was presented to address a problem which is often ignored in B cell receptor amplification studies.

While amplification failed to provide high fidelity V-gene reproducibility, it did provide an increased depth of sequencing for CDR3 detection; detecting over 100x more CDR3s than our original analysis. While the goal of our project was to analyze the entire repertoire, which more heavily weights the importance of V-gene usage, there are situations where depth of sequencing becomes the more important factor such as in cancer detection or specific goals to identify rare clones.

With a validated workflow, we assessed the repertoire of three different groups of animals; the conventionally housed, unchallenged “normal” C57BL/6J mouse (Chapter 4), “normal” C57BL/6Tac mice flown aboard the ISS (Chapter 5), and C57BL/6J mice subjected to AOS and challenged with TT and CpG (Chapter 6). For each of these mouse populations, we examined multiple aspects of the repertoire including V-, D-, an J-gene segment usage, constant region usage, CDR3 length, and CDR3 overlap.

There was high similarity among pools of mice. Our results are similar to those presented by Greiff et al.<sup>44</sup> who showed some level of genetic determination in the normal repertoire, but detectable changes in the repertoire due to antigen exposure. This predetermination could be due to a variety of factors including histone deposition, epigenetic effects, or even mutations in the RSS sequences. All have been shown to play a role in control of VDJ recombination.<sup>43,44,64</sup> However, while the gene segment usage among the different mouse populations correlated well, there was lower correlation among V/J pairing and CDR3 usage. These data suggest that while genetic usage may be predetermined, CDR3 generation is still unique to the individual as splice formation, and N- and P- nucleotide additions are still randomly generated. This has also been shown in human twin studies.<sup>289</sup>

There were no changes in the repertoire between mice flown aboard the ISS for a short duration (~21 days) and ground animals. This may be due to the turnover time for B cells (about five to six weeks), so that approximately half the repertoire hadn't changed from during the experimental period. The flight animals are likely undergoing V/D/J recombination with strong genetic influences, so it may take even longer than five to six weeks to see noticeable differences in the repertoire. Interestingly, the V-gene segment usage in these animals differed greatly from that found in normal, 9- to 11-weekold C57BL/6J mice. These differences could be due to age (9-11 weeks vs 32 weeks) or due to sub-strain (Tac vs. J) differences. These mouse substrains have been separated for several decades with concomitant genetic drift. Additional experiments will be necessary to tease out whether age or strain differences are responsible. Greiff noted that



differences between C57BL/6 mice and Balb/c mice were minimal in V gene usage.<sup>44</sup> Therefore, this might suggest age does influence the repertoire.<sup>12,79</sup>

In the AOS study we examined the effects of skeletal unloading and the concomitant physiological changes in the mice on the challenged murine antibody repertoire. Eight treatment groups included a 2x2x2 experimental design with AOS, tetanus toxoid and CpG as the three possible treatments. Our results show small changes to gene segment usage and pairing, and constant region usage. There are multiple reasons why we may not see a dramatic shift in the repertoire, even though a robust immune response was measured by TT-specific IgG. We only challenged animals once, measuring the primary response. As this study was designed to be part of a larger study that later examined the secondary response, this decision was a logical option. We also collected samples two weeks after antigen challenge. We may be seeing a decrease in the transcript levels of anti-TT specific antibodies as the antibody proteins have already been made and cleared from the animal. Also important is the antigen chosen. While TT is a human-relevant vaccine, it also is a large molecule, and multiple known epitopes. Some have hypothesized that the TT-specific repertoire is comprised of between 50-100 idiotypes and that these repertoires may vary greatly from individual to individual and from experiment to experiment.<sup>62,96</sup> This may preclude the formation of a dominant idiomere or CDR3. To support this hypothesis, when we attempted to enrich for TT-specific B cells using magnetic beads we still did not select for a single or group of dominant idiotypes among three independent biological samples; even when examining only class switched antibodies.

The data generated from the sequencing of over 40 independent biological samples is considerable. Multiple questions remain to be explored. We have not addressed heavy/light chain pairing. We can make some assumptions on heavy/light chain pairing based on overall frequency of sequence usage. However, previous work has shown that the same heavy chain can pair with multiple light chains.<sup>91</sup> Developing an algorithm to predict possible heavy/light chains pairs combined with 3D modeling to confirm binding ability and frequency may provide some insight into heavy/light chain pairs. There has been some work on heavy/light chains pairings in TT responses<sup>98</sup> by using single-cell sequencing techniques, but little bioinformatic work on how to pair heavy and light chains that were not paired during amplification or sequencing.

Another question which could be explored is what are the mutation rates within V-gene segments and specifically the impact of AOS and antigen challenge on that process? Somatic hypermutation is known to occur at specific hot spots, specifically in CDR regions.<sup>54</sup> Does this change with AOS and antigenic challenge? Since CpG causes random activation of B cells, mutational changes may occur differently than we may see in an antigen-specific response. Additionally, there may be effects from AOS. While our lab is already examining mutational rates, we are looking at overall and location effects, rather than examining effects by each V-gene segment specifically. Mutational rates could increase in one V-gene segment, but not in another.

Another interesting question to explore is what are the common CDR3s among all our individual biological samples? We have multiple CDR3s that occur in these datasets. Are these particular idiotypes vital to the immune system? How important is it to maintain these antibodies

in the repertoire? CDR3s such as CARGAYW, CARGSYW, CARGGYW, which share highly similar sequences, are detected in most of our datasets, and some, such as CMRYSNYWYFDVW are found even across strains. Other commonly detected CDR3s include CARESNYNYAMDYW and CARYAPYYFDYW which occur in high frequencies in multiple data sets. It is likely since they occur in multiple samples, from animals housed in multiple locations and fed different foods, that they are important. One hypothesis is that they are reacting to some form of self-antigen, a ubiquitous antigen, or a common member of the microbiome. Further examination of these CDR3s could include V-, D-, and J-gene segment usage to determine if multiple V/D/J combinations create the same CDR3s. Additional analysis on these CDR3s, and other CDR3s we have detected, such as the suspected anti-TT CDR3s, could include isotype analysis. We have briefly examined isotype/CDR3 pairing in one set of hyper-vaccinated TT+CpG mice but we have not further explored this. It may prove particularly useful interesting to detect likely anti-TT idiotypes that occur in both the IgM and class-switched populations and do some comparisons.

The decision not to amplify our cDNA samples leaves a complete RNASeq library at our disposal. These data have been or will be deposited in the NASA Genelab and will be available to others. Other potential questions to explore include looking at the expression of stress-associated or cytokine-related transcripts. Previous work has shown changes in cytokine levels in response to spaceflight and/or AOS.<sup>156,171</sup> We have only briefly explored cytokine levels in some of our samples and have yet to explore them in our AOS datasets. These measurements could occur both in blood and tissue level analyses. Examination of the T cell receptor (TCR)

repertoire is also a possibility. Changes to T-cell populations have been noted in spaceflight and AOS studies, so it is possible that we may see shifts in the TCR repertoire.<sup>175</sup>

One of the important lessons learned from this work, is that TT produces a wide ranging, and diverse antibody response. While TT is an important human vaccine, and its use is justified, the large surface area of the molecule with multiple epitopes is difficult to characterize. A single epitope antigen, such as 4-hydroxy-3-nitro-phenylacetyl conjugated to hen egg lysozyme (NP-HEL) as has been used by Greiff et al<sup>44</sup> might generate dominant idiotypes that can be identified in multiple animals. At least the Greiff group was able to identify related if not identical clones among several animals. Other antigens such as ovalbumin and hepatitis B surface antigen are also candidates. These antigens are all significantly smaller (under 300AAs, compared to the almost 1,000AA of the TT) and would likely cause a more narrowed antibody response.

Tied to the difficulties of TT-specific responses, additional experiments could focus on the isolation of TT-specific CDR3s from hyper-vaccinated animals. Additional technologies will be necessary on the ISS before these experiments become practical. However, single cell isolation and sequencing technologies are being developed and may not be too far in the future of space flight.

While we were able to detect subtle shifts, our works do not approach the biological significance of these shifts. More research, specifically those involving challenge of either unloaded animal or animals in flight, would provide more conclusive results. Due the large size of the TT molecule, multiple binding epitopes that do not provide a neutralization effect are

likely present. If the subtle changes detected to impact binding to one of the neutralizing epitopes, those changes would have high levels of biological significance. The high variability in anti-TT responses would likely limit this to single animal responses, rather than a population level effect. However, decreased responses to one antigen could easily decrease responses to other antigens, which has further reaching implications.

While our studies include animals housed at a variety of locations (Manhattan, KS, Loma Linda, CA, Kennedy Space Center, FL), multiple strains, and fed different foods, we do not have adequate controls to account for any variation in response due to these differences. While we see little differences across high frequency V-gene segments (Figure 7.1) we do find some significant variations in usage. Biological relevance and cause of these shifts are currently unknown and would require additional research.

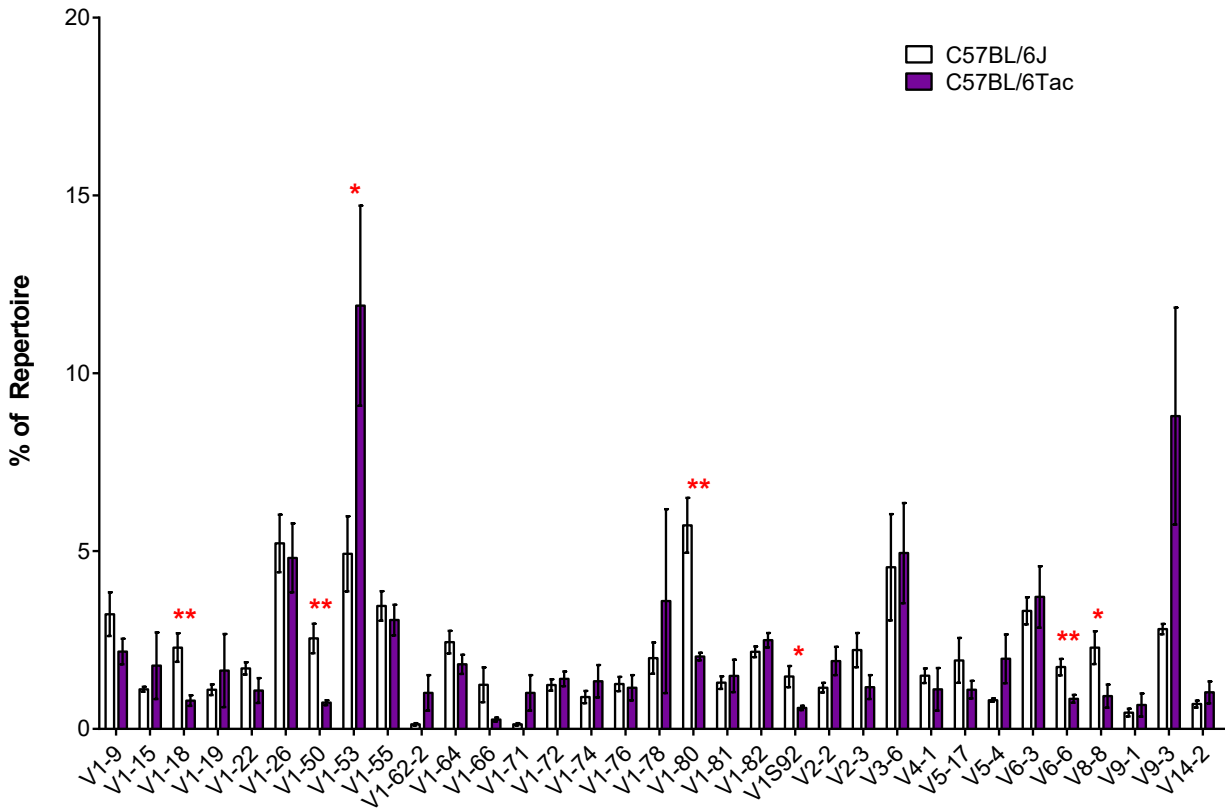
Future work also includes the assessment of the bone marrow B cell repertoire from the AOS animals described in my work. In addition, an experiment to parallel the AOS study is tentatively scheduled to be done aboard the ISS in 2019. The secondary response is also planned during that mission. We hypothesize that the second challenge will further narrow the B-cell repertoire and show a stronger repertoire shift.

In conclusion, this work has laid the foundation for future studies of repertoire assessment on Earth and in space. We take a unique approach in sequencing unamplified whole tissues. We have also exposed a potential issue with repertoire studies using primers for amplification. Our ground-based models have provided a starting point for direct comparisons

between AOS and flight data, a rare opportunity. The development of high through-put sequencing has revolutionized our ability to make repertoire assessments and continued refinement of technology will improve our understanding of the biology.

## Figures

**Figure 7.1. High frequency V-Gene segment usage between young C57BL/6 and aged C57BL/6Tac mice**



V-gene segments found in the top ten most common of C57BL/6J mice (three normal mouse pools (Chapter 4) plus four untreated animal from the AOS study (Chapter6)) and C57BL/6Tac mice (Chapter 5) represented by mean with standard error. \*P<0.05, \*\*P<0.01, \*\*\*P<0.001.

## References

- 1 Cooper, M. D., Peterson, R. D. & Good, R. A. Delineation of the thymic and brusal lymphoid systems in the chicken. *Nature* **205**, 143-146 (1965).
- 2 Strebhardt, K. & Ullrich, A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat Rev Cancer* **8**, 473-480, doi:10.1038/nrc2394 (2008).
- 3 Owen, J. A., Punt, J., Stranford, S. A. & Jones, P. P. *Kuby Immunology*. 7th edn, (2013).
- 4 Potter, M. Structural correlates of immunoglobulin diversity. *Surv Immunol Res* **2**, 27-42 (1983).
- 5 Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 3628-3632 (1976).
- 6 Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E. & Weissman, I. L. Identification of a lineage of multipotent hematopoietic progenitors. *Development (Cambridge, England)* **124**, 1929-1939 (1997).
- 7 Lai, A. Y. & Kondo, M. Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *J Exp Med* **203**, 1867-1873, doi:10.1084/jem.20060697 (2006).
- 8 Cory, S. Masterminding B Cells. *Journal of immunology (Baltimore, Md. : 1950)* **195**, 763-765, doi:10.4049/jimmunol.1501277 (2015).
- 9 Meffre, E., Casellas, R. & Nussenzweig, M. C. Antibody regulation of B cell development. *Nat Immunol* **1**, 379-385, doi:10.1038/80816 (2000).
- 10 Graf, T. Immunology: blood lines redrawn. *Nature* **452**, 702-703, doi:10.1038/452702a (2008).



- 11 Rodriguez-Fraticelli, A. E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212, doi:10.1038/nature25168  
<https://www.nature.com/articles/nature25168#supplementary-information> (2018).
- 12 Labrie, J. E., 3rd, Borghesi, L. & Gerstein, R. M. Bone marrow microenvironmental changes in aged mice compromise V(D)J recombinase activity and B cell generation. *Semin Immunol* **17**, 347-355, doi:10.1016/j.smim.2005.05.012 (2005).
- 13 Schatz, D. G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat Rev Immunol* **11**, 251-263, doi:10.1038/nri2941 (2011).
- 14 Collins, A. M., Wang, Y., Roskin, K. M., Marquis, C. P. & Jackson, K. J. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**, doi:10.1098/rstb.2014.0236 (2015).
- 15 Melchers, F. *et al.* Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells. *Immunol Rev* **175**, 33-46 (2000).
- 16 Sun, X. *et al.* Antibody repertoire development in fetal and neonatal piglets. XXII. lambda Rearrangement precedes kappa rearrangement during B-cell lymphogenesis in swine. *Immunology* **137**, 149-159, doi:10.1111/j.1365-2567.2012.03615.x (2012).
- 17 Gebert, C. *et al.* Chromosome choice for initiation of V-(D)-J recombination is not governed by genomic imprinting. *Immunology and cell biology* **95**, 473-477, doi:10.1038/icb.2017.1 (2017).
- 18 Ebert, A., Hill, L. & Busslinger, M. Spatial Regulation of V-(D)J Recombination at Antigen Receptor Loci. *Advances in immunology* **128**, 93-121, doi:10.1016/bs.ai.2015.07.006 (2015).
- 19 DeKosky, B. J. *et al.* Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proceedings of the National Academy of Sciences* **113**, E2636-E2645, doi:10.1073/pnas.1525510113 (2016).

- 20 de Wildt, R. M., van Venrooij, W. J., Winter, G., Hoet, R. M. & Tomlinson, I. M. Somatic insertions and deletions shape the human antibody repertoire. *Journal of molecular biology* **294**, 701-710, doi:10.1006/jmbi.1999.3289 (1999).
- 21 Lo, C. G., Lu, T. T. & Cyster, J. G. Integrin-dependence of lymphocyte entry into the splenic white pulp. *J Exp Med* **197**, 353-361 (2003).
- 22 Stein, J. V. & Nombela-Arrieta, C. Chemokine control of lymphocyte trafficking: a general overview. *Immunology* **116**, 1-12, doi:10.1111/j.1365-2567.2005.02183.x (2005).
- 23 Cyster, J. G. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annual review of immunology* **23**, 127-159, doi:10.1146/annurev.immunol.23.021704.115628 (2005).
- 24 Mori, M. *et al.* IL-4 promotes the migration of circulating B cells to the spleen and increases splenic B cell survival. *Journal of immunology (Baltimore, Md. : 1950)* **164**, 5704-5712 (2000).
- 25 Allman, D. M., Ferguson, S. E., Lentz, V. M. & Cancro, M. P. Peripheral B cell maturation. II. Heat-stable antigen(hi) splenic B cells are an immature developmental intermediate in the production of long-lived marrow-derived B cells. *Journal of immunology (Baltimore, Md. : 1950)* **151**, 4431-4444 (1993).
- 26 Allman, D. *et al.* Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *Journal of immunology (Baltimore, Md. : 1950)* **167**, 6834-6840 (2001).
- 27 Baumgarth, N. A Hard(y) Look at B-1 Cell Development and Function. *The Journal of Immunology* **199**, 3387-3394, doi:10.4049/jimmunol.1700943 (2017).
- 28 Owen, J. J., Cooper, M. D. & Raff, M. C. In vitro generation of B lymphocytes in mouse foetal liver, a mammalian 'bursa equivalent'. *Nature* **249**, 361-363 (1974).

- 29 Hansen, A. *et al.* VH/VL gene expression in polyreactive-antibody-producing human hybridomas from the fetal B cell repertoire. *Experimental and clinical immunogenetics* **11**, 1-16 (1994).
- 30 Hayakawa, K., Hardy, R. R., Herzenberg, L. A. & Herzenberg, L. A. Progenitors for Ly-1 B cells are distinct from progenitors for other B cells. *J Exp Med* **161**, 1554-1568 (1985).
- 31 Gearhart, P. J., Sigal, N. H. & Klinman, N. R. Production of antibodies of identical idiotype but diverse immunoglobulin classes by cells derived from a single stimulated B cell. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 1707-1711 (1975).
- 32 Davis, M. M. *et al.* An immunoglobulin heavy-chain gene is formed by at least two recombinational events. *Nature* **283**, 733-739 (1980).
- 33 Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature* **286**, 676-683 (1980).
- 34 Stavnezer, J. & Schrader, C. E. IgH chain class switch recombination: mechanism and regulation. *Journal of immunology (Baltimore, Md. : 1950)* **193**, 5370-5378, doi:10.4049/jimmunol.1401849 (2014).
- 35 Grusby, M. J., Johnson, R. S., Papaioannou, V. E. & Glimcher, L. H. Depletion of CD4<sup>+</sup> T cells in major histocompatibility complex class II-deficient mice. *Science* **253**, 1417-1420 (1991).
- 36 Lorenz, M., Jung, S. & Radbruch, A. How cytokines control immunoglobulin class switching. *Behring Institute Mitteilungen*, 97-102 (1995).
- 37 Janda, A., Bowen, A., Greenspan, N. S. & Casadevall, A. Ig Constant Region Effects on Variable Region Structure and Function. *Frontiers in Microbiology* **7**, 22, doi:10.3389/fmicb.2016.00022 (2016).

- 38 Ippolito, G. C. *et al.* Antibody repertoires in humanized NOD-scid-IL2Rgamma(null) mice and human B cells reveals human-like diversification and tolerance checkpoints in the mouse. *PLoS One* **7**, e35497, doi:10.1371/journal.pone.0035497 (2012).
- 39 Meng, W. *et al.* An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol*, doi:10.1038/nbt.3942 (2017).
- 40 Galson, J. D. *et al.* BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines. *Immunology and cell biology* **93**, 885-895, doi:10.1038/icb.2015.57 (2015).
- 41 Feeney, A. J., Tang, A. & Ogwaro, K. M. B-cell repertoire formation: role of the recombination signal sequence in non-random V segment utilization. *Immunol Rev* **175**, 59-69, doi:10.1111/j.1600-065X.2000.imr017508.x (2000).
- 42 Kidd, M. J., Jackson, K. J. L., Boyd, S. D. & Collins, A. M. DJ pairing during VDJ recombination shows positional biases that vary between individuals with differing IGHD locus immunogenotypes. *Journal of immunology (Baltimore, Md. : 1950)* **196**, 1158-1164, doi:10.4049/jimmunol.1501401 (2016).
- 43 Bolland, D. J. *et al.* Two Mutually Exclusive Local Chromatin States Drive Efficient V(D)J Recombination. *Cell reports* **15**, 2475-2487, doi:10.1016/j.celrep.2016.05.020 (2016).
- 44 Greiff, V. *et al.* Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development. *Cell reports* **19**, 1467-1478, doi:10.1016/j.celrep.2017.04.054 (2017).
- 45 Rettig, T. A., Ward, C., Pecaut, M. J. & Chapes, S. K. Validation of Methods to Assess the Immunoglobulin Gene Repertoire in Tissues Obtained from Mice on the International Space Station. *Gravitational and space research : publication of the American Society for Gravitational and Space Research* **5**, 2-23 (2017).
- 46 Slifka, M. K., Antia, R., Whitmire, J. K. & Ahmed, R. Humoral Immunity Due to Long-Lived Plasma Cells. *Immunity* **8**, 363-372, doi:10.1016/S1074-7613(00)80541-5.

- 47 Jourdan, M. *et al.* An in vitro model of differentiation of memory B cells into plasmablasts and plasma cells including detailed phenotypic and molecular characterization. *Blood* **114**, 5173-5181, doi:10.1182/blood-2009-07-235960 (2009).
- 48 Manz, R. A., Thiel, A. & Radbruch, A. Lifetime of plasma cells in the bone marrow. *Nature* **388**, 133-134, doi:10.1038/40540 (1997).
- 49 Hammarlund, E. *et al.* Plasma cell survival in the absence of B cell memory. *Nature communications* **8**, 1781, doi:10.1038/s41467-017-01901-w (2017).
- 50 Jones, D. D., Wilmore, J. R. & Allman, D. Cellular Dynamics of Memory B Cell Populations: IgM<sup>+</sup> and IgG<sup>+</sup> Memory B Cells Persist Indefinitely as Quiescent Cells. *Journal of immunology (Baltimore, Md. : 1950)* **195**, 4753-4759, doi:10.4049/jimmunol.1501365 (2015).
- 51 Liu, M. *et al.* Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**, 841-845, doi:10.1038/nature06547 (2008).
- 52 Odegard, V. H. & Schatz, D. G. Targeting of somatic hypermutation. *Nat Rev Immunol* **6**, 573-583, doi:10.1038/nri1896 (2006).
- 53 Sakaguchi, N., Maeda, K. & Kuwahara, K. Molecular mechanism of immunoglobulin V-region diversification regulated by transcription and RNA metabolism in antigen-driven B cells. *Scand J Immunol* **73**, 520-526, doi:10.1111/j.1365-3083.2011.02557.x (2011).
- 54 Cui, A. *et al.* A Model of Somatic Hypermutation Targeting in Mice Based on High-Throughput Ig Sequencing Data. *Journal of immunology (Baltimore, Md. : 1950)* **197**, 3566-3574, doi:10.4049/jimmunol.1502263 (2016).
- 55 Sheng, Z. *et al.* Gene-Specific Substitution Profiles Describe the Types and Frequencies of Amino Acid Changes during Antibody Somatic Hypermutation. *Frontiers in Immunology* **8**, 537, doi:10.3389/fimmu.2017.00537 (2017).

- 56 de Wildt, R. M., Hoet, R. M., van Venrooij, W. J., Tomlinson, I. M. & Winter, G. Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire. *Journal of molecular biology* **285**, 895-901, doi:10.1006/jmbi.1998.2396 (1999).
- 57 Neuberger, M. S. *et al.* Antibody diversification and selection in the mature B-cell compartment. *Cold Spring Harb Symp Quant Biol* **64**, 211-216 (1999).
- 58 Ward, C. *et al.* Effects of spaceflight on the immunoglobulin repertoire of unimmunized C57BL/6 mice. *Life Sciences in Space Research* **16**, 63-75, doi:<https://doi.org/10.1016/j.lssr.2017.11.003> (2018).
- 59 Wilson, P. C. *et al.* Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J Exp Med* **187**, 59-70 (1998).
- 60 Briney, B. S., Willis, J. R. & Crowe, J. E. Location and length distribution of somatic hypermutation-associated DNA insertions and deletions reveals regions of antibody structural plasticity. *Genes and Immunity* **13**, 523-529 (2012).
- 61 Kitaura, K. *et al.* Different Somatic Hypermutation Levels among Antibody Subclasses Disclosed by a New Next-Generation Sequencing-Based Antibody Repertoire Analysis. *Frontiers in Immunology* **8**, 389, doi:10.3389/fimmu.2017.00389 (2017).
- 62 Poulsen, T. R., Jensen, A., Haurum, J. S. & Andersen, P. S. Limits for antibody affinity maturation and repertoire diversification in hypervaccinated humans. *Journal of immunology (Baltimore, Md. : 1950)* **187**, 4229-4235, doi:10.4049/jimmunol.1000928 (2011).
- 63 Stubenrauch, K. *et al.* Characterization of murine anti-human Fab antibodies for use in an immunoassay for generic quantification of human Fab fragments in non-human serum samples including cynomolgus monkey samples. *Journal of pharmaceutical and biomedical analysis* **72**, 208-215, doi:10.1016/j.jpba.2012.08.023 (2013).
- 64 Feeney, A. J. Factors that influence formation of B cell repertoire. *Immunol Res* **21**, 195-202, doi:10.1385/ir:21:2-3:195 (2000).

- 65 Greiff, V. *et al.* A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. *Genome medicine* **7**, 49, doi:10.1186/s13073-015-0169-8 (2015).
- 66 Carmack, C. E., Shinton, S. A., Hayakawa, K. & Hardy, R. R. Rearrangement and selection of VH11 in the Ly-1 B cell lineage. *J Exp Med* **172**, 371-374 (1990).
- 67 Yang, Y. *et al.* Distinct mechanisms define murine B cell lineage immunoglobulin heavy chain (IgH) repertoires. *eLife* **4**, e09083, doi:10.7554/eLife.09083 (2015).
- 68 Rother, M. B. *et al.* Decreased IL7R $\alpha$  and TdT expression underlie the skewed immunoglobulin repertoire of human B-cell precursors from fetal origin. *Scientific reports* **6**, 33924, doi:10.1038/srep33924 (2016).
- 69 Prechl, J. A generalized quantitative antibody homeostasis model: antigen saturation, natural antibodies and a quantitative antibody network. *Clinical & translational immunology* **6**, e131, doi:10.1038/cti.2016.90 (2017).
- 70 Weksler, M. E. Changes in the B-cell repertoire with age. *Vaccine* **18**, 1624-1628 (2000).
- 71 Frolich, D. *et al.* Secondary immunization generates clonally related antigen-specific plasma cells and memory B cells. *Journal of immunology (Baltimore, Md. : 1950)* **185**, 3103-3110, doi:10.4049/jimmunol.1000911 (2010).
- 72 Sundling, C. *et al.* Single-cell and deep sequencing of IgG-switched macaque B cells reveal a diverse Ig repertoire following immunization. *Journal of immunology (Baltimore, Md. : 1950)* **192**, 3637-3644, doi:10.4049/jimmunol.1303334 (2014).
- 73 Galson, J. D., Pollard, A. J., Truck, J. & Kelly, D. F. Studying the antibody repertoire after vaccination: practical applications. *Trends in immunology* **35**, 319-331, doi:10.1016/j.it.2014.04.005 (2014).
- 74 Strauli, N. B. & Hernandez, R. D. Statistical inference of a convergent antibody repertoire response to influenza vaccine. *Genome medicine* **8**, 60, doi:10.1186/s13073-016-0314-z (2016).

- 75 Adamson, P. J. *et al.* Proteomic analysis of influenza haemagglutinin-specific antibodies following vaccination reveals convergent immunoglobulin variable region signatures. *Vaccine*, doi:10.1016/j.vaccine.2017.08.053 (2017).
- 76 Jiang, N. *et al.* Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci Transl Med* **5**, 171ra119, doi:10.1126/scitranslmed.3004794 (2013).
- 77 Khurana, S. *et al.* Human antibody repertoire after VSV-Ebola vaccination identifies novel targets and virus-neutralizing IgM antibodies. *Nat Med* **22**, 1439-1447, doi:10.1038/nm.4201 (2016).
- 78 Ademokun, A. *et al.* Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging cell* **10**, 922-930 (2011).
- 79 de Bourcy, C. F. *et al.* Phylogenetic analysis of the human antibody repertoire reveals quantitative signatures of immune senescence and aging. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 1105-1110, doi:10.1073/pnas.1617959114 (2017).
- 80 Walsh, S. H. & Rosenquist, R. Immunoglobulin gene analysis of mature B-cell malignancies: reconsideration of cellular origin and potential antigen involvement in pathogenesis. *Medical oncology (Northwood, London, England)* **22**, 327-341, doi:10.1385/mo:22:4:327 (2005).
- 81 Kodo, H., Gale, R. P. & Saxon, A. Antibody synthesis by bone marrow cells in vitro following primary and booster tetanus toxoid immunization in humans. *J Clin Invest* **73**, 1377-1384, doi:10.1172/jci111341 (1984).
- 82 Volk, W. A., Bizzini, B., Snyder, R. M., Bernhard, E. & Wagner, R. R. Neutralization of tetanus toxin by distinct monoclonal antibodies binding to multiple epitopes on the toxin molecule. *Infect Immun* **45**, 604-609 (1984).
- 83 Dietz, V., Galazka, A., van Loon, F. & Cochi, S. Factors affecting the immunogenicity and potency of tetanus toxoid: implications for the elimination of neonatal and non-neonatal tetanus as public health problems. *Bull World Health Organ* **75**, 81-93 (1997).



- 84 Schauer, U. *et al.* Levels of antibodies specific to tetanus toxoid, Haemophilus influenzae type b, and pneumococcal capsular polysaccharide in healthy children and adults. *Clinical and diagnostic laboratory immunology* **10**, 202-207 (2003).
- 85 Amanna, I. J., Carlson, N. E. & Slifka, M. K. Duration of humoral immunity to common viral and vaccine antigens. *The New England journal of medicine* **357**, 1903-1915, doi:10.1056/NEJMoa066092 (2007).
- 86 van Riet, E. *et al.* Cellular and humoral responses to tetanus vaccination in Gabonese children. *Vaccine* **26**, 3690-3695, doi:10.1016/j.vaccine.2008.04.067 (2008).
- 87 Pobre, K. *et al.* Carrier priming or suppression: understanding carrier priming enhancement of anti-polysaccharide antibody response to conjugate vaccines. *Vaccine* **32**, 1423-1430, doi:10.1016/j.vaccine.2014.01.047 (2014).
- 88 Fotinou, C. *et al.* The crystal structure of tetanus toxin Hc fragment complexed with a synthetic GT1b analogue suggests cross-linking between ganglioside receptors and the toxin. *The Journal of biological chemistry* **276**, 32274-32281, doi:10.1074/jbc.M103285200 (2001).
- 89 Faber, C. *et al.* Three-dimensional structure of a human Fab with high affinity for tetanus toxoid. *Immunotechnology* **3**, 253-270 (1998).
- 90 Schatz, D. *et al.* Aging and the immune response to tetanus toxoid: diminished frequency and level of cellular immune reactivity to antigenic stimulation. *Clinical and diagnostic laboratory immunology* **5**, 894-896 (1998).
- 91 Meijer, P. J. *et al.* Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. *Journal of molecular biology* **358**, 764-772, doi:10.1016/j.jmb.2006.02.040 (2006).
- 92 Persson, M. A., Caothien, R. H. & Burton, D. R. Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 2432-2436 (1991).

- 93 de Kruif, J. *et al.* Human immunoglobulin repertoires against tetanus toxoid contain a large and diverse fraction of high-affinity promiscuous V(H) genes. *Journal of molecular biology* **387**, 548-558, doi:10.1016/j.jmb.2009.02.009 (2009).
- 94 DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* **31**, 166-169, doi:10.1038/nbt.2492 (2013).
- 95 Truck, J. *et al.* Identification of antigen-specific B cell receptor sequences using public repertoire analysis. *Journal of immunology (Baltimore, Md. : 1950)* **194**, 252-261, doi:10.4049/jimmunol.1401405 (2015).
- 96 Lavinder, J. J. *et al.* Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2259-2264, doi:10.1073/pnas.1317793111 (2014).
- 97 Poulsen, T. R., Meijer, P. J., Jensen, A., Nielsen, L. S. & Andersen, P. S. Kinetic, affinity, and diversity limits of human polyclonal antibody responses against tetanus toxoid. *Journal of immunology (Baltimore, Md. : 1950)* **179**, 3841-3850 (2007).
- 98 Sorouri, M., Fitzsimmons, S. P., Aydanian, A. G., Bennett, S. & Shapiro, M. A. Diversity of the antibody response to tetanus toxoid: comparison of hybridoma library to phage display library. *PLoS One* **9**, e106699, doi:10.1371/journal.pone.0106699 (2014).
- 99 Lavinder, J. J., Horton, A. P., Georgiou, G. & Ippolito, G. C. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Curr Opin Chem Biol* **24**, 112-120, doi:10.1016/j.cbpa.2014.11.007 (2015).
- 100 Engstrom, P. E., Nava, S., Mochizuki, S. & Norhagen, G. Quantitative analysis of IgA-subclass antibodies against tetanus toxoid. *Journal of immunoassay* **16**, 231-245 (1995).
- 101 Wiley, S. R. *et al.* Targeting TLRs expands the antibody repertoire in response to a malaria vaccine. *Sci Transl Med* **3**, 93ra69, doi:10.1126/scitranslmed.3002135 (2011).
- 102 Reed, S. G., Orr, M. T. & Fox, C. B. Key roles of adjuvants in modern vaccines. *Nat Med* **19**, 1597-1608, doi:10.1038/nm.3409 (2013).

- 103 Desmet, C. J. & Ishii, K. J. Nucleic acid sensing at the interface between innate and adaptive immunity in vaccination. *Nat Rev Immunol* **12**, 479-491, doi:10.1038/nri3247 (2012).
- 104 Cariappa, A., Liou, H. C., Horwitz, B. H. & Pillai, S. Nuclear factor kappa B is required for the development of marginal zone B lymphocytes. *J Exp Med* **192**, 1175-1182 (2000).
- 105 Gerondakis, S. & Siebenlist, U. Roles of the NF- $\kappa$ B Pathway in Lymphocyte Development and Function. *Cold Spring Harbor Perspectives in Biology* **2**, a000182, doi:10.1101/cshperspect.a000182 (2010).
- 106 Ehrlich, A. K. *et al.* Local Delivery of the Toll-Like Receptor 9 Ligand CpG Downregulates Host Immune and Inflammatory Responses, Ameliorating Established Leishmania (Viannia) panamensis Chronic Infection. *Infect Immun* **85**, doi:10.1128/iai.00981-16 (2017).
- 107 Kachura, M. A. *et al.* A CpG-Ficoll Nanoparticle Adjuvant for Anthrax Protective Antigen Enhances Immunogenicity and Provides Single-Immunization Protection against Inhaled Anthrax in Monkeys. *Journal of immunology (Baltimore, Md. : 1950)* **196**, 284-297, doi:10.4049/jimmunol.1501903 (2016).
- 108 Lofano, G. *et al.* Oil-in-Water Emulsion MF59 Increases Germinal Center B Cell Differentiation and Persistence in Response to Vaccination. *Journal of immunology (Baltimore, Md. : 1950)* **195**, 1617-1627, doi:10.4049/jimmunol.1402604 (2015).
- 109 Tschumper, R. C. *et al.* Comprehensive assessment of potential multiple myeloma immunoglobulin heavy chain V-D-J intracлонаl variation using massively parallel pyrosequencing. *Oncotarget* **3**, 502-513, doi:10.18632/oncotarget.469 (2012).
- 110 Sahota, S. S., Leo, R., Hamblin, T. J. & Stevenson, F. K. Myeloma VL and VH gene sequences reveal a complementary imprint of antigen selection in tumor cells. *Blood* **89**, 219-226 (1997).
- 111 Kosmas, C. *et al.* Molecular analysis of immunoglobulin genes in multiple myeloma. *Leuk Lymphoma* **33**, 253-265, doi:10.3109/10428199909058425 (1999).

- 112 Bakkus, M. H., Heirman, C., Van Riet, I., Van Camp, B. & Thielemans, K. Evidence that multiple myeloma Ig heavy chain VDJ genes contain somatic mutations but show no intraclonal variation. *Blood* **80**, 2326-2335 (1992).
- 113 Kosmas, C. *et al.* Somatic hypermutation of immunoglobulin variable region genes: focus on follicular lymphoma and multiple myeloma. *Immunol Rev* **162**, 281-292 (1998).
- 114 Hojjat-Farsangi, M. *et al.* Immunoglobulin heavy chain variable region gene usage and mutational status of the leukemic B cells in Iranian patients with chronic lymphocytic leukemia. *Cancer Sci* **100**, 2346-2353, doi:10.1111/j.1349-7006.2009.01341.x (2009).
- 115 Rosenquist, R. *et al.* Clonal evolution as judged by immunoglobulin heavy chain gene rearrangements in relapsing precursor-B acute lymphoblastic leukemia. *Eur J Haematol* **63**, 171-179 (1999).
- 116 Logan, A. C. *et al.* High-throughput VDJ sequencing for quantification of minimal residual disease in chronic lymphocytic leukemia and immune reconstitution assessment. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 21194-21199, doi:10.1073/pnas.1118357109 (2011).
- 117 Gawad, C. *et al.* Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. *Blood* **120**, 4407-4417, doi:10.1182/blood-2012-05-429811 (2012).
- 118 Langerak, A. W. *et al.* High-Throughput Immunogenetics for Clinical and Research Applications in Immunohematology: Potential and Challenges. *Journal of immunology (Baltimore, Md. : 1950)* **198**, 3765-3774, doi:10.4049/jimmunol.1602050 (2017).
- 119 Six, A. *et al.* The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. *Frontiers in Immunology* **4**, 413, doi:10.3389/fimmu.2013.00413 (2013).
- 120 Calis, J. J. & Rosenberg, B. R. Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends in immunology* **35**, 581-590, doi:10.1016/j.it.2014.09.004 (2014).

- 121 Greiff, V., Miho, E., Menzel, U. & Reddy, S. T. Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. *Trends in immunology* **36**, 738-749, doi:10.1016/j.it.2015.09.006 (2015).
- 122 Baum, P., Venturi, V. & Price, D. Wrestling with the repertoire: the promise and perils of next generation sequencing for antigen receptors. *European Journal of Immunology* **42**, 2834-2839 (2012).
- 123 Bolotin, D. A. *et al.* Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur J Immunol* **42**, 3073-3083, doi:10.1002/eji.201242517 (2012).
- 124 Hou, X. L., Wang, L., Ding, Y. L., Xie, Q. & Diao, H. Y. Current status and recent advances of next generation sequencing techniques in immunological repertoire. *Genes Immun* **17**, 153-164, doi:10.1038/gene.2016.9 (2016).
- 125 Lee, J. *et al.* Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat Med* **22**, 1456-1464, doi:10.1038/nm.4224 (2016).
- 126 Carlson, C. S. *et al.* Using synthetic templates to design an unbiased multiplex PCR assay. *Nature communications* **4**, 2680, doi:10.1038/ncomms3680 (2013).
- 127 Arnaout, R. *et al.* High-resolution description of antibody heavy-chain repertoires in humans. *PLoS ONE* **6**, e22365 (2011).
- 128 Bashford-Rogers, R. J. M. *et al.* Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome research* **23**, 1874-1884 (2013).
- 129 Briney, B., Le, K., Zhu, J. & Burton, D. R. Clonify: unseeded antibody lineage assignment from next-generation sequencing data. *Scientific reports* **6**, 23901, doi:10.1038/srep23901 (2016).

- 130 Best, K., Oakes, T., Heather, J. M., Shawe-Taylor, J. & Chain, B. Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. *Scientific reports* **5**, 14629, doi:10.1038/srep14629 (2015).
- 131 van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res* **322**, 12-20, doi:10.1016/j.yexcr.2014.01.008 (2014).
- 132 Turchaninova, M. A. *et al.* High-quality full-length immunoglobulin profiling with unique molecular barcoding. *Nature protocols* **11**, 1599-1616, doi:10.1038/nprot.2016.093 (2016).
- 133 Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* **32**, 158-168, doi:10.1038/nbt.2782 (2014).
- 134 Benichou, J., Ben Hamo, R., Louzoun, Y. & Efroni, S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* **135**, 183-191 (2012).
- 135 Boyd, S. D. & Joshi, S. A. High-Throughput DNA Sequencing Analysis of Antibody Repertoires. *Microbiology spectrum* **2**, doi:10.1128/microbiolspec.AID-0017-2014 (2014).
- 136 Robasky, K., Lewis, N. E. & Church, G. M. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* **15**, 56-62, doi:10.1038/nrg3655 (2014).
- 137 Wang, Y., Chen, W., Li, X. & Cheng, B. Degenerated primer design to amplify the heavy chain variable region from immunoglobulin cDNA. *BMC bioinformatics* **7 Suppl 4**, S9, doi:10.1186/1471-2105-7-s4-s9 (2006).
- 138 Berdoz, J., Monath, T. P. & Kraehenbuhl, J. P. Specific amplification by PCR of rearranged genomic variable regions of immunoglobulin genes from mouse hybridoma cells. *PCR methods and applications* **4**, 256-264 (1995).

- 139 Sun, Y., Liu, H. Y., Mu, L. & Luo, E. J. Degenerate primer design to clone the human repertoire of immunoglobulin heavy chain variable regions. *World journal of microbiology & biotechnology* **28**, 381-386, doi:10.1007/s11274-011-0830-3 (2012).
- 140 Lees, W. D. & Shepherd, A. J. Studying Antibody Repertoires with Next-Generation Sequencing. *Methods in molecular biology (Clifton, N.J.)* **1526**, 257-270, doi:10.1007/978-1-4939-6613-4\_15 (2017).
- 141 Robinson, W. H. Sequencing the functional antibody repertoire--diagnostic and therapeutic discovery. *Nature reviews. Rheumatology* **11**, 171-182, doi:10.1038/nrrheum.2014.220 (2015).
- 142 Cole, C., Volden, R., Dharmadhikari, S., Scelfo-Dalbey, C. & Vollmers, C. Highly Accurate Sequencing of Full-Length Immune Repertoire Amplicons Using Tn5-Enabled and Molecular Identifier-Guided Amplicon Assembly. *Journal of immunology (Baltimore, Md. : 1950)* **196**, 2902-2907, doi:10.4049/jimmunol.1502563 (2016).
- 143 Shlemov, A. *et al.* Reconstructing Antibody Repertoires from Error-Prone Immunosequencing Reads. *Journal of immunology (Baltimore, Md. : 1950)*, doi:10.4049/jimmunol.1700485 (2017).
- 144 Koenig, D. W. & Pierson, D. L. Microbiology of the Space Shuttle water system. *Water science and technology : a journal of the International Association on Water Pollution Research* **35**, 59-64 (1997).
- 145 Castro, V. A., Thrasher, A. N., Healy, M., Ott, C. M. & Pierson, D. L. Microbial characterization during the early habitation of the International Space Station. *Microbial ecology* **47**, 119-126, doi:10.1007/s00248-003-1030-y (2004).
- 146 Ott, C. M., Bruce, R. J. & Pierson, D. L. Microbial characterization of free floating condensate aboard the Mir space station. *Microbial ecology* **47**, 133-136, doi:10.1007/s00248-003-1038-3 (2004).
- 147 Novikova, N. *et al.* Survey of environmental biocontamination on board the International Space Station. *Research in microbiology* **157**, 5-12, doi:10.1016/j.resmic.2005.07.010 (2006).

- 148 Ilyin, V. K. Microbiological status of cosmonauts during orbital spaceflights on Salyut and Mir orbital stations. *Acta astronautica* **56**, 839-850 (2005).
- 149 Taylor, P. W. & Sommer, A. P. Towards rational treatment of bacterial infections during extended space travel. *International journal of antimicrobial agents* **26**, 183-187, doi:10.1016/j.ijantimicag.2005.06.002 (2005).
- 150 Gushanas, T. First Look at Findings of NASA Twins Study. doi:/feature/how-stressful-will-a-trip-to-mars-be-on-the-human-body-we-now-have-a-peek-into-what-the-nasa (2017).
- 151 Nickerson, C. A. *et al.* Microgravity as a novel environmental signal affecting Salmonella enterica serovar Typhimurium virulence. *Infect Immun* **68**, 3147-3152 (2000).
- 152 Wilson, J. W. *et al.* Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16299-16304, doi:10.1073/pnas.0707155104 (2007).
- 153 Rosado, H., Doyle, M., Hinds, J. & Taylor, P. W. Low-shear modelled microgravity alters expression of virulence determinants of Staphylococcus aureus. *Acta astronautica* **66**, 408-413, doi:<http://dx.doi.org/10.1016/j.actaastro.2009.06.007> (2010).
- 154 Leys, N. M., Hendrickx, L., De Boever, P., Baatout, S. & Mergeay, M. Space flight effects on bacterial physiology. *Journal of biological regulators and homeostatic agents* **18**, 193-199 (2004).
- 155 Klaus, D., Simske, S., Todd, P. & Stodieck, L. Investigation of space flight effects on Escherichia coli and a proposed model of underlying physical mechanisms. *Microbiology (Reading, England)* **143** ( Pt 2), 449-455, doi:10.1099/00221287-143-2-449 (1997).
- 156 Sonnenfeld, G. & Shearer, W. T. Immune function during space flight. *Nutrition (Burbank, Los Angeles County, Calif.)* **18**, 899-903 (2002).
- 157 Hoff, P. *et al.* Effects of 60-day bed rest with and without exercise on cellular and humoral immunological parameters. *Cellular & molecular immunology* **12**, 483-492, doi:10.1038/cmi.2014.106 (2015).



- 158 Taylor, G. R. & Janney, R. P. In vivo testing confirms a blunting of the human cell-mediated immune mechanism during space flight. *Journal of leukocyte biology* **51**, 129-132 (1992).
- 159 Licato, L. L. & Grimm, E. A. Multiple interleukin-2 signaling pathways differentially regulated by microgravity. *Immunopharmacology* **44**, 273-279 (1999).
- 160 Stein, T. P. Nutrition and muscle loss in humans during spaceflight. *Advances in space biology and medicine* **7**, 49-97 (1999).
- 161 Chapes, S. K. & Ganta, R. R. Mouse infection models for space flight immunology. *Advances in space biology and medicine* **10**, 81-104 (2005).
- 162 Pecaut, M. J. *et al.* Is spaceflight-induced immune dysfunction linked to systemic changes in metabolism? *PLoS One* **12**, e0174174, doi:10.1371/journal.pone.0174174 (2017).
- 163 Borchers, A. T., Keen, C. L. & Gershwin, M. E. Microgravity and immune responsiveness: implications for space travel. *Nutrition (Burbank, Los Angeles County, Calif.)* **18**, 889-898 (2002).
- 164 Crucian, B. *et al.* Terrestrial stress analogs for spaceflight associated immune system dysregulation. *Brain, behavior, and immunity* **39**, 23-32, doi:10.1016/j.bbi.2014.01.011 (2014).
- 165 Chapes, S. K., Mastro, A. M., Sonnenfeld, G. & Berry, W. D. Antiorthostatic suspension as a model for the effects of spaceflight on the immune system. *Journal of leukocyte biology* **54**, 227-235 (1993).
- 166 Steffen, J. M. *et al.* A suspension model for hypokinetic/hypodynamic and antiorthostatic responses in the mouse. *Aviation, space, and environmental medicine* **55**, 612-616 (1984).
- 167 Musacchia, X. J., Steffen, J. M., Fell, R. D. & Dombrowski, J. Physiological comparison of rat muscle in body suspension and weightlessness. *The Physiologist* **30**, S102-105 (1987).

- 168 Fleming, S. D., Rosenkrans, C. F., Jr. & Chapes, S. K. Test of the antiorthostatic suspension model on mice: effects on the inflammatory cell response. *Aviation, space, and environmental medicine* **61**, 327-332 (1990).
- 169 Morey-Holton, E. R. & Arnaud, S. B. Skeletal responses to spaceflight. *Advances in space biology and medicine* **1**, 37-69 (1991).
- 170 Hargens, A. R. *et al.* Long-term measurement of muscle function in the dog hindlimb using a new apparatus. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society* **1**, 284-291, doi:10.1002/jor.1100010308 (1984).
- 171 Crucian, B. E. *et al.* Plasma cytokine concentrations indicate that in vivo hormonal regulation of immunity is altered during long-duration spaceflight. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* **34**, 778-786, doi:10.1089/jir.2013.0129 (2014).
- 172 Stein, T. P., Schluter, M. D. & Leskiw, M. J. Cortisol, insulin and leptin during space flight and bed rest. *Journal of gravitational physiology : a journal of the International Society for Gravitational Physiology* **6**, P85-86 (1999).
- 173 Fleming, S. D., Edelman, L. S. & Chapes, S. K. Effects of Corticosterone and Microgravity on Inflammatory Cell Production of Superoxide. *Journal of leukocyte biology* **50**, 69-76, doi:10.1002/jlb.50.1.69 (1991).
- 174 Crucian, B. E., Cubbage, M. L. & Sams, C. F. Altered cytokine production by specific human peripheral blood cell subsets immediately following space flight. *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* **20**, 547-556, doi:10.1089/10799900050044741 (2000).
- 175 Gridley, D. S. *et al.* Spaceflight effects on T lymphocyte distribution, function and gene expression. *Journal of applied physiology (Bethesda, Md. : 1985)* **106**, 194-202, doi:10.1152/japplphysiol.91126.2008 (2009).

- 176 Gaignier, F. *et al.* Three weeks of murine hindlimb unloading induces shifts from B to T and from th to tc splenic lymphocytes in absence of stress and differentially reduces cell-specific mitogenic responses. *PLoS One* **9**, e92664, doi:10.1371/journal.pone.0092664 (2014).
- 177 Lescale, C. *et al.* Hind limb unloading, a model of spaceflight conditions, leads to decreased B lymphopoiesis similar to aging. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **29**, 455-463, doi:10.1096/fj.14-259770 (2015).
- 178 Cogoli, A. & Tschopp, A. Lymphocyte reactivity during spaceflight. *Immunology today* **6**, 1-4, doi:10.1016/0167-5699(85)90151-3 (1985).
- 179 Taylor, G. R., Neale, L. S. & Dardano, J. R. Immunological analyses of U.S. Space Shuttle crewmembers. *Aviation, space, and environmental medicine* **57**, 213-217 (1986).
- 180 Armstrong, J. W., Balch, S. & Chapes, S. K. Interleukin-2 therapy reverses some immunosuppressive effects of skeletal unloading. *Journal of applied physiology (Bethesda, Md. : 1985)* **77**, 584-589 (1994).
- 181 Cogoli, A. Signal transduction in T lymphocytes in microgravity. *Gravitational and space biology bulletin : publication of the American Society for Gravitational and Space Biology* **10**, 5-16 (1997).
- 182 Lewis, M. L. The cytoskeleton, apoptosis, and gene expression in T lymphocytes and other mammalian cells exposed to altered gravity. *Advances in space biology and medicine* **8**, 77-128 (2002).
- 183 Hughes-Fulford, M., Chang, T. T., Martinez, E. M. & Li, C. F. Spaceflight alters expression of microRNA during T-cell activation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **29**, 4893-4900, doi:10.1096/fj.15-277392 (2015).
- 184 Gueguinou, N. *et al.* Stress response and humoral immune system alterations related to chronic hypergravity in mice. *Psychoneuroendocrinology* **37**, 137-147, doi:10.1016/j.psyneuen.2011.05.015 (2012).

- 185 Ortega, M. T., Lu, N. & Chapes, S. K. Evaluation of in vitro macrophage differentiation during space flight. *Advances in space research : the official journal of the Committee on Space Research (COSPAR)* **49**, 1441-1455, doi:10.1016/j.asr.2012.02.021 (2012).
- 186 Armstrong, J. W., Kirby-Dobbels, K. & Chapes, S. K. The effects of rM-CSF and rIL-6 therapy on immunosuppressed antiorthostatically suspended mice. *Journal of applied physiology (Bethesda, Md. : 1985)* **78**, 968-975 (1995).
- 187 Morey-Holton, E. R. & Globus, R. K. Hindlimb unloading of growing rats: a model for predicting skeletal changes during space flight. *Bone* **22**, 83s-88s (1998).
- 188 Ortega, M. T. *et al.* Shifts in bone marrow cell phenotypes caused by spaceflight. *Journal of applied physiology (Bethesda, Md. : 1985)* **106**, 548-555, doi:10.1152/jappphysiol.91138.2008 (2009).
- 189 Nash, P. V. & Mastro, A. M. Variable lymphocyte responses in rats after space flight. *Exp Cell Res* **202**, 125-131 (1992).
- 190 Morey-Holton, E. R., Halloran, B. P., Garetto, L. P. & Doty, S. B. Animal housing influences the response of bone to spaceflight in juvenile rats. *Journal of Applied Physiology* **88**, 1303-1309, doi:10.1152/jappl.2000.88.4.1303 (2000).
- 191 Bascove, M., Huin-Schohn, C., Gueguinou, N., Tschirhart, E. & Frippiat, J. P. Spaceflight-associated changes in immunoglobulin VH gene expression in the amphibian *Pleurodeles waltl*. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **23**, 1607-1615, doi:10.1096/fj.08-121327 (2009).
- 192 Gould, C. L. & Sonnenfeld, G. Enhancement of viral pathogenesis in mice maintained in an antiorthostatic suspension model: coordination with effects on interferon production. *Journal of biological regulators and homeostatic agents* **1**, 33-36 (1987).
- 193 Belay, T., Aviles, H., Vance, M., Fountain, K. & Sonnenfeld, G. Effects of the hindlimb-unloading model of spaceflight conditions on resistance of mice to infection with *Klebsiella pneumoniae*. *The Journal of allergy and clinical immunology* **110**, 262-268 (2002).

- 194 Voss, E. W., Jr. Prolonged weightlessness and humoral immunity. *Science* **225**, 214-215 (1984).
- 195 Rettig, T. A., Ward, C., Bye, B. A., Pecaut, M. J. & Chapes, S. K. Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. *PLoS One* **13**, e0190982, doi:10.1371/journal.pone.0190982 (2018).
- 196 de Bono, B., Madera, M. & Chothia, C. VH gene segments in the mouse and human genomes. *Journal of molecular biology* **342**, 131-143, doi:10.1016/j.jmb.2004.06.055 (2004).
- 197 Saul, F. A. & Poljak, R. J. Crystal structure of human immunoglobulin fragment Fab New refined at 2.0 Å resolution. *Proteins* **14**, 363-371, doi:10.1002/prot.340140305 (1992).
- 198 Haidar, J. N. *et al.* Backbone flexibility of CDR3 and immune recognition of antigens. *Journal of molecular biology* **426**, 1583-1599, doi:10.1016/j.jmb.2013.12.024 (2014).
- 199 Lu, J. *et al.* IgG variable region and VH CDR3 diversity in unimmunized mice analyzed by massively parallel sequencing. *Molecular immunology* **57**, 274-283, doi:10.1016/j.molimm.2013.09.008 (2014).
- 200 Xu, J. L. & Davis, M. M. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* **13**, 37-45 (2000).
- 201 Durnova, G., Kaplansky, A. & Portugalov. Effect of a 22-day space flight on the lymphoid organs of rats. *Aviat. Space Environ. Med.* **47**, 588-591 (1976).
- 202 Serova, L. V. Weightlessness effects on resistance and reactivity of animals. *The Physiologist* **23**, S22-26 (1980).
- 203 Grindeland, R., Popova, I., Vasques, M. & Arnaud, S. Cosmos 1887 mission overview: effects of microgravity on rat body and adrenal weights and plasma constituents. *FASEB J.* **4**, 105-109 (1990).
- 204 G, J., J, M., T, F., N, H. & G, Z. in *43rd Congress of the International Astronautical Federation*,. 1-8.

- 205 Grove, D., Pishak, A. & Mastro, A. The effect of a 10-day space flight on the function, phenotype, and adhesion molecule expression of splenocytes and lymph node lymphocytes. *Exp. Cell Res.* **219**, 102-109 (1995).
- 206 Udden, M. M. *et al.* Blood volume and erythropoiesis in the rat during spaceflight. *Aviation, space, and environmental medicine* **66**, 557-561 (1995).
- 207 Allebban, Z. *et al.* Effects of spaceflight on rat erythroid parameters. *J. Appl. Physiol.* **81**, 117-122 (1996).
- 208 Congdon, C. C. *et al.* Lymphatic tissue changes in rats flown on Spacelab Life Sciences-2. *J Appl Physiol* **81**, 172-177 (1996).
- 209 Wronski, T. J. *et al.* Lack of an effect of space flight on bone mass and bone formation in group-housed rats. *J. Applied Physiol.* **85**, 279-285 (1998).
- 210 Chapes, S. K., Simske, S. J., Sonnenfeld, G., Miller, E. S. & Zimmerman, R. J. Effects of spaceflight and PEG-IL-2 on rat physiological and immunological responses. *J. Applied Physiol.* **86**, 2065-2076 (1999).
- 211 Pecaut, M. J., Simske, S. J. & Fleshner, M. Spaceflight induces changes in splenocyte subpopulations: effectiveness of ground-based models. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **279**, R2072-2078 (2000).
- 212 Merrill, A., Jr., Wang, E., Mullins, R., Grindeland, R. & Popova, I. Analyses of plasma for metabolic and hormonal changes in rats flown aboard COSMOS 2044. *J. Appl. Physiol.* **73**, 132S-135S (1992).
- 213 Lesnyak, A. *et al.* Immune changes in test animals during spaceflight. *J. Leukoc. Biol.* **54**, 214-226 (1993).
- 214 Meehan, R., Whitson, P. & Sams, C. The role of psychoneuroendocrine factors on spaceflight-induced immunological alterations. *J. Leukoc. Biol.* **54**, 236-244 (1993).

- 215 Stein, T. P. & Schluter, M. D. Excretion of IL-6 by astronauts during spaceflight. *The American journal of physiology* **266**, E448-452, doi:10.1152/ajpendo.1994.266.3.E448 (1994).
- 216 Blanc, S. *et al.* Energy and water metabolism, body composition, and hormonal changes induced by 42 days of enforced inactivity and simulated weightlessness. *The Journal of clinical endocrinology and metabolism* **83**, 4289-4297, doi:10.1210/jc.83.12.4289 (1998).
- 217 Stowe, R. P. *et al.* Leukocyte subsets and neutrophil function after short-term spaceflight. *Journal of leukocyte biology* **65**, 179-186 (1999).
- 218 Stowe, R. P., Pierson, D. L. & Barrett, A. D. Elevated stress hormone levels relate to Epstein-Barr virus reactivation in astronauts. *Psychosomatic medicine* **63**, 891-895 (2001).
- 219 Stowe, R. P., Mehta, S. K., Ferrando, A. A., Feeback, D. L. & Pierson, D. L. Immune responses and latent herpesvirus reactivation in spaceflight. *Aviat. Space Environ. Med.* **72**, 884-891. (2001).
- 220 Konstantinova, I. V., Antropova, E. N., Legen'kov, V. I. & Zazhirei, V. D. Reactivity of lymphoid blood cells in the crew of Soiuz-6, Soiuz-7 and Soiuz-8 spacecraft before and after flight. *Kosmicheskaiia biologiiia i meditsina* **7**, 35-40 (1973).
- 221 Mandel, A. D. & Balish, E. Effect of space flight on cell-mediated immunity. *Aviation, space, and environmental medicine* **48**, 1051-1057 (1977).
- 222 Taylor, G. & Dardano, J. US/USSR space biology and medicine: Human cellular immune responsiveness following space flight. *Aviat. Space and Environ. Med.* **54**, S55-S59 (1983).
- 223 A, C., B, B., O, M. & E, H. in *Biorack on Spacelab D1* Vol. ESA SP-1091 89-100 (1990).
- 224 Sonnenfeld, G. *et al.* Effects of spaceflight on levels and activity of immune cells. *Aviation, space, and environmental medicine* **61**, 648-653 (1990).
- 225 Hughes-Fulford, M. Altered cell function in microgravity. *Experimental gerontology* **26**, 247-256 (1991).

- 226 Sonnenfeld, G. *et al.* Space flight alters immune cell function and distribution. *J. Appl. Physiol.* **73**, 191S-195S (1992).
- 227 Fuchs, B. & Medvedev, A. Countermeasures for ameliorating in-flight immune dysfunction. *J. Leukoc. Biol.* **54**, 245-252 (1993).
- 228 Grigoriev, A. I. *et al.* Main medical results of extended flights on space station Mir in 1986-1990. *Acta astronautica* **29**, 581-585 (1993).
- 229 Bikle, D. D., Harris, J., Halloran, B. P. & Morey-Holton, E. Altered skeletal pattern of gene expression in response to spaceflight and hindlimb elevation. *The American journal of physiology* **267**, E822-827, doi:10.1152/ajpendo.1994.267.6.E822 (1994).
- 230 Miller, E., Koebel, D. & Sonnenfeld, G. Influence of spaceflight on the production of interleukin-3 and interleukin-6 by rat spleen and thymus cells. *J. Appl. Physiol.* **78**, 810-813 (1995).
- 231 Sonnenfeld, G. *et al.* Effect of space flight on cytokine production and other immunologic parameters of rhesus monkeys. *J. Interferon and Cytokine Res.* **16**, 409-415 (1996).
- 232 Sonnenfeld, G. *et al.* Spaceflight and development of immune responses. *J Appl Physiol* **85**, 1429-1433 (1998).
- 233 Berry, C. A. Summary of medical experience in the Apollo 7 through 11 manned spaceflights. *Aerospace medicine* **41**, 500-519 (1970).
- 234 Meehan, R. T. *et al.* Alteration in human mononuclear leucocytes following space flight. *Immunology* **76**, 491-497 (1992).
- 235 Allebban, Z. *et al.* Effects of Spaceflight on the Number of Rat Peripheral-Blood Leukocytes and Lymphocyte Subsets. *Journal of leukocyte biology* **55**, 209-213 (1994).
- 236 Ichiki, A. *et al.* Effects of spaceflight on rat peripheral blood leukocytes and bone marrow progenitor cells. *J. Leukoc. Biol.* **60**, 37-43 (1996).



- 237 Sonnenfeld, G. The immune system in space, including Earth-based benefits of space-based research. *Current pharmaceutical biotechnology* **6**, 343-349 (2005).
- 238 Crucian, B. *et al.* Alterations in adaptive immunity persist during long-duration spaceflight. *NPJ microgravity* **1**, 15013, doi:10.1038/npjmgrav.2015.13 (2015).
- 239 Fitzgerald, W. *et al.* Immune suppression of human lymphoid tissues and cells in rotating suspension culture and onboard the International Space Station. *In vitro cellular & developmental biology* (2009).
- 240 Figliozzi, G. M. New Rodent Residence Elevates Research To Greater Heights. doi:/mission\_pages/station/research/news/rodent\_research (2015).
- 241 Aoki-Ota, M., Torkamani, A., Ota, T., Schork, N. & Nemazee, D. Skewed primary Igkappa repertoire and V-J joining in C57BL/6 mice: implications for recombination accessibility and receptor editing. *Journal of immunology (Baltimore, Md. : 1950)* **188**, 2305-2315, doi:10.4049/jimmunol.1103484 (2012).
- 242 Kaplinsky, J. *et al.* Antibody repertoire deep sequencing reveals antigen-independent selection in maturing B cells. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2622-2629, doi:10.1073/pnas.1403278111 (2014).
- 243 Kramer, J. M. *et al.* Analysis of IgM antibody production and repertoire in a mouse model of Sjogren's syndrome. *Journal of leukocyte biology* **99**, 321-331, doi:10.1189/jlb.2A0715-297R (2016).
- 244 Greiff, V. *et al.* Quantitative assessment of the robustness of next-generation sequencing of antibody variable gene repertoires from immunized mice. *BMC Immunol* **15**, 40, doi:10.1186/s12865-014-0040-5 (2014).
- 245 Carey, J. B., Moffatt-Blue, C. S., Watson, L. C., Gavin, A. L. & Feeney, A. J. Repertoire-based selection into the marginal zone compartment during B cell development. *J Exp Med* **205**, 2043-2052, doi:10.1084/jem.20080559 (2008).

- 246 Schelonka, R. L. *et al.* Categorical selection of the antibody repertoire in splenic B cells. *Eur J Immunol* **37**, 1010-1021, doi:10.1002/eji.200636569 (2007).
- 247 Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome biology* **12**, R112, doi:10.1186/gb-2011-12-11-r112 (2011).
- 248 Racanelli, V. *et al.* Antibody V(h) repertoire differences between resolving and chronically evolving hepatitis C virus infections. *PLoS One* **6**, e25606, doi:10.1371/journal.pone.0025606 (2011).
- 249 Fan, Q. *et al.* Plasmablast Response to Primary Rhesus Cytomegalovirus (CMV) Infection in a Monkey Model of Congenital CMV Transmission. *Clinical and vaccine immunology : CVI* **24**, doi:10.1128/cvi.00510-16 (2017).
- 250 Nivarthi, U. K. *et al.* Mapping the Human Memory B Cell and Serum Neutralizing Antibody Responses to Dengue Virus Serotype 4 Infection and Vaccination. *Journal of virology* **91**, doi:10.1128/jvi.02041-16 (2017).
- 251 Faham, M. *et al.* Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia. *Blood* **120**, 5173-5180, doi:10.1182/blood-2012-07-444042 (2012).
- 252 Gerasimov, E., Zelikovsky, A., Mandoiu, I. & Ionov, Y. Identification of cancer-specific motifs in mimotope profiles of serum antibody repertoire. *BMC bioinformatics* **18**, 244, doi:10.1186/s12859-017-1661-5 (2017).
- 253 Alamyar, E., Duroux, P., Lefranc, M.-P. & Giudicelli, V. in *Immunogenetics* Vol. 882 *Methods in Molecular Biology* (eds Frank T. Christiansen & Brian D. Tait) Ch. 32, 569-604 (Humana Press, 2012).
- 254 *IMGT Repertoire (IG and TR)*,  
<[http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus\\_musculus&group=IGHV\\_strains](http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus_musculus&group=IGHV_strains)> (2017).

- 255 Hansen, K. D., Brenner, S. E. & Dudoit, S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res* **38**, e131, doi:10.1093/nar/gkq224 (2010).
- 256 van Gurp, T. P., McIntyre, L. M. & Verhoeven, K. J. Consistent errors in first strand cDNA due to random hexamer mispriming. *PLoS One* **8**, e85583, doi:10.1371/journal.pone.0085583 (2013).
- 257 Weber, C. F. & Kuske, C. R. Comparative assessment of fungal cellobiohydrolase I richness and composition in cDNA generated using oligo(dT) primers or random hexamers. *Journal of microbiological methods* **88**, 224-228, doi:10.1016/j.mimet.2011.11.016 (2012).
- 258 Zheng, W., Chung, L. M. & Zhao, H. Bias detection and correction in RNA-Sequencing data. *BMC bioinformatics* **12**, 290, doi:10.1186/1471-2105-12-290 (2011).
- 259 Zhang, Q. *et al.* Regulation of production of mucosal antibody to pneumococcal protein antigens by T-cell-derived gamma interferon and interleukin-10 in children. *Infect Immun* **74**, 4735-4743, doi:10.1128/iai.00165-06 (2006).
- 260 Lou, Z., Casali, P. & Xu, Z. Regulation of B Cell Differentiation by Intracellular Membrane-Associated Proteins and microRNAs: Role in the Antibody Response. *Frontiers in Immunology* **6**, 537, doi:10.3389/fimmu.2015.00537 (2015).
- 261 Shahaf, G. *et al.* Antigen-driven selection in germinal centers as reflected by the shape characteristics of immunoglobulin gene lineage trees: a large-scale simulation study. *J Theor Biol* **255**, 210-222, doi:10.1016/j.jtbi.2008.08.005 (2008).
- 262 Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575-581 (1983).
- 263 Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* **19**, 981-992 (1980).

- 264 Tonegawa, S. Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 203-207 (1976).
- 265 Kabat, E. A., Wu, T. T. & Bilofsky, H. Evidence supporting somatic assembly of the DNA segments (minigenes), coding for the framework, and complementarity-determining segments of immunoglobulin variable regions. *J Exp Med* **149**, 1299-1313 (1979).
- 266 Xu, Z., Zan, H., Pone, E. J., Mai, T. & Casali, P. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat Rev Immunol* **12**, 517-531, doi:10.1038/nri3216 (2012).
- 267 Ippolito, G. C. *et al.* Forced usage of positively charged amino acids in immunoglobulin CDR-H3 impairs B cell development and antibody production. *J Exp Med* **203**, 1567-1578, doi:10.1084/jem.20052217 (2006).
- 268 Jiang, N. *et al.* Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5348-5353, doi:10.1073/pnas.1014277108 (2011).
- 269 Reddy, S. T. *et al.* Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol* **28**, 965-969, doi:10.1038/nbt.1673 (2010).
- 270 Gray, S. A. *et al.* Selection of therapeutic H5N1 monoclonal antibodies following IgVH repertoire analysis in mice. *Antiviral research* **131**, 100-108, doi:10.1016/j.antiviral.2016.04.001 (2016).
- 271 Parameswaran, P. *et al.* Convergent antibody signatures in human dengue. *Cell host & microbe* **13**, 691-700, doi:10.1016/j.chom.2013.05.008 (2013).
- 272 van Belzen, N. *et al.* Detection of minimal disease using rearranged immunoglobulin heavy chain genes from intermediate- and high-grade malignant B cell non-Hodgkins lymphoma. *Leukemia* **11**, 1742-1752 (1997).

- 273 Bashford-Rogers, R. J. *et al.* Eye on the B-ALL: B-cell receptor repertoires reveal persistence of numerous B-lymphoblastic leukemia subclones from diagnosis to relapse. *Leukemia* **30**, 2312-2321, doi:10.1038/leu.2016.142 (2016).
- 274 Zuckerman, N. S. *et al.* Ectopic GC in the thymus of myasthenia gravis patients show characteristics of normal GC. *Eur J Immunol* **40**, 1150-1161, doi:10.1002/eji.200939914 (2010).
- 275 Tan, Y. C. *et al.* High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clinical immunology (Orlando, Fla.)* **151**, 55-65, doi:10.1016/j.clim.2013.12.008 (2014).
- 276 Briney, B. S., Willis, J. R., McKinney, B. A. & Crowe, J. E. High-throughput antibody sequencing reveals genetic evidence of global regulation of the naive and memory repertoires that extends across individuals. *Genes and Immunity* **13**, 469-473 (2012).
- 277 Boyd, S. *et al.* Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *The journal of immunology* **184**, 6986-6992 (2010).
- 278 Watson, C. T. *et al.* Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *Am J Hum Genet* **92**, 530-546, doi:10.1016/j.ajhg.2013.03.004 (2013).
- 279 Sasso, E. H., Van Dijk, K. W. & Milner, E. C. Prevalence and polymorphism of human VH3 genes. *Journal of immunology (Baltimore, Md. : 1950)* **145**, 2751-2757 (1990).
- 280 Milner, E. C., Hufnagle, W. O., Glas, A. M., Suzuki, I. & Alexander, C. Polymorphism and utilization of human VH Genes. *Annals of the New York Academy of Sciences* **764**, 50-61 (1995).
- 281 Wang, Y. *et al.* Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics* **63**, 259-265, doi:10.1007/s00251-010-0510-8 (2011).
- 282 Huerkamp, M. J. It's in the bag: Easy and medically sound rodent gas anesthesia induction. *Tech Talk* **5**, 3 (2000).

- 283 Alamyar, E., Giudicelli, V., Li, S., Duroux, P. & Lefranc, M.-P. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome res* **8**, 26 (2012).
- 284 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).
- 285 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).
- 286 Kono, N. *et al.* Deciphering antigen-responding antibody repertoires by using next-generation sequencing and confirming them through antibody-gene synthesis. *Biochem Biophys Res Commun* **487**, 300-306, doi:10.1016/j.bbrc.2017.04.054 (2017).
- 287 Choi, N. M. *et al.* Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *Journal of immunology (Baltimore, Md. : 1950)* **191**, 2393-2402, doi:10.4049/jimmunol.1301279 (2013).
- 288 Klein-Schneegans, A. S., Kuntz, L., Fonteneau, P. & Loor, F. Serum concentrations of IgM, IgG1, IgG2b, IgG3 and IgA in C57BL/6 mice and their congenics at the *lpr* (lymphoproliferation) locus. *Journal of autoimmunity* **2**, 869-875 (1989).
- 289 Glanville, J. *et al.* Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 20066-20071, doi:10.1073/pnas.1107498108 (2011).
- 290 Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Statistics in medicine* **20**, 3123-3157 (2001).
- 291 Chapman, D. G. & University of California, B. *Some properties of the hypergeometric distribution with applications to zoological sample censuses.* (University of California Press, 1951).

- 292 Kunik, V., Peters, B. & Ofran, Y. Structural consensus among antibodies defines the antigen binding site. *PLoS computational biology* **8**, e1002388, doi:10.1371/journal.pcbi.1002388 (2012).
- 293 Sela-Culang, I., Kunik, V. & Ofran, Y. The structural basis of antibody-antigen recognition. *Frontiers in Immunology* **4**, 302, doi:10.3389/fimmu.2013.00302 (2013).
- 294 Banga, S. *et al.* Impact of acute malaria on pre-existing antibodies to viral and vaccine antigens in mice and humans. *PLoS One* **10**, e0125090, doi:10.1371/journal.pone.0125090 (2015).
- 295 Boyd, S. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Science translational medicine* **1**, 12 (2009).
- 296 DeKosky, B. J. *et al.* In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* **21**, 86-91, doi:10.1038/nm.3743 (2015).
- 297 Busse, C. E., Czogiel, I., Braun, P., Arndt, P. F. & Wardemann, H. Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *Eur J Immunol* **44**, 597-603, doi:10.1002/eji.201343917 (2014).
- 298 Rykova, M. P., Antropova, E. N., Larina, I. M. & Morukov, B. V. Humoral and cellular immunity in cosmonauts after the ISS missions. *Acta astronautica* **63**, 697-705 (2008).
- 299 Crucian, B. *et al.* Immune system dysregulation occurs during short duration spaceflight on board the space shuttle. *Journal of clinical immunology* **33**, 456-465, doi:10.1007/s10875-012-9824-7 (2013).
- 300 Globus, R. K. & Morey-Holton, E. Hindlimb unloading: rodent analog for microgravity. *Journal of applied physiology (Bethesda, Md. : 1985)* **120**, 1196-1206, doi:10.1152/jappphysiol.00997.2015 (2016).
- 301 Nickerson, C. A. *et al.* Low-shear modeled microgravity: a global environmental regulatory signal affecting bacterial gene expression, physiology, and pathogenesis. *Journal of microbiological methods* **54**, 1-11 (2003).

- 302 Chapes, S. K., Simske, S. J., Forsman, A. D., Bateman, T. A. & Zimmerman, R. J. Effects of space flight and IGF-1 on immune function. *Advances in space research : the official journal of the Committee on Space Research (COSPAR)* **23**, 1955-1964 (1999).
- 303 Gridley, D. S. *et al.* Changes in mouse thymus and spleen after return from the STS-135 mission in space. *PLoS One* **8**, e75097, doi:10.1371/journal.pone.0075097 (2013).
- 304 Pecaut, M. J. *et al.* Genetic models in applied physiology: selected contribution: effects of spaceflight on immunity in the C57BL/6 mouse. I. Immune population distributions. *Journal of applied physiology (Bethesda, Md. : 1985)* **94**, 2085-2094, doi:10.1152/jappphysiol.01052.2002 (2003).
- 305 Wei, L. X., Zhou, J. N., Roberts, A. I. & Shi, Y. F. Lymphocyte reduction induced by hindlimb unloading: distinct mechanisms in the spleen and thymus. *Cell research* **13**, 465-471, doi:10.1038/sj.cr.7290189 (2003).
- 306 Baqai, F. P. *et al.* Effects of spaceflight on innate immune function and antioxidant gene expression. *Journal of applied physiology (Bethesda, Md. : 1985)* **106**, 1935-1942, doi:10.1152/jappphysiol.91361.2008 (2009).
- 307 Armstrong, J. W. *et al.* Skeletal unloading causes organ-specific changes in immune cell responses. *Journal of applied physiology (Bethesda, Md. : 1985)* **75**, 2734-2739 (1993).
- 308 Cogoli, A., Tschopp, A. & Fuchs-Bislin, P. Cell sensitivity to gravity. *Science* **225**, 228-230 (1984).
- 309 Nash, P. *et al.* Effect of space flight on lymphocyte proliferation and interleukin-2 production. *J. Appl. Physiol.* **73**, 186S-190S (1992).
- 310 Lesnyak, A. *et al.* Effect of SLS-2 spaceflight on immunologic parameters of rats. *J. Appl. Physiol.* **81**, 178-182 (1996).
- 311 Cooper, D., Pride, M. W., Brown, E. L., Risin, D. & Pellis, N. R. Suppression of antigen-specific lymphocyte activation in modeled microgravity. *In vitro cellular & developmental biology* **37**, 63-65, doi:10.1290/1071-2690(2001)037<0063:soasla>2.0.co;2 (2001).



- 312 Cogoli-Greuter, M. Effect of Gravity Changes on the Cytoskeleton in Human Lymphocytes. *Gravitational and Space Biology Bulletin* **17**, 27-38 (2004).
- 313 Chang, T. T. *et al.* The Rel/NF-kappaB pathway and transcription of immediate early genes in T cell activation are inhibited by microgravity. *Journal of leukocyte biology* **92**, 1133-1145, doi:10.1189/jlb.0312157 (2012).
- 314 Sanzari, J. K. *et al.* Leukocyte activity is altered in a ground based murine model of microgravity and proton radiation exposure. *PLoS One* **8**, e71757, doi:10.1371/journal.pone.0071757 (2013).
- 315 Hwang, S. A., Crucian, B., Sams, C. & Actor, J. K. Post-Spaceflight (STS-135) Mouse Splenocytes Demonstrate Altered Activation Properties and Surface Molecule Expression. *PLoS One* **10**, e0124380, doi:10.1371/journal.pone.0124380 (2015).
- 316 Martinez, E. M., Yoshida, M. C., Candelario, T. L. & Hughes-Fulford, M. Spaceflight and simulated microgravity cause a significant reduction of key gene expression in early T-cell activation. *American journal of physiology. Regulatory, integrative and comparative physiology* **308**, R480-488, doi:10.1152/ajpregu.00449.2014 (2015).
- 317 Tauber, S. *et al.* Signal transduction in primary human T lymphocytes in altered gravity during parabolic flight and clinostat experiments. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* **35**, 1034-1051, doi:10.1159/000373930 (2015).
- 318 Alt, F. W. *et al.* Ordered rearrangement of immunoglobulin heavy chain variable region segments. *The EMBO journal* **3**, 1209-1219 (1984).
- 319 Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C. & Mathis, D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* **261**, 1175-1178 (1993).
- 320 Komori, T., Okada, A., Stewart, V. & Alt, F. W. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* **261**, 1171-1175 (1993).

- 321 Wardemann, H. & Busse, C. E. Novel Approaches to Analyze Immunoglobulin Repertoires. *Trends in immunology* **38**, 471-482, doi:10.1016/j.it.2017.05.003 (2017).
- 322 Tan, Y. G. *et al.* Clonal Characteristics of Circulating B Lymphocyte Repertoire in Primary Biliary Cholangitis. *Journal of immunology (Baltimore, Md. : 1950)* **197**, 1609-1620, doi:10.4049/jimmunol.1600096 (2016).
- 323 Jiang, Y. *et al.* VDJ-Seq: Deep Sequencing Analysis of Rearranged Immunoglobulin Heavy Chain Gene to Reveal Clonal Evolution Patterns of B Cell Lymphoma. *Journal of visualized experiments : JoVE*, e53215, doi:10.3791/53215 (2015).
- 324 Montesinos-Rongen, M., Purschke, F., Kuppers, R. & Deckert, M. Immunoglobulin repertoire of primary lymphomas of the central nervous system. *Journal of neuropathology and experimental neurology* **73**, 1116-1125, doi:10.1097/nen.0000000000000133 (2014).
- 325 Boxio, R., Dournon, C. & Fripiat, J. P. Effects of a long-term spaceflight on immunoglobulin heavy chains of the urodele amphibian *Pleurodeles waltl*. *Journal of applied physiology (Bethesda, Md. : 1985)* **98**, 905-910, doi:10.1152/jappphysiol.00957.2004 (2005).
- 326 Bascove, M., Gueguinou, N., Schaerlinger, B., Gauquelin-Koch, G. & Fripiat, J. P. Decrease in antibody somatic hypermutation frequency under extreme, extended spaceflight conditions. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**, 2947-2955, doi:10.1096/fj.11-185215 (2011).
- 327 Huin-Schohn, C. *et al.* Gravity changes during animal development affect IgM heavy-chain transcription and probably lymphopoiesis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **27**, 333-341, doi:10.1096/fj.12-217547 (2013).
- 328 Loder, F. *et al.* B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals. *J Exp Med* **190**, 75-89 (1999).
- 329 Menzel, U. *et al.* Comprehensive evaluation and optimization of amplicon library preparation methods for high-throughput antibody sequencing. *PLoS One* **9**, e96727, doi:10.1371/journal.pone.0096727 (2014).

- 330 Ghislin, S., Ouzren-Zarhloul, N., Kaminski, S. & Frippiat, J. P. Hypergravity exposure during gestation modifies the TCRbeta repertoire of newborn mice. *Scientific reports* **5**, 9318, doi:10.1038/srep09318 (2015).
- 331 Schenten, V., Gueguinou, N., Baatout, S. & Frippiat, J. P. Modulation of Pleurodeles waltl DNA polymerase mu expression by extreme conditions encountered during spaceflight. *PLoS One* **8**, e69647, doi:10.1371/journal.pone.0069647 (2013).
- 332 Garcia, K. C. *et al.* CD8 enhances formation of stable T-cell receptor/MHC class I molecule complexes. *Nature* **384**, 577-581, doi:10.1038/384577a0 (1996).
- 333 Cancro, M. P. *et al.* B cells and aging: molecules and mechanisms. *Trends in immunology* **30**, 313-318, doi:10.1016/j.it.2009.04.005 (2009).
- 334 Fulcher, D. A. & Basten, A. B cell life span: a review. *Immunology and cell biology* **75**, 446-455, doi:10.1038/icb.1997.69 (1997).
- 335 Grove, D. S., Pishak, S. A. & Mastro, A. M. The effect of a 10-day space flight on the function, phenotype, and adhesion molecule expression of splenocytes and lymph node lymphocytes. *Exp Cell Res* **219**, 102-109 (1995).
- 336 Allebban, Z. *et al.* Effects of spaceflight on rat erythroid parameters. *J Appl Physiol* **81**, 117-122 (1996).
- 337 Wronski, T. J. *et al.* Lack of effect of spaceflight on bone mass and bone formation in group-housed rats. *J Appl Physiol* **85**, 279-285 (1998).
- 338 Blanc, S. *et al.* Counteraction of spaceflight-induced changes in the rat central serotonergic system by adrenalectomy and corticosteroid replacement. *Neurochem Int* **33**, 375-382 (1998).
- 339 Taylor, G. R. & Dardano, J. R. Human cellular immune responsiveness following space flight. *Aviation, space, and environmental medicine* **54**, S55-59 (1983).

- 340 Gould, C. L., Lyte, M., Williams, J., Mandel, A. D. & Sonnenfeld, G. Inhibited interferon-g but normal interleukin-3 production from rats flown on the space shuttle. *Aviation, Space, and Environmental Med* **58**, 983-986 (1987).
- 341 Cogoli, A., Bechler, B., Mueller, O. & Hunzinger, E. in *BioRack on Spacelab D1* Vol. ESA SP-1091 89-100 (European Space Agency, 1990).
- 342 Nash, P. V. *et al.* Effect of spaceflight on lymphocyte proliferation and interleukin-2 production. *J Appl Physiol* **73**, 186S-190S (1992).
- 343 Sonnenfeld, G. *et al.* Spaceflight alters immune cell function and distribution. *J Appl Physiol* **73**, 191S-195S (1992).
- 344 Fuchs, B. B. & Medvedev, A. E. Countermeasures for ameliorating in-flight immune dysfunction. *Journal of leukocyte biology* **54**, 245-252 (1993).
- 345 Sonnenfeld, G. Immune responses in space flight. *International Journal of Sports Medicine* **19**, S195-S202 (1998).
- 346 Sonnenfeld, G. *et al.* Spaceflight and development of immune responses. *J Appl Physiol* **85**, 1429-1433 (1998).
- 347 Manser, T., Huang, S. Y. & Gefer, M. L. Influence of clonal selection on the expression of immunoglobulin variable region genes. *Science* **226**, 1283-1288 (1984).
- 348 Manser, T., Wysocki, L. J., Gridley, T., Near, R. I. & Gefer, M. L. The molecular evolution of the immune response. *Immunology today* **6**, 94-101, doi:10.1016/0167-5699(85)90024-6 (1985).
- 349 Manser, T. The efficiency of antibody affinity maturation: can the rate of B-cell division be limiting? *Immunology today* **11**, 305-308 (1990).
- 350 Chevillard, C., Ozaki, J., Herring, C. D. & Riblet, R. A Three-Megabase Yeast Artificial Chromosome Contig Spanning the C57BL Mouse Igh Locus. *The Journal of Immunology* **168**, 5659-5666 (2002).

- 351 *Igh immunoglobulin heavy chain complex [Mus musculus (house mouse)] - Gene - NCBI*, <<https://www.ncbi.nlm.nih.gov/pubmed/>> (2017).
- 352 Kimzey, S. L. in *Biomedical Results from Skylab* Vol. NASA SP-377 (eds R. S. Johnston & L. F. Dietlein) 249-282 (NASA, 1977).
- 353 Stowe, R. P., Sams, C. F. & Pierson, D. L. Effects of mission duration on neuroimmune responses in astronauts. *Aviation, space, and environmental medicine* **74**, 1281-1284 (2003).
- 354 Atkinson, W., Wolfe, C. & Hamborsky, J. (Public Health Foundation, Washington, D.C., 2012).
- 355 Vogel, F. R. Improving Vaccine Performance with Adjuvants. *Clinical Infectious Diseases* **30**, S266-S270, doi:10.1086/313883 (2000).
- 356 Gavin, A. L. *et al.* Adjuvant-enhanced antibody responses in the absence of toll-like receptor signaling. *Science* **314**, 1936-1938, doi:10.1126/science.1135299 (2006).
- 357 Krieg, A. M. Lymphocyte activation by CpG dinucleotide motifs in prokaryotic DNA. *Trends in microbiology* **4**, 73-76, doi:10.1016/0966-842X(96)81515-0 (1996).
- 358 Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740-745, doi:10.1038/35047123 (2000).
- 359 Klinman, D. M. *et al.* Activation of the innate immune system by CpG oligodeoxynucleotides: immunoprotective activity and safety. *Springer seminars in immunopathology* **22**, 173-183 (2000).
- 360 Scheule, R. K. The role of CpG motifs in immunostimulation and gene therapy. *Advanced drug delivery reviews* **44**, 119-134 (2000).
- 361 Klinman, D. M. Immunotherapeutic uses of CpG oligodeoxynucleotides. *Nat Rev Immunol* **4**, 249-258, doi:10.1038/nri1329 (2004).
- 362 Steinhagen, F., Kinjo, T., Bode, C. & Klinman, D. M. TLR-based immune adjuvants. *Vaccine* **29**, 3341-3355, doi:10.1016/j.vaccine.2010.08.002 (2011).

- 363 Takeshita, F. *et al.* Cutting edge: Role of Toll-like receptor 9 in CpG DNA-induced activation of human cells. *Journal of immunology (Baltimore, Md. : 1950)* **167**, 3555-3558 (2001).
- 364 Hornung, V. *et al.* Quantitative expression of toll-like receptor 1-10 mRNA in cellular subsets of human peripheral blood mononuclear cells and sensitivity to CpG oligodeoxynucleotides. *Journal of immunology (Baltimore, Md. : 1950)* **168**, 4531-4537 (2002).
- 365 Diwan, M., Tafaghodi, M. & Samuel, J. Enhancement of immune responses by co-delivery of a CpG oligodeoxynucleotide and tetanus toxoid in biodegradable nanospheres. *Journal of controlled release : official journal of the Controlled Release Society* **85**, 247-262 (2002).
- 366 Bode, C., Zhao, G., Steinhagen, F., Kinjo, T. & Klinman, D. M. CpG DNA as a vaccine adjuvant. *Expert Rev Vaccines* **10**, 499-511, doi:10.1586/erv.10.174 (2011).
- 367 Coffman, R. L., Sher, A. & Seder, R. A. Vaccine adjuvants: putting innate immunity to work. *Immunity* **33**, 492-503, doi:10.1016/j.immuni.2010.10.002 (2010).
- 368 Cao, Y. *et al.* An optimized assay for the enumeration of antigen-specific memory B cells in different compartments of the human body. *J Immunol Methods* **358**, 56-65, doi:10.1016/j.jim.2010.03.009 (2010).
- 369 Morey-Holton, E. & Wronski, T. J. Animal models for simulating weightlessness. *The Physiologist* **24**, S45-S48 (1981).
- 370 GRCm38.p6 - Genome - Assembly - NCBI, <<https://www.ncbi.nlm.nih.gov/pubmed/>> (2017).
- 371 Maldonado, M. A. *et al.* The role of environmental antigens in the spontaneous development of autoimmunity in MRL-lpr mice. *Journal of immunology (Baltimore, Md. : 1950)* **162**, 6322-6330 (1999).

372 Bauer, K. *et al.* Diversification of Ig heavy chain genes in human preterm neonates prematurely exposed to environmental antigens. *Journal of immunology (Baltimore, Md. : 1950)* **169**, 1349-1356 (2002).

373 Halliley, J. L. *et al.* Long-Lived Plasma Cells Are Contained within the CD19(-) CD38(hi)CD138(+) Subset in Human Bone Marrow. *Immunity* **43**, 132-145, doi:10.1016/j.immuni.2015.06.016 (2015).

## Appendix A - Supplemental figures

### Chapter 3 Supplemental Figures

#### Appendix A.1. Comparison of V-gene segment usage with original and KSU/IMGT bioinformatics.

	KSU	Com1 (Orig)	Com1 (IMGT)	Com 2 (Orig)	Com 2 (IMGT)
IGHV1-80	6.2	6.1	2.9	6.3	6.5
IGHV2-6	3.7	1.0	0.2		0.0
IGHV2-6-8	3.7	4.5	0.2	1.9	2.0
IGHV1-26	3.4	0.0	3.0	6.2	6.4
IGHV1-9	3.0	4.9	3.2	5.0	5.2
IGHV1-53	2.8	4.0	4.4	6.4	6.7
IGHV9-3	2.7	2.8	2.6	0.9	0.9
IGHV1-18	2.6		2.9	2.1	2.2
IGHV1-55	2.4	2.0	2.3	4.2	4.3
IGHV1-50	2.4		1.1	2.0	2.0
IGHV6-3	2.4	5.7	3.3	1.7	1.8
IGHV3-6	2.0	3.1	2.7	4.9	5.1
IGHV1-64	1.9	2.8	3.5	4.6	4.7
IGHV1-76	1.9	3.1	1.2	4.0	4.2
IGHV6-6	1.9	3.3	1.9	1.7	1.8
IGHV4-1	1.8	0.0	0.6		0.0
IGHV1-82	1.7	2.0	3.0	1.7	1.8
IGHV1S95	1.7		0.0		
IGHV1S92	1.6				
IGHV1-81	1.6	2.1	2.3		0.0
IGHV1-72	1.5	2.3	3.1	2.3	2.4
IGHV8-8	1.5	2.5	1.0	0.7	0.7
IGHV2-2	1.5	2.0	2.3	0.7	0.7
IGHV5-17	1.5	3.0	2.9	0.6	0.6
IGHV1-22	1.5	1.9	2.3	2.4	2.5
IGHV1-69	1.5	1.0	1.0	3.4	3.5
IGHV1S108	1.5				
IGHV7-3	1.4	1.1	0.8	2.6	2.7
IGHV10-1	1.4	2.2	1.9	4.2	4.4
IGHV10-3	1.3	1.5	1.1	1.7	1.8
IGHV2-3	1.2	2.0	1.3	2.3	2.4
IGHV14-4	1.2	1.2	1.9	1.8	1.8
IGHV11-2	1.2	2.5	0.1	0.1	0.1
IGHV1-52	1.1		0.4		0.0
IGHV9-1	1.1	0.9	0.6	0.4	0.5
IGHV1-78	1.0	1.1	0.7	1.2	1.2
IGHV14-2	1.0	1.3	1.9	0.1	0.1
IGHV2-9	0.9	0.7	0.9	0.8	0.9
IGHV8-12	0.9	1.2	0.7	0.5	0.6
IGHV1-19	0.9	1.6	1.7	1.6	1.7
IGHV1-61	0.8		0.4	0.5	0.5
IGHV1-74	0.8	0.9	0.6	0.9	0.9
IGHV1-15	0.8	0.7	1.4		0.0
IGHV1-59	0.8	0.0	1.1	0.3	0.3

	KSU	Com1 (Orig)	Com1 (IMGT)	Com 2 (Orig)	Com 2 (IMGT)
IGHV5-4	0.8	1.1	1.3	1.2	1.3
IGHV1-7	0.8	0.7	0.9	1.1	1.1
IGHV2-9-1	0.8	0.6	1.2		0.0
IGHV1-39	0.8	1.7	2.0	0.3	0.4
IGHV1-75	0.7	1.3	1.4	1.0	1.1
IGHV5-16	0.7	0.9	0.5	0.4	0.4
IGHV1-4	0.6	0.4	0.4	1.0	1.0
IGHV9-2	0.6	0.5	0.4	0.2	0.2
IGHV3-8	0.6	0.7	0.5	0.1	0.1
IGHV2-5	0.6		0.6		
IGHV1-12	0.5	0.7	0.4	0.4	0.4
IGHV14-1	0.5	0.4	0.7	0.2	0.2
IGHV14-3	0.5	0.6	0.7	0.5	0.5
IGHV1-54	0.5	0.7	1.1	0.3	0.3
IGHV1-42	0.5	0.9	1.1	0.5	0.6
IGHV9-4	0.5	0.4	0.4	0.1	0.1
IGHV1-5	0.4	0.7	0.8	0.5	0.5
IGHV3-1	0.4	0.4	0.5		0.0
IGHV1-66	0.4	0.7	1.0	0.7	0.7
IGHV1-84	0.4	0.3	0.4	0.2	0.3
IGHV2-4	0.4	0.2	0.2	0.3	0.3
IGHV8-5	0.3	0.5	0.3	0.2	0.2
IGHV5-6	0.3	0.0	1.1	0.1	0.1
IGHV5-9-1	0.3		0.3	0.2	0.3
IGHV1-85	0.3	0.5	0.8	0.6	0.6
IGHV1-58	0.3	0.5	0.3	0.3	0.3
IGHV1-62-3	0.3	0.1	0.1		
IGHV3-5	0.3	0.1	0.1	0.2	0.2
IGHV1S96	0.3	0.0			
IGHV1-20	0.3	0.2	0.3	0.2	0.2
IGHV1-47	0.3	0.3	0.4		0.0
IGHV1S5	0.2	0.0	0.0		0.0
IGHV1-62-2	0.2	1.0	0.6		
IGHV1-71	0.2		0.6	0.1	0.1
IGHV1-63	0.2	0.4	0.5	0.2	0.2
IGHV5-15	0.2	0.4	0.1	0.5	0.6
IGHV12-3	0.2	0.2	0.1		0.0
IGHV13-2	0.2	0.3	0.2		0.0
IGHV1-11	0.2	0.3	0.1	0.1	0.1
IGHV5-9	0.2	0.4	0.4	0.2	0.2
IGHV1-36	0.2	0.3	0.4	0.2	0.2
IGHV1-34	0.2		3.0	0.1	0.1
IGHV1-77	0.2	0.3	0.4	0.2	0.2
IGHV5-12	0.2	0.3	0.3	0.2	0.2



	KSU	Com1 (Orig)	Com1 (IMGT)	Com 2 (Orig)	Com 2 (IMGT)
IGHV5S21	0.2		0.3		0.0
IGHV8S9	0.2				
IGHV1S87	0.1				
IGHV1-56	0.1	0.1	0.1	0	0.0
IGHV1-67	0.1	0.0	0.1		0.0
IGHV8-2	0.1	0.0	0.0		
IGHV5-2	0.1	0.1	0.2	0.1	0.1
IGHV11-1	0.1	0.0	0.0		0.0
IGHV1-49	0.1	0.1	0.1		0.0
IGHV1S103	0.1		0.0		
IGHV1S107	0.1				
IGHV1S100	0.1				
IGHV3-4	0.1	0.1	0.1		0.0
IGHV1S65	0.1				
IGHV1-31	0.1	0.1	0.2		0.0
IGHV15-2	0.1	0.1	0.1		0.0
IGHV8-11	0.1	0.0	0.0		0.0
IGHV3-3	0.1	0.0	0.0		0.0
IGHV7-4	0.1	0.1	0.1	0.2	0.2
IGHV2-7	0.1	0.1	0.0		
IGHV16-1	0.0				
IGHV1-43	0.0	0.0	0.0		0.0
IGHV1S67	0.0		0.0		
IGHV1S68	0.0				
IGHV7-2	0.0	0.0			
IGHV3S7	0.0		0.0		
IGHV1-23	0.0	0.0	0.2		
IGHV1-37	0.0	0.0	0.1		0.0
IGHV8-6	0.0	0.0	0.0		0.0
IGHV8-4	0.0	0.0	0.0		
IGHV6-7	0.0	0.0			
IGHV1-14	0.0	0.0	0.0		0.0
IGHV8-8-1	0.0	0.0			
IGHV6-4	0.0	0.0	0.0		0.0
IGHV8-13	0.0	0.0			
IGHV1-62-1	0.0				
IGHV5-12-4	0.0	0.0	0.0		
IGHV8-9	0.0	0.0			
IGHV5S24	0.0				
IGHV1-17-1	0.0				
IGHV6-5	0.0	0.0	0.0		
IGHV13-1	0.0				
IGHV1-24	0.0				

Com1 and Com2 V-gene segment usage with company specific bioinformatics (original) and after processing through the KSU bioinformatic pipeline (IMGT). The highest value percent of repertoire is dark read while the lowest are white. Black boxes represent no detected reads (true zero). Rounded zeros are represented as 0.0.

## Chapter 4 Supplemental Figures

### Appendix A.2. V gene segment rankings among the three mouse pools for both IgH (A) and Igk (B).

A

	Pool 1	Pool 2	Pool 3
V1-80*01	7.83	7.00	3.34
V1-26*01	4.29	3.96	3.86
V6-3*01	2.91	2.74	4.93
V1-53*01	3.80	2.96	2.73
V1-55*01	3.57	2.40	3.34
V3-6*01	2.68	2.28	3.81
V1-9*01	3.00	3.54	1.85
V1-18*01	4.05	3.09	1.21
V9-3*01	2.46	2.72	2.46
V1-50*01	1.43	2.32	3.89
V1-64*01	1.88	2.08	3.64
V6-6*01	2.72	2.22	1.51
V1-82*01	2.06	1.97	1.96
V8-8*01	2.22	1.74	1.94
V1-78*01	2.66	1.17	1.51
V1-81*01	1.76	1.77	1.24
V7-3*01	1.68	1.64	1.45
V1-72*01	1.32	1.41	1.90
V10-1*01	1.62	1.62	1.37
V1-22*01	1.50	1.77	1.24
V5-17*01	1.38	1.75	1.37
V2-2*01	1.16	1.75	1.40
V14-4*01	1.50	1.30	1.43
V1-76*01	0.87	2.22	1.11
V2-3*01	1.54	1.40	1.24
V4-1*01	1.09	2.05	1.02
V1S92*01	1.43	0.99	1.49
V11-2*01	1.59	1.39	0.86
V1S95*01	1.35	0.97	1.50
V1-69*01	0.86	1.40	1.40
V1-39*01	1.04	0.85	1.36
V1-52*01	0.77	0.98	1.46
V1-15*01	1.05	0.96	1.06
V10-3*01	0.78	1.47	0.80
V2-9*01	1.39	0.82	0.83
V1-75*01	1.30	0.72	0.89
V2-9-1*01	0.93	0.78	1.13
V1-19*01	0.74	1.07	1.01
V14-2*01	1.03	0.99	0.80
V8-12*01	1.03	1.00	0.78
V1-66*01	0.35	0.47	1.92
V1-59*01	0.91	0.53	1.20
V5-4*01	0.99	0.91	0.65
V5-16*01	0.89	0.83	0.74
V1-7*01	0.80	0.88	0.76
V1S108*01	0.13	0.74	1.57
V5-9-1*02	0.31	0.30	1.74
V1-74*01	0.45	0.62	1.18
V2-6*01	0.02	1.98	0.24
V2-6-8*01	0.02	1.98	0.23
V9-1*01	0.43	0.79	0.93
V1-63*01	1.69	0.23	0.17
V1-42*01	0.77	0.44	0.80
V9-4*01	1.06	0.52	0.41
V1-61*01	0.44	0.60	0.87
V3-1*01	0.92	0.44	0.51
V1-54*01	0.82	0.47	0.59
V14-3*01	0.60	0.58	0.68
V1-4*01	0.56	0.68	0.43
V2-5*01	0.39	0.67	0.54
V3-8*01	0.45	0.67	0.41
V1-12*01	0.54	0.62	0.34
V1-85*01	0.67	0.30	0.45

	Pool 1	Pool 2	Pool 3
V5-6*01	0.39	0.37	0.54
V1-5*01	0.25	0.51	0.41
V1-84*01	0.50	0.26	0.41
V8-5*01	0.43	0.32	0.39
V13-2*01	0.59	0.24	0.30
V14-1*01	0.32	0.52	0.28
V9-2*01	0.14	0.32	0.56
V5S21*01	0.07	0.11	0.70
V1-47*01	0.34	0.28	0.25
V1-58*01	0.35	0.29	0.21
V1-34*01	0.30	0.22	0.31
V3-5*01	0.27	0.33	0.21
V2-4*01	0.27	0.39	0.15
V1-20*01	0.19	0.31	0.24
V12-3*01	0.30	0.25	0.16
V5-12*01	0.17	0.17	0.32
V1-77*01	0.20	0.19	0.19
V1S96*01	0.01	0.18	0.36
V1-62-2*01	0.13	0.10	0.32
V1-71*01	0.13	0.10	0.32
V5-9*01	0.23	0.17	0.14
V1-36*01	0.16	0.23	0.15
V1-11*01	0.11	0.23	0.08
V1-62-3*01	0.02	0.02	0.39
V5-15*01	0.07	0.23	0.08
V5-2*01	0.15	0.12	0.10
V8-2*01	0.09	0.14	0.12
V1S5*01	0.00	0.04	0.29
V8S9*01	0.04	0.13	0.15
V1-67*01	0.07	0.14	0.10
V7-4*01	0.18	0.07	0.06
V3-4*01	0.14	0.08	0.08
V1S87*01	0.00	0.11	0.17
V1S103*01	0.00	0.06	0.21
V1S65*01	0.06	0.07	0.13
V15-2*01	0.07	0.07	0.10
V1-49*01	0.05	0.09	0.11
V1-31*01	0.08	0.07	0.08
V1-56*01	0.03	0.04	0.16
V1S107*01	0.04	0.07	0.10
V11-1*01	0.01	0.12	0.05
V3-3*01	0.09	0.07	0.03
V1S100*01	0.02	0.04	0.10
V1-43*01	0.04	0.05	0.06
V1-23*01	0.05	0.04	0.06
V8-11*01	0.01	0.05	0.08
V2-7*01	0.04	0.06	0.04
V1-37*01	0.06	0.04	0.03
V16-1*01	0.03	0.05	0.05
V6-7*01	0.04	0.02	0.05
V1S67*01	0.01	0.03	0.07
V1S68*01	0.01	0.03	0.06
V7-2*01	0.03	0.04	0.03
V3S7*01	0.01	0.03	0.04
V8-6*01	0.00	0.02	0.05
V8-4*01	0.02	0.03	0.02
V6-4*01	0.01	0.02	0.03
V8-9*01	0.02	0.01	0.02
V1-62-1*01	0.00	0.01	0.04
V1-14*01	0.01	0.02	0.02
V8-8-1*01	0.01	0.02	0.02
V5-12-4*01	0.01	0.01	0.02
V8-13*01	0.00	0.00	0.04

	Pool 1	Pool 2	Pool 3
V5-9*04	0.00	0.01	0.02
V6-5*01	0.01	0.01	0.01
V13-1*01	0.01	0.01	0.01
V1-17-1*01	0.01	0.01	0.00
V5S24*01	0.01	0.01	0.00
V1-24*01	0.00	0.00	0.00

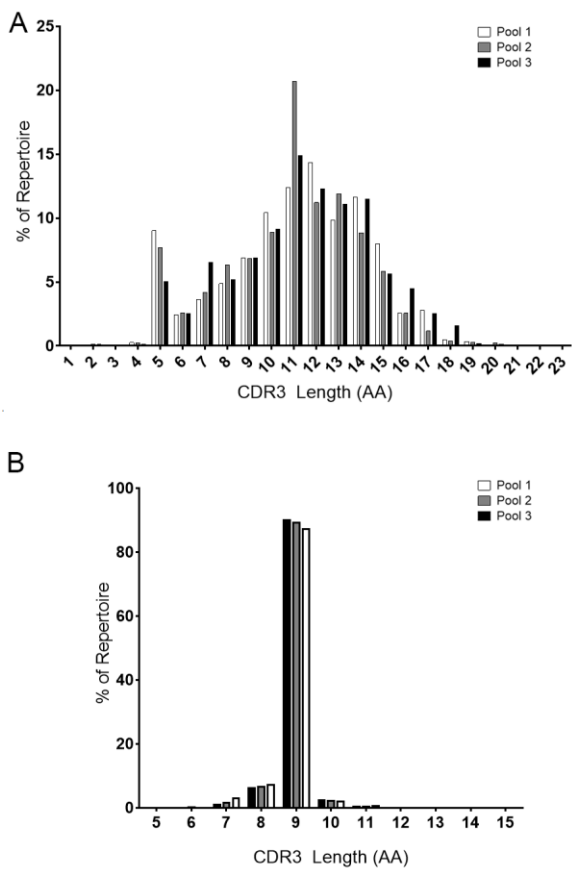
B

	Pool 1	Pool 2	Pool 3
V1-110*01	4.80	13.44	6.15
V1-117*01	8.13	5.30	7.09
V5-39*01	10.33	1.03	2.84
V10-96*01	2.69	5.04	4.77
V4-55*01	4.20	3.87	3.62
V1-135*01	3.62	2.50	5.33
V3-4*01	3.40	2.11	2.71
V5-43*01	3.25	1.71	1.53
V10-94*01	0.97	3.45	2.04
V6-15*01	1.50	1.84	3.00
V14-111*01	2.02	2.28	1.97
V12-44*01	1.67	3.09	1.41
V3-2*01	2.24	1.23	2.69
V19-93*01	2.17	2.36	1.63
V2-137*01	1.99	1.44	2.50
V12-46*01	2.22	2.33	1.35
V8-30*01	1.60	1.85	2.40
V4-59*01	1.92	1.59	1.98
V8-24*01	2.25	1.57	1.35
V9-120*01	1.76	1.58	1.70
V4-72*01	2.05	1.08	1.21
V6-23*01	1.04	1.68	1.59
V6-17*01	1.44	1.55	1.28
V14-126*01	1.63	1.38	1.22
V4-70*01	0.65	1.12	2.37
V15-103*01	1.02	1.87	1.23
V8-27*01	1.39	1.33	1.09
V4-57*01	2.20	0.65	0.78
V5-48*01	1.10	0.92	1.21
V16-104*01	1.02	0.96	1.15
V2-109*01	0.99	0.94	0.90
V17-127*01	1.06	0.82	0.92
V6-32*01	0.72	1.06	0.92
V4-53*01	1.01	1.02	0.63
V8-21*01	0.38	0.61	1.66
V8-19*01	0.86	0.92	0.81
V6-25*01	0.73	0.63	1.23
V17-121*01	0.97	0.68	0.80
V4-68*01	0.58	1.04	0.82
V9-124*01	0.83	0.95	0.62
V4-86*01	0.53	1.07	0.80
V12-89*01	0.96	0.65	0.75
V12-41*01	0.66	0.58	1.06
V4-91*01	0.68	0.58	0.78
V3-5*01	0.65	0.54	0.80
V4-63*01	0.45	1.30	0.22
V3-12*01	0.41	0.78	0.74
V3-10*01	0.54	0.55	0.81
V2-112*01	0.40	1.05	0.40
V8-28*01	0.68	0.49	0.51
V6-20*01	0.53	0.72	0.38
V4-57-1*01	0.41	0.57	0.54
V14-100*01	0.54	0.51	0.46
V6-13*01	0.30	0.26	0.81
V1-122*01	0.51	0.32	0.50
V5-45*01	0.60	0.31	0.39
V3-7*01	0.27	0.55	0.38
V4-61*01	0.61	0.29	0.31
V13-85*01	0.28	0.36	0.50
V8-16*01	0.28	0.49	0.32
V1-99*01	0.61	0.22	0.25
V4-79*01	0.20	0.23	0.64
V13-84*01	0.29	0.41	0.36

	Pool 1	Pool 2	Pool 3
V12-98*01	0.23	0.47	0.29
V4-71*01	0.03	0.39	0.55
V4-74*01	0.34	0.33	0.28
V4-80*01	0.23	0.32	0.36
V6-14*01	0.38	0.22	0.29
V4-50*01	0.32	0.26	0.30
V1-88*01	0.26	0.35	0.25
V7-33*01	0.09	0.25	0.28
V4-73*01	0.03	0.28	0.31
V11-125*01	0.10	0.23	0.26
V6-32*02	0.11	0.27	0.20
V4-58*01	0.15	0.19	0.21
V1-133*01	0.14	0.12	0.26
V4-69*01	0.11	0.14	0.27
V10-95*01	0.12	0.17	0.16
V9-129*01	0.11	0.20	0.14
V4-78*01	0.05	0.21	0.18
V3-1*01	0.09	0.17	0.18
V8-18*01	0.09	0.07	0.22
V12-38*01	0.11	0.10	0.15
V4-90*01	0.14	0.10	0.09
V8-23-1*01	0.08	0.13	0.10
V4-62*01	0.04	0.12	0.14
V8-34*01	0.05	0.12	0.12
V12-e*01	0.02	0.14	0.12
V4-81*01	0.09	0.11	0.08
V9-123*01	0.10	0.08	0.09
V14-130*01	0.08	0.08	0.07
V3-3*01	0.05	0.07	0.08
V4-51*01	0.06	0.07	0.06
V3-9*01	0.02	0.06	0.09
V1-131*01	0.01	0.06	0.10
V2-a*01	0.02	0.08	0.07
V4-52*01	0.03	0.05	0.08
V4-54*01	0.03	0.05	0.08
V4-92*01	0.07	0.05	0.04
V6-29*01	0.04	0.07	0.03
V1-132*01	0.05	0.03	0.06
V5-37*01	0.02	0.05	0.05
V18-36*01	0.04	0.04	0.03
V8-26*01	0.01	0.03	0.03
V20-101-2*01	0.01	0.01	0.02
V6-c*01	0.00	0.02	0.01
V6-d*01	0.01	0.01	0.00
V1-35*01	0.00	0.01	0.01
V6-b*01	0.00	0.01	0.01

The most abundant gene segment is ranked as one. Dark red indicates higher rank moving to blue, of lower rank.

Appendix A.3. CDR3 length for IgH (A) and Igκ (B) by pool.



The percent of repertoire CDR3 lengths from each mouse pools are displayed.

**Appendix A.4. Rankings of CDR3 sequences shared by all three mouse pools were uniform in both IgH (A) and Igκ (B).**

**A**

	Pool 1	Pool 2	Pool 3
CARGAYW	1	2	3
CARDYYGSSWYFDVW	18	6	6
CMRYSNYWYFDVW	4	11	15
CARGGYW	16	5	18
CARGTYW	45	18	7
CMRYGNYWYFDVW	37	15	23
CARGYYFDYW	45	9	22
CARGDYW	9	26	45
CARDSNWYFDVW	39	34	19
CARGPYW	13	158	34
CARDNWDWYFDVW	189	20	55
CMRYSSYWYFDVW	189	14	62
CARW	125	73	170
CARGGFAYW	97	158	137
CAKKGAMDYW	97	79	299
CARDYYGSSYWYFDVW	266	192	55
CARRLDYW	189	228	170
CMRYGSSYWYFDVW	266	26	299
CTTVRYW	125	301	170
CARPYDYW	28	301	299
CAQMRGFAYW	125	464	40
CARRFDYW	413	228	96
CMRYGSSWYFDVW	125	192	449
CARDDGYWYFDVW	189	301	299
CARRYGSSYWYFDVW	125	464	219
CTTLRYW	266	112	449
CAKNWDYW	189	464	219
CARDYDYWYFDVW	413	192	299
CARRDYGSSYWYFDVW	24	918	52
CARYGPYFDYW	62	31	933
CARIYYGSSYWYFDVW	266	464	299
CARDEWFAYW	760	79	299
CARHYGSSYWYFDVW	760	84	299
CARDWDYWYFDVW	266	464	449
CARGDGFYW	97	192	933
CARHYGSSYWYFDVW	266	63	933
CARGYYW	760	57	449
CAKGDYGSSWFAYW	266	130	933
CARLYYGSSYWYFDVW	97	301	933
CTRGYYFDYW	760	158	449
CARGDGYFDYW	760	464	219
CARGGDYW	760	464	219
CARGAMDYW	189	464	933
CARYDGYFDYW	413	301	933
CTVYGGSTWFAYW	413	301	933
CARRDYW	760	464	449
CARSYYFDYW	760	464	449
CAREGDYDWDYFDVW	760	918	52
CARDYYGSSGYFDVW	17	918	933
CTRVAYW	760	192	933
CTRWDYW	760	192	933
CARYAMDYW	760	228	933
CARDYYGSSFDYW	760	918	299
CARYSNYFDYW	760	918	299
CANYGSSYWYFDVW	760	301	933
CARSLDYW	760	301	933
CASELGGFAYW	760	301	933
CASPNDWYFDVW	760	301	933
CAREDYW	189	918	933
CARDGFAYW	266	918	933
CARKLDYW	760	464	933
CARNWDYAMDYW	760	464	933
CARYYYGSSYWYFDVW	760	464	933

	Pool 1	Pool 2	Pool 3
CARSYYFDYW	760	464	449
CAREGDYDWDYFDVW	760	918	52
CARDYYGSSGYFDVW	17	918	933
CTRVAYW	760	192	933
CTRWDYW	760	192	933
CARYAMDYW	760	192	933
CARDYYGSSFDYW	760	918	299
CARYSNYFDYW	760	918	299
CANYGSSYWYFDVW	760	301	933
CARSLDYW	760	301	933
CASELGGFAYW	760	301	933
CASPNDWYFDVW	760	301	933
CAREDYW	189	918	933
CARDGFAYW	266	918	933
CARKLDYW	760	464	933
CARNWDYAMDYW	760	464	933
CARYYYGSSYWYFDVW	760	464	933
CARSGTDYW	413	918	933
CARDFDYW	760	918	933
CARDGSSYWYFDVW	760	918	933
CARDYFDYW	760	918	933
CARESNNYFDYW	760	918	933
CARRDWDYFDVW	760	918	933
CARSYYYAMDYW	760	918	933
CARYGNYAMDYW	760	918	933
CARYSNYAMDYW	760	918	933
CATGFAYW	760	918	933
CTGLYYFDYW	760	918	933
CTYYYGSSDFDYW	760	918	933



B

	Pool 1	Pool 2	Pool 3
CQWSSYPLTF	13	6	2
CQWSSNPPTF	2	18	7
CQWSSYPPTF	6	3	24
CQHYSTPLTF	7	12	16
CQGNLTPWTF	21	7	12
CQSNEDPRTF	11	21	10
CFQGSHPWTF	17	12	18
CFQGSHPYTF	16	9	22
CQHYSTPYTF	24	14	13
CQYSSYPLTF	26	15	11
CQYNSYPLTF	33	15	5
CQSNSWPLTF	20	26	8
CQNGHSFPLTF	1	55	1
CQSNEDPYTF	14	25	25
CLQHGSPYTF	19	17	29
CLQYDEFYTF	37	10	20
CSQSTHVPYTF	24	23	21
CLQYASSPYTF	21	24	26
CFQGSHPVPLTF	5	30	38
CLQYDNLWTF	18	18	37
CSQSTHVPWTF	23	20	31
CQYNSYPYTF	45	22	16
CMQHLEYPYTF	43	40	9
CQWSSNPLTF	39	26	28
CQHYSTPWTF	35	33	32
CQSNSWPYTF	29	46	27
CQWSSYPPTF	28	40	34
CQGNLTPYTF	27	34	42
CQSNSWPPTF	3	89	30
CQWNYPLTF	94	8	22
CMQHLEYPPTF	30	43	59
CQYSKLPWTF	82	4	52
CLQYDNLVTF	69	38	33
CQGNLTPRTF	34	44	63
CLQSDNMLTF	45	53	44
CSQSTHVPPTF	77	1	65
CQHYSTPRTF	71	34	39
CQSNEDPWTF	82	51	15
CLQHGSPPTF	48	31	73
CWQGTHTFPWTF	77	40	35
CQSKVPWTF	102	55	4
CWQGTHTFPYTF	52	71	39
CWQGTHTFPRTF	43	67	54
CHQYLSWTF	47	48	70
CQHHYGTPTYTF	60	53	53
CQNGHSFPYTF	15	94	57
CQGGQSYPLTF	87	38	41
CQWSSNPYTF	73	37	59
CWQGTHTFPQTF	58	64	48
CQHHYGTPLTF	39	73	59
CQYSSYPWTF	56	103	14
CFQGSYPLTF	66	5	103
CFQGSHPPTF	58	47	70
CHQYLSYTF	35	62	78
CQHYSTPPTF	66	61	48
CLQYASSPWTF	52	59	65
CQNDYSYPLTF	52	73	54
CQWSSNPPTF	50	55	77
CQHFWGTPYTF	48	70	68
CQGGQSYPWTF	73	32	83
CQYSGYPLTF	97	48	45
CQSKVPRTF	7	51	138
CMQHLEYPPTF	52	72	74

	Pool 1	Pool 2	Pool 3
CQSNEDPPTF	30	120	48
CLQSDNMPYTF	60	76	69
CQHFWGTPWTF	41	69	97
CQQRSSYPLTF	69	96	42
CQWSSYPYTF	38	45	124
CFQGSHPRTF	30	82	97
CLQSDNLPLTF	68	79	63
CSQSTHVPPTF	60	96	67
CQGNLTPPTF	87	36	103
CQHFWGTPPTF	122	10	100
CAQNLPLWTF	91	50	93
CQQHNEYPWTF	64	86	89
CQHHYGTPTF	94	2	144
CLQYASYPRTF	50	78	122
CQQHNEYPPTF	84	76	93
CQYSSYPYTF	97	101	62
CQYNSYPPTF	131	91	46
CQYSSYPLTF	56	93	130
CLQSDNLPYTF	97	103	80
CLQYDEFPLTF	106	82	93
CQYSSYPYTF	128	73	80
CQQHNEYPYTF	110	87	85
CQYSKLPYTF	108	100	74
CFQGSHPPTF	102	103	78
CQDYSSPYTF	136	67	80
CQQRSSYPPTF	12	143	130
CLQYASSPPTF	87	91	112
CQNGHSFPPTF	10	198	90
CSQSTHVPPTF	91	110	97
CQSKVPYTF	84	112	103
CQSNEDPLTF	102	131	70
CHQRSSWTF	175	126	3
CQSNSWPHTF	117	64	124
CLQYASSPLTF	128	103	83
CQSNSWPPTF	122	111	85
CWQGTHTFPPTF	150	58	113
CGQSYSPYTF	106	120	103
CHQYLSRPTF	94	112	130
CQHYSTPPTF	122	123	92
CQGGQSYPYTF	144	63	138
CQHFWGTPRTF	79	138	130
CQHHYGTPTWTF	117	131	100
CWQGTHTFPHTF	91	138	122
CQNDHSYPYTF	73	163	117
CSQSTHVPRTF	110	126	117
CWQGTHTFPLTF	84	159	110
CQGGSSPLTF	128	112	116
CQHFWSTPWTF	112	188	57
CQHHYGTPTF	117	131	113
CQYWPYTF	144	129	88
CQYSSYPPTF	175	143	46
CAQNLPLPYTF	112	138	117
CQNVLPYTF	64	143	160
CQNDHSYPLTF	136	109	124
CQGGSSPRTF	97	191	93
CQDYSSPLTF	144	108	130
CQHFWGTPPTF	102	89	196
CVQYQFPYTF	131	112	144
CQHFWGTPPTF	117	96	182
CQNGHSFPRTF	4	224	167
CHQYLSLTF	158	131	107
CQHSRELPLTF	225	87	85
CQDYSSPWTF	136	149	117

B

	Pool 1	Pool 2	Pool 3
CQONNEDPWTF	136	138	128
CQNVLPSTPFTF	87	182	138
CQNVLPSTPWTF	153	149	107
CLOYDEFWPFTF	169	82	160
CLOYDEFPTF	79	170	163
CLOYDNLRTF	144	103	167
CQHFWSPTPYTF	131	182	102
CQQLVEYPFTF	163	29	223
CQOWSGYPFTF	136	129	150
CQOSKEVPPTF	175	59	188
CLQSDNMPFTF	136	151	138
CQOWSSNPWTF	122	198	111
CQONNEDPFTF	188	198	51
CQHSRELPYTF	158	151	144
CLQSDNLPFTF	153	166	138
CQQHLHIPYTF	153	131	176
CQNGHSFPWTF	9	214	238
CHQRSSYPWTF	131	182	150
COQGNTLPLTF	122	147	196
CLOYASYPYTF	158	159	150
CAQFYSYPLTF	214	124	130
CQNGHSFPPTF	71	239	158
CFQSNYLPYTF	73	179	223
CQNDYSYPYTF	169	96	213
CQHHYGTPTF	188	131	160
CQONNEDPYTF	175	166	138
CSQSTHVPPWTF	188	163	128
CQOSNEDPPTF	150	159	171
CQHFWSVTPRTF	158	217	107
CQQRSSYPPTF	158	170	163
CQQLVEYPYTF	243	119	130
CAQNLELWTF	243	138	113
CKQSYNLWTF	271	217	6
CQQGSSIPFTF	169	170	155
CLQHWNYPLTF	163	179	155
CQQYWSTPLTF	203	170	124
CGQSYSPYTF	214	170	117
CLOYDNLLTF	108	198	196
CQONNEDPRTF	153	126	223
CLOYDNLLYTF	203	112	188
CQQYHSYPLTF	41	224	238
CQNDYSYPFTF	150	198	158
CLOYDNLLWTF	214	82	213
CLOYASYPWTF	163	198	150
CQQYYSYPRTF	112	198	203
CHQWSSYPLTF	243	239	36
CQQYWSTPWTF	169	154	196
CLQTHQWPWTF	175	143	203
CLQVTHVPYTF	175	159	188
CVQYAQFPWTF	188	163	176
CLOYDNLFTE	79	214	238
CQQLYSTPYTF	243	112	182
CQQRSSYPYTF	188	191	163
CSQSTHVPPYTF	163	182	203
CLOYASSPPTF	175	208	167
CQHSWEIPLTF	243	95	213
CLOYASSPRTF	188	182	182
CHQRSSYTF	188	28	339
CQOGNTLPFTF	188	188	182
CQQHNEYPTF	243	264	54
CQHSWEIPYTF	225	166	182
COQNSWPRTF	122	191	261
CLQHGESPWTF	112	217	251

	Pool 1	Pool 2	Pool 3
CQOFTSSPYTF	203	182	196
CQQYYSYPFTF	336	101	144
CQOWSSNPPIF	203	217	171
CQQYSKLPFTF	112	271	223
CSQSTHVPTF	136	252	223
CQQYSKLPRTF	304	239	74
CQHSRELPFTF	203	239	176
CGQSYSPFTF	304	64	261
CQHSRELPWTF	304	124	203
CQOGQSYPTF	243	112	278
CQQDYSSPFTF	188	224	223
CFQSNYLPFTF	188	252	203
CLOYDNLLRTF	203	224	223
CQQLYSTPLTF	304	170	188
CHQRSSYPYTF	188	252	223
CHQWSSYPTF	243	170	261
CQQYSSYPTF	214	285	176
CQQYWSTPFTF	271	170	238
CVQYAQFPRTF	243	239	203
CLOYDEFPRTF	225	252	213
CQONNEDPLTF	336	191	163
CLOYASYPFTF	336	80	278
CQQYSKLPFTF	243	191	261
CKQSYNLYTF	243	285	171
CQQLVEYPLTF	225	198	278
CKQSYNLFTF	271	120	312
COQNSWPPWTF	271	214	223
CAQNLELPRTF	214	285	213
CAQNLELPLTF	304	208	203
CHQWSSYPYTF	225	321	171
CQQLVEYPRTF	188	321	213
CQQLYSTPWTF	225	239	261
CHQYHRSPLTF	175	239	312
CQHSRELPPTF	374	131	223
CQQYYSYPPTF	225	233	278
CKQSYNLLTF	271	252	223
CQOWSSNPRTF	442	80	238
CQHFWSVTPRTF	188	239	339
CQOGSSIPYTF	203	224	339
CQQDYSSPPTF	225	264	278
CLOYDNLLFTF	336	154	278
CQNVLPSTPPTF	374	264	130
CQHSWEIPWTF	225	285	261
CLOYDEFPTF	243	217	339
CQOWSSFPTF	136	147	518
CQOFTSSPWTF	243	321	238
CSQSTHVPLTF	304	208	290
CQQRKVPWTF	336	208	261
CQQYSSYPWTF	225	208	373
CLQHWNYPPFTF	117	387	312
CQQYSKLPFTF	271	233	312
CQQYSGYPYTF	243	285	290
CQQSYAPLTF	304	264	251
CLOYDEFPPYTF	163	321	339
CQQYHSYPPTF	271	321	238
CHQWSSYPPTF	304	271	261
CQQYWSTPPTF	304	271	261
CQNVLPSTPRTF	175	239	425
CFQGSYPYTF	271	233	339
CHQWSSYPWTF	336	271	238
CQHFWDTPRTF	214	321	312
CQQDYSSPRTF	225	285	339
CSQSTHVPPFTF	374	239	238



B

	Pool 1	Pool 2	Pool 3
CHQRSSYPPTF	203	427	223
CQOYWSTPRTF	243	321	290
CVQYAQFPLTF	271	271	312
CQOYNSYPRTF	374	285	196
CQNVLSTPLTF	271	252	339
CKQAYDVPYTF	304	299	261
CQQWSSFPYTF	203	154	518
CQOYSSYPRTF	243	321	312
CLQVTHVPLTF	131	472	278
CHQYHRSPPTF	225	285	373
CQOWNSYPPTF	271	472	144
CSQSTHVWTF	336	299	261
CQQGNLTRTF	374	387	150
CLQHSYLPYTF	304	357	251
CQQWSGYPTYF	243	357	312
CQQSKEVPPTF	574	321	19
CHQYHRSPYTF	243	299	373
CQNDHSYPPTF	336	357	223
CHQYHRSPWTF	304	191	425
CQQGNLTLYTF	271	224	425
CLQVTHVPWTF	336	299	290
CQOWTYPLTF	336	217	373
CQOYNSYPWTF	442	299	188
CQHFWSPTPTF	271	321	339
CHQRSSFTF	188	321	425
CQQWSNYPYTF	442	154	339
CQHFWSPTLTF	336	387	213
CVQYAQFPPTF	336	387	213
CHQYHRSPPTF	214	299	425
CHQWSSYRTF	374	239	339
CQQWSSYPRTF	271	357	339
CQQWSSYPWTF	336	321	312
CALWYSNHLVF	304	357	312
CQQWSSSPPTF	243	357	373
CQQSKEVPLTF	442	357	176
CFQGSHPVPTF	304	252	425
CQOYSGYPPTF	304	387	290
CLQRNAYPLTF	374	321	290
CQQRSSYPWTF	442	299	251
CQQNSWQPQYTF	271	357	373
CSQSTHVYTF	271	357	373
CQOYSSSPLTF	374	208	425
CLQHWNYPYTF	175	321	518
CQOFTSSPSTF	203	472	339
CAQNLPLPTF	442	387	188
CQQNNEDPPTF	442	264	312
CQQLYSTPRTF	442	299	278
CQQGNLTWTF	574	271	176
CQQNSWQPQTF	374	357	290
CHQWSSYPPTF	175	427	425
CLQHGESPPTF	442	357	238
CQOYHSYPRTF	374	427	238
CLQFYEFPLTF	374	299	373
CLQYASYPLTF	442	321	290
CQQRKVPYTF	442	321	290
CQQWSGYPLTF	442	299	312
CLQGTQHPYTF	574	285	196
CQQNSWPTLTF	304	387	373
CQOYSGYPWTF	336	357	373
CQQHYSTPTF	374	321	373
CQQDYSSYTF	243	321	518
CQOYNNYPLTF	574	271	238
CQOFTSSPPTF	574	321	203

	Pool 1	Pool 2	Pool 3
CQOYSSYRTF	163	427	518
CHQYLSLFTF	442	299	373
CQQWSSYPITF	243	357	518
CQHSRELPTF	442	472	213
CGQSYSPPTF	225	387	518
CAQNLPLPTF	574	321	238
CQQWSSSPLTF	374	387	373
CQHSWEIPPTF	574	252	312
CHQRSSYPLTF	374	427	339
CLQRNAYPYTF	243	472	425
CLQGTQHPPTF	442	387	312
CQQWSNYPPTF	225	233	684
CQHSWEIPPTF	442	427	290
CQHFWNTPYTF	442	472	251
CFQGSHPVPTF	442	299	425
CLQYDELYTF	243	561	373
CQQNSWPF	336	472	373
CQQNSWQPQTF	336	472	373
CQOYHSYPYTF	442	427	312
CHQYLSSETF	574	357	251
CSQSTHVLTF	574	271	339
CQOQSYSPRTF	442	321	425
CQOYSSYLTf	442	387	373
CHQRSSYPCTF	304	561	339
CFQGSGYPTYF	374	151	684
CVQGTHTFPYTF	374	321	518
CQQWSSYPPTF	442	357	425
CWQGTHERTF	442	357	425
CQQWNYPYTF	271	271	684
CALWYSNHWVF	304	239	684
CQHFWSPTPTF	336	387	518
CQQRSSYPRTF	574	299	373
CQOYHSYPWTF	169	561	518
CQQWSSSPPTF	442	472	339
CQQWNYPLYTF	442	387	425
CQQNSWLTF	271	472	518
CGQSYSYPTF	574	321	373
CAQLLELPYTF	336	561	373
CQQLYSTPPTF	374	472	425
CLQGTQHPPTF	304	285	684
CQQGNLTLPWTF	304	285	684
CLQFYEFPYTF	374	387	518
CQQWSSNPPYTF	442	427	425
CQQWSSNPPTWTF	574	357	373
CFQSNYLPPTF	574	427	312
CQOYNSYLTf	442	357	518
CHQYLSSTF	374	427	518
CQOYNSYPLTF	574	387	373
CQQWSSDPPTF	225	427	684
CHQYLSYTF	442	472	425
CQOYSSSPLTF	271	729	339
CQOYSTYPYTF	574	427	339
CQOYSSYYTF	442	387	518
CQOYSSYPTF	442	387	518
CLQVTHVPPTF	374	561	425
CQOHLHIPWTF	271	729	373
CVQGTHTFPPTF	574	427	373
CVQGTHTFPPTF	442	561	373
CKQSYNLPTF	574	285	518
CQNGHSFPF	225	729	425
CQOYSGYPLTF	304	561	518
CLQVTHVPPTF	574	472	339
CQOYSSYPLTF	574	472	339



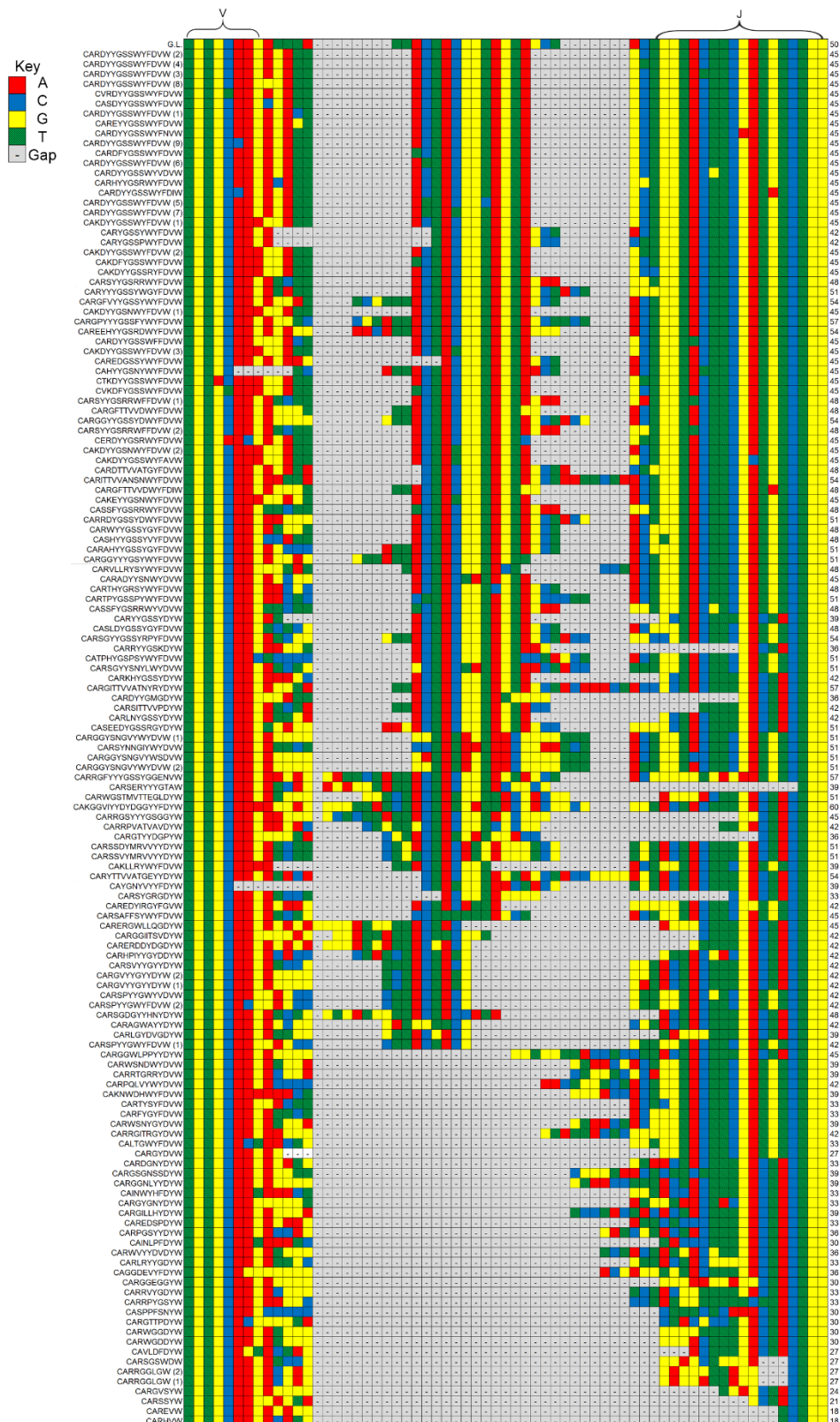
B

	Pool 1	Pool 2	Pool 3
CQQYYSYLT	574	387	425
CQWSTYPFT	442	264	684
CWQGTHTPF	336	387	684
CQWSSNPPMYT	304	427	684
CHQRSSYPT	574	472	373
CKQSYNLR	442	299	684
CLOHSYLP	574	427	425
CAQNLPL	442	472	518
CQYSKLW	442	472	518
CKQAYDVLP	374	387	684
CHQYHRSP	442	321	684
CQWSSDPYT	574	198	684
CSQSTHIPW	304	729	425
CLOYDNLLPW	574	729	167
CQHSWEIP	574	472	425
CQNDHSPW	574	561	339
CVQGTHTPL	574	561	339
CQYQNKLPW	442	357	684
CQCGSSIP	374	427	684
CQYNSYPHT	336	729	425
CQQRKVPST	574	561	373
CLOYDNLLT	442	729	339
CQYYTYPW	442	561	518
CQYYSYRT	374	729	425
CFQGSHPV	374	472	684
CLOYASSLT	374	472	684
CQYTSYPL	574	285	684
CQHLHIPPT	304	561	684
CHQRSSYP	442	427	684
CSQSTHVP	574	299	684
CCQGSHPVLP	574	561	425
CQFTSSPL	574	561	425
CQSYSAPT	574	561	425
CVQYQFPPT	574	561	425
CHOWSSYHT	574	472	518
CLOYDSLW	574	472	518
CQGNLPLPL	574	472	518
CQWNYPLT	574	321	684
CWQGTHTPT	336	729	518
CHQYHRSPPT	442	729	425
CQQRSSYP	442	729	425
CQWNNYPLT	442	729	425
CSQSTHVP	442	729	425
CQCGSSSPYT	442	472	684
CQWSSNPQYT	442	472	684
CQYYSYP	442	472	684
CQHYSTP	574	729	312
CQYNSYP	374	561	684
CQNDYSPPT	374	729	518
CQSNEDPPW	374	729	518
CLOYDEF	214	729	684
CQWNYPT	225	729	684
CFQGSHPVST	574	561	518
CLOYDNLTW	574	561	518
CMQHLECPYT	574	561	518
CQYNSYPLYT	574	561	518
CQHYSTW	574	729	373
CQYNSYPLA	574	729	373
CQYSGYPT	574	729	373
CLOYASSPYMYT	442	561	684
CQWNSNPYT	442	561	684
CHQYLSST	574	729	425
CLOYDNLLPT	574	729	425

	Pool 1	Pool 2	Pool 3
CQWSSYPLT	574	729	425
CALWYSNHFVV	574	472	684
CQHLHIPPT	574	472	684
CSQSTHVPW	374	729	684
CGQSYSPRT	574	561	684
CLOHGESP	574	561	684
CLOSDNMYT	574	561	684
CLOYDEFMYT	574	561	684
CQWSSNPPW	574	561	684
CFQGSHPW	574	729	518
CHQYLSSTP	574	729	518
CLOYASSPT	574	729	518
CMQOLEYPYT	574	729	518
CQQRKEVPW	574	729	518
CQQRSSYPPLT	574	729	518
CQYNSYTYT	574	729	518
CQYSKLTW	574	729	518
CFQSNYLPYT	442	729	684
CLOYDEFPLT	442	729	684
CQHHYGTW	442	729	684
CQYNTYPLT	442	729	684
CCQGSHPVYT	574	729	684
CLHYDEFW	574	729	684
CLOGTYYPRT	574	729	684
CQNDHSPRT	574	729	684
CQHLHIPPLT	574	729	684
CQHYSTPHT	574	729	684
CQSIEDPYT	574	729	684
CQSNEDRT	574	729	684
CQNSWPAL	574	729	684
CQSYEDPW	574	729	684
CQWSSSPIT	574	729	684
CQYWSTALT	574	729	684
CQYWSTMYT	574	729	684

The most abundant CDR3 sequence is ranked as one. Dark red indicates higher rank moving to blue, of lower rank.

# Appendix A.5. Full CDR3 nucleotide alignment of IgH gene combination examined in Figure 7 (IGHV1-26, IGHD1-1, IGHJ1).



## Chapter 5 Supplemental Figures

### Appendix A.6. V-Gene Segment Heat maps

A

	G1	G2	G3	F1	F2	F3		Avg G	Avg F	Avg T		G1	G2	G3	F1	F2	F3	
V1-53*01	12.78	8.74	23.71	10.15	4.99	8.83		15.08	7.99	11.53		V1-53*01	1	2	1	2	2	1
V9-3*01	10.07	22.08	5.81	2.79	3.10	7.88		12.65	4.59	8.62		V9-3*01	2	1	2	8	6	2
V1-26*01	3.87	3.20	3.88	9.70	3.32	5.63		3.65	6.22	4.93		V1-26*01	4	5	4	3	5	4
V3-6*01	3.24	5.20	4.58	11.29	1.86	2.07		4.34	5.07	4.71		V3-6*01	6	3	3	1	14	12
V1-78*01	0.72	0.73	0.90	0.69	16.63	1.75		0.78	6.36	3.57		V1-78*01	36	31	28	34	1	17
V6-3*01	3.40	2.35	2.13	1.49	4.30	7.48		2.63	4.43	3.53		V6-3*01	5	9	11	13	3	3
V5-4*01	0.96	1.10	1.38	5.45	1.30	1.26		1.15	2.67	1.91		V5-4*01	30	21	16	5	20	19
V1-15*01	0.25	1.04	0.90	6.46	1.09	0.68		0.73	2.75	1.74		V1-15*01	64	23	27	4	29	36
V1-19*01	5.93	0.42	0.71	0.42	0.81	0.83		2.35	0.69	1.52		V1-19*01	3	52	37	47	36	31
V1-55*01	2.53	3.64	2.71	2.96	1.84	4.76		2.96	3.19	3.07		V1-55*01	10	4	7	7	15	5
V1-82*01	1.63	1.63	2.83	2.44	2.88	2.83		2.03	2.72	2.37		V1-82*01	17	13	6	10	7	10
V1-9*01	1.97	1.83	2.44	3.56	1.24	1.21		2.08	2.00	2.04		V1-9*01	13	10	9	6	23	24
V1-80*01	1.55	1.72	2.16	2.42	1.81	1.99		1.81	2.07	1.94		V1-80*01	20	11	10	11	16	15
V1-72*01	1.83	1.28	1.95	1.45	2.43	2.02		1.69	1.96	1.83		V1-72*01	15	18	13	16	9	13
V2-2*01	1.13	1.05	2.92	1.36	1.25	3.18		1.70	1.93	1.82		V2-2*01	26	22	5	17	22	9
V1-64*01	1.86	2.59	1.26	1.04	2.18	1.68		1.90	1.63	1.77		V1-64*01	14	8	18	24	11	18
V1-81*01	1.17	0.69	0.84	1.07	1.19	3.74		0.90	2.00	1.45		V1-81*01	25	32	31	22	25	6
V1-74*01	2.97	0.76	0.81	0.96	0.53	2.57		1.51	1.35	1.43		V1-74*01	7	29	32	26	53	11
V9-1*01	1.57	2.75	0.77	0.19	2.28	0.64		1.70	1.03	1.37		V9-1*01	18	7	34	67	10	39
V1-62-2*01	2.58	0.44	0.23	0.19	0.68	3.31		1.09	1.39	1.24		V1-62-2*01	8	50	65	65	42	7
V1-71*01	2.58	0.44	0.23	0.19	0.68	3.31		1.08	1.39	1.24		V1-71*01	8	51	66	65	42	7
V5-6*01	1.66	0.85	1.41	0.96	1.27	1.25		1.31	1.16	1.23		V5-6*01	16	27	15	27	21	20
V11-2*01	1.57	1.49	0.77	1.07	1.31	1.13		1.28	1.17	1.22		V11-2*01	19	15	34	23	19	27
V2-9-1*01	0.43	1.46	0.96	0.47	1.94	2.01		0.95	1.47	1.21		V2-9-1*01	48	16	26	45	13	14
V14-4*01	1.25	0.54	0.96	1.56	0.78	1.91		0.92	1.42	1.17		V14-4*01	23	40	25	12	37	16
V1-52*01	1.19	1.19	1.17	1.13	1.19	1.08		1.19	1.13	1.16		V1-52*01	24	20	19	20	24	29
V1-76*01	0.48	0.40	1.12	1.49	2.74	0.55		0.66	1.59	1.13		V1-76*01	46	55	20	14	8	45
V2-3*01	2.22	0.39	0.98	0.57	1.64	0.80		1.20	1.00	1.10		V2-3*01	11	56	24	37	17	33
V4-1*01	0.84	0.51	0.34	0.36	4.10	0.30		0.56	1.59	1.07		V4-1*01	32	45	57	51	4	56
V14-2*01	0.81	0.51	2.61	1.20	0.67	0.60		1.31	0.83	1.07		V14-2*01	33	44	8	18	45	40
V1-75*01	0.86	0.79	1.63	0.93	0.87	1.21		1.09	1.00	1.05		V1-75*01	31	28	14	28	33	22
V1-22*01	0.32	1.58	0.65	2.51	0.68	0.52		0.85	1.24	1.04		V1-22*01	58	14	42	9	41	47
V5-17*01	0.56	1.33	1.98	0.62	1.05	0.67		1.29	0.78	1.04		V5-17*01	43	17	12	36	32	37
V9-2*01	1.38	3.00	0.71	0.07	0.15	0.41		1.70	0.21	0.95		V9-2*01	21	6	36	83	73	49
V1-39*01	0.55	0.94	0.60	1.45	1.08	0.94		0.70	1.16	0.93		V1-39*01	44	24	43	15	30	30
V1-50*01	0.99	1.24	0.78	0.81	0.87	0.75		1.00	0.81	0.91		V1-50*01	29	19	33	32	34	34
V8-8*01	2.12	0.47	0.57	1.11	0.59	0.26		1.05	0.66	0.85		V8-8*01	12	48	45	21	50	61
V1-69*01	1.00	0.62	0.87	0.63	0.66	1.21		0.83	0.83	0.83		V1-69*01	28	34	29	35	46	23
V6-6*01	0.68	0.59	1.12	0.54	1.17	0.71		0.80	0.81	0.80		V6-6*01	40	36	20	38	27	35
V5-16*01	1.12	0.52	0.42	0.47	1.06	1.19		0.69	0.91	0.80		V5-16*01	27	43	53	43	31	25
V1-7*01	0.69	0.28	0.53	0.84	2.09	0.26		0.50	1.07	0.78		V1-7*01	39	67	46	31	12	60
V1-18*01	0.33	0.76	1.34	0.97	0.71	0.53		0.81	0.74	0.78		V1-18*01	56	30	17	25	39	46
V1-42*01	0.71	0.54	1.00	0.91	0.65	0.81		0.75	0.79	0.77		V1-42*01	38	41	23	29	47	32
V8-12*01	1.31	0.59	1.00	0.52	0.46	0.57		0.97	0.52	0.74		V8-12*01	22	37	22	39	57	44
V5-9-1*02	0.28	0.66	0.37	0.28	1.45	1.16		0.44	0.96	0.70		V5-9-1*02	62	33	55	58	18	26
V1-59*01	0.64	0.92	0.87	0.35	0.68	0.57		0.81	0.54	0.67		V1-59*01	41	25	30	54	44	43
V1-61*01	0.77	0.59	0.67	1.18	0.44	0.32		0.67	0.65	0.66		V1-61*01	34	37	41	19	59	54
V2-9*01	0.35	0.49	0.36	0.45	1.16	1.13		0.40	0.91	0.66		V2-9*01	54	46	56	46	28	28
V7-3*01	0.76	0.59	0.43	0.35	0.78	0.65		0.59	0.59	0.59		V7-3*01	35	39	51	55	38	38
V2-5*01	0.18	1.64	0.17	0.49	0.46	0.27		0.67	0.40	0.53		V2-5*01	68	12	71	42	58	59
V14-3*01	0.31	0.53	0.50	0.85	0.59	0.36		0.45	0.60	0.52		V14-3*01	59	42	49	30	51	50
V3-1*01	0.52	0.41	0.51	0.76	0.50	0.35		0.48	0.54	0.51		V3-1*01	45	54	48	33	54	51
V10-1*01	0.33	0.60	0.53	0.38	0.61	0.50		0.49	0.50	0.49		V10-1*01	57	35	47	49	48	48
V9-4*01	0.19	0.14	0.22	0.14	0.69	1.23		0.18	0.69	0.43		V9-4*01	66	79	67	73	40	21
V1-54*01	0.41	0.39	0.29	0.27	0.82	0.31		0.36	0.47	0.41		V1-54*01	49	57	60	60	35	55
V10-3*01	0.35	0.27	0.70	0.52	0.21	0.22		0.44	0.32	0.38		V10-3*01	55	68	39	41	65	64
V1-77*01	0.36	0.31	0.18	0.28	0.48	0.59		0.28	0.45	0.37		V1-77*01	53	65	69	59	56	41
V1-34*01	0.71	0.33	0.39	0.34	0.19	0.21		0.48	0.25	0.36		V1-34*01	37	61	54	56	67	65
V14-1*01	0.59	0.48	0.57	0.20	0.16	0.14		0.55	0.17	0.36		V14-1*01	42	47	44	64	72	71
V5-12*01	0.12	0.31	0.14	0.09	1.19	0.22		0.19	0.50	0.34		V5-12*01	74	64	74	78	25	63
V1-12*01	0.29	0.16	0.27	0.39	0.49	0.29		0.24	0.39	0.32		V1-12*01	60	73	61	48	55	57
V3-8*01	0.46	0.39	0.21	0.37	0.28	0.16		0.35	0.27	0.31		V3-8*01	47	57	68	50	61	70
V1-47*01	0.14	0.35	0.13	0.26	0.55	0.34		0.21	0.38	0.29		V1-47*01	72	60	76	61	52	52



A

	G1	G2	G3	F1	F2	F3
V1-85*01	0.11	0.24	0.49	0.23	0.09	0.58
V1-66*01	0.15	0.18	0.25	0.52	0.23	0.27
V1-20*01	0.18	0.32	0.43	0.09	0.17	0.33
V1-5*01	0.12	0.27	0.23	0.47	0.18	0.19
V1-36*01	0.11	0.92	0.07	0.16	0.07	0.08
V1-31*01	0.41	0.17	0.68	0.02	0.03	0.02
V1-58*01	0.10	0.47	0.30	0.21	0.12	0.12
V1-4*01	0.37	0.29	0.17	0.08	0.19	0.18
V1-63*01	0.09	0.08	0.71	0.09	0.08	0.19
V5-9*01	0.13	0.15	0.18	0.15	0.40	0.14
V1-84*01	0.29	0.16	0.26	0.07	0.20	0.11
V1-62-3*01	0.40	0.31	0.29	0.04	0.02	0.02
V1-11*01	0.09	0.36	0.05	0.04	0.15	0.23
V13-2*01	0.27	0.11	0.10	0.12	0.22	0.10
V8-5*01	0.23	0.12	0.10	0.12	0.18	0.13
V2-6*01	0.15	0.14	0.14	0.35	0.06	0.02
V2-6-8*01	0.15	0.14	0.13	0.35	0.06	0.02
V3-5*01	0.12	0.15	0.08	0.08	0.25	0.14
V3-4*01	0.03	0.04	0.02	0.04	0.61	0.04
V12-3*01	0.08	0.11	0.13	0.15	0.09	0.18
V2-4*01	0.09	0.41	0.04	0.07	0.04	0.08
V5-15*01	0.06	0.06	0.10	0.13	0.07	0.10
V1-37*01	0.02	0.08	0.05	0.33	0.04	0.02
V1-23*01	0.04	0.09	0.10	0.16	0.04	0.02
V1-56*01	0.08	0.16	0.10	0.02	0.02	0.01
V5-2*01	0.01	0.05	0.04	0.15	0.03	0.05
V8-2*01	0.10	0.04	0.01	0.02	0.06	0.03
V8-11*01	0.09	0.05	0.05	0.04	0.02	0.01
V1-49*01	0.03	0.06	0.03	0.01	0.03	0.01
V1-43*01	0.05	0.01	0.03	0.02	0.01	0.03
V7-4*01	0.01	0.02	0.02	0.01	0.03	0.05
V1-67*01	0.03	0	0.02	0.02	0.06	0.02
V15-2*01	0.03	0.02	0.03	0.03	0.02	0.01
V1-8*01	0.04	0	0.02	0.02	0.03	0.02
V3-3*01	0.01	0.02	0.02	0.02	0.04	0.00
V8-9*01	0.01	0.04	0.01	0.01	0.01	0.00
V1-21-1*01	0.03	0	0.04	0.00	0.01	0.00
V6-7*01	0.01	0.03	0.02	0.01	0.01	0.01
V2-7*01	0.01	0.02	0.01	0.01	0.03	0.01
V1-17-1*01	0.01	0.04	0.02	0.00	0.00	0
V7-2*01	0.01	0.01	0.01	0.02	0.01	0.01
V16-1*01	0.01	0.02	0.01	0.01	0.01	0.00
V1-16*01	0.01	0.04	0.02	0.00	0	0
V1-27*01	0.03	0	0.03	0.00	0.01	0.00
V1-24*01	0.01	0.04	0.02	0.00	0	0
V8-6*01	0.02	0.03	0.00	0.00	0.01	0.01
V1-14*01	0.04	0.01	0.00	0.01	0.00	0.00
V6-4*01	0.00	0.01	0.01	0.01	0.02	0.00
V8-4*01	0.01	0.01	0.01	0.01	0.02	0.00
V5-12-4*01	0.01	0.02	0.01	0.00	0.01	0.01
V11-1*01	0.01	0.01	0.01	0.00	0.00	0.00
V1-51*01	0.01	0	0.01	0.00	0.01	0.00
V1-62-1*01	0.01	0.02	0.01	0.00	0	0
V1-48*01	0.01	0	0.02	0.00	0.00	0.00
V6-5*01	0.00	0.00	0.01	0.01	0.00	0.01
V5-21*01	0.00	0	0.01	0.01	0.01	0.01
V5-1*01	0.01	0	0.01	0.00	0.01	0.00
V3-7*01	0.00	0	0.01	0.01	0.00	0.00
V1-13*01	0.00	0	0.00	0.01	0	0.00
V1-19-1*01	0.00	0	0.01	0.00	0.00	0.00
V1-32*01	0.00	0	0.01	0.00	0.00	0.00
V1-35*01	0.00	0	0.01	0.00	0.00	0.00
V12-2*01	0.00	0	0.00	0.00	0.00	0
V1-60*01	0	0	0.00	0	0.00	0.00

Avg G	Avg F	Avg T
0.28	0.30	0.29
0.19	0.34	0.27
0.31	0.20	0.26
0.21	0.28	0.25
0.37	0.10	0.23
0.42	0.02	0.22
0.29	0.15	0.22
0.27	0.15	0.21
0.29	0.12	0.21
0.15	0.23	0.19
0.23	0.13	0.18
0.33	0.03	0.18
0.17	0.14	0.15
0.16	0.15	0.15
0.15	0.15	0.15
0.14	0.14	0.14
0.14	0.14	0.14
0.12	0.16	0.14
0.03	0.23	0.13
0.11	0.14	0.12
0.18	0.06	0.12
0.08	0.10	0.09
0.05	0.13	0.09
0.07	0.08	0.07
0.11	0.02	0.06
0.03	0.07	0.05
0.05	0.04	0.05
0.06	0.02	0.04
0.04	0.02	0.03
0.03	0.02	0.03
0.02	0.03	0.03
0.02	0.03	0.02
0.03	0.02	0.02
0.02	0.02	0.02
0.02	0.02	0.02
0.02	0.01	0.02
0.02	0.01	0.01
0.01	0.01	0.01
0.02	0.00	0.01
0.01	0.01	0.01
0.02	0.01	0.01
0.02	0.00	0.01
0.02	0.00	0.01
0.02	0.00	0.01
0.02	0.01	0.01
0.02	0.01	0.01
0.01	0.01	0.01
0.01	0.01	0.01
0.01	0.01	0.01
0.01	0.00	0.01
0.01	0.01	0.01
0.01	0.00	0.01
0.01	0.00	0.01
0.00	0.01	0.01
0.01	0.00	0.01
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00

	G1	G2	G3	F1	F2	F3
V1-85*01	77	70	50	62	76	42
V1-66*01	71	71	63	40	63	58
V1-20*01	67	62	51	77	71	53
V1-5*01	74	69	64	44	70	66
V1-36*01	78	26	84	69	80	80
V1-31*01	50	72	40	91	94	86
V1-58*01	79	49	58	63	75	75
V1-4*01	52	66	72	80	68	68
V1-63*01	81	86	38	79	78	67
V5-9*01	73	76	70	70	60	73
V1-84*01	61	74	62	82	66	76
V1-62-3*01	51	63	59	88	100	89
V1-11*01	83	59	85	85	74	62
V13-2*01	63	83	80	76	64	77
V8-5*01	65	81	78	75	69	74
V2-6*01	69	78	73	52	83	90
V2-6-8*01	70	80	75	52	83	90
V3-5*01	74	77	83	81	62	72
V3-4*01	94	94	99	86	49	83
V12-3*01	86	82	77	71	77	69
V2-4*01	82	53	90	84	87	79
V5-15*01	87	88	79	74	79	78
V1-37*01	99	85	86	57	86	93
V1-23*01	91	84	82	68	85	87
V1-56*01	85	74	80	95	98	96
V5-2*01	117	89	88	72	92	82
V8-2*01	80	91	105	92	81	85
V8-11*01	84	90	87	86	95	99
V1-49*01	92	87	91	98	92	94
V1-43*01	88	106	92	90	109	84
V7-4*01	103	101	95	98	89	81
V1-67*01	96	113	97	93	82	88
V15-2*01	95	100	94	89	99	97
V1-8*01	89	113	102	95	90	92
V3-3*01	105	103	103	93	88	115
V8-9*01	100	91	107	103	101	112
V1-21-1*01	93	113	89	112	107	109
V6-7*01	115	97	103	104	102	103
V2-7*01	103	103	116	108	90	99
V1-17-1*01	108	91	96	116	119	123
V7-2*01	110	107	110	97	103	98
V16-1*01	115	99	105	101	104	110
V1-16*01	108	94	100	122	124	123
V1-27*01	96	113	93	118	109	114
V1-24*01	105	96	101	122	124	123
V8-6*01	98	98	124	125	106	102
V1-14*01	90	110	125	105	114	106
V6-4*01	119	109	115	101	96	106
V8-4*01	112	111	118	98	97	110
V5-12-4*01	101	102	118	111	109	103
V11-1*01	105	108	109	115	114	115
V1-51*01	101	113	112	112	104	105
V1-62-1*01	111	105	108	122	124	123
V1-48*01	114	113	98	116	114	112
V6-5*01	124	112	118	108	113	95
V5-21*01	124	113	113	105	107	99
V5-1*01	112	113	110	112	112	115
V3-7*01	124	113	118	105	118	115
V1-13*01	122	113	126	108	124	106
V1-19-1*01	119	113	113	125	121	119
V1-32*01	119	113	117	120	121	119
V1-35*01	118	113	122	120	121	119
V12-2*01	122	113	123	118	114	123
V1-60*01	127	113	126	127	119	122



B

	G1	G2	G3	F1	F2	F3
V5-39*01	17.87	35.97	21.10	16.89	17.08	20.64
V3-4*01	15.33	8.62	12.80	4.70	3.43	9.05
V1-117*01	2.42	2.23	2.34	8.60	4.03	1.97
V2-137*01	1.57	1.86	12.18	2.54	1.71	1.37
V1-110*01	2.15	1.84	1.88	1.98	2.68	8.65
V10-96*01	2.39	2.37	2.04	5.31	3.29	1.90
V4-61*01	0.25	0.41	0.28	0.30	13.91	0.56
V4-55*01	0.96	2.84	2.27	4.33	0.50	1.24
V14-111*01	2.40	1.26	2.16	2.43	1.04	2.67
V6-25*01	9.26	0.54	0.56	0.79	0.26	0.30
V14-126-1*01	2.50	2.54	1.25	1.49	1.86	1.74
V6-15*01	1.16	2.61	1.39	2.13	1.55	1.95
V3-2*01	1.48	2.04	0.96	1.34	3.01	1.88
V12-44*01	1.98	1.21	1.05	2.47	1.64	1.66
V4-68*01	0.50	0.67	3.27	0.54	1.20	3.70
V6-17*01	1.54	0.98	0.61	1.58	3.99	0.80
V1-135*01	1.13	1.12	2.72	0.89	2.17	1.30
V4-57-1*01	0.95	1.39	2.40	1.03	1.29	2.18
V19-93*01	1.89	0.96	0.83	2.56	0.84	1.04
V8-24*01	0.89	0.75	1.40	2.46	1.66	0.94
V6-13*01	0.34	0.19	0.11	6.98	0.13	0.08
V3-5*01	1.75	1.58	0.82	1.50	0.66	1.21
V4-59*01	0.69	0.75	1.49	1.08	2.54	0.87
V6-20*01	1.85	0.64	0.65	1.77	0.88	1.15
V16-104*01	3.47	0.47	0.40	0.90	0.89	0.65
V5-43*01	0.65	0.93	0.65	0.51	0.68	3.03
V6-23*01	0.45	0.83	1.43	2.15	0.91	0.51
V12-46*01	0.62	0.99	1.16	0.61	1.08	1.75
V4-91*01	0.44	0.52	0.63	0.62	1.28	2.66
V17-127*01	0.60	1.06	0.80	0.85	0.84	1.79
V6-32*01	0.75	0.68	0.82	0.63	2.12	0.78
V9-120*01	1.55	0.69	0.90	0.80	1.03	0.82
V4-53*01	0.48	0.92	0.35	0.26	1.04	2.58
V3-12-1*01	1.60	0.73	0.56	0.92	0.54	1.08
V9-124*01	0.23	0.54	1.62	1.25	0.75	0.92
V17-121*01	0.45	1.09	0.34	0.41	1.26	1.47
V13-85*01	1.07	0.42	1.69	0.42	0.53	0.47
V3-10*01	0.71	0.59	0.83	0.46	1.26	0.51
V10-94*01	0.34	0.61	0.49	1.70	0.53	0.40
V8-30*01	0.52	0.65	0.56	0.70	0.82	0.30
V8-19*01	0.78	0.70	0.40	0.28	0.71	0.59
V2-109*01	0.51	0.54	0.55	0.50	0.78	0.34
V3-7*01	0.23	0.38	0.73	0.26	0.83	0.78
V4-72*01	0.42	1.23	0.41	0.48	0.31	0.31
V8-27*01	0.53	0.44	0.51	0.32	0.86	0.49
V13-84*01	0.70	0.90	0.21	0.40	0.47	0.42
V6-29*01	0.01	0.46	0.56	0.55	0.62	0.61
V15-103*01	0.25	0.24	0.44	0.47	0.92	0.46
V4-63*01	1.04	0.78	0.13	0.19	0.39	0.17
V4-86*01	0.29	0.44	0.33	0.39	0.50	0.41
V4-70*01	0.43	0.37	0.39	0.32	0.34	0.49
V8-28*01	0.54	0.28	0.39	0.36	0.35	0.36
V8-18*01	1.79	0.04	0.05	0.07	0.04	0.05
V14-100*01	0.29	0.29	0.31	0.30	0.43	0.26
V12-41*01	0.19	0.38	0.21	0.41	0.42	0.19
V12-98*01	0.22	0.10	0.38	0.74	0.08	0.14
V9-123*01	0.41	0.02	0.02	0.09	0.03	1.07
V4-80*01	0.12	0.23	0.30	0.50	0.28	0.21
V12-89*01	0.18	0.21	0.27	0.57	0.23	0.16
V5-45*01	0.16	0.42	0.11	0.22	0.38	0.30
V8-21*01	0.16	0.29	0.33	0.23	0.16	0.36
V1-99*01	0.36	0.47	0.11	0.16	0.11	0.19
V6-14*01	0.41	0.14	0.23	0.13	0.31	0.16

Avg G	Avg F	Avg T
24.98	18.20	21.59
12.25	5.73	8.99
2.33	4.87	3.60
5.20	1.87	3.54
1.96	4.44	3.20
2.27	3.50	2.88
0.31	4.92	2.62
2.02	2.02	2.02
1.94	2.05	1.99
3.45	0.45	1.95
2.10	1.69	1.90
1.72	1.88	1.80
1.49	2.08	1.79
1.42	1.92	1.67
1.48	1.81	1.65
1.04	2.12	1.58
1.66	1.45	1.56
1.58	1.50	1.54
1.23	1.48	1.35
1.01	1.69	1.35
0.22	2.39	1.30
1.38	1.12	1.25
0.98	1.50	1.24
1.05	1.27	1.16
1.44	0.81	1.13
0.74	1.41	1.08
0.91	1.19	1.05
0.93	1.15	1.04
0.53	1.52	1.03
0.82	1.16	0.99
0.75	1.18	0.96
1.05	0.88	0.96
0.58	1.29	0.94
0.96	0.85	0.90
0.80	0.97	0.89
0.63	1.05	0.84
1.06	0.47	0.76
0.71	0.74	0.73
0.48	0.88	0.68
0.58	0.61	0.59
0.63	0.53	0.58
0.53	0.54	0.54
0.44	0.62	0.53
0.69	0.37	0.53
0.49	0.56	0.53
0.60	0.43	0.52
0.34	0.60	0.47
0.31	0.62	0.46
0.65	0.25	0.45
0.35	0.43	0.39
0.40	0.38	0.39
0.40	0.36	0.38
0.63	0.06	0.34
0.30	0.33	0.31
0.26	0.34	0.30
0.23	0.32	0.28
0.15	0.40	0.28
0.22	0.33	0.28
0.22	0.32	0.27
0.23	0.30	0.27
0.26	0.25	0.26
0.31	0.16	0.23
0.26	0.20	0.23

	G1	G2	G3	F1	F2	F3
V5-39*01	1	1	1	1	1	1
V3-4*01	2	2	2	5	5	2
V1-117*01	6	7	7	2	3	10
V2-137*01	16	9	3	8	13	19
V1-110*01	9	10	11	14	8	3
V10-96*01	8	6	10	4	6	12
V4-61*01	55	49	55	54	2	37
V4-55*01	24	3	8	6	44	21
V14-111*01	7	13	9	11	24	6
V6-25*01	3	38	35	29	57	54
V14-126-1*01	5	5	18	19	12	16
V6-15*01	20	4	17	13	16	11
V3-2*01	19	8	21	20	7	13
V12-44*01	10	15	20	9	15	17
V4-68*01	40	33	4	37	21	4
V6-17*01	18	20	32	17	4	31
V1-135*01	21	16	5	26	10	20
V4-57-1*01	25	12	6	23	17	9
V19-93*01	11	21	24	7	32	26
V8-24*01	26	28	16	10	14	27
V6-13*01	51	65	67	3	74	75
V3-5*01	14	11	26	18	39	22
V4-59*01	31	27	14	22	9	29
V6-20*01	12	35	29	15	29	23
V16-104*01	4	43	43	25	28	34
V5-43*01	32	22	30	38	38	5
V6-23*01	44	25	15	12	27	38
V12-46*01	33	19	19	34	22	15
V4-91*01	45	41	31	33	18	7
V17-127*01	34	18	27	27	31	14
V6-32*01	28	32	25	32	11	33
V9-120*01	17	31	22	28	25	30
V4-53*01	42	23	49	58	23	8
V3-12-1*01	15	29	34	24	41	24
V9-124*01	58	39	13	21	36	28
V17-121*01	43	17	50	47	20	18
V13-85*01	22	48	12	45	43	42
V3-10*01	29	37	23	43	19	39
V10-94*01	52	36	40	16	42	46
V8-30*01	38	34	33	31	34	53
V8-19*01	27	30	44	57	37	36
V2-109*01	39	40	37	39	35	49
V3-7*01	57	51	28	58	33	32
V4-72*01	47	14	42	41	54	50
V8-27*01	37	45	38	52	30	41
V13-84*01	30	24	59	49	46	44
V6-29*01	93	44	36	36	40	35
V15-103*01	56	59	41	42	26	43
V4-63*01	23	26	62	64	49	63
V4-86*01	53	46	51	50	44	45
V4-70*01	46	53	45	53	53	40
V8-28*01	36	57	46	51	52	48
V8-18*01	13	84	77	75	83	80
V14-100*01	54	54	53	55	47	56
V12-41*01	62	52	58	48	48	60
V12-98*01	61	74	48	30	78	68
V9-123*01	48	92	87	73	86	25
V4-80*01	69	60	54	40	56	59
V12-89*01	63	61	56	35	59	66
V5-45*01	64	47	69	61	50	52
V8-21*01	64	56	52	60	66	47
V1-99*01	50	42	66	66	76	61
V6-14*01	49	69	57	69	55	66

B

	G1	G2	G3	F1	F2	F3	Avg G	Avg F	Avg T		G1	G2	G3	F1	F2	F3
V4-74*01	0.49	0.25	0.12	0.21	0.14	0.10	0.28	0.15	0.22	V4-74*01	41	58	65	63	69	74
V8-16*01	0.09	0.39	0.18	0.13	0.17	0.31	0.22	0.20	0.21	V8-16*01	71	50	61	68	64	51
V2-112*01	0.14	0.11	0.51	0.22	0.16	0.12	0.26	0.17	0.21	V2-112*01	67	73	39	62	67	72
V4-79*01	0.55	0.15	0.07	0.12	0.10	0.22	0.26	0.15	0.20	V4-79*01	35	68	75	70	77	57
V1-88*01	0.09	0.06	0.39	0.12	0.13	0.28	0.18	0.18	0.18	V1-88*01	72	81	47	71	71	55
V18-36*01	0.00	0.15	0.20	0.17	0.22	0.17	0.12	0.18	0.15	V18-36*01	98	67	60	65	60	65
V4-50*01	0.08	0.29	0.10	0.09	0.20	0.13	0.15	0.14	0.15	V4-50*01	74	55	72	73	62	70
V3-1*01	0	0.06	0.12	0.29	0.21	0.18	0.06	0.23	0.14	V3-1*01	100	80	64	56	61	62
V1-122*01	0.22	0.19	0.10	0.06	0.17	0.05	0.17	0.09	0.13	V1-122*01	60	66	71	79	65	79
V1-133*01	0.05	0.11	0.09	0.07	0.35	0.11	0.08	0.18	0.13	V1-133*01	76	71	73	77	51	73
V4-69*01	0.09	0.21	0.03	0.06	0.13	0.22	0.11	0.14	0.12	V4-69*01	73	62	86	80	73	58
V9-129*01	0.03	0.14	0.02	0.42	0.07	0.03	0.06	0.17	0.12	V9-129*01	79	70	88	46	79	87
V7-33*01	0.09	0.09	0.10	0.06	0.20	0.12	0.10	0.13	0.11	V7-33*01	70	76	70	78	63	71
V4-58*01	0.12	0.11	0.11	0.04	0.14	0.13	0.11	0.10	0.11	V4-58*01	68	72	67	81	70	69
V14-130*01	0.04	0.21	0.07	0.15	0.13	0.04	0.11	0.11	0.11	V14-130*01	78	63	74	67	72	82
V11-125*01	0.15	0.09	0.12	0.07	0.12	0.07	0.12	0.09	0.10	V11-125*01	66	77	63	76	75	76
V4-54*01	0.01	0.01	0.01	0.45	0.01	0.04	0.01	0.17	0.09	V4-54*01	90	96	91	44	95	85
V4-90*01	0.02	0.04	0.06	0.03	0.15	0.17	0.04	0.12	0.08	V4-90*01	84	85	76	83	68	64
V12-38*01	0.03	0.19	0.05	0.04	0.06	0.07	0.09	0.05	0.07	V12-38*01	81	64	79	82	82	77
V10-95*01	0.06	0.09	0.03	0.12	0.07	0.04	0.06	0.08	0.07	V10-95*01	75	78	84	72	79	83
V3-9*01	0.23	0.02	0.05	0.01	0.04	0.01	0.10	0.02	0.06	V3-9*01	59	93	78	95	85	96
V4-71*01	0.01	0.05	0.00	0.00	0.24	0.01	0.02	0.08	0.05	V4-71*01	88	83	98	99	58	94
V4-62*01	0.05	0.08	0.01	0.01	0.01	0.06	0.05	0.03	0.04	V4-62*01	77	79	92	93	94	78
V8-23-1*01	0.02	0.10	0.00	0.02	0.03	0.05	0.04	0.03	0.04	V8-23-1*01	85	75	99	89	89	81
V4-51*01	0.03	0.03	0.04	0.01	0.06	0.02	0.03	0.03	0.03	V4-51*01	79	89	81	91	81	91
V5-37*01	0.02	0.05	0.03	0.02	0.04	0.02	0.04	0.03	0.03	V5-37*01	85	82	83	90	84	90
V4-81*01	0.02	0.03	0.03	0.02	0.03	0.04	0.03	0.03	0.03	V4-81*01	87	88	85	88	86	84
V3-3*01	0.01	0.03	0.04	0.02	0.03	0.02	0.03	0.02	0.03	V3-3*01	91	87	82	87	88	92
V4-78*01	0.02	0.03	0.01	0.02	0.02	0.04	0.02	0.03	0.02	V4-78*01	83	90	94	86	92	85
V4-92*01	0.01	0.03	0.02	0.03	0.02	0.03	0.02	0.02	0.02	V4-92*01	92	91	89	84	93	88
V8-34*01	0.01	0.01	0.02	0.02	0.03	0.02	0.01	0.02	0.02	V8-34*01	89	95	89	85	89	92
V1-35*01	0.00	0.03	0.04	0.01	0.01	0.01	0.03	0.01	0.02	V1-35*01	99	86	80	93	96	95
V4-73*01	0.01	0.01	0.01	0.01	0.03	0.02	0.01	0.02	0.02	V4-73*01	95	98	93	92	91	89
V1-132*01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	V1-132*01	94	96	97	96	97	97
V1-131*01	0.00	0.02	0.01	0.01	0.00	0.00	0.01	0.00	0.01	V1-131*01	96	94	96	97	100	100
V8-26*01	0.02	0.00	0	0.00	0.00	0.00	0.01	0.00	0.01	V8-26*01	82	99	100	100	98	98
V20-101-2*01	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	V20-101-2*01	97	100	95	97	99	98

(A) VH- and (B) V $\kappa$ -gene segment usage in animals within ground (G) and flight (F) treatment groups are presented as percent of repertoire and rank. V-gene segments are listed by rank order (most frequent to least frequent). Dark red indicates higher percent of repertoire or rank moving to blue, lower percent of repertoire or rank. V-gene segments with identical ranks are displayed as ties.

## Appendix A.7. D- and J-Gene Segment and Constant Region Heat maps

A

	G1	G2	G3	F1	F2	F3		Avg G	Avg F	Avg T		G1	G2	G3	F1	F2	F3	
D1-1*01	39.20	46.80	30.60	40.24	41.34	40.16		38.87	40.58	39.73		D1-1*01	1	1	2	1	1	1
Undeter	25.84	24.02	36.60	25.36	31.09	24.74		28.82	27.06	27.94		Undeter	2	2	1	2	2	2
D2-4*01	8.61	5.58	7.43	10.19	4.51	6.80		7.21	7.17	7.19		D2-4*01	3	5	4	3	5	5
D2-3*01	5.10	6.63	11.39	6.71	4.44	8.53		7.71	6.56	7.13		D2-3*01	6	4	3	5	6	3
D4-1*01	7.83	5.31	5.26	7.92	7.83	7.25		6.13	7.66	6.90		D4-1*01	4	6	5	4	3	4
D2-5*01	6.82	7.02	3.56	4.85	4.64	6.22		5.80	5.24	5.52		D2-5*01	5	3	6	6	4	6
D3-1*01	3.64	1.75	2.49	1.87	2.63	2.83		2.63	2.44	2.54		D3-1*01	7	8	7	8	7	7
D3-2*02	2.30	1.77	1.87	2.08	2.58	1.77		1.98	2.14	2.06		D3-2*02	8	7	8	7	8	8
D6-2*02	0.46	1.00	0.26	0.23	0.46	0.63		0.58	0.44	0.51		D6-2*02	9	9	10	10	9	10
D5-1*01	0.11	0	0.39	0.45	0.33	0.35		0.17	0.38	0.27		D5-1*01	10	11	9	9	10	11
D5-5*01	0.09	0.11	0.13	0.11	0.13	0.73		0.11	0.32	0.22		D5-5*01	11	10	11	11	11	9

B

	G1	G2	G3	F1	F2	F3		Avg G	Avg F	Avg T		G1	G2	G3	F1	F2	F3	
J2*01	29.35	27.01	28.75	35.26	27.19	40.43		28.37	34.29	31.33		J2*01	2	2	2	1	2	1
J1*03	30.25	40.25	19.77	19.68	20.69	24.88		30.09	21.75	25.92		J1*03	1	1	4	3	3	2
J4*01	19.66	13.09	22.60	18.92	34.70	18.78		18.45	24.13	21.29		J4*01	4	4	3	4	1	3
J3*01	20.69	19.49	28.80	26.03	16.58	15.90		22.99	19.50	21.25		J3*01	3	3	1	2	4	4
U	0.04	0.16	0.08	0.11	0.84	0.01		0.09	0.32	0.21		<6	5	5	5	5	5	5

C

	G1	G2	G3	F1	F2	F3		Avg G	Avg F	Avg T		G1	G2	G3	F1	F2	F3	
J5*01	37.27	48.07	32.56	39.64	31.87	40.19		39.30	37.23	38.27		J5*01	1	1	1	1	2	1
J2*01	25.08	20.37	20.98	23.99	33.69	27.43		22.15	28.37	25.26		J2*01	3	3	3	2	1	2
J1*01	26.74	21.43	28.76	21.02	21.59	20.09		25.64	20.90	23.27		J1*01	2	2	2	3	3	3
J4*01	9.77	8.45	15.71	13.49	10.65	10.54		11.31	11.56	11.44		J4*01	4	4	4	4	4	4
U	1.14	1.68	1.99	1.86	2.20	1.76		1.60	1.94	1.77		U	5	5	5	5	5	5

D

	G1	G2	G3	F1	F2	F3
IgM	82.81	80.03	67.44	68.07	62.38	68.18
IgG	12.27	17.66	29.61	28.82	25.56	29.08
IgA	4.58	1.35	2.32	2.54	11.26	2.08
IgD	0.31	0.95	0.56	0.53	0.79	0.64
IgE	0.04	0.02	0.07	0.03	0.01	0.02

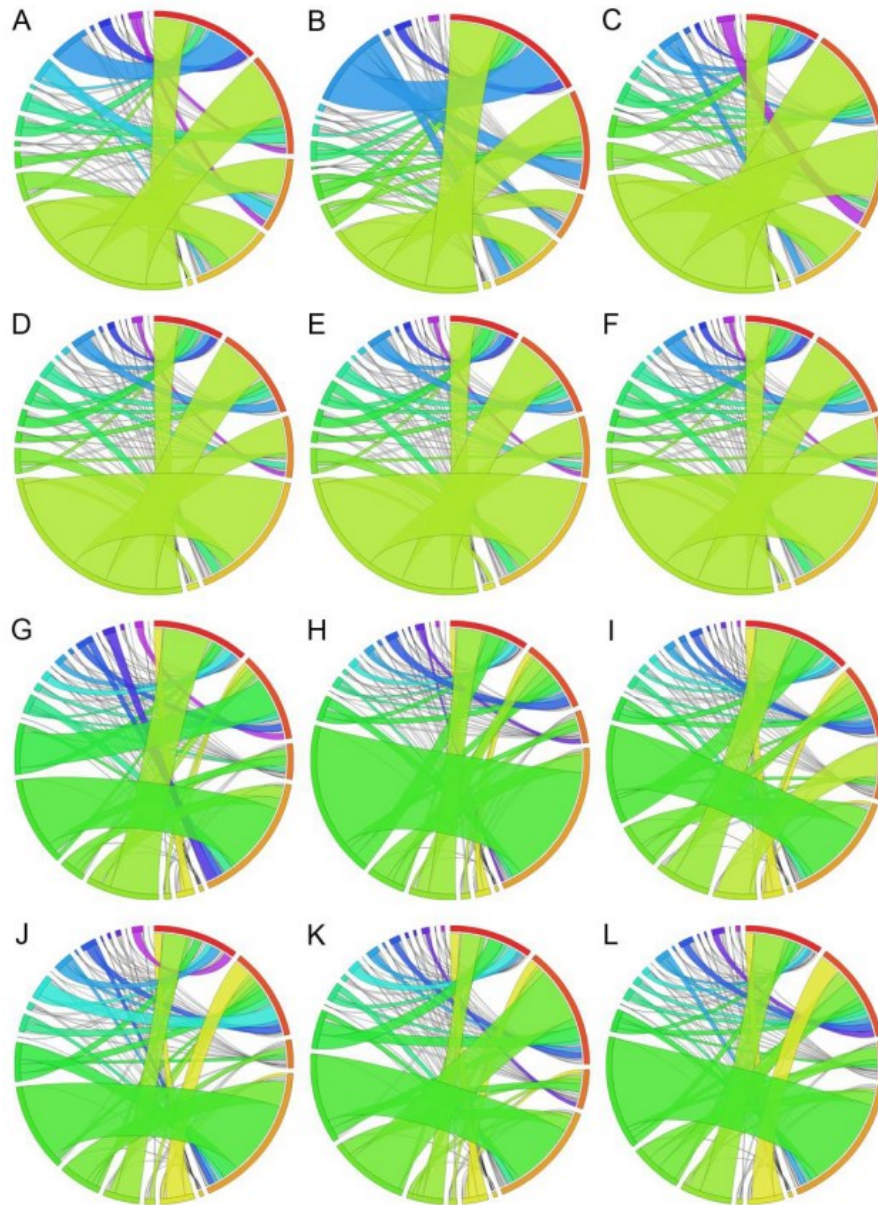
	Avg G	Avg F	Avg T
	76.76	66.21	71.48
	19.85	27.82	23.83
	2.75	5.29	4.02
	0.61	0.66	0.63
	0.04	0.02	0.03

	G1	G2	G3	F1	F2	F3
IgM	1	1	1	1	1	1
IgG	2	2	2	2	2	2
IgA	3	3	3	3	3	3
IgD	4	4	4	4	4	4
IgE	5	5	5	5	5	5

(A) D-gene segment, (B) JH-gene segment, (C) J $\kappa$ -gene segment, (D) IgH constant region usage in animals within ground (G) and flight (F) treatment groups are presented as percent of repertoire and rank. Dark red indicates higher percent of repertoire or rank moving to white, lower percent of repertoire or rank.



## Appendix A.8. V/J Combinations of Individual Animals



(A-C) IgH V/J pairings from G1, G2, and G3 respectively. (D-F) IgH V/J pairings from F1, F2, and F3 respectively. Circos plots are read clockwise starting at the 12 o'clock position starting with J1 (red), J2, J3, J4, U, V1 (lime green), V2, V3, V4, V5, V6, V7, V8 (light blue), V9, V10, V11, V12, V13, V14, and V15 (sliver, no color). (G-I) Igk V/J pairings from G1, G2, and G3 respectively. (J-L) Igk V/J pairings from F1, F2, and F3 respectively. Circos plots are read clockwise starting at the 12 o'clock position with J1 (red), J2, J4, J5, U, V1 (yellow), V2, V3, V4, V5, V6, V7 (sliver, no color), V8, V9, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19 (light purple).



## Appendix A.9. Top V(D)J Gene Family Combinations

A	<b>G1</b>	<b>%</b>	<b>G2</b>	<b>%</b>	<b>G3</b>	<b>%</b>	<b>Avg G</b>	<b>%</b>	<b>SD</b>	<div>Shared among 3 mice</div> <div>Shared among 2 mice</div> <div>Unique to 1 mouse</div>
	V9/D1/J1	8.40	V9/D1/J1	10.5	V1/U/J3	13.1				
	V1/U/J4	7.90	V1/D1/J1	9.89	V1/U/J2	5.92				
	V1/D1/J2	6.95	V1/D1/J2	4.83	V1/D2/J2	5.61				
	V1/D1/J1	5.50	V1/U/J2	3.81	V1/D1/J2	4.83				
B	V1/U/J2	3.93	V1/U/J4	3.25	V1/U/J4	4.23	<b>Avg F</b>	<b>%</b>	<b>SD</b>	
C	<b>F1</b>	<b>%</b>	<b>F2</b>	<b>%</b>	<b>F3</b>	<b>%</b>	<b>Avg G</b>	<b>%</b>	<b>SD</b>	
	V1/D1/J2	9.63	V1/D1/J4	20.3	V1/D1/J2	9.09				
	V3/D1/J3	8.02	V1/U/J3	4.87	V9/D1/J1	6.10				
	V1/D1/J1	6.82	V1/U/J2	4.59	V1/U/J2	5.45				
	V1/U/J2	6.09	V1/D1/J2	4.30	V1/D2/J2	4.90				
D	V5/D2/J2	4.95	V1/D1/J1	3.69	V1/D1/J3	3.59	<b>Avg F</b>	<b>%</b>	<b>SD</b>	
E	<b>G1</b>	<b>%</b>	<b>G2</b>	<b>%</b>	<b>G3</b>	<b>%</b>	<b>Avg G</b>	<b>%</b>	<b>SD</b>	
	V1/J2	16.80	V1/J1	14.3	V1/J3	19.1				
	V1/J4	11.70	V1/J2	12.3	V1/J2	18.9				
	V1/J3	10.6	V9/J1	12.3	V1/J4	11				
	V1/J1	9.74	V1/J4	9.27	V1/J1	7.65				
F	V9/J1	8.75	V1/J3	6.22	V2/J4	4.19	<b>Avg F</b>	<b>%</b>	<b>SD</b>	
G	<b>F1</b>	<b>%</b>	<b>F2</b>	<b>%</b>	<b>F3</b>	<b>%</b>	<b>Avg G</b>	<b>%</b>	<b>SD</b>	
	V1/J2	20.7	V1/J4	26.3	V1/J2	22.7				
	V1/J1	12.2	V1/J2	14	V1/J1	10.20				
	V1/J4	12	V1/J3	9.82	V1/J3	9.99				
	V1/J3	10.3	V1/J1	6.99	V1/J4	9.11				
H	V3/J3	10.3	V2/J4	3.58	V9/J1	6.3	<b>Avg F</b>	<b>%</b>	<b>SD</b>	
I	<b>G1</b>	<b>%</b>	<b>G2</b>	<b>%</b>	<b>G3</b>	<b>%</b>	<b>Avg G</b>	<b>%</b>	<b>SD</b>	
	V5/J5	20.0	V5/J5	34.6	V5/J5	18.2				
	V3/J1	13.9	V3/J1	7.70	V3/J1	10.6				
	V6/J2	10.5	V4/J5	4.97	V2/J4	9.32				
	V3/J5	3.67	V5/J1	4.31	V5/J1	6.66				
J	V4/J5	3.15	V3/J2	3.60	V4/J5	6.31	<b>Avg F</b>	<b>%</b>	<b>SD</b>	
K	<b>F1</b>	<b>%</b>	<b>F2</b>	<b>%</b>	<b>F3</b>	<b>%</b>	<b>Avg G</b>	<b>%</b>	<b>SD</b>	
	V5/J5	21.7	V5/J5	19.0	V5/J5	25.8				
	V4/J5	7.28	V4/J2	15.5	V3/J1	9.07				
	V3/J1	7.06	V3/J1	5.81	V1/J2	6.26				
	V1/J2	5.49	V6/J1	4.22	V4/J2	5.41				
L	V6/J5	4.55	V6/J2	3.93	V4/J4	3.8	<b>Avg F</b>	<b>%</b>	<b>SD</b>	

(A, B) Top 5 V/D/J combinations for IgH ground (A) and flight (B) mice with average and standard deviation (SD) per treatment group. (C, D) Top 5 V/J combinations for IgH ground (C) and flight (D) mice. (E, F) Top 5 V/J combinations for Igk ground (E) and flight (F) mice with average and standard deviation (SD) per treatment group. Gene combinations are color coded to show overlap within the top five most common combinations per treatment group. Combinations found within all three mice are blue, within two mice are green, and unique to a single mouse as white.

## Chapter 6 Supplemental Tables

### Appendix A.10. Average V-, D-, and J-Gene Segment, Constant Region Usage, and CDR3 AA Length for Statistical Differences Among Treatment Groups

	No AOS				AOS			
	No TT		TT		No TT		TT	
	No CpG	CpG	No CpG	CpG	No CpG	CpG	No CpG	CpG
VH Gene Segment Usage (Avg $\pm$ SEM)								
V11-2 <sup>c</sup>	0.55	0.51	0.51	0.31	0.48	0.41	0.58	0.35
	$\pm 0.06$	$\pm 0.14$	$\pm 0.03$	$\pm 0.04$	$\pm 0.08$	$\pm 0.05$	$\pm 0.06$	$\pm 0.08$
V1-12 <sup>e,f</sup>	0.46	0.34	0.29	1.08	0.69	0.35	0.39	0.48
	$\pm 0.08$	$\pm 0.01$	$\pm 0.04$	$\pm 0.30$	$\pm 0.09$	$\pm 0.04$	$\pm 0.07$	$\pm 0.09$
V1-15 <sup>d</sup>	1.19	0.88	0.7	0.96	0.66	0.7	1.47	1.97
	$\pm 0.09$	$\pm 0.11$	$\pm 0.14$	$\pm 0.21$	$\pm 0.02$	$\pm 0.08$	$\pm 0.42$	$\pm 0.84$
V1-31 <sup>b</sup>	0.06	0.14	0.07	0.06	0.14	0.09	0.04	0.06
	$\pm 0.02$	$\pm 0.05$	$\pm 0.02$	$\pm 0.01$	$\pm 0.05$	$\pm 0.02$	$\pm 0.01$	$\pm 0.01$
V1-36 <sup>c</sup>	0.15	0.56	0.15	0.21	0.2	0.24	0.17	0.31
	$\pm 0.02$	$\pm 0.19$	$\pm 0.03$	$\pm 0.06$	$\pm 0.03$	$\pm 0.04$	$\pm 0.02$	$\pm 0.08$
V14-1 <sup>g</sup>	0.13	0.26	0.62	0.12	0.45	0.27	0.17	0.61
	$\pm 0.02$	$\pm 0.07$	$\pm 0.21$	$\pm 0.02$	$\pm 0.27$	$\pm 0.04$	$\pm 0.05$	$\pm 0.23$
V15-2 <sup>d</sup>	0.04	0.04	0.03	0.05	0.06	0.03	0.08	0.05
	$\pm 0.00$	$\pm 0.01$	$\pm 0.01$	$\pm 0.01$	$\pm 0.02$	$\pm 0.01$	$\pm 0.01$	$\pm 0.01$
V1-63 <sup>e</sup>	0.8	0.17	0.6	0.19	0.15	0.19	0.25	0.15
	$\pm 0.32$	$\pm 0.02$	$\pm 0.22$	$\pm 0.03$	$\pm 0.01$	$\pm 0.01$	$\pm 0.08$	$\pm 0.03$
V1-64 <sup>f</sup>	2.37	2.71	3.33	2.45	2.47	2.63	2.74	1.97
	$\pm 0.38$	$\pm 0.06$	$\pm 0.39$	$\pm 0.17$	$\pm 0.23$	$\pm 0.35$	$\pm 0.48$	$\pm 0.12$
V1-69 <sup>f</sup>	1.81	0.87	0.9	1.4	1.32	1.14	0.85	0.81
	$\pm 0.31$	$\pm 0.14$	$\pm 0.21$	$\pm 0.19$	$\pm 0.29$	$\pm 0.16$	$\pm 0.07$	$\pm 0.10$
V1-74 <sup>b</sup>	1.01	0.93	0.40	0.57	0.67	0.74	0.68	0.64
	$\pm 0.23$	$\pm 0.21$	$\pm 0.06$	$\pm 0.06$	$\pm 0.10$	$\pm 0.17$	$\pm 0.11$	$\pm 0.12$
V1-76 <sup>a,b,d</sup>	1.16	0.97	0.63	0.45	0.44	0.60	0.53	0.57
	$\pm 0.18$	$\pm 0.18$	$\pm 0.16$	$\pm 0.08$	$\pm 0.08$	$\pm 0.06$	$\pm 0.06$	$\pm 0.09$
V1-78 <sup>d</sup>	2.15	1.46	1.19	1.11	0.73	0.83	1.14	1.38
	$\pm 0.64$	$\pm 0.14$	$\pm 0.31$	$\pm 0.20$	$\pm 0.10$	$\pm 0.10$	$\pm 0.19$	$\pm 0.39$

V1-80 <sup>c</sup>	5.48 ± 0.90	5.89 ± 0.86	7.54 ± 1.95	4.87 ± 1.52	7.11 ± 0.89	4.32 ± 0.67	7.29 ± 1.54	2.78 ± 0.74
V1-85 <sup>b</sup>	0.68 ± 0.31	0.41 ± 0.04	0.26 ± 0.05	0.23 ± 0.02	0.44 ± 0.06	0.24 ± 0.06	0.27 ± 0.02	0.21 ± 0.03
V2-3 <sup>d</sup>	2.84 ± 0.62	1.06 ± 0.14	0.95 ± 0.11	0.84 ± 0.11	0.98 ± 0.12	0.80 ± 0.19	2.30 ± 1.34	1.01 ± 0.13
V3S7 <sup>e</sup>	0.01 ± 0.00	0.02 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
V4-1 <sup>a</sup>	1.58 ± 0.26	1.08 ± 0.12	1.53 ± 0.25	1.19 ± 0.07	1.81 ± 0.50	2.26 ± 0.50	1.87 ± 0.57	2.78 ± 0.81
<b>DH Gene Segment Usage (Avg ± SEM)</b>								
D3-1 <sup>d</sup>	3.64 ± 0.98	2.37 ± 0.19	2.18 ± 0.18	2.02 ± 0.29	2.14 ± 0.15	2.59 ± 0.03	3.07 ± 0.42	2.94 ± 0.39
D5-5 <sup>b</sup>	0.30 ± 0.09	0.22 ± 0.04	0.18 ± 0.04	0.15 ± 0.02	0.38 ± 0.11	0.2 ± 0.05	0.17 ± 0.01	0.16 ± 0.03
<b>JH Gene Segment Usage (Avg ± SEM)</b>								
J2 <sup>c,d</sup>	28.06 ± 1.09	29.66 ± 1.06	28.33 ± 1.33	35.35 ± 2.87	28.52 ± 2.65	38.95 ± 4.32	27.24 ± 1.11	27.06 ± 2.92
J3 <sup>e</sup>	18.68 ± 0.86	22.24 ± 1.07	17.49 ± 1.64	19.34 ± 3.06	19.62 ± 0.95	17.63 ± 0.95	19.98 ± 0.89	15.05 ± 1.34
<b>Constant Region Usage (Avg ± SEM)</b>								
IgA <sup>c</sup>	3.66 ± 0.72	6.74 ± 1.24	3.02 ± 0.61	8.81 ± 2.11	3.18 ± 0.25	4.32 ± 1.06	3.41 ± 1.62	4.62 ± 1.47
IgG <sup>a</sup>	10.85 ± 0.90	11.83 ± 2.25	12.48 ± 2.15	12.01 ± 1.06	8.99 ± 0.98	8.66 ± 0.25	7.18 ± 0.72	10.41 ± 2.26
IgM <sup>a</sup>	82.58 ± 1.72	78.67 ± 2.71	81.74 ± 1.89	76.27 ± 3.34	85.00 ± 0.84	84.47 ± 1.09	86.13 ± 3.00	81.95 ± 2.84
<b>H-CDR3 AA Length (Avg ± SEM)</b>								
5 <sup>c</sup>	3.37 ± 0.98	4.08 ± 0.45	7.66 ± 1.86	3.76 ± 1.33	5.72 ± 0.63	3.27 ± 0.54	5.77 ± 1.22	2.13 ± 0.42
7 <sup>b,d,e</sup>	4.45 ± 0.70	6.87 ± 0.67	3.20 ± 0.27	3.52 ± 0.34	3.92 ± 0.34	3.92 ± 0.29	4.05 ± 0.15	3.96 ± 0.17
9 <sup>d,f</sup>	7.94 ± 0.19	8.13 ± 0.67	5.08 ± 0.49	7.36 ± 0.35	6.97 ± 0.35	6.59 ± 1.13	6.12 ± 0.30	8.88 ± 1.41
13 <sup>g</sup>	11.44 ± 0.81	13.09 ± 0.46	14.55 ± 1.26	11.18 ± 1.47	13.84 ± 1.54	10.48 ± 0.51	10.87 ± 1.30	10.38 ± 0.34

14 <sup>a</sup>	15.17 ± 1.83	13.95 ± 0.55	17.09 ± 2.31	17.07 ± 3.09	13.72 ± 1.92	10.63 ± 1.26	11.33 ± 1.78	8.66 ± 0.57
-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	----------------

a - P<0.05 for a main effect of AOS

b - P<0.05 for a main effect of TT

c - P<0.05 for a main effect of CpG

d - P<0.05 for an interaction effect of AOSxTT

e - P<0.05 for an interaction effect of AOSxCpG

f - P<0.05 for an interaction effect of TTxCpG

g - P<0.05 for an interaction effect of AOSxTTxCpG

**Appendix A.11. Average V- and J-Gene Segment Usage and CDR3 AA Length for Statistical Differences Among Treatment Groups**

	No AOS				AOS			
	No TT		TT		No TT		TT	
	No CpG	CpG	No CpG	CpG	No CpG	CpG	No CpG	CpG
<b>VK Gene Segment Usage (Avg <math>\pm</math> SEM)</b>								
V1-132 <sup>a</sup>	0.03 $\pm$ 0.00	0.03 $\pm$ 0.00	0.02 $\pm$ 0.01	0.02 $\pm$ 0.00	0.04 $\pm$ 0.01	0.04 $\pm$ 0.01	0.03 $\pm$ 0.01	0.04 $\pm$ 0.01
V12-46 <sup>b</sup>	2.52 $\pm$ 0.46	1.87 $\pm$ 0.16	1.39 $\pm$ 0.17	0.94 $\pm$ 0.12	2.73 $\pm$ 1.02	2.23 $\pm$ 0.53	2.11 $\pm$ 0.34	1.00 $\pm$ 0.05
V4-55 <sup>c</sup>	7.02 $\pm$ 2.18	4.55 $\pm$ 0.85	13.47 $\pm$ 3.59	3.95 $\pm$ 1.74	10.51 $\pm$ 2.06	4.53 $\pm$ 1.24	7.39 $\pm$ 1.81	2.07 $\pm$ 0.46
V4-57-1 <sup>f</sup>	1.27 $\pm$ 0.25	0.50 $\pm$ 0.08	0.40 $\pm$ 0.06	0.61 $\pm$ 0.11	1.93 $\pm$ 1.06	0.41 $\pm$ 0.06	0.64 $\pm$ 0.30	1.02 $\pm$ 0.18
V4-62 <sup>f</sup>	0.08 $\pm$ 0.01	0.06 $\pm$ 0.00	0.04 $\pm$ 0.00	0.05 $\pm$ 0.01	0.11 $\pm$ 0.04	0.04 $\pm$ 0.00	0.05 $\pm$ 0.02	0.08 $\pm$ 0.00
V4-70 <sup>b</sup>	2.75 $\pm$ 0.99	2.92 $\pm$ 0.70	1.64 $\pm$ 0.33	0.41 $\pm$ 0.08	1.7 $\pm$ 0.38	1.45 $\pm$ 0.74	1.80 $\pm$ 0.57	0.62 $\pm$ 0.06
V4-86 <sup>a</sup>	1.12 $\pm$ 0.27	0.86 $\pm$ 0.12	1.41 $\pm$ 0.43	1.02 $\pm$ 0.21	1.97 $\pm$ 0.56	2.22 $\pm$ 0.53	1.54 $\pm$ 0.45	2.86 $\pm$ 0.81
V5-39 <sup>c</sup>	3.73 $\pm$ 1.75	4.59 $\pm$ 1.82	1.62 $\pm$ 0.46	13.78 $\pm$ 4.86	2.62 $\pm$ 0.76	6.49 $\pm$ 2.39	1.95 $\pm$ 0.60	6.62 $\pm$ 4.13
V5-48 <sup>c</sup>	1.02 $\pm$ 0.08	2.20 $\pm$ 0.28	1.08 $\pm$ 0.16	1.40 $\pm$ 0.30	1.16 $\pm$ 0.15	1.38 $\pm$ 0.43	1.13 $\pm$ 0.11	1.22 $\pm$ 0.16
V6-14 <sup>f</sup>	0.31 $\pm$ 0.07	0.38 $\pm$ 0.04	0.46 $\pm$ 0.13	0.27 $\pm$ 0.09	0.22 $\pm$ 0.05	0.54 $\pm$ 0.16	0.39 $\pm$ 0.05	0.30 $\pm$ 0.11
V6-25 <sup>f</sup>	0.51 $\pm$ 0.08	0.74 $\pm$ 0.24	1.17 $\pm$ 0.35	0.80 $\pm$ 0.16	0.54 $\pm$ 0.05	0.78 $\pm$ 0.23	1.13 $\pm$ 0.28	0.47 $\pm$ 0.06
V8-19 <sup>c</sup>	0.59 $\pm$ 0.06	0.79 $\pm$ 0.08	0.51 $\pm$ 0.12	0.76 $\pm$ 0.06	0.55 $\pm$ 0.04	0.84 $\pm$ 0.13	0.62 $\pm$ 0.07	0.60 $\pm$ 0.09
V8-23-1 <sup>d</sup>	0.11 $\pm$ 0.03	0.11 $\pm$ 0.03	0.07 $\pm$ 0.02	0.08 $\pm$ 0.01	0.07 $\pm$ 0.01	0.08 $\pm$ 0.01	0.12 $\pm$ 0.01	0.14 $\pm$ 0.04
V8-30 <sup>f</sup>	0.99 $\pm$ 0.18	1.36 $\pm$ 0.35	1.49 $\pm$ 0.50	1.56 $\pm$ 0.32	1.01 $\pm$ 0.18	8.7 $\pm$ 3.37	1.49 $\pm$ 0.25	1.11 $\pm$ 0.09
<b>JK Gene Segment Usage (Avg <math>\pm</math> SEM)</b>								
U <sup>b</sup>	0.76 $\pm$ 0.08	0.99 $\pm$ 0.13	0.67 $\pm$ 0.04	0.85 $\pm$ 0.06	0.97 $\pm$ 0.05	1.05 $\pm$ 0.11	0.84 $\pm$ 0.20	0.68 $\pm$ 0.03

	<b>K-CDR3 AA Length (Avg ± SEM)</b>							
7 <sup>b</sup>	2.23 ± 0.85	3.46 ± 0.67	1.86 ± 0.37	0.89 ± 0.08	2.03 ± 0.44	2.20 ± 0.84	2.08 ± 0.56	0.91 ± 0.09
8 <sup>d</sup>	8.33 ± 1.05	6.02 ± 0.66	6.27 ± 0.51	6.94 ± 0.90	5.71 ± 0.37	7.40 ± 1.54	7.09 ± 0.65	5.42 ± 0.38
10 <sup>g</sup>	2.97 ± 0.49	2.94 ± 0.29	1.7 ± 0.32	2.16 ± 0.36	1.68 ± 0.37	2.28 ± 0.16	2.17 ± 0.41	2.98 ± 0.35

U – Undetermined gene segment

a - P<0.05 for a main effect of AOS

b - P<0.05 for a main effect of TT

c - P<0.05 for a main effect of CpG

d - P<0.05 for an interaction effect of AOSxTT

e - P<0.05 for an interaction effect of AOSxCpG

f - P<0.05 for an interaction effect of TTxCpG

g - P<0.05 for an interaction effect of AOSxTTxCpG

# Appendix A.12. Average V/J Pairings for Heavy Chain for Statically Different Pairings

	No AOS	AOS	No TT	TT	No CpG	CpG
V10-3/J2 <sup>c</sup>	0.15 ± 0.04	0.15 ± 0.04	0.15 ± 0.04	0.15 ± 0.05	0.11 ± 0.03	0.19 ± 0.05
V11-2/J1 <sup>c</sup>	0.50 ± 0.03	0.54 ± 0.02	0.57 ± 0.02	0.47 ± 0.02	0.62 ± 0.02	0.41 ± 0.02
V1-12/J2 <sup>a</sup>	0.09 ± 0.03	0.16 ± 0.05	0.13 ± 0.04	0.11 ± 0.05	0.11 ± 0.04	0.14 ± 0.05
V1-22/J1 <sup>a</sup>	0.20 ± 0.03	0.49 ± 0.13	0.31 ± 0.1.	0.37 ± 0.14	0.35 ± 0.09	0.34 ± 0.15
V1-26/U <sup>c</sup>	0.06 ± 0.04	0.08 ± 0.04	0.07 ± 0.04	0.07 ± 0.03	0.1 ± 0.05	0.04 ± 0.01
V14-1/J1 <sup>b</sup>	0.03 ± 0.02	0.14 ± 0.07	0.14 ± 0.07	0.02 ± 0.01	0.14 ± 0.07	0.04 ± 0.02
V15-2/J1 <sup>c</sup>	0.02 ± 0.0	0.03 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.01	0.01 ± 0.0
V1-53/J2 <sup>c</sup>	1.14 ± 0.24	1.23 ± 0.29	1.07 ± 0.26	1.30 ± 0.26	0.95 ± 0.15	1.42 ± 0.30
V1-53/J3 <sup>c</sup>	0.65 ± 0.16	0.57 ± 0.10	0.64 ± 0.16	0.57 ± 0.10	0.50 ± 0.10	0.71 ± 0.14
V1-53/J4 <sup>c</sup>	1.70 ± 0.91	1.24 ± 0.73	1.76 ± 0.74	1.19 ± 0.89	2.09 ± 1.04	0.86 ± 0.34
V1-55/U <sup>a</sup>	0.01 ± 0.00	0.03 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.00
V1-63/J2 <sup>a</sup>	0.07 ± 0.02	0.04 ± 0.01	0.05 ± 0.02	0.05 ± 0.02	0.06 ± 0.02	0.05 ± 0.01
V1-74/J2 <sup>b</sup>	0.29 ± 0.09	0.30 ± 0.07	0.32 ± 0.09	0.26 ± 0.06	0.23 ± 0.04	0.35 ± 0.09
V1-76/J1 <sup>a,b</sup>	0.32 ± 0.17	0.12 ± 0.02	0.31 ± 0.17	0.12 ± 0.02	0.25 ± 0.16	0.19 ± 0.09
V1-76/J4 <sup>a</sup>	0.35 ± 0.11	0.19 ± 0.03	0.32 ± 0.1	0.22 ± 0.07	0.29 ± 0.10	0.25 ± 0.09
V1-77/J4 <sup>a</sup>	0.03 ± 0.01	0.04 ± 0.01	0.04 ± 0.01	0.03 ± 0.01	0.03 ± 0.01	0.03 ± 0.01
V1-85/U <sup>b</sup>	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.01 ± 0.01	0.00 ± 0.00

V2-3/J4 <sup>a,b</sup>	0.99 ± 0.63	0.41 ± 0.09	0.99 ± 0.63	0.4 ± 0.09	0.94 ± 0.64	0.46 ± 0.11
V3-8/J2 <sup>a,b</sup>	0.10 ± 0.03	0.16 ± 0.06	0.09 ± 0.02	0.16 ± 0.06	0.12 ± 0.06	0.13 ± 0.06
V5-15/J1 <sup>c</sup>	0.03 ± 0.02	0.05 ± 0.03	0.03 ± 0.02	0.05 ± 0.03	0.06 ± 0.04	0.02 ± 0.01
V5-17/J1 <sup>c</sup>	0.27 ± 0.08	0.31 ± 0.13	0.28 ± 0.08	0.30 ± 0.13	0.20 ± 0.06	0.37 ± 0.13
V6-3/U <sup>b</sup>	0.04 ± 0.02	0.03 ± 0.02	0.02 ± 0.01	0.05 ± 0.02	0.04 ± 0.02	0.03 ± 0.02
V8-4/J2 <sup>a</sup>	0.00 ± 0.00	0.01 ± 0.01	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.01	0.01 ± 0.01
V9-2/J4 <sup>a,c</sup>	0.04 ± 0.02	0.02 ± 0.01	0.03 ± 0.01	0.04 ± 0.02	0.02 ± 0.01	0.04 ± 0.02

U – Undetermined J-gene segment

a - P<0.05 for a main effect of AOS

b - P<0.05 for a main effect of TT

c - P<0.05 for a main effect of CpG



### Appendix A.13. Average V/J Pairings for Kappa Chain for Statically Different Pairings

	No AOS	AOS	No TT	TT	No CpG	CpG
V10-95/U <sup>a</sup>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
V11-125/U <sup>c</sup>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
V12-46/J2 <sup>c</sup>	0.59 ± 0.27	0.38 ± 0.06	0.60 ± 0.26	0.37 ± 0.06	0.62 ± 0.26	0.35 ± 0.05
V12-46/J5 <sup>b</sup>	0.27 ± 0.21	0.40 ± 0.32	0.55 ± 0.37	0.12 ± 0.03	0.50 ± 0.39	0.17 ± 0.06
V14-100/U <sup>b</sup>	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.00 ± 0.00	0.01 ± 0.01	0.01 ± 0.01
V14-111/J2 <sup>a</sup>	0.94 ± 0.29	0.63 ± 0.11	0.91 ± 0.28	0.66 ± 0.15	0.87 ± 0.28	0.70 ± 0.16
V19-93/J1 <sup>c</sup>	1.25 ± 0.54	1.20 ± 0.43	1.27 ± 0.52	1.19 ± 0.45	1.61 ± 0.61	0.85 ± 0.19
V19-93/U <sup>a</sup>	0.02 ± 0.00	0.01 ± 0.00	0.02 ± 0.01	0.01 ± 0.00	0.02 ± 0.01	0.01 ± 0.00
V2-109/J2 <sup>b</sup>	0.20 ± 0.10	0.18 ± 0.07	0.26 ± 0.10	0.13 ± 0.03	0.17 ± 0.06	0.22 ± 0.10
V3-12/U <sup>c</sup>	0.01 ± 0.01	0.01 ± 0.00	0.02 ± 0.01	0.01 ± 0.00	0.01 ± 0.00	0.02 ± 0.01
V3-7/J1 <sup>b</sup>	0.11 ± 0.06	0.09 ± 0.03	0.13 ± 0.06	0.07 ± 0.01	0.07 ± 0.02	0.12 ± 0.06
V4-51/U <sup>a</sup>	0.00 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
V4-53/J5 <sup>c</sup>	0.34 ± 0.07	0.29 ± 0.04	0.32 ± 0.06	0.32 ± 0.06	0.27 ± 0.05	0.36 ± 0.06
V4-55/J5 <sup>c</sup>	3.17 ± 2.16	2.81 ± 1.47	2.79 ± 1.47	3.19 ± 2.15	4.73 ± 2.24	1.25 ± 0.57
V4-55/U <sup>c</sup>	0.17 ± 0.09	0.18 ± 0.06	0.16 ± 0.05	0.19 ± 0.10	0.24 ± 0.09	0.10 ± 0.04
V4-58/J1 <sup>c</sup>	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.01
V4-58/J4 <sup>a,b</sup>	0.07 ± 0.03	0.14 ± 0.05	0.07 ± 0.05	0.14 ± 0.05	0.09 ± 0.03	0.12 ± 0.06
V4-63/J5 <sup>a</sup>	0.13 ± 0.06	0.44 ± 0.27	0.23 ± 0.13	0.34 ± 0.26	0.28 ± 0.16	0.29 ± 0.25

V4-73/U <sup>b,c</sup>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
V4-86/J4 <sup>b</sup>	0.32 ± 0.22	0.56 ± 0.34	0.18 ± 0.10	0.7 ± 0.35	0.46 ± 0.24	0.42 ± 0.33
V4-86/U <sup>a</sup>	0.02 ± 0.01	0.04 ± 0.02	0.03 ± 0.02	0.03 ± 0.02	0.03 ± 0.02	0.03 ± 0.01
V4-91/J5 <sup>a</sup>	0.42 ± 0.10	0.25 ± 0.04	0.36 ± 0.11	0.32 ± 0.06	0.31 ± 0.08	0.36 ± 0.09
V5-39/J1 <sup>c</sup>	1.09 ± 0.65	0.85 ± 0.68	0.94 ± 0.62	1.00 ± 0.69	0.39 ± 0.27	1.55 ± 0.80
V5-48/J1 <sup>c</sup>	0.19 ± 0.10	0.11 ± 0.06	0.20 ± 0.10	0.11 ± 0.04	0.08 ± 0.02	0.22 ± 0.11
V6-14/U <sup>a</sup>	0.00 ± 0.00	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.01
V6-17/J5 <sup>c</sup>	0.26 ± 0.08	0.26 ± 0.09	0.28 ± 0.09	0.25 ± 0.07	0.20 ± 0.04	0.33 ± 0.10
V6-25/J1 <sup>b</sup>	0.26 ± 0.12	0.23 ± 0.12	0.15 ± 0.02	0.34 ± 0.15	0.25 ± 0.13	0.24 ± 0.11
V6-29/J4 <sup>b</sup>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
V7-33/U <sup>b</sup>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
V8-16/J1 <sup>c</sup>	0.06 ± 0.04	0.05 ± 0.02	0.05 ± 0.04	0.05 ± 0.02	0.03 ± 0.01	0.07 ± 0.04
V8-18/J4 <sup>b</sup>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
V8-21/J2 <sup>c</sup>	0.13 ± 0.05	0.12 ± 0.02	0.13 ± 0.04	0.12 ± 0.02	0.10 ± 0.02	0.15 ± 0.04
V8-21/U <sup>b,c</sup>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
V8-24/J4 <sup>c</sup>	0.22 ± 0.07	0.23 ± 0.12	0.21 ± 0.07	0.23 ± 0.12	0.15 ± 0.03	0.30 ± 0.12
V9-123/J5 <sup>c</sup>	0.04 ± 0.02	0.06 ± 0.02	0.05 ± 0.02	0.05 ± 0.03	0.03 ± 0.01	0.07 ± 0.03

U – Undetermined J-gene segment

a - P<0.05 for a main effect of AOS

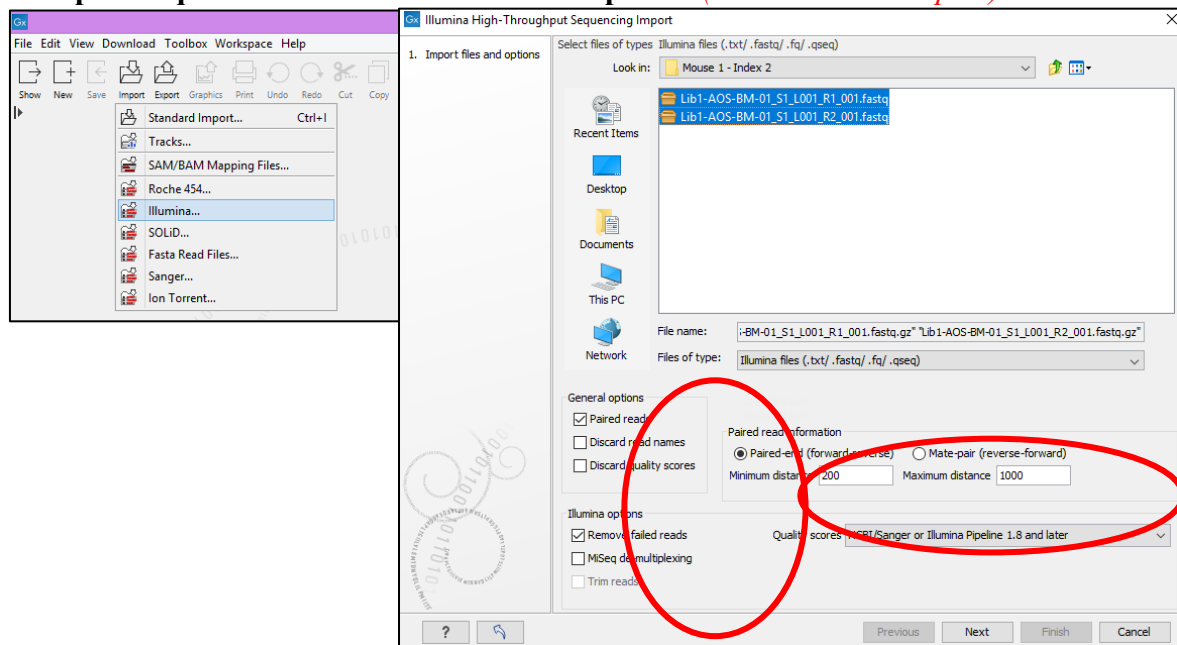
b -  $P < 0.05$  for a main effect of TT

c -  $P < 0.05$  for a main effect of CpG

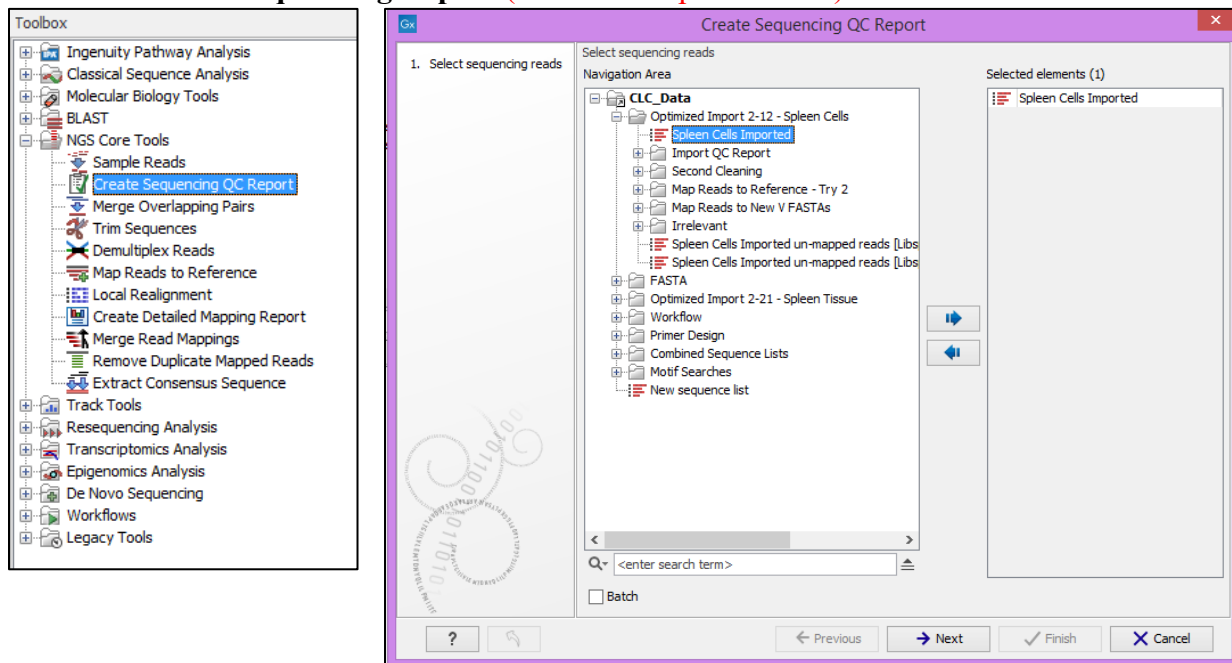
## Appendix B - Standard operating procedures for repertoire analysis

### Part I – Illumina Sequencing Import to CLC and Quality Cleaning

#### 1. Import sequences from Illumina reads as paired *(Save as Initial Import)*

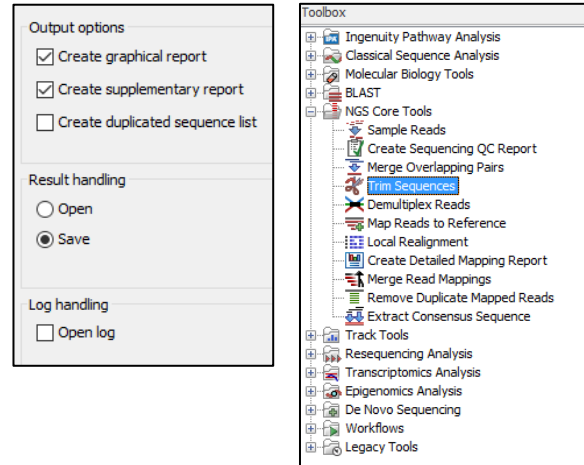


#### 2. Generate Sequencing Report *(Save into Reports Folder)*

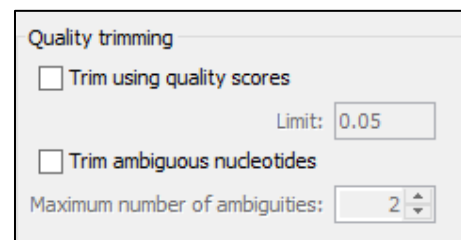


## Remove 12 5' bps

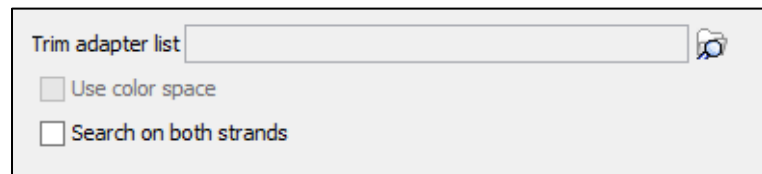
- Select the trim sequences tool
- Select the imported reads to clean



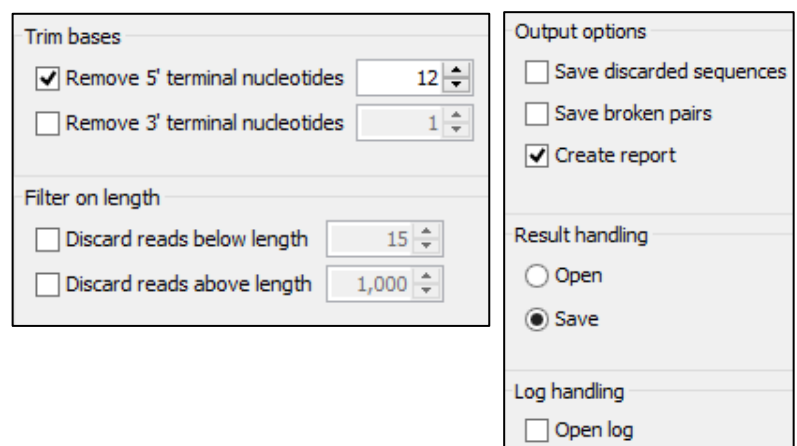
## Deselect quality and ambiguous reads



- Deselect adaptor trim



- Select removal of 5' bases
- Set to remove 12 5' bases
- Create report and save with appropriate name/folder (*Save as "-12 bp"*)

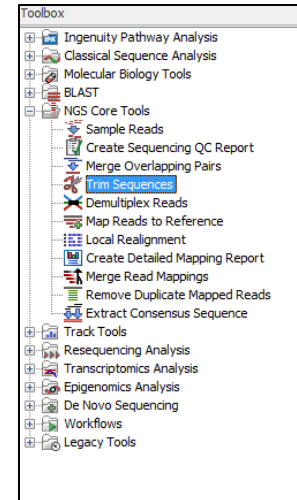


### 3. Generate QC Sequencing Report

- See item 2
- Select newly generated -12bp sequencing set

### 4. Quality Trim/Adaptor Trim/Remove Sequences

- Select the trim sequences tool
- Select the -12bp sequence set



- Select and set quality trim to 0.03 and trim ambiguous nucleotides at 2

The image shows the configuration interface for the 'Quality trimming' tool. It has two checked options: 'Trim using quality scores' with a 'Limit' of 0.03, and 'Trim ambiguous nucleotides' with a 'Maximum number of ambiguities' of 2.

- Select the appropriate adaptor list  
(this number is provided by our sequencing personnel)

The image shows the configuration interface for the 'Adaptor List' tool. It includes a 'Trim adapter list' dropdown set to 'Adapter List', a 'Use color space' checkbox, and a checked 'Search on both strands' checkbox. Below is a 'Preview' section with a table showing the results of a search on 1,000 reads.

Name	Matches found	Reads discarded	Nucleotides removed	Avg. length
Universal	15	0	276	216.7
Index 5	30	1	515	216.6
Index 19	30	1	515	216.6
PolyA-Removal	66	0	399	216.5
Short Index 5	9	0	41	216.9
Short Index 19	15	0	60	216.9
Total	129	1	1,259	215.9

- e. Select discard any sequences less than 40bps
- f. Deselect the remove 5' terminal nucleotides box

Trim bases	
<input type="checkbox"/> Remove 5' terminal nucleotides	<input type="text" value="12"/>
<input type="checkbox"/> Remove 3' terminal nucleotides	<input type="text" value="1"/>
Filter on length	
<input checked="" type="checkbox"/> Discard reads below length	<input type="text" value="40"/>
<input type="checkbox"/> Discard reads above length	<input type="text" value="1,000"/>

- g. Create a report and save with appropriate name/folder *(Save as Mouse X – Quality Cleaned (paired))*

Output options
<input type="checkbox"/> Save discarded sequences
<input type="checkbox"/> Save broken pairs
<input checked="" type="checkbox"/> Create report
Result handling
<input type="radio"/> Open
<input checked="" type="radio"/> Save
Log handling
<input type="checkbox"/> Open log

**THIS DATA FILE WILL BECOME YOUR FILE USED FOR ALL SUBSEQUENT DATA**

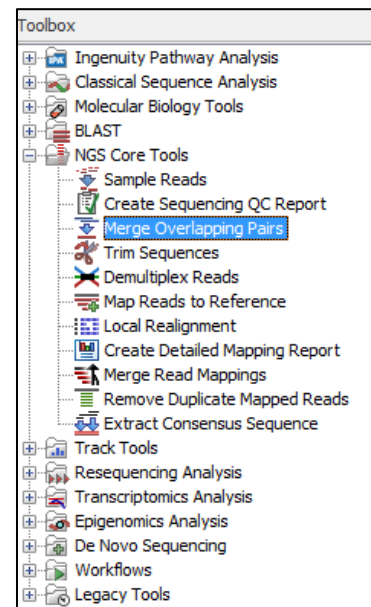
Once the mapping is confirmed, the initial import and -12bp data files can be deleted

## 5. Generate Sequencing Report

- a. See item 2
- b. Select newly generated “cleaned” sequencing set

## 6. Merge Overlapping Reads

- Select the merge overlapping reads tool
- Select the cleaned sequencing data set

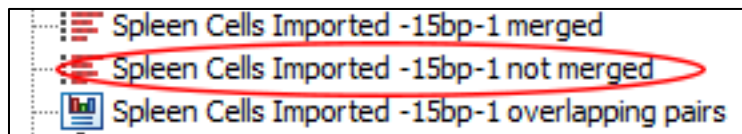


- Use default settings

Alignment scores	
Mismatch cost	2
Gap cost	3
Maximum unaligned end mismatches	0
Minimum score	8

- Generate report and save with appropriate name/folder  
(Save as "Mouse X – Quality Cleaned (paired)")

- Discard any unmerged reads



Output options	
<input checked="" type="checkbox"/>	Create report
Result handling	
<input type="radio"/>	Open
<input checked="" type="radio"/>	Save
Log handling	
<input type="checkbox"/>	Open log

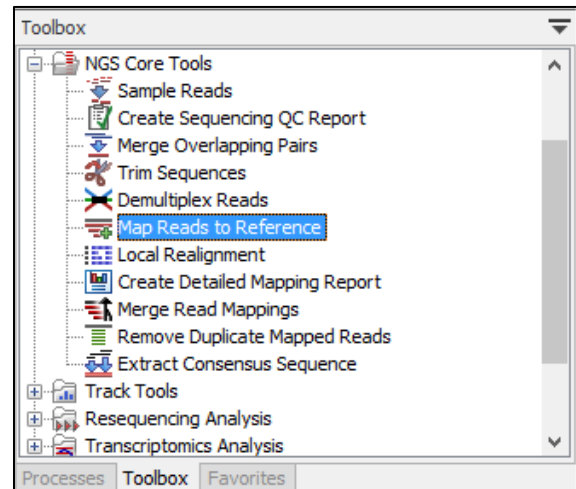


## Part II – Mapping Cleaned Sequences to V Gene Segments and Locus for

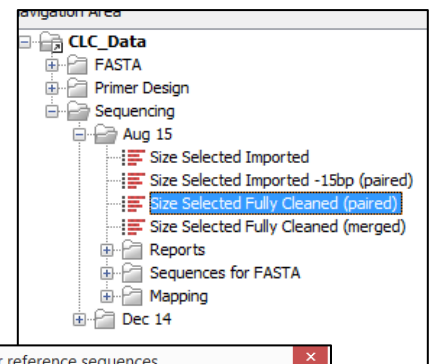
### IMGT Submission

#### 1. Map paired reads to “V alleles”

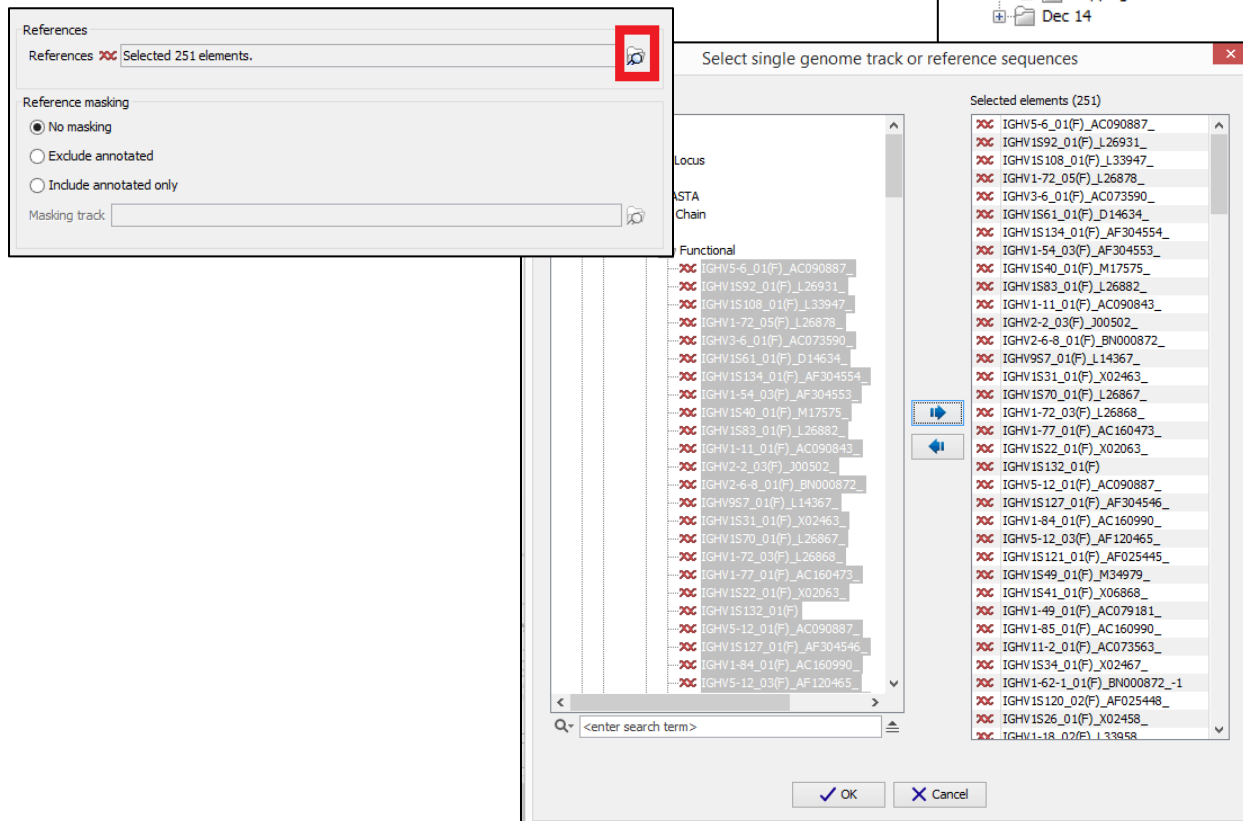
- Select the map reads to reference tool



- Select the “cleaned” data file (paired)



- Select all V allele FASTA files



- d. Use default mapping settings
- e. Select Output options and save with appropriate name/folder (ex. *“Cells Dec 14 V Aligned Paired Mapped”*)
- f. Select create stand-alone read mappings
- g. Select create report

Mapping options		Output options	
Read alignment		<input type="radio"/> Create reads track <input checked="" type="radio"/> Create stand-alone read mappings <input checked="" type="checkbox"/> Create report <input type="checkbox"/> Collect un-mapped reads	
Mismatch cost	<input type="text" value="2"/>		
<input checked="" type="radio"/> Linear gap cost <input type="radio"/> Affine gap cost			
Insertion cost	<input type="text" value="3"/>		
Deletion cost	<input type="text" value="3"/>		
Insertion open cost	<input type="text" value="6"/>		
Insertion extend cost	<input type="text" value="1"/>		
Deletion open cost	<input type="text" value="6"/>		
Deletion extend cost	<input type="text" value="1"/>		
Length fraction	<input type="text" value="0.5"/>		
Similarity fraction	<input type="text" value="0.8"/>		
<input type="checkbox"/> Global alignment <input checked="" type="checkbox"/> Color space alignment Color error cost <input type="text" value="3"/>		Result handling <input type="radio"/> Open <input checked="" type="radio"/> Save	
<input checked="" type="checkbox"/> Auto-detect paired distances		Log handling <input type="checkbox"/> Open log	
Non-specific match handling			
<input type="radio"/> Map randomly <input checked="" type="radio"/> Ignore			

## 2. Map merged reads to “V alleles”

- a. Follow the above steps using the cleaned merged file.

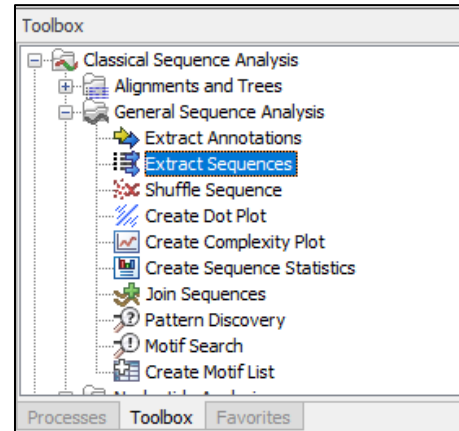
## 3. Map Merged and Paired Reads to “IgH Locus”

- a. Follow instructions as above but select the IgH locus FASTA file for 1c for both paired and merged files

## 4. Repeat steps 1 to 3 for any additional mapping (IgK, IgL)

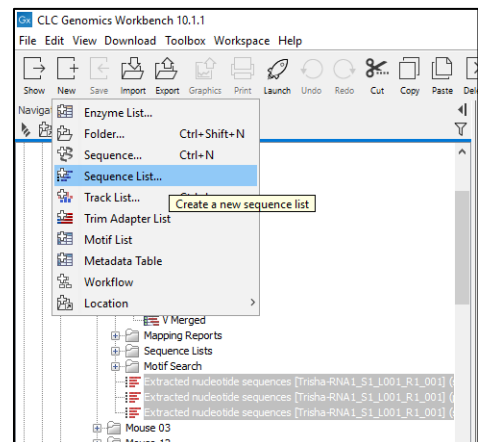
**5. Extract the sequences that have mapped and save with appropriate name/folder (ex. “Cells Dec 14 V Aligned Paired”)**

- Select all of the mapping results
- Select the extract sequences tool and extract to new sequence list



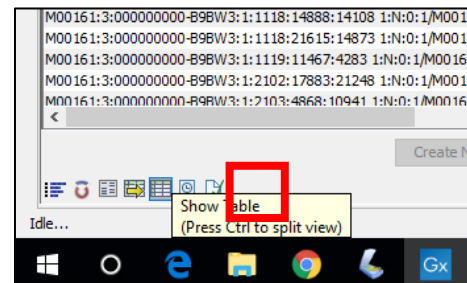
**6. Combine Lists**

- Highlight the extracted sequences and select new sequence list

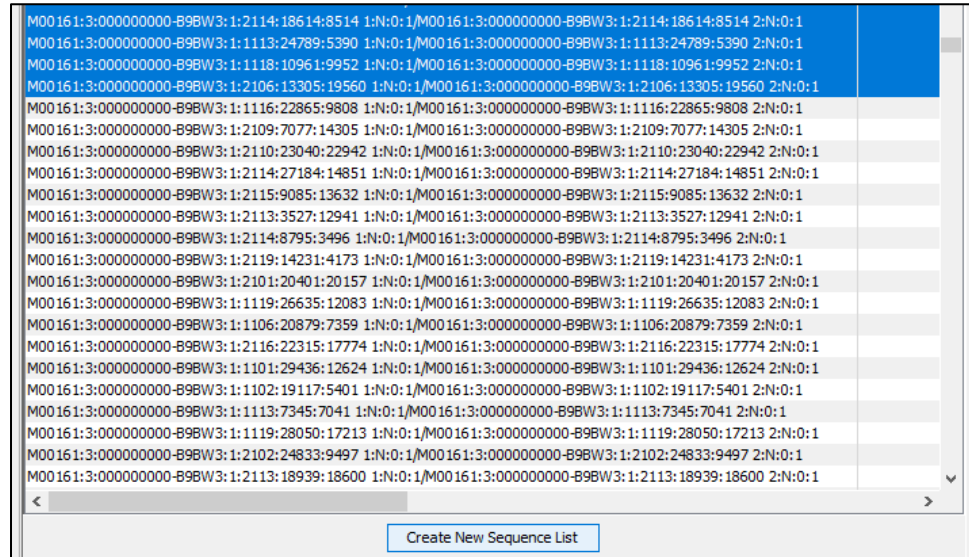


**7. Break the mapped reads into sets of <500k sequences for IMGT Submission**

- Double click the new sequence list to show the results
- Use the show table button to obtain the list of sequences
- The total number of sequences in the list will be displayed at the top of the table

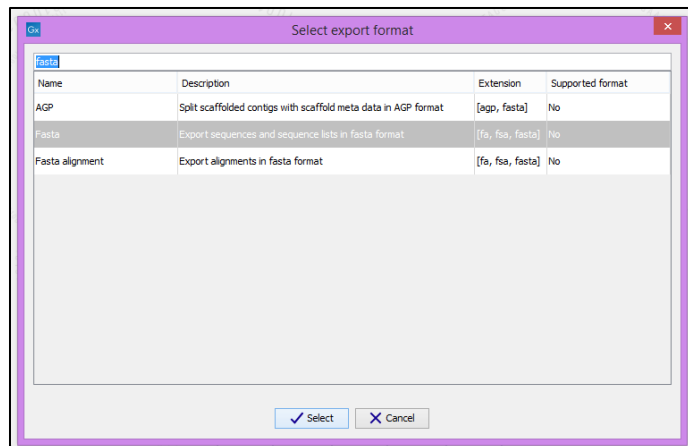
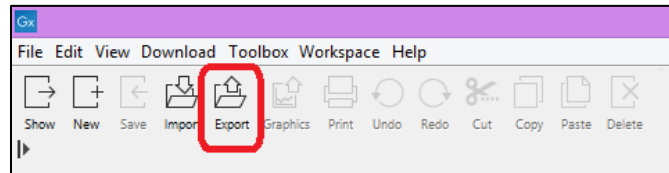


- d. If >500k sequences, select multiple sequences using the shift button and press the create new sequence list button.
- e. Check that this new list has <500k sequences
- f. Save the list as “mouse X – pt x”



## 8. Export FASTA File

- a. Select the sequence lists to export
- b. Select the export button
- c. Select FASTA format for export
- d. Save to location (on the computer, rather than in CLC)

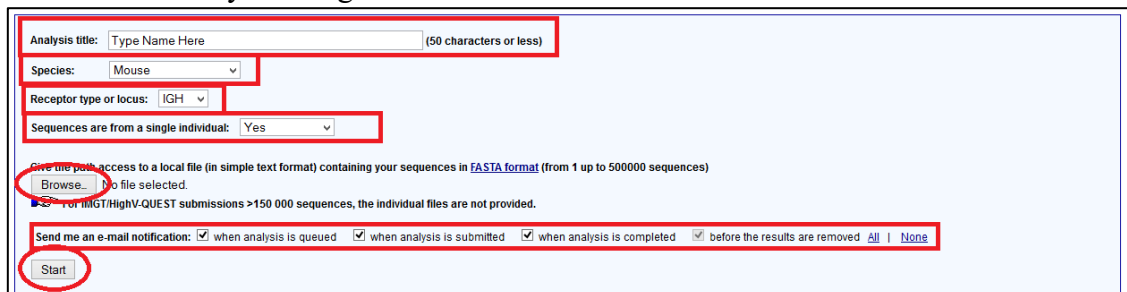
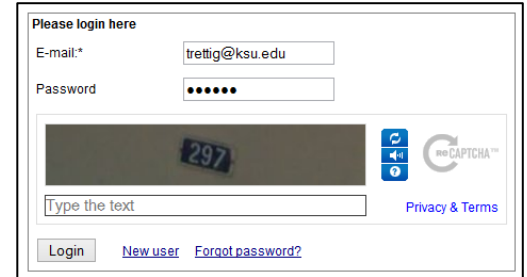
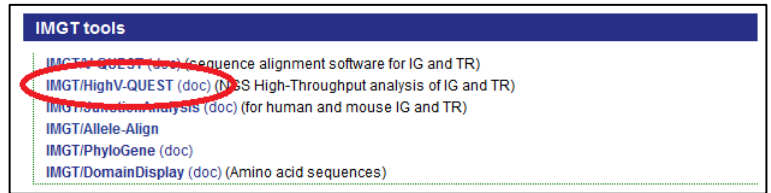


**THIS IS THE FILE THAT WILL BE SUBMITTED TO IMGT FOR FUTHER PROCESSING**

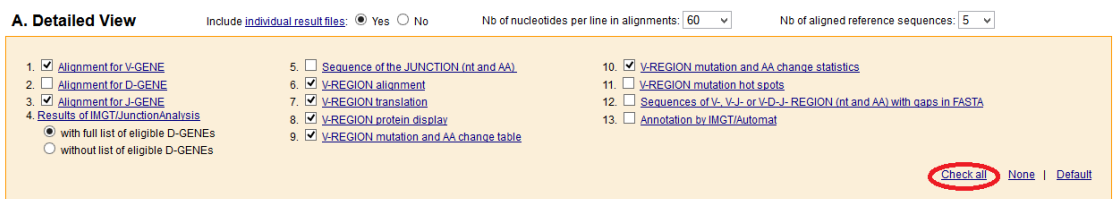
## Part III – Submission of FASTAs to IMGT and retrieval of IMGT Results

### 1. Submission of files for V, D, J identification

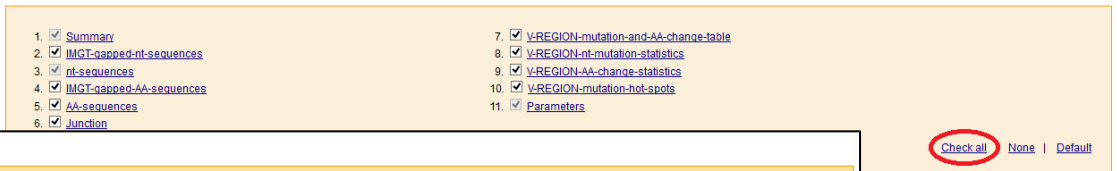
- Select the High-V quest link
- Log in
- Name file (files will be deleted after a few weeks, so a complicated name is not required)
- Select species *Mus musculus* (house mouse)
- Select locus as IGH
- Select yes for sequences are from a single individual
- Select exported FASTA file from CLC
- Select e-mail notifications
- In Detailed View – select “check all”
- This only applies to <150k sequences
- In Files in CSV – select “check all”
- Leave Advanced parameters as default
- Submit file by clicking “Start”



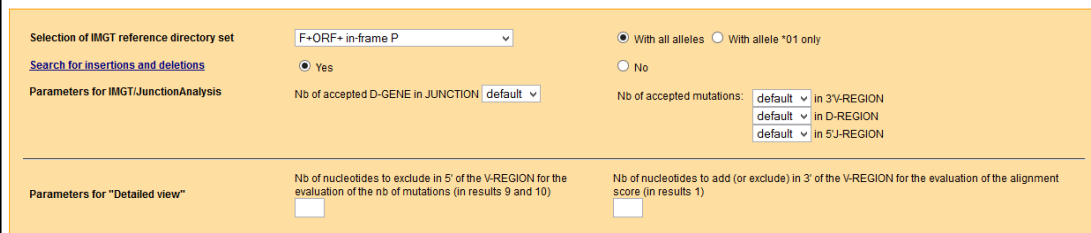
#### Display results



#### B. Files in CSV
























#### Advanced parameters



## 2. Download data from IMGT

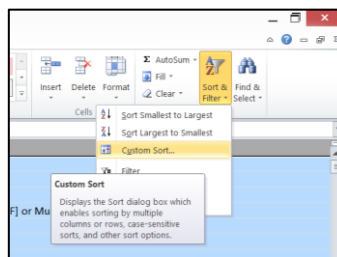
- Login to the HighV-Quest page
- Select analysis history
- Download files
- Save files in folder “1 – IMGT Output”

Here is information about your analysis history							
Title	User	Status	Submission date	Nb of sequences	IMGT/V-QUEST reference directory		Actions
					Species	Receptor type (or locus)	
SS Paired		queued	2015-08-26 15:14:35 <a href="#">CET</a>	74868	Mus_musculus	IGH	Not completed 
SS Merged		queued	2015-08-26 15:11:39 <a href="#">CET</a>	33291	Mus_musculus	IGH	Not completed 
SS All Sequences		completed	2015-08-25 21:06:17 <a href="#">CET</a>	74868	Mus_musculus	IGH	(33.63 MB, 74884 files)  
SS V Gene		completed	2015-08-25 21:04:14 <a href="#">CET</a>	28926	Mus_musculus	IGH	(26.88 MB, 28940 files)  
SS IgH		completed	2015-08-25 21:03:04 <a href="#">CET</a>	46130	Mus_musculus	IGH	(5.17 MB, 46145 files)  
SS IgH Paired (real)		completed	2015-08-25 20:56:09 <a href="#">CET</a>	46130	Mus_musculus	IGH	(5.16 MB, 46145 files)  
SS IgH Paired		completed	2015-08-25 20:18:52 <a href="#">CET</a>	48019	Mus_musculus	IGH	(3.53 MB, 48034 files)  
SS IgH Merged		completed	2015-08-25 20:17:41 <a href="#">CET</a>	22709	Mus_musculus	IGH	(2.46 MB, 22723 files)  
SS V Merged		queued	2015-08-25 18:38:18 <a href="#">CET</a>	13329	Mus_musculus	IGH	Not completed 
SS V Paired		completed	2015-08-25 18:36:53 <a href="#">CET</a>	29046	Mus_musculus	IGH	(27.06 MB, 29060 files)  
All Tissue		completed	2015-06-08 17:48:06 <a href="#">CET</a>	58950	Mus_musculus	IGH	(22.44 MB, 58965 files)  
All Cells		completed	2015-06-08 17:46:08 <a href="#">CET</a>	39398	Mus_musculus	IGH	(17.88 MB, 39412 files)  
Note: this table is not dynamic, please click <a href="#">here</a> to refresh the table							

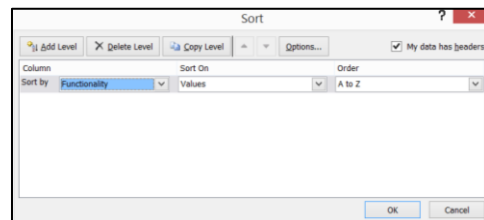
## Part IV – Determination of Constant Region (IgH Only)

### 1. Generate a new FASTA file for

- Open the 1\_Summary file the IMGT output in Excel and delineate on tab
- Use the sort function to select only sequence IDs and Sequences from functionally productive antibodies
- Repeat this step until you have compiled all of the productive sequences for the mouse



CLC from



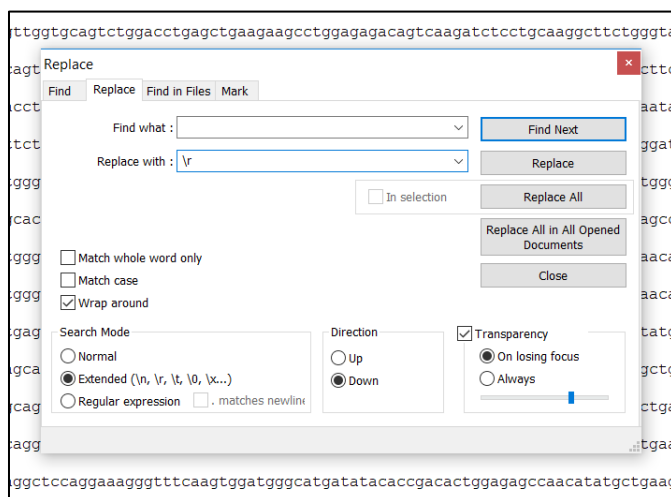
### 2. Generate a new worksheet with the following data

- Column A - >
- Column B - Sequence ID (copy and paste)
- Column C - =A2&B2
- Column D - Sequence (copy and paste)

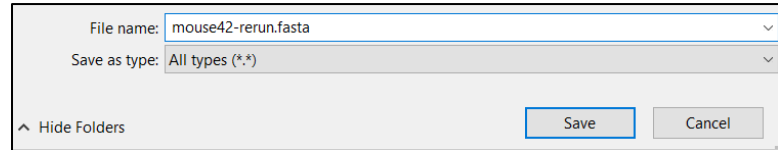
	A	B	C
1			
2	>	2111:11321:19496_1:N:0	>2111:11321:19496_1:N:0
3	>	2106:10911:13957_1:N:0	>2106:10911:13957_1:N:0
4	>	2117:14975:6400_1:N:0	>2117:14975:6400_1:N:0
5	>	1112:22502:7631_2:N:0	>1112:22502:7631_2:N:0
6	>	2102:13533:17668_1:N:0	>2102:13533:17668_1:N:0
7	>	2105:23230:7358_1:N:0	>2105:23230:7358_1:N:0
8	>	1114:24265:15785_1:N:0	>1114:24265:15785_1:N:0

### 3. Format the FASTA File

- Copy columns C and D to notepad++
- Highlight the gap between the sequence ID and the sequence
- Use the replace tool (Control + H)
- Replace the highlight with \r
  - Make sure the extended search mode is on

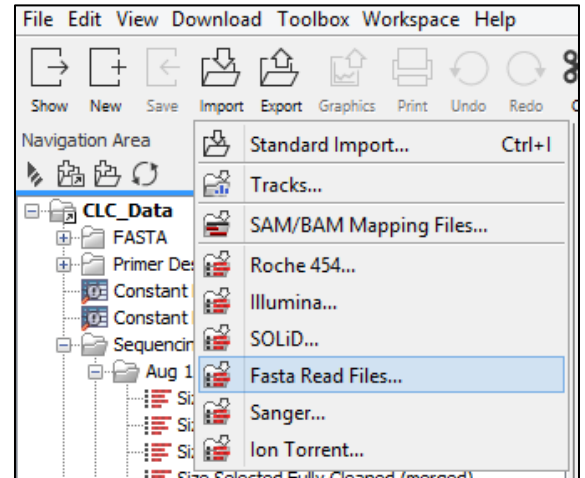


#### 4. Save as a .fasta file



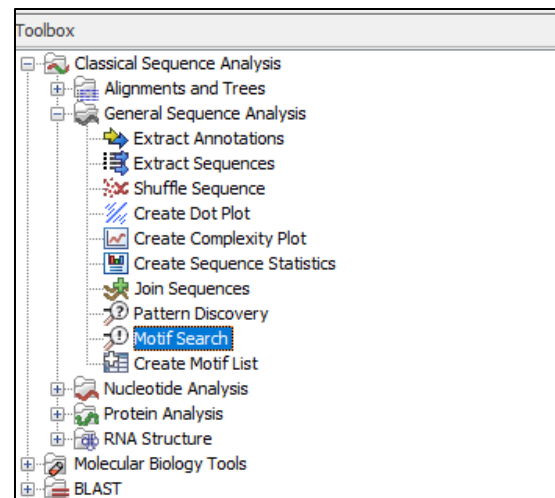
#### 5. Import the new FASTA file in CLC

- a. Select the import fasta read files tool

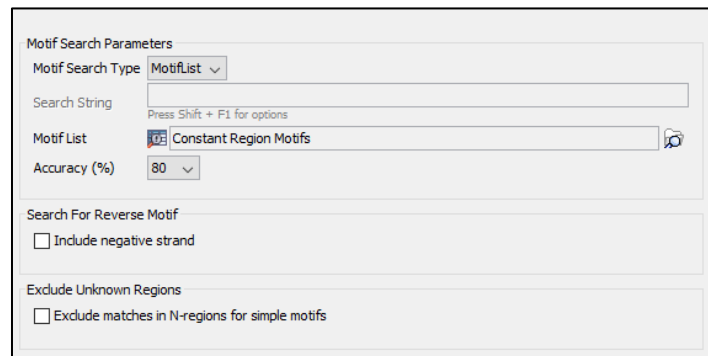


#### 6. Run the motif search

- a. Select the file you wish to search and then the motif search tool
- b. Select next on the first screen to confirm file

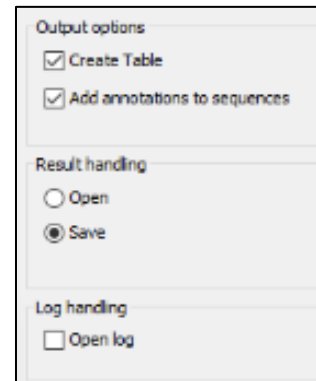


- c. Select motiflist
- d. Select the constant region motif list
- e. Use 80% accuracy





f. Check Create Table and Add Annotations to Sequences



Output options

☒ Create Table

☒ Add annotations to sequences

Result handling

☐ Open

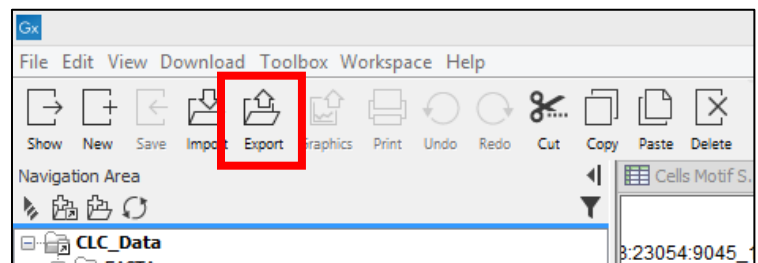
☒ Save

Log handling

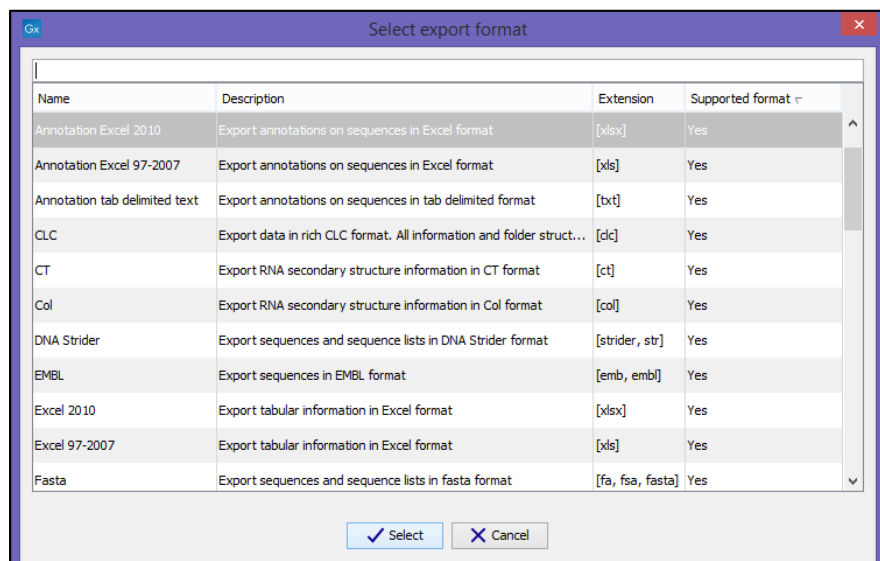
☐ Open log

## 7. Export the motif data file to Excel

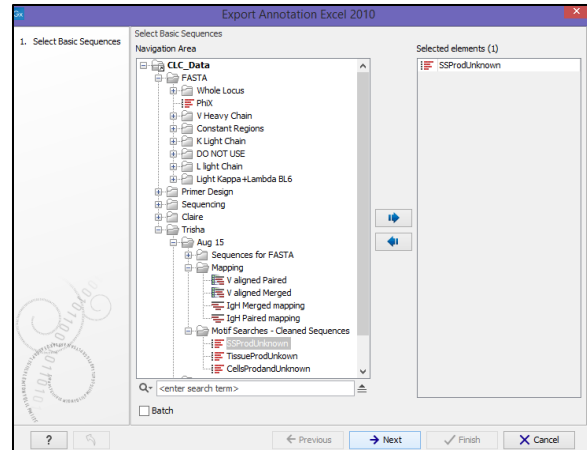
a. Select the export button



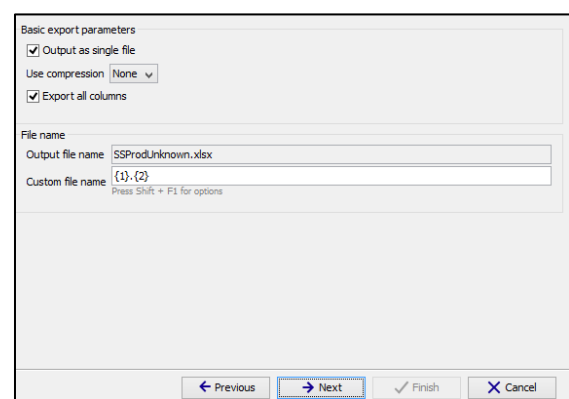
b. Select the format as annotation Excel 2010



- c. Select your files and hit next
  - i. You may combine multiple motif search tables into one Excel file



- d. Name your file and select next, then finish



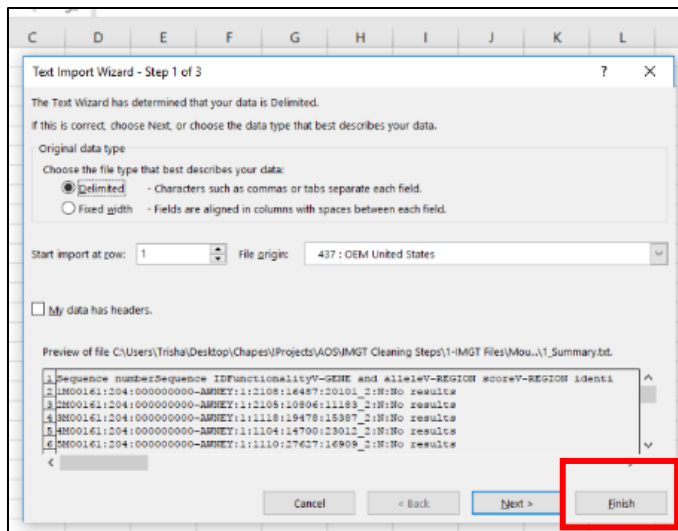
**SAVE POINT – Save this file as “Mouse X” in Folder “Motif Search Results”**

## Part V – Reformatting of IMGT Data, Duplicate Removal, and Functionality

### Assignment

1. Combine the columns from the multiple IMGT outputs outlined in Appendix 6 for productive and unknown reads only.

- a. Open the files extracted from the zipped IMGT output in Excel
- b. Select the finish button to properly format the data
- c. Make sure that you carefully maintain the correct sequence IDs with the correct results



2. Reset the sequence numbers

- a. Fill down the first three numbers
- b. Select all three cells
- c. Double click the small square at the bottom right of the bottom number to fill down

Sequence	Sequence	Functional V-GENE	ar V-RE
1	M00161:2:unknown	( Musmus IC	
2	M00161:2:unknown	( Musmus IC	
3	M00161:2:unknown	( Musmus IC	
787	M00161:2:unknown	( Musmus IC	
788	M00161:2:unknown	( Musmus IC	
789	M00161:2:unknown	( Musmus IC	
790	M00161:2:unknown	( Musmus IC	
791	M00161:2:unknown	( Musmus IC	

**SAVE POINT – Save this file as “Mouse X – Raw Data” in Folder “2 – IMGT Raw Data”**

### 3. Create the truncated ID column

- Duplicate the Sequence ID column and rename Trunc ID
- Add one empty column to the right of the Trunc ID column

1	A	B	C	D	E
1	Sequence	Sequence ID	Trunc ID		Functional V-G
2	1	M00161:204:000000000-AWNEY:1:1118:13432:12043_2:N	M00161:204:000000000-AWNEY:1:1118:13432:12043_2:N		unknown (Mus
3	2	M00161:204:000000000-AWNEY:1:2105:24964:12318_2:N	M00161:204:000000000-AWNEY:1:2105:24964:12318_2:N		unknown (Mus
4	3	M00161:204:000000000-AWNEY:1:2116:17468:4600_2:N:0	M00161:204:000000000-AWNEY:1:2116:17468:4600_2:N:0		unknown (Ma
5	787	M00161:204:000000000-AWNEY:1:2101:3769:17582_2:N:0	M00161:204:000000000-AWNEY:1:2101:3769:17582_2:N:0		unknown (Ma
6	788	M00161:204:000000000-AWNEY:1:1104:7310:12989_2:N:0	M00161:204:000000000-AWNEY:1:1104:7310:12989_2:N:0		unknown (Ma
7	789	M00161:204:000000000-AWNEY:1:1106:13834:6405_2:N:0	M00161:204:000000000-AWNEY:1:1106:13834:6405_2:N:0		unknown (Ma
8	790	M00161:204:000000000-AWNEY:1:1109:7421:21630_2:N:0	M00161:204:000000000-AWNEY:1:1109:7421:21630_2:N:0		unknown (Ma

- Select the Trunc ID column and use the find and replace tool to replace the Illumina specific code associated with the run. Leave the “Replace With” box empty.

Find and Replace

Find Replace

Find what: M00161:204:000000000-AWNEY:1:

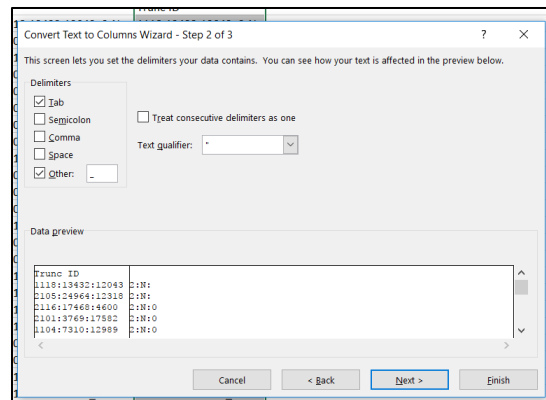
Replace with:

Options >>

Replace All Replace Find All Find Next Close

- Use the text to columns tool to remove the trailing information

- i. Select delimited
- ii. Delimit using “other” and type an underscore ( \_ )
- iii. Select finished



- e. Delete the column containing the trailing IDs

#### 4. Remove “(see comment)” from the entire worksheet

- a. Make sure to include the space in the removal

### IGH ONLY STEPS – light chain proceed to save point

#### 5. Copy in the CLC generated motif search as in step 1

- a. Organize all of the results into two columns on a single page
  - i. A = Sequence
  - ii. B = Description
  - iii. Any other columns may be discarded

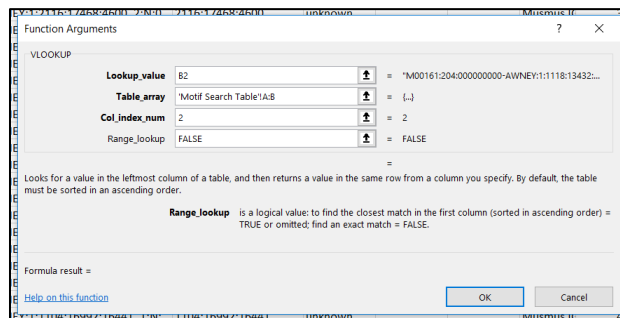
#### 6. Align the motifs to their corresponding sequence ID

- a. In the raw data tab create two new columns
  - i. One labeled Constant Region
  - ii. One blank

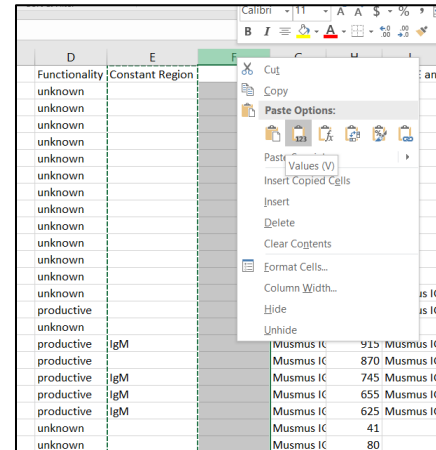
- b. In the Constant Region column use the VLOOKUP function to ID the constant regions to their sequence ID

- i. Lookup value = the sequence ID on the raw data page

- ii. Table array = the two columns where the data is stored in the motif search page



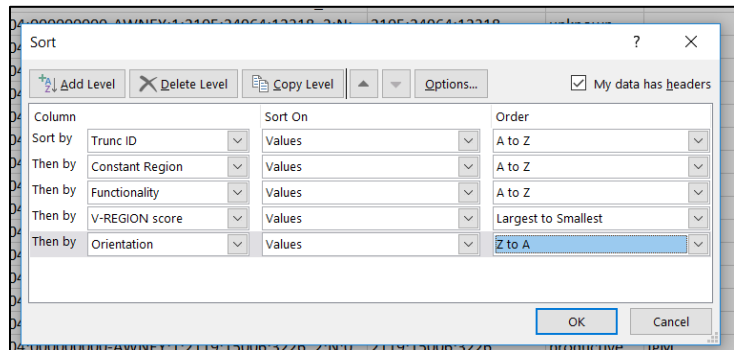
- iii. Col\_Index\_num = the column's data that you want reported (in this case the description)
- iv. Range\_lookup = false (only return this data if the lookup value is found)
- c. Add an "IFERROR" function to the outside of the established VLOOKUP formula
  - i. **=IFERROR(VLOOKUP('Raw Data'!B2,'Motif Search'!A:B,2,FALSE),"" )**
    1. This tells excel to return a blank cell if a constant region isn't found
- d. Fill the equation down using the box in the bottom right corner of the cell (*Note: this is a time/processor heavy equation step*)
- e. **CONFIRM THAT THE EQUATION HAS FILLED ALL THE WAY DOWN**
- f. Use the paste values function to paste the VLOOKUP results (*Note: this is a time/processor heavy step*)
- g. Delete the column containing the VLOOKUP results



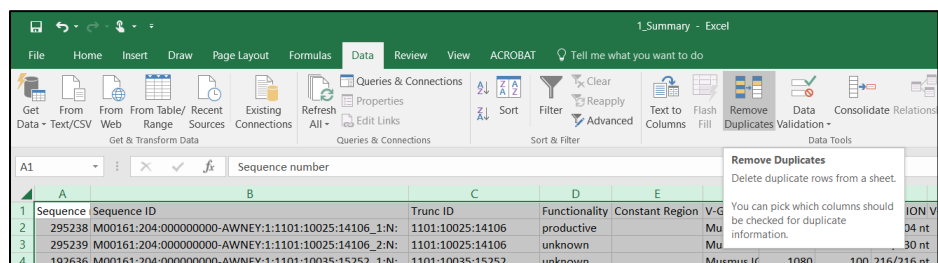
**SAVE POINT – Save this file as “Mouse X – For Sort” in Folder “3 – For Sort”**

## 7. Perform sort on sequences

- a. Trunc ID (A → Z)
- b. Constant Region (A → Z)
- c. Functionality (A → Z)
- d. V-Region Score (Largest → Smallest)
- e. Orientation (Z → A)



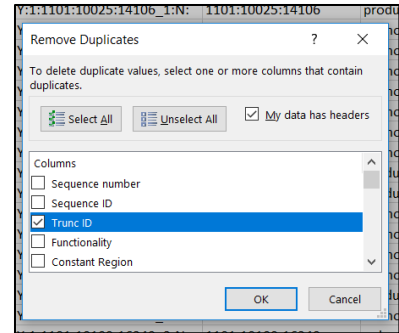
## 8. Remove Duplicates



- a. Highlight the entire sheet
- b. Use the remove duplicate tool

- c. Select remove duplicates based on the Trunc ID Column

**SAVE POINT – Save this file as “Mouse X –Dups Removed” in Folder “4 – Dups Removed”**



**9. Repeat steps 7 through 9 with the “For Sort” data file to confirm the same duplicate removal results were accurate**

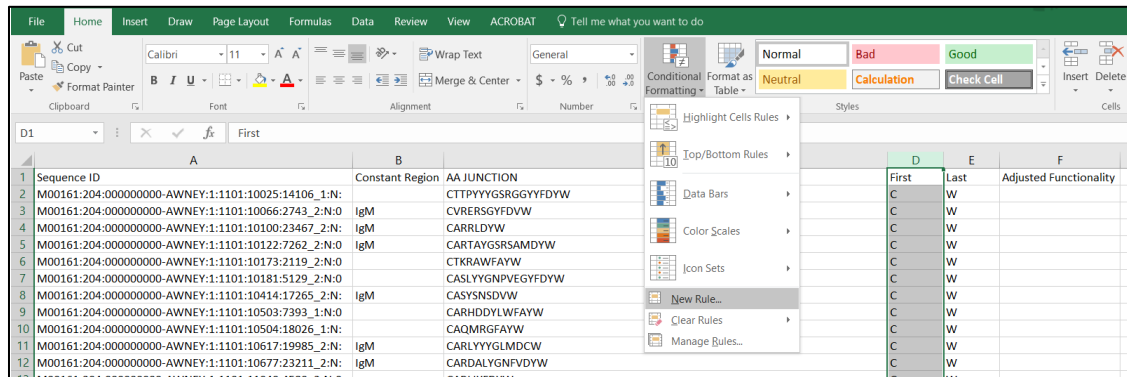
- a. If results agree, continue, if not continue to determine correct duplicate removal results
- b. Delete the Trunc ID column and the V-Region Score column in the duplicates removed file

**10. Adjust functionality**

- a. Create new spreadsheet
  - i. A = Sequence ID
  - ii. B = Constant Region (IgH Only)
  - iii. C = AA Junction
  - iv. D = First
  - v. E = Last
  - vi. F = Adjusted Functionality
- b. Remove “ (see Functionality comment)” from the AA Junction column using the replace tool as in step 4. Be cuse to include the space before the parentheses.
- c. Generate the First column
  - i. =LEFT(C2,1) where C2 is the AA Junction sequence
- d. Generate the Last column
  - i. =RIGHT(C2,1) where C2 is the AA Junction sequence
- e. Fill results down

	A	B	C	D	E	F
1	Sequence ID	Constant Region	AA JUNCTION	First	Last	Adjusted Functionality
2	M00161:204:000000000-AWNEY:1:1101:10025:14106_1:N:		CTTPYYGSRGGYFDYW			
3	M00161:204:000000000-AWNEY:1:1101:10066:2743_2:N:0	IgM	CVRRSGYFDYW			
4	M00161:204:000000000-AWNEY:1:1101:10100:23467_2:N:	IgM	CARRLDYW			
5	M00161:204:000000000-AWNEY:1:1101:10122:7262_2:N:0	IgM	CARTAYGSRAMDYW			
6	M00161:204:000000000-AWNEY:1:1101:10173:2119_2:N:0		CTKRAWFAYW			
7	M00161:204:000000000-AWNEY:1:1101:10181:5129_2:N:0		CASLYGNPVEGYFDYW			
8	M00161:204:000000000-AWNEY:1:1101:10414:17265_2:N:	IgM	CASYNSDQVW			
9	M00161:204:000000000-AWNEY:1:1101:10503:7393_1:N:0		CARHDDYLWFAYW			
10	M00161:204:000000000-AWNEY:1:1101:10504:18026_1:N:		CAQMRGFAYW			
11	M00161:204:000000000-AWNEY:1:1101:10617:19985_2:N:	IgM	CARLYYSLMDCW			
12	M00161:204:000000000-AWNEY:1:1101:10677:23211_2:N:	IgM	CARDALYGNFVDYW			
13	M00161:204:000000000-AWNEY:1:1101:11040:4580_2:N:0		CARLYFDYW			
14	M00161:204:000000000-AWNEY:1:1101:11111:24720_1:N:		CARYYYGSGYW			

f. Select one column and use the conditional formatting tool to ID non-standard



motifs

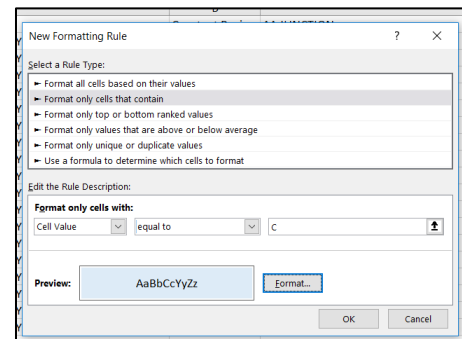
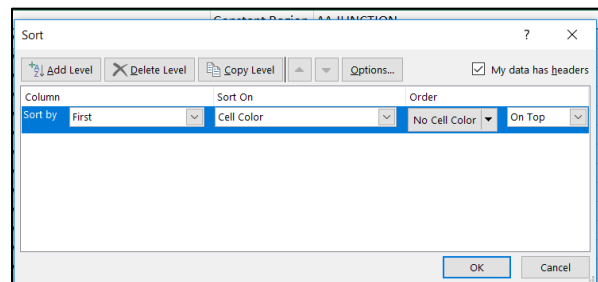
i. Heavy Chain Motif

1. First = C
2. Last = W

ii. Light Chain Motif

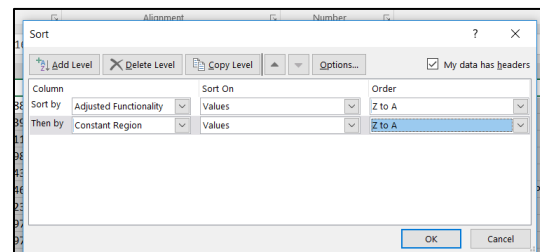
1. First = C
2. Last = F

- g. Set the new rule to format cells that do not contain equal to the motif and set the format to a color (doesn't matter which color)
- h. Repeat for the other part of the motif
- i. Fill the "adjusted functionality" column with "productive"
- j. Sort the data by cell color in the first column with the no cell color cells on top
- k. Rows that do not fit the motif are labeled "Unknown"
- l. Repeat for the "last" column labeling anything that does not fit the motif as "Unknown"



## IGH ONLY STEPS – light chain proceed to step 11

- m. Sort the cells
  - i. Adjusted functionality (Z → A)
  - ii. Constant Region (Z → A)
- n. Any unknown sequence that has a constant region of IgG, IgE, or IgA should be switched to productive adjusted functionality.
  - i. IgM and IgD constant regions should remain unknown





72	M00161:204:000000000-AWNEY:1:2114:24626:17225_1:N:	IgM	WMRYGSSYWFYDVW	W	W	Unknown
73	M00161:204:000000000-AWNEY:1:2115:8927:18892_2:N:0	IgM	SARRAYDYDAHYAMDYVW	S	W	Unknown
74	M00161:204:000000000-AWNEY:1:2118:9819:11685_1:N:0	IgM	WARDYYW	W	W	Unknown
75	M00161:204:000000000-AWNEY:1:1112:24099:21499_2:N:	IgG	CARLYYYGLMDCC	C	C	Productive
76	M00161:204:000000000-AWNEY:1:2105:6169:13078_2:N:0	IgG	CARWGYRGGYFVVC	C	C	Productive
77	M00161:204:000000000-AWNEY:1:2111:8198:21563_2:N:0	IgG	CARSTYCNYPQGYFYVC	C	C	Productive
78	M00161:204:000000000-AWNEY:1:1105:3623:9027_1:N:0:	IgG	GAQMRGFAYW	G	W	Productive
79	M00161:204:000000000-AWNEY:1:1107:17685:5846_1:N:0	IgG	YGSRGYYFDYW	Y	W	Productive
80	M00161:204:000000000-AWNEY:1:2111:25384:7799_1:N:0	IgG	YGSRGYYFDYW	Y	W	Productive
81	M00161:204:000000000-AWNEY:1:2105:25867:8382_2:N:0	IgD	CAQKGGDVGMDDYC	C	C	Unknown
82	M00161:204:000000000-AWNEY:1:2102:18146:22211_2:N:	IgD	RTRVGVWYFVW	R	W	Unknown
83	M00161:204:000000000-AWNEY:1:2117:21153:7722_2:N:0	IgA	CARDSWAFAYC	C	C	Productive

## 11. Reassign the Adjusted Functionality

- Created 2 new columns, one labeled adjusted functionality and one blank
- Use the VLOOKUP function to reassign the correct functionality to the sequence ID

Lookup_value	B2	=	"M00161:204:000000000-AWNEY:1:1101:10025:...
Table_array	Sheet5!A:F	=	{...}
Col_index_num	6	=	6
Range_Lookup	FALSE	=	FALSE
= "Productive"			

- Lookup\_value = Sequence ID on the main sheet
  - Table\_array = The entire table in the adjusted functionality sheet
  - Col\_index\_num = 6 ("adjusted functionality" column)
  - Range\_Lookup = FALSE
- Add an IFERROR function to keep unknown functionality as unknown
    - =**IFERROR**(VLOOKUP(B2,Sheet5!A:F,6,FALSE),"Unknown")
  - Spot check to make sure that some of the productive cells have become unknown in the "adjusted functionality" column
    - Use of a capitol letter for the adjusted functionality helps make this check easier, IMGT's output uses a lowercase letter.
  - Use the paste values in 6F to copy and paste the column with the VLOOKUP in it into a blank column
  - Delete the old "functionality" column and the column containing the VLOOKUP equation

	A	B	C	D	
1	Sequence ID	Sequence ID	Adjusted Functionality	Constant Region	V-G
2	295238	M00161:204:000000000-AWNEY:1:1101:10025:14106_1:N:	Productive		Mu
3	192636	M00161:204:000000000-AWNEY:1:1101:10035:15252_1:N:	Unknown		Mu
4	254002	M00161:204:000000000-AWNEY:1:1101:10041:16805_1:N:	Unknown		Mu
5	174468	M00161:204:000000000-AWNEY:1:1101:10048:13224_1:N:	Unknown		Mu
6	62471	M00161:204:000000000-AWNEY:1:1101:10066:2743_2:N:0	Productive	IgM	Mu

**SAVE POINT – Save this file as “Mouse X –Final” in Folder “5 - Final”**

**This is the file you will use for all further analysis. Depending on the analysis required, it may be easiest to copy and paste the data out of this file that you need into new workbooks. Work to minimize the number of equations, conditional formatting, and pivot tables running in your workbooks. This will increase the speed at which Excel functions and reduce crashes. Once an equation is complete, copy and paste special (values) the data to make processing run smoother.**

## Part VI – Characterization of Data (V, D, J, Constant, Functionality, CDR3 Length)

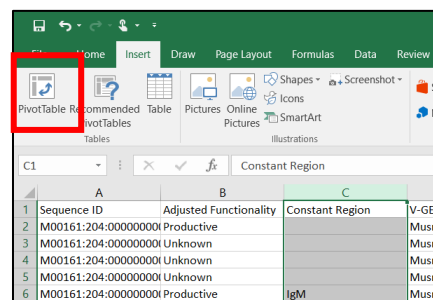
### 1. Copy the required data into a new workbook in Excel

- Copy the sequence ID data with the data you wish to analyze to a new workbook
- All reads are used for assessing V gene segment usage
- Productive only reads are used for assessing D, J, and CDR3 Length

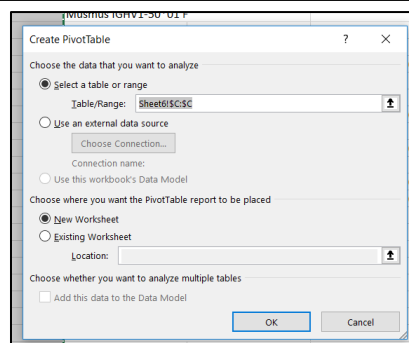
### 2. Calculate the number of productive and unknown functionality reads

	A	B	C	D	E	F	G
1	Sequence ID	Adjusted Functionality	Constant Region	V-GENE and allele	J-GENE and allele	D-GENE and allele	CDR3-IMGT length
2	M00161:204:000000000	Productive		Musmus IGHV14-1*01 F	Musmus IGHJ2*01 F	Musmus IGHD1-1*01 F	16
3	M00161:204:000000000	Unknown		Musmus IGHV1-50*01 F		X	
4	M00161:204:000000000	Unknown		Musmus IGHV1-7*01 F		X	
5	M00161:204:000000000	Unknown		Musmus IGHV1-64*01 F, or Musmus IGHV1-64*02 [F]		X	
6	M00161:204:000000000	Productive	IgM	Musmus IGHV10-3*01 F	Musmus IGHJ1*03 F	Musmus IGHD1-1*01 F	10
7	M00161:204:000000000	Unknown		Musmus IGHV1-18*01 F		X	
8	M00161:204:000000000	Unknown		Musmus IGHV1-80*01 F		X	
9	M00161:204:000000000	Productive	IgM	Musmus IGHV1-22*01 F	Musmus IGHJ4*01 F	Musmus IGHD1-2*01 F	6
10	M00161:204:000000000	Unknown		Musmus IGHV1-9*01 F		X	
11	M00161:204:000000000	Productive	IgM	Musmus IGHV2-2*01 F	Musmus IGHJ4*01 F	Musmus IGHD1-1*01 F	13
12	M00161:204:000000000	Unknown		Musmus IGHV1-50*01 F		X	
13	M00161:204:000000000	Productive		Musmus IGHV6-6*01 F	Musmus IGHJ3*01 F	Musmus IGHD3-2*01 F	8
14	M00161:204:000000000	Productive		Musmus IGHV1-26*01 F	Musmus IGHJ2*01 F, or	Musmus IGHD2-1*01 F	15
15	M00161:204:000000000	Unknown		Musmus IGHV1-50*01 F		X	
16	M00161:204:000000000	Unknown		Musmus IGHV1-78*01 F		X	

- Highlight the Constant Region column



- Insert a pivot table



- c. Set Adjusted Functionality as the row and values
  - i. Click on the adjusted functionality in the top box and drag it down for rows and value

Row Labels	Count of Adjusted Functionality
Productive	8610
Unknown (blank)	27677
<b>Grand Total</b>	<b>36287</b>

### 3. Copy and paste the working list of strain specific gene segments into the workbook

- a. See Appendix 8 to 11 for the gene segment lists into its own tab

	A	B	C	D	E
1	Sequence ID	V-GENE and allele			
2	M00161:204:000000000-AWNEY:1:1101:10025:14	Musmus IGHV14-1*01 F			
3	M00161:204:000000000-AWNEY:1:1101:10035:15	Musmus IGHV1-50*01 F			
4	M00161:204:000000000-AWNEY:1:1101:10041:16	Musmus IGHV1-7*01 F			
5	M00161:204:000000000-AWNEY:1:1101:10048:13	Musmus IGHV1-64*01 F, or Musmus IGHV1-64*02 [F]			
6	M00161:204:000000000-AWNEY:1:1101:10066:27	Musmus IGHV10-3*01 F, or Musmus IGHV10-3*02 F or Musmus IGHV10-3*03 F			
7	M00161:204:000000000-AWNEY:1:1101:10086:64	Musmus IGHV1-18*01 F			
8	M00161:204:000000000-AWNEY:1:1101:10092:14	Musmus IGHV1-80*01 F			
9	M00161:204:000000000-AWNEY:1:1101:10100:23	Musmus IGHV1-22*01 F			
10	M00161:204:000000000-AWNEY:1:1101:10109:16	Musmus IGHV1-9*01 F			
11	M00161:204:000000000-AWNEY:1:1101:10122:72	Musmus IGHV2-2*01 F			
12	M00161:204:000000000-AWNEY:1:1101:10139:20	Musmus IGHV1-50*01 F			
13	M00161:204:000000000-AWNEY:1:1101:10173:21	Musmus IGHV6-6*01 F			
14	M00161:204:000000000-AWNEY:1:1101:10181:51	Musmus IGHV1-26*01 F			
15	M00161:204:000000000-AWNEY:1:1101:10196:17	Musmus IGHV1-50*01 F			
16	M00161:204:000000000-AWNEY:1:1101:10220:42	Musmus IGHV1-78*01 F			

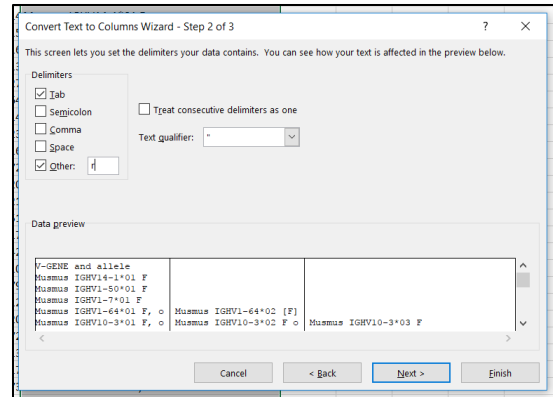
### 4. Assess V Gene Segment usage

- a. Copy the sequence IDs and V Gene segments to a new worksheet
- b. Remove extra characters from the V gene segment columns in this order using the replace tool
  - i. Remove the words “Musmus”
  - ii. Letters = o, r, f, p
  - iii. A single space
  - iv. Symbols = (), [], and a comma
  - v. IgK/IgL only – replace musspr with mussp

	A	B	C	D
1	Sequence ID	V-GENE and allele		
2	M00161:204:000000000-AWNEY:1:1101:10025:14106_1:N:	IGHV14-1*01		
3	M00161:204:000000000-AWNEY:1:1101:10035:15252_1:N:	IGHV1-50*01		
4	M00161:204:000000000-AWNEY:1:1101:10041:16805_1:N:	IGHV1-7*01		
5	M00161:204:000000000-AWNEY:1:1101:10048:13224_1:N:	IGHV1-64*01	IGHV1-64*02	
6	M00161:204:000000000-AWNEY:1:1101:10066:2743_2:N:0	IGHV10-3*01	IGHV10-3*02	IGHV10-3*03
7	M00161:204:000000000-AWNEY:1:1101:10086:6446_2:N:0	IGHV1-18*01		
8	M00161:204:000000000-AWNEY:1:1101:10092:14574_1:N:	IGHV1-80*01		
9	M00161:204:000000000-AWNEY:1:1101:10100:23467_2:N:	IGHV1-22*01		
10	M00161:204:000000000-AWNEY:1:1101:10109:16249_1:N:	IGHV1-9*01		
11	M00161:204:000000000-AWNEY:1:1101:10122:7262_2:N:0	IGHV2-2*01		
12	M00161:204:000000000-AWNEY:1:1101:10139:20232_2:N:	IGHV1-50*01		
13	M00161:204:000000000-AWNEY:1:1101:10173:2119_2:N:0	IGHV6-6*01		
14	M00161:204:000000000-AWNEY:1:1101:10181:5129_2:N:0	IGHV1-26*01		
15	M00161:204:000000000-AWNEY:1:1101:10196:17711_1:N:	IGHV1-50*01		
16	M00161:204:000000000-AWNEY:1:1101:10220:4230_1:N:0	IGHV1-78*01		
17	M00161:204:000000000-AWNEY:1:1101:10283:7934_1:N:0	IGHV14-4*01		
18	M00161:204:000000000-AWNEY:1:1101:10301:12926_2:N:	IGHV1520*01	IGHV1521*01	
19	M00161:204:000000000-AWNEY:1:1101:10328:20861_1:N:	IGHV1-22*01		
20	M00161:204:000000000-AWNEY:1:1101:10346:7237_2:N:0	IGHV1-64*01		
21	M00161:204:000000000-AWNEY:1:1101:10347:13086_1:N:	IGHV1-80*01		
22	M00161:204:000000000-AWNEY:1:1101:10414:17265_2:N:	IGHV3-6*01		
23	M00161:204:000000000-AWNEY:1:1101:10503:7393_1:N:0	IGHV5-2*01		
24	M00161:204:000000000-AWNEY:1:1101:10504:18026_1:N:	IGHV8-5*01		

- c. Select the V Gene column and use the text to column tool outlined in Part V Step 3d
- d. Delineate on the letter r

*Note: Any characters remaining in attached to the V gene assignments will break the VLOOKUP so make sure to do a quick visual check*



- e. Align all of the V gene segments with their corresponding IDs into two columns

*Note: You can do this however you want to, but I*

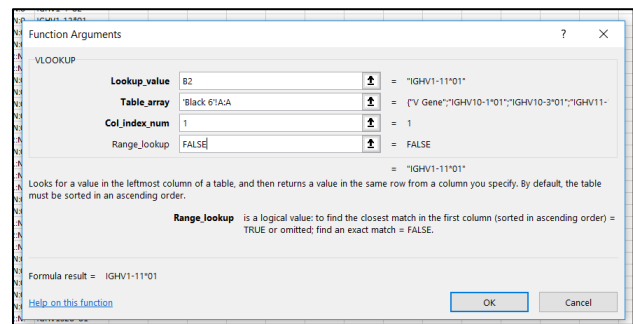
Sequence ID	A	B	C	D
1	MO0161204-000000000-AWNEY-1:1101:10025:14106_1.N:	Musmus IGHV14-1*01 F		
2	MO0161204-000000000-AWNEY-1:1101:10025:15252_1.N:	Musmus IGHV1-50*01 F		
3	MO0161204-000000000-AWNEY-1:1101:10041:15805_1.N:	Musmus IGHV1-7*01 F		
4	MO0161204-000000000-AWNEY-1:1101:10048:13224_1.N:	Musmus IGHV1-64*01 F, o	Musmus IGHV10-3*02 F, o	
5	MO0161204-000000000-AWNEY-1:1101:10066:2743_2.N:0	Musmus IGHV10-3*02 F, o	Musmus IGHV10-3*02 F, o	Musmus IGHV10-3*03 F
6	MO0161204-000000000-AWNEY-1:1101:10086:6446_2.N:0	Musmus IGHV1-18*01 F		
7	MO0161204-000000000-AWNEY-1:1101:10092:14574_1.N:	Musmus IGHV1-80*01 F		
8	MO0161204-000000000-AWNEY-1:1101:10102:23467_2.N:	Musmus IGHV1-22*01 F		
9	MO0161204-000000000-AWNEY-1:1101:10109:16249_1.N:	Musmus IGHV1-9*01 F		
10	MO0161204-000000000-AWNEY-1:1101:10122:7262_2.N:0	Musmus IGHV2-2*01 F		
11	MO0161204-000000000-AWNEY-1:1101:10139:20332_2.N:	Musmus IGHV1-50*01 F		
12	MO0161204-000000000-AWNEY-1:1101:10173:2119_2.N:0	Musmus IGHV6-6*01 F		
13	MO0161204-000000000-AWNEY-1:1101:10181:5129_2.N:0	Musmus IGHV1-26*01 F		
14	MO0161204-000000000-AWNEY-1:1101:10196:17711_1.N:	Musmus IGHV1-50*01 F		
15	MO0161204-000000000-AWNEY-1:1101:10220:4230_1.N:0	Musmus IGHV1-78*01 F		
16	MO0161204-000000000-AWNEY-1:1101:10276:10805_1.N:	Musmus IGHV1-80*01 F		
17	MO0161204-000000000-AWNEY-1:1101:10283:7934_1.N:0	Musmus IGHV14-4*01 F		
18	MO0161204-000000000-AWNEY-1:1101:10301:12026_2.N:	Musmus IGHV1520*05 F, o	Musmus IGHV1521*01 F	
19	MO0161204-000000000-AWNEY-1:1101:10320:20661_1.N:	Musmus IGHV1-22*01 F		
20	MO0161204-000000000-AWNEY-1:1101:10346:7237_3.N:0	Musmus IGHV1-64*01 F		

*have found this method to be by far the least time consuming*

- i. Sort by the last column that contains data (for this example, column N). This will put everything that has data in this column at the top
- ii. Highlight the ENTIRE ROW containing data in this last column
- iii. **Copy** and **insert** these rows right below where the last row containing data in column N is
- iv. In column N highlight all of the cells you just inserted
- v. **Cut** these cells and **paste** them in in column A corresponding to their sequence IDs
- vi. Repeat this process with this segment of cells with the data in M until all of the data is in a single column.
- vii. Once all of the data is in a single column, delete the information in above in columns C through N.
- viii. Restart at the sorting step with data in column M.
- ix. Repeat until all of the V gene data is in a single column.

- f. Perform VLOOKUP for B6 specific genes

- i. Lookup\_Value = Cell containing V gene
- ii. Table\_array = column containing B6 specific V genes
- iii. Col\_index\_num = 1
- iv. Range\_Lookup = False



v. Fill this formula down for all V genes in column

- vi. Copy the sequence IDs and VLOOKUP results to two new columns using the paste values function
- vii. Select the newly pasted sequence ID and VLOOKUP values. Select both columns and use the sort function on the VLOOKUP column and sort A → Z to find results with an #N/A result
- viii. Delete all sequence IDs with an #N/A in their VLOOKUP result

	A	B	C
1	Sequence ID	V-GENE and allele	VLOOKUP
2	M00161.204.000000000-AWNEY:1.2105.20019.13004_1.N	IGHV1-11*01	IGHV1-11*01
3	M00161.204.000000000-AWNEY:1.2106.12794.9768_2.N	IGHV1-50*01	IGHV1-50*01
4	M00161.204.000000000-AWNEY:1.1103.18291.24645_2.N	IGHV1-14*01	IGHV1-14*01
5	M00161.204.000000000-AWNEY:1.2108.16154.2806_1.N	IGHV1-4*02	#N/A
6	M00161.204.000000000-AWNEY:1.1101.24887.2802_2.N	IGHV1-12*01	IGHV1-12*01
7	M00161.204.000000000-AWNEY:1.1101.18081.2184_1.N	IGHV1-53*02	#N/A
8	M00161.204.000000000-AWNEY:1.2116.4066.15509_2.N	IGHV1-19*01	IGHV1-19*01
9	M00161.204.000000000-AWNEY:1.1114.25121.13576_2.N	IGHV1-15*01	#N/A
10	M00161.204.000000000-AWNEY:1.1110.16482.20880_2.N	IGHV1-18*02	#N/A
11	M00161.204.000000000-AWNEY:1.2115.18298.9906_2.N	IGHV2-2-1*01	#N/A
12	M00161.204.000000000-AWNEY:1.2115.14008.3622_2.N	IGHV10-1*01	IGHV10-1*01
13	M00161.204.000000000-AWNEY:1.1110.18805.6662_2.N	IGHV9-12*01	IGHV9-12*01
14	M00161.204.000000000-AWNEY:1.1107.3457.10651_2.N	IGHV5-12-1*01	#N/A
15	M00161.204.000000000-AWNEY:1.1114.7124.10620_1.N	IGHV5-12*03	#N/A
16	M00161.204.000000000-AWNEY:1.2105.19069.14787_2.N	IGHV5-12-2*01	#N/A
17	M00161.204.000000000-AWNEY:1.2118.24460.2658_2.N	IGHV1-82*01	IGHV1-82*01
18	M00161.204.000000000-AWNEY:1.1101.10025.14106_1.N	IGHV14-1*01	IGHV14-1*01
19	M00161.204.000000000-AWNEY:1.1101.10041.16805_1.N	IGHV1-7*01	IGHV1-7*01
20	M00161.204.000000000-AWNEY:1.1101.10048.13224_1.N	IGHV1-64*01	IGHV1-64*01
21	M00161.204.000000000-AWNEY:1.1101.10066.2743_2.N	IGHV10-9*01	IGHV10-9*01
22	M00161.204.000000000-AWNEY:1.1101.10086.6446_2.N	IGHV1-18*01	IGHV1-18*01
23	M00161.204.000000000-AWNEY:1.1101.10092.14574_1.N	IGHV1-80*01	IGHV1-80*01
24	M00161.204.000000000-AWNEY:1.1101.10100.23467_2.N	IGHV1-22*01	IGHV1-22*01
25	M00161.204.000000000-AWNEY:1.1101.10109.16249_1.N	IGHV1-9*01	IGHV1-9*01
26	M00161.204.000000000-AWNEY:1.1101.10122.7262_2.N	IGHV2-2*01	IGHV2-2*01
27	M00161.204.000000000-AWNEY:1.1101.10173.2119_2.N	IGHV6-6*01	IGHV6-6*01
28	M00161.204.000000000-AWNEY:1.1101.10181.5129_2.N	IGHV1-26*01	IGHV1-26*01
29	M00161.204.000000000-AWNEY:1.1101.10220.4250_1.N	IGHV1-78*01	IGHV1-78*01
30	M00161.204.000000000-AWNEY:1.1101.10283.7934_1.N	IGHV14-4*01	IGHV14-4*01
31	M00161.204.000000000-AWNEY:1.1101.10301.12926_2.N	IGHV1520*01	#N/A
32	M00161.204.000000000-AWNEY:1.1101.10414.17265_2.N	IGHV9-6*01	IGHV9-6*01
33	M00161.204.000000000-AWNEY:1.1101.10500.7393_1.N	IGHV5-2*01	IGHV5-2*01
34	M00161.204.000000000-AWNEY:1.1101.10504.18026_1.N	IGHV8-5*01	IGHV8-5*01
35	M00161.204.000000000-AWNEY:1.1101.10546.3050_1.N	IGHV9-2*01	IGHV9-2*01
36	M00161.204.000000000-AWNEY:1.1101.10566.7008_1.N	IGHV5-9-1*02	IGHV5-9-1*02
37	M00161.204.000000000-AWNEY:1.1101.10617.10988_2.N	IGHV6-1*01	IGHV6-1*01

g. Use the COUNTIF function to determine if a single sequence ID has multiple VLOOKUP results

- i. =COUNTIF(E:E,E2)
- E:E = the column of all sequence IDs to search in
  - E2 = the sequence ID being searched for
- ii. Fill down for all sequence IDs (*Note: this may be a time/processing power intensive step*)

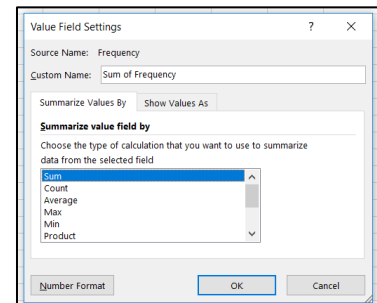
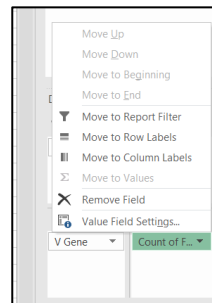
E	F	G
Seq ID	V Gene	COUNTIF
93240	IGHV10-1*01	1
1304	IGHV10-1*01	1
232274	IGHV10-1*01	1
232332	IGHV10-1*01	1
157120	IGHV10-1*01	1
157079	IGHV10-1*01	1
172365	IGHV10-1*01	1
93237	IGHV10-1*01	2
157331	IGHV10-1*01	2
8442	IGHV10-1*01	5
172287	IGHV10-1*01	5
92954	IGHV10-1*01	6
87114	IGHV10-1*01	6
180178	IGHV10-1*01	1
1290	IGHV10-1*01	1
257676	IGHV10-1*01	1
1293	IGHV10-1*01	1
257665	IGHV10-1*01	1
180051	IGHV10-1*01	1
180120	IGHV10-1*01	1
1334	IGHV10-1*01	1
1319	IGHV10-1*01	1
180090	IGHV10-1*01	1
257677	IGHV10-1*01	1
362622	IGHV10-1*01	1

h. Generate two new columns for pivot table construction

- i. Frequency Column
- =1/G2 where G2 = the result of the COUNTIF Function
- ii. V Gene Column
- =F1 where F1 = the V Gene
- iii. Fill down for all sequence IDs (*Note: this may be a time/processing power intensive step*)

E	F	G	H	I	J
Seq ID	V Gene	COUNTIF		Frequency V Gene	
93240	IGHV10-1*01	1		1	IGHV10-1*01
1304	IGHV10-1*01	1		1	IGHV10-1*01
232274	IGHV10-1*01	1		1	IGHV10-1*01
232332	IGHV10-1*01	1		1	IGHV10-1*01
157120	IGHV10-1*01	1		1	IGHV10-1*01
157079	IGHV10-1*01	1		1	IGHV10-1*01
172365	IGHV10-1*01	1		1	IGHV10-1*01
93237	IGHV10-1*01	2		0.5	IGHV10-1*01
157331	IGHV10-1*01	2		0.5	IGHV10-1*01
8442	IGHV10-1*01	5		0.2	IGHV10-1*01
172287	IGHV10-1*01	5		0.2	IGHV10-1*01
92954	IGHV10-1*01	6		0.166667	IGHV10-1*01
87114	IGHV10-1*01	6		0.166667	IGHV10-1*01
180178	IGHV10-1*01	1		1	IGHV10-1*01
1290	IGHV10-1*01	1		1	IGHV10-1*01
257676	IGHV10-1*01	1		1	IGHV10-1*01
1293	IGHV10-1*01	1		1	IGHV10-1*01
257665	IGHV10-1*01	1		1	IGHV10-1*01
180051	IGHV10-1*01	1		1	IGHV10-1*01
180120	IGHV10-1*01	1		1	IGHV10-1*01
1334	IGHV10-1*01	1		1	IGHV10-1*01
1319	IGHV10-1*01	1		1	IGHV10-1*01
180090	IGHV10-1*01	1		1	IGHV10-1*01
257677	IGHV10-1*01	1		1	IGHV10-1*01
362622	IGHV10-1*01	1		1	IGHV10-1*01

- i. Create a pivot table as outlined in step 2b with the frequency and V gene columns
  - i. Rows = V Gene
  - ii. Values = Frequency
  - iii. Filters = Frequency
  - iv. Change the value setting from COUNT of frequency to SUM of frequency by clicking on the down arrow and selecting Value Field Settings
  - v. Select sum



- j. Isolate frequency of 1 and 0.5 V Genes for tabulation
  - i. Click in the down arrow at the top by the frequency filter
  - ii. Check the select multiple items box at the bottom
  - iii. Check only the 0.5 and 1 options

V Gene	Sum of Frequency
IGHV10-1*01	127
IGHV10-3*01	66
IGHV11-1*01	3
IGHV11-11*01	9
IGHV11-2*01	63
IGHV11-12*01	33.5
IGHV11-13*01	3
IGHV11-14*01	0.5
IGHV11-15*01	89
IGHV11-17-1*01	3.5
IGHV11-18*01	385.5
IGHV11-19*01	68
IGHV11-19-1*01	2
IGHV11-20*01	41
IGHV12-1*01	8
IGHV12-21*01	1
IGHV12-21-1*01	9.5
IGHV12-2*01	5
IGHV12-22*01	109.5
IGHV12-3*01	18
IGHV12-23*01	3.5
IGHV12-24*01	1
IGHV12-25*01	5

- k. Copy the two resulting columns out of the pivot table and into the Excel worksheet and calculate the % of Repertoire

- i. Column C  $= (B2/B\$161) * 100$ 
  1. B2 = Freq of the V Gene
  2. B161 = Sum of all Freq
    - a. \$ to hold the cell static

	A	B	C
1	V Gene	Freq	% of Rep
2	IGHV10-1*01	127	1.677675
3	IGHV10-3*01	66	0.871863
4	IGHV11-1*01	3	0.03963
5	IGHV11-11*01	9	0.11889
6	IGHV11-2*01	63	0.832232
7	IGHV11-12*01	33.5	0.442536
8	IGHV11-13*01	3	0.03963
9	IGHV11-14*01	0.5	0.006605
10	IGHV11-15*01	89	1.175694
11	IGHV11-17*01	3.5	0.046235
12	IGHV11-18*01	385.5	5.09247
13	IGHV11-19*01	68	0.898283
14	IGHV11-19*02	2	0.02642
15	IGHV11-20*01	41	0.541612
16	IGHV12-1*01	8	0.10568
17	IGHV12-1*02	1	0.01321
18	IGHV12-1*03	9.5	0.125495
19	IGHV12-2*01	5	0.06605
20	IGHV12-2*02	109.5	1.446499
21	IGHV12-3*01	18	0.237781
22	IGHV12-3*02	3.5	0.046235
23	IGHV12-4*01	1	0.01321
24	IGHV12-5*01	5	0.06605
25	IGHV12-6*01	308	4.068692
26	IGHV12-7*01	1	0.01321
27	IGHV12-8*01	17.5	0.231176
28	IGHV13-1*01	1	0.01321
29	IGHV13-1*02	9	0.11889

## 5. Assess J Gene Segment usage

- a. Proceed as with step 4 with two minor alterations
  - i. Replace all cells containing “Less than six nucleotides aligned P or F” to Undetermined
  - ii. Only filter for J gene segments that have a frequency of 1

## 6. Assess CDR3 Lengths

- a. Delineate the “CDR-IMGT lengths” column on “.” to separate. The last number is the CDR3 length
- b. Delete the first two columns
- c. Generate a pivot table as outlined previously in Step 2 using the CDR3 Length column

## IGH ONLY STEPS – light chain proceed to Save Point

## 7. Assess D Gene Segment usage

- a. Copy the Sequence IDs and D Gene column into a new tab
- b. Clean the D gene segments as in step 4d
- c. Sort both columns by the D Gene column
- d. Any sequence ID that does not have a D gene identified label as “Undetermined”
- e. Perform the VLOOKUP as outlined in step 4g with the D gene segments
- f. Copy and paste the results into two new columns as paste values
- g. Find and replace any #N/A cells with “Undetermined”

O	P	Q
Row Labels	Count of D Gene	
IGHD1-1*01	594	
IGHD2-3*01	167	
IGHD2-4*01	160	
IGHD2-5*01	145	
IGHD3-1*01	55	
IGHD3-2*02	51	
IGHD4-1*01	144	
IGHD5-1*01	8	
IGHD5-5*01	10	
IGHD6-2*02	5	
UNDETERMINED	697	
(blank)		
<b>Grand Total</b>	<b>2036</b>	



- h. Generate pivot table as in step 4i using only the D gene column
- i. Calculate the % Repertoire as outlined in Step 4k

**8. Assess Constant Region Usage**

- a. Generate a pivot table as outlined previously in Step 2 using the Constant Region column

**SAVE POINT** – Save the data in small enough work books to be manageable and still contain the data required for processing. Typically we save V, D, J, Constant, and CDR3 length in the same workbook.

## **Sub Appendix A – Programs Required**

- CLC Genomics Workbench
  - <https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/>
- Notepad++
  - <https://notepad-plus-plus.org/>
- Microsoft Excel
  - Windows based program
- PeaZip
  - <http://www.peazip.org/>

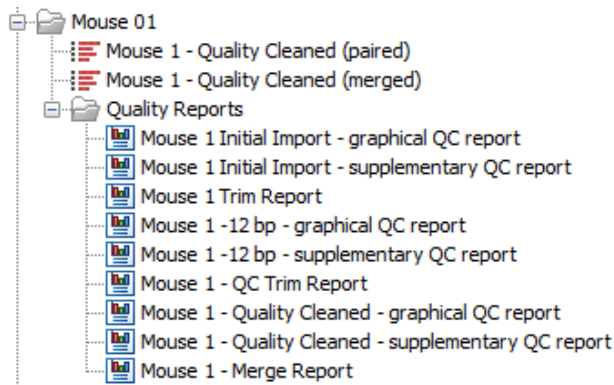
## Sub Appendix 2 - List of Resources

- CLC User Manual
  - [http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/753/index.php?manual=Introduction\\_CLC\\_Genomics\\_Workbench.html](http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/753/index.php?manual=Introduction_CLC_Genomics_Workbench.html)
- Notepad++ User Manual
  - [http://docs.notepad-plus-plus.org/index.php/Main\\_Page](http://docs.notepad-plus-plus.org/index.php/Main_Page)
- Immunogenetics
  - <http://www.imgt.org/>
- NCBI
  - IgH Locus - <https://www.ncbi.nlm.nih.gov/gene/111507>
  - IgK Locus - <https://www.ncbi.nlm.nih.gov/gene/243469>
  - IgL Locus - <https://www.ncbi.nlm.nih.gov/gene/111519>

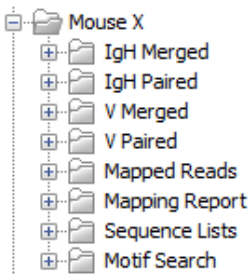
## Sub Appendix 3 – Current Naming and Sorting Conventions for CLC

File Generation	Naming Convention
First file imported into CLC for cleaning (will be deleted)	Initial Import
Initial Import minus 5' 12bp	-12
-12 after quality cleaning	Mouse X – Quality Controlled (paired)
Quality Controlled file after merging	Mouse X – Quality Controlled (merged)
Mapping results and reports	IgH/IgK/IgL Paired, IgH/IgK/IgL Merged, V Paired, V Merged
Combined list of extracted mapping reads	Mouse X in folder “sequence lists”
Broken down lists of mapped reads to IMGT Submission	Mouse X – pt 1

Example of final folder after completing cleaning reads in CLC



Example of folder set up for CLC processing after cleaning for IgH



## **Sub Appendix 4 – Current Naming Conventions for IMGT Processing**

### **Folders**

1 – IMGT Output	The zipped files downloaded directly from IMGT
2 – IMGT Raw Data	The combination of all productive and unknown reads with all of the IMGT Raw Data
3 – For Sort	The adjusted IMGT files with constant regions attached (IgH only) and Trunc ID column
4 – Dups Removed	The results after duplicates have been removed
5 – Final	The results after duplicate removal and functionality assessments (This is the file that will be used for analysis)

## Sub Appendix 5 – Important IMGT Links

- IMGT HighV-Quest
  - <http://www.imgt.org/HighV-QUEST/login.action>
- Heavy Chain V Gene Segments by Strain
  - [http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus\\_musculus&group=IGHV\\_strains](http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus_musculus&group=IGHV_strains)
- Kappa Chain V Gene Segments by Strain
  - [http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus\\_musculus&group=IGKV\\_strains](http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus_musculus&group=IGKV_strains)
- Lambda Chain V Gene Segments by Strain
  - [http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus\\_musculus&group=IGLV](http://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=Mus_musculus&group=IGLV)
- Base for identifying gene segments
  - Contains information on chromosome location, strain specificity, etc.
  - <http://www.imgt.org/IMGTrepertoire/LocusGenes/>

## Sub Appendix 6 – Data Collected from IMGT Output

<b>Output 1: Use all EXCEPT F-I, K-M, 0-R, V, X, Z, AD (<u>delete</u> the following)</b>	<b>Output 6: Use columns J-AD (<u>include</u> the following)</b>
(F) V-REGION identity %	(J) 3'V-REGION
(G) V-REGION identity nt	(K) P3'V
(H) V-REGION identity % (with ins/del events)	(L) N-REGION
(I) V-REGION identity nt (with ins/del events)	(M) N1-REGION
(K) J-REGION score	(N) P5'D
(L) J-REGION identity %	(O) D-REGION
(M) J-REGION identity nt	(P) P3'D
(O) D-REGION reading frame	(Q) P5'D1
(P) CDR1-IMGT length	(R) D1-REGION
(Q) CDR2-IMGT length	(S) P3'D1
(R) CDR3-IMGT length	(T) N2-REGION
(V) JUNCTION frame	(U) P5'D2
(X) Functionality comment	(V) D2-REGION
(Z) J-GENE and allele comment	(W) P3'D2
(AD) Deleted n nt	(X) N3-REGION
	(Y) P5'D3
<b>Output 2: Use columns J-R (<u>include</u> the following)</b>	(Z) D3-REGION
(J) FR1-IMGT	(AA) P3'D3
(K) CDR1-IMGT	(AB) N4-REGION
(L) FR2-IMGT	(AC) P5'J
(M) CDR2-IMGT	(AD) 5'J-REGION
(N) FR3-IMGT	

(O) CDR3-IMGT	<b>Output 7: Use columns F-K (<u>include</u> the following)</b>
(P) JUNCTION	(F) FR1-IMGT
(Q) J-REGION	(G) CDR1-IMGT
(R) FR4-IMGT	(H) FR2-IMGT
	(I) CDR2-IMGT
<b>Output 3: Not used</b>	(J) FR3-IMGT
	(K) CDR3-IMGT
<b>Output 4: Not used</b>	
	<b>Output 8: Not Used</b>
<b>Output 5: Not used</b>	
	<b>Output 10: Not used</b>
	<b>Output 11: Not used</b>



## Sub Appendix 7 – Checklists for Data Processing

### Part I

Task	Initial	Date
Import sequences from Illumina Read file		
Generate sequencing report		
Remove 12 5' bps		
Generate sequencing report of new file		
Quality Trim/Adaptor Trim/Remove Sequences		
Generate sequencing report of new file		
Merge overlapping reads		

### Part II

Task	Initial	Date
Map paired reads to V gene sequences		
Map merged reads to V gene sequences		
Map paired reads to locus		
Map merged reads to locus		
Repeat mapping as needed (IgH, IgK, IgL)		
Extract mapped sequences from each mapping and combine in new list		
Break new list into parts <500k sequences		
Export list parts as FASTA files		

### Part III

Task	Initial	Date
Submit files to IMGT High-V Quest		
Download Files from IMGT		

## Part IV

Task	Initial	Date
<b>Create FASTA Tags in Excel</b>		
Copy sequence IDs and sequences of all productive (any type) sequences		
<b>Generate a New Worksheet</b>		
Column B = Sequence IDS		
Column D = RNA Sequence		
In column A fill down the > symbol		
In column C enter the equation =A1&B1 and fill down		
<b>Finalize FASTA Formatting</b>		
Copy columns C and D to Notepad++		
Replace the "Tab" break between the sequence and the name with \r		
Save file as .fasta		
<b>Motif Search</b>		
Import Fasta to CLC		
Run Motif search		
Save results at excel 2010 file		

## Part V

Task	Initial	Date
<b>Remove Duplicated Reads</b>		
Combine all productive and unknown functionality reads from all IMGT files		
Set new sequence numbers		
<b>SAVE POINT - NEW FILE</b>		
Create truncated ID column		
Remove " (See Comment)" throughout the worksheet		
Align the Motif Search results (IGH Only)		
<b>CONFIRM FORMULAS HAVE COPIED CORRECTLY AND ALL ADDITIONAL CHARCTERS AND SPACES ARE REMOVED</b>		
<b>SAVE POINT - NEW FILE</b>		
Sort sequences		
· Truncate (A -> Z)		
· Constant Region (A -> Z) (IGH Only)		
· Functionality (A -> Z)		
· V Gene Score (Large -> Small)		
· Orientation (Z -> A)		
Remove Duplicates		
<b>SAVE POINT - NEW FILE</b>		
Repeat Duplicate Removal from last save point to confirm removed duplicates correct		
<b>IF REMOVAL MATCHES, CONTINUE BELOW, IF NOT, DETERMINE ERROR</b>		
<b>Adjust Functionality</b>		
Copy CDR3 AA sequences and their sequence IDs from all productive functionality sequences to new page		
Copy Constant region associated with CDR3s to the new page (IGH Only)		

Remove (See Functionality Comment) from CDR3s		
Create one column with the first letter in the CDR3		
Create one column with the last letter in the CDR3		
Use the Conditional Formatting to isolate CDR3s not matching the C-xx-F Motif (IgK Only)		
Use the Conditional Formatting to isolate CDR3s not matching the C-xx-W Motif (IgH Only)		
Create "Adjusted Functionality" Column		
Any CDR3 matching the C-xx-F Motif is termed "Productive" (IgK Only)		
Any CDR3 not matching the C-xx-F Motif is termed "Unknown" (IgK Only)		
Any CDR3 matching the C-xx-W Motif is termed "Productive" (IgH Only)		
Any CDR3 not matching the C-xx-W motif of IgA, IgE, or IgG constant region is termed "Productive"		
Any CDR3 not matching the C-xx-W motif of IgD or IgM constant region or NO constant region is termed "Unknown"		
Use the Vlookup to reassociate the new "Adjusted Functionality" column back into the main spreadsheet		
Delete old "Functionality" column		
<b>SAVE POINT - NEW FILE (Use this file for all following steps)</b>		

## Part IV

Task	Initial	Date
<b>Determine Productive and Unknown Read Numbers</b>		
Create a pivot table of the Adjusted Functionality Column		
<b>Total Productive and Unknown Numbers Collected Here</b>		
Import current B6 gene list		
· Confirm gene segments are in A -> Z order		
<b>V Gene Segment Usage</b>		
Arrange all productive and unknown V gene segments into single column with Sequence ID		
Perform VLookup for B6 V gene segments		
Create Pivot table of V gene segments		
Weight V gene usage with 0.5 (2 possibilities) or 1 (1 possibility) only		
Calculate % of Repertoire		
<b>J Gene Segment Usage</b>		
Arrange all productive J genes into a single column with Sequence ID		
Convert all <6 nucleotide P or F statements to "undetermined"		
Perform VLookup for B6 J Genes		
Create pivot table of J gene segments		
Calculate % of Repertoire		
<b>Calculate CDR3 Lengths</b>		
Deliniate the CDR3 Length column on "."		
Create pivot table		
Calculate % of repertoire		
<b>D Gene Segment Usage (IgH Only)</b>		
Arrange all productive D genes into single column with Sequence No		
Label all empty cells "undetermined"		
Perform VLookup for B6 D gene		

Label all N/A results “undetermined”		
Create Pivot Table of D Gene Segments		
Calculate D gene % of Repertoire		
<b>Constant Region Usage (IgH Only)</b>		
Arrange all productive constant regions into a single column		
Create pivot table		
Calculate % of repertoire		

## Sub Appendix 8 – Functional Heavy B6 Gene Segments (IMGT)

V10-1*01	V1-42*01	V1-72*01	V2-6*01	V6-5*01	D1-1*01
V10-3*01	V14-3*01	V1-74*01	V2-6-8*01	V6-6*01	D2-3*01
V11-1*01	V1-43*01	V1-75*01	V2-7*01	V6-7*01	D2-4*01
V1-11*01	V14-4*01	V1-76*01	V2-9*01	V7-2*01	D2-5*01
V11-2*01	V1-47*01	V1-77*01	V2-9-1*01	V7-3*01	D2-6*01
V1-12*01	V1-49*01	V1-78*01	V3-1*01	V7-4*01	D2-7*01
V1-14*01	V1-5*01	V1-80*01	V3-3*01	V8-11*01	D2-8*01
V1-15*01	V1-50*01	V1-81*01	V3-4*01	V8-12*01	D3-1*01
V1-16*01	V15-2*01	V1-82*01	V3-5*01	V8-13*01	D3-2*02
V1-17-1*01	V1-52*01	V1-84*01	V3-6*01	V8-2*01	D4-1*01
V1-18*01	V1-53*01	V1-85*01	V3-8*01	V8-4*01	D5-2*01
V1-19*01	V1-54*01	V1-9*01	V3S7*01	V8-5*01	D5-3*01
V1-20*01	V1-55*01	V1S100*01	V4-1*01	V8-6*01	D5-4*01
V1-22*01	V1-56*01	V1S103*01	V5-12*01	V8-8*01	D5-5*01
V12-3*01	V1-58*01	V1S107*01	V5-12-4*01	V8-8-1*01	D5-6*01
V1-23*01	V1-59*01	V1S108*01	V5-15*01	V8-9*01	D6-1*01
V1-24*01	V16-1*01	V1S5*01	V5-16*01	V8S9*01	D6-2*01
V1-26*01	V1-61*01	V1S65*01	V5-17*01	V9-1*01	D6-3*01
V13-1*01	V1-62-1*01	V1S67*01	V5-2*01	V9-2*01	D6-4*01
V1-31*01	V1-62-2*01	V1S68*01	V5-4*01	V9-3*01	
V13-2*01	V1-62-3*01	V1S87*01	V5-6*01	V9-4*01	
V1-34*01	V1-63*01	V1S92*01	V5-9*01		
V1-36*01	V1-64*01	V1S95*01	V5-9*04		
V1-37*01	V1-66*01	V1S96*01	V5-9-1*02		
V1-39*01	V1-67*01	V2-2*01	V5S21*01		J1*03
V1-4*01	V1-69*01	V2-3*01	V5S24*01		J2*01
V14-1*01	V1-7*01	V2-4*01	V6-3*01		J3*01
V14-2*01	V1-71*01	V2-5*01	V6-4*01		J4*01

## Sub Appendix 9 – Functional Heavy B6 V-Gene Segments in Chromosomal

### Order (NCBI Order)

<b>5' End</b>	V1-59	V1-18	V3-4	V2-6-8
V1-85	V1-58	V1-17-1	V7-4	V2-9-1
V1-84	V8-8	V1-15	V3-3	V5-12-4
V1-82	V1-56	V1-14	V14-4	V5-9-1
V1-81	V1-55	V1-12	V7-3	V2-6
V1-80	V1-54	V1-11	V9-3	V5-12
V1-78	V8-6	V1-9	V9-2	V2-5
V1-77	V1-53	V15-2	V9-1	V5-9
V1-76	V1-52	V1-7	V16-1	V2-4
V1-75	V1-50	V10-3	V14-3	V5-6
V1-74	V8-5	V1-5	V11-12	V2-3
V1-72	V1-49	V1-4	V3-2	V5-4
V1-71	V8-4	V10-1	V4-2	V2-2
V8-12	V1-47	V6-7	V14-2	V5-2
V1-69	V1-43	V6-6	V11-1	<b>3' End</b>
V1-67	V1-42	V6-5	V3-1	
V1-66	V1-39	V6-4	V4-1	
V8-11	V1-37	V6-3	V14-1	
V1-64	V1-36	V12-3	V7-2	
V1-63	V1-34	V13-2	V7-1	
V8-9	V1-31	V3-8	V2-9	
V1-62-3	V1-26	V9-4	V5-17	
V1-62-2	V1-22	V3-6	V5-16	
V1-62-1	V1-20	V13-1	V5-15	
V1-61	V1-19	V3-5	V2-7	



## Sub Appendix 10 – Functional Kappa B6 Gene Segments in Chromosomal

### Order

<b>5' End</b>	V10-95	V4-58	V8-24
V2-137	V10-94	V4-57-1	V6-23
V1-135	V19-93	V4-57	V8-21
V1-133	V4-92	V4-55	V6-20
V1-132	V4-91	V4-53	V8-19
V14-130	V4-90	V4-51	V6-17
V17-127	V12-89	V4-50	V8-16
V14-126	V1-88	V5-48	V6-15
V11-125	V4-86	V12-46	V6-14
V9-124	V13-85	V5-45	V6-13
V9-123	V13-84	V12-44	V3-12
V1-122	V4-81	V5-43	V3-10
V17-121	V4-80	V12-41	V3-9
V9-120	V4-79	V5-39	V3-7
V1-117	V4-78	V12-38	V3-5
V2-112	V4-74	V5-37	V3-4
V14-111	V4-73	V18-36	V3-3
V1-110	V4-72	V18-34	V3-2
V2-109	V4-71	V7-33	V3-1
V16-104	V4-70	V6-32	<b>3' End</b>
V20-101-2	V4-69	V8-30	
V14-100	V4-68	V6-29	
V1-99	V4-63	V8-28	
V12-98	V4-61	V8-27	
V10-96	V4-59	V6-25	

## **Sub Appendix 11 – Functional Lambda B6 Gene Segments in Chromosomal**

### **Order**

5' End

V2\*02

V3-02

V1\*01

J2

J4


J3

J1


3' End


# Appendix C - Copyright Releases


## Chapter 3 Copyright Release


From Jamie Foster <jfoster@ufl.edu> 


Subject **Re: Copyright Release Form for Doctoral Dissertation**


To Trisha Rettig 


 Reply

 Forward

 Archive

 Junk

 Delete

 More

3/14/2018 2:34 P

Hi Trisha

Just heard back from the old editor and since this is for a non-commercial purposes no copy right form is needed.

so happy dissertation publishing

Cheers  
Jamie

On Mar 14, 2018, at 2:12 PM, Trisha Rettig <[trettig@ksu.edu](mailto:trettig@ksu.edu)> wrote:

Hello Dr. Foster,

I am following up on the message below as a copyright release is required for graduation.

Please let me know if you have any questions or concerns.

Thank you,

Trisha

----- Forwarded Message -----

**Subject:**Copyright Release Form for Doctoral Dissertation  
**Date:**Tue, 6 Mar 2018 19:47:14 -0600  
**From:**Trisha Rettig <[trettig@ksu.edu](mailto:trettig@ksu.edu)>  
**To:**[jfoster@ufl.edu](mailto:jfoster@ufl.edu)

Dear Dr. Foster,

I am a graduate student in the laboratory of Stephen K. Chapes at Kansas State University. I am writing to request a copyright released to include the published article entitled "Validation of Methods to Assess the Immunoglobulin Gene Repertoire in Tissues Obtained from Mice on the International Space Station" by Rettig, TA et al. as a chapter in my Ph.D. dissertation.

I have included a more complete letter and a generic copyright release form as an attachment.

Thank you in advance, and please let me know if you have any questions.

Thank you,

Trisha

--  
Trisha Rettig  
Graduate Teaching Assistant | PhD Candidate  
Kansas State University | 242 Chalmers

<Copyright Release - GSR.docx>

## Chapter 4 Copyright Release

From plosone <plosone@plos.org> ☆ Reply Forward Archive Junk Delete More ▾

Subject **FW: Copyright Release Form for Doctoral Dissertation [ ref\_00DU0Ifis\_5000BhjmSh:ref ]** 3/9/2018 6:10 PM

To Trisha Rettig ☆

Dear Dr Rettig,

Thank you for your message. PLOS ONE publishes all of the content in the articles under an open access license called "CC-BY." This license allows you to download, reuse, reprint, modify, distribute, and/or copy articles or images in PLOS journals, so long as the original creators are credited (e.g., including the article's citation and/or the image credit). Additional permissions are not required. You can read about our open access license here: <http://journals.plos.org/plosone/s/licenses-and-copyright>

There are many ways to access our content, including HTML, XML, and PDF versions of each article. Higher resolution versions of figures can be downloaded directly from the article.

Thank you for your interest in PLOS ONE and for your continued support of the Open Access model. Please do not hesitate to be in touch with any additional questions.

Kind regards,

Guzyal Gabitova

PLOS | OPEN FOR DISCOVERY  
Guzyal Gabitova | Publications Assistant  
1160 Battery Street, Suite 225, San Francisco, CA 94111  
[ggabitova@plos.org](mailto:ggabitova@plos.org)  
[plos.org](http://plos.org) | [Facebook](#) | [Twitter](#) | [Blog](#)

Case Number: 05668870

----- Original Message -----  
From: Joerg Heber [[jheber@plos.org](mailto:jheber@plos.org)]  
Sent: 3/6/2018  
To: [plosone@plos.org](mailto:plosone@plos.org)  
Subject: FW: Copyright Release Form for Doctoral Dissertation

From: Trisha Rettig [[trettig@ksu.edu](mailto:trettig@ksu.edu)]  
Date: Tuesday, March 6, 2018 at 17:48  
To: Joerg Heber [[jheber@plos.org](mailto:jheber@plos.org)]  
Subject: Copyright Release Form for Doctoral Dissertation

Dear Dr. Heber,  
I am a graduate student in the laboratory of Stephen K. Chapes at Kansas State University. I am writing to request a copyright released to include the published article entitled "Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing" by Rettig, T.A. et al. as a chapter in my Ph.D. dissertation.  
I have included a more complete letter and a generic copyright release form as an attachment.  
Thank you in advance, and please let me know if you have any questions.  
Thank you,  
Trisha



--



Trisha Rettig

Graduate Teaching Assistant | PhD Candidate

Kansas State University | 242 Chalmers  
ref:\_00DU0Ifis\_5000BhjmSh:ref

## Chapter 5 Copyright Release

From: Tom Hei <tkh1@cumc.columbia.edu>   
Subject: **Re: Fwd: Copyright Release Form for Doctoral Dissertation**  
To: Trisha Rettig 

 Reply  Forward  Archive  Junk  Delete  More

3/15/2018 10:19 A

Dear Trisha: Just heard from our Journal Manager:

" Elsevier authors retain the right to include their work within their dissertation or thesis (so long as it is not published commercially), along with certain other personal uses, without the formal need to seek permission from the publisher. Elsevier only requires that authors include a full acknowledgement and, if appropriate, a link to the final published version of the paper hosted on Science Direct. For further information, the author can refer to the Journal Author Rights section of the copyright policy page of the Elsevier website (<https://www.elsevier.com/about/our-business/policies/copyright>)."

I hope that this is helpful. Once again, congratulations on the completion of your thesis work.

Tom Hei



COLUMBIA UNIVERSITY  
MEDICAL CENTER

Tom K. Hei, Ph.D.,  
Professor and Vice-Chairman of Radiation Oncology  
Associate Director, Center for Radiological Research  
Professor of Environmental Health Sciences  
630 West 168th Street, VC11-205  
New York, NY 10032  
Phone: (212)-305-8462, Fax: (212)-305-3229  
Email: [tkh1@cumc.columbia.edu](mailto:tkh1@cumc.columbia.edu)  
<http://www.crr-cu.org/hei.htm>

On 3/15/2018 10:50 AM, Tom Hei wrote:

Dear Trisha: My sincere apology for the delay in getting back to you as I was at the NCI for an intramural program review for the past few days and had just been back last night. I have forwarded your request to our Publisher in Amsterdam for her to sign off. I do not believe that there will be any problem with the request.

My heart felt congratulations to your upcoming graduation, Dr. Trisha Rettig. Well done !

Regards, Tom Hei

On 3/14/2018 2:11 PM, Trisha Rettig wrote:

Hello Dr. Hei,

I am following up on the following copyright release form, as I am need of this for graduation requirements.

Please let me know if you have any questions or concerns.

Thank you,